CREATING AND PURIFYING AN OBSERVATION INSTRUMENT USING THE GENERALIZABILITY THEORY

CONSTRUYENDO Y DEPURANDO UN INSTRUMENTO DE OBSERVACIÓN MEDIANTE LA TEORÍA DE LA GENERALIZABILIDAD

ELENA RODRÍGUEZ-NAVEIRAS, ÁFRICA BORGES Y MARÍA CADENAS

Universidad de La Laguna

Abstract

The control of quality of data it is one of the most relevant aspects in observational researches. The Generalizability Theory (GT) provides a method of analysis that allows us to isolate the various sources of error measurement. At the same time, it helps us to determine the extent to which various factors can change and analyze the effect on the generalizability coefficient. In the work shown here, there are two studies aimed to creating and purifying an observation instrument, Observation Protocol in the Teaching Functions (Protocolo de Funciones Docentes, PROFUNDO, v1 and v2), for behavioral assessment which has been carried out by instructors in a social-affective out-of-school program. The reliability and homogeneity studies are carried out once the instrument has been created and purified. The reliability study will be done through the GT method taking both codes (c) and agents (a) as differential facets in. The generalization will be done through observers using a crossed multi-faceted design (A × O × C). In the homogeneity study the generalization facet will be done through codes using the same design that the reliability study.

Key words: Generalizability Theory, observational instrument, observational methodology, reliability, homogeneity.

Resumen

El control de la calidad de los datos es uno de los aspectos más relevantes en las investigaciones observacionales. La Teoría de la Generalizabilidad (TG) proporciona un método de análisis que nos permite aislar las fuentes de variación debidas al error aleatorio. Al mismo

Correspondencia: Elena Rodríguez-Naveiras. Universidad de La Laguna. Campus de Guajara s/n. Facultad de Psicología. Email: naveiras@ull.es

tiempo, nos ayuda a determinar el grado en que ciertos factores pueden cambiar y analizar su efecto mediante el coeficiente de generalizabilidad. En este trabajo, se presentan dos estudios que tratar de construir y depurar un instrumento de observación, el Protocolo de Funciones Docentes (PROFUNDO, v1 y v2), que trata de evaluar la conducta realizada por los instructores in programas socio-afectivos fuera de la escuela. Se han realizado estudios de fiabilidad y homogeneidad mediante TG considerando como facetas diferenciales tanto los códigos (c), como los agentes (a). La generalización se llevará a cabo a través de los observadores usando un diseño cruzado multifaceta (A × O × C). En el estudio de homogeniedad, la faceta de generalizabilidad se llevará a cabo a través de los códigos, usando el mismo diseño que en el estudio de fiabilidad.

Palabras claves: Teoría de la Generalizabilidad, instrumento de observación, metodología observacional, fiabilidad, homogeneidad.

Introduction

The program assessment is an essential activity to do in order to value and to take decisions in the education and psychology intervention programs. However, the program assessments don´t follow only one procedure but there are several approaches when evaluating programs (Newcomer, Hatry, & Wholey, 2010). Moreover, there are so many assessment aspects that must be evaluated that a single approach like the quantitive and quality perspective are specially complementary at this stage. Thus, while the cost estimation or the changes check-up due to the program can be tackled from a quality point of view, aspects like the study of contents (Anguera, 2008) or aspects like taking into account the opinions of specific skateholders (like the sponsors or the program designers) are carried out through intensive interviews (Rossi, Friendman, & Lipsey, 2004).

However, there is a type of assessment where the choosing methodology is the observational one, in which a quality approach is combined with a quantification. It is the training assessment when the interest is focused on analyzing the participants´ behavior in both cases, the analysis of the students´ interaction and the behavior of people who make the program (Pedrosa & Borges, 2012). When it is an educational program, the assessment of the teachers behavior becomes essential as he is the one who leads the educational process.

