Psicothema

Psicothema 2015, Vol. 27, No. 4, 416-423
doi: 10.7334/psicothema2015.11

ISSN 0214 - 9915 CODEN PSOTEG

www.psicothema.com

# Further validity evidence of the Behavioral Inhibition Observation System (BIOS)

Sergi Ballespí Sola[1], Maria Claustre Jané Ballabriga[1] and Maria Dolors Riba Lloret[2]
[1] Departament de Psicologia Clínica i de la Salut and [2] Departament de Psicobiologia i Metodologia de les Ciències de la Salut

## Abstract

**Background:** The Behavioral Inhibition Observation System (BIOS) is a brief clinician-report scale for detecting behavioral inhibition (BI) from direct observation. This study aims to compare the validity coefficients obtained in the natural context of use of the BIOS (i.e., a clinical situation) with those obtained using the BIOS after standardized observation. **Method:** The participants were 74 randomly selected preschool children who were exposed to systematic observation. **Results:** The results indicate excellent internal consistency ($\alpha$ = .91) and moderate to good inter-rater reliability for all items (ICC from .55 to .88). The correlations with observational measures of BI ranged from .40 to .70, and were mostly equivalent to those of the previous study. The correlations with parents', teachers', and clinicians' measures of BI and related constructs ranged from .30 to .60, and were also equivalent to those obtained in the natural context of use of the BIOS (i.e., clinical situation). **Conclusions:** The validity coefficients obtained with the BIOS in a non-structured natural observation are mostly equivalent to those obtained in an experimental situation, thus supporting that the BIOS is a cost-efficient instrument for measuring BI from observation in a clinical situation.

***Keywords:*** Behavioral inhibition; anxiety; assessment; direct observation; validity.

## Resumen

***Evidencias adicionales de validez del Behavioral Inhibition Observation System (BIOS).*** **Antecedentes:** El Protocolo de Observación de la Inhibición Conductual (BIOS) es una breve escala para clínicos destinada a detectar la inhibición conductual (behavioral inhibition o BI) mediante observación directa. El objetivo de este estudio es comparar los coeficientes de validez obtenidos en el contexto natural de aplicación del BIOS (i.e., en situación clínica) con los obtenidos al utilizar el BIOS tras una observación estandarizada. **Método:** Los participantes fueron 74 preescolares seleccionados aleatoriamente y expuestos a observación sistemática. **Resultados:** Los resultados indican una excelente consistencia interna ($\alpha$=.91) y fiabilidad inter-jueces de moderada a buena para todos los ítems (ICC de .55 a .88). Las correlaciones con medidas observacionales de BI oscilaron entre .40 y .70 y en su mayoría fueron equivalentes a las del estudio anterior. Las correlaciones con las medidas de BI y constructos afines de padres, maestros y clínicos oscilaron entre .30 y .60 y también fueron equivalentes a las obtenidas en el contexto natural de uso del BIOS (i.e., en situación clínica). **Conclusiones:** Los coeficientes de validez obtenidos con el BIOS en observación natural no estructurada son mayoritariamente equivalentes a los obtenidos en situación experimental, demostrando así que el BIOS es un instrumento coste-eficiente para medir BI a partir de la observación en situación clínica.

***Palabras clave:*** Inhibición conductual; ansiedad; evaluación; observación directa; validez.

Behavioral inhibition (BI) is the temperamental trait (i.e., a biologically determined tendency) that predisposes one to react with anxiety to novel, uncertain, and changing situations (Kagan, Reznick, & Snidman, 1988). In uncertain *social* situations, inhibited children show an anxiety and avoidance profile very similar to that of shyness and social phobia (SP) (Stein, Ono, Tajima, & Muller, 2004). Thus, given that they share the tendency to distress in social interactions, BI, shyness and SP are conceptually related and part of the Social Anxiety Spectrum (Schneier, Blanco, Antja, & Liebowitz, 2002), although they differ in: 1) specificity (i.e., BI not only refers to social situations, such as shyness and SP, but also includes non-social contexts); 2) complexity (i.e., shyness and SP include a cognitive component); 3) severity (i.e., ranging from a temperamental disposition to psychopathology); and 4) the observation time across development (i.e., BI is mostly detected at the preschool age, when temperament is more clearly observable; the term shyness is used to refer to social anxiety from the school stage onwards, once the cognitive component appears; and SP mostly starts in adolescence).

The chronological relationship between BI and SP explains that, in addition to being phenomenologically related, BI can be an important risk factor for SP (Chronis-Tuscano et al., 2009; Hirshfeld-Becker et al., 2007), and, given the comorbidity between social anxiety and depression, BI also increases risk for depression (Beesdo et al., 2007). Given these psychopathological implications, BI is considered an early precursor of social anxiety,

Received: January 13, 2015 • Accepted: June 4, 2015
Corresponding author: Sergi Ballespí Sola
Departament de Psicologia Clínica i de la Salut
Edificio B, Campus de Bellaterra
Cerdanyola del Vallès, Barcelona (Spain)
08193 Bellaterra
e-mail: sergi.ballespi@uab.cat

so efforts to detect this trait in early stages of development have been enhanced.

