The Development of a Short Form of the Sensitivity to Punishment and Sensitivity to Reward Questionnaire

Andrew Cooper[1] and Rapson Gomez[2]

[1]Department of Psychology, Goldsmiths, University of London, UK
[2]School of Psychology, University of Tasmania, Australia

Abstract. The Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ) has been proposed as a measure of the behavioral approach system (BAS) and behavioral inhibition system (BIS). Previous research with the SPSRQ has highlighted potential problems with the factor structure of the measure and individual item properties. The aim of the current studies was to use factor analytic and item response theory (IRT) methods to examine the psychometric properties of the SPSRQ. A further aim was to develop a short version of the SPSRQ. In Study 1, 393 adult participants completed the SPSRQ. The results from this study highlighted problems with the factor structure and item properties that had been noted in previous research. On this basis, a short form of the measure was proposed. In Study 2, the short form of the SPSRQ was tested with an independent sample (N = 327). These analyses suggested the short form of the SPSRQ had an improved factor structure, good item properties, and acceptable reliability.

Keywords: sensitivity to punishment, sensitivity to reward, factor analysis, item response theory, reinforcement sensitivity theory

Jeffrey Gray's reinforcement sensitivity theory (RST) of personality (Corr, 2004; Gray & McNaughton, 2000; Pickering et al., 1997) has proven to be an influential biologically based model of personality. In the original model of RST, Gray proposed a behavioral approach system (BAS), a behavioral inhibition system (BIS), and a flight-fight system (FFS). The BAS was presumed to be sensitive to conditioned signals of reward or nonpunishment and was initially identified with trait impulsivity, while the BIS was presumed to be sensitive to conditioned signals of punishment or frustrative nonreward and was initially identified with trait anxiety. A third system, the FFS, mediated responses to unconditioned aversive stimuli. These systems were presumed to underlie individual differences in learning, motivation, and emotion. It should be noted that there have been recent substantive changes to RST, particularly in relation to the BIS and FFS (now labeled the fight-flight-freeze system [FFFS]; for more detail see McNaughton & Corr, 2004). The effect these changes might have on self-report personality measures within RST remains an open question. Corr (2004) has suggested that combined FFFS/BIS functioning may broadly reflect punishment sensitivity. It may be feasible, however, to dissociate the FFFS and BIS at the personality level pending further research (for a more detailed discussion see Cooper, Perkins, & Corr, 2007; Smillie, Pickering, & Jackson, 2006).

As RST was initially developed largely on the basis of animal research, one of the key issues in the RST literature has been measurement with human data, in particular self-report questionnaire measurement (Corr, 2001). While a number of measures of trait anxiety and impulsivity have been used as proxy measures of the BIS and BAS, there have been several recent attempts to develop specific self-report measures for RST. Two such measures are the Behavioral Inhibition System/Behavioral Approach System Scales (BIS/BAS Scales; Carver & White, 1994) and the Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ; Torrubia, Avila, Moltó, & Caseras, 2001).

There is now reasonable support for the construct validity and hypothesized four-factor structure of the BIS/BAS Scales (Cogswell, Alloy, van Dulmen, & Fresco, 2006; Cooper, Gomez, & Aucote, 2007). However, a recent item response theory (IRT) analysis of the BIS/BAS Scales did highlight some deficiencies in the scales, particularly in the BAS-Reward Responsiveness scale (Gomez, Cooper, & Gomez, 2005). Existing data also suggest some problems with the psychometric properties of the SPSRQ, particularly in relation to its proposed factor structure (Cogswell et al.; O'Connor, Colder, & Hawk, 2004). Although it can be argued that this is not critical because the BIS/BAS Scales are presently available for RST research, we would disagree: The BIS/BAS Scales and the SPSRQ measure different aspects of the behavioral approach and inhibition systems. More specifically, the BIS/BAS Scales tap generalized sensitivities to rewards and punishment, while the SPSRQ taps sensitivity to punishments and rewards related to specific situations. The focus on specific situations may

Journal of Individual Differences 2008; Vol. 29(2):90–104
DOI 10.1027/1614-0001.29.2.90

be one reason for the difficulty in extracting unidimensional sensitivity to reward (SR) and sensitivity to punishment (SP) factors. Indeed, Beauducel, Kersting, and Liepmann (2005) used a multitrait-multimethod design to show that the situational context factors in the SPSRQ accounted for a reasonable amount of variance. Given this previous research, our aims in the current studies were to use IRT and factor analytic techniques to further analyze the SPSRQ. In addition, given that several previous studies have identified problematic items in the SPSRQ, on the basis of our analyses and that of previous studies, we propose a viable shorter form of the scales.

In brief, the SPSRQ comprises a SP scale and a SR scale (Torrubia et al., 2001). The SP scale comprises 24 items, rated either *yes* or *no*, assessing reactivity to punishment in aversive situations, as well as cognitive processes produced by cues for punishment and failure. An example of an item is "Are you often afraid of new or unexpected situations?" The SR scale also has 24 items, rated either *yes* or *no*. These items assess sensitivity to reward in a variety of situations. An example of an item is "Do you like to compete and do everything you can to win?" In the initial study that reported the development of the SPSRQ, Torrubia et al. (2001) conducted separate principal component analyses with varimax rotation for males and females. In both cases the two expected factors (SR and SP) were obtained. The SPSRQ demonstrated good internal and test-retest reliability. The scales also showed independence in that they correlated close to zero with each other, a finding that has been reported by others (Caseras, Avila, & Torrubia, 2003; O'Connor et al., 2004; Smillie & Jackson, 2005). This finding has commonly been used to support the validity of RST measures (but see Pickering, in press, for a detailed analysis of this issue using a formal modeling approach).

Studies have examined the convergent and discriminant validity of the SPSRQ. In terms of relating to other BIS and BAS scales, the SP scale has generally correlated positively and significantly with other BIS measures (Caseras et al., 2003; O'Connor et al., 2004; Smillie & Jackson, 2005; Torrubia et al., 2001). The SR scale has tended to correlate significantly and positively with other BAS measures and has generally had nonsignificant correlations with BIS measures (Caseras et al., 2003; O'Connor et al., 2004; Smillie & Jackson, 2005; Torrubia et al., 2001). In a factor analysis of a wide range of BIS and BAS measures, the SP scale clearly loaded on a factor with other BIS measures, and did not load on any of the BAS factors (Caseras et al., 2003). Similarly, the SR scale loaded strongly on a factor that Caseras et al. labeled Reward Interest. Taken together, these findings support the convergent and discriminant validity of the SP and SR scales.

A number of studies have examined how the SPSRQ relates to Eysenck's Neuroticism and Extraversion dimensions. Torrubia et al. (2001) found that the SP scale correlated significantly and negatively with Extraversion and significantly and positively with Neuroticism. The correlation with Neuroticism was stronger than that with Extraversion. This indicates reasonably good concurrent validity based on Gray's rotation of the Eysenck personality dimensions (Pickering, Corr, & Gray, 1999). The SR scale correlated significantly and positively with both Extraversion and Neuroticism. The correlation with Extraversion was stronger than that with Neuroticism. Thus, the SR scale also showed good concurrent validity (although see Smillie et al., 2006, for a more recent discussion of how Extraversion may relate to the BAS). Other studies have largely supported the good concurrent validity of the SPSRQ in relation to the Eysenck dimensions (Caseras et al., 2003).

