

---

## THE CONVERGENT VALIDITY OF THE ROMANIAN VERSION OF THE BEHAVIORAL ASSESSMENT OF CHILDREN: A MULTITRAIT-MULTIMETHOD ANALYSIS

---

*Dragoș Iliescu<sup>1,2</sup>, Cătălin Nedelcea<sup>1</sup>, Florina Golu<sup>1</sup>, and Andrei Ion<sup>1,2\*</sup>*

---

<sup>1</sup> Department of Psychology, Faculty of Psychology and Educational Sciences, Universitatea din București, Bucharest, Romania.

<sup>2</sup> Assessment and Individual Differences – AID Lab, Department of Psychology, University of Bucharest, Romania.

---

### Abstract

This study assessed the validity of the Behavior Assessment System for Children 2nd Edition, by providing evidence based on a multitrait-multimethod (MTMM) analysis performed on data obtained with a culturally adapted form of the test. The study included two Romanian samples (161 adolescents aged 12-18 years and 91 children aged 8-12 years) that were each assessed with all the three age corresponding forms (self, parent, and teacher) of the test. We used a confirmatory factor analytic framework in order to test the convergent and discriminant validity of BASC-2, by generating a CFA-based multi-trait, multi-method latent factor correlation matrix. Results showed that, for both the Adolescent and Child forms, the coefficients falling under the validity diagonal (mono-trait, hetero-method) have the highest median values, compared to the coefficients encountered in the monomethod block and those in the heteromethod-heterotrait triangles. The study provides evidence for the construct validity of the multi-trait, multi-rater assessment system that is the basis of the BASC-2.

**Keywords:** multi-rater assessment, multi-trait-multi-method, confirmatory factorial analysis.

This paper discusses the convergent and discriminant validity of multi-rater behavioral assessment systems. The focal issue is of interest in light of current

---

\* Correspondence concerning this article should be addressed Andrei Ion, Department of Psychology, University of Bucharest, 90th Panduri Ave., Bucharest, 030018, Romania.  
Tel/Fax: 0040 314 253445.  
Email: andrei.ion@fpse.unibuc.ro

discussions regarding both (a) the fundamental question of the measurement quality of multi-informant assessments in clinical practice and (b) methods of choice in assessing said measurement quality for the more complex multi-form and multi-rater assessment systems, especially considering the newer evolutions in statistical modeling. This paper discusses and illustrates a statistical approach to this problem by performing a multitrait-multimethod (MTMM) confirmatory factor analysis (CFA) on data collected with the Behavior Assessment System for Children 2nd Edition (BASC-2; Reynolds & Kamphaus, 2004).

Multi-rater behavioral rating systems are broadband mental health assessments that are typically used for assessing aspects pertaining to social, emotional, and behavioral problems (Shapiro & Heick, 2004). From a measurement point of view, the main feature of these comprehensive assessment systems is that they employ multi-trait, multi-method, and multi-informant assessments (Haynes & O'Brien, 2000), collecting information about the existence and/or frequency of specific thoughts, feelings and behaviors corresponding to different traits, across multiple settings (Ramsay et al., 2002). Examples of such measures are the Achenbach System of Empirically Based Assessment (ASEBA), the Conners Rating Scales - Revised (CRS-R), the Behavior Assessment System for Children 2nd Edition (BASC-2), and others (Ramsay et al., 2002).

All sources of validity are important when evaluating the adequacy of clinical interpretations based on the data collected from informants, but construct validity is an especially crucial element in this context (Kane, 2013), comprising such aspects as factor structure, convergent and discriminant validity. Convergent validity refers to the agreement between different measures/methods that estimate the same construct(s). Discriminant validity captures the extent to which different constructs are distinctive from one another. Both these forms of validity are typically tested via the MRMM approach pioneered by Campbell & Fiske (1959), in which different assessment methods that capture the same trait(s) are expected to correlate highly with one another, but less strongly with measures that focus on different traits.