The observational methodology is flexible, accurate as well as, it is clear in time and effort since a constant flow of behavioral patterns must be transfered into data. This is done by using the instruments which allow this transfer, which means the observational instruments. In observational methodology, the lack of standardized instruments is quite frequent except for those instruments which are already standardized (Cerezo, 1991, 2000; Trenado & Cerezo, 2004). This fact makes this type of assessment difficult when practising because making an observational instrument for each situation which guarantees the quality in both, reliability and homogeneity, requires longer time for the assessment process as well as the fact that it makes it more difficult.

The observational methodology, which needs the highest level of naturalness in the context where it develops, requires the highest level of rigor and appropriateness of measuring instruments to capture the behavioral flow. Therefore, it is essential to guarantee the data quality, making it essential to carry out appropriate training to observers to ensure the reliability of the records (Anguera, 1988, 2003). Thus, the observer will have a greater conceptual, empirical and technological management to perform his or her task (Anguera, 2003).

However, this type of methodology can imply mistakes. The most important ones are the categorization and the coding of behavior, due to the observers or problems due to technical means of register (Behar & Riba, 1993; Ibañez, 1993; Quera, 1991; Quera & Behar, 1997).

To determine the reliability of observational codifications, different rates or ratios have been suggested. The kappa of Cohen Coefficient (1960) studies the match between observers, measuring the degree of agreement between

them. The maximum possible match is equal to 1 whereas the minimum possible match is equal to 0.

Moreover, the Generalizability Theory (GT) has turned out to be specially suitable to determine the data quality as an alternative method like assessing the psychometric properties of direct observation. Based on the domain sampling approach, GT assumes that any observation of behavior represents a random sample of behavior from a hypothetical domain or universe of all possible behavior. Developed by Cronbach, Gleser, Nanda & Rajaratnam (1972), GT represents a set of techniques that can be used for assessing the degree to which a given set of measurements of an individual generalize to a more extensive set of measurements of that individual.

The GT provides a suitable framework to find the estimates of reliability and margin of error, presenting a globalizing nature to suit the particular conditions of each object of measurement (Blanco, Castellano, & Hernández Mendo, 2000; Blanco, Sastre, & Escolano, 2010; Máiquez, Blanco, Rodrigo, & Vermaes, 2000).

Through the generalizability theory we can check the homogeneity in the categories of the observational instrument which is the extend to which the categories are differential factors among them (Blanco, 1989; Castellano, Hernández Mendo, Gómez, Fontetxa, & Bueno, 2000). Mutual exclusiveness is a basic requirement for homogeneity. When elaborating an observational instrument it must be taken into that the category system requires exclusiveness and accurateness and the field format requires mutual exclusiveness (Anguera, 1979, 1991; Anguera & Blanco, 2003; Castellano, 2000; Gorospe, 2005).

The objective of this work is to present the creating and purifying process of an observational instrument for measuring the behavior performance of instructors in a high ability out-of-school program. In order to do so, two studies were carried out. The first study focused on determining the data quality and the homogeneity from the observation instrument, Observation Protocol in the Teaching Func-

tions (PROFUNDO, v1) and the second study focused on the purifying process of the observation instrument, leading to the second version of the same one (PROFUNDO, v2).

Study 1

Method

Participants

The participants of this study are the children who took part in the Program to Complete Development of High Ability Children developed during the academic year 2006/2007. In particular in two levels of intervention: «the children level», made up by four boys and four girls aged from 4 to 6 years old and «the preadolescent level» made up by nine boys and three girls aged from 9 to 12 years old.

The second group of participants is made up by the instructors who implemented the program. Each level of intervention was guided by two instructors, a main instructor that was responsible for the program (in this case, a 27 year old one for «the children level», and a 25 year old one for «the preadolescent level», both graduates in Psychology) and a secondary instructor that helped with the main instructor´s task (a 20 year old one for «the children level» and a 19 year old one for «the preadolescent level», both students in Psychology).

Finally, the third group of participants is made up by three observers that have participated in the observation training, these were two students in Psychology and a graduate in Psychology.

Procedure

The Program to Complete Development of High Ability Children (PIPAC) was set up in 2004 at La Laguna University. This is an out-of-school program whose aim is the complete development in high ability children. It is carried out throughout the academic year in fort-

night sessions which last one hour and are mainly concentrated on social-affective aspects.

