salud mental

Diagnostic operationalization and phenomenological heterogeneity in psychiatry: the case of attention deficit hyperactivity disorder

Giovanni Abrahão Salum,[1,2] Ary Gadelha,[1,3] Guilherme Vanoni Polanczyk,[1,4] Eurípedes Constantino Miguel,[1,4] Luis Augusto Rohde[1,2,4]

[1] Instituto Nacionais de Ciência e Tecnologia de Psiquiatria do desenvolvimento para Crianças e Adolescentes, Brazil.

[2] Department of Psychiatry, Hospital de Clinicas de Porto Alegre, Universidade Federal de Rio Grande do Sul, Porto Alegre, Brazil.

[3] Department of Psychiatry, Universidade Federal de São Paulo, São Paulo, Brazil.

[4] Department of Psychiatry, Escola de Medicina, Universidade de São Paulo, São Paulo, Brazil.

Correspondence:
Luis Augusto Rohde
Hospital de Clínicas de Porto Alegre, Federal University of Rio Grande do Sul. Ramiro Barcelos 2350, Porto Alegre, Brazil, 90035-003.
Phone: +55 51 3359 - 8094
Email: gsalumjr@gmail.com

ABSTRACT

Introduction. We assessed the impact of polythetic conceptualizations of mental disorders on the validity and reliability of psychiatric diagnosis, with a specific focus on two levels of heterogeneity: phenomenological and pathophysiological. **Objective.** We investigated this issue using attention deficit hyperactivity disorder (ADHD) as an example. **Method.** We examined individuals from two samples enriched for psychopathology (*n* = 1 255 children in Porto Alegre and 1 257 children in São Paulo, Brazil). We conducted a series of data analyses to investigate phenomenological heterogeneity, including confirmatory factor analysis. We also investigated pathophysiological heterogeneity using symptom-level regressions between ADHD symptoms and four neurocognitive processes consistently linked to ADHD (working memory, inhibitory control, intra-subject variability in reaction times, and temporal processing). Lastly, we assessed the performance of polythetic systems for reliability testing inter-rater and test-rest reliability of two well-known symptomatic scales. **Results.** Among the 116 200 possible combinations of symptoms to achieve DSM symptomatic threshold for categorical ADHD diagnosis, we found 173 combinations in the two independent samples, and only four were replicated in both samples (2.3%). We also found that the number of ADHD symptoms is a poor indicator of variation in the general ADHD latent trait. Overall, symptoms did not have specific profiles of associations with any of the neurocognitive processes. Reliability analyses revealed that increasing the number of items augments overall reliability of measurements. **Discussion and conclusion.** Our findings illustrate both potential benefits and problems inherent to the polythetic system for ADHD. Implications for the search of mechanisms underlying psychiatric disorders are discussed.

Keywords: Attention, psychiatry, diagnosis.

RESUMEN

Antecedentes. Evaluamos el impacto de las conceptualizaciones politécnicas de los trastornos mentales en la validez y la fiabilidad del diagnóstico psiquiátrico, con un enfoque específico en dos niveles de heterogeneidad: fenomenológico y fisiopatológico. **Objetivos.** Investigamos este problema utilizando el trastorno por déficit de atención e hiperactividad (TDAH) como ejemplo. **Método.** Examinamos individuos de dos muestras enriquecidas por psicopatología (*n* = 1 255 niños en Porto Alegre y 1 257 niños en São Paulo, Brasil). Llevamos a cabo una serie de análisis de datos para investigar la heterogeneidad fenomenológica, incluido el análisis factorial confirmatorio. También investigamos la heterogeneidad fisiopatológica utilizando regresiones al nivel de síntomas entre los síntomas del TDAH y cuatro procesos neurocognitivos consistentemente vinculados al TDAH (memoria de trabajo, control inhibitorio, variabilidad intrasujeto en tiempos de reacción y procesamiento temporal). Por último, evaluamos el rendimiento de los sistemas politéticos para la prueba de confiabilidad interevaluador y la confiabilidad *test-rest* de dos escalas sintomáticas bien conocidas. **Resultados.** Entre las 116 200 posibles combinaciones de síntomas para alcanzar el umbral sintomático del DSM para el diagnóstico categórico de TDAH, encontramos 173 combinaciones en las dos muestras independientes y sólo cuatro se replicaron en ambas muestras (2.3%). También encontramos que la cantidad de síntomas de TDAH no es un buen indicador de la variación en el rasgo latente general del TDAH. En general, los síntomas no tenían perfiles específicos de asociaciones con ninguno de los procesos neurocognitivos. Los análisis de confiabilidad revelaron que aumentar el número de artículos aumenta la confiabilidad general de las mediciones. **Discusión y conclusión.** Nuestros hallazgos ilustran tanto los beneficios potenciales como los problemas inherentes al sistema politécnico para el TDAH. Se discuten las implicaciones para la búsqueda de mecanismos subyacentes a los trastornos psiquiátricos.

Palabras clave: Atención, psiquiatría, diagnóstico.

INTRODUCTION

In May 2013, the American Psychiatric Association published the fifth edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). The DSM-5 was initially planned to integrate findings from neuroscience to the diagnostic criteria (Hyman, 2007). Nevertheless, the DSM-5 Task Force soon realized the complexities and limitations associated with including biomarkers (e.g., genetic, imaging, blood) into the diagnostic system and the new version of the manual conserved its descriptive-phenomenological nature. This decision was mostly driven by the observation that evidence suggesting a potential role for biological markers of mental disorders were restricted to group level differences, and none of them had sufficient validity at the individual level to demonstrate clinical utility (Kapur, Phillips, & Insel, 2012). Among a variety of reasons that may be responsible for this lack of translation, "heterogeneity" in the way we classify psychiatric disorders is considered essential (Kapur et al., 2012; Sonuga-Barke, 2013; 2010).

Psychiatric disorders are defined in terms of polythetic operationalized diagnostic criteria, i.e., a combination of a certain number of symptoms (a collection of behaviors, emotions, thoughts, and sensory phenomena) that needs to be perceived by the individual and/or by others as causing significant impairment. They are called polythetic because each diagnosis shares a number of characteristics which occur commonly in members of a group but none of which is essential for group membership. The inclusion of polythetic operationalized diagnostic criteria to our current classificatory manuals dates back to the publication of the Feighner criteria (McLaughlin & Nolen-Hoeksema, 2011; Nolen-Hoeksema & Watkins, 2011), which formed the basis for the development of the Research Diagnostic Criteria, which in turn were central to the development of the DSM-III. Since then, all subsequent versions of the DSM adopted the polythetic system for a variety of diagnoses, allowing for phenotypic variation in the symptom manifestations of a disorder as a way to provide more diagnostic flexibility or, in other words, increase "coverage" (Merport & Recklitis, 2012). In modern psychometric words, the polythetic system is built under the idea that endorsements of diagnostic criteria are only fallible markers of underlying latent constructs that explain symptomatic aggregation.

The adoption of polythetic systems has important implications for both the validity and the reliability of mental disorders. Due to its inclusive nature, polythetic systems introduce a great deal of variability in the clinical description of psychiatric syndromes. This great amount of variability may relate to both problems in the validity due to "true heterogeneity" at several levels, and reliability, that is intrinsically related to "measurement error" of a given diagnosis.

With respect to validity, it is important to bear in mind that psychiatric disorders are likely to be fuzzy "kinds of things" like "species", populations with central paradigmatic and more marginal members. Therefore, heterogeneity in any type of classification is to be expected. Different clinical presentations can be a result of true heterogeneity in at least three levels: the etiological level, the pathophysiological level, and the phenomenological level (Marco et al., 2009; McLoughlin et al., 2009). The etiological level refers to how a given clinical condition can be caused by many different combinations of sufficient sets of etiological factors. The pathophysiological level refers to how a given clinical condition can be a result of distinct pathophysiological processes. The phenomenological level refers to how a given unique clinical condition may be described differently by different subjects. A schematic representation of the various levels of heterogeneity is depicted in Supplementary Figure 1 at https://www.ufrgs.br/prodah/site/wp-content/uploads/2018/11/Supplementary-Figures-SaludMental.pdf.

With respect to reliability, diagnostic criteria are fallible since endorsers are imperfect observers and reporters of symptoms. Variability may be introduced by a variety of factors such as poor clinicians' abilities, poor wording, lack of transcultural sensitivity, and memory bias, among others, that may indicate "measurement error." In addition, the assessment of reliability is further complicated by the wax and waning of some psychiatric symptoms and also to different perspectives of observers regarding symptom presence or absence and the degree they affect subjects, which certainly represent much more than measurement error (Penninx et al., 2011).