Different methodological or analytical innovations have been devised to test these principles (e.g., Hau, Wen & Cheng, 2005; Marsh, 1988; Marsh & Byrne, 1993). One such development is the use of confirmatory factor analysis (CFA) for generating the MTMM matrix. This particular form of CFA, CFA-MTMM, involves the specification of latent factors corresponding to both traits and methods. The latent factors capturing different methods for the same traits, and the latent factors comprising traits assessed via the same method are informative for the construct validity of the respective traits, respectively for a measure's discriminant validity and for the identification of potential method-effects. As important advantages of CFA-MTMM we mention the fact that it separates trait, method and error components while testing the assumptions of the underlying model (Eid et al., 2008), and that the analysis partitions the variance into true variance (trait or method) and error variance, thus correlating latent and not observed variables in the MTMM matrix.

During the past decade, the BASC-2 has emerged as one of the most popular multi-rater behavioral assessment systems in both research and clinical settings (Karr & Garcia-Barrera, 2016). Its items measure a wide array of constructs, tapping into both the clinical and the adaptive functioning area. The four-factor measurement architecture of the BASC-2 has been replicated across several empirical studies (e.g., Splett et al., 2017), and has been shown to be culturally invariant, with item bias (differential item functioning) occurring for a relatively low number of items (Dever, Raines, & Dowdy, 2016). A recent survey of school psychology test usage reported that the BASC Teacher Rating Scales (TRS) was the most frequently used measure in school psychology practice (Benson et al., 2019). Despite its popularity, empirical research regarding the measure's construct validity is generally scarce. Canivez et al., (2021) examined the latent factor structure of the TRS, providing some evidence supporting its construct validity. However, we did not identify any empirical investigations tackling the construct validity for the overall behavioral assessment system and for its individual components, such as the TRS, Parent Rating Scales (PRS) and Self Rating Scales (SRS). Additionally, previous empirical investigation tapping into its construct validity were conducted at the scale level, rather than at the item level, thus limiting our understanding regarding the relationships between items and latent factors among the system's various components.

### *The Present Study*

In this paper we conducted a CFA-MTMM in order to investigate the convergent and discriminant validity of one of the most frequently used assessment systems in school psychology practice (Benson et al., 2019), the BASC-2. We focused on two distinct samples (one of children and one of adolescents), that were each assessed with all the three age-corresponding forms (self, parent, and teacher) of the BASC-2.

## **Method**

### *Participants*

Sample 1 consists of 161 focal adolescents, among them 58 girls (36.02%), aged between 12 and 18 years ( $M = 16.30$ ,  $SD = 1.74$ ). These participants provided self-assessments and were also assessed by one of their parents (74.53% by mothers) and by one teacher. Sample 2 consists of 91 focal children, among them 60 girls (65.93%), aged between 8 and 12 years ( $M = 10.25$ ,  $SD = 1.22$ ). These participants provided self-assessments and were also assessed by one of their parents (81.32% by mothers) and by one teacher.

*Procedure*

Data was collected in Bucharest, the capital city of Romania, by sixteen school psychologists, who approached households based on a route sampling schema. If children in the targeted age range were present in the household, the parent and self-assessment forms were administered to the parent and child respectively. The parent was then asked to contact one of the teachers of the child and obtain his or her assessment. A total of 4000 households were approached in order to collect the 252 records in our two samples. Children up to 9-10 years only have one teacher in the Romanian educational system. Children above 9-10 years have several teachers; one of them is the class principal and this teacher provided the current assessment. The teacher assessments were handed by the teachers back to the parents in a closed envelope and were later collected by the researchers. The parents received a copy of the interpretative report for their child.