Apart from the initial study reported by Torrubia et al. (2001), there have been two studies that have examined the factor structure of the SPSRQ (Cogswell et al., 2006; O'Connor et al., 2004). O'Connor et al. used confirmatory factor analysis (CFA) to test the two-factor structure of the scale and found poor model-data fit. A subsequent exploratory factor analysis (EFA) highlighted six items from the SP scale and seven items from the SR scale that either did not load substantially on the proposed factor, or loaded on both factors. These items are indicated in Table 2. The authors removed these items and tested the confirmatory fit of the revised model on three independent samples. The revised model did not, however, show acceptable fit for two of the samples and only marginally acceptable fit for the third sample. Cogswell et al. also examined the factorial structure of the SPSRQ. They also found poor confirmatory fit for the full model. They subsequently conducted an EFA on the data, and found eleven problematic items (see Table 2), of which nine items were those shown to be problematic by O'Connor et al. Cogswell et al. removed these nine items and used CFA to test this model on an independent sample. Despite this, the trimmed model did not show particularly good fit. The findings in these studies indicate that the current factorial structure of the SPSRQ is less than optimal. Indeed, a careful examination of the initial Torrubia et al. study does show relatively low factor loadings (< 0.30) for several SP and SR items. Overall, therefore, it would appear that support for the SPSRQ is mixed. While there appears to be reasonably good support for the convergent, discriminant, and concurrent validities of the scales in the SPSRQ, there are problems with its factor structure.

The current paper reports on two studies. The aim of Study 1 was to examine the psychometric properties of the SPSRQ using factor analysis and IRT to examine item and test properties of the scales. As far as we are aware, no published study has yet evaluated the SPSRQ using IRT. IRT can offer significant advantages over classical test theory methods in the analysis of self-report measures (Embretson & Reise, 2000). IRT analyses can provide important test and item level information that can greatly assist in highlighting poorly functioning items and, thus, can aid the development of a scale. On the basis of results from previous research and the current study, a further aim of Study 1 was to propose a viable short version of the SPSRQ. As noted in previous research (e.g., Cogswell et al., 2006), it is important to test potential revised versions

of the SPSRQ and other BIS/BAS-related measures in ongoing attempts to improve the psychometric properties of RST measures. One possible form of revision is to remove items that have been shown to have poor psychometric properties across multiple studies. Hence, a viable short form of the SPSRQ may contribute to this ongoing process. Having derived a shorter version of the SPSRQ in Study 1, the aim of Study 2 was to examine the psychometric properties of this version using CFA and IRT with an independent sample of participants. The convergent, discriminant- and concurrent validity of the short version was also tested by examining correlations with other widely used BIS/BAS measures and the Eysenck personality dimensions.

Study 1

Method

Participants

In all, 393 participants completed the SPSRQ, of which 148 were male and 245 were female. All participants were students at the University of Ballarat, a university in regional Victoria, Australia. The age of the participants ranged from 18 years to 51 years, with a mean of 23.62 years (*SD* = 5.99). The mean age for males was 23.63 (*SD* = 5.74) and the mean age for females was 23.61 (*SD* = 6.15).

Measure

The SPSRQ is a 48-item measure and comprises an SP scale and an SR scale. The SP and SR scales have 24 items each. The SPSRQ has a dichotomous yes/no response format. Individuals are given a score of 1 for a "yes" response to an item and a score of 0 for a "no" response to an item. Thus, scores can range from 0 to 24 for the full SP and SR scales. Both the SP and SR scales have been shown to have good reliability (Torrubia et al., 2001). In the current study, the SP scale had a Cronbach's α value of 0.84 and the SR scale had a Cronbach's α value of 0.82.

Procedure

Following approval from the University of Ballarat Human Research Ethics Committee, participants were recruited via notices placed around the University of Ballarat campus. Some participants participated as part of a credit requirement for a unit in first-year psychology. All testing took place in a classroom environment. Participants were tested in groups of 5 to 20 where they completed the SPSRQ. Participants were informed by the experimenter that they would complete a personality questionnaire, and to follow carefully the instructions written at the top of the questionnaire. Following completion of the SPSRQ, participants were debriefed and thanked for their time.

Data Analysis

All CFA analyses were conducted using the MPlus 4.2 software program (Muthen & Muthen, 2007). As the SPSRQ uses a categorical (dichotomous) response format, all CFA models were tested using mean and variance adjusted weighted least squares (WLSMV) estimation of the tetrachoric correlation matrix. Model fit was ascertained using the minimum fit function χ^2. With large sample sizes, the value of χ^2 can potentially lead to the model being rejected with only trivial differences between the data and model-implied covariances. Hence, fit was also examined using three practical fit indices. They were the root mean square error of approximation (RMSEA; Steiger, 1990), the comparative fit index (CFI; Bentler, 1990), and the standardized root mean square residual (SRMR). The RMSEA provides a measure of model fit relative to the population covariance matrix when the complexity of the model is also taken into account. RMSEA values of < .05 are taken as good fit, .05 to .08 as moderate fit, .08 to .10 as marginal fit, and > .10 as poor fit (Browne & Cudeck, 1993; Hu & Bentler, 1999). The CFI provides a measure of the fit of the hypothesized model relative to the independent model, with values usually ranging from 0.00 to 1.00. For the CFI, values between .90 and .95 indicate acceptable fit, and values above .95 indicate good fit. SRMR is the standardized difference between the observed covariance and predicted covariance. A value of zero indicates perfect fit, and values less than .08 indicate acceptable fit (Hu & Bentler, 1999).

EFA was conducted using MPlus 4.2 (Muthen & Muthen, 2007). Factors were extracted from the sample tetrachoric correlation matrix using WLSMV estimation with Varimax rotation. The number of factors extracted was based on a visual inspection of the scree plot and the minimum average partial (MAP; Velicer, 1976) test.

The IRT analyses were conducted using the BILOG-MG 3 software program (Zimowski, Muraki, Mislevy, & Bock, 2003). As the SPSRQ has a dichotomous response scale, an IRT model appropriate for these scales is the 2 Parameter Logistic Model (2PLM) (Birnbaum, 1968). For this model, logistic curves, called item characteristic curves (ICCs), are generated for each item, showing the probability of a positive response to the item as a function of the underlying trait. The mathematical properties of the curves provide the parameters for evaluation of the items. The two item parameters utilized in the 2PLM are the item threshold (or item difficulty) and the item discrimination. The threshold parameter (β) indicates the point on the scale of the latent trait where a person has a 0.5 probability of responding positively to the item, while the item discrimination parameter (α) is the ability of an item to discriminate people of different levels of the underlying trait below and above the threshold parameter (Steinberg & Thissen, 1995). Accord-

ing to Baker (2001), α values of 0.01 to 0.24 are very low, 0.25 to 0.64 are low, 0.65 to 1.34 are moderate, 1.35 to 1.69 are high, and more than 1.7 are very high. Very low and low α values are generally undesirable, as it indicates that the item does not discriminate well between individuals differing in latent trait values.

IRT models also provide information functions for each item and for all items together. These are called the item information function (IIF) and the test information function (TIF), respectively. The IIF indicates the effectiveness of an item to measure the trait at different levels of the trait, while the TIF provides the effectiveness of the test to measure the trait at different levels of the trait. IRT also provides the standard error of measurement (*SEM*) of the TIF. As the SEM of a TIF is the inverse of the square root of the TIF, the *SEM* values can be viewed as indicators of the precision of the test at different trait levels (Embretson & Reise, 2000). Since the *SEM* is related to reliability, the TIF can be viewed in terms of reliability. Prior to using IRT, the scales were assessed for unidimensionality and model-data fit. Unidimensionality was assessed by examining the fit of one-factor CFA models for the SP and SR scales separately. BILOG-MG 3 (Zimowski et al., 2003) was used to test model-data fit. Model fit can be tested by examining the difference between the –2 log likelihood values for nested models. The difference in –2 log likelihood values can be tested for significance using a χ^2 distribution, with degrees of freedom equal to the number of parameters constrained in the nested model. As the 1 Parameter Logistic Model (1PLM) is nested within the 2PLM, we compared the fit of the 2PLM to the 1PLM. A significant χ^2 value implies the 2PLM provides better model fit. At the item level, χ^2 goodness-of-fit statistics are based on the observed and expected response frequencies for each item. A nonsignificant χ^2 value indicates good model-data fit for that particular item.

Table 1. Means, standard deviations, skewness, and kurtosis for the SP and SR scales.

	SP	SR
Mean	11.56	10.64
SD	5.36	4.83
Skewness	0.08	0.09
Kurtosis	–0.88	–0.76

Results

Descriptive Analysis

Table 1 shows basic descriptive statistics for the full SP and SR scales. The observed total scores found for both the SP and SR scales in the current study are similar to the normative scores provided in Torrubia et al. (2001).