To assess BI (Kagan, Reznick, & Snidman, 1987), pioneering studies used the Behavioral Inhibition Paradigm, an artificial situation designed to permit the systematic observation of BI in the laboratory. This paradigm confronted the child with strange people, objects, and tasks that were designed to evoke the major signs of BI identified by Kagan and colleagues (i.e., cessation of play and vocalizations, avoidance behaviors, long latencies before interacting with the unfamiliar, signs of negative affect, and proximity to the mother; Kagan et al., 1987, 1988). The original paradigm was followed by several adaptations (e.g., Asendorpf, 1994; Rubin, Hastings, Steward, Henderson, & Chen, 1997) and attempts to conduct the observation in natural environments (e.g., Broberg, Lamb, & Hwang, 1990; Fordham, & Stevenson-Hinde, 1999).

However, the need for more cost-effective methods led to the emergence of specific psychometric scales for assessing BI: the Behavioral Inhibition Questionnaire (Bishop, Spence, & McDonald, 2003), the Preschool Inhibition Scale (Ballespí, Jané, Riba, & Domènech-Llaberia, 2003), the Behavioral Inhibition Scale (van Brakel, & Muris, 2006), and the Behavioral Inhibition Scale for children aged 3 to 6 (Ballespí, Jané, & Riba, 2012c). These scales were designed to obtain a BI measurement from parents and teachers or from the child him- or herself. However, there is no instrument for obtaining a psychometric measure of BI from professionals in child mental health (e.g., clinical and educational psychologists, pediatricians, psychiatrists). Such a measure would be valuable for various reasons.

First, a short BI scale would allow experts in child development to evaluate changes in temperament, the risk for impairment or psychopathology, and the efficacy of interventions in cases of high BI. Second, discrepancies among the common informants of children's behavior (i.e., parents, teachers, and the child him- or herself) are well known (De Los Reyes and Kazdin, 2005), so a systematic measure by an expert could provide a valuable additional perspective, and would allow professionals to compare their scores with those of other informants. Finally, a measure of BI from the clinician's point of view could be especially ecological considering that the first contacts of the child with a clinician tend to be strange or uncertain situations in which the child is faced with an unfamiliar person. The clinician could take advantage of this situation to easily assess the child's reactions to the unfamiliar.

Given the usefulness of having such a measure and that no currently available instrument provides a psychometric measurement of BI by experts in child development, the Behavioral Inhibition Observation System (BIOS) was developed (Ballespí, Jané, & Riba, 2013). The BIOS is an 8-item scale that can be completed in only 3 minutes. In a preliminary study (Ballespi et al., 2013), the BIOS showed good internal consistency ($\alpha$ = .88), adequate test-retest reliability in an interval of 3 weeks ($r$ = .66), and correlations ranging from .36 to .55 with other measures of BI, consistent with those reported among different sources in previous studies (Achenbach, Dumenci, & Rescorla, 2002; Ballespí et al., 2012; Bishop et al., 2003; Kagan et al., 1987; van Brakel, & Muris, 2006).

Thus, previous results support the reliability and validity of the BIOS scores. However, the BIOS has been designed to rate BI after a single and short observation of the child in a natural situation, and this is a very different observational context from those reported by previous studies based on extensive observations in the lab, in the school, or at home. So there is the remaining question of to what extent the psychometric properties of the BIOS in the natural context of use are different from those of the same instrument in a more optimal observational context.

The first aim of the current study is to analyze the validity and reliability of the BIOS used in standardized laboratory situation specifically designed to elicit the same "major signs" of BI on which the BIOS' items are based. In the preliminary study, the clinicians rated the BIOS after a clinical interview, that is, after a single instance of direct contact with the child in a social situation. In contrast, in the current study, two raters will use the instrument after repeatedly observing and micro-analyzing the child's behavior in an experimental situation. It is expected that higher validity and reliability coefficients will be obtained. Specifically, the more complete, accurate, and structured situation of observation should lead to higher internal consistency, higher correlations between the BIOS scores and the observational parameters and higher correlations to same measures of BI and to related concepts than in the preliminary study. In this sense, it is predicted that correlations with other measures of BI will be higher than those with phenomenologically related concepts, such as withdrawal and SP. In contrast, externalizing problems have shown empirical and theoretical independence from BI (e.g., see Denissen, & Asendorpf, 2008), so zero correlations are expected with measures of aggressiveness, hyperactivity, and oppositional or disruptive behavior.

The second aim of the study is to provide further evidence of the reliability of the BIOS by reporting inter-raters' degree of agreement.

## Method

### *Participants*

The participants were recruited from 3- to 6-year-old children in the school system in the province of Barcelona, their parents (i.e., the parent more closely involved in the child's care), and their teachers. Of the 365 children whose parents agreed to participate, the consent forms to be included in the standardized observation were obtained in 329 cases (91%). The only exclusion criterion was the presence of intellectual disability or autism spectrum disorder. Thirty-two families declined to participate because they did not want their children to be filmed. Among the 32 children whose parents declined to participate in the observation, there was one child with high level of BI from the parents' (but not the teachers') ratings and one child with a high level of BI from the teachers' (but not the parents') ratings.

Of those 329 children whose parents signed the informed consent to be observed, 74 preschoolers (37 boys and 37 girls) were randomly selected (equalizing for sex) to participate in the behavioral inhibition observational paradigm. All of the children were Caucasian and were between 40 and 76 months old (mean = 57, $SD$ = 10.6). They all spoke Catalan, Spanish or both languages at home. Nine percent of the participants came from families with low or medium-low socioeconomic levels, 23% came from medium levels, 45% came from medium-high levels, and 23% came from high socioeconomic levels, based on the Hollingshead index.

