

Title:

Scenes for Social Information Processing in Adolescence: Item and factor analytic procedures for psychometric appraisal

Authors

Paula Vagos, Faculty of Psychology and Educational Sciences, University of Coimbra (Portugal), paulavagos@fpce.uc.pt,

Daniel Rijo, Faculty of Psychology and Educational Sciences, University of Coimbra (Portugal), drijo@fpce.uc.pt

Isabel M. Santos, Department of Education, University of Aveiro (Portugal), isabel.santos@ua.pt

Funding source

This work was funded by a research scholarship by Fundação para a Ciência e Tecnologia, Portugal and the European Social Fund (SFRH / BPD / 72299 / 2010).

Contact information for the corresponding author:

Paula Vagos

Institucional address: Faculdade de Psicologia e Ciências da Educação, Rua do Colégio Velho, 3001-802 Coimbra

Email contact: paulavagos@fpce.uc.pt

Voice contact: 00361 965114841

Abstract

Relatively little is known about measures used to investigate the validity and applications of social information processing (SIP) theory. The *Scenes for Social Information Processing in Adolescence* (SSIPA) include items built using a participatory approach to evaluate the attribution of intent, emotion intensity, response evaluation and response decision steps of SIP. A sample of 802 Portuguese adolescents were evaluated (61.5% female; mean of 16.44 years old) using this instrument. Item analysis and exploratory and confirmatory factor analytic procedures were used for psychometric examination. Two measures for attribution of intent were produced, including hostile and neutral; along with three emotion measures, focused on negative emotional states; eight response evaluation measures, and four response decision measures, including prosocial and impaired social behavior. All of these measures achieved good internal consistency values and fit indicators. Boys seemed to favor and choose overt and relational aggression behaviors more often; girls conveyed higher levels of neutral attribution, sadness, and of assertiveness and passiveness favoring and choosing. The SSIPA achieved adequate psychometric results and seems a valuable alternative to evaluating SIP, even if it is essential to continue investigation into its internal and external validity.

Keywords

Social information processing, Attribution of intent, Emotion, Response evaluation and decision, Psychometrics

Social information processing (SIP; Crick & Dodge, 1994) theories intend to provide a comprehensive framework on which to support the study, understanding and promotion of adjusted social behaviors. The pertinence and applicability of this model and its developments have been established by robust findings associating the social information processing steps to several behavioral patterns (Crick & Dodge, 1994, 1996), including prosocial behavior (Nelson & Crick, 1999), but particularly aggressive behavior (Calvete & Orue, 2010, 2012; de Castro, Veerman, Koops, Bosch, & Monshouwer, 2002; Harper, Lemerise, & Caverly, 2010). Commonly, these findings have been found using hypothetical self-referent situation interviews or questionnaires, most of which, nevertheless, lack a sound psychometric evaluation, particularly construct validity based on internal factor structure, which leads to the exercise of caution in the interpretation of their results. Also, different steps of SIP are evaluated using separate instruments, making assessment protocols longer and more difficult to answer to (Erdley, Rivera, Shepherd, & Holleb, 2010), particularly with younger participants.

Therefore, there is still an evident need to improve on the quality of how we evaluate and come to conclusions based on the SIP underlying different types of social behaviors. The current work intends to address this need, by psychometrically evaluating the results obtained from a community sample of adolescents, using a newly developed instrument, the *Scenes for Social Information Processing in Adolescence* (SSIPA). Its items were developed on the basis of a participatory approach using focus groups, as a way to access and specifically evaluate adolescents' social experiences and their SIP (Vagos, Rijo, & Santos, 2013), and address the attribution of intent, emotion intensity, and response evaluation and decision on four social behavior options. Unlike previous instruments, the SSIPA sees SIP as inherent and probably differentiated across different types of social behaviors (and not only aggression), and includes items that may further ascertain this assumption.

The SSIPA was developed to consider the principal cognitive steps proposed by the SIP model (Crick & Dodge, 1994), which are said to influence each other and mediate between a social event and the behavioral response that is given to it. They are: 1) encoding of internal and external clues taken as consequences or characteristic of the social event; 2) interpretation or assigning meaning to the stimuli received, according to previous social experiences, schemata and scripts; 3) clarification of the goals desired from that social event, which can be personal or social gains; 4) response search, where several responses are retrieved from the personal knowledge database if they are considered as potentially useful in the current social event, 5) response evaluation, by evaluating the expected outcomes and consequences and the appropriateness of any given response option, and the personal ability to carry out that response option, and finally 6) behavioral decision and enactment. Such implicit cognitive processes thus possibly determine social behavior patterns that may result in maladaptive or adaptive interpersonal cycles (Horowitz, 1991). More recently, two assumptions have been proposed in connection to this classical theory, which are also considered by the SSIPA. Namely, the interference of emotions on the rational and cognitive processing of social information (de Castro, 2004), and the definition of evaluation criteria that may characterize response evaluation and lead to response decision (Fontaine & Dodge, 2006), particularly for adolescents. These evaluation criteria are: 1) initial acceptability of the response, considering if it is generally acceptable and applicable, and if it is congruent with personal values and standards, 2) likelihood that the individual can successfully and effectively perform the response option, and evaluation of its social and moral value, 3) outcome expectancy, its likelihood and value to the self, and 4) comparison of response options and selection of the best evaluated option.

Such systematic evaluation of the SIP has not been provided by the short list of psychometrically sound assessment instruments available for evaluating SIP in children,

namely the *Social-Cognitive Assessment Profile* (Hughes, Meehan, & Cavell, 2004), the *Social Information Processing Interview* (de Castro, 2000), and the *Social Information Processing Application* (Kupersmidt, Stelter, & Dodge, 2011). Instruments for evaluating social information processes in adolescence are even scarcer. *The adolescent stories* (Godwin & Maumary, 2004) are one of these options, and they have previously been used with adequate internal consistency values (Fontaine et al., 2010; Fontaine, Yang, Dodge, Bates, & Pettir, 2008). Nevertheless, a thorough study of its internal structure and construct validity has not been reported. The only instrument specific for adolescents known to the authors that was psychometrically evaluated is the *Cuestionario del procesamiento de la información social* (Calvete & Orue, 2009), which was adapted from the *Social Information Processing Interview*. The *likert* type version of this instrument evaluates hostile attribution of intent, anger, and selection of physical or verbal aggressive responses. Its results have shown adequate internal consistency values and a three factor internal structure. Still, this instrument does not consider emotions other than anger that may be associated with aggressive and non-aggressive behavior, nor does it take into account the developments made to the response evaluation step of the model. Additionally, its developmental validity may be questioned: it resulted from adapting a measure built to evaluate children, whose social demands are differently presented and solved in comparison to adolescents. Given that a developmental perspective on social information processing has been validated (Fontaine, Yang, Dodge, Bates, & Pettir, 2008), conclusions based on childhood or adulthood should not be transposed to adolescence. Concomitantly, adolescence may be an important period of life to study SIP given that such processing is at the same time well-established but not yet rigid (Horowitz, 1991), allowing for an accurate understanding, and, simultaneously, an effective intervention on its possible vulnerabilities. Therefore, the need to evaluate adolescents using

psychometrically sound instruments designed specifically for them should not be overlooked, as it may contribute to the continued substantiation of the SIP model

The purpose of the current work is, therefore, to psychometrically evaluate results obtained with the SSIPA, particularly in terms of item analysis based on corrected item-total and inter-item correlations (Murphy & Davidshofer, 2001), and of internal factor structure analysis, based on exploratory and/or confirmatory factor analysis when appropriate (Fabrigar, Wegener, MacCallum, & Strahan, 1999). The quality of the items being entered into a factorial analysis is paramount, as the subsequent factorial analysis will more likely result in a statistical and theoretical valid factor solution (Floyd & Widaman, 1995); thus, combining item and factor analytic approaches is an adequate and essential step in test construction (Kline, 2000). Given the theoretical framework for developing this instrument, we expect to find specific measurement models pertaining to four steps of the SIP: 1) Attribution of intent will be evaluated by hostile attribution and neutral attribution measures; 2) Emotion intensity will be evaluated by anger, sadness, and shame measures; 3) Response evaluation will be evaluated in connection with assertiveness, passiveness and aggression, which will in turn include overt and relational aggression; and 4) Response decision will be evaluated in relation with assertiveness, passiveness and aggression, which will in turn include overt and relational aggression.

Method

Participants

The participants for this study were 802 adolescents, being 65.1% ($n = 522$) female and 34.9% ($n = 280$) male. Their ages varied between 15 and 20 years old, with a mean age of 16.44 years old ($SD = 0.99$)¹. The complete sample attended secondary school and was uniformly distributed by school grade: 36% attended the 10th grade, 28.3% attended the 11th

¹ The mean age for female and male students was significantly different ($t(796) = -2.34; p = 0.02$), with girls being older ($M = 16.49, SD = 1.06$) than boys ($M = 16.32, SD = .085$).

grade (n = 227), and 35.3% attended the 12th grade (n = 283). Three female participants did not provide information on their school year (0.5%). The majority of the participants had never been retained in the same school grade before (62.5%; n = 501), while 37.2% had been retained between one to four times (n = 300). One boy did not provide information on his history of grade retentions (0.1%). To determine each student's socioeconomic status (SES), the profession of the students' parents was coded, according to the Portuguese Profession Classification²(Instituto do Emprego e Formação Profissional, 1994). The majority of the sample came from a low SES (53.1%, n = 426), 40.5% (n = 325) came from a medium SES, and a minority came from a high SES (3%; n = 24). Twenty seven students (19 boys and 8 girls) did not provide information on their parents' profession, therefore, their SES could not be categorized.