*Measures*

The BASC-2 (Reynolds & Kamphaus, 2004) is a system for the multidimensional assessment of behavior in children and adolescents, developed on a multimethod framework. For the current study data was collected using the Child (-C) and Adolescent (-A) versions of the BASC-2, collecting teacher (TRS), parent (PRS) and self-assessments (SRP). The present study used the Romanian version of the BASC-2, which was adapted to the Romanian language and culture (Reynolds et al., 2011) through a process that closely followed the International Guidelines on Test Adaptation of the International Test Commission (ITC, 2005). In some countries and cultures, the next form of this family of tests, namely the BASC-3 (Reynolds & Kamphaus, 2015), was already adapted and adopted – still we consider the usage of the BASC-2 in this context to be relevant on two accounts. First, we use the current data to illustrate not so much the validity of this particular test or form of a test, but rather in order to illustrate an approach regarding the internal validity of multi-rater systems in general. Second, a test is not necessarily obsolete because a new version of the test has been published by the test publisher – the International Test Commission's (ITC) Guidelines for Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal (International Test Commission, 2015) make this quite clear. Irrespective of the publication of the BASC-3, many countries and cultures still use the BASC-2, as these multi-rater systems are notoriously difficult to adapt culturally and especially to standardize.

*Analytical approach*

To estimate convergent and discriminant validity, we focused on those scales that were present in at least two of the three different-method forms for each age group (parent, teacher, or self-report). Scales that were included only in a single

form were excluded: Activities of daily living (present only in the parent forms), Learning Problems and Study Skills (present only in the teacher forms). A total number of 13 scales remained; these are all present in both the parent and teacher forms and 5 of them are also present in the self-report forms.

The convergent and discriminant validity of the BASC-2 was tested via confirmatory factorial analysis (CFA), following the general approach outlined by Kenny and Kashy (1992). The analyses were conducted with Mplus version 7 (Muthen & Muthen, 2008). The CFA-MTMM analyses were conducted at the item level, each item loading onto its corresponding trait or method dimension, respectively. To tackle the effects of potentially abnormality in the data, we used an MLR estimation. We generated a CFA-based multi-trait, multi-method latent factor correlation matrix, separately for the Child and Adolescent data.

## Results

Table 1 reports the scales included in the study. The CFA-MTMM correlation matrix is outlined in Tables 2 and 3, for the Child and Adolescent forms, respectively. A summary of the intercorrelations encountered across the various diagonals (Campbell, & Fiske, 1959) was included in Table 4. For both forms, the coefficients falling under the validity diagonal (mono-trait, hetero-method) have the highest median values, compared to the coefficients encountered in the monomethod block and those in the heteromethod-heterotrait triangles.

**Table 1.** The BASC-2 scales included in the study

	Scale name	PRS-C and PRS-A	TRS-C and TRS-A	SRP-C and SRP-A
1	Hyperactivity	x	x	x
2	Aggression	x	x	-
3	Conduct Problems	x	x	-
4	Anxiety	x	x	x
5	Depression	x	x	x
6	Somatization	x	x	-
7	Attention Problems	x	x	x
8	Atypicality	x	x	x
9	Withdrawal	x	x	-
10	Adaptability	x	x	-
11	Social Skills	x	x	-
12	Leadership	x	x	-
13	Functional Communication	x	x	-

*Note:* The Externalizing Problems Composite contains scales 1, 2, 3; the Internalizing Problems Composite contains scales 4, 5, 6; the Behavioral Symptoms Index contains scales 1, 2, 5, 7, 8, 9; the Adaptive Skills Composite contains scales 10, 11, 12, 13.

*Convergent validity (heteromethod-monotrait)*

The median correlation in the TRS-C and PRS-C block was  $r = .34$ . Overall, the correlations within this block tend to be moderate, being in line with the premises of the MTMM approach. The TRS-C and SRP-C overlapped in only five dimensions, and all displayed significant correlations, with a median correlation of  $r = .20$ . For the PRS-C and SRP-C block only four dimensions overlapped, and all of them displayed significant correlations, with the lowest correlation at  $r = .13$  ( $p < .01$ ). The median correlation for the TRS-A and PRS-A block was  $r = .31$ . Generally, the correlations observed within the validity diagonal were higher compared to those encountered in the other remaining blocks. All the six overlapping dimensions for the TRS-A and SRP-A block were significant, being small to moderate. For the PRS-A and SRP-A block all the convergent validities for the five dimensions were in the  $r = .24$ -.29 range.