Few studies have investigated the "trade-offs" of polythetic diagnostic systems for validity and reliability of mental disorders and more specifically for attention deficit hyperactivity disorder (ADHD). First, with respect to validity, specific implications of the operationalization of criteria for heterogeneity are not often discussed. Regarding phenomenological heterogeneity, Olbert, Gala, and Tupler (2014) assessed the number of combinations of symptoms for the diagnoses of several disorders. The authors demonstrated that there are 116 200 possible combinations of symptoms to fulfill DSM-IV ADHD diagnosis. However, few studies evaluated how many of these combinations can be found in real samples. Also, it is not clear how symptom count strategy relates to the underlying latent traits that are thought to underlie symptom endorsement. In addition, few studies investigated pathophysiological heterogeneity at the symptom level. Second, the implications of the polythethic systems to reliability are not well studied either. For example, few studies assess how increasing the number of symptoms for a given trait impacts on test-retest and informant reliability.

Here, we demonstrate the implications of the current operationalization of mental disorders for validity (focusing on heterogeneity) and reliability of ADHD (American Psychiatric Association, 1994). First, we investigated phenomenological heterogeneity assessing the number of possible

combinations of symptoms to achieve DSM ADHD diagnosis. Then, we investigated how the symptom count strategy relates to the latent ADHD construct in terms of variation in the general factor. We advance further investigating pathophysiological heterogeneity of ADHD at the symptom level with four well-known neurocognitive validators. Finally, we compared test-retest and informant reliability at the symptom level and at the dimensional level. We performed these four-related analyses using a large community sample of 6-12 year old children from a middle-income country.

METHOD

Ethic statement

This study was approved by the ethics committee of the University of São Paulo (IORG0004884, project IRB registration number: 1132/08). Written consent was obtained from all parents of participants, and verbal assent was obtained from all children.

Brazilian High-Risk Cohort for psychiatric disorders

This report is part of a large community school-based study – the Brazilian High-Risk Cohort (Salum et al., 2015). A total of 57 schools from two cities (22 in Porto Alegre and 35 in São Paulo) participated in screening and enrollment procedures. From this pool of 9.937 interviews, we selected two subgroups: a random ($n = 958$) and high-risk stratum ($n = 1\,524$). For subjects in the random-selection stratum, a simple randomization procedure from school directories was used, without replacement of non-available subjects. Selection for the high-risk stratum involved a risk-prioritization procedure based on family history and current psychiatric symptoms. Further information can be found elsewhere (Salum et al., 2015).

Psychiatric diagnosis

The psychiatric diagnosis was established using the Development and Well-Being Assessment (DAWBA) (Goodman, Ford, Richards, Gatward, & Meltzer, 2000). The DAWBA is a structured interview administered by lay interviewers, which also contains the Strength and Difficulties Questionnaire (SDQ) (a 25-item scale enquiring about behavioral and emotional difficulties) and recorded verbatim responses of any reported problems. Verbatim responses and structured questions are carefully evaluated by psychiatrists, which confirm or refute the diagnosis. All questions are closely related to DSM-IV diagnostic criteria and focus on current problems causing significant distress or social impairment. The DAWBA has been translated to several languages,

and for the present study the Brazilian Portuguese version (Fleitlich-Bilyk & Goodman, 2004) was administered to the biological parents of all children included in the project. Administrations were performed in accordance with previously reported procedures (Goodman, Ford, Richards, et al., 2000). Nine psychiatrists performed the rating procedures. All were trained and supervised by a senior child psychiatrist. A second child psychiatrist rated a total of 200 interviews and the kappa values between raters for ADHD was high (.72).

Child Behavioral Checklist

The Child Behavior Checklist (CBCL) (Achenbach & Rescorla, 2001) is a widely-used questionnaire assessing children's behavior and emotional problems. Lay interviewers administered the CBCL Version for School Aged Children (6-18 year old version). Several studies provided evidences of validity and reliability of the instrument across distinct cultures (Rescorla et al., 2012). Parents rate each item based on a three-point scale: (0) Not True, (1) Sometimes/Somewhat True, and (2) Very True/Often True. For this specific study we used the ADHD scale from the DSM-IV scales (Rescorla et al., 2012).

Strengths and Difficulties Questionnaire

The Strength and Difficulties Questionnaire (SDQ) (Goodman, Ford, Simmons, Gatward, & Meltzer, 2000) is a 25-item scale assessing behavioral and emotional difficulties, as well as their resultant impairment and distress. Parents and teachers rate each item based on a three-point scale: (0) Not True, (1) Somewhat True, and (0) Certainly True. The instrument has shown to be reliable and valid across distinct cultures (Anselmi, Fleitlich-Bilyk, Menezes, Araujo, & Rohde, 2010; Woerner et al., 2004). For this specific study we used the Hyperactivity scale (five items).

Neurocognitive tasks

ADHD has been implicated with a variety of neurocognitive deficits such as behavioral inhibition (Hofmann & Smits, 2008), working memory (Hidalgo, Tupler, & Davidson, 2007), intra-subject reaction time variability (Salum et al., 2012; Telzer et al., 2008), and temporal processing (Costello, Compton, Keeler, & Angold, 2003; Ward, 1974). The battery used included the following tests: a) Two-choice Reaction Time (2C-RT) (Hogan, Vargha-Khadem, Kirkham, & Baldeweg, 2005); b) ConflictControl Task (CCT) (Hogan et al., 2005); c) Go/No-Go (GNG) (Bitsakou, Psychogiou, Thompson, & Sonuga-Barke, 2008); d) Digit span: this is a sub-test of WISC-III (Wechsler, 2002); e) Corsi blocks task (Vandierendonck, Kemps, Fastame, & Szmalec, 2004); f) Time Anticipation tasks – 400ms and 2000ms (TA) Toplak

& Tannock, 2005); and g) Duration Discrimination (DDT) (Toplak, Rucklidge, Hetherington, John, & Tannock, 2003). Description of each test can be found elsewhere (Salum et al., 2015).

Statistical analysis

Phenomenological heterogeneity

We assessed phenomenological heterogeneity through two approaches. First, using combinatorial analyses. Among the 116 220 possible combinations of symptoms that generate the same diagnosis of ADHD described previously by Olbert et al. (2014), we examined the frequency of these combinations in individuals with ADHD diagnosis in two samples enriched for psychopathology: one from Porto Alegre and the other from São Paulo. Second, we used Confirmatory Factor Analysis. The bifactor model provides a way to simultaneously conceptualize both the communality and specificity of symptoms from separate domains (Brunner, Nagy, & Wilhelm, 2012; Castellanos et al., 2005; Glaser, Thomas, Joyce, Castellanos, & Gerhardt, 2005; Krueger et al., 2002). The model comprises a single general factor accounting for covariation among all symptoms along with separate, specific factors of inattention, hyperactivity, and possibly impulsivity that vary orthogonally with the general factor. The bifactor model better fits with multiple pathway theoretical conceptualizations of the disorder, accounting more clearly for disorder heterogeneity (Nigg, Willcutt, Doyle, & Sonuga-Barke, 2005; Sonuga-Barke, 2005). Previous studies investigating correlated, second-order, and bifactor structures of ADHD symptoms provide evidence in favor of a bifactor model of ADHD (Dumenci, McConaughy, & Achenbach, 2004; Gibbins, Toplak, Flora, Weiss, & Tannock, 2011; Martel, Roberts, Gremillion, Von Eye, & Nigg, 2011; Martel, Von Eye, & Nigg, 2010; Toplak et al., 2009).

A bifactor model with one general factor and three specific factors was fitted to polychoric correlations among the DAWBA items using mean- and variance-adjusted weighted least squares (WLSMV) estimator implemented with Mplus 7.0 (Muthén & Muthén, 2012). The goodness of fit was assessed through the following fit indices: chi-square, CFI (comparative fit index), TLI (Tucker-Lewis Index), and RMSEA (root mean square error of approximation). To demonstrate good fit to the data, previous literature suggests that an estimated model should have an RMSEA of near or below .06, and CFI and TLI near or above .95 (Hu & Bentler, 1999). The bifactor model was the model with the better fit among tested models in the sample (Porto Alegre: FP = 72, X2(117) = 365.291, CFI = .993, TLI = .991, RMSEA = .041 90% CI .036, .046; São Paulo: FP = 72, X2(117) = 330.126, CFI = .995, TLI = .993, RMSEA = .038 90% CI .033, .043). For further details see Ref. (Merport, Bober, Grose, & Recklitis, 2012). This analysis was used to compare the ADHD latent trait with the symptom count strategy. For this purpose, only the general ADHD factor was used for analysis as it was the best reliable proxy of the latent ADHD severity.