Instruments

Scenes for Social Information Processing in Adolescence (SSIP)

This instrument was conceived having as its basis a participatory (American Psychological Association, 1999) and phenomenological (Vogt, King, & King, 2004) approach to item development and evaluation, and intends to evaluate the interpretation and response evaluation and decision (Fontaine & Dodge, 2006) steps of the SIP model (Crick & Dodge, 1994), as well as emotions which may be present and concomitant with SIP. Accordingly, it includes four dimensions (*i.e.*, interpretation, emotion intensity, response evaluation and response decision), which are evaluated in connection to six hypothetical provocative scenes, presented in a first person and gender neutral perspective; all scenes and items pertaining to them were originally developed and presented to the participants in

² Examples of professions in the high socioeconomic status groups are judges, higher education teachers, or M.D.s; for the medium socioeconomic status group are nurses, psychologists, or school teachers; for the low socioeconomic group are farmers, cleaning staff, or undifferentiated worker. When the mother and fathers' professions were classified into different socioeconomic status, the highest SES coding was attributed to the family.

Portuguese (see Appendix A for the English version³ of the instrument, after item and factor structure analysis). These scenes included overtly and relationally provocative situations, because the nature of the provocation of ambiguous events may demand different SIP strategies and the selection of different forms of behavioral responses (Crick, Grotpeter, & Bigbee, 2002; Xie, Cairns, & Cairns, 2002).

The respondent is asked to imagine that the event was happening to him or her, and then rate the probability of one hostile and one neutral attribution for each scene. The respondent is then asked to rate the intensity of three negative emotions that might be present on those scenes: anger, sadness and shame. Next, he or she is presented with four options of social behavior, thus minimizing the possibility that the response choice will be biased by lack of options activated from previous social experiences, and asked to rate each one of them according to several evaluation criteria. The four behavioral options pertain to assertiveness, passiveness, overt aggression and relational aggression. The evaluation criteria are⁴: internal congruence (*i.e.*, How much would this behavior represent who you are?); response valuation (*i.e.*, How good or bad do you think this is, as a way of acting?); response self-efficacy (*i.e.*, How capable are you of acting like this?); personal outcome expectancy (*i.e.*, How would you feel about yourself if you acted like this?); and social outcome expectancy (*i.e.*, How much would other people like you if you acted like this?). Finally, the decision on a specific type of response is evaluated based on the probability of response. The respondent is asked to rate all response options because they may not be mutually exclusive: the same individual may ponder different types of behaviors when faced with different ambiguous/ provocative events or perpetrators (Sumrall, Ray, & Tidwell, 2000).

Procedure

³ For the development of the English version of the instrument, one of the authors and one independent researcher fluent in English individually translated the Portuguese version to English. Discrepancies between these two translated versions were agreed upon with the help of another author of the manuscript, who acted as third party translator.

⁴ The initial acceptability criteria is considered to be fulfilled, because the behavioral options were directly taken from adolescents' verbalizations of common and acceptable responses in each one of the scenes (Vagos et al., 2013)

This study was approved by the national committee for the evaluation of ethics and procedures for studies conducted in school settings. Afterwards, authorization was sought and given by the participating schools and by the parents of participants under 18 years old. One member of the research team went to each school and classroom to request the voluntary participation of students, to whom the confidentiality of the data was guaranteed. Information on the research was presented in an introductory sheet accompanying the instrument, where socio-demographic information was also asked. The questionnaires took about 20-25 minutes to complete.

Data analysis was conducted using SPSS (v18.0), MPlus (Muthén, & Muthén, 2010) and R (3.0.1; R Development Core Team, 2013). SPSS was used for corrected item-total correlation (*i.e.*, correlation between one item and the sum of the items that compose the measure to which the item belongs, excluding the item itself) and inter-item bivariate correlation, and descriptive analysis. Values ranging from 0.30 to 0.70 and from 0.20 to 0.50 were considered acceptable for the corrected item-total correlations and for the inter-item correlations, respectively (Ferketich, 1991).

MPlus was used to perform exploratory factor analysis (EFA) on each set of items proposed for the evaluation of each type of attribution of intent, each type of emotion and each type of response decision. Given that EFA has not been previously reported for measures addressing these particular constructs, nor has a conceptual framework been given for such constructs so that its measures might be confirmed via confirmatory factor analysis (CFA), exploratory analysis was, at this point, the most appropriate option (Costello & Osborne, 2005; Fabrigar et al., 1999). Parallel analysis (PA) was taken into account as indicative of the dimensionality of these measures, given that it has been shown to be one of the most accurate methods for determining the number of factors to retain (Glorfeld, 1995). In addition, the overall fit of the exploratory measurement models was evaluated based on a two-index

approach proposed by Hair Jr., Black, Babin and Anderson (2005), which takes into account the sample size and number of items in each analysis (*i.e.*, ≤ 6 items in these cases).

Acceptable fit was based on achieving values for *Root Mean Square Error of Approximation* (RMSEA) lower than .07 in combination with *Comparative Fit Index* (CFI) values higher than .97. When the PA and the fit indicators seemed to be inconsistent in relation to the number of factors to retain, PA was given priority, and the factorial solution was further examined to better understand and consequently solve these inconsistencies. Measurement models were, therefore, only established after abiding by the PA suggestion and simultaneously achieving acceptable overall fit indicators.

Mplus was also used to perform CFA analysis on the measurement models for the response evaluation measures, in testing the conceptual framework proposed by Fontaine and colleagues (2006; 2010). Further analysis on these measures included both EFA and CFA approaches, to explore and verify the models best fitting the data taken from the current sample. For these measures, which would have a maximum of 30 items, acceptable fit was considered based on RMSEA values lower than .07 combined with CFI values higher than .92⁵ (Hair Jr. et al., 2005).

Once the measurement models were defined, internal consistency analyses were carried out on all measures using the ordinal alpha, given that it is more appropriate than the *Cronbach Alpha*, when analyzing likert-type (*i.e.*, ordinal) measures. The ordinal and the Cronbach Alpha are conceptually similar (Gadermann, Guhn, Zumbo, & Columbia, 2012), and so values representing modest reliability (*i.e.*, 0.70) were deemed acceptable (Nunnally, 1978).

⁵ Hair Jr. et al (2005) suggest, as an alternative, considering *Standardized Root Mean Square Residual* (SRMR) values lower than .08 combined with CFI values higher than .92. Given that our data deviated from the normal distribution, the SRMR would not be calculated, and so this combination of fit indicators was not considered in the present work.

Furthermore, and following the measurement models having been established via either EFA or CFA, factor invariance of these models in relation to gender was tested, using a CFA based and forward approach as described by Dimitrov (2010): configural, then metric and then scalar invariance were examined. Configural invariance indicates that the same basic factor structure is stable across groups and so was analyzed evaluating the fit of the measurement models separately for boys and girls. Metric invariance determines that loadings are similar across boys and girls, of each item on its corresponding factor for first-order measurement models and also of the first order on the second order factors in the case of second-order measurement models. Finally, scalar invariance adds to the constraint of loading equality, the imposition that variables' thresholds have to be invariant across groups for first-order measurement models and also the invariance of first-order factor means in the case of second-order measurement models (Dimitrov, 2010). For this testing, a unit loading constraint on the 1st item of each factor was used for scaling purposes for first order measurement models; for the second order models, a unit variance constraint in addition to a unit loading constraint was used on the 1st order factor of the second order factors for scaling purposes (Kline, 2011).

For the purpose of construct validity, results obtained for boys and girls on the measures found for the SSIPA were compared; if results pertaining to group differences found with the SSIPA are in line with gender differences found in the literature, this may add to the evidence that the SSIPA evaluates its proposed constructs. Following factor invariance analyses, a latent mean comparison approach was used (Dimitrov, 2006), to accommodate for only partial invariance being found for some of the measures. The male group was taken as the reference group, and so its mean was fixed to zero.

Results

Item analytic procedures

Inter-item and corrected item-total correlations were conducted as the first step in item evaluation, intending to pinpoint items that would be problematic to use in the following EFA or CFA, by being too highly (*i.e.*, redundant) or too little (*i.e.*, irrelevant) associated with the remaining items or with the total score. Items were considered problematic if they consistently showed borderline results or if they simultaneously showed low inter-item and low corrected item-total correlation. In the first case, they were recorder for further interpretation of the potential misfit in factorial solutions; in the second case, they were excluded from the subsequent analyses. With clarity in mind, only results pertaining to problematic items are presented next⁶.

For the six items intending to evaluate neutral attribution of intent, two correlations were below the cutoff value of .20 for inter-item correlation: $r = .14$ between items taken from scenes 1 and 6 and $r = .19$ between items taken from scenes 1 and 4. Item 1 refers to a relationally provocative scene whereas items 4 and 6 refer to overtly provocative scenes. None of these items simultaneously surpassed the cutoff criteria for corrected item-total correlation. As for the six items intending to evaluate hostile attribution of intent, all abided by the cutoff criteria for both inter-item and corrected item-total correlations

All six items intending to evaluate anger and shame abided by the cutoff criteria for both inter-item and corrected item-total correlations. As for sadness, only the correlation between items taken from scenes 1 (relational) and 2 (overt) was below the cutoff value of .20 ($r = .139$); both items nevertheless abided by the corrected item-total cutoff criteria.