*Instruments*

The Behavioral Inhibition Observation System (BIOS) is a brief clinician-report scale created to measure the degree of BI through direct observation. The BIOS was created for professionals with a good knowledge of children's normative behavior (e.g., child psychologists and psychiatrists, educational psychologists, pediatricians), who would be able to rate the BIOS after an individual meeting with the child (e.g., the first visit with a psychologist, an interview or assessment situation). The BIOS consists of 7 items scored on a 5-point Likert scale and based on the major signs of BI defined by Kagan et al. (1987), and the B8 indicator. The B8 indicator is a final item in the form of a 7-point scale to offer a global impression of the child's degree of BI. Previous findings support that the BIOS is structured in one dimension that explains 60% of the total variance and has good internal consistency (Cronbach's $\alpha$ = .88; Ballespí et al., 2013). The internal consistency of the BIOS for the current sample is provided in Table 1.

The *Behavioral Inhibition Scale for Children Aged 3 to 6* (BIS 3-6) is a dimensional instrument used to measure BI in children aged 3 to 6 with good psychometric properties (Ballespí et al., 2012c). It consists of 35 items, to which parents and teachers respond using five response categories (*almost never, rarely, sometimes, often, almost always*). The items' content is based on the major signs described by Kagan and colleagues (1987), but items related to social BI predominate. The scale is structured on a single dimension of BI, which explains 43% of the total variability and shows good internal consistency (Cronbach's $\alpha$ = .95). The internal consistency in the current sample was excellent for both parents' and teachers' scores ($\alpha$ = .96 in both cases).

The *Child Behavior Checklist 1½-5* (CBCL/1½-5) for parents and the *Caregiver-Teacher Report Form for Ages 1½-5* (C-TRF) are standardized instruments included in the Achenbach System of Empirically Based Assessment (ASEBA, Achenbach, & Rescorla, 2000) with available Spanish versions. Both instruments have 7 clinical dimensions, 3 second-order factors and 5 Diagnostic and Statistical Manual (DSM) scales. In the current study, we used the scores of the withdrawal scale to check the convergent validity of the BIOS scores. According to the manual, the Withdrawal scale shows Cronbach's alpha values of .75 and .83 for the CBCL and the C-TRF, respectively. We used the scores of the Aggression Problems scale (with Cronbach's alpha values of .92 and .98 for the CBCL and the C-TRF, respectively), and the scores of the DSM-based ADHD scale (with Cronbach's alpha values of .78 and .98) to analyze the discriminant validity of the BIOS scores.

The *Early Child Inventory 4* (ECI-4) (Sprafkin, & Gadow, 1996) is an inventory of 108 DSM-IV-based symptoms (DSM-IV) (APA, 1994) for preschool-aged children. Previous studies using the Spanish version showed appropriate psychometric properties, with Cronbach's alpha values ranging from .62 to .94 (e.g., Viñas et al., 2008). In the present study, we used the dimensional scores of the Social Phobia, Oppositional Defiant Disorder, and Conduct Disorder scales to analyze the validity of the BIOS scores. The Cronbach's alpha coefficients for the current sample were, for parents' and teachers' data respectively, .77 and .83 for the Oppositional Defiant Disorder scale, .70 and .92 for the Conduct Disorder scale, and .68 and .57 for the Social Phobia scale.

*Procedure*

In order to analyze the psychometric properties of the BIOS in a context of systematic observation, the current study compares the BI ratings provided using the BIOS by the raters who analyze the behavioral inhibition paradigm, with four sources of data: 1) parents, 2) teachers, 3) clinicians who individually interview children, 4) the observational parameters derived from the behavioral inhibition paradigm.

*The Behavioral Inhibition observational paradigm*

The observational paradigm had to be adapted to the school environment for the parents to give their consent because the labs are twenty miles from the city and most parents were not prepared to make the trip. This solution made it possible to obtain data from a standardized situation without having to observe the participants in the laboratory. The current version of the observational paradigm was based on previous adaptations (see Ballespí et al., 2013 for a more extended review), and it required two researchers, lasted approximately 20 minutes and consisted of five phases (i.e., Play Rug, New Toy, Interview, Box, and Cognitive Stress).

*Data collection from parents, teachers, and clinicians*

The task of completing the questionnaires was assigned to the parent who spent more time with the child, so mothers elected to answer in more than 90% of the cases. To gather data for further research, every child also had an individual meeting with a clinician at the school. The meeting lasted approximately 30 minutes and consisted of an initial interview to help the child gain confidence and subsequent administration of language, psychopathology, and cognitive development tests. Immediately after every child's assessment, the clinician completed the BIOS to rate the child's BI. This part of data collection is further described elsewhere (Ballespí et al., 2013).

*The exposure to the observational paradigm*

The schools provided the best available space that met the requirements of the observation paradigm (i.e., an empty classroom, the gym or the library). Although the spaces assigned were different in each case, they were all sufficient to create the same standardized situation. The observational paradigm was carried out approximately 10 days after the meeting with the clinician. Thus, the mean of the time interval between the interview situation from which the clinicians scored the BIOS and the standardized situation from which the researchers did so was $M$ = 9.28 days ($SD$ = 6.05, $Range$ = 25). The holiday period in Spain, the eventual illness of some children and attendance at summer camp resulted in 6 children being observed more than 20 days after the clinicians' meeting. However, because BI is considered a constitutional trait, it was not expected to vary in such a short time.