Pathophysiological heterogeneity

Confirmatory factor analysis was also used to derive a four-factor model of cognition for ADHD using four of the best neurocognitive deficits associated with the disorder: inhibitory based executive function (Hofmann & Smits, 2008), working memory (Hidalgo et al., 2007), intra-subject reaction time variability (Salum et al., 2012; Telzer et al., 2008), and temporal processing (Costello et al. 2003; Ward, 1974). Fit indexes for the model are as follows: X2 = 752.281 (df = 322, p < .001), RMSEA = .024, 90% CI (.021 - .026), CFI = .994; TLI = .994. The indicators for each domain are as follows: (1) Inhibitory-based executive function: percentage of failed inhibitions in the incongruent trials of the CCT and number of commission errors in the GNG task; (2) Working Memory: the level at which the participant failed to correctly repeat the sequences on two consecutive trials at one level of difficulty in Digit Span and Corsi blocks tasks; (3) Intra-subject variability in Reaction Times: the mean intra-subject variability in the reaction times of the 2C-RT, congruent trials of the CCT and in the go trials of the GNG task; (4) and Temporal Processing: the mean percentage of total hits in the 400ms anticipation task, the mean percentage of too early responses in the 2000ms task and the average of the last five reversal values.

Associations between each ADHD symptom and the four neuropsychological domains were investigated using path analysis in separate models for each ADHD symptom (observed variables) and using the four latent factors representing the four neurocognitive domains.

Reliability analysis

Intra-class correlation coefficients and Spearman correlation coefficients were used for both temporal stability and cross-informant reliability analyses. Temporal stability was measured by the DSM-IV ADHD scale of the CBCL in a sub-sample of 772 subjects with a time-lag of one to 17 months. Cross-informant reliability analysis was performed with the Hyperactivity scale from the SDQ in a sub-sample of 1177 subjects that had both parental and teacher data. For both analyses, we compared the performance of each of the items in predicting itself and all other items.

RESULTS

Validity analysis

Phenomenological heterogeneity - Combinatorial analysis

We examined the frequency of symptomatic combinations in individuals with ADHD diagnosis in two samples en-

riched for psychopathology. Groups of 1 255 children in Porto Alegre and 1 257 children in São Paulo, as described above, composed the two samples. A total of 118 and 71 children in Porto Alegre and São Paulo, respectively, had a formal diagnosis of ADHD. From the 189 ADHD cases, we found a total of 173 combinations of the ADHD symptomatic profiles. Therefore, only 16 (8.4%) children with ADHD had a shared profile of symptom combination with another child. In addition, only four out of the 173 combinations were found in both samples (2.3%) (Figure 1; Panel A). A patient-by-patient matrix with the total sample comparing the percentage of symptom agreement taking patient by-patient revealed that the median agreement between symptoms

was 61%, with 30% of the sample showing an agreement lower than half of the symptoms (Figure 1; Panel B).

Phenomenological heterogeneity - Latent trait vs. symptom count

Another way of looking at heterogeneity in ADHD is in terms of its dimensional latent trait. We investigated the associations between the latent ADHD trait and the symptom count using a bifactor model through a Confirmatory Factor Analysis (CFA). Using this model, we investigated the relationship between the symptom count approach and the general factor of the bifactor model that accounts for ADHD severity. We can observe that subjects with the same

PANEL A

n	Combinations	n	Combinations	n	Combinations	n	Combinations	n	Combinations
1	AAAAAAAAAAAAAPPPPP	36	AAPPPAAAPAPPPPPPA	71	APPPPAAAPAPPAAPPPP	106	PPAPPPAPPPAPPPPPPP	141	PPPPPAAPPAPPAAAAPA
2	AAAAAAAAAAAAPPAPPPP	37	AAPPPPAPPPPPPAPAPA	72	APPPPAAPPAPAPPPPPP	107	PPAPPPPPPPPAAPPPPP	142	PPPPPAAPPPPPPPPPPP
3	AAAAAAAAAAAPAAPPPPP	38	AAPPPPPPPPPPPPPPPP	73	APPPPAPPAPAAAAAPPA	108	PPAPPPPPPPPAPPPPPP	143	PPPPPAPAAAAPPPPPAPP
4	AAAAAAAAAAAPAPPAPPP	39	APAAAAAAAPAAPAPPPP	74	APPPPAPPPPPPPPPPPP	109	PPPAAAPPAPPPPPPPPA	144	PPPPPAPPAPPPPPPPPPP
5	AAAAAAAAAAAAPAPPPPP	40	APAAAAAAAPPPPPPAPP	75	APPPPPPPAAAAPPPPPP	110	PPPAAAPPPAPPAAAAPA	145	PPPPPAPPPPAAAAPPPA
6	AAAAAAAAAAPAPAAPPPP	41	APAAAAAPPPPPAPPPP	76	PAAAAAAAAAAAPPPPPA	111	PPPAAPPPPPAAPPPPPP	146	PPPPPAPPPPAPPPPPPP
7	AAAAAAAAAAPPAAPPAPP	42	APAAAAAPAPAAPPPPPA	77	PAAAAAAAPPPPPPPPPA	112	PPPAAPPPPPPAAPPAPP	147	PPPPPAPPPPPPPPPAPA
8	AAAAAAAAAAPPPAPAPPP	43	APAAAAAPAPAPPPPPPP	78	PAAAAAAPPPPAPPPPPP	113	PPPAAPPPPPPPPPPPPP	148	PPPPPAPPPPPPPPPPPA
9	AAAAAAAAAAPPPPPAAPP	44	APAAAAPPAAPPPPPPPP	79	PAAAAPPPAPPPPPPPPP	114	PPPAPAAPAAPPPPPPPP	149	PPPPPAPPPPPPPPPPPP
10	AAAAAAAAAAPPPPPPPPP	45	APAAAPAAAPAPPAPPP	80	PAAPAAAAPPAPPPPPPA	115	PPPAPAPAAPPPPPPPPP	150	PPPPPPAAAAAAAAPPPA
11	AAAAAAAAAPAAPPPPPPP	46	APAAAAPAAPPPPPPPP	81	PAAPAPAAAPPPPPPPPP	116	PPPAPAPAPAAAAPAAAA	151	PPPPPPAAAPAPPPAAPP
12	AAAAAAAAAPPAPPPAAPP	47	APAAAAPPPPPPPAPPPP	82	PAAPPPAPPAPAPAPPPP	117	PPPAPAPAPPAAAAAPPP	152	PPPPPPAAAPPAPPPPPP
13	AAAAAAAAPAAPPAPPAPP	48	APAAAPAAAAAPPPPPA	83	PPAAAAAAAPAPPPPAPA	118	PPPAPPPPPAAPPPPPPP	153	PPPPPPAPPAAAAAAAAA
14	AAAAAAAPPPPAPPPPPPA	49	APAAAPAAAAPPAPPPPP	84	PPAAAAAPAPAPPPPAPA	119	PPPAPPPPPAPAPPPAAP	154	PPPPPPAPPAPPAAAPPA
15	AAAAAAPAPAAPPPPAAPP	50	APAAAPAAPPPPPPPAAP	85	PPAAAAAPPAPPPPPAPP	120	PPPPAAAPPAPPPPPPP	155	PPPPPPAPPPAAPAAPPP
16	AAAAAAPAPPAPAAPPPPP	51	APAAAPAPPPAPPPAPP	86	PPAAAAPAAPAAPPPAPP	121	PPPPAAPAPAAPPAPPP	156	PPPPPPAPPPAAPPAAPA
17	AAAAAAPPPAAPPPPPPPP	52	APAAAPPAPAPPPPPPPA	87	PPAAAAPPAPPAPAPPP	122	PPPPAAPPAPPAPPAPPA	157	PPPPPPPAAAAAAAAAPPP
18	AAAAAAPPPPPAPAPPPP	53	APAAPPPPPPAAPPPPPP	88	PPAAAAPPPPAPAPPPP	123	PPPPAAPPPAPAAPAPAA	158	PPPPPPPAPAAAAAPPAPA
19	AAAAAAPPPPPPPAAPPPA	54	APAPAAAAAPPPPPPAPP	89	PPAAAPAPPPAAPPAPP	124	PPPPAAPPPPPPPPAPPP	159	PPPPPPPPAPPPPPPPPPP
20	AAAAAAPPPPPPPPPPPP	55	APAPAAAAPPAPPPPPPP	90	PPAAAPAPPPPPPPAPP	125	PPPPAPAPAPPAPAAAA	160	PPPPPPPPAAAAAAAAA
21	AAAAAPAAPPPPPPPPPPA	56	APAPAAAAPPPAPAAPPPP	91	PPAAAPPAAPPAPPPPA	126	PPPPAPPAAAAAAAAPA	161	PPPPPPPPPAAAPAAAPA
22	AAAAAPPPPPAPPPPPPPP	57	APAPAAAPPAAPPPPPPA	92	PPAAAPPPPAAAAPPPPP	127	PPPPAPPPAAAAAPAAAA	162	PPPPPPPPPAAAAPPAPAP
23	AAAPAAAPPPPPPPPAPP	58	APAPPAPPPAPAPAPPA	93	PPAAAPPPPPPPPPPPPP	128	PPPPAPPPPAAAPAAAA	163	PPPPPPPPPAPPPAPAPA
24	AAAPAAAPPPPPPPPPPPA	59	APPAAAAAAPPAAPPPPP	94	PPAAPPPAPAAPPAAAPA	129	PPPPAPPPPPPPPAPPAAP	164	PPPPPPPPPAPPPPAPPA
25	AAAPAPAPPPPPPPPPPP	60	APPAAAAAAPPPPPPPPP	95	PPAAPPPPPPPPPPPAPA	130	PPPPAPPPPPPPPPAAPPP	165	PPPPPPPPPAPPPPPPPP
26	AAAPAPPPPAAPPAPPPP	61	APPAAAAPPAAPPPPPPP	96	PPAPAAAPPPPPPPAPPP	131	PPPPAPPPPPPPPPPPPP	166	PPPPPPPPPPAAPPPPPP
27	AAAPPAPAPAPPAPPPPP	62	APPAAAPPPAPPPPPPPP	97	PPAPAAPPPPPAPAPPA	132	PPPPPAAAAAPAPPPPP	167	PPPPPPPPPPAPPPAPPP
28	AAAPPAPPAAAAPPPPPP	63	APPAAPAAAPPPPPPPPP	98	PPAPAPPPPAPPPPPAAP	133	PPPPPAAAAPPPPPPPP	168	PPPPPPPPPPAPPPPPPP
29	AAAPPPAPAPPAPPPPPP	64	APPAAPPPPAAAAAAAA	99	PPAPAPPPPPPAAPPPPA	134	PPPPPAAAAPPPPPPPP	169	PPPPPPPPPPPAPPPPPP
30	AAPAAAPPPAAPPAPPPP	65	APPAPAPPAPPAPPPPPA	100	PPAPPAAPAAAAAAPA	135	PPPPPAAAPAAAAPPPP	170	PPPPPPPPPPPPPPPPPA
31	AAPAPAAAAPPPPPAPP	66	APPPAAPPPAPPPAPPPP	101	PPAPPAAPPPAPPPPPPP	136	PPPPPAAAPPPPPPPPP	171	PPPPPPPPPPPPPPPPAPP
32	AAPAPPPPAAAAAAPAP	67	APPPAPAAPAPPPPPPPP	102	PPAPPAPPPPPPPPPPP	137	PPPPPAAPAAAAAAAPPP	172	PPPPPPPPPPPPPPPPPPA
33	AAPPAAAPPPPPPPPPAA	68	APPPPAPPPPPPPPPPP	103	PPAPPPAAPPPAPPPAPA	138	PPPPPAAPAPAPPPPPPP	173	PPPPPPPPPPPPPPPPPP
34	AAPPAPPPPAAPAPAPPA	69	APPPPAPPPPPPPPPPPP	104	PPAPPPAPAAAAAAPPPP	139	PPPPPAAPPAAAAAAAAA		
35	AAPPPAAAAPPAPPPPPP	70	APPPPAAAPAAPPPPPP	105	PPAPPPAPPPAPPPAAP	140	PPPPPAAPPAAPPPPPA		