To ensure the correct use of the observational instrument, a group of six observers was trained in six different sessions which lasted four hours each.

Instrument

Observation Protocol in the Teaching Functions (PROFUNDO, v1)

In order to assess the instructors´ behavior in the PIPAC program, the work and research group in high abilities at La Laguna University came up with an ad hoc observational instrument called Observation Protocol in the Teaching Functions (PROFUNDO) based on the combination of field formats and category systems (Anguera & Blanco, 2003; Blanco et al., 2006 named by Gil Galve, 2008).

It is about an observation instrument which is categorized and made up by a generic scope which includes six macro-categories based on Hernández-Jorge´s (2005) teaching functions: organization function, teaching communicative function, behavioral control function, motivation function, orientation and assessment function and interaction function. These macro-categories are spread in 13 criteria which are accurate specifications of them, operating with 30 observable behavioral codes.

To complete the constant flow of the instructors´ behavior, it is necessary to include an instrumental category which collects both unobservable behavior (those which are not in the scope of the camera) and other types of behavior (those which are not connected to the teaching functions) (see chart 1).

Chart 1

Observation Protocol in the Teaching Functions (PROFUNDO), first version.

MACRO-CATEGORIES Functions	CRITERIA	CODES (*)	
1. Organization Function	External organization	Pupils`organization	(OA)
		Organization of the teaching material	(OD)
2. Teaching Communicative Function	Connection with previos sessions	Connection with previos sessions	(CA)
	Behavioral patterns when leading the activities	Individual explanation	(EI)
		Group explanation	(EG)
	Instructor`s participation	Instructor`s participation	(PM)
3. Motivation Function	Create useful expectations	Create expectations	(GE)
	Encouragement	Individual encouragement	(RI)
		Group encouragement	(RG)
4. Behavioral Control Function	Control resources	Individual negative control resources	(CI)
		Group negative control resources	(CG)
5. Orientation and Assessment Function	Instructor`s help or colaboration with the participants	Instructor`s help	(AM)
	Feedback or assessment	Individual assesment	(AI)
		Group assesment	(AG)
	Non verbal homework check-up	Non verbal check-up	(RN)
	Interaction between the instructor and the participants	Interactions not related to the activity	(IG)
		No answer from the instructor	(NR)
		Interrupts the participants	(IP)
	Interaction between instructors	Interaction between instructors	(IM)
		No answer to the other instructor	(NR)
6. Interaction Function		Answers to the instructor`s questions	(RM)
		Asks the intructor	(PM)
		Other participants`complaints	(QP)
	Participants` intervention	Participant`s own innitiative	(IPP)
		Participant`s positive assesment	(VP)
		Participants` interruptions	(DP)
		No answer (participants)	(NR)
		Participants`physical contact	(CP)
Instrumental Category	Other behavioral patterns		(X)
	Non-visible		(Y)

Note. * These acronyms remain in the Spanish terminology.

Data analysis

An SPSS 17.1 version and the GT program (Ysewijn, 1996) have been used for the statistic analysis.

Results

Once the training finished, the reliability and homogeneity study started. The three observers registered the same observation sessions

proving the match among observers through the Kappa of Cohen Coefficient. Two of the three observers were not experts and their results were compared to the expert observer´s results. These results did not reach the score of 0.80 (instead they got 0.71 and 0.73).

On the other hand, the data quality was also analyzed using the GT. The Observation Plan, designed for the calculation of the reliability to the observers and taking into account the whole observation session, consists of three facets. Two of them related to different levels (30 codes of behavior) and agents with fourteen levels which correspond to those who have been observed. That is to say, the instructors (the main and the secondary one) and each of the participants. And the last facet, which is that of the generalization where the observers are found. All of these lead to a crossed multifaceted observation plan A × O × C. The results of the Optimization, Estimation and Observation Plans have been obtained through the GT program (Ysewijn, 1996).

The homogeneity study was carried out by the GT. A cross facet designed is used with two facets of differentiation, these are, observers (o1, o2 and o3) and agents (with 14 levels), and the facet to generalize are the behavioral codes (with 30 levels).