Confirmatory Factor Analysis

Initially, the fit of the two-factor full SPSRQ model was examined. In this model, all of the SP items loaded on a single latent SP factor, and all of the SR items loaded on a single latent SR factor. The SP and SR factors were free to covary and there were no correlations between the residuals. As expected, model fit for the full measurement model was not acceptable, χ^2 (208, N = 393) = 535.48, $p < .001$; RMSEA = 0.063, CFI = 0.81; SRMR = 0.114. The latent SP and SR factors had a moderately low nonsignificant correlation (r = 0.17). Factor loadings for the SP items ranged from .15 to .84, with five SP items having factor loadings less than .35. Factor loadings for the SR items ranged from .22 to .73, with three SR items having factor loadings less than .35.

Table 2. Standardized factor loadings for a two-factor solution using varimax rotation

Item	Factor 1 (SP)	Factor 2 (SR)
1. Do you often refrain from doing something because you are afraid of it being illegal?[a,b]	0.15	0.01
2. Does the good prospect of obtaining money motivate you strongly to do some things?	–0.03	0.53
3. Do you prefer not to ask for something when you are not sure you will obtain it?	0.52	0.08
4. Are you frequently encouraged to act by the possibility of being valued in your work, in your studies, with your friends, or with your family?[a,b]	0.08	0.38
5. Are you often afraid of new or unexpected situations?	0.78	–0.07
6. Do you often meet people that you find physically attractive?[a,b]	–0.06	0.50
7. Is it difficult for you to telephone someone you do not know?	0.54	0.05
8. Do you like to take some drugs because of the pleasure you get from them?[a,b]	0.03	0.31
9. Do you often renounce your rights when you know you can avoid a quarrel with a person or an organization?[a]	0.44	0.08
10. Do you often do things to be praised?	0.23	0.59
11. As a child, were you troubled by punishments at home or in school?[a,b]	0.14	0.18
12. Do you like being the center of attention at a party or a social meeting?	–0.24	0.62
13. In tasks that you are not prepared for, do you attach great importance to the possibility of failure?	0.59	0.22

Table 2 (continued).

Item	Factor 1 (SP)	Factor 2 (SR)
14. Do you spend a lot of your time on obtaining a good image?	0.19	0.52
15. Are you easily discouraged in difficult situations?	0.67	0.09
16. Do you need people to show their affection for you all the time?[a,b]	0.37	0.38
17. Are you a shy person?[b]	0.65	–0.31
18. When you are in a group, do you try to make your opinions the most intelligent or the funniest?	0.05	0.64
19. Whenever possible, do you avoid demonstrating your skills for fear of being embarrassed?	0.77	0.04
20. Do you often take the opportunity to pick up people you find attractive?	–0.07	0.59
21. When you are with a group, do you have difficulties selecting a good topic to talk about?	0.57	–0.07
22. As a child, did you do a lot of things to get people's approval?	0.29	0.43
23. Is it often difficult for you to fall asleep when you think about things you have done or must do?[a]	0.43	0.10
24. Does the possibility of social advancement move you to action, even if this involves not playing fair?	0.08	0.50
25. Do you think a lot before complaining in a restaurant if your meal is not well prepared?	0.30	0.03
26. Do you generally give preference to those activities that imply an immediate gain?	0.11	0.58
27. Would you be bothered if you had to return to a store when you noticed you were given the wrong change?[a]	0.33	0.04
28. Do you often have trouble resisting the temptation of doing forbidden things?	0.02	0.55
29. Whenever you can, do you avoid going to unknown places?[b]	0.62	–0.05
30. Do you like to compete and do everything you can to win?	–0.11	0.62
31. Are you often worried by things that you said or did?	0.58	0.10
32. Is it easy for you to associate tastes and smells to very pleasant events?[a,b]	0.00	0.24
33. Would it be difficult for you to ask your boss for a raise?	0.46	0.01
34. Are there a large number of objects and sensations that remind you of pleasant events?[a,b]	0.05	0.31
35. Do you generally try to avoid speaking in public?	0.51	–0.30
36. When you start to play with a slot machine, is it often difficult for you to stop?[a,b]	0.16	0.37
37. Do you, on a regular basis, think that you could do more things if it was not for your insecurity or fear?	0.84	0.03
38. Do you sometimes do things for quick gains?	0.01	0.71
39. Comparing yourself to people you know, are you afraid of many things?	0.74	0.08
40. Does your attention easily stray from your work in the presence of an attractive stranger?	0.14	0.53
41. Do you often find yourself worrying about things to the extent that performance in intellectual abilities is impaired?	0.64	0.25
42. Are you interested in money to the point of being able to do risky jobs?	0.05	0.58
43. Do you often refrain from doing something you like in order not to be rejected or disapproved of by others?	0.65	0.30
44. Do you like to put competitive ingredients in all of your activities?	–0.16	0.59
45. Generally, do you pay more attention to threats than to pleasant events?[a]	0.25	0.39
46. Would you like to be a socially powerful person?	0.10	0.72
47. Do you often refrain from doing something because of your fear of being embarrassed?	0.83	0.02
48. Do you like displaying your physical abilities even though this may involve danger?	–0.08	0.59

MAP Test

Dimensions	Minimum average partial correlation
1	0.029
2	0.015
3	0.017
4	0.020
5	0.027
6	0.041
7	0.079
8	1.000

Note. Items are numbered according to their original position in the SPSRQ and will be referred to as such in the text. The SP scale comprises

Table 3. IRT parameter estimates and standard errors for the SP and SR scales

	Item parameter estimates	
SP Scale	α	β
3. ask for something*	0.60 (.08)	–0.45 (.12)
5. new situations	1.22 (.15)	0.14 (.07)
7. telephone	0.65 (.09)	0.31 (.11)
9. quarrel*	0.51 (.08)	0.74 (.15)
13. failure	0.73 (.10)	–0.35 (.10)
15. discouraged	0.92 (.11)	0.27 (.08)
17. shy	0.81 (.10)	0.14 (.09)
19. embarrassed	1.24 (.16)	0.25 (.07)
21. topic	0.70 (.10)	0.77 (.13)
23. fall asleep*	0.49 (.08)	–1.50 (.24)
25. complaining*	0.36 (.07)	–1.50 (.31)
27. return to store*	0.37 (.06)	0.22 (.17)
29. unknown places	0.78 (.11)	0.72 (.11)
31. worried	0.70 (.09)	–0.65 (.12)
33. boss/raise*	0.57 (.09)	–1.47 (.22)
35. speaking in public*	0.60 (.08)	–0.39 (.12)
37. insecurity	1.61 (.23)	–0.09 (.06)
39. afraid	1.09 (.14)	0.66 (.08)
41. intellectual abilities	0.85 (.11)	0.60 (.10)
43. rejected	0.87 (.11)	–0.03 (.08)
45. threats*	0.35 (.07)	2.32 (.46)
47. fear of embarrassment	1.48 (.21)	–0.23 (.06)
SR Scale	α	β
2. money	0.64 (.09)	–0.95 (.15)
4. valued*	0.44 (.07)	–1.87 (.31)
6. attractive*	0.56 (.08)	–0.82 (.16)
10. praised	0.72 (.10)	0.26 (.10)
12. center of attention	0.73 (.10)	0.82 (.12)
14. good image*	0.57 (.09)	–0.09 (.12)
18. opinions	0.83 (.12)	0.56 (.10)
20. pick up people	0.74 (.11)	1.09 (.14)
22. approval*	0.46 (.07)	0.54 (.15)
24. social advancement*	0.58 (.09)	0.21 (.11)
26. immediate gain	0.75 (.10)	–0.08 (.09)
28. temptation*	0.62 (.09)	1.09 (.15)
30. compete/win	0.76 (.10)	0.55 (.10)
36. slot machine*	0.52 (.10)	2.65 (.45)
38. quick gains	0.96 (.13)	0.00 (.08)
40. attractive stranger*	0.62 (.09)	0.17 (.11)
42. risky jobs*	0.84 (.14)	1.58 (.19)
44. competitive*	0.61 (.10)	0.81 (.13)
46. socially powerful	1.03 (.13)	–0.22 (.08)
48. physical abilities	0.78 (.12)	1.07 (.14)

Note. The items with an asterisk (*) are those removed on the basis of the IRT analyses. The remaining items form the SPSRQ-S. Standard errors for each parameter estimate are in brackets. α = discrimination parameter; β = threshold parameter.

