

Evaluación de programas educativos a través de metodología observacional, aplicado a un programa extraescolar para altas capacidades desarrollado en México

Trinidad Romero González, Elena Rodríguez-Naveiras, África Borges del Rosal
Universidad de La Laguna

RESUMEN

Se aborda la evaluación de la conducta de las monitoras del Programa Integral para Altas Capacidades (PIPAC) implementado en la Universidad de Guadalajara, México, mediante metodología observacional, que analiza de forma sistemática y objetiva la conducta espontánea que se da en un contexto natural a lo largo de un continuo. Se ha usado como instrumento de observación el Protocolo de Observación de Funciones Docentes (PROFUNDO.v2). Las conductas más frecuentes de las monitoras fueron Guía Individual, Guía Grupal dada por la monitora principal y conducta Inobservable, y del alumnado Responde al Docente y se Dirige al Monitor. En cuanto a conductas interactivas, resultan significativos los códigos pertenecientes a la función de orientación, e interacción del alumnado. Se observan indicadores de buenas prácticas docentes, puesto que las monitoras desarrollan una conducta participativa, activa e interactiva.

Palabras clave: Evaluación de programas, programa extraescolar, metodología observacional, altas capacidades.

Evaluation of educational programs using observational methodology, applied to an out-of-school program for children with high abilities in Mexico.

In this research, the assessment of behaviors developed by the instructors in the PIPAC program, implemented at the University of Guadalajara, Mexico, is addressed. Observational methodology has been applied, which is a process of systematic and objective recording of spontaneous behavior that occurs in a natural context and along a continuum. We used an observational instrument, the Observational Protocol of Teaching Functions (PROFUNDO.v.2). The most common behavior in both the main and the basic instructor is the Individual Guide, the Group Guide given by the principal instructor and Unobservable given by both instructors, and students Responds to Faculty Manager and Monitor. Regarding interactive behaviors, the significant codes are the guidance function, and interaction of students. Indicators of good teaching practices are observed, since the instructors develop a participatory, active and interactive behavior.

Key words: Program evaluation, out-the-school programs, observational methodology, high capacities.

Las altas capacidades intelectuales son objeto de estudio desde el inicio de la Psicología Científica, con una gran tradición en países como Estados Unidos, Alemania y Gran Bretaña. El alumnado de mayor talento, si bien como grupo no demuestran tener peor adaptación personal, escolar o social que sus compañeros (Borges, Hernández-Jorge y Rodríguez-Naveiras, 2011), se enfrentan a diversos problemas, tales como asincronía o bien expectativas poco realistas de padres y profesores (Pfeiffer y Blei, 2008).

Dentro de la escuela, su mayor velocidad de aprendizaje y profundidad en los temas que aprenden, un temprano acceso tanto a contenidos avanzados y complejos como a tareas escolares repetitivas, pueden conducirles a desmotivación, falta de atención aburrimiento y conductas no deseadas (Pérez y Domínguez, 2006).

Ambas problemáticas, escolares y socioafectivas, han llevado al desarrollo de programas educativos de intervención que solventen esta problemática, tanto en el entorno escolar como fuera de la escuela.

Los programas intraescolares más habituales son aceleración, enriquecimiento y agrupamiento (Acereda, 2002), pudiendo diferenciar en los extraescolares entre cognitivos, que promueven un nivel de eficacia intelectual adaptado a su talento, y socioafectivos, planteados desde una perspectiva psicoeducativa, para potenciar el desarrollo emocional, las habilidades sociales y las conductas de cooperación, trabajo en grupo y respeto a los demás (Borges y Rodríguez-Naveiras, 2011).

El Programa Integral para Altas Capacidades (PIPAC), de corte socioafectivo, fue creado en la Universidad de La Laguna en 2004 (Borges y Rodríguez-Naveiras, 2014), implementándose también recientemente en la

Universidad de Guadalajara y la Universidad Autónoma del Estado de Morelos. Su objetivo es el desarrollo integral de escolares con alta capacidad, llevándolo a cabo mediante actividades lúdicas en grupo, respondiendo a tres tipos de contenidos: intrapersonales, interpersonales y cooperativos (Borges y Rodríguez-Naveiras, 2011).