The data connected to the Estimation Plan for the three observers are found in chart number 2. As it can be seen, the reliability rates do not exceed 0.83. This low rate of reliability makes it advisable to calculate the optimization plan (see chart 3). The results show that in order to get an acceptable data quality of the observation instrument, it would be necessary to count on six observers at least. That is why, carried out a second study with seven observers. Moreover, a purifying process of the instrument is done.

The results obtained in the homogeneity study were high leading to a problem in the code discrimination (see chart number 4). These results show that there are difficulties when discriminating the codes, which means that the behavioral codes are not well differentiated among them. This fact makes us think that the most suitable procedure is the purifying of the instruments which allows us to differentiate and discriminate the codes in an observational instrument. Therefore, a purifying had to be done as the levels in both studies (reliability and homogeneity) weren´t acceptable.

Chart 2

Random Estimation Plan (AC/O)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
A	0.95004					
		O			0.02868	0.9
C	0.78484					
		AO	0.16858	5.3	0.16858	5.3
AC	13.61883					
		OC	0.14717	4.7	0.14717	4.6
		AOC	2.84314	90.0	2.84314	89.2
Sum of variances	15.35371		3.15889	100%	3.18757	100%
Standard deviation	3.91838		Relative SE: 1.77733		Absolute SE: 1.78538	
Coef_G relative	0.83					
Coef_G Absolute	0.83					

Chart 3

Random Optimization Plan by Observers (AC/O)

	Level	Univ	Level	Univ	Level	Univ	Level	Univ	Level	Univ
A	14	INF	14	INF	14	INF	14	INF	14	INF
O	3	INF	4	INF	5	INF	6	INF	7	INF
C	30	INF	30	INF	30	INF	30	INF	30	INF
Observ	1260		1680		2100		2520		2490	
Coef_G relative	0.82937		0.86632		0.89012		0.90672		0.91897	
Coef_G absolute	0.82808		0.86527		0.88923		0.90596		0.91829	

Chart 4

Measurement design for homogeneity OA/C

Source of variance	Differentiation variance	Relative error variance	% Relative	Absolute error variance	% Absolute
Sum of Variances	1.07644	0.59491	100%	0.60750	100%
Standard Deviation	1.03752	Relative SE	0,77130	Absolute SE	0.77942

Coef_G relative = 0.64
Coef_G absolute = 0.64

Study 2

Method

Participants

The participants in the second study were the same as in the first one. However, there were seven observers, two of whom were graduate in Psychology whereas the rest of them were students in Psychology.

Procedure

In order to assure the proper use and knowledge of the observation instrument and guarantee the data registration a ten session training, which lasted a total of thirty hours, was carried out.

Instrument

Observation Protocol in the Teaching Functions (PROFUNDO, v2)

The purifying process, which seems to be necessary due to the results obtained in the first study, was made following two criteria. In the first place, the number of agents turned out to be too high, which affected the reliability. As the main purpose is to concentrate on the behavior of the instructors who lead the program, only three agents are included in the end: the main instructor, the secondary instructor and the participants, being this last group taken as a whole as their behavior is only important due to their interaction with the instructors.

In the second place, a content analysis and a theoretical reflection were done, so as to determine which codes were the most relevant ones. At this stage, ten behavior codes were

crossed out for two reasons: the first one was the uncertainty in the definition of some codes which led to confusion, due to the inability from the observers´ side to tell some of these codes apart when it came down to registering the behavioral patterns. The second one was that in many situations the observers categorized same behavioral patterns in a wrong way as they tended to interpret specific behavioral patterns.

The second version of the instrument is made up by six macro-categories connected to the teaching functions which in turn spread into ten criteria, which lead us to 20 directly observable behavioral codes, and the instrumental category (see chart 5).