Given that the studies by both Cogswell et al. (2006) and O'Connor et al. (2004) jointly found that 9 of the original 48 SPSRQ items were problematic, a further CFA was then conducted on the SPSRQ, with these nine items removed. For this analysis there were 22 SP items loading on the SP factor, and 17 SR items loading on the SR factor. The SP and SR factors were free to covary and there were no correlations between the residuals. Relative to the two-factor model involving all of the SPSRQ items, this model showed better fit, χ^2 (175, N = 393) = 466.08, $p < .001$; RMSEA = 0.065, CFI = 0.84; SRMR = 0.110, however, these fit indices do not indicate acceptable model fit. Overall, the initial CFA models that were examined showed unacceptable global model fit and several items from both scales showed relatively low factor loadings. Similarly to previous research (Cogswell et al., 2006; O'Connor et al., 2004), we conducted an EFA to further explore the factor structure of the SPSRQ.

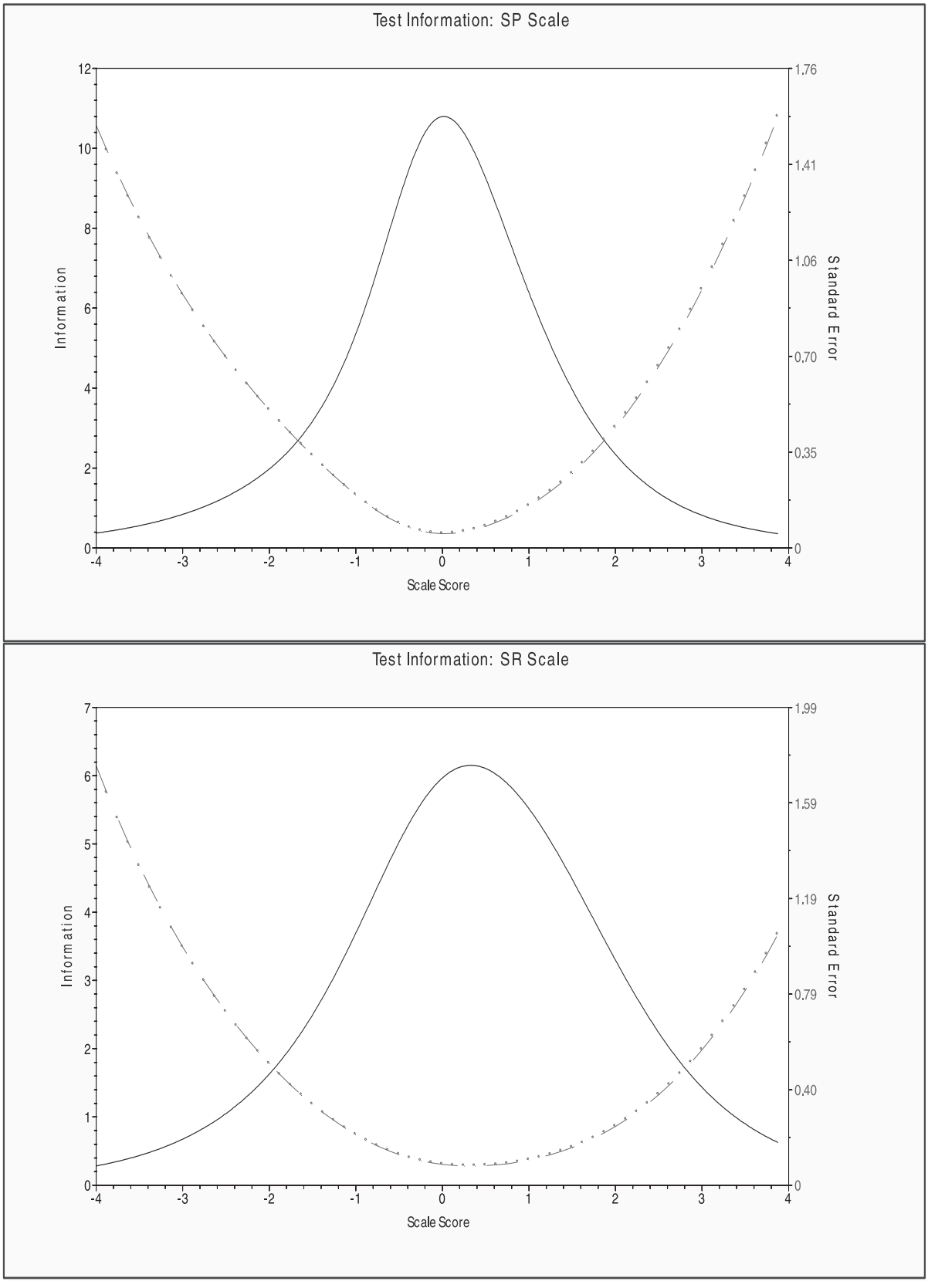

Figure 1. Test information and standard error of measurement curves for the SP and SR scales.

Exploratory Factor Analysis

The results of the MAP test and a visual inspection of the scree plot both suggested two factors should be extracted (the results of the MAP test are shown at the bottom of Table 2). The first five eigenvalues were 9.50, 7.16, 2.41, 2.27, and 1.82. The first two factors accounted for 17.6% and 15.3% of the variance, respectively. Table 2 shows the Varimax-rotated factor-loading matrix for all SPSRQ items. For comparison, Table 2 also indicates the items designated as problematic by Cogswell et al. (2006) and O'Connor et al. (2004). We decided to remove items for the subsequent IRT analysis that were deemed problematic across both previous studies, as well as our current EFA. In our study, we deemed an item factor-loading in the EFA problematic if it was below 0.32 (Tabachnick & Fidell, 2007), or if there was substantial cross-loading across both factors. On this initial basis, we removed Items 1, 8, 11, 16, 32, and 34 (see Table 2 for item content).

IRT Analysis

To test for the unidimensionality of the two scales, one-factor CFA models were tested separately for the trimmed SP and SR scales in MPlus 4.2, using WLSMV estimation of the tetrachoric correlation matrix. The model for the SP scale showed very good global fit, χ^2 (114, N = 393) = 173.70, $p < .0003$; RMSEA = 0.037, CFI = 0.97; SRMR = 0.071. The SP scale would appear to be sufficiently unidimensional. For the SR scale, the goodness-of-fit indexes were mixed, χ^2 (95, N = 393) = 244.67, $p < .0001$; RMSEA = 0.063, CFI = 0.88; SRMR = 0.093. The RMSEA and SRMR indicated acceptable fit, while the CFI did not. On the whole, we would argue the SR scale is sufficiently unidimensional to proceed with IRT analyses. Indeed, Reise, Smith, and Furr (2001) have suggested that strict forms of unidimensionality may be unrealistic for "high bandwidth" personality traits, as opposed to narrower cognitive constructs (see also Bolt, Hare, Vitale, & Newman, 2004).

The computer program BILOG-MG 3 (Zimowski et al., 2003) was used to test model-data fit. Initially, we tested whether the 2PLM fit the data better than the nested 1PLM for both the SP and SR scales. These tests showed that the 2PLM fit the data better than the 1PLM for the SP scale, χ^2 (22) = 209.92, $p < .0001$, and the SR scale, χ^2 (20) = 46.14, $p < .001$. In terms of fit at the item level, a nonsignificant χ^2 value for items indicates good model-data fit. All of the SR items had nonsignificant χ^2 values, and only Item 23[1] from the SP scale had a significant χ^2 value at $p < .05$. Inspection of the item-fit plots indicated good model-data fit. Overall, it appeared appropriate to fit the data with the 2PLM model, although it should be noted that these χ^2 values may be sensitive to sample size and capitalize on sample-specific characteristics.