*Discriminant validity within source (monomethod-heterotrait)*

The highest correlations in the PRS-C block were at  $r = .13$  ( $p < .05$ ), with most correlations being not significant and a median correlation in this block of  $.04$ . In the TRS-C block, the highest correlation emerged at  $r = -.24$  ( $p < .01$ ); 14 significant correlations were observed in this block out of the total of 89 correlations; the median correlation was  $r = -.01$ . Only one significant correlation was encountered within the SRP-C block, which had a median correlation of  $r = .07$ .

The highest correlation within the PRS-C block was  $r = .13$  ( $p < .05$ ), with most correlations being not significant; the median correlation in this block was  $.04$ . For the TRS-C block, only 3 out of 79 intercorrelations emerged as significant, and the median correlation was  $r = .07$ . Only one significant correlation was encountered in the SRP-C block ( $r = .14$ ;  $p < .05$ ), the median correlation was  $r = .07$ . The correlations observed in the SPR-C block were mostly not significant; the highest correlation was  $r = .14$  ( $p < .01$ ).

In the TRS-A block 32 traits correlated significantly with one another, representing more than one third of the total number of intercorrelations; the median correlation in the TRS-A mono-method block was  $r = .09$ . In the PRS-A block, 24 out of 79 dimensions had significant correlations, a pattern of intercorrelations similar to the one observed in the TRS-A block; the median correlation in the PRS-A mono-method block was  $r = .05$ . The intercorrelations within the SRP-A block had a median of  $r = .07$ ; only 9 out of 26 correlations were significant.

*Discriminant validity between source (heteromethod-heterotrait)*

In the PRS-C block the median correlation was  $r = .03$ , with the largest correlations at  $r = .13$  ( $p < .05$ ) and only six out of 78 correlations flagged as

significant ( $p < .05$ ). In the TRS-C block a total of 13 out of 78 correlations emerged as significant; the highest correlation was  $r = .19$  ( $p < .01$ ), and the median correlation in this block was  $r = -.01$ . Only 2 out of ten correlations were significant in the SPR-C block, the highest of them at  $r = .14$  ( $p < .01$ ); the median correlation in this block was  $r = .08$ . The highest correlation in the PRS-A block was  $r = -.24$  ( $p < .01$ ); 38 out of 78 correlations were significant and the median correlation in this block was  $r = .09$ . In the TRS-A block the median correlation was  $r = .05$ ; the largest correlation was  $r = .25$  ( $p < .001$ ), and 31 out of 78 correlations were significant. The median correlation within the SRP-A block was  $r = .12$ , the highest correlation was  $r = .35$  ( $p < .001$ ) and only six out of the total of 15 correlations were not significant.

## **Discussion**

### *Summary of findings*

Our study enhances the current understanding regarding the construct validity of a behavioral assessment system by analyzing the BASC-2 through the lens of CFA-MTMM. Taken together, our results illustrate the method, and show that most of the individual scales in the various BASC-2 rating forms (parent, teacher, self-report) showed convergent and discriminant validity, both at the between and within-sources level.

Several specific conclusions related to the validity of the BASC-2 also emerged. First, the mid-sized correlations from the hetero-method mono-trait blocks suggest that the sources (parent, teacher, self-reports) complement each other in assessing the various constructs. The partial overlap between the three sources when assessing the same constructs indicates the necessity to consider all three perspectives when making clinical judgements or when drawing conclusions, cautioning practitioners against relying on single source data as evidence for clinical judgments. More specifically, considering the mid-sized associations between the same constructs measured by different informants, it is likely that data collected from each informant will reflect a relatively unique construct-relevant variance. Thus, data collected from multiple informants should result in more precise, accurate and valid clinical judgements and decisions. Precision should be higher because the relative degree of overlap between the different sources should translate into increased reliability or less noise in measurement. Validity should improve because each source also captures construct variance that is unique, in addition to variance that is shared with one or several of the other sources. Overall, assessing clinical and adaptive behaviors in children and adolescents via the three different sources may therefore become a common practice in school psychology settings. Second, the mostly non-significant correlations encountered in both the method blocks and the source blocks show that the scales are conceptually distinct from one another, having

a minimal degree of overlap. Third, overall, the scale correlations encountered throughout the three areas (heteromethod-monotrait, monomethod-heterotrait and heteromethod-heterotrait) suggest that the BASC-2 scales have relatively unique variances that are statistically distinguishable from one another. Fourth, relatively similar scale correlations emerged in the corresponding scales across the two forms (child and adolescent).