Chart 5

Observation Protocol in the Teaching Functions (PROFUNDO), second version

MACRO-CATEGORIES: Functions	CRITERIA	CODES (*)	
1. Organization Function	External organization	Organization of the didactic context	(OD)
		Pupils´organization	(OA)
2. Teaching Communicative Function	Behavioral patterns when leading the activities	Individual explanation	(EI)
		Group explanation	(EG)
3. Motivation Function	Integration in the suggested activity	Instructor`s participation	(PM)
	Encouragement	Individual encouragement	(RI)
		Group encouragement	(RG)
4. Behavioral Control Function	Control resources	Individual negative control resources	(CI)
		Group negative control resources	(CG)
5. Orientation and Assessment Function	Guide, Assesment and Feedback	Individual guide	(GI)
		Group guide	(GG)
	Non verbal homework check-up	Non verbal check-up	(RN)
	Interaction between the instructor and the participants	General Interactions	(IG)
		No anwser	(NR)
6. Interaction Function	Interaction between instructors	Interaction between instructors	(IM)
	Participants` intervention	Answers to the instructors	(RP)
		Addresses to the instructor	(DM)
		Participants disruptions	(DP)
Instrumental Category		Other behavioral patterns	(X)
		Non-visible	(Y)

Note. * These acronyms remain in the Spanish terminology

Data analysis

The programs used for the statistic analysis are the 17.1 version of the **SPSS** and the GT program.

Results

Once the observation instrument was purified we carried out the reliability and homogeneity study. The same rates used in study 1

were calculated and the registers of the six observers were compared to those made by the expert observer. The Kappa rates (Cohen, 1960, 1968) obtained are higher in three of the observers getting rates equal or higher than 0,90 (0.87; 0.93, 0.93; 0.93; 0.83; 0.84). As we did in the previous study, this one was carried out through the GT. The Observation Plan of the reliability consists of three facets, being the following the differentiation ones: twenty behavioral codes, agents with three levels (the main instructor, the secondar the participants) and the generalization measurement object with the observers with seven levels (observers 1, 2,

3, 4, 5, 6 and 7), which lead to the crossed multi-facet observational plan, A × O × C.

A cross facet design is used in the homogeneity study A × O × C. In the measurement plan, the observers 7 levels (o1, o2, o3, o4, o5, o6 and o7) and the agents (with 3 levels, main instructor, secondary instructor and participants) are considered as facet of differentiation, and the Codes as facet of generalization or instrumentation (OA/C). Next, the data related to the estimation plan of the reliability A × O × C is shown in chart number 6. It can be observed that the reliability rate is over 0.90 when using 7 observers.

Chart 6

Random Estimation Plan (AC/O)

Source of variance	Differentiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
A	(0.00000)					
		O			0.19203	4.8
C	5.07239					
		AO	(0.00000)	0.0	(0.00000)	0.0
AC	248.46358					
		OC	0.08801	2.3	0.08801	2.2
		AOC	3.70311	97.7	3.70311	93,0
Sum of variances	253.53596		3.79112	100%	3.98315	100%
Standard deviation	15.92281		Relative SE: 1.94708		Absolute SE: 1.99578	
Coef_G relative	0.99					
Coef_G absolute	0.98					

Later, the optimization plan lets us state that adequate reliable rates can be obtained if the number of observers is lowered (see chart 7).

Chart 7

Random Optimization Plan (AC/O)

	Level	Univ	Level	Univ	Level	Univ	Level	Univ	Level	Univ
A	3	INF	14	INF	14	INF	14	INF	14	INF
O	7	INF	2	INF	3	INF	4	INF	5	INF
C	20	INF	20	INF	20	INF	20	INF	20	INF
Observ	420		120		180		240		300	
Coef_G relative	0.98527		0.95027		0.96629		0.97450		0.97950	
Coef_G absolute	0.98453		0.94788		0.96464		0.97324		0.97848	

The results obtained in the homogeneity study shows that the different behavioral codes are well differentiated as low homogeneity rates are obtained. Therefore, it can be estimated that there is a goodness fit of categories highly significant (see chart 8).

Chart 8

Measurement design for homogeneity OA/C

Source of variance	Differentiation variance	Relative error variance	% Relative	Absolute error variance	% Absolute
Sum of Variances	1.37799	13.80006	100%	14.01598	100%
Standard Deviation	1.17388	Relative SE	3.71484	Absolute SE	3.74379
Relative Coefficient G = 0.09					
Absolute Coefficient G = 0.09					

Discussion and conclusions

The results shown in this study demonstrate the utility of GT when our aim is to purify an observational instrument. At the same time, they let us make important decisions.