No obstante, no solo es necesario la creación de programas dirigidos al alumnado más capaz, sino que es preciso su evaluación, pudiéndose esgrimir las siguientes razones: a) es necesario comprobar su utilidad, ya que la ausencia de datos que lo confirmen los hace vulnerables; b) es difícil mantener y mejorar la calidad de un programa que no se somete a valoración; y c) la evaluación permite conocer si el programa produce otros efectos, y si pudieran tener consecuencias (Borland, 2003).

Por lo general, se encuentran valoraciones positivas de la intervención extraescolar en diversos estudios (Borges, Hernández-Jorge y Rodríguez-Naveiras, 2006; Delcourt, Cornell, y Goldberg, 2007; López-Aymes, 2005; Rodríguez-Naveiras, Borges y Hernández-Jorge, 2010; VanTassel-Baska, Bracken, Feng, y Brown, 2009), siendo menos frecuente que la evaluación arroje resultados adversos. En este sentido, Adelson (2010) no encuentra efectos en matemáticas y lectura como resultado de la implementación de programas destinados a superdotados.

La evaluación de programas es multimodal, pudiéndose diferenciar en función del grado de implicación del evaluador, el momento del desarrollo del programa en el que se lleva a cabo o por su finalidad, que puede ser sumativa, cuando se trata de valorar el impacto de la intervención y formativa,

cuando se estudia el proceso llevado a cabo en el programa (Borges y Rodríguez-Naveiras, 2012).

En lo que respecta a la evaluación formativa, uno de los aspectos fundamentales a estudiar es el desempeño docente del monitor que implementa el programa, debido a la relevancia que tiene el enseñante en el proceso de aprendizaje (Baker et al., 2010; Martínez, 2013). Ello exige establecer

un marco teórico que permita categorizar la conducta del educador. El Modelo de Funciones Docentes de Hernández (1991), ampliado luego por Hernández-Jorge (2005), explica de forma comprensiva y parsimoniosa el comportamiento del docente, estableciendo cómo su conducta responde a ocho funciones para el desarrollo del proceso de enseñanza-aprendizaje (véase tabla número 1).

Tabla 1. Modelo de Funciones Docentes (Hernández-Jorge, 2005)

Funciones Docentes	Descripción
Organización	Planificación previa de la enseñanza y la organización sobre el contexto.
Comunicabilidad docente	Capacidad para comunicar contenidos de forma comprensible.
Motivación	Capacidad del docente para estimular al alumnado al aprendizaje.
Control Comportamental	Relacionado con gestión, orden y disciplina.
Orientación y Asesoría	Ayudas que otorga el docente al aprendizaje del alumnado.
Interacción y Cooperación	Corregir errores y ampliar la información.
Personalización	Adaptación de la enseñanza a diferentes cualidades del alumnado.
Evaluación	Evaluación de los resultados del programa de formación.

La evaluación obliga a la utilización de instrumentos de medida. La de la competencia del docente se ha realizado por diversos procedimientos: portafolios, encuestas a su alumnado, pruebas al docente o mediante observación de su conducta (Díaz-Hernández, 2014; Martínez, 2013).

Cuando el objetivo es evaluar comportamientos, la metodología de elección es la observacional, que es flexible y rigurosa, a la vez que permite obtener mucha y muy rica información. Posibilita estudiar el comportamiento humano en situaciones naturales y cuasi-naturales (Blanco, Sastre y Escolano, 2010), así como analizar las interacciones que se producen en el aula, focalizando y definiendo el objeto de estudio (Mayorga y López, 2005).

En el ámbito de altas capacidades en la evaluación del comportamiento del docente mediante observación, Van Tassel-Baska, Bracken, Feng y Brown (2009) encuentran que los profesores demuestran ser competentes en las materias que imparten, utilizando algunos profesores estrategias instruccionales y diferencias individuales con sus alumnos. Rodríguez-Naveiras (2011) evaluó a las monitoras de un programa extraescolar de corte socioafectivo, encontrando que las monitoras mostraban un comportamiento muy participativo, activo e interactivo, ya que los patrones más frecuentes eran las asesorías de las educadoras, seguidas de respuesta de los participantes, sea por iniciativa propia o respondiendo a intervenciones de las primeras.