Concerning the response evaluation criteria, all thirty items intending to evaluate assertiveness and overt aggression simultaneously fulfilled the inter-item and the corrected item-total cutoff values. For passiveness, items taken from scenes 1, 5 and 6 had correlation values lower than the cutoff value of $r = .20$ for inter-item correlations, which may indicate

⁶ Complete results may be requested from the first author.

that they are not addressing similar constructs to those evaluated by items pertaining to other scenes; all items nevertheless abided by the cutoff criteria for corrected item-total correlation. For relational aggression, the item evaluating internal congruence for scene 5 correlated lower than the cutoff value of .20 with the item evaluating social outcomes for scene 1 and with the item evaluating personal outcomes for scene 2; this item nonetheless abided by the cutoff criteria for corrected item-total correlation. It is worth noting that items evaluating internal congruence for all scenes attained the lowest correlation values for inter-item (particularly with items intending to evaluate social goals) and corrected item-total correlations. Those items may, subsequently, represent problematic items that seem neither to strongly correlate with the remaining items of the scale nor with a possible complete score.

All six items intending to evaluate response decision for three (*i.e.*, assertiveness, overt aggression and relational aggression) out of the four behavioral options under study simultaneously abided by the inter-item and corrected item-total cutoff values. As for passiveness, the item pertaining to scene 5 achieved very low correlations with the five remaining items ($r < .20$) and with the total score ($r = .174$). It was, therefore, excluded from the remaining analyses.

Factor analysis

In order to select the best estimator for EFA and CFA, multivariate kurtosis and multivariate skewness were tested based on Mardia's equations (Mardia, 1970); the Henze-Zirkler's multivariate normality test was also used (Henze & Zirkler, 1990). All tests were significant ($p < .001$) for each set of six items evaluating neutral and hostile attribution of intent, anger, shame and sadness, and response decision for assertiveness, overt aggression and relational aggression. These tests were also significant for the set of five items evaluating response decision for passiveness and for each set of thirty items evaluating response evaluation for assertiveness, passiveness, overt aggression and relational aggression. None of

the data was, therefore, multivariate normal, and consequently a robust *Weighted Least Squares* estimator was used (Flora & Curran, 2004).

Items were included in the factor solution if they had $\lambda \geq .32$ in only one factor and cross-loadings $\leq .32$. Items not matching these criteria were dropped (Tabachnick & Fidell, 2006). For items with only one $\lambda \geq .32$, the cross-loading values were not considered for the composition of the factor. With clarity in mind, figures representing the 9 parallel analyses conducted within this study are not present and only the fit indicators of the retained exploratory or confirmatory factor models are presented; similarly, only the $\Delta\chi^2$ but not the fit indicators for metric and scalar invariance are presented⁷.

Attribution of intent measures

The six items that evaluate neutral attribution of intent were submitted to EFA. PA indicated that only one factor should be retained, and a one-factor solution fitted the data very well ($RMSEA = .048$, CI for $RMSEA = .027, .070$; $CFI = .99$; λ varied between .414 for item taken from scene 1 and .643 for item taken from scene 4). This one-factor measurement model, however, did not achieved acceptable fit for the male sample ($RMSEA = .11$, CI for $RMSEA = .076, .146$; $CFI = .94$), leading us to further explore the items (*i.e.*, inter-item and corrected item-total correlations) for the male and female samples separately.

For the male sample, item 1 presented lower than acceptable correlations with items 2 ($r = .15$, non-significant), 4 ($r = .17, p = .004$) and 6 ($r = 0.04$, non-significant), and also with the total score of six items ($r = 0.27, p < .001$). Using and EFA on the male sample alone, and excluding item 1 still resulted in only a two-factor solution achieving acceptable fit, when PA suggested retained only one factor. Item 6 was subsequently excluded, because it was the only item loading on the first factor. This resulted in an one-factor solution presenting acceptable fit for the male sample ($RMSEA = .031$, CI for $RMSEA = .000, .127$; $CFI = .99$; λ varied

⁷ A complete result description may be requested from the first author.

between .616 for item taken from scene 3 and .663 for item taken from scene 4), which was further verified by CFA ($RMSEA = .031$, CI for $RMSEA = .000, .0127$; $CFI = .99$). For the female sample, both a six-item one-factor solution and this four-item one-factor solution fit the data very well ($RMSEA = .044$, CI for $RMSEA = .012, .073$; $CFI = .99$ and $RMSEA = .059$, CI for $RMSEA = .000, .012$; $CFI = .99$, respectively). Measurement invariance was subsequently tested on the four-item one-factor solution, and results showed full metric (M1⁸ – M0⁹: $\Delta\chi^2 = 4.22$, $df = 3$, $p = 0.23$) and full scalar invariance (M2¹⁰ – M1: $\Delta\chi^2 = 5.76$, $df = 4$, $p = 0.22$).

The four-item one-factor model also achieved acceptable fit using the complete sample ($RMSEA = .000$, CI for $RMSEA = .000, .050$; $CFI = 1.00$), and in addition attained an acceptable ordinal alpha value (Table 1), which was all in all very similar to the ordinal alpha value attained for a six-item one-factor model (.73). Therefore, and for the sake of simplicity when using this measure to evaluate neutral attribution, the four-item one-factor model seems the optimal solution.

The six items evaluating hostile attribution of intent were submitted to EFA. PA indicated that only one factor should be retained, but only a two factor solution achieved an acceptable fit ($RMSEA = .035$, CI for $RMSEA = .00, .071$; $CFI = .99$). We proceeded with analyzing the loadings on a two-factor solution that might be preventing a one-factor solution from adjusting, and found that item 4 presented a negative loading for the first factor. Excluding this item resulted in a five-item one-factor solution achieving acceptable fit ($RMSEA = .032$, CI for $RMSEA = .000, .070$; $CFI = .99$). Switching to a CFA approach, this measurement model fitted acceptably in what concerns both boys ($RMSEA = .057$, CI for $RMSEA = .000, .011$; $CFI = .99$) and girls ($RMSEA = .043$, CI for $RMSEA = .000, .085$; $CFI =$

⁸ M1 represents the model with full equality constraints for all loading values.

⁹ M0 represents the baseline unconstrained model.

¹⁰ M2 represents the model with full equality constraints for all loading and threshold/ intercept values.

.99), thus indicating configural invariance of this model across gender. Full metric invariance (M1– M0: $\Delta\chi^2 = 6.90$, $df = 4$, $p = 0.14$) and full scalar invariance (M2 – M1: $\Delta\chi^2 = 7.53$, $df = 5$, $p > 0.18$) were also found for this measurement model, as well as an acceptable ordinal alpha value (Table 1).

Emotion measures

The six items pertaining to anger were used for EFA. PA indicated that only one factor should be retained, but only a two-factor solution attained acceptable fit ($RMSEA = .003$, CI for $RMSEA = .000, .053$; $CFI = 1.00$). Analyzing the loading values for the two-factor solution to better understand what might be contributing to the poor adjustment of a one-factor solution, we again found a negative loading for item taken from scene 4, which was, consequently, excluded. This resulted in an exploratory one-factor solution achieving acceptable fit ($RMSEA = .0035$, CI for $RMSEA = .000, .067$; $CFI = .99$; λ varied between .571 for item taken from scene 2 and .791 for item taken from scene 3), which also was confirmed to fit acceptably to the male ($RMSEA = .040$, CI for $RMSEA = .000, .099$; $CFI = .99$) and female samples separately ($RMSEA = .027$, CI for $RMSEA = .000, .071$; $CFI = .99$), thus indicating configural invariance. Full metric invariance (M1 – M0: $\Delta\chi^2 = 8.34$, $df = 4$, $p = 0.08$) and full scalar invariance (M2 – M1: $\Delta\chi^2 = 9.09$, $df = 5$, $p = 0.10$) were subsequently found for this measurement model, in addition to an acceptable ordinal alpha value (Table 1).

The six items pertaining to sadness were submitted to an EFA, and the results of PA and overall fit were evaluated sequentially, after excluding item 6 due to cross-loading ($\lambda = .439$ for the first factor and $\lambda = .394$ for the second factor) and then item 4 as it presented a Heywood case (i.e., negative residual variance) and it was the only one with $\lambda > .32$ for the first factor. PA on an EFA using items 1 through 3 and item 5 suggested retaining only one factor, in accordance with a one factor solution producing acceptable fit ($RMSEA = .000$, CI for $RMSEA = .000, .063$; $CFI = 1.00$), with loadings ranging from .430 (item taken from scene

2) to .838 (item taken from scene 3). This four-item one-factor measurement model attained acceptable ordinal alpha value (Table 1) and fitted very well for boys ($RMSEA = .000$, CI for $RMSEA = .000, .063$; $CFI = 1.00$) and for girls ($RMSEA = .071$, CI for $RMSEA = .017, .123$; $CFI = .99$), thus pointing to configural invariance. Full metric invariance was achieved across gender for this one-factor measurement model (M1 – M0: $\Delta\chi^2 = 6.66$, $df = 3$, $p = 0.083$), as well as partial scalar invariance after allowing the threshold for the first response category of item 1 to vary across groups (M2P¹¹ – M1: $\Delta\chi^2 = 4.14$, $df = 3$, $p = 0.25$).