Note: A, Absent; P, Present; Symptom order: Fidgets, can't remain seated, Runs or climbs when shouldn't, Can't play quietly, Can't calm down, Blurts out answers, Can't wait for a turn, Butts into conversations or games, Unstoppable talk, Careless mistakes/inattentive, Loses interest, Doesn't listen, Doesn't finish task, Poor self-organization, Avoids tasks needing thought, Loses things, Distractible, Forgetful.
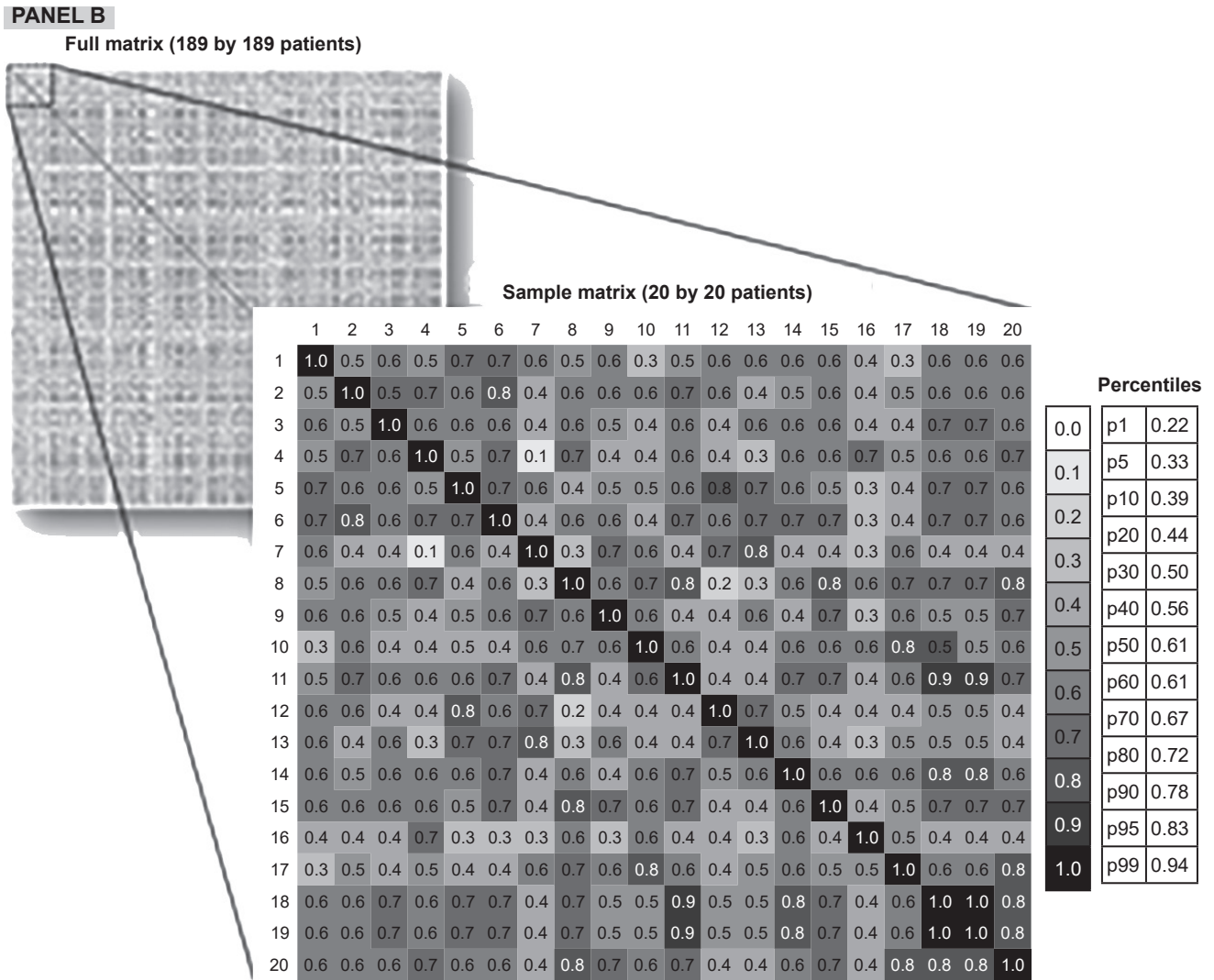
PANEL B

Full matrix (189 by 189 patients)

