

## **ESTIMACIÓN DE LA PRECISIÓN DE UN PROGRAMA EDUCATIVO MEDIANTE LA TEORÍA DE LA GENERALIZABILIDAD**

**Natali Larraz<sup>1</sup>**

*Universidad de Zaragoza, España*

**Pedro Allueva**

*Universidad de Zaragoza, España*

**Ángel Blanco-Villaseñor**

*Universidad de Barcelona, España*

### **Resumen**

El objetivo de este estudio es presentar un modelo de evaluación de la eficiencia y probar la precisión de un programa de intervención educativa para el desarrollo de las habilidades metacognitivas. Para comprobar la validez de la estructura del diseño y con el fin de poder generalizar los resultados, se ha aplicado la Teoría de la Generalizabilidad. Esta teoría asume otras fuentes de variación además de las debidas a las diferencias individuales y a la intervención, como el número de ejercicios, el número de personas y el número de grupos. Tras su aplicación hemos podido demostrar que el diseño del programa es adecuado a pesar de que los resultados obtenidos no fueron significativos y se demuestra cómo, con ese mismo diseño, pero ampliando el número de ejercicios obtendremos resultados significativos.

*Palabrasclave:* Teoría de la Generalizabilidad; Validez; Evaluación de programas; Variabilidad

### **Accuracy estimation in the evaluation of an educational intervention program according to the Generalizability Theory Abstract**

The aim of this study is to present a model for evaluating the efficiency and the accuracy of an intervention educative program. To prove the validity of the design and the structure of the study in order to generalize its results it has been applied the Generalizability Theory. This theory assumes other sources of variation to assess its internal validity in a study, as well as due to individual differences and intervention, as the number of exercises, the number of participants and number of groups of the study. After its application, the results obtained shows that if the number of exercises to develop metacognitive skills have been increased, the effects of the program would have been significant between the two study groups.

*Key words:* Generalizability Theory; Validity, Program Evaluation; Variability

---

<sup>1</sup> Correspondence about this article should be addressed to Natalia Larraz, email: nlarraz@unizar.es.

El ámbito de la evaluación de programas, suele estar unido generalmente al ámbito de las ciencias sociales. Los programas de intervención social, suelen aplicarse con el fin de obtener un beneficio, resolver un problema y comprobar la eficacia de un método innovador, de un tratamiento o una intervención. En este sentido, se deben obtener unos criterios de evaluación que garanticen la eficacia y la eficiencia del programa implementado para definir en qué medida el programa es útil, eficaz, válido y fiable. Es decir, para observar en qué medida ha alcanzado los objetivos del programa y se pueden generalizar los efectos que persigue, se deben seguir unos criterios de evaluación que garanticen su eficacia, validez y fiabilidad.

Los programas que se aplican en el ámbito de las ciencias sociales, generalmente utilizan los diseños cuasi-experimentales con el fin de acercarse a una realidad social que impide generalmente la posibilidad de utilizar las condiciones que se aplican en los diseños experimentales, ya que parten de la necesidad observada de diseñar un modelo de intervención en el que se trabaja con grupos no equivalentes encontrados en la investigación social. Los diseños cuasi-experimentales, se utilizan para referirse a los diseños de investigación debidamente controlados pero que carecen de alguna de las características esenciales de un diseño experimental, que son la aleatorización (asignación al azar de los participantes a las diferentes condiciones del estudio), la manipulación (producción deliberada y artificial de la variable independiente) y el control (exclusión de factores que influyen en los resultados) (Arnaú, 1990).

Dentro del ámbito de la evaluación de programas, el diseño se define como “la estrategia integral del proceso y, en consecuencia, el curso de acción o secuencia de decisiones acerca de cómo recoger, ordenar y analizar los datos” (Anguera, 1995, p. 149) y está subordinado al cumplimiento de los fines del programa.

En la evaluación de programas, se ha de valorar la eficacia y la eficiencia del mismo (Máiquez, Blanco-Villaseñor y Capote, 2000). Cuando se evalúa su eficacia, se analiza si el programa ha logrado los objetivos y si se han producido los cambios esperados en relación a los objetivos, metodología y contenidos del mismo. Si se evalúa la eficiencia del programa, se pretende averiguar si se pueden generalizar algunos de los parámetros estructurales del programa para mejorar su aplicabilidad en futuras ocasiones. De este modo se pueden analizar los recursos empleados, la estructura del programa con el fin de adaptar y mejorar su aplicación para futuras intervenciones.