An EFA was conducted using the six items referring to shame. PA suggested retaining only one factor, but only a two-factor solution fulfilled the fit criteria ($RMSEA = .011$, CI for $RMSEA = .000, .055$; $CFI = 1.00$). Subsequently, we considered the loadings of the two factor solution to pinpoint items that might be preventing a one-factor solution from adjusting. Item 1 and then item 3 were excluded, as they presented presenting negative loading values, rendering an exploratory one-factor solution acceptable ($RMSEA = .038$, CI for $RMSEA = .000, .088$; $CFI = .99$; λ varied between 0.678 for the item taken from scene 2 and .820 for the item taken from scene 5). This factor solution was also confirmed to fit acceptably to the male ($RMSEA = .000$, CI for $RMSEA = .000, .103$; $CFI = 1.00$) and female samples ($RMSEA = .046$, CI for $RMSEA = .000, .108$; $CFI = .99$) separately, thus indicating configural invariance. Full metric (M1 – M0: $\Delta\chi^2 = 2.72$, $df = 3$, $p = 0.44$) and partial scalar invariance (M2P – M1: $\Delta\chi^2 = 2.66$, $df = 3$, $p = 0.05$) were also found, after allowing the threshold for the second response category of item 4 to vary between groups. This four-item one-factor solution attained an acceptable ordinal alpha value (Table 1).

Response evaluation measures

Measurement models for individual response evaluation of assertiveness, passiveness, overt aggression and relational aggression individually were proposed based on the theoretical

¹¹ M2P represents the model with partial equality constraints for loading and intercept values, where at least one threshold/intercept value was allowed to vary between groups.

premises that sustained these response scales (Fontaine & Dodge, 2006) and that have been used to assume scales to evaluate each criteria for response evaluation (Fontaine et al., 2010). These models included five measures, each pertaining to one evaluation criteria: response moral valuation, self-efficacy, personal outcome, and social outcome. These models did not achieve acceptable fit, thus questioning the manner in which these measures have been used in previous studies.

Considering these results, and without a theoretical framework for this analysis, we proceeded with separate EFA for the thirty items that compose the response evaluation measures for assertiveness, passiveness, overt aggression, and relational aggression, aiming to explore the best measurement models for this data. The only solutions achieving acceptable fit criteria consisted of seven factors for each of the behavioral options, in which no item loaded above .32 on the seventh factor. The remaining six factors seem to organize into scenes. Given that the six factor solution did not achieve an acceptable fit, but the seventh factor solution was non-interpretable, a CFA approach was then used to investigate an empirical hypothesis based on the results previously reported for attribution of intent and emotion intensity (*i.e.*, a one factor model) and a theoretically based hypothesis considering the type of provocation (*i.e.*, a two-factor model, for relationally and overtly provoked behavior; Sumrall, Ray, & Tidwell, 2000). None of them achieved acceptable fit; the best fit was always found for the two-factor solution. Modification indices for these solutions indicated that associations between items belonging to the same scene or same evaluation criteria were to be considered, mirroring the EFA analysis results. Results also indicated low loading values for items evaluating internal congruence, which were pinpointed as problematic for attaining the lowest inter-item and item-total correlations. These items were, thus, excluded from the remaining analysis.

Consequently, two new higher order models were proposed: a) type of provocation (relational *versus* overt) as higher order factors for scenes as first-order factors; and b) type of provocation (relational *versus* overt) as higher order factors for 5 evaluation criteria as first-order factors. Only the first models presented acceptable fit indicators for all the response evaluation measures, namely assertiveness (Figure 1.A), passiveness (Figure 1.B) overt aggression (Figure 1.C) and relational aggression measures (Figure 1.D). Both measures (*i.e.*, relationally and overtly provoked) for each type of response option (*i.e.*, assertiveness, passiveness, overt aggression and relational aggression) always surpassed the .70 cutoff value for internal consistency analyses (Table 1).

The second order measurement model for assertiveness fitted very well for boys ($RMSEA = .069$, CI for $RMSEA = .061, .076$; $CFI = .97$) and for girls ($RMSEA = .075$, CI for $RMSEA = .070, .080$; $CFI = .97$), thus indicating configural invariance. Full metric invariance was found across gender (M1 – M0: $\Delta\chi^2 = 34.04$, $df = 24$, $p = .084$), as well as full scalar invariance (M2 – M1: $\Delta\chi^2 = 32.66$, $df = 23$, $p = 0.08$). In the case of passiveness, the model fitted very well for the male ($RMSEA = .074$, CI for $RMSEA = .066, .081$; $CFI = .96$) and for the female sample ($RMSEA = .059$, CI for $RMSEA = .054, .064$; $CFI = .97$), thus indicating configural invariance. Full metric invariance was achieved (M1 – M0: $\Delta\chi^2 = 22.48$, $df = 18$, $p = .21$), as well as full scalar invariance (M2 – M1: $\Delta\chi^2 = 10.56$, $df = 23$, $p = 0.98$). The same measurement model for overt aggression fitted very well for boys ($RMSEA = .058$, CI for $RMSEA = .050, .065$; $CFI = .98$) and girls ($RMSEA = .043$, CI for $RMSEA = .037, .048$; $CFI = .99$), thus indicating configural invariance. Full metric invariance (M1 – M0: $\Delta\chi^2 = 14.32$, $df = 17$, $p = .64$) and partial scalar invariance were also achieved for this model (M2P – M1: $\Delta\chi^2 = 33.48$, $df = 23$, $p = 0.07$), after allowing the threshold for the first response category of item 20 to vary between groups. The second order measurement model for relational aggression fitted very well for the male ($RMSEA = .062$, CI for $RMSEA = .055, .070$; $CFI = .98$) and the

female sample separately ($RMSEA = .052$, CI for $RMSEA = .047, .057$; $CFI = .99$), thus indicating configural invariance. Full metric (M1 – M0: $\Delta\chi^2 = 14.32$, $df = 17$, $p = .64$) and full scalar invariance was achieved between groups (M2 – M1: $\Delta\chi^2 = 24.45$, $df = 23$, $p = 0.37$).

Response decision measures

The six items intending to evaluate assertiveness were entered into an EFA. PA suggested retaining one factor, contrasting with only a two-factor solution achieving acceptable fit. In trying to define a homogeneous one-factor solution, item 5 was excluded, because it presented a Heywood case and was the only item with $\lambda > .32$ for the second factor. The remaining five items were again subjected to an EFA which produced a one-factor acceptable solution ($RMSEA = .057$, CI for $RMSEA = .030, .086$; $CFI = .99$), in accordance with the PA, but it nonetheless did not abide by the overall fit criteria, for either the male ($RMSEA = .072$, CI for $RMSEA = .019, .124$; $CFI = .99$) or the female sample ($RMSEA = .071$, CI for $RMSEA = .038, .108$; $CFI = .99$). Again, item analyses were not informative, and so EFA were performed separately for the male and female sample. PA always suggested retaining only one factor, but the two-factor solution was the only one achieving acceptable fit. For boys, items 5 and 4 were sequentially excluded, as they had negative loading values. This resulted in an acceptable one-factor solution for boys ($RMSEA = .000$, CI for $RMSEA = .000, .082$; $CFI = .99$), attaining an ordinal alpha value of .69, which nonetheless did not fit well for girls. In turn, the EFA for girls resulted in excluding items 5 and 3 due to negative loadings, leading to an acceptable one-factor solution ($RMSEA = .000$, CI for $RMSEA = .000, .076$; $CFI = 1.00$) with an ordinal alpha value of .71. Again, this measurement model did not adjust for boys, and so measurement invariance regarding gender could not be ascertained for the measurement models underlying boys' and girls' responses to the response decision on assertiveness.

For passiveness, item 5 was excluded *a priori*, following item analytic procedures.

The remaining five items were subjected to EFA. PA suggested retaining one factor and the one-factor solution produced acceptable fit indicators ($RMSEA = .067$, CI for $RMSEA = .041$, $.096$; $CFI = .98$).

This five-item one-factor measurement mode did not, however, fit acceptably to the male sample ($RMSEA = .088$, CI for $RMSEA = .042$, $.139$; $CFI = .97$). Item analysis undertaken separately by gender was not informative and so we proceeded with EFA for boys and girls separately. In both cases, the item taken from scene 6 proved problematic, presenting a negative variance and being the only one loading on the second factor, when a two-factor solution was the only one achieving acceptable fit. The exclusion of item 6 resulted in an exploratory and confirmatory acceptable fit for boys ($RMSEA = .000$, CI for $RMSEA = .000$, $.011$; $CFI = 1.00$) and for girls ($RMSEA = .000$, CI for $RMSEA = .000$, $.072$; $CFI = 1.00$), and was confirmed as a good fit for the complete sample ($RMSEA = .030$, CI for $RMSEA = .000$, $.082$; $CFI = .99$). This four-item one-factor solution attained a very close to acceptable consistency value (Table 1). Partial metric invariance for measurement model was established (M1P¹² – M0: $\Delta\chi^2 = 4.64$, $df = 2$, $p = 0.09$), after allowing the loading of the item vary between boys and girls. Subsequent full scalar invariance (M2 – M1P: $\Delta\chi^2 = 3.99$, $df = 4$, $p = 0.41$) was also found.

