The Youth Self Report: validity of the Russian version

Georgy Vasin^{1,*}, Marina Lobaskova¹ and Elena Gindina¹

¹Psychological Institute of Russian Academy of Education, Laboratory of developmental behaviour genetics, Moscow, Russia

Abstract. The Youth Self Report (YSR) is a popular questionnaire for measuring behavioral and emotional problems in adolescents. In this study we investigate the validity of the Russian adaptation of this questionnaire on a sample of nonreferred Russian-speaking adolescents. Our results show that the theoretical factor structure fits the data well and the scale scores have adequate reliability. We also demonstrate the validity of select scale scores with a correlation analysis. We conclude that the Russian version of the YSR has good psychometric properties, although some issues need to be taken into consideration in its practical application.

1 Introduction

The Youth Self Report (YSR) was developed by T. Achenbach as part of the Achenbach System of Empirically Based Assessment (ASEBA, [1]). It is a self-reported measure of behavioral and emotional problems for 11-18 year old adolescents. The first version of the YSR was developed in 1987 based on a factor analytic study of recorded behavioral symptoms of 1000 clinically referred youths. The YSR scores behavioral and emotional problems on a continuous scale, providing an alternative to the binary scoring of the DSM.

The YSR factor structure consists of 8 syndrome scales: Anxious/Depressed, Withdrawn/Depressed, Somatic Complaints, Social Problems, Thought Problems, Attention Problems, Rule-Breaking Behavior and Aggressive Behavior. For a more general description of emotional and behavioral problems, two syndrome groupings are available: Externalizing and Internalizing problems, as well as a Total Problems scale (see Figure 1 for details).

Fig. 1. The model tested in this study.

The statistically derived taxonomy of psychopathology reflected in the YSR has been very

popular with researchers. It has been replicated in over 23 societies for adolescents [2, 3]. Ivanova already reported the validity of this taxonomy for Russian-speaking adults [4, 5]. In this study we present evidence of this taxonomy holding for a sample of Russian-speaking adolescents as well. We also provide other evidence of the validity of YSR scores, since updating the validity evidence of questionnaires is a vital part of their application in practice.

The Russian version of the YSR was adapted from the 1999 revision by members of the Laboratory of Developmental Behavioral Genetics at the Psychological Institute of the Russian Academy of Education. A reverse translation procedure was used.

2 Method

For this study we sampled 522 adolescents from 5 Russian cities. Participants were picked to cover the full range of ages supported by the YSR, from 11 to 18 years old. 223 participants were female (54%). We did not specifically seek out referred adolescents for our study. In addition to the YSR, the 11-15 year olds in our study (376 participants in total) filled out the Early Adolescent Temperament Questionnaire – Revised (EATQ-R; [6]), their results were used to study the concurrent validity of several YSR scales.

In this validity study we analyzed the construct validity and reliability of all the YSR scales, as well as convergent and concurrent validity of some of the scales. We performed confirmatory factor analysis, internal consistency and correlation analysis with the EATQ-R scale scores, the procedures for all of which are discussed in more detail below.

Confirmatory factor analysis is a useful tool in validity studies. Since the theoretical factor structure is already defined by the test framework, it provides us

* Corresponding author: georgy.vasin@pirao.ru

with a structural model to test against the empirical data. Good model fit is evidence of construct validity. Following the procedure outlined in Ivanova and colleagues' paper [2], we did not include the total problems scale into our model. The final model included 79 items scored 0 (not true), 1 (sometimes true) or 2 (often true) and 10 latent variables (8 syndrome scales as well as internalizing and externalizing syndrome groups). Item 81 was removed from the model due to convergence issues.

Model parameters were estimated with the weighted least squares method (WLS) and delta parameterization [7]. The WLS was shown to be effective for data with categorical variables with few categories [8, 9]. Model fit was assessed with the following coefficients: root-mean-square error of approximation (RMSEA; [10]), Tucker-Lewis Index (TLI; [11]), Comparative Fit Index (CFI; [12]) and weighted root mean residuals (WRMR [13]). Good model fit is indicated by $RMSEA \leq 0.06$, $TLI \geq 0.86$, $CFI \geq 0.86$ and $WRMR \leq 1$ [13]. Yu and Muthen (2002, [12]) identified RMSEA as the best performing index for WLS estimation, so we treated it as the primary fit statistic in our study.

Reliability is a prerequisite of validity. We measured the reliability of the YSR scales using 3 coefficients of internal consistency. High internal consistency usually indicates high reliability. Chronbach's Alpha is traditionally reported in reliability studies, we used the standardized version which is calculated by averaging correlation coefficients between all items within a scale [14]. However, given the heterogeneity of the YSR scales, we also included two McDonald's Omega coefficients into our analysis. McDonald's Omega is calculated based on factor loadings obtained from second order exploratory factor analysis. The total Omega coefficient (ω_t) is based on the sum of squared loadings on all second order factors, while the hierarchical Omega coefficient (ω_h) is based only on the squared loadings on the first order factor [14]. All 3 coefficients range from 0 to 1, values above 0.6 indicate adequate internal consistency.