El objetivo de esta investigación es doble. Por una parte, evaluar el comportamiento de las monitoras desarrollado en un programa de intervención extraescolar para alumnado de altas capacidades. Por otra, realizar una validación cruzada del Protocolo de Observación de Funciones Docentes

(PROFUNDO, Rodríguez-Naveiras, 2011) con el fin de comprobar si se adecua a la evaluación de educadores en un entorno distinto al que fue creado.

MÉTODO

Diseño

La metodología utilizada en este estudio es observacional, con un diseño nomotético-seguimiento-multidimensional (Blanco, Losada y Anguera, 2003).

Participantes

Se han observado los comportamientos de dos monitoras, la principal (encargada de guiar y dinamizar el grupo) y la básica (encargada de apoyar a la monitora principal), ambas son estudiantes de la Licenciatura en Psicología en la Universidad de Guadalajara (México) y becarias de investigación del programa de verano Delfín 2012. Los participantes en el programa fueron dos niñas y tres niños, con edades comprendidas entre los 4 y 6 años, todos ellos diagnosticados como alumnos sobresalientes.

Como observadoras participaron una Doctora en Psicología de 34 años y una estudiante de psicología, de 23 años.

Instrumentos

El instrumento de observación utilizado fue el Protocolo de Observación de Funciones Docentes (PROFUNDO, v.2) (Rodríguez-Naveiras, 2011), creado para la evaluación de las monitoras en un contexto de intervención psicoeducativa, basado en el modelo de funciones docentes de Hernández-Jorge (2005). En la tabla 2 se expone la estructura del protocolo de observación utilizado.

Tabla 2. *Protocolo de Observación de Funciones Docentes (PROFUNDO.v2)*

MACROCATEGORÍAS Funciones	CRITERIOS	CÓDIGOS		
Organización	Organización externa	Organización del contexto didáctico	OD OA	
		Organización del alumnado		
Comunicabilidad Docente	Conductas de dirección de las actividades	Exposición individual	EI	
		Exposición grupal	EG	
Motivadora	Integración en la actividad propuesta	Participación del monitor	PM	
		Refuerzo.		
Control Comportamental	Contingencias de control	Refuerzo individual	RI	
		Refuerzo grupal	RG	
	Guía, asesoría y retroalimentación.	Control individual	CI	
		Control grupal	CG	
Orientación y Asesoramiento	Revisión no verbal de la tarea.	Guía individual	GI	
		Guía grupal	GG	
		Revisión No Verbal.	RN	
Interacción	Interacción entre monitor y participantes	Interacciones generales	IG	
		No responde	NR	
	Intervención de los participantes	Interacciones entre monitores	Interacciones entre monitores	IM
			Se dirige al monitor	DM
Categorías Instrumentales		Responde al monitor	RP	
			DP	
		Otros comportamientos	X	
		Inobservable	Y	

Las sesiones se registraron con una cámara de vídeo SONY HDR CX100 Handycam, realizándose la codificación con el software de registro de conductas Augen.1 (Montero y Montero, 2011).

Procedimiento

El Programa Integral para Altas Capacidades (PIPAC) se implementó en la Universidad de Guadalajara, México durante el verano de 2012.

Antes de realizar la codificación se llevó a cabo un entrenamiento en observación, con una duración de 10 horas, según el procedimiento estandarizado desarrollado por el equipo de investigación (Cadenas, Rodríguez y Díaz, 2012; Rodríguez-Naveiras, 2011).

De las 18 actividades que se desarrollaron en el programa a lo largo de las sesiones, se seleccionaron seis, con una duración de entre 3 y 15 minutos, en función de los resultados de optimización obtenidos mediante la Teoría de la Generalizabilidad (véase apartado 4.3). Para su selección se tuvo en cuenta que fueran claramente observables y que los diálogos fueran fácilmente audibles.

Análisis de datos

El cálculo de la fiabilidad entre observadores se llevó a cabo con el coeficiente de fiabilidad Kappa de Cohen (Cohen, 1960, 1968), mediante el programa SPSS, versión 15, y con la Teoría de la Generalizabilidad (TG) (Cronbach, Gleser, Nanda y Rajaratnam, 1972), usando el programa SAGT v.1.0 (Ramos, Hernández-Mendo, Pastrana y Blanco-Villaseñor, 2012).

El Estudio de Homogeneidad y Optimización se realizó a través de la TG empleando el programa SAGT v.1.0 (Ramos, Hernández-Mendo, Pastrana y Blanco-Villaseñor, 2012).