*Analysis of the observational data and the use of the BIOS after analyzing the observational paradigm*

Each of the 74 selected children was exposed to the adaptation of the behavioral inhibition paradigm described above. The observation sessions were filmed and subsequently analyzed by

two independent raters. Neither of these raters made observations of the children in the previous phase of the study when the initial ratings were gathered. Raters repeatedly processed the recordings of the behavioral inhibition paradigm to obtain four types of observational indicators: latencies (e.g., *time it takes the child to start to play*), time (e.g., *total time of positive affect*), percentages (e.g., *fraction of time speaking or complaining*), and other (rates and frequencies, such as the *amount of spontaneous comments*). These indicators are based on the major signs of BI defined by Kagan et al. (1987) and are those used in previous adaptations of the paradigm (see Ballespí et al., 2012b, 2013, for a more extensive review). Immediately after the child's responses in the observational paradigm were completely analyzed, each rater filled out the BIOS. The training for rating the BIOS consisted of complete theoretical information about the BI construct, and double-blind administration of the scale in clinical situation with children different from those of this study.

*Data analysis*

The data were refined and processed with the statistics package SPSS 17.0 (SPSS, 2007). Based on previous studies (see Ballespí et al., 2013), some specific observational measurements derived from the standardized paradigm were summarized in three general indicators: *Latency of contact with the four objects*, *Latency of response to the interview questions, and Percentage of total speech time* (see Table 3). To homogenize the direction of responses to the BIOS, the inverse items were recoded in a direct format. Cronbach's alpha index was calculated. To evaluate validity in relation to other measures, Pearson's linear correlation coefficient and Spearman's rank order correlation index (when assumptions for Pearson's coefficient were not met) were calculated. Inter-rater reliability was evaluated using the intra-class correlation coefficient (ICC). Differences between pairs of correlation coefficients were tested using Fisher's Z transformation (Snedecor & Cochran 1989).

## Results

*Internal consistency*

As Table 1 shows, all of the items had correlations with both the scale and the B8 index, ranging from .50 to .90. The Cronbach's Alpha coefficient indicated excellent internal consistency in both applications ($\alpha$ = .91). The third column indicates that removing any item did not affect the alpha index. This was true even for Items 1 and 5, which seemed to be the least consistent. Therefore, all of the items contributed to the internal consistency of the BIOS.

*Inter-rater reliability*

The results shown in Table 2 indicate good inter-rater reliability for the BIOS scores. The ICC showed values of .80 and .94 for the B8 indicator and the BIOS total score, respectively. The inter-rater correlations ranged from .80 to .90 for all of the items, except Items 1 (*Latency of response*) and 7 (*Speech tone and volume*), which showed values of .54 and .55, respectively. Table 2 also presents the item-to-item correlations among the raters' applications of the BIOS (after viewing the recordings of the observation paradigm) and the clinicians' applications (after individually interviewing each child in the school). The results showed moderate correlations ranging from .30 to .60 for all items, except Item 5 (*Negative affect*), which does not show significant correlations. For the global indexes (the B8 and the BIOS total score), the correlations indicated moderate convergence between the clinical and the research contexts of BIOS application.

*Relationship with conceptually related constructs*

Table 3 shows moderately-highly significant correlations with all of the observational BI indicators. Specifically, the correlations ranged from .40 to .80 and were always in the expected direction. Table 4 shows that the BIOS scores for both applications had correlations with other measures of the same construct (i.e.,

*Table 1*
Internal consistency of the Behavioral Inhibition Observation System (BIOS)

| | Rater 1 | | | Rater 2 | | |
|---|---|---|---|---|---|---|
| | Corrected Item-Total correlation | Item-B8 correlation | Alpha if item deleted | Corrected Item-Total correlation | Item-B8 correlation | Alpha if item deleted |
| Item 1 - Latency of response - (*How long did it take the child to answer the interviewer's questions?*) | .57 | .60 | .91 | .67 | .67 | .90 |
| Item 2 - Amount of speech - (*How much did the child talk during the interview?)* | .73 | .77 | .89 | .72 | .71 | .89 |
| Item 3 - Tension/Discomfort - (*To what extent did he/she look tense or uncomfortable?*) | .89 | .92 | .87 | .88 | .85 | .87 |
| Item 4 - Positive Affect - (*To what extent did the child show signs of positive affect?*) | .81 | .82 | .88 | .82 | .76 | .88 |
| Item 5 - Negative Affect - (*To what extent did the child show signs of negative affect?*) | .53 | .67 | .91 | .57 | .67 | .91 |
| Item 6 - Latency to gain confidence - (*How long did it take for the child to trust the interviewer?*) | .90 | .92 | .87 | .85 | .82 | .88 |
| Item 7 - Speech tone and volume - (*How were the tone and volume of the child's voice?*) | .67 | .69 | .90 | .69 | .64 | .90 |
| | | Cronbach's Alpha: .91 | | | Cronbach's Alpha: .91 | |