The first study shows that the reliability and homogeneity rates are not very satisfactory when using both the Kappa and the GT coefficients. The most important factor in using GT is the fact that it allows us to make important decisions on how to improve these results through the Optimization Plan. In this specific case, the solution was found by increasing the number of observers.

However, this one would not be the only possible solution or the most adequate one. The one shown is this case leads us to a purifying of the instrument as the results of this study show. A reduction was made in both, the number of agents to take part and the number of necessary codes, to successfully collect the instructors´ behaviors which are necessary to codify their performance as teachers in the program. The reduction in codes has let us prove the resources efficiently as the new 20 codes of the instrument let us register the instructors` and the PIPAC program participants` behavior in a clear and reliable way. Moreover, this simplification has let us verify that the behavioral codes are well differentiated among them as significant rates are obtained in the homogeneity study.

It has been possible to determine that the reliability in the different observation sessions reach an acceptable number once the instrument has been simplified. In the following study and after doing the Optimization Plan, it can be concluded that only two observers are more than enough for the observational registers to be reliable. It is also important to highlight that the homogeneity of the codes show a highly significant goodness fit in this second study. This leads us to conclude that the observational instrument has the generalizability, reliability and adjustment requirements needed, which allow the accurate and reliable study of the instructors´ behavior in an out-of-school program.

It is clear the importance to develop systematical instruments that can be used in many contexts and not only in one specific research. It is widely known that the observational methodology usually lacks of the creation of measurement instruments whose use is specific for the research for which it was created but the fact of adjusting observational instrument that can be used in similar contexts to that in which it was created, it will diminish the cost of the research in time and effort. At the same time, the use of the observational methodology for more applied studies like the process evaluation in educative contexts which is the case shown in this study will be more affordable. The interest of making standardized observational instrument which can be used in many different contexts has been demonstrated in

contexts of familiar interaction (Cerezo, 1991, 2000; Trenado & Cerezo, 2004).

Once the final instrument has been purified and the data quality and the homogeneity of the codes have been proved, the behavior of the instructors in the PIPAC program has been registred which enables us to analyze throughly how the program is developed and directed by the instructors. At this stage improvements and modifications can be done by the program assessment basing them on a throughly study of their development (Rodríguez-Naveiras, 2011; Rodríguez-Naveiras, Borges, & Hernández-Jorge, 2010).

On the other hand, and connected to what has been mentioned previously, the instrument presented here, and to the extent in which we can register the instructors` behaviors, can be adapted for its use in other contexts such as intervention programs of psycho-educative characteristics and programs of formal education in different educative levels. This latter one is nowadays being developed (Rodríguez, Díaz, & Cadenas, 2011).

References

Acereda, A. (2002). *Niños superdotados* [*gifted children*]. Madrid, Spain: Pirámide.

Alonso, J. A. & Benito, Y. (1996). *Superdotados*: *Adaptación escolar y social en Secundaria* [*Gifted: Social and Secondary School Adaptation*]. Barcelona, Spain: Narcea.

Anguera, M. T. (1988). *Observación en la escuela* [*Observation in School*]. Barcelona, Spain: Graó.

Anguera, M. T. (2003). La observación [The observation]. En C. Moreno Rosset (Ed.), *Evaluación psicológica. Concepto, proceso y aplicación en las áreas del desarrollo y de la inteligencia* [*Psychological evaluation. Concept, process and implementation in the areas of development and intelligence*] (pp. 271-308). Madrid, España: Sanz y Torres.

Anguera, M. T. (2008). Evaluación de programas desde la metodología cualitativa [*Program Evaluation from qualitative methodology*]. *Acción Psicológica, 5*(2), 87-101.

Arancibia, V., Lissi, M. R., & Narea, M. (2008). Impact in the school system of a strategy for identifying and selecting academically talented students: the experience of Program PENTA-UC. *High Ability Studies.19*(1), 53-65.

