

Caregivers' Agreement and Validity of Indirect Functional Analysis: A Cross Cultural Evaluation Across Multiple Problem Behavior Topographies

Javier Virues-Ortega · David Segui-Duran ·
Alberto Descalzo-Quero · Jose Julio Carnerero ·
Neil Martin

© Springer Science+Business Media, LLC 2010

Abstract The Motivation Assessment Scale is an aid for hypothesis-driven functional analysis. This study presents its Spanish cross-cultural validation while examining psychometric attributes not yet explored. The study sample comprised 80 primary caregivers of children with autism. Acceptability, scaling assumptions, internal consistency, factor structure, inter-assessor reliability and agreement, and known-group validity analyses were performed. Scaling assumptions, internal consistency (Cronbach alpha of 0.75) and factor structure were satisfactory other than for the Escape domain which demonstrated low internal consistency (0.65), inadequate scaling assumptions (multitrait analysis, 50% success rate) and did not constitute a separate factor. Caregivers' agreement for the primary function reached 73.9% and known group-validity hypotheses

across behavior topographies were partially met. The clinical appropriateness of the scale is discussed.

Keywords Motivation assessment scale · Autism · Functional analysis · Indirect methods

Introduction

Functional analysis methodology aims to assess the motivation (function) of problem behavior in, primarily, individuals with developmental disabilities. Identifying the function(s) of self-injurious, aggressive, stereotypic and other forms of problem behavior is crucial to the development of effective treatment strategies (Herzinger and Campbell 2007). The functional analysis literature suggests that the most common functions of problem behavior in developmental disabilities are: (a) social attention: social responses that follow the behavior (e.g., verbal reprimands, statements of concern, other forms of social attention and

Authors are indebted to the Autism Association of Málaga (Asociación Autismo Málaga) for volunteering their resources for this project. Dr. Rafaela Caballero (Hospital Universitario Virgen de la Macarena, and University of Seville) facilitated data collection in Seville.

J. Virues-Ortega (✉)
ABA España, Apto. 44068, 28029 Madrid, Spain
e-mail: jvortega@cienciaconducta.com; jvirues@isciii.es

D. Segui-Duran
Children Mental Health Unit, Hospital Universitario Carlos Haya, Málaga, Spain

A. Descalzo-Quero
Children Mental Health Unit, Hospital Universitario Virgen de las Nieves, Granada, Spain

J. J. Carnerero
Centro Al-Mudaris, Córdoba, Spain

N. Martin
European Association for Behaviour Analysis, London, UK

Present Address:
D. Segui-Duran
Community Mental Health Unit, Hospital de Antequera, Málaga, Spain

Present Address:
A. Descalzo-Quero
Community Mental Health Unit, Hospital de Baza, Granada, Spain

contact); (b) escape or avoidance: the problem behavior is effective in terms of escaping or avoiding work demands or other forms of aversive stimulation; (c) tangible: the problem behavior increases the likelihood of accessing preferred items; and (d) sensory: the behavior produces sensory stimulation for the individual (e.g., tactile, kinesthetic, visual) (Iwata et al. 1982/1994). Various methods have been developed to detect which of these functions may be the most relevant in a given case.

There are three general approaches to functional assessment: indirect assessment, descriptive analysis and functional analysis. Descriptive and functional analyses both require systematic observation and recording of specific environmental and behavioral events. Descriptive analysis simply correlates behavior with naturally occurring events in the environment (e.g., attention by caregivers, escape from academic demands). However, during a functional analysis, the operant contingencies are contrived through several analogue conditions so that the environmental causes of problem behavior may become more apparent (Iwata et al. 1982/1994).

Operant theory suggests that problem behavior, like other learned behavior, can be tied to specific antecedent and consequent events through a cause-effect relationship. Therefore, if the environmental conditions most commonly associated with specific behaviors are identified then they may be arranged as different conditions to experimentally test whether specific conditions are more reliably associated with some problem behaviors than others. If a specific behavior pattern is found to covary consistently with one of these conditions, it may be concluded that a relevant function of a given behavior has been identified. This type of experimental functional analysis allows the detection of the motivation of problem behavior in a very high percentage of cases (Iwata et al. 1994a, b). In addition, functional analysis outcomes lead to function-specific approaches to treatment, which have been found to be more effective than interventions based on other forms of behavioral assessment (Herzinger and Campbell 2007).

Functional analysis is a complex methodology and practitioners may not have the resources available (in terms of training, assessment time, access to a controlled setting etc.) to fully implement such an assessment. Descriptive analysis, on the other hand, provides correlational evidence with low specificity and sensitivity when weighed against functional analysis (Thompson and Iwata 2007; Hall 2005). As a result, indirect methods have evolved as a third approach to the functional assessment of problem behavior.

Indirect methods of assessment are frequently presented in the form of questionnaires that systematically pose questions related to the more frequent functions of problem behavior. Although indirect methods are less precise than experimental functional analysis, their level of

correspondence with functional analysis has been found to be superior to descriptive analysis (Hall 2005). Specifically, a number of studies have shown the level of correspondence between indirect assessments and functional analysis to be above 75% (Hall 2005; Durand and Crimmins 1988; Wasano et al. 2009). In addition, indirect methods of assessment may serve several functions themselves during assessment. First, they point to the likely behavioral functions that may be later tested within a functional analysis. Second, they provide a systematic way of collecting information on problem behavior from parents, teachers and caregivers. Although, some indirect methods were initially developed to study self-injurious behavior, this approach has been extended to apply to various forms of aggressive, stereotypic and disruptive behavior (see Table 1). However, several problems with indirect methods have been reported including low inter-rater reliability (Zarcone et al. 1991) and poor performance with regard to low-frequency behavior (Paclawskyj et al. 2001).