*Table 2*
Inter-rater reliability of the Behavioral Inhibition Observation System (BIOS) and item-to-item convergence with the clinicians' ratings

| Indicator | *ICC*[1] | *95% CI* | *r* Rater1-Clin[2] | *95% CI* | *r* Rater2-Clin[2] | 95% CI |
|---|---|---|---|---|---|---|
| Item 1 - Latency of response | .54* | .36 to .69 | .42* | .21 to .59 | .33* | .11 to .52 |
| Item 2 - Amount of speech | .80* | .70 to .87 | .59* | .42 to .72 | .39* | .18 to .57 |
| Item 3 - Tension/discomfort | .87* | .80 to .91 | .37* | .16 to .55 | .35* | .13 to .54 |
| Item 4 - Positive Affect | .83* | .74 to .89 | .55* | .37 to .69 | .46* | .26 to .62 |
| Item 5 - Negative Affect | .84* | .91 to .96 | .08 | | .11 | |
| Item 6 - Latency to gain confidence | .88* | .81 to .92 | .58* | .41 to .71 | .56* | .38 to .70 |
| Item 7 - Speech tone and volume | .55* | .37 to .69 | .44* | .24 to .61 | .37* | .16 to .55 |
| B8 Indicator | .80* | .69 to .87 | .46* | .26 to .62 | .37* | .16 to .55 |
| BIOS Total Score | .94* | .90 to .96 | .55* | .37 to .69 | .51* | .32 to .66 |

[1]Item-to-item intra-class correlation coefficients between both current applications of the BIOS in the research context (i.e., between Rater 1's and Rater 2's applications). [2]Pearson's or Spearman's correlation coefficients among the raters' and the clinicians' applications of the BIOS. The raters completed the BIOS after viewing recordings of the standardized paradigm. The clinicians rated the BIOS after interacting with the child in an interview situation. In all cases, n = 74. *p<.01

from parents, teachers or clinicians) ranging from .30 to .60. The correlations with the other rater's scores or with the same rater's B8 indicator were approximately .90. The correlations with parents' and teachers' measures of related constructs fell primarily between .20 and .50, and most reached significance. This result contrasts with the correlations obtained with nonrelated constructs, which were all non-significant. Overall, Tables 3 and 4 show that the correlations progressively decreased as the conceptual relation between BI and the construct assessed by the external measure also progressively decreased. Most correlations of Tables 3 and 4 do not differ from those obtained when the BIOS was used in its natural context of use (i.e., after child-clinician interaction), with the exception of those regarding interaction with objects and affect expression at the end of Table 4.

*Table 3*
Correlations among the BIOS scores and observational measurements of the same construct

| | BIOS' RATER 1 | | | | | | BIOS' RATER 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | TOTAL SCORE | | | B8 | | | TOTAL SCORE | | | B8 | | |
| | *r* | *n* | *95% CI* | *r* | *n* | *95% CI* | *r* | *n* | *95% CI* | *r* | *n* | *95% CI* |
| LATENCIES[1] | | | | | | | | | | | | |
| Latency of 1st spontaneous comment | .70*• | 74 | .56 to .80 | .65*· | 74 | .50 to .77 | .70*· | 74 | .56 to .80 | .58* | 74 | .41 to .71 |
| Latency of 8th spontaneous comment | .78* | 74 | .67 to .86 | .78* | 74 | .67 to .86 | .77* | 74 | .66 to .85 | .65* | 74 | .50 to .77 |
| Latency of R to interview questions[2] | .65* | 73 | .49 to .77 | .63* | 73 | .47 to .75 | .67* | 74 | .52 to .78 | .57* | 73 | .39 to .71 |
| Latency of contact with the four objects[3] | .55*•• | 71 | .36 to .69 | .55*• | 71 | .36 to .69 | .69*••• | 71 | .54 to .80 | .64*•• | 71 | .48 to .76 |
| SPEECH | | | | | | | | | | | | |
| Percentage of time of total speech[4] | -.73* | 74 | -.82 to -.60 | -.73*• | 74 | -.60 to -.82 | -.68* | 74 | -.54 to -.79 | -.63* | 74 | -.47 to -.75 |
| Frequency of spontaneous comments | -.63* | 74 | -.47 to -.75 | -.67* | 74 | -.52 to -.78 | -.57* | 74 | -.39 to -.71 | -.54* | 74 | -.36 to -.68 |
| PLAY | | | | | | | | | | | | |
| Total time playing with toys | -.49*· | 74 | -.30 to -.65 | -.57*• | 74 | -.39 to -.71 | -.58*• | 74 | -.41 to -.71 | -.62*•• | 74 | -.46 to -.74 |
| Number of toys explored | -.51*• | 74 | -.32 to -.66 | -.55*•• | 74 | -.37 to -.69 | -.57*•• | 74 | -.39 to -.71 | -.60*•• | 74 | -.43 to -.73 |
| AFFECT | | | | | | | | | | | | |
| Percentage of time of positive affect | -.73*• | 74 | -.60 to -.82 | -.80*••• | 74 | -.70 to -.87 | -.71*• | 74 | -.57 to -.81 | -.72*•• | 74 | -.59 to -.82 |
| Percentage of time of negative affect | .41*· | 74 | .20 to .58 | .50*• | 74 | .31 to .65 | .43*· | 74 | .22 to .60 | .52*• | 74 | .33 to .67 |

. [1]Unit of measurement for latencies and time: seconds. [2]The average of the latencies of response to the 7 interview questions. [3]The average of the latencies of contact with the rug, the first toy, the bird, and the box. [4]The percentage of the total time the child spent talking, both spontaneously and in response to the researcher. *The correlation is statistically significant ($p$<.01). The correlation is significantly higher than that obtained when the BIOS was used in the natural context (i.e., after the child-clinician interaction; ·$p$<.05, •$p$<.01, ••$p$<.001, •••$p$<.0001).