Behar, J. & Riba, C. (1993). Sesgos del observador y de la observación [Observer and Observation bias]. En M. T. Anguera (Ed.), *Metodología Observacional en la investigación psicológica (Vol. 2, Fundamentación 2*, pp. 11-148). Barcelona, Spain: PPU.

Blanco, A. (1993). Fiabilidad, precisión, validez y generalización de los diseños observacionales [Reliatbility, accuracy, validity and generalization of observational designs]. En M. T. Anguera (Ed.), *Metodología Observacional en la investigación psicológica* (Vol. 2 Fundamentación, pp 151-261). Barcelona, Spain: PPU.

Blanco, A. & Anguera, M. T. (2000). Evaluación de la calidad en el registro del comportamiento: Aplicación a deportes de equipo [Quality assessment in the registration behavior: Application to Team Sports]. En E. Oñate, F. García-Sicilia and L. Ramallo (Eds.), *Métodos numéricos en Ciencias Sociales* (pp. 30-48). Barcelona, Spain: Centro Internacional de Métodos Numéricos en Ingeniería (CIMNE).

Blanco, A., Castellano, J., & Hernández Mendo, A. (2000). Generalizabilidad de las observaciones en la acción del juego en el fútbol. *Psicothema*, *12*(2), 81-86.

Blanco, A., Sastre, S., & Escolano, E. (2010). Desarrollo ejecutivo temprano y Teoría de la Generalizabilidad: bebés típicos y prematuros. *Psicothema, 22*(2), 221-226.

Borges, A., Rodríguez-Naveiras, E., Hernández-Jorge, C., & Fernández, R. (2007, June). *Programa Integral para Altas Capacidades: Evaluación Formativa 2004-006* [*Comprehensive Program for High Capacities: Formative Assessment 2004-2006*]. Comunicación presentada en la IX Conferencia Española y I Encuentro Iberoamericano de Biometría, Salamanca.

Borges, A., Hernández-Jorge, C., & Rodríguez-Naveiras, E. (2011). Evidencias contra el mito de la inadaptación de las personas con altas capacidades intelectuales [Evidence against the myth of the maladjustment of people with high intellectual abilities]. *Psicothema, 23*, 362-367.

Borges, A., & Rodríguez-Naveiras, E. (2012). Programas de intervención en altas capacidades intelectuales y su evaluación [Intervention programs in high intellectual abilities and their assessment]. En D. Valadez (Ed.), *Alumnos superdotados y talentosos, identificación, evaluación e intervención. Una guía para docentes.* México: Manual Moderno.

Cerezo, M. A. (Ed.). (1991). *Interacciones familiares: Un sistema de evaluación observacional- SOC-III [Family interactions: An observational assessment system-SOC-III].* Madrid, Spain: Mepsa.

Cerezo, M. A. (2000). *Interacción familiar. Un sistema de evaluación observacional. SOC III edición bilingüe (CD-ROM) [Family interaction. An observational assessment system. SOC III bilingual].* Valencia, Spain: Publicaciones de la Universidad de Valencia.

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: Wiley.

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.

Díaz, O. & Bravo, A. (2002). Programa de Enriquecimiento de ASAC (Asociación de altas capacidades) [ASAC Enrichment Program (gifted Association)]. *Faisca. Revista de Altas Capacidades*, *9*, 11-125.

Gil, A. (2008) *Los porteros de Fútbol, ¿se comportan como sistemas complejos?* (Tesis Doctoral). Universidad de Barcelona.

Heller, K. A. (2009). Gifted education from the german perspective. In D. J. Matthews (Ed.), *The routledge international companion to gifted education* (pp. 61-67). New York, NY US: Routledge/Taylor y Francis.

Heller, K. A. & Reimann, R. (2002). Theoretical and methodological problems of a 10-year follow-up program evaluation study. *European Journal of Psychological Assessment, 18*(3), 229-241.

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

Ibáñez, I. (1993). La observación. En W. Peñate, P. Matud, I. Ibáñez (Eds.), *Evaluación Psicológica: conceptos y técnicas de análisis* (pp. 225-260). Valencia: PROMOLIBRO (Promoción del Libro Universitario).