Para obtener las frecuencias absolutas y relativas de los eventos se usó el programa SPSS versión 15.0, utilizando el programa GSEQ v.5.1. (Bakeman y Quera, 1996) para el cálculo del análisis secuencial de los datos.

RESULTADOS

Tras el entrenamiento, se calculó la fiabilidad interobservadores mediante dos coeficientes: Kappa (Cohen, 1960, 1968) y Teoría de la Generalizabilidad (TG) (Cronbach, Gleser, Nanda y Rajaratnam, 1972). Ambos coeficientes se complementan, pues mientras el primero permite establecer el porcentaje de acuerdo entre observadores, controlando el efecto del azar, la TG analiza la variabilidad de las puntuaciones observadas según fuentes separadas de variabilidad (Blanco-Villaseñor, 1991), siendo una medida idónea para la identificación de problemas en un sistema categorial y su interacción.

El coeficiente Kappa fue de 0,87, que supera los valores criterio establecidos por Fleiss (1981).

El diseño empleado en la TG fue multifaceta cruzado, AGxCxO, partiendo de un Plan de Observación que comprende tres facetas de efectos aleatorios, dos de diferenciación: códigos (C), con veinte niveles (los códigos de conducta directamente observables) y agentes (AG), con tres niveles: monitora principal, monitora básica y participantes, así como una de generalización, observadores (O), con dos niveles, relativos a las dos observadoras.

Los resultados del Plan de Medida se muestran en la tabla 3 y corroboran la fiabilidad inter observadoras.

Tabla 3. *Plan de medida AGC/O*

Fuente	Varianza de diferenciación	Fuente	Varianza de error relativa	% relativo	Varianza de error absoluta	% absoluto
		O			0.209	5.645
AG	0.000	OAG	0.000	0.000	0.000	0.000
C	8.565	OC	0.516	14.795	0.516	13.960
AGC	99.620	OAGC	2.973	85.205	2.973	80.395
Coefficiente G relativo	0.969					
Coefficiente G absoluto	0.967					

A continuación se procedió a realizar el estudio de homogeneidad, para garantizar que el instrumento PROFUNDO v.2 (Rodríguez-Naveiras, 2011) es discriminativo en otro contexto distinto al que se creó, en concreto en la evaluación de las monitoras del programa PIPAC implementado en la Universidad de Guadalajara, México.

Se siguió un diseño multifaceta cruzado AGxCxO, partiendo de un *Plan de Observación* con tres facetas de efectos aleatorios, dos de diferenciación: agentes (AG), con tres niveles, que se corresponden con la monitora principal, la monitora básica y los participantes; y observadores (O), con dos niveles, observador 1 y observador 2. La faceta a generalizar son los códigos (C), con veinte niveles, los 20 códigos de conducta directamente observables, dando lugar a un diseño multifaceta cruzado AGxCxO. El Plan de Medida se presenta en la tabla 4, mostrando que el instrumento es discriminativo aplicado a la evaluación de monitoras del programa desarrollado en la Universidad de Guadalajara, México.

Con el objetivo de determinar el número mínimo de actividades necesarias para garantizar la obtención de datos fiables se realizó un análisis de optimización a través de la TG, con un diseño multifaceta cruzado, CxAGxA, de efectos aleatorios, siguiendo el Plan de Medida CAG/A, con tres facetas, dos de diferenciación: códigos (C), con veinte niveles (códigos); y agentes (AG), con tres niveles: monitora principal, básica y participantes. La faceta a generalizar son las actividades (A), que presentan en cinco niveles (tabla número 5).

Como se puede observar, con seis actividades se alcanzan coeficientes de generalizabilidad aceptables para situaciones de evaluación de programas, ya que se superan el criterio establecido por Hintze y Matthews (2004), siendo el coeficiente superior a 0.70.