The six items that evaluate overt aggression were submitted to an EFA. PA suggested retaining one factor and the one-factor solution achieved acceptable fit indicators ($RMSEA = .063$, CI for $RMSEA = .043$, $.084$; $CFI = .99$). This measure nonetheless did not fit acceptably in what concerns the male sample ($RMSEA = .108$, CI for $RMSEA = .075$, $.145$; $CFI = .97$). Due to the fact that item analysis was not informative on problematic items, we proceeded with an EFA using the male sample only. Items 3 and then 5 were excluded as they presented

¹² M1P represents the model with a partial constraint of equality of loading values, where at least one loading value was allowed to vary between groups.

a negative loading value on a two-factor solution, and, therefore, they could be preventing a one-factor solution from adjusting. The resulting four-item one-factor model had acceptable fit for boys ($RMSEA = .052$, CI for $RMSEA = .000, .140$; $CFI = .99$). For girls, both a six-item one-factor solution and the four-item one-factor solution seemed to acceptably fit the data ($RMSEA = .050$, CI for $RMSEA = .021, .079$; $CFI = .99$ and $RMSEA = .070$, CI for $RMSEA = .020, .129$; $CFI = .99$, respectively). Thus, we proceeded with the multi-group analysis considering the four-item one-factor solution, and found full metric invariance (M1 – M0: $\Delta\chi^2 = 1.48$, $df = 3$, $p = 0.68$) and partial scalar invariance after allowing the threshold for the first response category of item 6 to vary between groups (M2P – M1: $\Delta\chi^2 = 4.47$, $df = 3$, $p = 0.22$). This four-item one-factor measurement model also acceptably fitted the data from the complete sample ($RMSEA = .045$, CI for $RMSEA = .000, .094$; $CFI = .99$), in addition to attaining acceptable ordinal alpha values for the complete sample (Table 1), similarly to that obtained for the six-item one-factor solution (.89). Therefore, and having simplicity in mind when using this measure to evaluate neutral attribution, the four-item one-factor model seems the optimal solution.

For an EFA on the six items that evaluate relational aggression, PA suggested retaining one factor. A one-factor solution presented acceptable fit indicators ($RMSEA = .068$, CI for $RMSEA = .048, .089$; $CFI = .99$) for the complete sample, but was unacceptable for the male sample alone ($RMSEA = .117$, CI for $RMSEA = .083, .125$; $CFI = .97$). Item analyses were not informative on which could be the problematic items. Proceeding with EFA on the male sample, only a two factor solution achieved acceptable fit indicators, though PA suggested retaining one factor. Item 5 had a very low loading value for both factors in that solution, and so was excluded. An EFA with the remaining five items attained acceptable fit indicators with the male sample ($RMSEA = .062$, CI for $RMSEA = .000, .115$; $CFI = .99$). For the female sample, both the six-item one-factor solution and the five-item one-factor solution

achieved acceptable fit ($RMSEA = .051$, CI for $RMSEA = .023, .079$; $CFI = .99$ and $RMSEA = .026$, CI for $RMSEA = .000, .070$; $CFI = .99$), respectively. Testing the factorial invariance of the five-item one-factor solution resulted in full metric invariance (M1 – M0: $\Delta\chi^2 = 0.88$, $df = 4$, $p = 0.92$) and partial scalar invariance after allowing the threshold for the second response category of item 3 to vary between groups (M2P – M1: $\Delta\chi^2 = 8.08$, $df = 4$, $p = 0.09$). This five-item one-factor measurement model also acceptably fitted the data from the complete sample ($RMSEA = .046$, CI for $RMSEA = .017, .077$; $CFI = .99$), in addition to attaining an acceptable ordinal alpha value (Table 1), higher than that obtained for the six-item one-factor solution (.82). Therefore, and having simplicity in mind, when using this measure to evaluate neutral attribution, the four-item one-factor model seems the optimal solution.

Descriptive analysis

Descriptive measures could only be computed after the measurement models for all measures had been defined. They are presented in Table 1 for the seventeen measures found to be evaluated by the SSIPA, for the complete sample and differentiated by gender.

Normality analyses suggest that four of these measures deviate from the normal distribution: shame, and response evaluation and decision for overtly and relationally provoked overt and relational aggression.

Concerning gender comparisons, we first present the results for the measures that obtained full metric and full scalar invariance, including the mean value for the comparative group (*i.e.*, girls, *versus* the reference group, boys, to whom the mean response value was placed at 0.00; Dimitrov, 2006). For the attribution measures, girls presented higher values for the neutral attribution (.181, $p = .002$; *cohen d* = .23) and for hostile attribution (non-significant). In what concerns emotion intensity, girls reported more anger (non-significant). Girls also endorsed a better evaluation of the assertive behavior, when relationally provoked

(0.128, $p = 0.014$, *cohen d* = .21) and when overtly provoked (non-significant), and of the passive behavior when relationally provoked (non-significant) and when overtly provoked (0.199, $p < .001$, *cohen d* = .32). On the contrary, boys thought better of the relational aggression behavior option, when relationally provoked (-0.293 , $p < .001$, *cohen d* = .36) and when overtly provoked (0.314, $p < .001$, *cohen d* = .35). The effect sizes for these comparisons were small to medium.

For the measures attaining only partial invariance (though it usually represented a minority of differential functioning items), we analyzed the latent mean comparison significance value when full invariance *versus* partial invariance was considered, to ascertain if results were stable across the two conditions, and therefore robust. For sadness, the difference was significantly different in both conditions ($p < .001$; *cohen d* = .35), with a mean of 0.270 for the full invariance condition and a mean of 0.240 for the partial invariance condition. For shame, the difference was always non-significant. For the response evaluation of overt aggression, the difference was significant for the relationally provoked ($p = .001$; *cohen d* = .33) and overtly provoked ($p < .001$; *cohen d* = .47) scenarios, in both conditions. For the full invariance condition, mean values were -0.217 for the relationally provocative and -0.402 for the overtly provocative scenarios; for the partial invariance condition, they were -0.197 and -0.402, respectively. Thus, in both cases, boys favored the over aggressive behavior option. Girls reported a significantly higher probability of deciding on a passive behavioral option ($p < .001$, *cohen d* = .31), either for the full (mean = 0.167) or the partial metric invariance condition (mean = 0.229). In the case of choosing an overt aggressive behavior, the difference was significant in both the full and partial invariance conditions ($p < .001$; *cohen d* = .48), favoring boys, with a mean value of -0.271 for the full and of -0.384 for the partial invariance condition. Lastly, in the case of choosing a relationally aggressive behavior, the difference was significant in both conditions ($p = .001$; *cohen d* = .36), favoring

boys, with a mean value of -0.306 for the full and of -0.298 for the partial invariance condition. The effect sizes for these measures were medium to large. The significance level for all measures was the same across invariance conditions, and the difference in the mean values for the comparative group (*i.e.*, girls) was always very small, pointing to consistency and reliability in the results obtained for these measures when comparing means, despite the partial invariance across gender.

Discussion

The SIP model proposes that several cognitive steps take place in an interchangeable manner when any social event is encountered to determine the final behavioral response that is enacted in such events (Crick & Dodge, 1994). Strong evidence has early and continuously been presented for this model, and this has led to several advances, the most significant of which being the consideration of emotional interference in SIP (Lemerise & Arsenio, 2000), and the definition of different evaluation criteria for any behavioral response options (Fontaine & Dodge, 2006). Despite this vast research, the instruments used in this area usually do not obey the standards for psychological testing (APA, 1999), since their development process and their psychometric characteristics are scarcely reported.

The goal of this work was to evaluate the psychometric quality of the results of an instrument built to evaluate four steps of SIP in adolescence, namely attribution of intent, emotion intensity, response evaluation and response decision. To achieve the goals of this study, two approaches were used, one based on item analysis, to guarantee the quality of the items (*i.e.*, discriminability and contribution to the complete constructs), and one based on factor structure analysis of the instrument, to better ascertain which constructs might be under evaluation. Internal consistency was also considered as an indicator of item quality and of the homogeneity of the constructs under evaluation. Finally, gender comparisons were undertaken, in order to provide preliminary norms for score interpretation, as well as to gather

evidence of construct validity, by analyzing if results were in line with what had been previously found with instruments addressing similar or theoretically associated constructs.

The item analyses showed that most items seemed to be pertinent and address the same constructs amongst themselves and as a whole (Murphy & Davidshofer, 2001). The exceptions were items from the passive behavioral option from scene 5, which did not sufficiently correlate with the remaining items and with the total score of the scale. The original wording of this item (*do nothing and go to the movie on my own*) implied inactivity and initiative at the same time, making it possibly contrary to the idea of passiveness portrayed by the passive behavior options of the remaining scenes, which were connected with simply doing nothing and trying to remain unnoticed, not taking any initiative.

Factorial analysis on the items produced several measures which are closely in line with our hypothesized measurement models, and that may be better discussed by construct. Most of these models were equally applicable to the evaluation of girls' and boys' experiences; some of them presented only partial invariance, though only a minority of items were functioning differently and results taken from latent mean comparison seem robust and pointing to the same findings regardless of considering full or partial invariance for these measures. Therefore, partial invariance did not seem to have an influence on the reliability of the results obtained from mean comparisons, which, therefore, will not be discussed in light of this condition. In contrast, one measure produced different measurement models for boys and girls. This complete variance of results by gender will be duly discussed.