Table 3 shows the discrimination and the threshold parameter values for the SP and SR scales. Tables 4 and 5 show the item and test information values and the reliability at different levels of their latent traits for the SP and SR scales, respectively. Figure 1 shows the TIF curves for the SP and SR scales.

The SP Scale

Table 3 shows that a number of the SP items had discrimination parameters in the moderate range (from 0.65 to 1.34). There were a number of items that had low discrimination parameters, however, namely, Items 3, 9, 23, 25, 27, 33, 35, and 45. Table 3 further shows that the threshold parameters for most items were relatively close to the mean value of the latent trait. Item 45 had a very large threshold parameter, indicating it was a difficult item to endorse. Items 23, 25, and 33 had very low threshold parameters, indicating they were easy items to endorse.

Table 4 shows that many of the SP items had acceptable item information values across the full range of the latent trait. There were a number of items that had low IIF values across most of the latent trait range: these included Items 3, 9, 23, 25, 27, 33, and 45. Table 4 and Figure 1 show the TIF values and *SEM* values for the SP scale. These indicate the SP scale has very good measurement precision for those within 1 *SD* unit above and below the mean, with a relatively peaked distribution between these values.

The SR Scale

Table 3 shows that many of the SR items had discrimination parameter values in the moderate range. There were a number of items with low discrimination parameters, however, namely, Items 4, 6, 14, 22, 24, 28, 36, 40, and 44. It further shows that while many of the items had threshold parameter values relatively close to the mean of the latent trait, Items 4 and 6 had relatively low threshold parameters, and Items 20, 28, 36, 42, and 48 had relatively high threshold parameters.

Table 5 shows that while some SR items had acceptable item information values across the full range of the latent trait, Items 4, 6, 14, 20, 22, 24, 28, 36, 42, 44, and 48 generally had low to very low IIF values across the range of the latent trait. When compared to the SP scale, some of the SR items tended to have information curves peaking at +1 *SD* units above the mean, rather than at the mean of the latent trait. This indicates these items had higher measurement precision for those at moderately high to very high on the latent trait. Table 5 and Figure 1 show the TIF values

[1] Item numbers referred to in the text in both Study 1 and 2 can be identified by referring to Table 2. The numbering of items in Table 2 is consistent with the numbering provided in Torrubia et al. (2001).

Table 4. Item and test information functions for the SP scale at various levels of the trait

	Estimated trait						
Items	–3.0	–2.0	–1.0	0	1.0	2.0	3.0
3. ask for something	0.01	0.15	0.24	0.25	0.16	0.07	0.03
5. new situations	0.01	0.04	0.33	1.01	0.53	0.08	0.01
7. telephone	0.03	0.08	0.19	0.29	0.26	0.14	0.06
9. quarrel	0.03	0.05	0.11	0.16	0.18	0.14	0.08
13. failure	0.05	0.16	0.33	0.37	0.20	0.07	0.02
15. discouraged	0.01	0.07	0.26	0.56	0.45	0.14	0.03
17. shy	0.02	0.09	0.27	0.47	0.34	0.13	0.03
19. embarrassed	0.01	0.04	0.28	0.98	0.63	0.11	0.01
21. topic	0.02	0.05	0.14	0.27	0.34	0.21	0.08
23. fall asleep	0.12	0.17	0.17	0.13	0.07	0.03	0.02
25. complaining	0.08	0.09	0.09	0.08	0.05	0.04	0.02
27. return to store	0.04	0.06	0.09	0.10	0.10	0.07	0.05
29. unknown places	0.01	0.05	0.15	0.33	0.43	0.23	0.08
31. worried	0.08	0.20	0.34	0.32	0.15	0.06	0.02
33. boss/raise	0.14	0.22	0.22	0.16	0.07	0.03	0.01
35. speaking in public	0.06	0.14	0.24	0.26	0.16	0.08	0.03
37. insecurity	0.01	0.01	0.53	1.88	0.34	0.02	0.01
39. afraid	0.01	0.02	0.14	0.54	0.78	0.25	0.04
41. intellectual abilities	0.01	0.05	0.17	0.41	0.48	0.22	0.06
43. rejected	0.03	0.11	0.34	0.54	0.32	0.10	0.02
45. threats	0.01	0.02	0.04	0.06	0.08	0.09	0.09
47. fear of embarrassment	0.01	0.07	0.69	1.54	0.26	0.02	0.01
TIF	0.84	1.97	5.35	10.79	6.39	2.34	0.82
SEM	6.34	3.46	1.33	0.36	1.05	3.01	6.46
R	0	0.49	0.81	0.91	0.84	0.57	0

Note. TIF = test information function; *SEM* = standard error of measurement; R = reliability.

Table 5. Item and test information functions for the SR scale at various levels of the trait

	Estimated trait						
Items	–3.0	–2.0	–1.0	0	1.0	2.0	3.0
2. money	0.10	0.21	0.29	0.24	0.11	0.04	0.02
4. valued	0.12	0.14	0.13	0.09	0.05	0.03	0.01
6. attractive	0.09	0.17	0.22	0.20	0.12	0.05	0.02
10. praised	0.03	0.08	0.22	0.36	0.31	0.14	0.05
12. center of attention	0.01	0.04	0.13	0.29	0.38	0.24	0.09
14. good image	0.05	0.11	0.19	0.24	0.18	0.10	0.04
18. opinions	0.01	0.05	0.18	0.40	0.45	0.21	0.06
20. pick up people	0.01	0.03	0.10	0.24	0.39	0.29	0.12
22. approval	0.03	0.06	0.11	0.14	0.15	0.11	0.07
24. social advancement	0.04	0.09	0.17	0.24	0.21	0.12	0.05
26. immediate gain	0.03	0.12	0.29	0.41	0.26	0.10	0.03
28. temptation	0.01	0.04	0.10	0.20	0.30	0.23	0.12
30. compete/win	0.02	0.06	0.18	0.36	0.39	0.19	0.06
36. slot machine	0.01	0.01	0.03	0.06	0.12	0.18	0.19
38. quick gains	0.02	0.09	0.37	0.68	0.37	0.09	0.02
40. attractive stranger	0.04	0.09	0.20	0.29	0.24	0.12	0.05
42. risky jobs	0.01	0.01	0.05	0.16	0.43	0.47	0.21
44. competitive	0.02	0.05	0.13	0.25	0.32	0.21	0.09
46. socially powerful	0.02	0.12	0.49	0.76	0.29	0.06	0.01
48. physical abilities	0.01	0.02	0.10	0.25	0.44	0.31	0.12
TIF	0.67	1.62	3.69	5.96	5.52	3.30	1.43
SEM	3.49	1.78	0.74	0.30	0.38	0.86	1.98
R	0	0.38	0.73	0.83	0.82	0.70	0.30

Note. TIF = test information function; *SEM* = standard error of measurement; *R* = reliability.

and curve for the SR scale. These indicate the SR scale has very good measurement precision for those just below the mean of the latent trait to those +1.5 *SD* units above the latent trait mean. The SR TIF curve peaks at a higher point on the latent trait than the SP TIF curve, and is optimally informative for those around +0.5 *SD* units above the latent trait mean.

Scale Revision

The IRT analyses above suggest that several items in both scales perform relatively poorly. We would suggest that it is beneficial to have a revised version of the SPSRQ that removes these poorly performing items as part of an ongoing revision process. On this basis, we removed items that had low item information values, hence, low measurement precision (Fraley, Waller, & Brennan, 2000), as well as items that had very large positive or negative threshold values (notably in the SR scale), and low discrimination parameters. Table 3 indicates the items removed on the basis of the IRT analyses. The remaining items in Table 3 form a potential short form of the SPSRQ (SPSRQ-S). There were 14 items remaining in the SP scale and 10 items remaining in the SR scale. It should be noted that only one of the items included in the SPSRQ-S was shown to be problematic in one of the previous studies (Item 29 in Cogswell et al., 2006). The full and short SP scales were significantly correlated ($r = 0.96$, $p < .0001$), as were the full and short SR scales ($r = 0.91$, $p < .0001$). As noted above, the full SP and SR scales had Cronbach's α reliabilities of 0.84 and 0.82, respectively. In comparison, the short forms of the SP and SR scales had Cronbach's α reliabilities of 0.85 and 0.75, respectively.