*Table 4*
Correlations among the BIOS scores and psychometric measurements of the same and other constructs

| | BIOS' RATER 1 | | | | | | BIOS' RATER 2 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | TOTAL SCORE | | | B8 INDICATOR | | | TOTAL SCORE | | | B8 INDICATOR | | |
| | *r* | *n* | *95% CI* | *r* | *n* | *95% CI* | *r* | *n* | *95% CI* | *r* | *n* | *95% CI* |
| SAME CONSTRUCT | | | | | | | | | | | | |
| Clinicians' BIOS PT | .55* | 74 | .37 to .69 | .52* | 74 | .33 to .67 | .51* | 74 | .32 to .66 | .44* | 74 | .24 to .61 |
| Clinicians' BIOS B8 | .45* | 74 | .25 to .62 | .45* | 74 | .25 to .62 | .41* | 74 | .20 to .58 | .36* | 74 | .14 to .55 |
| Parents' BIS 3-6[a] | .38* | 73 | .16 to .56 | .33 | 73 | .11 to .52 | .42* | 73 | .21 to .59 | .41* | 73 | .20 to .59 |
| Teacher's BIS 3-6 | .63*• | 73 | .47 to .75 | .54* | 73 | .35 to .69 | .55*· | 73 | .37 to .69 | .54* | 73 | .35 to .69 |
| Other rater's B8 Indicator | .86* | 74 | .79 to .91 | .87* | 74 | .80 to .92 | .92* | 74 | .88 to .95 | .87* | | .80 to .92 |
| Same rater's B8 Indicator | .96* | 74 | .94 to .98 | | | | .89* | 74 | .83 to .93 | | | |
| RELATED CONSTRUCTS | | | | | | | | | | | | |
| Parents' CBCL Withdrawal scale[b] | .28 | 37 | | .07 | 37 | | .34 | 37 | .02 to .60 | .17 | 37 | |
| Teacher's C-TRF Withdrawal scale | .34 | 37 | .02 to .60 | .19 | 37 | | .31 | 37 | | .37 | 37 | .05 to .62 |
| Parents' ECI-4 Social Phobia scale | .37* | 68 | .14 to .56 | .27 | 68 | .03 to .48 | .34* | 68 | .11 to .54 | .32 | 68 | .09 to .52 |
| Teacher's ECI-4 Social Phobia scale | .51* | 73 | .32 to .66 | .43* | 73 | .22 to .60 | .45* | 73 | .25 to .62 | .43* | 73 | .22 to .60 |
| NON-RELATED CONSTRUCTS | | | | | | | | | | | | |
| Parents' CBCL Aggressiveness | .07 | 37 | | .08 | 37 | | .14 | 37 | | .16 | 37 | |
| DSM-ADHD[1] | .08 | 37 | | .10 | 37 | | .14 | 37 | | .27 | 37 | |
| Teacher's TRF Aggressiveness | .15 | 37 | | .05 | 37 | | .07 | 37 | | .15 | 37 | |
| DSM-ADHD | -.03 | 37 | | .03 | 37 | | -.07 | 37 | | .08 | 37 | |
| Parents' ECI-4 ODD[2] | .11 | 68 | | .08 | 68 | | .13 | 68 | | .12 | 68 | |
| CD[3] | .12 | 68 | | .10 | 68 | | .09 | 68 | | .19 | 68 | |
| Teacher's ECI-4 ODD | .12 | 73 | | -.12 | 73 | | -.12 | 73 | | -.05 | 73 | |
| CD | .06 | 73 | | .06 | 73 | | .02 | 73 | | .09 | 73 | |

[1]Scale of attention deficit/hyperactivity disorder, based on DSM criteria. [2]Oppositional defiant disorder scale. [3]Conduct disorder scale. *The correlation is statistically significant ($p<.01$). The correlation is significantly higher than that obtained when the BIOS was used in the natural context (i.e., after the child-clinician interaction; ·$p<.05$, •$p<.01$). [a] Sample size decreases from 74 due to missing values from parents and teachers in some instruments. [b] In the case of CBCL and C-TRF, n = 37 is not due to attrition, but it is due to the fact that only participants of a half of the schools were asked to rate this instruments in the context of another study.

## Discussion

The BIOS was designed to measure BI from the child's reactions to the first contact with a professional (e.g., a clinician). Thus, the BIOS aims to take advantage of an unfamiliar situation with a stranger and uncertain tasks (e.g., an interview, assessments) to ensure an easy and brief application (i.e., 3 minutes) because it is expected to be used at the end of a clinical session with the child.

The conditions required by the BIOS to be applied make this instrument very cost-efficient, but raise the question of whether this minimalist context of observation (i.e., a single natural and non-structured observation) is enough to gather a valid and reliable measurement of BI. To answer this question, the principal aim of this study was to further explore the psychometric properties of the BIOS, this time, in a more optimal observational context, in order to verify the psychometric properties when a more accurate observation and a deeper analysis of the "major signs" of BI are possible. In this case, the BIOS is used with an adaptation of the Behavioral Inhibition Paradigm, that is, a standardized situation specifically designed to elicit the same "major signs" of BI on which the BIOS items are based. The results extend the previous findings and add evidence for the validity and reliability of the BIOS scores.

Most general observational indexes showed moderate-to-high correlations (from .55 to .73) with the BIOS scores. These results reinforce those of previous research that refer to correlations ranging between .2 and .5 across different situations and informants (Achenbach et al., 2002; Ballespí et al., 2003; Bishop et al., 2003; van Brakel et al., 2004; De Los Reyes & Kazdin, 2005). In addition, all of the correlations obtained with the observational indicators were in the expected direction. This result supports that the BIOS' scores are consistent with the characteristic profile of BI (Kagan et al., 1987; 1988).