Attribution of intent was measured by neutral attribution on the one hand and hostile attribution on the other. The consideration of neutral and hostile attribution as separate measures is in line with the notion of negative and positive interpretation of social events not laying on the opposite ends of a single continuum for, for example, socially anxious individuals (Huppert, Foa, Furr, Filip, & Mathews, 2003). In relation to aggression and

prosocial behavior however, the evaluation of these dimensions simultaneously has not been considered, and so these are new findings. Previous research used mutually exclusive response options, and consistently found that aggression associates with hostile attribution (de Castro et al., 2002), and some evidence has been found for prosocial adolescents presenting a more neutral or even slightly positive attribution of intent (Nelson & Crick, 1999). Using the SSIPA will allow testing these assumptions, based on the co-existence of both neutral and hostile attribution styles, in different social behavior groups. For instance, it seems predictable, based on previous findings, that aggressive adolescents present high hostile attribution in conjunction with low neutral attribution, and that prosocial adolescents present the opposite pattern, but the social behaviors of adolescents who consider both attributions as equally low or highly probable may also be informative and remains unclear. This may, indeed, represent cognitive flexibility and balance of positive and negative thoughts, representative of psychological and social adaptability (Elliott & Lassen, 1997). Regarding socio-demographic differences, adolescent girls significantly endorsed a more neutral attribution style in comparison with boys, which was in line with previous findings (Nelson & Crick, 1999).

SIP has seldom been studied in relation to emotion, even if emotion may serve as its antecedent and/or consequence (Crick & Dodge, 1996). When it has been studied, it has focused on the ability of aggressors to manage or cope with their emotions (*e.g.*, Marsee & Frick, 2007; Prinstein, Boergers, & Vernberg, 2001), rather than pinpointing specific emotions associated with interpersonal provocation. Previous works suggest that different emotions (*i.e.* angry or upset) are highly correlated and so would be best evaluated by one single measure, even if they may be distinguishable by the type of provocation of the probe-scenario (Crick et al., 2002). The present findings point to single measures underlying the three type of negative emotions under study, namely, sadness, shame and anger. These

emotions may be serving different purposes when dealing with provocation. Sadness may be a more diffuse and prevalent emotion, signifying a potential loss, whereas shame is associated with not having fulfilled personal standards, and anger is caused by experiencing an offense from others against the self (Lazarus, 2006). Diverse events may simultaneously activate different emotions, which in turn may have an impact on several behaviors practice by the self and others, and so they should not be considered as mutually exclusive or in isolation. For example, experiencing shame has been put forward as being connected with attacking others (*i.e.*, overt aggression; Elison, Lennon, & Pulos, 2006), and so has anger (Calvete & Orue, 2010, 2012; Castro & Merk, 2005). As for differences based on gender, girls experienced significantly more sadness than boys, concurring with the findings that girls are usually sadder than boys (Kubik, Lytle, Birnbaum, Murray, & Perry, 2003); for anger and shame, the differences between boys and girls were not significant.

The importance of contextual clues in appraising different types of social behavior became evident in the measurement models for the response evaluation measures, which were organized into overtly and relationally provoked responses. These findings diverge from using the different criteria put forward as underlying the response evaluation steps of the SIP as single measures (Fontaine et al., 2010); such measures had not been scrutinized statistically, perhaps because they lack internal consistency to stand on their own (Bailey & Ostrov, 2008). The present findings recommend their combined use as a single measure, albeit distinguished by type of provocation, and question the viability of the use of criteria for the evaluation of possible behavioral options, at least when they are examined using self-report instruments, instead of, for instance, an interview format as in Fontaine and Dodge (2006).

Future works may better ascertain the pertinence of these evaluation criteria measures for the explanation and prediction of different types of social behavior, following what has been undertaken for aggression and particularly for the normative beliefs or moral value

attributed to such behavior (Werner & Hill, 2010; Werner & Nixon, 2005). Considering our items' analytic procedures and the theoretical approaches to these concepts, we would expect that the response valuation criteria might more strongly explain assertiveness, which focuses on the suitability of the response itself for mutually satisfying the interests of all interacting parties (Rakus, 1991); self-efficacy in explaining passiveness, because perceptions of self-efficacy may be highly motivational of behavior (Bandura, 1982), and so the reverse may also be true; and expected outcomes in explaining aggression, given that it may have been learned as the best way to achieve satisfaction (Bandura, 1983).

Response decision measures were grouped in one-factor for all behavioral options under scrutiny. The response decision measure for assertiveness was differently constituted for boys and girls. The item *Calmly ask why they hadn't talked to me* only defined assertiveness for boys, whereas the item *Call that person and calmly tell him/her to be more careful in the future so it wouldn't happen again* only defined assertiveness for girls. Both items (as all the remaining items measuring assertiveness) concern the display of negative feelings. This type of assertive behavior includes actually expressing negative feelings and then asking for a behavioral change (Rakus, 1991); this second component of the behavior is only explicitly stated in the assertive option that differently defined assertiveness for girls and not for boys (*i.e., so it wouldn't happen again*). We may speculate that, because adolescent girls value social proximity while also being more anxious about behaving assertively, generally and while displaying negative feelings in particular, and actually doing it less (Bridges, Sanderman, Breukers, Ranchor, & Arrindell, 1991; Vagos, Pereira, & Arrindell, 2014), they care more about doing it in such a way as to clearly mention the stability of the relationship, whereas for boys, who are more relaxed about their assertiveness, such a reference is seen as unnecessary, and even the expression of some femininity. Surprisingly, the measurement invariance of assertiveness instruments has seldom been assessed, and so

this finding is distinctive and may serve as a trigger for future work aiming to define what differentiates women's and men's assertiveness.

The other response decision measures were similarly defined by gender. Passiveness such as it is evaluated in the SSIPA referred to inactivity in social events, rather than submissive, security or avoidance behaviors, which are commonly described as social behaviors (McManus, Sacadura, & Clark, 2008); overt aggression represented mainly verbal and direct aggression; relational aggression incorporated the social and relational aspects of this type of aggression, including behaviors intended to cause damage to others' general social reputation and behaviors aiming to exclude others' from significant relations (Archer & Coyne, 2005).

Male children (Werner & Hill, 2010) and early adolescents (Nelson & Crick, 1999) have been found to favor overt and relational aggression in comparison to girls. According to our findings, this preference seems continues into late adolescence, both for response evaluation and for response decision. In the same line, boys have also been found to behave more aggressively than girls, in a sample similar in culture and age to the one currently used (Vagos, Rijo, Santos, & Marsee, 2014), being it either overtly or relationally (Crick & Grotpeter, 1995). Girls on the other hand have been found to be more passive socially (Cunha, Pinto-Gouveia, & Soares, 2007), which is also in line with the present findings. The SIP evaluation of the assertive and passive types of response has not been previously addressed, but given the strong association which is theoretically and empirically documented between response evaluation and response decision (Crick & Dodge, 1996; Fontaine et al., 2010), one would expect that the more you practice it, the more you evaluate it favorably, and, therefore, girls not only choose passive responses more frequently but also favored them the most in comparison to boys. The same might be true for assertiveness. Though we cannot compare boys and girls based on our measures for choosing an assertive response, previous research

has found girls to act more assertively (Vagos, Pereira, et al., 2014) and so they would also be expected to favor this type of behavior.

Overall, the measures derived from the SSIPA are in line with research undertaken in connection with the SIP theory. In comparison with other measures developed for and used with adolescents (namely Calvete & Orue, 2009), the SSIPA presents several advantages, but also some limitations. One of these limitations concerns the issue that perhaps the scenes used in the SSIPA were too specific to Portuguese adolescence. The fact that they were originally taken from existing international instruments and only two out of six were adapted (Vagos et al., 2013), may be used to support their universality, but future research should determine this, namely by using focus groups to try to understand pertinent and common social events in non-Portuguese adolescents. Also, the fact that twice as many items represent aggression (*versus* assertiveness and passiveness), may consequently increase the aggressive “by chance” response option, and, therefore, should also be considered in future work. Finally, due to the structure of the SSIPA (*i.e.*, the items and response options it includes, the order in which they are presented, and the fact that it is a self-response questionnaire), one could argue that it may be evaluating a more reflexive and controlled SIP, rather than a more automatic and genuine SIP, which may make unique contributions to the prediction of individual differences, particularly in aggression (Fontaine, 2007).

The advantages of the SSIPA, on the other hand, are manifested in the fact that it addresses different types of attribution, making it more possible to distinguish among social behavior groups. Moreover, it evaluates the response evaluation and decision phases and it includes assertive and passive behavior options, making it more adequate for evaluating SIP in relation to social maladjustment and adjustment in adolescence. Additionally, these measures combine content validity and pertinence for the intended purpose and targeted population (guaranteed by the developmental process of this instrument; Vagos et al., 2013)

with other requisite properties, namely psychometric, thus having the potential to make a substantial contribution to the literature and to psychological assessment (Vogt et al., 2004). This research represents a promising and continued effort in the development and thorough qualitative and quantitative examination of an instrument for addressing attribution of intent, emotional intensity, response evaluation and response decision as steps of SIP. This effort and its ensuing results are encouraging, though still preliminary. Further research on this instrument is indispensable (and currently underway), namely to address its internal validity by structural equation modeling of the association between its measures according to a SIP framework, and to address its external validity in relation to convergent and divergent measures of different types of social behavior.