In order to draw conclusions with respect to our overarching objective, i.e., analyzing the relevance of the BASC-2 in guiding test-based interpretations and clinical judgements, we refer to the four conditions outlined by Benson et al. (2018): (1) test-based scores should provide a good representation of the target construct; (2) test-based scores should be conceptually distinct from one another; (3) test-based scores should be replicable across different methods and datasets; and (4) test-based scores should exhibit relatively unique variance, not overlapping with conceptually-similar constructs. Our dataset did not allow us to test whether the BASC-2 scores respect the first condition outlined by Benson et al. (2018). A few dimensions, such as Hyperactivity and Somatization displayed larger than expected correlations across the various blocks outlined by the MTMM approach, indicating a slightly larger degree of overlap between the contents measured by these dimensions and other traits, deviating from the second condition. However, our CFA-MTMM analysis suggests that BASC-2 test-scores meet the conditions outlined by Benson et al. (2018) for generating meaningful scale-derived interpretations.

By conducting the CFA-MTMM analysis at the item level we addressed a limitation that was identified in other empirical investigations as well (e.g., Canivez et al., 2021), especially in light of the fact that the BASC-2 and BASC-3 technical manuals report only scale-level CFA analyses. We did not identify any other studies that deployed the CFA-MTMM analysis framework on the BASC-2 or the recently revised version, BASC-3. Consequently, our paper expands our understanding regarding the relationships of methods and constructs in a multiple informant assessment system.

From a methodological point of view, we point out that the CFA approach to MTMM is underrepresented in the psychometric literature, despite its obvious advantages, such as distinguishing trait and method effects from measurement error variance or isolating trait or method effects by correlating latent traits or method variables. Our paper illustrates how MTMM analyses can be conducted in a CFA framework.

### *Limitations*

Caution must be taken when interpreting our conclusions as they were marked by several limitations. First, CFA-MTMM is a “data-hungry” technique typically requiring large sample sizes. However, considering the complexity of collecting multi-source data, our samples are rather small, and the results must be interpreted with caution. Second, our observations could be affected by the



convenience approach through which we sampled the participants: opportunity sampling can in no way ensure representativity or diversity in samples. Third, we did not collect any additional measure for tapping into the constructs measured by the BASC-2; this would have greatly enhanced our understanding regarding the complex construct-method inter-relationships, so it is a direction worth considering in future endeavors.

## **Conclusion**

Our data suggests that the BASC-2 is a useful tool for school psychologists, clinicians, teachers, and other practitioners in mental health. It can appropriately guide effective strategies and intervention plans, can be included with confidence in assessment and intervention services, and can be an important starting point for supplementary assessment, if needed.

## **Author Note**

We have no known conflict of interest to disclose.

## **Implications**

Our findings argue in favor of using data from multiple informants, in our case children, parents and teachers, for measuring various behavioral dimensions (both clinical and adaptive). These sources not only add to the measure's overall validity, but the sources complement each other, suggesting that the practice of collecting data from such multiple sources enhances the validity of multi-informant, multi-dimensions behavioral assessment systems.

## **References**

- Benson, N. F., Floyd, R. G., Kranzler, J. H., Eckert, T. L., Fefer, S. A., & Morgan, G. B. (2019). Test use and assessment practices of school psychologists in the United States: Findings from the 2017 National Survey. *Journal of School Psychology, 72*, 29-48. <https://doi.org/10.1016/j.jsp.2018.12.004>
- Benson, N. F., Beaujean, A. A., McGill, R. J., & Dombrowski, S. C. (2018). Revisiting Carroll's survey of factor-analytic studies: Implications for the clinical assessment of intelligence. *Psychological Assessment, 30*, 1028–1038. <https://doi.org/10.1037/pas0000556>.

- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 56(2), 81–105. <https://doi.org/10.1037/h0046016>
- Canivez, G. L., von der Embse, N. P., & McGill, R. J. (2021). Construct validity of the BASC-3 teacher rating scales: Independent hierarchical exploratory factor analyses with the normative sample. *School Psychology*, 36(4), 235. <https://doi.org/10.1037/spq0000444>
- Dever, B. V., Raines, T. C., & Dowdy, E. (2016). Factor structure and differential item functioning of the BASC-2 BESS Spanish Language Parent Form. *School Psychology Quarterly*, 31(2), 213. <https://doi.org/10.1037/spq0000131>
- Eid, M., Nussbeck, F. W., Geiser, C., Cole, D. A., Gollwitzer, M., & Lischetzke, T. (2008). Structural equation modeling of multitrait-multimethod data: Different models for different types of methods. *Psychological Methods*, 13, 230–253. <https://doi.org/10.1037/a0013219>
- Hau, K. T., Wen, Z., & Cheng, Z. (2005). *Structural equation model and its applications*. (2nd ed.) Beijing: Educational Science Publishing House.
- Haynes, S. N., & O'Brien, W. H. (2000). *Principles and Practice of Behavioral Assessment*. New York: Plenum Publishers.
- International Test Commission (2005). *International Guidelines on Test Adaptation*. [www.intestcom.org]
- International Test Commission (2015). *International Guidelines for Practitioner Use of Test Revisions, Obsolete Tests, and Test Disposal*. [www.intestcom.org]
- Kane, M. (2013). The argument-based approach to validation. *School Psychology Review*, 42(4), 448–457. <https://doi.org/10.1080/02796015.2013.12087465>
- Karr, J. E., Garcia-Barrera, M. A., Holdnack, J. A., & Iverson, G. L. (2017). Using multivariate base rates to interpret low scores on an abbreviated battery of the Delis–Kaplan executive function system. *Archives of Clinical Neuropsychology*, 32(3), 297–305. <https://doi.org/10.1093/arclin/acw105>
- Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*, 112(1), 165–172. <https://doi.org/10.1037/0033-2909.112.1.165>
- Marsh, H. W., & Byrne, B. M. (1993). Confirmatory factor analysis of multitrait-multimethod self-concept data: Between-group and within-group invariance constraints. *Multivariate Behavioral Research*, 28(3), 313–349. [https://doi.org/10.1207/s15327906mbr2803\\_2](https://doi.org/10.1207/s15327906mbr2803_2)
- Marsh, H. W. (1988). Multitrait-Multimethod Analyses. In J. P. Keeves (Ed.), *Educational Research Methodology, Measurement, and Evaluation: An International Handbook* (pp. 570–578). Oxford: Pergamon.
- Muthén, L.K. and Muthén, B.O. (1998–2017). *Mplus User's Guide*. Eighth Edition. Los Angeles, CA: Muthén & Muthén.
- Ramsay, M. C., Reynolds, C. R., & Kamphaus, R. W. (2002). *Essentials of behavioral assessment*. New York: John Wiley & Sons.
- Reynolds, C. R., & Kamphaus, R. W. (2004). *Behavior Assessment System for Children, Second Edition*. Minnesota: AGS Publishing.

- Reynolds, C. R. & Kamphaus, R.W. (2015). Behavior Assessment System for Children (3rd ed.). Bloomington, MN: Pearson.
- Reynolds, C. R., Kamphaus, R. W., Ion, A., Mitrofan, M., & Iliescu, D. (2011). *Manual tehnic pentru Sistemul de Evaluare Comportamentală pentru Copii [Technical Manual for the Behavior Assessment System for Children]*. Bucharest: OS Romania.
- Shapiro, E. S., & Heick, P. F. (2004). School Psychologist Assessment Practices in the Evaluation of Students Referred for Social/Behavioral/Emotional Problems. *Psychology in the Schools, 41*(5), 551- 561. <https://doi.org/10.1002/pits.10176>
- Splett, J. W., Raborn, A., Lane, K. L., Binney, A. J., & Chafouleas, S. M. (2017). Factor analytic replication and model comparison of the BASC-2 Behavioral and Emotional Screening System. *Psychological assessment, 29*(12), 1543. <https://doi.org/10.1037/pas0000458>