Several methods for the indirect assessment of behavioral functions have been developed including the Functional Analysis Screening Tool, FAST (Iwata 1996) and Questions about Behavioral Function, QABF (Paclawskyj et al. 2000). However, the Motivation Assessment Scale, MAS (Durand and Crimmins 1988) is the one that has been researched more extensively—the QABF has significantly fewer studies exploring its psychometric attributes and FAST validation has not yet been published. Psychometric studies comparing the MAS with the QABF have shown that both instruments are psychometrically comparable (Paclawskyj et al. 2001; Shogren and Rojahn 2003). However, criterion validity with experimental functional analysis was slightly superior for QABF as compared to the MAS in the study by Paclawskyj et al. (coincidence: 38.5% vs. 53.9%). The authors have not been able to determine whether any other method of functional assessment has been translated into Spanish and validated (personal communication, ABA Spain Executive Board, affiliated chapter of the Association for Behavior Analysis International in Spain). Given the existing research already conducted with the MAS, and the fact that it is already being used in several applied programs in Spain, the MAS was selected for validation in this study.

The MAS was developed to assess prevalent causes of problem behavior through the assistance of third-party respondents. The scale has 16 items, rated over a 7-point Likert scale, and four items for each of four domains related to a frequent function of problem behavior. For example, a question designed to test whether the behavior was maintained by Sensory consequences is “Would the behavior still occur if there were nobody around?”, to test a possible Escape function, “Does the behavior frequently

Table 1 Psychometric properties of the motivation assessment scale in the literature

	N	Topography	Reliability				Validity		
			Cronbach alpha	Agreement (%) ^a	Inter-rater ^b	Test–retest	Construct ^c	Convergent ^d	Criterion % (n) ^e
Present study	80	Various	0.65–0.82	73.9	0.37–0.61	–	3 (AI, A, T)	–	–
Bihm et al. 1991	118	Various	–	–	–	–	4 (AI, E, A, T)	–	–
Duker and Sigafoos 1998	86	Various	0.68–0.87	66.7	–0.19 to 0.87	–	4 (E, T)	–	–
Durand and Crimmins 1988	50	Self-injury	–	–	0.80–0.95	0.82–0.99	–	–	100.0 (8)
Freeman et al. 2001	91	Various	0.67–0.87	–	–	–	–	0.66–0.76	–
Joosten and Bundy 2008	67	Stereotypic	–	–	–	–	2 (AI)	–	–
Kearney 1994	42	Various	–	–	0.35–0.82	–	–	–	–
Kearney et al. 2006	315	Various	0.71–0.85	–	–	–	3 (E, A, T)	–	–
Paclawskyj et al. 2001	13	Various	–	–	–	–	–	0.51–0.86	38.5 (13)
Shogren and Rojahn 2003	31	Various	0.80–0.96	–	0.35–0.73	0.71–0.88	–	0.73–0.89	–
Sigafoos et al. 1994	18	Aggressive	–	44.4	–0.67 to 0.72	–	–	–	–
Singh et al. 1993	60	Self-injury	–	–	–	–	4 (AI, E, A, T)	–	–
Singh et al. 1993	96	Self-injury	–	–	–	–	3 (AI)	–	–
Sprent and Connelly 1996	47	Self-injury	0.68–0.84	70.0	0.31–0.57	–	–	–	–
Wasano et al. 2009	3	Pica	–	–	–	–	–	–	100.0 (3)
Zarcone et al. 1991	55	Self-injury	–	29.1	~0.41 ^f	–	–	–	–

A Attention; AI Alone; E Escape; T Tangible

^a Main function identification

^b Range or mean Spearman or Pearson correlation coefficients across raters for the four scale domains

^c Number of factors and corresponding theoretical domains in parenthesis

^d Correlation with Questions About Behavior Function

^e Main function agreement with experimental functional analysis

^f Interpolated from the number of participants across 0.1 correlation intervals. Full references available in the reference section

follows when a difficult task is requested from him?”, for Attention, “Is the behavior more likely when he is no longer receiving attention?”, and for Tangible (i.e. the behavior gains access to tangible items), “Does the behavior stops when you give him a preferred toy or activity?”. Functional analysis research has suggested that these four functional domains are the most frequent motivational functions of problem behavior.

The MAS scale’s reliability and validity have been explored in previous studies (Table 1). However, specific features of the scale related to its clinical appropriateness and usability have not yet been investigated. These include specific aspects of scaling assumptions (multitrait analysis), parental agreement in the identification of behavioral function, known-group validity across behavior topographies, and cross-cultural validity. The cross-cultural validation of the Spanish version of the MAS will be useful given the large population of potential users as suggested by the number of Spanish native speakers (~400,000,000; Central Intelligence Agency 2009). It will also provide the opportunity for the examination of the psychometric attributes not yet explored, including acceptability, scaling

assumptions, and agreement with regard to the main identification of function across caregivers. These features will indicate the extent to which the scale has been designed properly and provide reliable information for use in clinical rather than research settings. Comparability with the original version is ensured by means of the following two methods: (a) use of translation and back-translation, (b) implementation of general statistical procedures (including reliability and construct validity analyses) that have been used in previous psychometric studies.