References

- Archer, J., & Coyne, S. M. (2005). An integrated review of indirect, relational, and social aggression. *Personality and Social Psychology Review*, 9(3), 212–230.
doi:10.1207/s15327957pspr0903_2
- American Psychological Association (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Bailey, C. A., & Ostrov, J. M. (2008). Differentiating forms and functions of aggression in emerging adults: Associations with hostile attribution biases and normative beliefs. *Journal of Youth and Adolescence*, 37(6), 713–722. doi:10.1007/s10964-007-9211-5
- Bridges, K. R., Sanderman, R., Breukers, P., Ranchor, A., & Arrindell, W. A. (1991). Sex differences in assertiveness on the U.S. version of the scale for interpersonal behavior (SIB). *Personality and Individual Differences*, 12(12), 1239–1243. doi:10.1016/0191-8869(91)90195-H
- Calvete, E., & Orue, I. (2009). Evaluación del procesamiento de la información social en adolescentes españoles y su asociación con la conducta agresiva. *Behavioral Psychology*, 17(3), 523–542.
- Calvete, E., & Orue, I. (2010). Cognitive schemas and aggressive behavior in adolescents: The mediating role of social information processing. *The Spanish Journal of Psychology*, 13(1), 190–201.
- Calvete, E., & Orue, I. (2012). Social information processing as a mediator between cognitive schemas and aggressive behavior in adolescents. *Journal of Abnormal Child Psychology*, 40(1), 105–117. doi:10.1007/s10802-011-9546-y
- Castro, B. O. De, & Merk, W. (2005). Emotions in social information processing and their relations with reactive and proactive aggression in referred aggressive boys. *Journal of Clinical Child and Adolescent Psychology*, 34(1), 105–116. Retrieved from http://www.tandfonline.com/doi/abs/10.1207/s15374424jccp3401_10

- Crick, N. R., & Dodge, K. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115*(1), 74–101. doi:10.1111/j.1467-8624.1996.tb01778.x
- Crick, N. R., & Dodge, K. A. (1996). Social information-processing mechanisms in reactive and proactive aggression. *Child Development*, *67*(3), 993–1002. doi:10.1111/j.1467-8624.1996.tb01778.x
- Crick, N. R., & Grotpeter, J. K. (1995). Relational aggression, gender and social-psychological adjustment. *Child Development*, *66*(3), 710–722. doi:10.1111/j.1467-8624.1995.tb00900.x
- Crick, N. R., Grotpeter, J. K., & Bigbee, M. A. (2002). Relationally and physically aggressive children's intent attributions and feelings of distress for relational and instrumental peer provocations. *Child Development*, *73*(4), 1134–1142. doi:10.1111/1467-8624.00462462
- Cunha, M., Pinto-Gouveia, J., & Soares, I. (2007). Natureza, frequência e consequências dos medos sociais na adolescência: Dados na população portuguesa. *Psychologica*, *44*, 207–236.
- De Castro, B. O. (2004). The development of social information processing and aggressive behaviour: Current issues. *European Journal of Developmental Psychology*, *1*(1), 87–102. doi:10.1080/17405620444000058
- De Castro, B. O., Veerman, J. W., Koops, W., Bosch, J. D., & Monshouwer, H. J. (2002). Hostile attribution of intent and aggressive behavior: A meta-analysis. *Child Development*, *73*(3), 916–934. doi:10.1111/1467-8624.00447
- Dimitrov, D. M. (2006). Comparing groups on latent variables: a structural equation modeling approach. *Work (Reading, Mass.)*, *26*(4), 429–36. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/16788262>

- Dimitrov, D. M. (2010). Testing for Factorial Invariance in the Context of Construct Validation. *Measurement and Evaluation in Counseling and Development*, 43(2), 121–149. doi:10.1177/0748175610373459
- Elison, J., Lennon, R., & Pulos, S. (2006). Investigating the compass of shame: the Development of the compass of shame scale. *Social Behavior and Personality*, 34, 221–238
- Elliott, C. H., & Lassen, M. K. (1997). A schema polarity model for case conceptualization, intervention, and research. *Clinical Psychology-Science and Practice*, 4(1), 12–28. doi:10.1111/j.1468-2850.1997.tb00095.x
- Erdley, C. A., Rivera, M. S., Shepherd, E. J., & Holleb, L. J. (2010). Social-cognitive models and skills. In D. W. Nagle, D. J. Hansen, C. A. Erdley, & P. J. Norton (Eds.), *Practitioner's guide to empirically based measures of social skills* (pp. 21–35). London: ABCT Clinical Assessment Series. doi:10.1007/798-1-4419-0609-0_2
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272–299. doi:10.1037//1082-989X.4.3.272
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods*, 9(4), 466–91. doi:10.1037/1082-989X.9.4.466
- Floyd, F. J., & Widaman, K. F. (1995). Factor analysis in the development and refinement of clinical assessment instruments. *Psychological Assessment*, 7(3), 286–299.
- Fontaine, R. G. (2007). Toward a conceptual framework of instrumental antisocial decision-making and behavior in youth. *Clinical Psychology Review*, 27(5), 655–675. doi:10.1016/j.cpr.2007.01.012

- Fontaine, R. G., & Dodge, K. A. (2006). Real-time decision making and aggressive behavior in youth: A heuristic model of response evaluation and decision (RED). *Aggressive Behavior*, 32(6), 604–624. doi:10.1002/ab.20150
- Fontaine, R. G., Tanha, M., Yang, C., Dodge, K. A., Bates, J. E., & Pettit, G. S. (2010). Does Response Evaluation and Decision (RED) Mediate the Relation between Hostile Attributional Style and Antisocial Behavior in Adolescence? *Journal of Abnormal Child Psychology*, 38(5), 615–626. doi:10.1007/s10802-010-9397-y
- Fontaine, R. G., Yang, C., Dodge, K., Bates, J., & Pettit, G. (2008). Testing an individual system model of response evaluation and decision (RED) and antisocial behavior across adolescence. *Child Development*, 79(2), 462–475.
- Ferketich, S. (1991). Focus on psychometrics. Aspects of item analysis. *Research in Nursing & Health*, 14(2), 165-168.
- Gadermann, A. M., Guhn, M., Zumbo, B. D., & Columbia, B. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data : A conceptual , empirical , and practical guide, 17(3).
- Glorfeld, L.W. (1995). An improvement on Horn’s parallel analysis methodology for selecting the correct number of factors to retain. *Educational and Psychological Measurement*, 55, 377-393.
- Godwin, J., & Maumary, A. (2004). Adolescent stories - Fast track project technical report. Retrieved from www.fasttrackproject.org
- Instituto do Emprego e Formação Profissional (1994). *Classificação nacional de profissões* [Nacional Classification of Professions]. Lisbon, Portugal: IEFP.
- Hair Jr., J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2005). *Multivariate data analysis* (7th ed.). Upper Saddle River: Prentice Hall.

Harper, B. D., Lemerise, E. a, & Caverly, S. L. (2010). The effect of induced mood on children's social information processing: goal clarification and response decision.

Journal of Abnormal Child Psychology, 38(5), 575–86. doi:10.1007/s10802-009-9356-7

Horowitz, M. J. (1991). *Person schemas and maladaptive interpersonal patterns*. London: The University of Chicago Press.

Huppert, J. D., Foa, E. B., Furr, J. N., Filip, J. C., & Mathews, A. (2003). Interpretation Bias in Social Anxiety: A Dimensional Perspective. *Cognitive Therapy and Research*, 27(5), 569–577. doi:10.1023/A:1026359105456

Kilne, P. (1993). Test construction: Factor analytic and item analytic methods. In P. Kline (Ed.), *The handbook of psychological testing* (pp. 134–153). London, UK: Routledge.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). NY: The Guilford Press.

Kubik, M. Y., Lytle, L. a, Birnbaum, A. S., Murray, D. M., & Perry, C. L. (2003). Prevalence and correlates of depressive symptoms in young adolescents. *American Journal of Health Behavior*, 27(5), 546–53. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/14521250>

Kupersmidt, J. B., Stelter, R., & Dodge, K. A. (2011). Development and validation of the Social Information Processing Application: A web-based measure of social information processing patterns in elementary school-age boys. *Psychological Assessment*, 23(4), 834–847. doi:10.1037/a0023621

Lazarus, R. S. (2006). Emotions and interpersonal relationships: toward a person-centered conceptualization of emotions and coping. *Journal of Personality*, 74(1), 9–46. doi:10.1111/j.1467-6494.2005.00368.x

- Lemerise, E. A., & Arsenio, W. F. (2000). An integrated model of emotion processes and cognition in social information processing. *Child Development, 71*(1), 107–118.
doi:10.1111/1467-8624.00124
- Marsee, M. A., & Frick, P. J. (2007). Exploring the cognitive and emotional correlates to proactive and reactive aggression in a sample of detained girls. *Journal of Abnormal Child Psychology, 35*(6), 969–981. doi:10.1007/s10802-007-9147-y
- McManus, F., Sacadura, C., & Clark, D. M. (2008). Why social anxiety persists: An experimental investigation of the role of safety behaviours as a maintaining factor. *Journal of Behavior Therapy and Experimental Psychiatry, 39*(2), 147–161.
doi:10.1016/j.jbtep.2006.12.002
- Murphy, K. R., & Davidshofer, C. O. (2001). Item analysis. In K. R. Murphy & C. O. Davidshofer (Eds.), *Psychological testing - Principles and applications* (pp. 192–214). New Jersey, USA: Prentice-Hall, Inc.
- Muthén, L.K., & Muthén, B. O. (2010). *Mplus user's guide* (6th ed.). Los Angeles, CA: Muthén & Muthén.
- Nelson, D. A., & Crick, N. R. (1999). Rose-colored glasses: Examining the social information-processing of prosocial young adolescents. *The Journal of Early Adolescence, 19*(1), 17–38. doi:10.1177/0272431699019001002
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). NY: McGraw-Hill.
- Prinstein, M. J., Boergers, J., & Vernberg, E. M. (2001). Overt and relational aggression in adolescents: Social-psychological adjustment of aggressors and victims. *Journal of Clinical Child Psychology, 30*(4), 479–491. doi:10.1207/s15374424jccp3004_05
- Rakus, R. (1991). *Assertive behavior - Theory, research and training*. Londres, Inglaterra: Routledge.