Tabla 4. *Plan de medida OAG/C*

Fuente	Varianza de diferenciación	Fuente	Varianza de error relativa	% relativo	Varianza de error absoluta	% absoluto
O	0.418					
AG	0.000					
OAG	0.000					
		C			0.428	7.438
		OC	0.052	0.969	0.052	0.897
		AGC	4.981	93.453	4.981	86.503
		OAGC	0.297	5.578	0.297	5.163
Coefficiente G relativo	0.073					
Coefficiente G absoluto	0.068					

Tabla 5. *Plan de Optimización de las actividades*

Sesiones	5	6	7	8
Coef. G relativo	0.682	0.720	0.750	0.775
Coef. G absoluto	0.671	0.710	0.740	0.765

Una vez garantizada la calidad del dato, se procedió a analizar los datos de las sesiones seleccionadas. La metodología observacional, cuyo objetivo es la descripción y búsqueda de relación en comportamientos observados, permite dos tipos de análisis: macroanálisis y microanálisis (Fontes, García, Garriga, Pérez-Llantada, y Sarriá, 2001). A continuación se detalla cada uno de ellos.

Mediante el macroanálisis se describen las relaciones observadas a través de medidas globales, que, puesto que el tipo de datos es de eventos (Bakeman y Quera, 1996), el parámetro más adecuado es determinar su

frecuencia. En la tabla número 6 se recogen las frecuencias absolutas y relativas de los códigos de conductas de las monitoras, principal y básica y en la tabla 7 la de los participantes.

Como se puede observar, la conducta más frecuente emitida por ambas monitoras, principal y básica, es *Guía Individual* (GI), seguida por *Guía Grupal* (GG) emitida por la monitora principal. Le sigue en frecuencia la conducta *Inobservable* (Y), emitida con la misma frecuencia tanto por la monitora principal como por la básica, lo que señala dificultades técnicas en el registro.

Tabla 6. Frecuencias absolutas y relativas de las conductas de las monitoras.

Códigos	Monitora Principal		Monitora Básica	
	Frecuencia Absoluta	Frecuencia Relativa	Frecuencia Absoluta	Frecuencia Relativa
OA	3	,01	0	,00
OD	0	,00	0	,00
EI	2	,01	0	,00
EG	11	,03	0	,00
PM	0	,00	0	,00
IM	7	,02	6	,02
RI	13	,03	12	,03
RG	1	,00	0	,00
CI	1	,00	0	,00
CG	2	,01	1	,00
GI	86	,23	38	,10
GG	29	,08	4	,01
RN	0	,00	0	,00
X	1	,00	8	,02
Y	26	,07	26	,07
IG	2	,01	0	,00
NR	0	,00	0	,00

Tabla 7. Frecuencias absolutas y relativas de las conductas de los participantes.

CÓDIGOS	Participantes	
	Frecuencia Absoluta	Frecuencia Relativa
IG	2	,01
NR	0	,00
DM	24	,06
RP	72	,19
DP	0	,00

La conducta más frecuente emitida por los participantes es, la *Respuesta del Participante* (RP), seguida por el código *Se Dirige al Monitor* (DM).

El procedimiento empleado en este estudio para el análisis secuencial es el de retardos, mediante el programa informático GSEQ 5.1 (Bakeman, Quera & Gnisci, 2009).

El método de retardos sirve para detectar patrones de conductas que tienen lugar con una probabilidad mayor lo que cabría esperar por azar, estableciendo una conducta antecedente o criterio, a partir de la cual se contabilizan las conductas que le siguen, consecuentes, dándose dependencia positiva $Z > 1,96$ cuando la probabilidad de ocurrencia de conducta consecuente es mayor a la que se pudiera dar por el azar, y negativa, cuando la conducta criterio inhibe la consecuente con probabilidad mayor al azar, $Z < -1,96$.

Se ha considerado conducta criterio toda aquella cuya frecuencia relativa sea superior a 0,10, emitida por cualquiera de los tres agentes (monitora principal, básica y participantes) y como consecuente cualquiera con frecuencia superior a 0.

En la tabla número 8 se presentan los resultados del análisis secuencial, dándose

solo patrones significativos en el caso de una conducta emitida por la monitora principal, en concreto *Guía Individual*, que son indicaciones o instrucciones del monitor, referidas a un participante en concreto, para responder a cómo se está realizando la actividad. Le siguen dos comportamientos: *Responde al Monitor* (RP), emitida por los participantes, que son respuestas a las preguntas o comentarios del monitor, o bien *Refuerzo Individual* (RI) emitida por la monitora principal, que se codifica cuando el monitor elogia, recompensa o presenta al alumno como un ejemplo a seguir. No hay conductas inhibitorias que sigan a la criterio. El residual hace referencia al valor de Z, presentando entre paréntesis el número de veces que la conducta consecuente sigue a la conducta criterio en relación a cuantas veces se da ésta en el periodo observado.