In addition to the confirmatory factor analysis and internal consistency analysis, in some cases we were also able to test the convergent and concurrent validity of the YSR scale scores with a correlation analysis. Over 70% of the participants in our study filled out the EATQ-R questionnaire. Even though EATQ-R measures temperament traits, it contains several scales that relate to the YSR scales: affiliation, fear, inhibitory control, shyness, aggression and depression. Since aggression and depression are measured by both questionnaires, their scores can be used to calculate the convergent validity of the respective YSR scales, with high positive correlations indicating high validity. We also tested the concurrent validity of the anxiety/depression, delinquency, aggression, attention problems and withdrawn YSR scales. All scale scores can be considered continuous variables, so Pearson correlation coefficients were used.

3 Results and discussion

3.1 Confirmatory factor analysis

The model parameters converged normally. Judging by the primary model fit indicator in our study, $RMSEA=0.033$, the test data fit the theoretical framework quite well. However the other fit indices did not show good fit: $CLI=0.774$; $TFI=0.776$; $WRMR=1.5$. Another indicator of mismatch between the empirical and theoretical factor structure was the high amount of item variance unaccounted for by the model (0.74 on average). Additionally, 7 items had nonsignificant factor loadings: 1, 45, 62, 74, 89, 103, 111.

Problems with the YSR factor structure have been well documented in literature [15, 16; 17]. Most often they have to do with poor model fit according to CFI and TLI. Achenbach addressed some of the criticism by pointing out that RMSEA is the most adequate fit statistic for the YSR data, while CFI and TLI may not perform as well on categorical data [18]. Another argument was that poor model fit may have been the result of several items not functioning as intended due to a mostly nonreferred sample [18]. Both of these arguments apply to our results as well. Ivanova et al. also reported nonsignificant factor loadings for a number of items in 4 of the 23 samples in their study, with Sweden having as much as 16. When analyzed together with our results, there does not seem to be any pattern in what items have nonsignificant loadings. Thus it is likely that the instances of nonsignificant factor loadings are culture specific. The low amount of item variance explained by the model might be accounted for by the YSR scales' heterogeneity (see 3.2).

3.2 Reliability

Internal consistency was within the acceptable range for all the YSR scales, except for the Social Problems scale. The internal consistency coefficients are presented in Table 1.

Table 1. Internal consistency of the YSR scales.

	Alpha	ω_h	ω_t
Internalizing	0,79	0,58	0,81
Withdrawn/Depressed	0,65	0,47	0,73
Somatic Complaints	0,77	0,58	0,81
Anxious/Depressed	0,78	0,57	0,81
Externalizing	0,85	0,56	0,87
Rule-Breaking Behavior	0,67	0,63	0,73
Aggressive Behavior	0,82	0,64	0,83
Social Problems	0,48	0,4	0,59
Thought Problems	0,65	0,47	0,74
Attention Problems	0,67	0,44	0,73
Total Problems	0,93	0,33	0,93

The 3 different coefficients offer a complex look at the internal consistency of the YSR scales. Chronbach's Alpha is considered the lower bound of scale reliability [19]. Chronbach's Alpha ranged from 0.48 to 0.93, with the Social Problems scale having the lowest internal consistency. All internal consistency coefficients are

functions of the number of items in a test, so naturally, scales with more items, such as Internalizing, Externalizing and Total Problems scales, had higher internal consistency.

The Omega coefficients provide more interesting results. As indicated by the high amount of unexplained item variance in our confirmatory factor analysis, each scale likely measures more than one latent construct. Given that scale scores are merely sums of all item scores, all these extra dimensions contribute to the scale score. These extra dimensions can be viewed as measurement bias, in which case the ω_h is a more appropriate representation of scale reliability. However, if we look at some of the items in the YSR scales, their heterogeneity appears intentional. For example, the Aggressive Behavior scale includes items pertaining to the adolescent's mood and temper, attacking others and breaking their things, as well as shouting and screaming. Part of their shared variance can certainly be attributed to aggressive behavior, but they also form secondary factors, such as mood-related questions, physical and verbal confrontation, etc. This would be especially impactful in nonreferred samples, since it is unlikely that many participants would exhibit all behavior problems at once. In this context, ω_h seems to be a more adequate representation of scale reliability, since it takes into account the secondary factors as well as the main factor. As shown in Table 1, ω_h and ω_t have highly contrasting values, which presents a flexible approach to assessing reliability of the YSR scales. This argument is best exemplified by the difference between ω_h and ω_t for the Total Problems scale, which we know is an amalgamation of all the YSR scales.