- Sumrall, S. G., Ray, G. E., & Tidwell, P. S. (2000). Evaluations of relational aggression as a function of relationship type and conflict setting. *Aggressive Behavior, 26*(2), 179–191. doi:10.1002/(sici)1098-2337(2000)26:2<179::aid-ab4>3.3.co;2-w
- Vagos, P., Pereira, A., & Arrindell, W. (2014). Validação da versão reduzida da escala de comportamento interpessoal para adolescentes Portugueses. *Psicologia: Reflexão E Crítica, 27*(3), 452–461. doi:10.1590/1678-7153.201427305
- Vagos, P., Rijo, D., & Santos, I. M. (2013). Assessing Social Information Processing in Adolescents: A phenomenological approach to item development. In *Abstract book of the the12th European Conference on Psychological Assessment* (p. 522). San Sebastian, Spain.
- Vagos, P., Rijo, D., Santos, I. M., & Marsee, M. a. (2014). Forms and Functions of Aggression in Adolescents: Validation of the Portuguese Version of the Peer Conflict Scale. *Journal of Psychopathology and Behavioral Assessment, 36*(4), 570–579. doi:10.1007/s10862-014-9421-6
- Vogt, D. S., King, D. W., & King, L. A. (2004). Focus groups in psychological assessment: Enhancing content validity by consulting members of the target population. *Psychological Assessment, 16*(3), 231–243.
- Werner, N. E., & Hill, L. G. (2010). Individual and peer group normative beliefs about relational aggression. *Child Development, 81*(3), 826–836. doi:10.1111/j.1467-8624.2010.01436.x.
- Werner, N. E., & Nixon, C. L. (2005). Normative Beliefs and Relational Aggression: An Investigation of the Cognitive Bases of Adolescent Aggressive Behavior. *Journal of Youth and Adolescence, 34*(3), 229–243. doi:10.1007/s10964-005-4306-3

Xie, H. L., Cairns, R. B., & Cairns, B. D. (2002). The development of social aggression and physical aggression: A narrative analysis of interpersonal conflicts. *Aggressive Behavior*, 28(5), 341–355. doi:10.1002/ab.80008

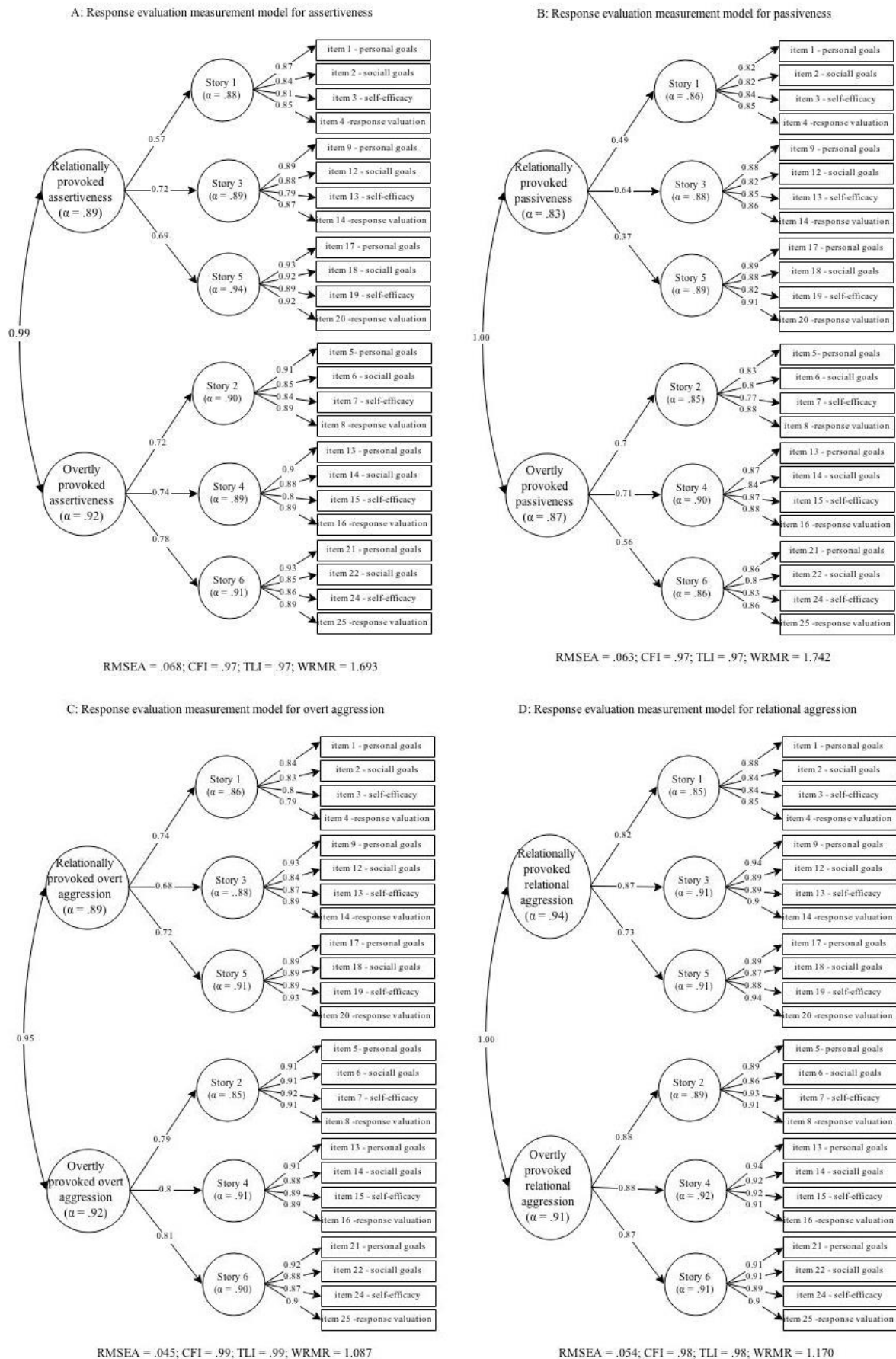


Figure 1: Second order measurement models for response evaluation measures

Table 1

Descriptive measures for the measures of the Scenes for Social Information Processing in Adolescence, for the complete sample and by gender

	Internal consistency	Complete sample			Male	Female
		M (SD)	Skewness ^b	Kurtosis ^c	M (SD)	M (SD)
Attribution measures						
4 Neutral	.72	12.97 (2.64)	-.037	.601	12.57 (2.81)	13.19 (2.53)
5 Hostile ^a	.73	11.84 (3.19)	.365	.650	11.74 (3.22)	11.90 (3.18)
Emotion measures						
5 Anger ^a	.80	10.52 (3.72)	.592	.109	10.43 (3.56)	10.57 (3.81)
4 Sadness	.79	9.42 (3.69)	.550	-.170	8.59 (3.44)	9.86 (3.75)
4 Shame ^a	.86	5.97 (2.67)	1.766	3.40	6.09 (2.73)	5.90 (2.64)
Response evaluation						
12 Relationally provoked assertiveness	.91	40.31 (8.55)	-.506	.625	39.15 (9.18)	40.93 (8.14)
12 Overtly provoked assertiveness ^a	.93	41.84 (8.97)	-.478	.429	40.37 (9.31)	42.62 (8.69)
12 Relationally provoked passiveness ^a	.86	60.64 (7.30)	-.248	.475	30.45 (8.29)	30.75 (6.71)
12 Overtly provoked passiveness	.89	33.11 (8.13)	-.272	.427	31.43 (8.65)	34.01 (7.68)
12 Relationally provoked overt aggression	.92	22.15 (7.46)	.773	.730	23.79 (8.24)	21.27 (6.86)
12 Overtly provoked overt aggression	.95	22.44 (8.65)	.891	.866	25.08 (9.64)	21.02 (7.71)
12 Relationally provoked relational aggression	.94	19.70 (7.34)	1.145	1.562	21.48 (8.50)	18.75 (6.44)
12 Overtly provoked relational aggression	.96	18.40 (7.53)	1.349	1.585	20.34 (8.65)	17.36 (6.63)
Response decision						
4 Assertiveness	-	-	-	-	12.81 (3.18)	12.67 (3.22)
4 Passiveness	.69	11.15 (3.26)	-.013	-.105	10.49 (3.37)	11.51 (3.15)
4 Relational aggression	.80	7.19 (2.94)	1.16	1.30	8.12 (3.38)	6.68 (2.54)
5 Overt aggression	.89	7.17 (2.92)	2.07	5.78	7.87 (3.39)	6.79(2.55)

Note: Numbers appearing before the name of the measure represent the number of items that constitute that measure. Factor score values underlying this descriptive analyses were computed by the sum of the numerical answers given by any subject to the items comprising each measure, according to the measurement models defined after EFA and/or CFA. Internal consistency values refer to the ordinal alpha values. All gender comparisons for all indicators were significant, except the ones for measures marked with ^a.

^bStandard error = .086; ^cStandard error = .172