Methods

Participants

The sample was comprised of 80 children and adolescents with developmental disabilities (95% male; mean age 4.4 years; range 3.5–17.0 years). The participants were those who had been offered services from three outpatient mental health units located in Granada, Malaga and Seville (in Southern Spain), and from an applied behavior-analytic

program in Cordoba (also in Southern Spain) between June 2006 and June 2008. All participants lived at home and were referred to the outpatient units for assessment by a primary care physician, or to the applied behavior analysis (ABA) program for behavioral treatment by their parents. Participants were from families of middle-low socioeconomic status with the exception of participants from the ABA program that were from middle-high socioeconomic status, as estimated from parental education and professional background. Participants at the ABA program received discrete-trial teaching, incidental teaching and language training programs based on Skinner's Analysis of Verbal Behavior (Greer and Ross 2007; Maurice et al. 2001). The inclusion criteria for participation in this study were: (a) a confirmed diagnosis of autism or pervasive developmental disability as diagnosed by an independent licensed psychiatrist or psychologist following DSM-IV TR criteria (American Psychiatric Association 2000), (b) presence of a behavior problem in the form of disruptive, aggressive, stereotypic or self-injurious behavior with an history of at least one year according to caregiver's report, (c) caregivers spending no less than 10 h weekly with the participant, and (d) caregivers were able to answer questions and they were willing to participate. Individuals with a concurrent sensory disability (e.g., blindness) or any other axis I diagnosis (e.g., Rett syndrome) were excluded. In addition to a diagnosis of autism (88.2% of sample) or pervasive developmental disabilities not otherwise specified (11.8% of sample), 55.0% of participants also had a diagnosis of mental retardation (18 mild, 13 moderate, 11 severe). Axis II diagnosis was not associated to behavior problem category (Kruskal–Wallis, $p > .1$).

Procedure

Upon admission to one of the services described above, caregivers were invited to participate in the study. Participating families signed an informed consent form and met with an independent licensed psychologist who had seen the child and who had access to the clinical records. The purpose of the meeting was to agree on the problem behavior of greatest concern for the caregivers and to appropriately define the behavior where further clarification was needed (e.g. "bad temper", "being upset all the time" etc.). In order to avoid bias, once the target behavior was established, an independent rater administered the MAS asking only for minor clarifications based on the instructions and item descriptors given. The rater ensured that the intent of the item was clear to the caregiver and when two caregivers were available, they were asked the questions in different rooms and by different raters.

Psychometric Analyses

Psychometric methods are listed below. Published standard criteria have been added in parenthesis to aid the interpretation of results:

1. *Cross-cultural adaptation* A modified version of the translation/back-translation method by Bullinger et al. (1998) in addition to the guidelines by the International Test Commission (2000) were followed. It should be noted that different language versions of a scale that has been cross-culturally adapted successfully shall be considered equivalent. There is an extensive literature and work conducted upon that premise (e.g., EuroQoL group 2010; World Health Organization 2001). A forward translation of the original U.S. English version into Spanish was conducted by two translators who were native Spanish speakers. The translators had experience of translating scientific and psychometric terminology but were not familiar with the questionnaire. The translators were asked to place emphasis on conceptual rather than literal equivalence with a level of comprehension at primary education reading level. Translators rated the difficulty of translating each item from 0 (not at all difficult) to 10 (extremely difficult) and wrote comments if necessary. The two translators met with one of the authors (ADQ) to agree on a common version and differences were debated until a consensus was reached. One of the authors with extensive experience of scientific translation (JVO) conducted the final backward-translation based on two preliminary versions prepared by two independent English-speaking professional translators. The conceptual equivalence between the original version and the final backward-translation was reviewed by a professional translator who did not have access to the original version before this point. Discrepancies were noted and the forward-translation was revised as appropriate. Finally, the forward-translation was pilot tested with a group of six participants exhibiting different problems behavior using two raters. After the assessment, raters and respondents were asked if they found any items difficult or confusing. Difficulties reported were noted and revisions were conducted as needed. Specifically, the document was reformatted to make it visually more attractive and readable (e.g. the font size was increased). Minor changes in word selection were also made to provide accessibility to low-education respondents. Additional details on the cross-cultural adaptation processes are available upon request.
2. *Data quality and acceptability* The percentage of total computable scale scores informs the quality of the data

set by summarizing the proportion of subjects for whom assessment results were usable (using a criterion of >90%; Martinez-Martin in press). Floor and ceiling effects indicate the percentage of subjects having the lowest and highest possible scores respectively. Calculating floor and ceiling effects inform the extent to which the complete range of variation of the construct is captured by the scale or is otherwise skewed at one end or another (Martinez-Martin in press). Skewness (criterion: ± 1) is an index of lack of symmetry in a data set (Mardia 1970).