Tabla 8. Resultados por códigos de las sesiones

Conducta criterio	Consecuente	
	Retardo 1	Residual
GI M-P	RI MP	2,74 (7/13)
	RP P	3,67 (28/72)

DISCUSIÓN

En primer lugar, y en cuanto a la calidad del dato, se puede afirmar que existe fiabilidad interobservadores con los dos estadísticos utilizados. Asimismo, se comprueba la homogeneidad del instrumento, lo que confirma la aplicabilidad del mismo en otros contextos fuera del que fue creado.

Un aspecto importante a destacar es la necesidad de hacer un uso eficiente de la

metodología observacional, facilitando así su aplicabilidad y disminuyendo uno de sus mayores problemas, lo exigente que resulta en tiempo y recursos. En este sentido, la TG permite determinar, en este caso, el número de actividades a codificar, que han sido seis, manteniendo estándares de generalizabilidad de resultados. No obstante, y puesto que solo se dan dos patrones significativos, es de cuestionar si no sería recomendable incrementar el número de actividades.

En un trabajo reciente (Rodríguez-Dorta, 2014) se usó un segundo criterio, además de lo arrojado por la optimización para seleccionar las sesiones a observar: continuar la codificación hasta que no aparezcan más patrones relevantes.

La conducta más frecuente de las monitoras es *Guía Individual* (GI), mientras que la conducta *Guía Grupal* (GG) solo aparece con una frecuencia relevante en la monitora principal. La mayor frecuencia en la guía por parte de las monitoras es congruente con los resultados encontrados en el trabajo de Rodríguez-Naveiras (2011).

Las conductas más frecuentes de los participantes son las interacciones con las monitoras, tanto como respuesta a las intervenciones de estas (RP), como por iniciativa propia (DM). Cabe señalar la ausencia de conductas disruptivas. A nivel de análisis secuencial, sólo hay patrones significativos cuando la conducta criterio es *Guía Individual* (GI) emitida por la monitora principal, dando lugar a dos patrones: o bien le sigue *Responde al monitor* (RP) por parte de los participantes o *Refuerzo Individual* (RI) por parte de la monitora principal. Ello pone de manifiesto que, mientras que el primer patrón señala una buena interacción entre monitora principal y participantes, el segundo resulta más cuestionable, pues se da refuerzo sin una conducta antecedente adecuada de los participantes.

Por otra parte se ha comprobado que la responsabilidad de la implementación del programa recae en la monitora principal, ya que la frecuencia de los códigos de conducta emitidos por ella es más alta que en la básica, y en que solo se ha podido extraer una estructura comportamental de una conducta criterio de dicha monitora. Este resultado, coincide con los hallados en Rodríguez-Naveiras (2011), y pone de manifiesto que asume su rol.

Esta investigación permite también extraer algunas reflexiones de importancia. La información que da el instrumento de observación permite informar a las monitoras acerca de sus conductas, para posteriormente realizar planes de mejora con respecto a las funciones que desarrollan.

Una limitación importante de este estudio es debida a factores técnicos. Se ha podido comprobar una alta frecuencia de aparición de las conductas inobservables (Y) por parte de ambas monitoras, debido principalmente a que, al contar con una única cámara de video, no se abarcaba todo el campo de visión del aula donde se desarrollan las actividades. Esto supone un problema importante, ya que produce pérdida de información, dificulta la interpretación de resultados y provoca la aparición de secuencias de conductas ambiguas y difíciles de interpretar. En un futuro sería conveniente poder contar con mayores medios, teniendo el aparataje necesario para ampliar la observación de toda la sesión. De contar sólo con una cámara, lo que suele ser lo más común, la recomendación fundamental a los implementadores del programa es estudiar cuidadosamente el lugar en el que ubicarla, para permitir un registro preciso, minimizando las inobservabilidades.

A modo de conclusión, los resultados permiten confirmar la validación cruzada del instrumento de observación, dado que ha permitido evaluar la conducta de monitores del Programa Integral para Altas Capacidades en su versión implementada en la Universidad de Guadalajara, México. Ello hace pensar que el instrumento PROFUNDO. v2 (Rodríguez-Naveiras, 2011) podría utilizarse en otros programas extraescolares. De hecho, en este momento se está realizando la evaluación del monitor de un taller de robótica (Valadez, Rodríguez-Naveiras y Borges, en preparación).