Discussion

The aim of Study 1 was to use factor analysis and IRT to examine the psychometric properties of the full SPSRQ. Following from this, a further aim was to develop a short form of the measure, the SPSRQ-S. Previous research had highlighted some significant problems with the factorial structure of the full measure and item properties for several items (Cogswell et al., 2006; O'Connor et al., 2004). Similarly, in the current study a CFA analysis of the full SPSRQ measurement model showed poor global model fit. A subsequent EFA analysis highlighted a number of items from the SR and SP scales with poor factor loadings on their hypothesized factor. These items were removed from the scale, and an IRT analysis was performed on the remaining items, with a view to identifying items with deficient item properties. Having identified deficient items, these items were removed to create a short version of the scale, the SPSRQ-S. The items removed via the EFA and IRT have largely been shown to be problematic in the previous research highlighted above. It is important that revised versions of measures be tested thoroughly in independent samples as part of an ongoing process. We begin this process by testing the SPSRQ-S in Study 2 below.

Study 2

Method

Participants

The participants were 327 individuals, of which 112 were male and 215 were female. All participants were students at the University of Ballarat. The age of the participants ranged from 18 years to 37 years, with a mean of 21.86 years (*SD* = 4.67). The mean age for males was 22.62 (*SD* = 5.21) and the mean age for females was 21.46 (*SD* = 4.32).

Measures

All participants completed the following measures:

- *The Sensitivity to Punishment and Sensitivity to Reward Scale (SPSRQ)*
  The SPSRQ (Torrubia et al., 2001) has been described extensively in the previous study. All results reported below are for the short form of the scale (SPSRQ-S).
- *The Eysenck Personality Questionnaire – Revised (EPQ-R) (Short version)*
  The EPQ-R (Eysenck & Eysenck, 1991) is a 48-item, dichotomous response questionnaire that provides scores for Extraversion, Neuroticism, Psychoticism, and Lie. The EPQ-R has been used extensively in past research and has good reliability and validity.
- *Spielberger Trait Anxiety Inventory (STAI)*
  The STAI (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) is a 20-item self-report measure of trait anxiety. Items were rated on a 4-point Likert-type response format, with a response of 1 indicating *almost never* and 4 indicating *almost always*. Items are summed to form a score for trait anxiety. This scale has been used widely as a proxy measure of the BIS in RST research. It has very good reported reliability and validity (Spielberger et al., 1983).
- *The Carver and White BIS/BAS Scales*
  The BIS/BAS Scales (Carver & White, 1994) are a 20-item measure comprised of a BIS Scale (seven items) and three BAS Scales: reward responsiveness (five items), drive (four items), and fun seeking (four items). The BIS/BAS Scales have a Likert-type response format, with a response of 1 indicating *very false for me* and a response of 4 indicating *very true for me*. The BIS/BAS Scales have been widely used in RST research, and have been shown to have acceptable validity and reliability (Cogswell et al., 2006; Caseras et al., 2003).

Table 6. Means, standard deviations, skewness, kurtosis, and Cronbach's α for all measures

	SP	SR	TA	E	N	P	BIS	RR	D	FS
Mean	6.15	5.22	41.00	8.56	6.00	2.82	19.91	16.94	10.18	11.76
SD	3.93	2.50	9.42	3.27	3.12	1.87	3.05	2.39	2.42	2.20
Skewness	0.24	–0.07	0.47	–0.76	–0.03	0.48	–0.21	–0.62	0.38	0.08
Kurtosis	–1.00	–0.81	–0.39	–0.36	–0.92	–0.48	0.15	0.24	0.11	–0.37
Cronbach's α	0.85	0.70	0.91	0.83	0.78	0.63	0.73	0.82	0.86	0.77

Note. SP = Sensitivity to Punishment, SR = Sensitivity to Reward, TA = Trait Anxiety, E = Extraversion, N = Neuroticism, P = Psychoticism, BIS = Behavioural Inhibition System, RR = Reward Responsiveness, D = Drive, FS = Fun Seeking.

Procedure

Following approval from the University of Ballarat Human Research Ethics Committee, participants were recruited via notices placed around the University of Ballarat campus. Some participants participated as part of a credit requirement for a unit in first-year psychology. All testing took place in a classroom environment. Participants were tested in groups of 10 to 20. Participants were given a booklet containing the personality questionnaires. The ordering of all personality questionnaires in the booklet was randomized. Participants were informed by the experimenter that they would complete a number of personality questionnaires, and to follow carefully the instructions written at the top of each questionnaire. Following completion of the questionnaires, participants were debriefed and thanked for their time.

Data Analysis

The CFA and IRT analyses conducted in Study 2 used the same analytic procedures described above in Study 1.

Results

Descriptive Analysis

Table 6 shows the means, standard deviations, skewness, kurtosis, and Cronbach's α values for each of the measures used in the study. All variables appeared to be normally distributed. The reported Cronbach's α value for the short SP scale was very similar to that for the full scale, while the short SR scale had a smaller Cronbach's α value than is generally reported for the full scale, although it was still acceptable.

Confirmatory Factor Analysis

The SPSRQ-S was tested for model fit. In this model, all of the SP items loaded on a single latent SP factor, and all of the SR items loaded on a single latent SR factor. The SP and SR factors were free to covary and there were no correlations between the residuals. Model fit for the SPSRQ-S showed unacceptable model fit, χ^2 (95, N = 327) = 231.31, $p < .0001$; RMSEA = 0.066, CFI = 0.88; SRMR = 0.107. The latent SP and SR factors had a low, nonsignificant, negative correlation (r = –0.07). All item factor-loadings for the SR and SP scales were above 0.40. Examination of the modification indices indicated that model fit would have been improved by freeing the path between item 10 and the SP latent factor.

IRT Analysis[2]

One-factor CFA models were tested for the short SP and SR scales. Model fit for the one-factor SP model was good, χ^2 (54, N = 327) = 75.94, $p > .01$; RMSEA = 0.035, CFI = 0.98; SRMR = 0.063. Model fit for the one-factor SR model was acceptable, χ^2 (27, N = 327) = 69.99, $p < .0001$; RMSEA = 0.070, CFI = 0.90; SRMR = 0.091. Both short scales appeared to be sufficiently unidimensional for IRT analyses.

IRT analyses were conducted for both scales. The 2PLM model was used to fit the data using the BILOG-MG software program. The χ^2 goodness-of-fit values produced by BILOG-MG for the SP and SR scales indicated good model-data fit. The SP items generally showed good item discrimination and threshold parameter values and relatively large item information values across the latent trait range. Item 29, however, had a relatively low discrimination parameter (0.53), and had relatively low item information values across the latent trait range. Figure 2 shows the TIF and *SEM* curves for the short SP scale. These indicate the TIF curve is not as sharply peaked as for the full scale, but covers a broader range of the latent trait more effectively.

IRT analyses of the SR items showed that they had generally good item discrimination and threshold parameters, and moderate item information values across the latent trait range. An exception was Item 2, which had a low discrimination parameter. Figure 2 shows the TIF and *SEM* curves for the short SR scale. These indicate the TIF curve peaks

[2] For reasons of space, only a brief description of the IRT analyses will be presented here. Details of the full IRT analyses can be obtained from the first author.