3. *Scale assumptions* An item-total corrected correlation is calculated by correlating the individual items scores with the scale total and subtracting the score of the item being analyzed. Moderate values for these data would indicate that items are sensitive to different aspects of the same construct with criterion values for $r \geq 0.30$ (Nunnally and Bernstein 1994). Multi-trait analysis determines the extent to which items correlate more strongly with their own domain (in this case, associated hypothesized function) than with other items in the scale. Differences in excess of twice the typical error of the correlation coefficient ($2/\sqrt{n}$) indicate the strength of the association (i.e. successful association). The success rate was computed as the percentage of successes across all possible item-domain correlation pairs corrected for overlap. Although there is no specific threshold to determine if the success rate is adequate, higher rates ought to be associated with distinctive and well-designed domains within the scale (Ware and Gandek 1998).
4. *Internal consistency* Internal consistency was measured using Cronbach's alpha (with a criterion of >0.70; Nunnally and Bernstein 1994), and by means of an item's homogeneity index. Item homogeneity was calculated from the mean of the inter-item correlations (using a criterion of ≥ 0.30 ; Eisen et al. 1979). Although both item homogeneity and Cronbach's alpha are correlated, they relate to different aspects of internal consistency. Whilst a high Cronbach's alpha indicates high total score variability accompanied by low intra-item variability, item homogeneity is weighted more heavily in terms of the items' similarity.
5. *Agreement across caregivers* The subsample where both caregivers (typically parents) were able to complete the MAS for their child's problem behavior was selected incidentally. In other words, this subsample was formed simply as a result of the participant being accompanied by both caregivers who were both willing to participate in the study. In order to warrant comparability of results with other previous studies of the interrater agreement of such indirect methods of functional assessment, the following methods were used. First, the percentage of agreement of the primary function identified was calculated using the standard of agreement for observational studies (criterion of $\geq 80\%$; Kennedy 2005). Second, agreement of the identification of the main function irrespective of random variation was calculated using Cohen's kappa with a criterion of ≥ 0.60 (Martinez-Martin in press). Third, Kendall's tau-b correlation for the rank-ordered functions was calculated with moderate correlations expected (range: 0.30–0.50, see Juniper et al. 1996). Fourth, Spearman's correlation coefficients for each functions total score were calculated, again with moderate correlations expected (criterion range of 0.30–0.50). The two last approaches, although less restrictive, do not require exact correspondence for the agreed primary function for the scale scores to correlate.
6. *Factor structure* Exploratory factor analysis was performed using the principal components method. Adequacy of the correlation matrix was studied using Bartlett's statistic (criterion, $p \leq 0.05$; Dziuban and Shirkey 1974) and the Kaiser–Meyer–Olkin test (KMO; criterion, ≥ 0.5 ; Lorenzo-Seva and Ferrando 2007). Parallel analysis was used to determine the number of dimensions to be retained (Horn 1965). The resulting model was submitted to an unweighted least square factor analysis plus varimax rotation. Loading per item (criterion, >0.30; Lorenzo-Seva and Ferrando 2007) and factor reliability estimates (criterion, ≥ 0.70 ; Mislavy and Bock 1990) were examined. Goodness of fit of the resulting model was assessed by means of the mean fit residuals (MFR, criterion: $< 1/\sqrt{n}$; Lorenzo-Seva and Ferrando 2007) and the root mean square residuals (RMSR, criterion: <0.06; Lorenzo-Seva and Ferrando 2007). In addition, goodness of fit of the original theoretical model was also assessed. Goodness of fit parameters help interpretation of the extent to which the data follows a pre-specified theoretical model, whether based on previous analyses or on the theoretical domains of the scale.
7. *Known-group validity* Known-group validity is a strategy that indirectly assesses the validity of a scale by demonstrating that scale scores vary systematically depending on known performances of the construct that it is intended to measure (Netemeyer et al. 2003). Domains scores were compared across four groups of problem behavior topographies—disruptive, aggressive, self-injurious and stereotypic behaviors. Disruptive behaviors included screaming and crying. Aggressive behaviors included property destruction, throwing and banging objects and aggression towards others (pushing, hitting, biting etc.). Self-injurious

behavior included biting, hitting and banging oneself. Stereotypic behavior encompassed perseverative and inflexible motor and vocal verbal behavior. Two of the authors coded each behavior problem in each category (JVO, DSD) with 94.2% agreement. In order to test known group validity, review articles summarizing large databases involving functional analysis were used to derive hypotheses on prevalent functions of behavior problem topographies (Hanley et al. 2003; Herzinger and Campbell 2007; Iwata et al. 1994a, b). From these it was hypothesized that the MAS would be likely to attribute a Sensory domain function for stereotypic behavior, and an Escape domain function for aggressive behaviors. These hypotheses were tested using non-parametric comparisons (Wilcoxon test).

All analyses were conducted using SPSS version 13 (LEAD Technologies, Inc., Charlotte, NC) and Factor version 7 (Lorenzo-Seva and Ferrando 2007) software.

Results

After minor wording changes, the original and back-translated versions were considered equivalent (the standard Spanish version is available upon request).

Table 2 presents a summary of the psychometric analyses. All data quality and acceptability measures, including total computable scores, floor, ceiling effects and skewness, were within the usable range. All corrected item-total

correlations were within the specified criterion limits. Multitrait analysis showed that most items correlated with their own domain function above other domains, further supporting scaling assumptions. However, multitrait analysis suggested a low rate of correspondence for the Escape domain. Specifically, the Escape domain was not clearly distinguishable from the Tangible and Attention domains with correlation ranges for the Escape and tangible items of 0.20–0.56, Escape and Attention items of 0.09–0.32, and Escape with Escape items of 0.31–0.56.

Internal consistency of the total scale and item homogeneity by domain were within the expected range (Alpha of 0.75 and item homogeneity between 0.37 and 0.53). Cronbach's alpha varied significantly across domains with a range of 0.65–0.82 (see Table 2). However, when Cronbach's alpha was analyzed by subscale, the Escape domain was shown to fall slightly below the criterion (0.65).