Los resultados aquí obtenidos ponen de manifiesto la utilidad de crear instrumentos de observación con probada bondad, que permitan aplicaciones prácticas, sin tener que desarrollar, para cada investigación, un protocolo específico, ya que supone una clara amenaza a la eficiencia.

REFERENCIAS

- Acereda, A. (2002). *Niños superdotados*. Madrid: Pirámide.
- Adelson, J. L. (2010). Examining the effects of gifted programming in mathematics and reading using the ecls-k. Pro Quest Information y Learning). *Dissertation Abstracts International Section A: Humanities and Social Sciences*, 70 (7), 9-16.
- Bakeman, R. y Quera, V. (1996). *Análisis de la interacción. Análisis Secuencial con SDIS y GSEQ*. Madrid: Ra-Ma.
- Bakeman, R., Quera, V. & Gnisci, A. (2009). Observer agreement for timed-event sequential data: A comparison of time-based and event-based algorithms. *Behavior Research Methods*, 41 (1), 137-147.
- Baker, E. L.; Barton, P. E.; Darling-Hammond, L.; Haertel, E.; Ladd, H. F.; Linn, R. L.; Ravitch, D.; Rothstein, R.; Shavelson, R. J. & Shepard, L. A. (2010). *Problems with the use of student test scores to evaluate teachers*. Washington, D. C.: Economic Policy Institute.
- Blanco-Villaseñor, A. (1991). La teoría de la generalizabilidad aplicada a diseños observacionales. *Revista Mexicana de Análisis de la Conducta*, 17(3), 23-63.
- Blanco, A.; Losada, J.L. y Anguera, M.T. (2003). Analytic techniques in observational designs in environment-behavior relation. *Medio Ambiente y Comportamiento Humano*, 4 (2), 111-126.
- Blanco, A.; Sastre, S. y Escolano, E. (2010). Desarrollo ejecutivo temprano y Teoría de la Generalizabilidad: bebés típicos y prematuros. *Psicothema*, 22 (2), 221 - 226.
- Borges, A.; Hernández-Jorge, C. y Rodríguez-Naveiras, E. (2006). *Evaluación formativa en una escuela de padres y madres de niños y niñas superdotados*. X Congreso Internacional de Educación Familiar. Las Palmas de Gran Canaria. España
- Borges, A.; Hernández-Jorge, C. y Rodríguez-Naveiras, E. (2011). Evidencias contra el mito de la inadaptación de las personas con altas capacidades intelectuales. *Psicothema*, 23, 362-367.
- Borges, A. y Rodríguez-Naveiras, E. (2011) Diferencias en comportamientos docentes entre monitoras de un programa socioafectivo extraescolar. *Revista de Investigación y Divulgación en Psicología y Logopedia*, 1(1), 21-29.
- Borges, A. y Rodríguez-Naveiras, E. (2012) Programas de Intervención en altas capacidades y su evaluación. En M.D. Valadez, J. Betencourt y Zabala, M.A. (Eds.) *Alumnos superdotados y talentosos* (pp. 397-408). México: Manual Moderno.
- Borges, A. y Rodríguez-Naveiras, E. (2014). *Una década del Programa Integral para Altas Capacidades: origen, desarrollo y evaluación*. IX Congreso Iberoamericano de Psicología. Lisboa.
- Borlan, J.H. (2003) Evaluating gifted programs: A broader perspective. En N. Colangelo y G.A. Davis. *Handbook of gifted education* (pp. 293-310). Boston: Allyn and Bacon.
- Cadenas, M.; Rodríguez, M. y Díaz, M. (2012). Los equipos de entrenamiento: una muestra para el estudio de los sesgos en fiabilidad entre parejas de observadores. *Revista de Investigación y Divulgación en*

Psicología y Logopedia de la Universidad de La Laguna, 2(2), 41-46. ISSN 2174-7571.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20, 37-46.

Cohen, J. (1968). Weighted kappa: Nominal scale agreement with provision for scaled disagreement of partial credit. *Psychological Bulletin*, 70, 213-220.

Cronbach, L.J.; Gleser, G.C.; Nanda, H y Rajaratnam, N. (1972). *The dependability of behavioral measurements: theory of generalizability for scores and profiles*. Nueva York: John Wiley and Sons.