López, V. & Sotillo, M. (2009). Giftedness and social adjustment: Evidence supporting the resilience approach in Spanish-speaking children and adolescents. *High Ability Studies, 20* (1), 39-53.

López de la Llave, A. & Pérez-Llantada, M. C. (2004). *Evaluación de programas en psicología aplicada(salud, intervenciones sociales, deporte, calidad)*. Madrid, Spain: Dykinson

Máiquez, M. L., Blanco-Villaseñor, A., Rodrigo, M. J., & Vermaes, I. (2000). La evaluación de la eficiencia en la intervención familiar: generalizabilidad y optimización del programa experiencial para padres. Psicothema, 12(4), 533-542.

Neber, H., & Heller, K. A. (2002). Evaluation of a summer-school program for highly gifted secondary-school students: The german pupils academy. *European Journal of Psychological Assessment, 18*(3), 214-228.

Newcomer, K., Hatry, H. & Wholey, J. (2010). Planning and designing useful evaluations. In, Wholey, J., Hatry, H. & Newcomer, K. (2010). *Handbook of practical program evaluation* (pp. 48-73). an Francisco: JOSSEY-BASS A Wiley Imprint.

Pedrosa, I. & Borges, A. (2012, july). *Study of children's interaction within the Comprehensive Program for Highly Gifted Students (PIPAC)*. V European Congress of Methodology. Santiago de Compostela.

Pereira, N., Peters, S. J. & Gentry, M. (2010). My class activities instrument as used in Saturday enrichment program evaluation. *Journal of Advanced Academics, 21*(4), 568-593.

Pfeiffer, S. I., & 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.

Prieto, M. D. & Ferrándiz, C. (2001). *Inteligencias Múltiples y curriculum escolar*. Málaga, Spain: Aljibe.

Quera, V. (1991). Muestreo y registro observacional. En M. T. Anguera (Ed.), *Metodología observacional en la investigación psicológica* (Vol. 1, pp.241-328). Barcelona, Spain: PPU.

Quera, V. & Behar, J. (1997). *La observación*. En, G. Buela-Casal & J. C. Sierra (Dirs). (1997). *Manual de evaluación psicológica. Fundamentos, técnicas y aplicaciones [Handbook of psycholocial evaluation. Foudations, techniques and applications]*. Madrid, Spain: Siglo Veintiuno.

Renzulli, J. S. (1977). The enrichment triad model: A plan for developing defensible programs for the gifted and talented: II. *Gifted Child Quarterly, 21*(2), 227-233.

Rodríguez-Naveiras, E; Borges, A & Hernández-Jorge, C. (2007, Febrery). *Problemas metodológicos en la evaluación de programas repetidos con los mismos sujetos: resultados de tres años de implementación del programa integral para altas capacidades*. Comunicación presentada en el X Congreso de Metodología de las Ciencias Sociales y de la Salud, celebrado en Barcelona.

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

Rodríguez-Naveiras, E; Borges, A., & Hernández-Jorge, C. (2010, july). *Protocolo de observación de funciones docentes (PROFUNDO) para la evaluación de proceso de la conducta de las monitoras del Programa Integral para Altas Capacidades*. Comunicación presentada en el VII Congreso Iberoamericano de Psicología celebrado en Oviedo.

Rodríguez, M., Díaz, M., & Cadenas, M. (2011). Estimación del tiempo de registro óptimo para la observación de las funciones docentes del profesor universitario. *Revista de Investigación y Divulgación de Psicología y Logopedia, 1*(1),10-15.

Trenado, R. M. & Cerezo, M. A. (2004): *Sistema de codificación de la interacción temprana materno-infantil, CITMI-R*. Documento no publicado, Universidad de Valencia.

Wholey, J. S., Hatry, H. P., & Newcomer, K. E. (2010) *Kandbook of practical program evaluation*. San Francisco, CA: Sage.

Ysewjin, P. (1996). *Software for generalizability studies*. Mimeografía.