Interrater reliability analyses were based on 57 independent assessments completed by secondary caregivers (45 fathers, 10 mothers, 2 formal caregivers). The distribution of problem behavior topographies was similar to that of the complete sample (18 disruptive, 12 aggressive, 15 self-injurious and 12 stereotypic topographies). Most secondary assessments were conducted in the Malaga outpatient mental health units (40 in Malaga, 4 in Granada, 7 in Seville, 6 in Cordoba). In order to prevent the inflation of the agreement data, seven cases with the same score for two or more functions were discarded; therefore agreement analyses were based on 50 cases. Percentage of agreement in the

Table 2 Psychometric attributes of the motivation assessment scale

	Criterion	Alone	Escape	Attention	Tangible	Total
Acceptability						
TCS (%)	≥90	91.25	98.75	98.75	100.00	93.75
Skewness	±111	0.12	0.44	0.85	0.09	–
Floor effect (%)	<15	2.50	11.25	13.75	3.75	–
Ceiling effect (%)	<15	0.00	0.00	0.00	1.25	–
Scaling assumptions						
Corrected item-total ^{a,b}	≥0.20	0.42–0.67	0.31–0.56	0.53–0.57	0.60–0.69	–
Multi-trait analysis (%SR)	≥80	91.67	50.00	83.33	83.33	–
Internal consistency						
Cronbach alpha	≥0.70	0.75	0.65	0.75	0.82	0.75
Items homogeneity	≥0.30	0.42	0.37	0.43	0.53	–
Inter-rater agreement ^b						
% agreement	≥80	65.0	75.0	50.0	68.2	73.9
Cohen kappa	≥0.60	–	–	–	–	0.57
Spearman <i>r</i>	≥0.30	0.61	0.48	0.37	0.42	0.45
Kendall tau-b	≥0.30	0.48	0.46	0.28	0.46	–

TCS Total computable scores; %SR Rate of successes

^a Spearman *r*

^b All Spearman *r* and Kendall tau-b rank correlation coefficients were significant ($p < 0.01$)

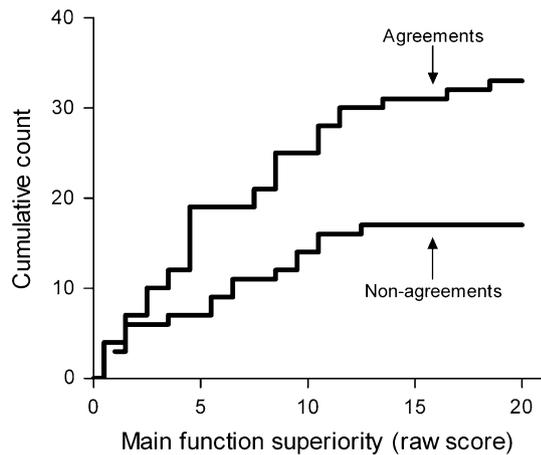


Fig. 1 Cumulative observer agreements and non-agreements in main function identification across the range of main function superiority (score units above second function)

identification of the primary function reached 73.9%, with some variation across functions (see Table 2). Figure 1 shows the number of cases of agreement and non-agreement accumulated linearly with no clear impact of how distinctive the primary function was compared to the secondary (i.e., score units above the second function identified). This suggests that the scale results may be interpretable even in cases where the primary function is not highly distinctive, which adds to the utility of the scale. Cohen’s kappa for interrater agreement was only 0.57. Kendall’s tau rank correlation indicated only a moderate correspondence between the behavioral functions ranked across informants (Kendall tau-b range 0.28–0.48). Finally, Spearman’s correlations between respondent scores ranged from moderate to high (range 0.37–0.61). Measures based on agreement (percentage of agreement, Cohen kappa) were slightly below the standard criteria while measures based on correlation (Spearman’s *r*, Kendall tau-b) were above the suggested criteria (see Table 2). Sensory, Escape and Tangible domain functions had the highest scores; whereas the Attention domain had the poorest in all estimates (Table 2).

Due to the excess of kurtosis, polychoric correlations were used as the basis for factor analysis (Muthen and Kaplan 1985). Barlett’s statistic ($p < 0.001$) and KMO test (0.80) suggested the correlation matrix was adequate. An initial principal components factor analysis plus parallel analysis suggested that three factors should be retained. A subsequent unweighted least squares factor analysis and varimax rotation was conducted forcing the model to three factors. The final three-factor solution accounted for a cumulative variance of 47.4% with reliability estimates of 0.77, 0.75 and 0.64 for Factors 1, 2 and 3 respectively. Items of the rotated matrix were examined for coincidence with the theoretical structure of the scale (see Fig. 2). Factors 1, 2 and 3 tended to match items for the domains of Attention,

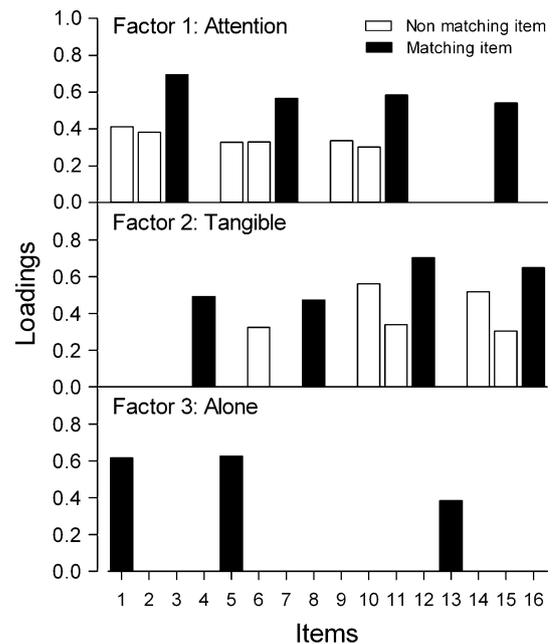


Fig. 2 Item loadings for the three-factor solution of the MAS. Items in Factors 1, 2 and 3 tend to match items of Attention, Tangible and Alone domains respectively. Loadings with absolute value lower than 0.300 have been omitted