Delcourt, M. A. B.; Cornell, D. G.; y Goldberg, M. D. (2007). Cognitive and affective learning outcomes of gifted elementary school students. *Gifted Child Quarterly*, 51(4), 359-381.

Díaz-Hernández, M. (2014). *Protocolo de Observación de Funciones Docentes en Universidad: Un instrumento para la evaluación de la conducta del profesorado universitario*. Tesis Doctoral. La Laguna: Universidad de La Laguna. Servicio de Publicaciones.

Fontes, S.; García, C.; Garriga, A.; Pérez-Llantada, M. C. y Sarriá, E. (2001). *Diseños de investigación en Psicología*. Madrid. UNED.

Fleiss, J. L. (1981). *Statistical methods for rates and proportions*. New York: John Wiley

Hernández, P. (1991). *Psicología de la Educación: corrientes actuales y teorías aplicadas*. Méjico: Trillas.

Hernández-Jorge, C. (2005). *Habilidades de comunicación para profesionales*. Tenerife: ARTE: Comunicación visual.

Hinze, J.M y Matthews, W.J (2004). The generalizability of systematic direct observations across time and setting:

A preliminary investigation of the psychometrics of behavioural observation. *School Psychology Review*, 33(2), 258-270.

López-Aymes, G. (2005). *Aplicación del programa: La aventura de aprender a pensar y a resolver problemas. Un estudio comparativo entre niños y niñas de diferentes capacidades intelectuales*. Tesis Doctoral no publicada. Universidad Complutense de Madrid.

Martínez, J. F. (2013). Combinación de las mediciones de la práctica y el desempeño docente: consideraciones técnicas y conceptuales para la evaluación docente. *Pensamiento Educativo. Revista de Investigación Educativa Latinoamericana*, 50(1), 4-20. DOI: 10.7764/PEL.50.1.2013.2

Mayorga, M. J. y López, A. (2005). Observar para evaluar al profesorado: una experiencia práctica. *XXI Revista de Educación*, 7, 167-181

Montero, J. y Montero, J. (2011). *Programa Software Augen, v. β*. Computer Business Solutions.

Pérez, L.F. y Domínguez, R. (2006) El concepto de superdotación como base de experiencia y propuestas. En L.F. Pérez (coord) *Alumnos con capacidad superior. Experiencias de intervención educativa*, (pp. 17-50). Madrid: Síntesis.

Pfeiffer, S. I. y Blei, S. (2008). Serving gifted students. In N. Mather (Ed.), *Evidence-based interventions for students with learning and behavioral challenges*. (pp. 336-358). New York, NY US: Routledge/Taylor y Francis Group.

Ramos, F. J.; Hernández-Mendo, A.; Pastrana, J. L. y Blanco-Villaseñor, A. (2012). *SAGT: Software para la Aplicación de la Teoría de la Generalizabilidad*. Proyecto fin de carrera para la titulación: Ingeniería Técnica en Informática de Gestión de la Escuela Técnica Superior de Ingeniería Informática.

Universidad de Málaga (España). (Registro SafeCreative Código: 1209202373502 Fecha 20-sep-2012 18:03 UTC).

Rodríguez-Dorta, M. (2014). *Evaluación de proceso del comportamiento docente en educación primaria y especial*. Tesis doctoral no publicada. Universidad de La Laguna.

Rodríguez-Naveiras, E.; Borges, A. y Hernández-Jorge, C. (2010, julio). An approach to the evaluation procedure: comparing the behavior of the program instructor. 12th Conference of European Council for High Ability "Perspectives on the evaluation of giftedness". Paris.

Rodríguez-Naveiras, E. (2011). PROFUNDO: *Un instrumento para la evaluación de proceso de un programa de altas capacidades*. La Laguna: Universidad de La Laguna. Soportes Audiovisuales e Informáticos.

Valadez, D.; Rodríguez-Naveiras, E. y Borges, A. (en preparación) *Evaluación del monitor de un taller de robótica, dirigido a alumnado de altas capacidades*.

VanTassel-Baska, J.; Bracken, B.; Feng, A. y Brown, E. (2009). A longitudinal study of enhancing critical thinking and reading comprehension in title I classrooms. *Journal for the Education of the Gifted*, 33(1), 7-37.