Sample matrix (20 by 20 patients)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1.0	0.5	0.6	0.5	0.7	0.7	0.6	0.5	0.6	0.3	0.5	0.6	0.6	0.6	0.6	0.4	0.3	0.6	0.6	0.6
2	0.5	1.0	0.5	0.7	0.6	0.8	0.4	0.6	0.6	0.6	0.7	0.6	0.4	0.5	0.6	0.4	0.5	0.6	0.6	0.6
3	0.6	0.5	1.0	0.6	0.6	0.6	0.4	0.6	0.5	0.4	0.6	0.4	0.6	0.6	0.6	0.4	0.4	0.7	0.7	0.6
4	0.5	0.7	0.6	1.0	0.5	0.7	0.1	0.7	0.4	0.4	0.6	0.4	0.3	0.6	0.6	0.7	0.5	0.6	0.6	0.7
5	0.7	0.6	0.6	0.5	1.0	0.7	0.6	0.4	0.5	0.5	0.6	0.8	0.3	0.4	0.7	0.7	0.6	0.7	0.7	0.6
6	0.7	0.8	0.6	0.7	0.7	1.0	0.4	0.6	0.6	0.4	0.7	0.6	0.7	0.7	0.7	0.3	0.4	0.7	0.7	0.6
7	0.6	0.4	0.4	0.1	0.6	0.4	1.0	0.3	0.7	0.6	0.4	0.7	0.8	0.4	0.4	0.3	0.6	0.4	0.4	0.4
8	0.5	0.6	0.6	0.7	0.4	0.6	0.3	1.0	0.6	0.7	0.8	0.2	0.3	0.6	0.8	0.6	0.7	0.7	0.7	0.8
9	0.6	0.6	0.5	0.4	0.5	0.6	0.7	0.6	1.0	0.6	0.4	0.4	0.6	0.4	0.7	0.3	0.6	0.5	0.5	0.7
10	0.3	0.6	0.4	0.4	0.5	0.4	0.6	0.7	0.6	1.0	0.6	0.4	0.4	0.6	0.6	0.6	0.8	0.5	0.5	0.6
11	0.5	0.7	0.6	0.6	0.6	0.7	0.4	0.8	0.4	0.6	1.0	0.4	0.4	0.7	0.7	0.4	0.6	0.9	0.9	0.7
12	0.6	0.6	0.4	0.4	0.8	0.6	0.7	0.2	0.4	0.4	0.4	1.0	0.7	0.5	0.4	0.4	0.4	0.5	0.5	0.4
13	0.6	0.4	0.6	0.3	0.7	0.7	0.8	0.3	0.6	0.4	0.4	0.7	1.0	0.6	0.4	0.3	0.5	0.5	0.5	0.4
14	0.6	0.5	0.6	0.6	0.6	0.7	0.4	0.6	0.4	0.6	0.7	0.5	0.6	1.0	0.6	0.6	0.6	0.8	0.8	0.6
15	0.6	0.6	0.6	0.6	0.5	0.7	0.4	0.8	0.7	0.6	0.7	0.4	0.4	0.6	1.0	0.4	0.5	0.7	0.7	0.7
16	0.4	0.4	0.4	0.7	0.3	0.3	0.3	0.6	0.3	0.6	0.4	0.3	0.6	0.4	0.4	1.0	0.5	0.4	0.4	0.4
17	0.3	0.5	0.4	0.5	0.4	0.4	0.6	0.7	0.6	0.8	0.6	0.4	0.5	0.6	0.5	0.5	1.0	0.6	0.6	0.8
18	0.6	0.6	0.7	0.6	0.7	0.7	0.4	0.7	0.5	0.5	0.9	0.5	0.5	0.8	0.7	0.4	0.6	1.0	1.0	0.8
19	0.6	0.6	0.7	0.6	0.7	0.7	0.4	0.7	0.5	0.5	0.9	0.5	0.5	0.8	0.7	0.4	0.6	1.0	1.0	0.8
20	0.6	0.6	0.6	0.7	0.6	0.6	0.4	0.8	0.7	0.6	0.7	0.4	0.4	0.6	0.7	0.4	0.8	0.8	0.8	1.0

Grayscale scale: 0.0 | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 | 0.6 | 0.7 | 0.8 | 0.9 | 1.0

Percentiles

p1	0.22
p5	0.33
p10	0.39
p20	0.44
p30	0.50
p40	0.56
p50	0.61
p60	0.61
p70	0.67
p80	0.72
p90	0.78
p95	0.83
p99	0.94

Figure 1. Phenotypic heterogeneity due to current diagnostic criteria.
Panel A – The 173 combinations of symptoms among 189 cases of ADHD and concordance between Porto Alegre and São Paulo samples (gray areas show combinations that were found in both samples).
Panel B – Patient by patient matrix showing concordance of Attention Deficit Hyperactivity Disorder (ADHD) symptoms among 189 ADHD children.

symptom count present a wide variation in the latent trait for both subjects with and without ADHD. This approach demonstrates that there is also heterogeneity in terms of severity of attention problems within patients with the same symtpom count (Figure 2).

Pathophysiological heterogeneity

Then, we investigated pathophysiological heterogeneity. Using a set of neuropsychological tests, we fitted a model with these four neurocognitive domains and investigated whether they would be associated individually with each one of the ADHD symptoms. We investigated associations at the symptom level and found in the bivariate analysis that most ADHD symptoms were associated with all four neurocognitive domains, except for most of the impulsivity items that seem to relate more specifically to intra-subject reaction time reliability but not to the other domains (Table 1). Taken together, these findings revealed that there is some level of pathophysiological heterogeneity at the dimensional level of ADHD, but this heterogeneity is not found at the symptomatic level. In other words, no symptom or group of symptoms were particularly associated with any neurocognitive domains.

Reliability

Item reliability analysis

Lastly, it is reasonable to think that investigating behaviors with single items may artificially inflate measurement error. To investigate such effects, we compared the reliability of
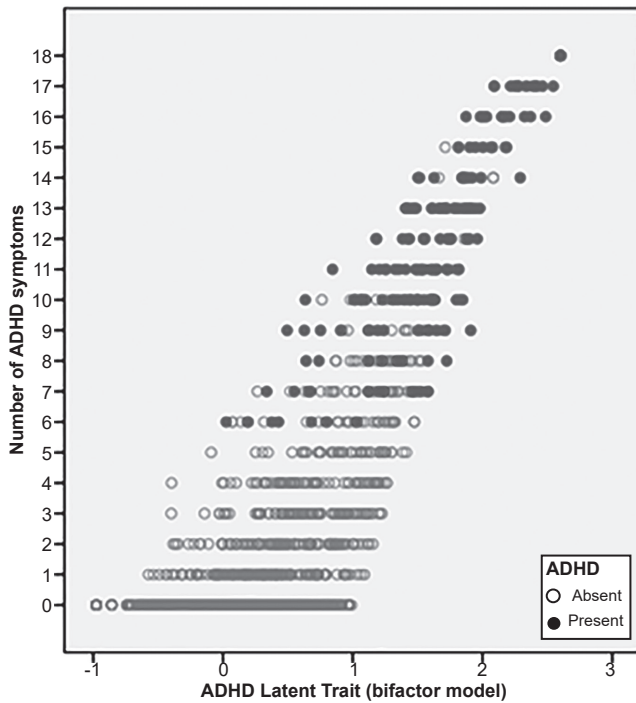
Figure 2. Comparison between Sum of Symptoms and Attention Deficit/Hyperactivity Disorder Latent Traits from the Confirmatory Factor Analysis (Bifactor model, general factor).

specific attention items against the reliability of the attention total scores formed by the sum of several items. We investigated the temporal stability of symptoms (test-retest reliability) and informant effects (inter-rater reliability) for two ADHD-related rating scales: the SDQ – Hyperactivity Scale and CBCL – DSM-IV ADHD Scale.

We can observe that reliability estimates from both temporal stability and cross-informant are better for attention scores if compared to attention items individually (Supplementary Figure 2 at https://www.ufrgs.br/prodah/site/wp-content/uploads/2018/11/Supplementary-Figures-SaludMental.pdf). Items significantly predict later endorsement of themselves, but also predict future endorsement of other items (Table 2). Item endorsements from one informant predict endorsement of the same item for the other informant, but also endorsement of other items by the second informant (Table 3). Nevertheless, it is important to note that item-item correlations are higher than correlation between the item and other items with respect to temporal stability (Table 2). Such effects were not observed for inter-rater reliability (Table 3), as can be noted by overlapping confidence intervals.

DISCUSSION AND CONCLUSION

We explored the various implications for validity (focusing on heterogeneity) and reliability of the current polythetic

Table 1
Associations between ADHD symptoms with four neurocognitive functions

Symptom level regressions	Working Memory			Reaction Time Variability			Inhibitory-Based EF			Temporal Processing		
	Est.	SE	p	Est.	SE	p	Est.	SE	p	Est.	SE	p
Hyperactivity items												
Fidgets	.055	.026	.030	.090	.025	<.001	.090	.024	<.001	.052	.026	.048
Cant remain seated	.056	.024	.019	.063	.023	.007	.095	.023	<.001	.075	.025	.003
Runs when shouldn't	.076	.027	.005	.102	.026	<.001	.095	.026	<.001	.073	.028	.009
Cant play quietly	.051	.026	.053	.082	.026	.002	.099	.025	<.001	.055	.028	.047
Cant calm down	.090	.029	.002	.080	.028	.005	.072	.028	.009	.079	.030	.008
Impulsivity items												
Blurts out answers	-.001	.027	.972	.052	.026	.051	.050	.026	.053	.001	.028	.972
Cant wait for a turn	.024	.028	.395	.065	.028	.018	.059	.027	.027	.019	.029	.504
Butts into conversations	.066	.025	.008	.053	.025	.031	.046	.024	.050	.046	.026	.077
Unstoppable talk	.048	.025	.053	.051	.025	.038	.032	.024	.186	.039	.026	.134
Inattentive items												
Careless mistakes	.116	.027	<.001	.112	.027	<.001	.105	.026	<.001	.099	.028	<.001
Loses interest	.131	.029	<.001	.081	.027	.003	.103	.026	<.001	.129	.029	<.001
Doesn't listen	.130	.028	<.001	.103	.027	<.001	.076	.026	.004	.118	.028	<.001
Doesn't finish task	.088	.026	.001	.077	.025	.002	.072	.024	.003	.094	.026	<.001
Poor self organization	.106	.025	<.001	.068	.025	.006	.043	.024	.070	.103	.026	<.001
Avoids tasks thought	.117	.025	<.001	.101	.024	<.001	.093	.024	<.001	.104	.026	<.001
Loses things	.085	.025	.001	.121	.025	<.001	.073	.024	.002	.067	.026	.010
Distractible	.130	.024	<.001	.114	.023	<.001	.119	.023	<.001	.110	.024	<.001
Forgetful	.117	.026	<.001	.123	.025	<.001	.066	.024	.006	.098	.026	<.001

Note: Estimations in Red indicate worse performance (p level = .05).