En este estudio se pretende evaluar la eficiencia de un programa para el desarrollo de las habilidades metacognitivas en estudiantes de

segundo curso de Educación Secundaria Obligatoria de un centro de Educación Infantil, Primaria y Secundaria de la ciudad de Zaragoza (España) llevado a cabo mediante un diseño cuasi-experimental con grupo control no equivalente y medidas antes y después de la intervención (Larraz, 2011).

El objetivo de evaluar la eficiencia del programa es observar en qué medida se pueden generalizar sus resultados si modificamos algunos de sus parámetros con el fin de mejorar su aplicabilidad para futuras ocasiones. Para ello, se ha empleado la Teoría de la Generalizabilidad (TG) desarrollada por Cronbach y colaboradores y que ha sido aplicada en diferentes estudios previos (Blanco-Villaseñor, 1989, 1993; Arnaú, Blanco-Villaseñor y Losada, 1991; Máiquez, Blanco-Villaseñor y Capote, 2000; Blanco-Villaseñor, Sastre y Escolano, 2010) con el fin de analizar y evaluar distintos aspectos de los mismos. Dichos estudios, han utilizado la base de la TG y han demostrado ser útiles para poder ampliar investigaciones a gran escala (Máiquez, Blanco-Villaseñor y Capote, 2000) o para estimar la precisión de los diseños observacionales en diferentes ámbitos, como el desarrollo infantil (Blanco-Villaseñor, Sastre y Escolano, 2010) o el deporte (Blanco-Villaseñor y Anguera, 2000; Anguera, Blanco-Villaseñor, Hernández Mendo y Losada, 2011; Usabiaga, Castellano, Blanco-Villaseñor y Casamichana, 2013), entre otros.

La generalizabilidad se puede definir como “el grado por el cual podemos generalizar un resultado en unas condiciones particulares a un valor teórico buscado” (Blanco-Villaseñor, 1993, p. 189). La TG es una teoría de los errores multifaceta de una medición conductual y permite averiguar la precisión o fiabilidad de la misma con el fin de generalizar los resultados obtenidos de las observaciones reales, por lo que pretende averiguar la fiabilidad, la validez y la precisión de una medida. “La esencia de esta teoría es el postulado de que en cualquier situación de medida existen múltiples (de hecho infinitas) fuentes de variación (denominadas facetas en dicha teoría)” (Blanco-Villaseñor, 1993, p. 155). En este sentido, se puede afirmar que existen otras fuentes de error, además de las analizadas en el estudio, que pueden afectar a la estimación de la generalizabilidad de los datos registrados mediante la observación sistemática, ya que el muestreo tiene lugar en diferentes etapas (sesiones, días, momentos, técnicas e instrumentos, etc.) y requiere de la estimación de los componentes de varianza, que pueden ser combinados para producir una o más estimaciones de coeficientes de generalizabilidad (Blanco-Villaseñor, 2001).

Por tanto, la TG trata de identificar el error de medida asociado a los componentes de varianza en función de otras facetas, además de las ya

analizadas como las diferencias individuales y la intervención, que son las debidas a las características del programa (número de ejercicios) y de los participantes (género y número de participantes). Con este fin, se pueden implementar estrategias que reduzcan estas fuentes de error sobre la medida. Las facetas que afectan a la medida en este estudio, pueden ser: a) el número de ejercicios fue seis, dicho parámetro equivale al número de sesiones, dado que en cada sesión se implementó un ejercicio; b) el género de los participantes, femenino y masculino; c) el número de participantes; así como el efecto de interacción de: a) ejercicios y participantes; b) participantes y género; c) ejercicios y género; d) participantes, ejercicios y género. Así, se puede determinar el modo que afecta a los resultados la modificación de dichas facetas. En este estudio las facetas que se estudian, en relación a la variabilidad, son las tres mencionadas anteriormente. Se han desestimado facetas como los instrumentos de evaluación, ya que se utilizó un mismo instrumento de test-retest de pruebas paralelas, y los momentos de intervención que consistieron en una sesión semanal. Esta variable no da lugar a ser incluida como una faceta de estudio ya que haría mención al número de ejercicios que equivale al número de sesiones implementadas.