Tangible and Sensory, respectively. Items for the Escape domain did not form a distinctive factor; on the contrary, they tended to combine with those of the Tangible domain in Factor 2. Goodness of fit of the three-factor model was adequate as assessed by the MFR (< 0.0001) and the RMSR (0.05). Whereas goodness of fit of the four-factor model was adequate (MFR < 0.0001 ; RMSR = 0.03), only the first two factors matched the theoretical domains of the scale (Tangible and Attention). Confirmatory factor analysis for the theoretical four-factor structure was not sustained.

Hypotheses regarding specific associations between the type of problem behavior and the function of that behavior were partially met (see Fig. 3). Participants with stereotypic behavior scored significantly higher in terms of the Sensory domain (Wilcoxon, $p < 0.001$). However, Tangible, rather than Escape, was the single behavior function scoring significantly higher for participants with aggressive behavior. Participants exhibiting self-injurious behavior scored significantly higher with regard to the Tangible domain compared to the Attention or Sensory domains (Wilcoxon, $p < 0.05$) although this was not one of the hypotheses specifically made.

Discussion

The psychometric attributes of the cross-culturally adapted version of the MAS were shown to be within or above the

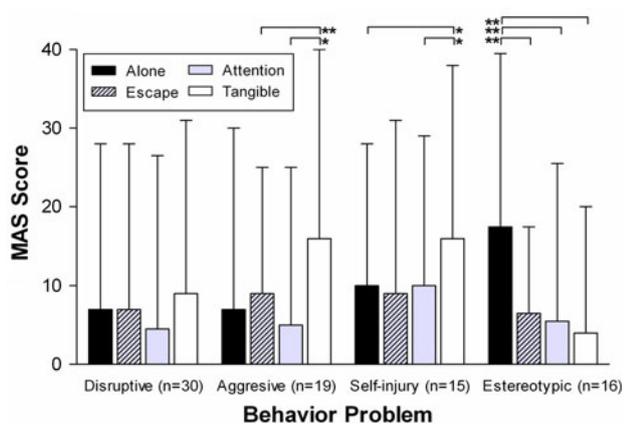


Fig. 3 MAS domains scores (median and rank) across behavior topographies. * $p < 0.05$; ** $p < 0.001$ (Wilcoxon)

range of the results observed in previous studies. This was true for internal consistency (Cronbach's alpha), interrater agreement (percentage agreement and Cohen's kappa), and Spearman's and Kendall's tau-b correlation coefficients across raters (see Table 1). These results suggest that the cross-validation was successful (in terms of replication) despite the fact that the performance of the original scale on specific psychometric attributes is far from optimal, and that alternative approaches to functional assessment may be superior with regard to those attributes (e.g., Cronbach alpha for QABF, cf. Nicholson et al. 2006). In addition, this study is the first to report specific scaling assumption features (multitrait analysis), caregiver agreement on the identification of behavioral functions and known-group validity across different behavior topographies.

Acceptability and scaling assumptions were satisfactory. Internal consistency by means of Cronbach's alpha, although in line with previous studies and rendering the scale suitable for research use, did not meet the highest standards for clinical use (Scientific Advisory Committee of the Medical Outcomes Trust 2002). Although agreement across assessors did not meet the suggested criterion of 80%, it was the highest reported so far for this instrument in an independent validation (see Table 1).

Figure 1 shows that cases of agreement and non-agreement accumulated linearly, regardless of how much the primary function scored above the second. This suggests that the scale results may be interpretable even in cases where the primary function is not highly distinctive. This finding adds to the usability of the scale. Both factor analysis and known group validity analyses were partially successful as three of the four theoretical domains were identified in the factor structure and one of the two hypotheses appeared to be confirmed.

A major shortcoming to the clinical usability of this scale had to do with the poor performance of the Escape domain. Moreover, while reliability of the Attention

domain was adequate, Attention was seldom scored as the primary function (8.8% of cases), which was also the case for Escape (6.3% of cases). A potential underreport of these functions becomes apparent when the distribution of behavior functions in our study is compared with that of studies summarizing large databases of experimental functional analyses. For example, Hanley et al. (2003) report Escape as the primary function for 34.2% of cases and Attention for 25.3%. The fact that the function of Attention was likely to have been underreported, but still demonstrated to be a key factor based on the psychometric profile, suggests that caregivers may be systematically and unsystematically biased against the Attention and Escape functions respectively.

Several studies have also found limited internal consistency and construct validity with regard to the Escape domain of the MAS (e.g., Freeman et al. 2001; Joosten and Bundy 2008; Spreat and Connelly 1996). Zarcone et al. (1991) found the Escape function reported as the primary function in only 6.3% of the cases, which also suggests a potential underreporting effect. This limitation in the performance of the Escape domain may explain the lack of correspondence between experimental functional analysis and other indirect methods. Of note, all non-matching outcomes between indirect assessments and experimental functional analysis reported by Paclawskyj et al. (2001) were attributable to a misidentification of the Escape function. More research involving both experimental functional analysis and indirect assessment methods is required to determine if indirect assessments are systematically biased against escape- and attention-maintained behavior. Any straightforward recommendation concerning the usability of the scale would be dependent on the outcome of that research. Meanwhile interventionists should show a higher degree of vigilance on items pertaining to the Escape domain (items 2, 6, 10, 14) in terms of supporting clinical decisions based on the outcome within this domain. In addition, future studies could also address this issue by providing further refinements to the wording of the items in these domains and also by assessing the impact of raters' training in assessment accuracy. It should be noted that these are not inherent features of the Spanish version of the MAS. In fact, as indicated above, previous studies suggest coincident findings.