The convergence with external measures of same and related constructs supports the idea that the BIOS scores

do indeed measure BI. Further support of this assumption is provided by the non-significant correlations with nonrelated constructs. These near-zero correlations show that the BIOS scores distinguish the construct that they intend to measure from phenomena that are theoretically unrelated to BI, such as externalizing disorders.

All of these results satisfy the predictions and reinforce the results of previous research (Ballespí et al., 2013). Interestingly, the validity coefficients obtained with the standardized observation paradigm used here mostly did not differ from those obtained in the natural context of use. Almost all correlations between the BIOS and measures of the same and related constructs were equivalent when the BIOS was used in a clinical situation and when it was used in the systematic observation. In addition, the correlations between the BIOS and almost a half of the observational indexes were equivalent in both contexts of use. This is interesting because the current situation of application of the BIOS differs from the clinical context in three points. While the clinicians observed the child *once* in a *natural* situation, here the researchers (a) rated the BIOS after micro-analyzing a *standardized* paradigm specially designed to elicit the same signs of BI in which the BIOS' items are based, (b) they *repeatedly* reviewed the recordings of the paradigm, and (c) the researchers who rated the BIOS were the same who *measured the observational indicators* derived from the paradigm. So it was expectable to obtain higher correlations than in the preliminary study.

However, only the indexes referred to non-social BI (i.e., *Latency of contact with the four objects, Total time playing with toys, Number of toys explored*), and those regarding affect expression showed higher correlations when the BIOS was used in the lab situation. Regarding interaction with objects, it should be considered that the standardized paradigm was designed to measure reactions to both social and non-social novel stimuli, while the natural situation was mostly a social situation (i.e., an interaction with a clinician). This would explain that the correlations with the speech indicators (i.e., latency of spontaneous comments, amount of speech) but not those with objects' interaction were equivalent in both contexts of use.

In addition, the items' content of the standardized situation is more similar to that of the BIOS and it was deliberately designed to be strange, whereas the meeting with the clinician is aimed to be friendly and it is usually more familiar for children, because children have already been assessed before. This would explain why correlations between the BIOS scores and the affect indicators were moderate-to-high in the standardized context and zero in the natural context of use.

The second aim of the present study was to analyze the inter-rater reliability of the BIOS scores. The correlation of .94 for the total score and item-to-item correlations ranging from .80 to .90 for most items indicate good inter-rater reliability. It should be noted that both researchers rated the BIOS after viewing the recordings of the paradigm several times and paying attention to the major signs of BI, so this may explain their precision in rating the BIOS and the high convergence of both applications. Only two values were lower than .80 and were for Items 1 (*Latency of response*) and 7 (*Speech tone and volume*). It is possible that more accurate training may be required to produce better agreement about what constitutes a long latency of response and a low tone of voice.

Overall, the present findings are important because they provide additional evidence for the validity and reliability of the BIOS scores. Moreover, the current results show that the validity coefficients of the BIOS in the natural context of use (i.e., after a single non-structured observation of child-clinician interaction) are mostly not different from those obtained in an experimental context (i.e., standardized observation). This is especially interesting considering that the coefficients of validity here obtained would have probably been even more similar to those from the clinical situation if the BIOS had not been rated after obtaining the observational indicators of the paradigm. Therefore, the present study contributes to show that, although the BIOS is aimed to be used after minimal observation, the BIOS scores are sufficiently valid to measure BI.

However, this study has some limitations. The first is the relatively small sample size. Although the current sample is similar to those used in previous studies (e.g., $n = 100$ in Bishop et al., 2003; $n = 59$ in van Brakel et al., 2004), this size does limit the power with regard to significant correlation coefficients. Moreover, the sample was drawn from a general population. Future research with clinical samples should complement the current results for the BIOS.

Although the current adaptation was consistent with previous ones (e.g., Asendorpf, 1994; Rubin et al., 1997), a third limitation is the artificiality of a standardized paradigm, such as the one used in the present study. The fourth limitation refers to the limited social content of the current paradigm, in which the child is confronted predominantly with strange objects and tasks and disguised adults. Fifth, although the paradigm had to be adapted to the school context to encourage parents to give their consent, this prevented the same degree of standardization that is possible in the laboratory. Sixth, the observational indicators used in the behavioral inhibition paradigm provide measurements of transient states of BI. Given that BI is a trait, stable measures should also be used to further explore the BIOS' validity.

Accordingly, future research should address these limitations by comparing the BIOS scores with more measures of trait BI. Structured interviews with parents and teachers may help capture this stability. Aggregate measures that combine the information from different sources (especially from parents, teachers and professionals, at these ages) or integrate measures from different moments in time should also determine the stable tendency toward anxious inhibition. Further research with more ecological observational data (i.e., more representative of the natural environment and the real situations in which the child shows BI) should also be conducted. Observations in natural contexts or the use of more social content in the standardized paradigm would be useful complements to the analysis of the BIOS' validity. Future research would also benefit from more extensive samples, especially for the analysis of cut-off scores. Finally, studies using clinical samples and longitudinal designs would allow the analysis of the predictive validity of the BIOS scores.

## References


Achenbach, T.M., & Rescorla, L.A. (2000). *Manual for the ASEBA Preschool Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth & Families.