Una de las fases esenciales de la TG que nunca fueron tratadas por Cronbach, Gleser, Nada y Rajaratnam (1972) fueron los planes o diseños de medida y que sí han sido tratadas con un notable interés por Cardinet, Johnson y Pini (2010), que suponen una extensión del trabajo de Cardinet en los últimos 25 años en lo que respecta a la intercambiabilidad de los diseños o planes de medida.

Este concepto, según Cardinet, Johnson y Pini (2010), ha sido denominado con el término de

simetría y permite que las facetas sean intercambiables, en algunos casos será el objeto de medida y en otros el error de medida, y así poder determinar en cada caso la magnitud de la variabilidad atribuible a cada posible fuente de variabilidad.

En relación al programa objeto de estudio, se ha observado en sus resultados un desarrollo significativo de la metacognición en el grupo experimental ( $p = .004 < .05$ ). Sin embargo, este desarrollo no es significativo con respecto al grupo control ( $p = .776 > .05$ ). El objetivo del programa consistió en desarrollar las habilidades metacognitivas de un grupo de estudiantes, y las variables objeto de estudio fueron el desarrollo de la metacognición como variable dependiente, y el programa de intervención como variable independiente. La variable de género se contempló como una variable independiente asignada, la cual no presentó diferencias significativas en los resultados del análisis realizado. El análisis de los datos se realizó mediante la prueba no paramétrica U de Mann-Whitney en la que se compararon las diferencias ente las ganancias de las puntuaciones pretest-postest en ambos grupos de estudio (ver tabla 1). Dado que con este programa no se logró un desarrollo significativo ( $p < .05$ ) de las habilidades metacognitivas, nos parece interesante comprobar si aumentando o modificando alguno de los parámetros del estudio (ejercicios, número de participantes y género), los resultados en relación al desarrollo de las habilidades metacognitivas pueden ser significativos. Para realizar dicha comprobación aplicamos la TG. En la tabla 1 se muestran los resultados obtenidos, con el programa de intervención, en el desarrollo de habilidades metacognitivas.

Tabla 1. *Resultados del desarrollo de habilidades metacognitivas obtenidas con el programa de intervención*

Grupo	Diferencia	Sig. Dif PF-PI		Diferencia	Sig. Dif. E-C	
	Intragrupal (PF-PI)	W de Wilcoxon	P	Intergrupal (E-C)	U Mann-Whitney	P
	M±SD		Sig.	M±SD		Sig.
Experimental (E)	6,4 ± 1,8		S	1,1 ± 2,5		NS
Control (C)	5,3 ± 1,7		S			

Nota: M: Media; SD: Desviación Típica; PF: Prueba final; PI: Prueba Inicial

## Método

### Participantes

El programa para desarrollar las habilidades metacognitivas se ha llevado a cabo en dos grupos-clase de estudiantes de segundo curso de Educación Secundaria Obligatoria de un centro de Educación Infantil, Primaria y

Secundaria de la ciudad de Zaragoza (Larraz, 2011). La TG se aplicó sobre la muestra del grupo experimental del citado programa, que es aquel que ha formado parte de la intervención y ha participado en el programa y se pueden comprobar las diferentes facetas que afectan a la medida de sus resultados. El

tamaño de la muestra de este estudio es de  $N=21$  participantes de ambos géneros: 10 mujeres y 11 varones. En términos de proporción, la variable de género, está compuesta por el 48% de mujeres y el 52% de varones. La edad media del grupo es de 13'2 años y la desviación típica de 0'58 años.

### Instrumentos

Para realizar el estudio sobre la generalizabilidad del diseño empleado se han tomado los datos de las puntuaciones obtenidas de los registros de los seis ejercicios realizados durante el programa que dan muestra de la evolución de los juicios metacognitivos a lo largo de su implementación. La escala de los juicios metacognitivos se ha basado en una escala tipo Lickert con un rango de 1-3 puntos. Los ejercicios que se llevaron a cabo consistieron en desarrollar las habilidades metacognitivas tomando medidas de antes, durante y después de la solución de diferentes problemas y tareas planteadas en el aula. Su número fue equivalente para cada uno de los participantes, por lo que cada participante realizó el mismo número de ejercicios en cada sesión de trabajo.