Another limitation of our study is that other important psychometric attributes were not assessed. Reproducibility (test-retest reliability), criterion validity with experimental functional analysis and minimum clinically significant difference, to mention a few, should be explored in the future. Additional psychometric features not explored in this study (due to the limited sample size) included topography-specific factor analysis and precision (standard error of measurement, McHorney and Tarlov 1995).

An additional flaw of our study is the gender distribution of participants (76 boys, 4 girls). However, it should be noted that both the prevalence of autism and problem behavior is higher among males and the proportion of males in our study does not depart considerably from other reported distributions (Fombonne 2005; Gimpel and Holand 2003, p. 12).

The results of the scale may be used as the basis for hypothesis-driven functional analysis, therefore improving the efficiency of the assessment process. Specifically, results suggestive of distinctive functions through indirect assessment could lead to single-function tests (e.g., Vollmer et al. 1995). This approach could not only increase efficiency but also increase the feasibility of conducting an experimental functional analysis on occasions where the behavior is construed as dangerous or occurs at relatively low rates. Under these circumstances, a full experimental functional analysis may not be practical or ethical (Iwata and Dozier 2008). However, practitioners and researchers should bear in mind that caregivers may under-report positive reinforcement contingencies in the form of attention and negative social reinforcement in the form of escape, and that this scale should not be used as the only means of functional assessment.

References

- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (4th ed. Text revised)*. Arlington VA: American Psychiatric Publishing.
- Bihm, E. M., Kienlen, T. L., Ness, M. E., & Poindexter, A. R. (1991). Factor structure of the motivation assessment scale for persons with mental retardation. *Psychological Reports, 68*, 1235–1238.
- Bullinger, M., Alonso, J., Apolone, G., Leplege, A., Sullivan, M., Wood-Dauphinee, S., et al. (1998). Translating health status questionnaires and evaluating their quality. *Journal of Clinical Epidemiology, 51*, 913–923.
- Central Intelligence Agency. (2009). World fact book. Retrieved June 30, 2009 from <http://devdata.worldbank.org/hnpstats/cd.asp>.
- Duker, P. C., & Sigafoos, J. (1998). The motivation assessment scale: Reliability and construct validity across three topographies of behavior. *Research in Developmental Disabilities, 19*, 131–141.
- Durand, M., & Crimmins, D. (1988). Identifying the variables maintaining self-injurious behavior. *Journal of Autism and Developmental Disorders, 18*, 99–117.
- Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin, 81*, 358–361.
- Eisen, M., Ware, J. E., Donald, C. A., & Brook, R. H. (1979). Measuring components of children's health status. *Medical Care, 17*, 902–921.
- EuroQoL group. (2010). EQ-5D: A standardized instrument for use as a measure of health outcome. Retrieved January 26, 2010 from www.euroqol.org.
- Fombonne, E. (2005). Epidemiology of autistic disorder and other pervasive developmental disorders. *Journal of Clinical Psychiatry, 66*(Suppl 10), 3–8.
- Freeman, K. A., Walker, M., & Kaufman, J. (2001). Psychometric properties of the questions about behavioral function scale in a child sample. *American Journal of Mental Retardation, 112*, 122–129.
- Gimpel, G. A., & Holand, M. L. (2003). *Emotional and behavioral problems of young children: Effective interventions in the preschool and kindergarten years*. New York: Guilford press.
- Greer, R. D., & Ross, D. E. (2007). *Verbal behavior analysis: Inducing and expanding new verbal capabilities in children with language delays*. Upper Saddle River, NJ: Allyn & Bacon.
- Hall, S. S. (2005). Comparing descriptive, experimental and informant-based assessments of problem behaviors. *Research in Developmental Disabilities, 26*, 514–526.
- Hanley, G. P., Iwata, B. A., & McCord, B. E. (2003). Functional analysis of behavior problem: A review. *Journal of Applied Behavior Analysis, 36*, 147–185.
- Herzinger, C. V., & Campbell, J. M. (2007). Comparing functional assessment methodologies: A quantitative synthesis. *Journal of Autism and Developmental Disorders, 37*, 1430–1445.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185.
- International Test Commission. (2000). Test adaptation guidelines. Retrieved June 20, 2009 from <http://www.intestcom.org/Guidelines/test+adaptation.php>.
- Iwata, B. A. (1996). *Functional analysis screening tool*. Gainesville, FL: Author.
- Iwata, B. A., Dorsey, M. F., Slifer, K. J., Bauman, K. E., & Richman, G. S. (1994a). Toward a functional analysis of self-injury. *Journal of Applied Behavior Analysis, 27*, 197–209. (Original document published 1982).
- Iwata, B. A., & Dozier, C. L. (2008). Clinical application of functional analysis methodology. *Behavior Analysis in Practice, 1*, 3–9.
- Iwata, B. A., Pace, G. M., Dorsey, M. F., Zarcone, J. R., Vollmer, T. R., Smith, R. G., et al. (1994b). The functions of self-injurious behavior: An experimental-epidemiological analysis. *Journal of Applied Behavior Analysis, 27*, 215–240.
- Joosten, A. V., & Bundy, A. C. (2008). The motivation of stereotypic and repetitive behavior: Examination of construct validity of the motivation assessment scale. *Journal of Autism and Developmental Disorders, 38*, 1341–1348.
- Juniper, E. F., Guyatt, G. H., & Jaeschke, R. (1996). How to develop and validate a new health-related quality of life instrument. In B. Spilker (Ed.), *Quality of life and pharmacoeconomics in clinical trials* (2nd ed., pp. 49–56). Philadelphia: Lippincott-Raven Publishers.
- Kearney, C. A. (1994). Interrater reliability of the motivation assessment scale: Another, closer look. *Journal of the Association for Persons with Severe Handicaps, 19*, 139–142.
- Kearney, C. A., Cook, C., Chapman, G., & Bensaheb, A. (2006). Exploratory and confirmatory factor analyses of the motivation assessment scale and resident choice assessment scale. *Journal of Developmental and Physical Disabilities, 18*, 2–11.
- Kennedy, C. H. (2005). *Single-case design for educational research*. Boston: Allyn & Bacon.
- Lorenzo-Seva, U., & Ferrando, P. J. (2007). FACTOR: A computer program to fit the exploratory factor analysis model. *Behavior Research Methods, Instruments, & Computers, 38*, 88–91.
- Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika, 57*, 519–530.
- Martinez-Martin, P. (in press). Composite rating scales. *Journal of Neurological Sciences*. Downloaded October 6, 2009 from <http://www.sciencedirect.com/science/article/B6T06-4X4XGRP-4/2/58a451ef09a9fe6c475aac66fbbe0f1a>.
- Maurice, C., Green, G., & Foxx, R. M. (2001). *Making a difference: Behavioral intervention for autism*. Austin, TX: Pro-Ed, Inc.