Table 2
Spearman Correlation matrix between times 1 and 2 for attention items and attention score (n = 772)

| | Time 2 (1 to 17 months after Time 1) | | | | | | | | Prediction | | Scale |
	Total Score	Item 4	Item 8	Item 10	Item 41	Item 78	Item 93	Item 104	Item-Item	Item-Other	
Time 1											
Total score	.539	.369	.423	.381	.411	.351	.263	.327			.0
Lower	.487	.305	.367	.323	.350	.284	.193	.259			.1
Upper	.595	.430	.478	.441	.473	.415	.337	.393			.2
Item 4	.430	.392	.386	.271	.312	.309	.149	.203	.392	.272	.3
Lower	.372	.327	.324	.204	.249	.240	.081	.133	.327	.205	.4
Upper	.485	.450	.451	.337	.379	.370	.219	.278	.45	.339	.5
Item 8	.440	.309	.406	.307	.297	.348	.175	.232	.406	.278	
Lower	.380	.242	.342	.244	.231	.279	.104	.159	.342	.210	
Upper	.501	.371	.471	.372	.364	.417	.247	.302	.471	.346	
Item 10	.446	.275	.344	.441	.290	.241	.216	.277	.441	.274	
Lower	.390	.210	.281	.382	.223	.171	.142	.206	.382	.206	
Upper	.508	.337	.410	.502	.360	.311	.290	.348	.502	.343	
Item 41	.399	.272	.301	.242	.415	.233	.208	.267	.415	.254	
Lower	.335	.203	.238	.174	.345	.164	.138	.195	.345	.185	
Upper	.460	.334	.364	.311	.482	.307	.275	.337	.482	.321	
Item 78	.400	.299	.389	.208	.266	.346	.126	.233	.346	.254	
Lower	.346	.236	.328	.145	.196	.280	.050	.169	.280	.187	
Upper	.460	.360	.452	.275	.341	.414	.200	.307	.414	.323	
Item 93	.261	.132	.107	.189	.269	.092	.260	.172	.260	.160	
Lower	.185	.059	.035	.120	.201	.022	.189	.098	.189	.089	
Upper	.334	.202	.183	.264	.338	.169	.330	.247	.330	.234	
Item 104	.344	.213	.191	.253	.298	.165	.199	.322	.322	.238	
Lower	.281	.146	.127	.191	.231	.096	.132	.253	.253	.172	
Upper	.406	.276	.261	.326	.370	.235	.272	.391	.391	.307	

Note: Att, Attention score; Item 4, Doesn't finish task; Item 8, Can't concentrate; Item 10, Can't remain seated; Item 41, Impulsive; Item 78, Inattentive; Item 93, Talks too much; Item 104, Noisy. All coefficients are significantly correlated (*p* value < .01).

operationalization of mental disorders, using ADHD as an example. We showed at the level of phenomenological heterogeneity that only 2.3% of the combinations were found in two independent samples, with only 30% of the sample showing an agreement higher than half of the symptoms. We then investigated the relationship between symptom counting with the severity of the latent trait of ADHD. We showed that there is a wide variation in severity in subjects showing the same symptom count, which may indicate the fragilities of the system to capture the severity of the trait. At the level of pathophysiological heterogeneity, we found no evidence that specific symptoms were associated with specific pathophysiological processes, and most ADHD symptoms were associated to all four pathophysiological processes investigated (except for impulsivity items). For reliability, we found evidence that attention scores (summing up items) were more reliable if compared to attention items individually for both temporal and cross-informant stability. In addition, items significantly predict later endorsement of themselves, but also predict future endorsement of other items.

The current operationalization of ADHD diagnostic criteria generates an enormous amount of diagnostic possibilities. We showed that different patterns of symptomatic combinations are found across samples and no obvious common pattern emerged from the analysis. It is important to note that what we demonstrate here is not specific to ADHD. Polythetic conceptualization is the soul of current psychiatric diagnosis and is found in the definition of most mental disorders. Other studies also demonstrated this amount of potential combinations for personality disorders (Cooper & Balsis, 2009; Cooper, Balsis, & Zimmerman, 2010). We are not arguing that there is this number of subtypes of ADHD out there. However, the issue of the "intrinsic heterogeneity" introduced by the diagnostic system is not often discussed in the literature, especially with respect to its impact to the search of biological markers for mental disorders.

Another important implication of the polythetic system is that it assumes that each symptom is "created equal," i.e., they have the same weight to the definition of the latent construct. We were able to show that individuals with the same symptom count in fact lie at very different points of the latent construct. Since ADHD is best conceptualized as a dimension rather than a category (Coghill & Sonuga-Barke, 2012), the issue of how we assess severity is crucial to the definition of diagnostic thresholds.

Table 3
Spearman Correlation matrix between parent and teacher rating of attention items and attention score (n = 1177)

	Teacher-rated						Prediction		
	Sum	Poor concentration	Restless	Fidgety	Good attention	Reflective	Item-Item	Item-Other	Scale
Parent-rated									
Total Score	.326	.315	.253	.231	.276	.214			.1
Lower	.270	.260	.198	.176	.221	.155			.2
Upper	.373	.367	.307	.286	.327	.265			.3
Poor concentration	.268	.274	.171	.153	.257	.179	.274	.190	.4
Lower	.216	.217	.113	.103	.200	.119	.217	.134	.5
Upper	.322	.332	.224	.210	.307	.232	.332	.243	
Restless	.250	.214	.204	.180	.210	.197	.204	.200	
Lower	.190	.155	.145	.126	.148	.143	.145	.143	
Upper	.303	.270	.255	.236	.265	.251	.255	.256	
Fidgety	.184	.159	.172	.177	.126	.107	.177	.141	
Lower	.127	.101	.112	.121	.067	.047	.121	.082	
Upper	.238	.213	.227	.234	.179	.160	.234	.195	
Good attention (i)	.289	.292	.220	.182	.256	.179	.256	.218	
Lower	.234	.239	.164	.123	.200	.122	.200	.162	
Upper	.340	.343	.274	.233	.310	.232	.310	.271	
Reflective (i)	.180	.199	.141	.131	.145	.102	.102	.154	
Lower	.121	.140	.085	..070	.085	.043	.043	.095	
Upper	.235	.255	.197	.186	.201	.159	.159	.210	

Note: (i) Inverse items.

We advanced the study of validity investigating pathophysiological heterogeneity. We were able to show that ADHD symptoms relate to all investigated neurocognitive domains (except for impulsivity items). This is consistent with the view that symptoms are a common final via of different dysfunctional mental processes. This is also consistent with the view that most diagnostic combinations generated by the polythetic system did not provide significantly different phenotypes with respect to the pathophysiological level, since they do not relate specifically to any of the four neuropsychological domains evaluated.

In contrast, results from reliability analysis showed that increasing the number of symptoms increases reliability of the latent trait, which is consistent with the idea that part of variability is due to a measurement error, instead of true heterogeneity. However, it is possible that general effects are being found in the literature, only because they are more reliable and not because specific effects are not real. It is well know that lack of reliability attenuates effect sizes and decreases the power of statistical tests, both of which compromise the ability to provide the evidence necessary to validate specific contributions (Kraemer & Thiemann, 1987; 1989).