### Procedimiento

Para llevar a cabo un estudio de generalizabilidad (G), se ha de escoger un modelo de estimación apropiado en función del tipo de muestra y de los diferentes niveles de las facetas de estudio. En este caso, se ha llevado a cabo un modelo de tres facetas (ejercicios, participantes y género). Para ello, en una primera fase denominada *Plan de observación*, se han de analizar cuáles son los objetos de medida y los instrumentos de medida. “Los objetos de medida admisibles constituyen la población objeto de estudio y los instrumentos de medida (las condiciones de evaluación en terminología de Cronbach) constituyen el universo de generalización” (Blanco-Villaseñor, 1993, p. 193). Los objetos de medida se han denominado *facetas de diferenciación* y los instrumentos de medida se han denominado *facetas de generalización*.

Las facetas de diferenciación del estudio G son el número de participantes, el número de ejercicios y el género, cada uno de ellos se compone de distintos niveles. Tanto las facetas de los participantes como de los ejercicios son facetas aleatorias ya que una faceta es aleatoria “si una muestra aleatoria simple de niveles observados se extrae de un conjunto infinito (o

hipotéticamente infinito) de niveles admisibles” (Blanco-Villaseñor, 2001, p. 22), y el género es una variable fija, ya que una faceta es fija “si los niveles admisibles son representados de manera exhaustiva en el plan de observación, es decir si los niveles observados agotan los niveles admisibles (Blanco-Villaseñor, 2001, p. 22). En la Tabla 2 se representan las facetas de diferenciación tomadas para este estudio.

Tabla 2.

<i>Facetas de la TG</i>			
Facetas	Etiqueta	Niveles	Universo
Participantes	P	21	Infinito
Ejercicios	E	6	Infinito
Género	G	2	2

La faceta de diferenciación se ha contrastado sobre el número de ejercicios (E) y la faceta de generalización, se ha probado sobre los registros de los ejercicios. A partir de estas estimaciones, se puede comprobar si resultaría más útil aumentar el número de ejercicios en posteriores aplicaciones del programa. El motivo de seleccionar los ejercicios como faceta de diferenciación es debido a que entre las fuentes de variación existente, los ejercicios son los que mayor varianza aportan sobre las medidas obtenidas, tal y como se explica a continuación.

En una segunda fase, se han de analizar los componentes de estimación del estudio. Se le denomina *Plan de estimación*. En ella se selecciona el modelo de estimación más apropiado que estará determinado por el modo de muestrear los niveles de cada faceta. Tras analizar la estimación de los componentes de varianza explicada de los ejercicios (E), de los participantes (P) y del género (G) (Tabla 3), la importancia que tienen los ejercicios sobre las medidas es del 38,5%; la importancia del número de participantes es del 11,5% y; la importancia que tiene la interacción de los ejercicios y los participantes (P\*E) es del 50%. El género, al ser una variable fija, no aporta variabilidad a los resultados, por lo que se desestima como faceta explicativa del estudio, con un 0% de estimación de varianza. Al igual que los efectos de interacción estimados relacionados con el género, que aportan una variabilidad nula sobre las medidas. El software utilizado para analizar los datos ha sido el EduG 6.0 (Cardinet, Johnson, y Pini, 2010).

Tabla 3. *Proporción de varianza explicativa de los componentes del estudio*

Fuente de variación o Facetas	SS	Df	Media cuadrática	Modelo aleatorio	Modelo mixto	Componentes corregidos	Porcentaje de Varianza	SE
P	10.00	20	0.50	0.02	0.02	0.02	<b>11.5</b>	0.01
E	18.00	5	3.60	0.08	0.08	0.08	<b>38.5</b>	0.04
G	0.00	1	0.00	0.00	0.00	0.00	<b>0.0</b>	0.00
P* E	21.00	100	0.21	0.10	0.10	0.10	<b>50.0</b>	0.01
P*G	0.00	20	0.00	0.00	0.00	0.00	<b>0.0</b>	0.00
E*G	0.00	5	0.00	0.00	0.00	0.00	<b>0.0</b>	0.00
P*E*G	0.00	100	0.00	0.00	0.00	0.00	<b>0.0</b>	0.00
Total	49.00	251					100%	

Nota: P: Participantes, E: Ejercicios; G: Género; SS: Suma de cuadrados; Df: Grados de libertad; SE: Error estándar.