- McHorney, C. A., & Tarlov, A. R. (1995). Individual-patient monitoring in clinical practice: Are available health status surveys adequate? *Quality of Life Research*, *4*, 293–307.
- Mislevy, R. J., & Bock, R. D. (1990). *BILOG 3 Item analysis and test scoring with binary logistic models*. Mooresville, NC: Scientific Software.
- Muthen, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology*, *38*, 171–189.
- Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: Issues and applications*. Thousand Oaks, CA: Sage.
- Nicholson, J., Konstantinidi, E., & Furniss, F. (2006). On some psychometric properties of the questions about behavioral function (QABF) scale. *Research in Developmental Disabilities*, *27*, 337–352.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York: McGraw-Hill.
- Paclawskyj, T. R., Matson, J. L., Rush, K. S., Smalls, Y., & Vollmer, T. R. (2000). Questions about behavioral function (QABF): A behavioral checklist for functional assessment of aberrant behavior. *Research in Developmental Disabilities*, *21*, 223–239.
- Paclawskyj, T. R., Matson, J. L., Rush, K. S., Smalls, Y., & Vollmer, T. R. (2001). Assessment of the convergent validity of the questions about behavioral function scale with analogue functional analysis and the motivation assessment scale. *Journal of Intellectual Disability Research*, *45*, 484–494.
- Scientific Advisory Committee of the Medical Outcomes Trust. (2002). Assessing health status and quality-of-life instruments: Attributes and review criteria. *Quality of Life Research*, *11*, 193–205.
- Shogren, K. A., & Rojahn, J. (2003). Convergent reliability and validity of the questions about behavioral function and the motivation assessment scale: A replication study. *Journal of Developmental and Physical Disabilities*, *15*, 367–375.
- Sigafoos, J., Kerr, M., & Roberts, D. (1994). Interrater reliability of the motivation assessment scale: Failure to replicate with aggressive behavior. *Research in Developmental Disabilities*, *15*, 333–342.
- Singh, N. N., Donatelli, L. S., Best, A., & Williams, D. E. (1993). Factor structure of the motivation assessment scale. *Journal of Intellectual Disability Research*, *37*, 65–74.
- Spreat, S., & Connelly, L. (1996). A reliability analysis of the motivation assessment scale. *American Journal on Mental Retardation*, *100*, 528–532.
- Thompson, R. H., & Iwata, B. A. (2007). A comparison of outcomes from descriptive and functional analyses of problem behavior. *Journal of Applied Behavior Analysis*, *40*, 333–338.
- Vollmer, T. R., Marcus, B. A., Ringdahl, J. E., & Roane, H. S. (1995). Progressing from brief assessments to extended experimental analyses in the evaluation of aberrant behavior. *Journal of Applied Behavior Analysis*, *28*, 561–576.
- Ware, J. E., & Gandek, B. (1998). Methods for testing data quality, scaling assumptions and reliability. *Journal of Clinical Epidemiology*, *51*, 945–952.
- Wasano, L. C., Borrero, J. C., & Kohn, C. S. (2009). Brief report: A comparison of indirect versus experimental strategies for the assessment of pica. *Journal of Autism and Developmental Disorders*, *18*, 99–117.
- World Health Organization. (2001). World health organization disability assessment schedule (2nd ed.). Retrieved January 26, 2010 from www.who.int/icidh/whodas/index.html.
- Zarcone, J., Rodgers, T., Iwata, B., Rourke, D., & Dorsey, M. (1991). Reliability analysis of the motivation assessment scale: A failure to replicate. *Research in Developmental Disabilities*, *12*, 349–360.