The DSM was not designed to capture the underlying pathophysiology of mental disorders (Kraemer, 2007). Nevertheless, pathophysiological research so far has invested an enormous amount of effort to uncover the "joints of nature" using the DSM vocabulary. Those questions have direct implications for what philosophical conceptualization we have about mental disorders. We are assuming here that psychiatric disorders are what Kendler (2008) call things with "mechanistic property cluster," i.e., "sets of symptoms that are connected through a system of causal relations." In this model, not all members need overlap in some single set of traits; rather, members are clustered near one another in a feature space because of developmental evolutionary and physiological causal mechanisms and constraints. This view encourages the thought that there are robust explanatory structures to be discovered underlying most psychiatric disorders. Therefore, although research trying to uncover these historically situated syndromes is plausible, we should shift from the question about the essences of psychiatric kinds to a quest for the complex and multi-level causal mechanisms that produce, underlie, and sustain mental disorders (Faraone, Kunwar, Adamson, & Biederman, 2009).

Another view is that "research world" should rupture with the current systems and that we need to emphasize the identification of biological processes mediating mental functions that cut across psychiatric diagnoses for advancing research (Kapur et al., 2012). Initiatives such as the Research Domain Criteria (RDoC) are promising to overcome such limitations and thereby sidestepping the issue of heterogeneity introduced by diagnostic systems (Sanislow et al., 2010). Nevertheless, this type of alternative may be especially important for new insights into therapeutics rather then advances in nosology.

Our study has some limitations. First, our analysis investigating the pathophysiological heterogeneity is restricted to four specific domains of cognition, and symptom-specific associations may be found with other domains of cognition. Second, reliability analysis was investigated using sub-scales of two scales and we do not have assessed test-retest and informant effects for all 18 ADHD symptoms. Lastly, our analysis is restricted to the phenomenological and pathophysiological levels and we do not evaluate the heterogeneity at the etiological level. Nevertheless, we were able to further advance our understanding about the implications of the polythetic system to validity and reliability using a variety of different statistical methods in a large sample of children from the community.

In conclusion, we demonstrated both strengths and weaknesses of the polythetic conceptualization of mental disorders in the current diagnostic systems using ADHD as a prototype. Advances in psychiatry will need a continuous effort to bridge clinicians and researchers together in order to understand mechanisms of mental disorders through continuous decomposition and reassembly (Kendler, 2008).

Funding

Conflicts of interests

Luis Augusto Rohde has received grant or research support from, served as a consultant to, and served on the speakers' bureau of Eli Lilly and Co., Janssen, Medice, Novartis and Shire. The ADHD and Juvenile Bipolar Disorder Outpatient Programs chaired by Dr Rohde have received unrestricted educational and research support from the following pharmaceutical companies: Eli Lilly and Co., Janssen, and Novartis. Dr Rohde has received authorship royalties from Oxford Press and ArtMed and travel grants from Shire to take part in the 2018 APA annual meeting and from Novartis to take part of the 2016 AACAP annual meeting. Guilherme Polanczyk is employed by the University of São Paulo. He receives grant or research support from CNPq, FAPESP, Fundação Maria Cecilia Souto Vidigal (FMCSV), Grand Challenges Canada, and the Bill and Melinda Gates Foundation. He has served as a consultant to Shire, Teva, and Medice; and has received royalties from Editora Manole. Giovanni Salum, Ary Gadelha, and Eurípides Miguel declare no conflicts of interest.

REFERENCES

Achenbach, T. M. & Rescorla, L. A. (2001). *Manual for the ASEBA school-age forms and profiles*. Burlington: University of Vermont.

American Psychiatric Association. (1994). *Diagnostic and Statistical Manual of Mental Disorders-Text Revision(4 ed.)*. Washington: American Psychiatric Association.

Anselmi, L., Fleitlich-Bilyk, B., Menezes, A. M., Araujo, C. L., & Rohde, L. A. (2010). Prevalence of psychiatric disorders in a Brazilian birth cohort of 11-year-olds. *Social Psychiatry and Psychiatric Epidemiology*, *45*(1), 135-142. doi: 10.1007/s00127-009-0052-2

Bitsakou, P., Psychogiou, L., Thompson, M., & Sonuga-Barke, E. J. (2008). Inhibitory deficits in attention-deficit/hyperactivity disorder are independent of basic processing efficiency and IQ. *J Neural Transm*, *115*(2), 261-268. doi: 10.1007/s00702-007-0828-z

Brunner, M., Nagy, G., & Wilhelm, O. (2012). A tutorial on hierarchically structured constructs. *Journal of Personality*, *80*(4), 796-846. doi: 10.1111/j.1467-6494.2011.00749.x

Castellanos, F. X., Sonuga-Barke, E. J., Scheres, A., Di Martino, A., Hyde, C., & Walters, J. R. (2005). Varieties of attention-deficit/hyperactivity disorder-related intra-individual variability. *Biological Psychiatry*, *57*(11), 1416-1423. doi: 10.1016/j.biopsych.2004.12.005

Coghill, D. & Sonuga-Barke, E. J. (2012). Annual research review: Categories versus dimensions in the classification and conceptualisation of child and adolescent mental disorders-implications of recent empirical study. *Journal of Child Psychology and Psychiatry*, *53*(5), 469-489. doi: 10.1111/j.1469-7610.2011.02511.x

Cooper, L. D. & Balsis, S. (2009). When less is more: How fewer diagnostic criteria can indicate greater severity. *Psychological Assessment*, *21*(3), 285-293. doi: 10.1037/a0016698

Cooper, L. D., Balsis, S., & Zimmerman, M. (2010). Challenges associated with a polythetic diagnostic system: criteria combinations in the personality disorders. *Journal of Abnormal Psychology*, *119*(4), 886-895. doi: 10.1037/a0021078

Costello, E. J., Compton, S. N., Keeler, G., & Angold, A. (2003). Relationships between poverty and psychopathology: A natural experiment. *JAMA*, *290*(15), 2023-2029. doi: 10.1001/jama.290.15.2023

Dumenci, L., McConaughy, S. H., & Achenbach, T. M. (2004). A Hierarchical Three-Factor Model of Inattention-Hyperactivity-Impulsivity Derived From the AttentionProblems Syndrome of the Teacher's Report Form. *School Psychology Review*, *33*(2), 287-301.

Faraone, S. V., Kunwar, A., Adamson, J., & Biederman, J. (2009). Personality traits among ADHD adults: implications of late-onset and subthreshold diagnoses. *Psychol Medicine*, *39*(4), 685-693. doi: 10.1017/S0033291708003917

Fleitlich-Bilyk, B. & Goodman, R. (2004). Prevalence of child and adolescent psychiatric disorders in southeast Brazil. *Journal of the American Academy of Child & Adolescent Psychiatry*, *43*(6), 727-734. doi: 10.1097/01.chi.0000120021.14101.ca

Gibbins, C., Toplak, M. E., Flora, D. B., Weiss, M. D., & Tannock, R. (2011). Evidence for a General Factor Model of ADHD in Adults. *Journal of Attention Disorders*, *16*(8), 635-644.

Glaser, P. E., Thomas, T. C., Joyce, B. M., Castellanos, F. X., & Gerhardt, G. A. (2005). Differential effects of amphetamine isomers on dopamine release in the rat striatum and nucleus accumbens core. *Psychopharmacology*, *178*(2-3), 250-258. doi: 10.1007/s00213-004-2012-6

Goodman, R., Ford, T., Richards, H., Gatward, R., & Meltzer, H. (2000). The Development and Well-Being Assessment: Description and initial validation of an integrated assessment of child and adolescent psychopathology. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, *41*(5), 645-655.

Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry*, *177*, 534-539.

Hidalgo, R. B., Tupler, L. A., & Davidson, J. R. (2007). An effect-size analysis of pharmacologic treatments for generalized anxiety disorder. *Journal of Psychopharmacology*, *21*(8), 864-872. doi: 10.1177/0269881107076996

Hofmann, S. G. & Smits, J. A. (2008). Cognitive-behavioral therapy for adult anxiety disorders: A meta-analysis of randomized placebo-controlled trials. *The Journal of clinical psychiatry*, *69*(4), 621-632.

Hogan, A. M., Vargha-Khadem, F., Kirkham, F. J., & Baldeweg, T. (2005). Maturation of action monitoring from adolescence to adulthood: an ERP study. *Developmental Science*, *8*(6), 525-534. doi: 10.1111/j.1467-7687.2005.00444.x

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, *6*(1), 1-55.

Hyman, S. E. (2007). Can neuroscience be integrated into the DSM-V? *Nature Reviews Neuroscience*, *8*(9), 725-732. doi: 10.1038/nrn2218

Kapur, S., Phillips, A. G., & Insel, T. R. (2012). Why has it taken so long for biological psychiatry to develop clinical tests and what to do about it? *Molecular psychiatry*, *17*(12), 1174-1179. doi: 10.1038/mp.2012.105

Kendler, K. S. (2008). Explanatory models for psychiatric illness. *American Journal of Psychiatry*, *165*(6), 695-702. doi: 10.1176/appi.ajp.2008.07071061

Kraemer, H. C. (2007). DSM categories and dimensions in clinical and research contexts. *International Journal of Methods in Psychiatric Research*, *16*(1), S8-S15. doi: 10.1002/mpr.211

Kraemer, H. C., & Thiemann, S. (1989). A strategy to use soft data effectively in randomized controlled clinical trials. *Journal of Consulting and Clinical Psychology*, *57*(1), 148-154.