### Resultados

Para obtener los resultados del estudio G, se ha calculado el coeficiente de generalizabilidad ( $E_Q^2$ ), que es un índice global de la varianza de puntuaciones universo relativas a la varianza de error, definido como “la proporción de variancia observada que es atribuible a la puntuación universo, es decir es la razón entre el valor esperado de la variancia de puntuaciones universo ( $\sigma^2_r$ ) y el valor esperado de la variancia de puntuaciones observadas ( $E \sigma^2_x$ )” (Blanco-Villaseñor, 1993, p. 201). Esta fase en la que se muestran los resultados, se ha denominado *Plan de optimización*. En esta fase, tratamos, a través de la información que nos proporcionan los análisis precedentes, de identificar la mejor adecuación de los procedimientos de medida en función del plan de observación inicial para lograr una optimización de la faceta de diferenciación y generalización del estudio, es decir, del número adecuado de ejercicios y del número de registros de medida respectivamente.

Según los resultados obtenidos de las optimizaciones para comprobar la precisión de la

generalización de los ejercicios (ver Tabla 4), se observa que el número óptimo de ejercicios que se debería aplicar para obtener niveles adecuados de generalización es de 40, el cual obtiene un valor aceptable de generalización de .90, lo que nos obligaría a realizar 1680 registros de medida. La precisión de la generalización con seis ejercicios, que es el número de ejercicios objeto de cuantificación que han sido aplicados en el programa, es baja ya que tan solo alcanza un valor de .58. Si se aumenta el número de ejercicios a 10, la precisión de la generalización sigue siendo baja, con un valor de .69, siendo necesarios 420 registros de medida. En el caso de aumentar a 20 ejercicios, la precisión aumenta considerablemente alcanzando un valor de .82, pero aún sigue sin ser un nivel aceptable de generalización. A partir de 30 ejercicios, se obtiene un valor de generalización de .87, próximos a los niveles adecuados. En este sentido, se estima que para que las diferencias entre un grupo y otro sean significativas se debería aumentar el número de ejercicios para el desarrollo de las habilidades metacognitivas a un total de 40.

Tabla 4. Resultados del estudio G

Facetas y Coeficiente G	Opción 1		Opción 2		Opción 3		Opción 4		Opción 5	
	N	Univ.	N.	Univ.	N	Univ.	N	Univ.	N.	Univ.
Participantes	21	INF	21	INF	21	INF	21	INF	21	INF
Ejercicios	<b>6</b>	INF	<b>10</b>	INF	<b>20</b>	INF	<b>30</b>	INF	<b>40</b>	INF
Género	2	2	2	2	2	2	2	2	2	2
Observaciones.	252		420		840		1260		1680	
Coef. G		<b>0.58</b>		<b>0.69</b>		<b>0.82</b>		<b>0.87</b>		<b>0.90</b>
Coef G redondeado		0.58		0.70		0.82		0.87		0.90
Coef G absoluto		0.44		0.56		0.72		0.79		0.84
Coef. G absoluto redondeado		0.44		0.57		0.72		0.80		0.84

Nota: N: Número de Facetas; INF: Infinito; Univ: Universo; Coef. G: Coeficiente de Generalizabilidad.

### Discusión

Los resultados obtenidos en el estudio de la generalizabilidad han sido muy positivos, dado que la evaluación de la eficiencia nos ha servido para comprobar que el diseño utilizado para el programa de intervención es adecuado, aunque insuficiente en cuanto al número de ejercicios, ya que para que las diferencias sean significativas ( $p < .05$ ) entre los grupos de estudio debería aumentarse el número de ejercicios a un mínimo de 40.

Tras realizar el estudio de la generalizabilidad del diseño del programa, analizar las facetas de diferenciación y generalización y la fuente de varianza que aporta cada una de ellas, para futuras aplicaciones conocemos los criterios que debemos tener en cuenta para optimizar sus resultados, lo que es más importante, si cabe, cuando los recursos son escasos. Con este aumento del número de ejercicios se obtendría un desarrollo significativo de las habilidades metacognitivas en el grupo experimental con respecto al grupo control lo que, a su vez, lograría mejorar la generalización de los resultados del programa aplicado.

Estos resultados implican que el programa implementado fue adecuado en cuanto a sus pretensiones, dado que demuestra que trabajando en esa línea se desarrollan las habilidades metacognitivas en los estudiantes de Secundaria y permite analizar las variables del mismo para optimizarlas en futuros estudios. Para mejorar la validez externa, la optimización, la eficacia y potencia de los resultados obtenidos y el diseño del programa en futuras intervenciones, la faceta que nos parece más adecuado contemplar como objeto de mejora es ampliar el número total de ejercicios a 40.