Achenbach, T.M., Dumenci, L., & Rescorla, L.A. (2002). Ten-year comparisons of problems and competencies for national samples of youth: Self, parent and teacher reports. *Journal of Emotional and Behavioral Disorders, 10*, 194-203.

American Psychiatric Association (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.

Asendorpf, J.B. (1994). The malleability of behavioral inhibition: A study of individual developmental functions. *Developmental Psychology, 30,* 912-919.

Ballespí, S., Jané, M.C., & Riba, M.D. (2012a). Who should report abnormal behavior at preschool age? The case of behavioral inhibition. *Child Psychiatry and Human Development, 43*, 48-69.

Ballespí, S., Jané, M.C., & Riba, M.D. (2012b). The Behavioural Inhibition Scale for children aged 3 to 6 (BIS 3-6): Validity based on its relation with observational measures. *Journal of Psychopathology and Behavioral Assessment, 34*, 487-496.

Ballespí, S., Jané, M.C., & Riba, M.D. (2012c). Parent and teacher ratings of temperamental disposition to social anxiety: the BIS 3-6. *Journal of Personality Assessment, 94,* 164-174.

Ballespí, S., Jané, M.C., & Riba, M.D. (2013). Reliability and validity of a brief clinician-report scale for screening behavioral inhibition. *Journal of Psychopathology and Behavioral Assessment, 35*, 321-334.

Ballespí, S., Jané, M.C., Riba, M.D., & Domènech-Llaberia, E. (2003) Preschool Behavioural Inhibition Scale – Teacher's form (PBIS-T): Psychometric properties. *Psicothema, 15*, 205-210.

Beesdo, K., Bittner, A., Pine, D. S., Stein, M. B., Hofler, M., Lieb, R., & Wittchen, H. U. (2007). Incidence of social anxiety disorder and the consistent risk for secondary depression in the first three decades of life. *Archives of General Psychiatry, 64*, 903-912.

Bishop, G., Spence, S.H., & McDonald, C. (2003). Can parents and teachers provide a reliable and valid report of behavioural inhibition? *Child Development, 6*, 1899-1917.

Broberg, A., Lamb, M.E., & Hwang, P. (1990). Inhibition: Its stability and correlates in sixteen- to forty-month-old children. *Child Development, 61,* 1153-1163.

Chronis-Tuscano, A., Degnan, K.A., Pine, D.S., Pérez-Edgar, K., Henderson, H.A., Díaz, Y., Raggi, V.L., & Fox, N.A. (2009). Stable early maternal report of behavioral inhibition predicts lifetime social anxiety disorder in adolescence. *Journal of the American Academy of Child and Adolescent Psyquiatry, 48*, 928-935.

De Los Reyes, A., & Kazdin, A.E. (2005). Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendation for further study. *Psychological Bulletin, 131*, 483-509.

Denissen, J. A. A., & Asendorpf, J. B. (2008). Childhood personality predicts long-term trajectories of shyness and aggressiveness in the context of demo- graphic transitions in emerging adulthood. *Journal of Personality, 76*, 67-99.

Fordham, K., & Stevenson-Hinde, J. (1999). Shyness, friendship quality, and adjustment during middle childhood. *Journal of Child Psychology and Psychiatry, 40*, 757-768.

Hirshfeld-Becker, D.R., Biederman, J., Henin, A., Faraone, S.V., Davis, S., Harrington, K., & Rosenbaum, J.F. (2007). Behavioral inhibition in preschool children at risk is a specific predictor of middle childhood social anxiety: A five-year follow-up. *Journal of Developmental & Behavioral Pediatrics, 28*, 225-233.

Kagan, J., Reznick, S.J, & Snidman, N. (1988). Biological bases of childhood shyness. *Science, 240,* 167-171.

Kagan, J., Reznick, S.J., & Snidman, N. (1987). The physiology and psychology of behavioral inhibition in children. *Child Development, 58,* 1459-1473.

Rubin, K.H., Hastings, P.D., Steward, S., Henderson, H.A., & Chen, X. (1997). The consistency and concominants of inhibition: Some of the children, all of the time. *Child Development, 68,* 467-483.

Schneier, F.R., Blanco, C., Antja, S.X., & Liebowitz, M.R. (2002). The social anxiety spectrum. *Psychiatric Clinics of North America, 25*, 757-774.

Snedecor, G. W., & Cochran, W. G. (1989). Statistical Methods. Ames: University Press.

Sprafkin, J., & Gadow, K.D. (1996). *Early Childhood Inventories Manual*. Stony Brook, NY: Checkmate Plus.

SPSS, Inc. [computer software]. (2007). SPSS versió 17.0. Chicago: SPSS, Inc.

Stein, D.J., Ono, Y., Tajima, O., & Muller, J.E. (2004). The social anxiety disorder spectrum. *Journal of Clinical Psychiatry, 65(supp 14)*, 27-33.

van Brakel, A.M.L., & Muris, P. (2006). A brief scale for measuring "behavioral inhibition to the unfamiliar" in children. *Journal of Psychopathology and Behavioral Assessment, 28*, 79-84.

Viñas, F., Jané, M.C., Canals, J., Ballespí, S., Domènech-Llaberia, E., & Esparó, G. (2008). Assessment of psychopathology in preschool age children through the Early Childhood Inventory-4 (ECI-4): Agreement among parents and teachers. *Psicothema, 20*, 481-486.