Kraemer, H. C. & Thiemann, S. (1987). *How Many Subjects? Statistical Power Analysis in Research*. Newbury Park: Sage Publications.

Krueger, R. F., Hicks, B. M., Patrick, C. J., Carlson, S. R., Iacono, W. G., & McGue, M. (2002). Etiologic connections among substance dependence, antisocial behavior, and personality: modeling the externalizing spectrum. *Journal of abnormal psychology*, *111*(3), 411-424.

Marco, R., Miranda, A., Schlotz, W., Melia, A., Mulligan, A., Muller, U., ... Sonuga-Barke, E. J. (2009). Delay and reward choice in ADHD: An experimental test of the role of delay aversion. *Neuropsychology*, *23*(3), 367-380. doi: 10.1037/a0014914

Martel, M. M., Roberts, B., Gremillion, M., Von Eye, A., & Nigg, J. T. (2011). External validation of bifactor model of ADHD: Explaining heterogeneity in psychiatric comorbidity, cognitive control, and personality trait profiles within DSM-IV ADHD. *Journal of abnormal child psychology*, *39*(8), 1111-1123. doi: 10.1007/s10802-011-9538-y

Martel, M. M., Von Eye, A., & Nigg, J. T. (2010). Revisiting the latent structure of ADHD: is there a 'g' factor? *Journal of child psychology and psychiatry and allied disciplines*, *51*(8), 905-914.

McLaughlin, K. A., & Nolen-Hoeksema, S. (2011). Rumination as a transdiagnostic factor in depression and anxiety. *Behaviour research and therapy*, *49*(3), 186-193. doi: 10.1016/j.brat.2010.12.006

McLoughlin, G., Albrecht, B., Banaschewski, T., Rothenberger, A., Brandeis, D., Asherson, P., & Kuntsi, J. (2009). Performance monitoring is altered in adult ADHD: a familial event-related potential investigation. *Neuropsychologia*, *47*(14), 3134-3142. doi: 10.1016/j.neuropsychologia.2009.07.013

Merport, A., Bober, S. L., Grose, A., & Recklitis, C. J. (2012). Can the distress thermometer (DT) identify significant psychological distress in long-term cancer survivors? A comparison with the Brief Symptom Inventory-18 (BSI-18). *Support Care in Cancer*, *20*(1), 195-198. doi: 10.1007/s00520-011-1269-7

Merport, A. & Recklitis, C. J. (2012). Does the brief symptom inventory-18 case rule apply in adult survivors of childhood cancer?: comparison with the symptom checklist-90. *Journal of pediatric psychology*, *37*(6), 650-659. doi: 10.1093/jpepsy/jss050

Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide*. Los Angeles: Muthén & Muthén.

Nigg, J. T., Willcutt, E. G., Doyle, A. E., & Sonuga-Barke, E. J. (2005). Causal heterogeneity in attention-deficit/hyperactivity disorder: do we need neuropsychologically impaired subtypes? *Biological Psychiatry*, *57*(11), 1224-1230. doi: 10.1016/j.biopsych.2004.08.025

Nolen-Hoeksema, S. & Watkins, E. R. (2011). A Heuristic for Developing Transdiagnostic Models of Psychopathology Explaining Multifinality and Divergent Trajectories. *Perspectives on Psychological Science*, *6*(6), 589-609. doi: 10.1177/1745691611419672

Olbert, C.M., Gala, G.J., & Tupler, L.A. (2014). Quantifying heterogeneity attributable to polythetic diagnostic criteria: theoretical framework and empirical application. *Journal of Abnormal Psychology*, *123*(2), 452-62.

Penninx, B. W., Nolen, W. A., Lamers, F., Zitman, F. G., Smit, J. H., Spinhoven, P., ... Beekman, A. T. (2011). Two-year course of depressive and anxiety disorders: results from the Netherlands Study of Depression and Anxiety (NESDA). *Journal of Affective Disorders*, *133*(1-2), 76-85. doi: 10.1016/j.jad.2011.03.027

Rescorla, L., Ivanova, M. Y., Achenbach, T. M., Begovac, I., Chahed, M., Drugli, M. B., ... Zhang, E. Y. (2012). International epidemiology of child and adolescent psychopathology ii: integration and applications of dimensional findings from 44 societies. *Journal of the American Academy of Child & Adolescent Psychiatry*, *51*(12), 1273-1283 e1278. doi: 10.1016/j.jaac.2012.09.012

Salum, G., Sergeant, J., Sonuga-Barke, E., Vandekerckhove, J., Gadelha, A., Pan, P., ... Rohde, L. A. (2012). Specificity of Basic Information Processing and Inhibitory Control in Attention Deficit/Hyperactivity Disorder (ADHD). *Psychological Medicine*, *44*(3), 617-631.

Salum, G., Gadelha, A., Pan, P. M., Moriyama, T. S., Graeff-Martins, A. S., Tamanaha, A. C., ... Rohde, L. A. (2015). High risk cohort study for psychiatric disorders in childhood: rationale, design, methods and preliminary results. *International Journal of Methods in Psychiatric Research*, *24*(1), 58-73. doi: 10.1002/mpr.1459

Sanislow, C. A., Pine, D. S., Quinn, K. J., Kozak, M. J., Garvey, M. A., Heinssen, R. K., ... Cuthbert, B. N. (2010). Developing constructs for psychopathology research: research domain criteria. *Journal of abnormal psychology*, *119*(4), 631-639. doi: 10.1037/a0020909

Sonuga-Barke, E. (2013). The challenge of mapping diagnostic categories onto developmental pathophysiology: DSM-6 anyone?. *Journal of Child Psychology and Psychiatry*, *54*(6), 601-602. doi: 10.1111/jcpp.12096

Sonuga-Barke, E. J. (2010). Disambiguating inhibitory dysfunction in attention-deficit/hyperactivity disorder: toward the decomposition of developmental brain phenotypes. *Biological Psychiatry*, *67*(7), 599-601. doi: 10.1016/j.biopsych.2010.01.017

Sonuga-Barke, E. J. (2005). Causal models of attention-deficit/hyperactivity disorder: from common simple deficits to multiple developmental pathways. *Biological Psychiatry*, *57*(11), 1231-1238. doi: 10.1016/j.biopsych.2004.09.008

Telzer, E. H., Mogg, K., Bradley, B. P., Mai, X., Ernst, M., Pine, D. S., & Monk, C. S. (2008). Relationship between trait anxiety, prefrontal cortex, and attention bias to angry faces in children and adolescents. *Biological Psychology*, *79*(2), 216-222. doi: 10.1016/j.biopsycho.2008.05.004

Toplak, M. E., Pitch, A., Flora, D. B., Iwenofu, L., Ghelani, K., Jain, U., & Tannock, R. (2009). The unity and diversity of inattention and hyperactivity/impulsivity in ADHD: Evidence for a general factor with separable dimensions. *Journal of Abnormal Child Psychology*, *37*(8), 1137-1150.

Toplak, M. E. & Tannock, R. (2005). Tapping and anticipation performance in attention deficit hyperactivity disorder. *Perceptual and Motor Skills*, *100*(3), 659-675.

Toplak, M. E., Rucklidge, J. J., Hetherington, R., John, S. C., & Tannock, R. (2003). Time perception deficits in attention-deficit/ hyperactivity disorder and comorbid reading difficulties in child and adolescent samples. *Journal of Child Psychology and Psychiatry*, *44*(6), 888-903.

Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British journal of psychology*, *95*(1), 57-79. doi: 10.1348/000712604322779460

Ward, A. J. (1974). Childhood psychopathology. A natural experiment in etiology. *Journal of the American Academy of Child & Adolescent Psychiatry*, *13*(1), 153-165.

Wechsler, D. (2002). *WISC-III: Escala de Inteligência Wechsler para Crianças: Manual*. Sao Paulo: Casa do Psicólogo.

Woerner, W., Fleitlich-Bilyk, B., Martinussen, R., Fletcher, J., Cucchiaro, G., Dalgalarrondo, P., ... Tannock, R. (2004). The Strengths and Difficulties Questionnaire overseas: evaluations and applications of the SDQ beyond Europe. *European child & adolescent psychiatry*, *13*(2), ii47-ii54. doi: 10.1007/s00787-004-2008-0