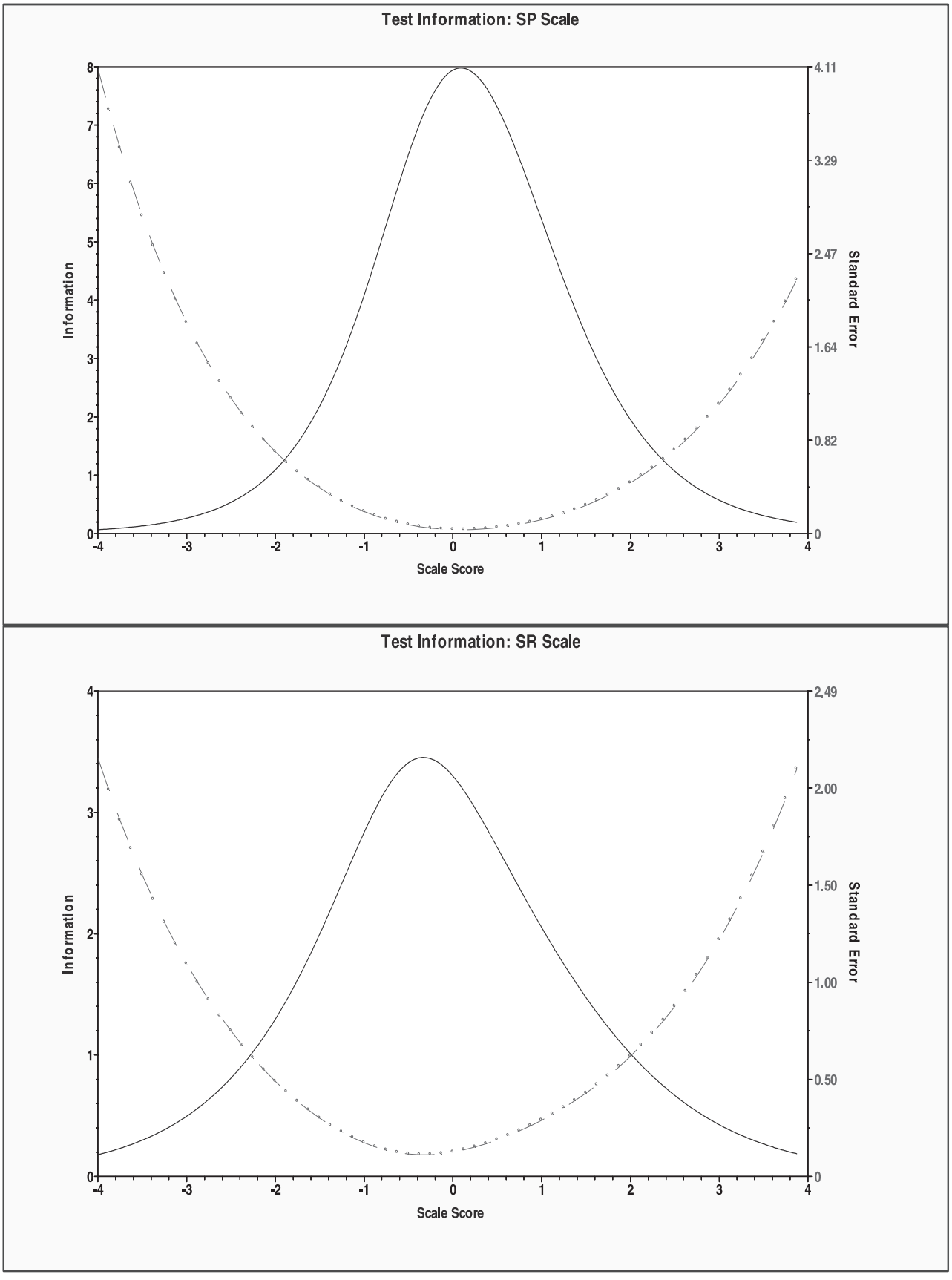

Figure 2. test information and standard error of measurement curves for the short SP and SR scales.

at approximately –0.3 *SD* units below the latent trait mean. In summary, the IRT analyses showed that the item and test properties for the SPSRQ-S are largely effective.

Convergent, Discriminant, and Concurrent Validity

The short SP and SR scales had a nonsignificant correlation of –0.04. Table 7 shows the correlations between the short SP and SR scales and the BIS/BAS Scales, the EPQ-R, and the STAI. The SP scale correlated positively and significantly with both the STAI and the BIS scale, and had low negative correlations with the BAS scales. It further had a significant positive correlation with Neuroticism, and a significant and negative correlation with Extraversion. The SR scale had significant and positive correlations with the BAS scales, but had low, nonsignificant correlations with the BIS and STAI measures. The SR scale had significant and positive correlations with Extraversion and Psychoticism and did not correlate significantly with Neuroticism.

Table 7. Correlations between the SPSRQ-S and the EPQ-R, BIS/BAS and STAI Measures

Measure	SP	SR
Extraversion	–.49**	.36**
Neuroticism	.66**	–.08
Psychoticism	–.18**	.10
Trait Anxiety	.72**	–.01
BIS	.60**	–.04
Reward Responsiveness	–.07	.29**
Drive	–.12*	.41**
Fun Seeking	–.18**	.47**

*$p < .01$, **$p < .001$.

Discussion

The aim of Study 2 was to examine the psychometric properties of the SPSRQ-S using an independent sample of participants. This involved using CFA to test the factorial structure; IRT to examine the item and test properties; and correlations with other commonly used BIS and BAS measures to assess the convergent, discriminant, and concurrent validity of the short scales. A CFA of the SPSRQ-S was initially undertaken to assess the fit of the full measurement model. Although this model showed better global fit indices when compared to the full SPSRQ, it did not quite reach the criteria normally accepted as good overall model fit (Hu & Bentler, 1999). This is consistent with previous research that has also found that CFA models of the SPSRQ, even with problematic items trimmed from the scales, do not show adequate model fit (Cogswell et al., 2006; O'Connor et al., 2004). Despite the lack of good global fit, the factor loadings in the SPSRQ-S CFA model were generally high. IRT analyses of the two scales showed that nearly all items had good item properties. Examination of the TIF and *SEM* curves for the SP and SR scales showed that both scales had relatively good measurement precision across the trait range from –1 *SD* unit below the mean to +1 *SD* unit above the mean. It should be noted, however, that the short SR scale had a somewhat lower Cronbach's α value when compared to that normally found for the full scale, although it was still acceptable. This may be at least partially because of the reduced number of items in the short measure.

The short SP and SR scales had a low intercorrelation, which is consistent with the design intentions of Torrubia et al. (2001), and consistent with previous research that has shown the full SP and SR scales to have a low intercorrelation (Caseras et al., 2003; O'Connor et al., 2004; see Pickering, in press, for an alternative argument). The short SP scale showed good convergent and discriminant validity. It tended to correlate positively with other BIS-related measures and did not correlate with BAS measures. It also showed good concurrent validity as it correlated positively with Neuroticism and negatively with Extraversion, which is similar to the pattern shown in previous research (Caseras et al., 2003). The short SR scale showed good convergent and discriminant validity. It correlated positively with other BAS-related measures and did not tend to correlate with BIS measures. It also correlated positively with Extraversion, but did not correlate significantly with Neuroticism. Overall, the findings from this study suggest the SPSRQ-S may be a viable alternative to the full scale.

General Discussion

The aims of the two studies were to examine the psychometric properties of the SPSRQ using both factor analytic and IRT-based methods, and to propose a potential short form of the measure. These aims were founded on the basis that previous research with the SPSRQ has highlighted potential problems with the measure, both in terms of factor structure and individual item properties (Cogswell et al., 2006; O'Connor et al., 2004). This study also appears to be the first that has used IRT to examine the psychometric properties of this particular scale. It has been argued earlier that IRT can potentially provide a more rigorous assessment of the psychometric properties of a scale. In Study 1, we identified problematic items from previous studies and used the results of our analyses to trim problematic items and develop a short version of the measure, the SPSRQ-S. The IRT analyses, in particular, showed that a number of items in the full scale were contributing very little to the overall measurement precision of the scales. In Study 2, an independent sample was used to examine the psychometric properties of the SPSRQ-S. This study showed that the SPSRQ-S has good psychometric properties that represent an improvement on the psychometric properties of the full scale, with the exception that the Cronbach's α of the SR scale was somewhat lower (although still acceptable). Cor-

relations between the SPSRQ-S and other commonly used measures of the BIS and BAS were largely consistent with previous research using the full measure (Caseras et al., 2003; O'Connor et al.; Torrubia et al., 2001).