Overall the YSR Russian version scales have acceptable internal consistency, with the exception of the Social Problems scale.

3.3 Correlation analysis

We tested the validity of select YSR scale scores by comparing them to EATQ-R scale scores in a correlation analysis.

Aggression scales from the YSR and the EATQ-R had a high positive correlation, $r=0.59$, $p<0.01$. The YSR Anxious/Depressed scale also had a high positive correlation with the EATQ-R Depression scale, $r=0.49$, $p<0.01$. The constructs measured by these pairs of scales are quite similar, so the positive correlations demonstrate convergent validity of said YSR scales.

The correlation analysis also provided evidence of concurrent validity of some of the YSR scale scores. Anxious/Depressed correlated with Affiliation ($r=0.22$, $p<0.01$) and Fear ($r=0.32$, $p<0.01$). The EATQ-R Affiliation scale measures the desire to make personal connections with others and Fear measures negative affectivity related to possible future stress or pain, so higher scores on these two scales should coincide with higher scores on the Anxious/Depressed YSR scale. The Inhibitory Control EATQ-R scale scores correlated negatively with Aggression ($r=-0.36$, $p<0.01$) and Attention Problems ($r=-0.3$, $p<0.01$). Inhibitory Control

is the ability to suppress inappropriate behavior, so higher control should be related to lower aggression and better attention (i.e. fewer problems). The YSR Withdrawn/Depressed was positively related to the EATQ-R Shyness ($r=0.41$, $p<0.01$). Since the former scale measures discomfort in social situations and the latter – a tendency to avoid other people, a positive correlation was expected. All of these results point towards concurrent validity of the respective YSR scale scores.

This analysis also demonstrated that the latent constructs measured by the YSR scales are not too wide, evidenced by the many nonsignificant correlation coefficients. For example, Affiliation and Somatic Complaints are theoretically unrelated and we can see it empirically as well ($r=0.04$, $p=0.58$).

Our study was limited by the lack of a contrasting referred sample.

4 Conclusion

In this study we examined the validity of the YSR scores for nonreferred Russian-speaking adolescents. Confirmatory factor analysis demonstrated good model fit, although it also uncovered some problems with the questionnaire's factor structure, such as 7 items with nonsignificant factor loadings. Seven out of 8 scales have adequate reliability. The exception is the Social Problems scale, we recommend interpreting its scores with caution when working with nonreferred adolescents. Correlation analysis also showed the validity of the Aggression, Attention Problems, Withdrawn/Depressed and Anxious/Depressed scale scores. We concluded that the Russian version of the YSR displayed adequate psychometric properties and can be used to study the behavioral and emotional problems of Russian-speaking adolescents.

Acknowledgements

This study was supported by the grant from the Russian Foundation for Humanities (grant RFH N 15-06-10724).

References

1. I T. M. Achenbach, *Integrative guide for the 1991 CBCL/4-18, YSR, and TRF profiles*, (Department of Psychiatry, University of Vermont, 1991)
2. M. Y. Ivanova et al., *J. Consult. Clin. Psychol.*, **75** (2007)
3. M. Y. Ivanova et al., *Journal of Clinical Child and Adolescent Psychology*, **36** (2007)
4. M. Y. Ivanova et al., *J Psychopathol Behav*, **37** (2015)
5. M. Y. Ivanova et al., *Int J Clin Health Psychol*, **15** (2015)
6. D. M. Capaldi, M. K. Rothbart, *J. Early Adolesc.*, **12** (1992)

7. B. Muthén, T. Asparouhov, Mplus web notes, **4** (2002)
8. D. B. Flora, P. J. Curran, Psychol. methods, **9** (2004)
9. M. Rhemtulla, P. É. Brosseau-Liard, V. Savalei, Psychol. methods, **17** (2012)
10. M. W. Browne, R. Cudeck, K. A. Bollen, J. S. Long, Sage focus editions, **154** (1993)
11. L. R. Tucker, C. Lewis, Psychometrika, **38** (1973)
12. P. M. Bentler, Psychological Bulletin, **107** (1990)
13. C.-Y. Yu, *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (Doctoral dissertation, University of California, Los Angeles, 2002)
14. W. Revelle, R. E. Zinbarg, Psychometrika, **74** (2009)
15. D. Drotar, R. E. Stein, E. C. Perrin, Journal of Clinical Child Psychology, **24** (1995)
16. B. G. Heubeck, JACP, **28** (2000)
17. G. M. Macmann, D. W. Barnett, E. J. Lopez, School Psychology Review, **22** (1993)
18. T. M. Achenbach, L. Rescorla, *ASEBA school-age forms & profiles* (Aseba Burlington, 2001)
19. N. M. Webb, R. J. Shavelson, E. H. Haertel, *Handbook of statistics* (Elsevier, 2006)