Si atendemos a una visión integradora de la validez de un estudio en el ámbito educativo, tenemos que tener en cuenta esta teoría para incluir otras fuentes de variación que pueden resultar imprescindibles para evaluar sus resultados, sobre todo si se considera el rendimiento educativo como estructuras organizadas de conocimientos

complejas, lo que impone un cambio en los estudios de evidencia de la validez que englobe a todos los procedimientos de validación (Mateo y Martínez, 2008).

Por tanto, en función de lo estimado y las características propias del análisis mostrado en este estudio, se puede afirmar que la teoría de la generalizabilidad es una buena herramienta para predecir los resultados de una investigación y puede ser útil para planificar el diseño de las mismas de un modo más preciso y exacto.

## Referencias

- Anguera, M.T., Blanco-Villaseñor, A., Hernández Mendo, A. y Losada, J.L. (2011). Diseños observacionales: ajuste y aplicación en psicología del deporte. *Cuadernos de psicología del deporte*, 11(2), 63-76.
- Anguera, M.T. (1995). Diseños. En R. Fernández-Ballesteros (Ed.), *Evaluación de programas. Una guía práctica en ámbitos sociales, educativos y de salud* (pp. 149-172). Madrid: Síntesis.
- Arnau, J. (1990). Metodología experimental. En M. T. Anguera, G. Arnau y J. Gómez, *Metodología de la investigación en ciencias del comportamiento* (pp. 7-122) Murcia: Servicio de Publicaciones de la Universidad de Murcia.
- Arnau, J., Blanco-Villaseñor, A., y Losada, J. L. (1991). Estimación de la precisión de un diseño multivariable de medidas repetidas. *Anales de psicología*, 7(1), 85-104.
- Blanco-Villaseñor, A. (1989). Fiabilidad y generalización de la observación conductual. *Anuario de psicología*, 43 (4), 7-32.
- Blanco-Villaseñor, A. (1993). Fiabilidad, precisión, validez y generalización de los diseños observacionales. En M. T. Anguera (Ed.), *Metodología observacional en la investigación psicológica*, Vol. 2 (pp. 151-261). Barcelona: PPU.
- Blanco-Villaseñor, A (2001). Generalizabilidad de Observaciones Uni y Multifaceta: Estimadores LS y ML. *Metodología de las Ciencias del Comportamiento*, 3 (2), 161-199.
- Blanco-Villaseñor, A. y Anguera, M.T. (2000). Evaluación de la calidad en el registro del comportamiento: Aplicación a deportes de equipo. En E. Oñate, F. García Sicilia y L. Ramallo (Eds.), *Métodos Numéricos en Ciencias Sociales* (pp. 30-48). Barcelona: CIMNE.
- Blanco-Villaseñor, A., Sastre, S. y Escolano, E. (2010). Desarrollo ejecutivo temprano y teoría de la generalizabilidad: bebés típicos y prematuros. *Psicothema*, 22(2), 221-226.
- Cardinet, J., Johnson, S., & Pini, G. (2010). *Applying Generalizability Theory using EduG*. New York: Routledge Academic.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. (1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. New York: John Wiley and Sons.
- Larraz, N. (2011). *Desarrollo de las habilidades creativas y metacognitivas en la Educación Secundaria Obligatoria. Programa de intervención*. Tesis Doctoral. Zaragoza: Universidad de Zaragoza. Departamento de Psicología y Sociología. Biblioteca de la Facultad de Económicas y Empresariales.
- Máiquez, M. L., Blanco-Villaseñor, A. y Capote, C. (2000). Estimación de la precisión en la evaluación de un programa de intervención familiar. *Psicothema*, 12(2), 352-357.
- Mateo, J y Martínez, F. (2008). *Medición y evaluación educativa*. La Muralla: Madrid.
- Usabiaga, O., Casellano, J., Blanco-Villaseñor, A. y Casamichana, D. (2013). La Teoría de la Generalizabilidad en las primeras fases del método observacional aplicado en el ámbito de la iniciación deportiva: calidad del dato y estimación de la muestra. *Revista de psicología del deporte*. 22 (1), 103-109.

Received: 06/05/2013  
Accepted: 23/11/2014