Overall, the findings from the two studies support previous research that has highlighted problems with the SPSRQ (Cogswell et al., 2006; O'Connor et al., 2004). It had been noted in these previous studies that future revision of the SPSRQ might be required. The revised measure developed in the current studies may be a starting point for this process. A strength of these studies is that we have used IRT methods, in addition to the more routinely used factor analytic methods, to highlight problems with the scale and suggest a revised version. Nonetheless, there are several limitations in the current studies that should be noted. First, the current studies only examined the SPSRQ-S in relation to other self-report questionnaire instruments and did not include behavioral or experimental correlates of the revised scale. This would be an important task for future studies. Second, both samples included in the current studies were comprised of Australian university students. It is important that the psychometric properties of the SPSRQ-S be examined using other samples. Third, it should be noted that the current studies used an English-language version of the measure, whereas the original scale used the Catalan language. As yet, no study appears to have examined measurement invariance across the two language versions of the scale, either full or alternate versions. This would be an important step for any future research. It has been noted in previous research that the SPSRQ may need further revision, particularly in terms of the language used in the English version of the scale (O'Connor et al., 2004). The current studies would also suggest that future revision of the scale, beyond removing problematic items, should be carried out.

In summary, the current studies have used factor analytic and IRT methods to create a short version of the SPSRQ. The results of the studies would suggest the SPSRQ-S may be a viable alternative in future use of the scales, or may at least be a starting point for more substantial revisions of the measure. The properties of the SPSRQ-S, should, however, be carefully examined across further independent samples.

References

Baker, F. (2001). *The basics of item response theory.* College Park, MD: ERIC Clearinghouse on Assessment and Evaluation, University of Maryland.

Beauducel, A., Kersting, M., & Liepmann, D. (2005). A multitrait-multimethod model for the measurement of Sensitivity to Reward and Sensitivity to Punishment. *Journal of Individual Differences, 26*, 168–175.

Bentler, P.M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*, 238–246.

Birnbaum, A. (1968). Some latent-trait models and their use in inferring an examinee's ability. In F.M. Lord & M.R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–479). Reading, MA: Addison-Wesley.

Bolt, D.M., Hare, R.D., Vitale, J.E., & Newman, J.P. (2004). A multigroup item response theory analysis of the Psychopathy Checklist-Revised. *Psychological Assessment, 16,* 155–168.

Browne, M.W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K.A. Bollen & J.S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Newbury Park, CA: Sage.

Carver, C.S., & White, T.L. (1994). Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS scales. *Journal of Personality & Social Psychology, 67*, 319–333.

Caseras, X., Ávila, C., & Torrubia, R. (2003). The measurement of individual differences in behavioral inhibition and behavioral activation systems: A comparison of personality scales. *Personality and Individual Differences, 34*, 999–1013.

Cogswell, A., Alloy, L.B., van Dulmen, M.H.M., & Fresco, D.M. (2006). A psychometric evaluation of behavioral inhibition and approach self-report measures. *Personality and Individual Differences, 40*, 1649–1658.

Cooper, A., Gomez, R., & Aucote, H. (2007). The Behavioural Inhibition System and Behavioural Approach System (BIS/BAS) Scales: Measurement and structural invariance across adults and adolescents. *Personality and Individual Differences, 43*, 295–305.

Cooper, A.J., Perkins, A.P., & Corr, P.J. (2007). A confirmatory factor analytic study of anxiety, fear, and behavioral inhibition system measures. *Journal of Individual Differences, 28*, 179–187.

Corr, P.J. (2001). Testing problems in J.A. Gray's personality theory: A commentary on Matthews and Gilliland (1999). *Personality and Individual Differences, 30*, 333–352.

Corr, P.J. (2004). Reinforcement sensitivity theory and personality. *Neuroscience and Biobehavioral Reviews, 28*, 317–332.

Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists.* Mahwah, NJ: Erlbaum.

Eysenck, H.J., & Eysenck, S.B.G. (1991). *Manual of the Eysenck Personality Scales (EPS Adult).* London: Hodder & Stoughton.

Fraley, R.C., Waller, N.G., & Brennan, K.A. (2000). An item response theory analysis of self-report measures of adult attachment. *Journal of Personality and Social Psychology, 78*, 350–365.

Gray, J.A., & McNaughton, N. (2000). *The neuropsychology of anxiety* (2nd ed.). Oxford: Oxford University Press.

Gomez, R., Cooper, A., & Gomez, A. (2005). An item response theory analysis of the Carver and White (1994) BIS/BAS scales. *Personality and Individual Differences, 39*, 1093–1103.

Hu, L.T., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.

McNaughton, N., & Corr, P.J. (2004). A two-dimensional neuropsychology of defense: Fear/anxiety and defensive distance. *Neuroscience and Biobehavioral Reviews, 28*, 285–305.

Muthen, L.K., & Muthen, B.O. (2007). *MPlus 4.2*. Los Angeles: Muthen & Muthen.

O'Connor, R.M., Colder, C.R., & Hawk, L.W.J. (2004). Confirmatory factor analysis of the Sensitivity to Punishment and Sensitivity to Reward Questionnaire. *Personality and Individual Differences, 37*, 985–1002.

Pickering, A.D. (in press). Formal and computational models of reinforcement sensitivity theory. In P.J. Corr (Ed.), *The reinforcement sensitivity theory of personality*. Cambridge, UK: Cambridge University Press.

Pickering, A.D., Corr, P.J., & Gray, J.A. (1999). Interactions and reinforcement sensitivity theory: A theoretical analysis of Rusting and Larsen (1997). *Personality and Individual Differences, 26*, 357–365.

Pickering, A.D., Corr, P.J., Powell, J.H., Kumari, V., Thornton, J.C., & Gray, J.A. (1997). Individual differences in reactions to reinforcing stimuli are neither black nor white: To what extent are they gray? In H. Nyborg (Ed.), *Scientific study of human nature: Tribute to Hans J. Eysenck at eighty* (pp. 36–67). Amsterdam: Pergamon/Elsevier.

Reise, S.P., Smith, L., & Furr, R.M. (2001). Invariance on the NEO PI-R Neuroticism scale. *Multivariate Behavioral Research, 36*, 83–110.

Smillie, L.D., & Jackson, C.J. (2005). The appetitive motivation scale and other BAS measures in the prediction of approach and active avoidance. *Personality and Individual Differences, 38*, 981–994.

Smillie, L.D., Pickering, A.D., & Jackson, C.J. (2006). The new reinforcement sensitivity theory: Implications for personality measurement. *Personality and Social Psychology Review, 10*, 320–335.

Steiger, J.H. (1990). Structural model evaluation and modification. *Multivariate Behavioral Research, 25*, 173–180.

Steinberg, L., & Thissen, D. (1995). Item response theory in personality research. In P.E. Shrout & S.T. Fiske (Eds.), *Personality research, methods, and theory: A Festschrift honoring Donald W. Fiske* (pp. 161–181). Mahwah, NJ: Erlbaum

Spielberger, C.D., Gorsuch, R., Lushene, R., Vagg, P.R., & Jacobs, G.A. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists Press.

Tabachnick, B.G., & Fidell, L.S. (2007). *Using multivariate statistics* (5th ed.). Boston: Pearson Education.

Torrubia, R., Ávila, C., Moltó, J., & Caseras, X. (2001). The Sensitivity to Punishment and Sensitivity to Reward Questionnaire (SPSRQ) as a measure of Gray's anxiety and impulsivity dimensions. *Personality and Individual Differences, 31*, 837–862.

Velicer, W.F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*, 321–327.

Zimowski, M.F., Muraki, E., Mislevy, R.J., & Bock, D.R. (2003). *BILOG-MG 3*. Lincolnwood, IL: Scientific Software International.

Andrew Cooper

Department of Psychology
Goldsmiths, University of London
New Cross, SE14 6NW
London
UK
E-mail a.cooper@gold.ac.uk

