Title

Longitudinal Measurement Invariance of the ASEBA Youth/Adult Self-Reports Across the Transition From Adolescence to Adulthood.

Authors

Moriarity, Daniel P
Mac Giollabhui, Naoise
Cardoso Melo, Dener
et al.

Publication Date

2024-04-18

DOI

10.1177/10731911241245875

Longitudinal Measurement Invariance of the ASEBA Youth/Adult Self-Reports Across the Transition from Adolescence to Adulthood

Daniel P. Moriarity, Ph.D.[1]; Naoise Mac Giollabhui, Ph.D.[2]; Dener Cardoso Melo[3], Catharina Hartman, Ph.D.[3]

[1]Department of Psychiatry and Biobehavioral Sciences, University of California, Los Angeles

[2]Department of Psychiatry, Massachusetts General Hospital/Harvard Medical School

[3]Department of Psychiatry, University of Groningen

Correspondence concerning this article should be addressed to Daniel P. Moriarity, E-mail: dmoriarity@mednet.ucla.edu

Funding: Daniel P. Moriarity was supported by National Research Service Award F32 MH130149 and grant #OPR21101 from the California Governor's Office of Planning and Research/California Initiative to Advance Precision Medicine.

Abstract

The ability to quantify within-person changes in mental health is central to the mission of clinical psychology. Typically, this is done using total or mean scores on symptom measures; however, this approach assumes that measures quantify the same construct, the same way, each time the measure is completed. Without this quality, termed longitudinal measurement invariance, an observed difference between timepoints might be partially attributable to changing measurement properties rather than changes in comparable symptom measurements. This concern is amplified in research using different forms of a measure across developmental periods due to potential differences in reporting styles, item-wording, and developmental context. This study provides the strongest support for the longitudinal measurement invariance of the Anxiety Scale, Depression/Affective Problems: Cognitive Subscale, and the ADHD Scale; moderate support for the Depression/Affective Problems Scale and the Somatic Scale, and poor support for the Somatic Symptoms Subscale of the Dutch Achenbach System of Empirically Based Assessment Youth Self-Report and Adult Self-Report in a sample of 1,309 individuals (N=1,090 population-based, N=219 clinic-based/referred to an outpatient clinic before age 11 years) across 6 waves of data (mean ages= 11 years at Wave 1 and 26 years at Wave 6).

Keywords: Measurement; measurement invariance; longitudinal; adolescent; adult; developmental; psychopathology

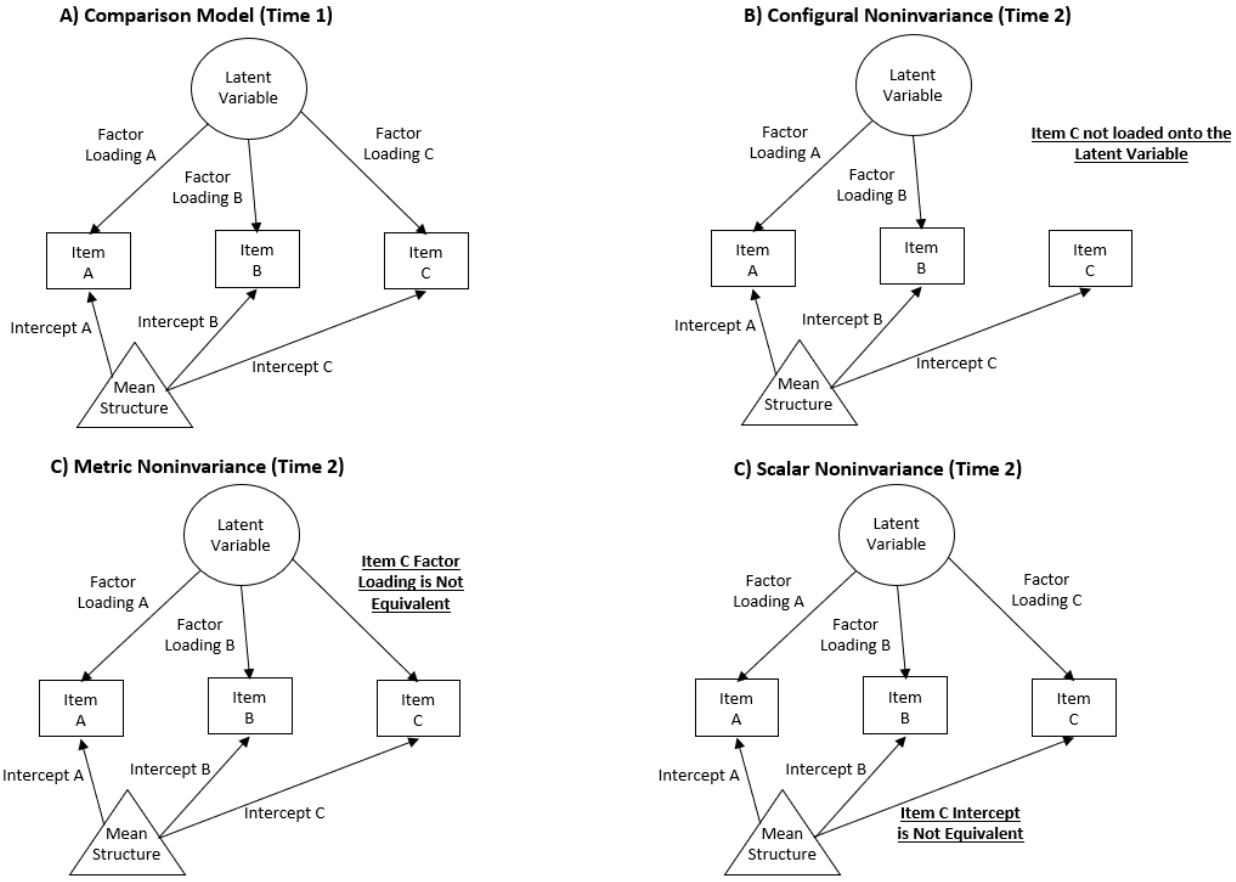
Introduction

Central to psychology's mission to examine mechanisms and treatments for psychopathology is the ability to measure change in symptoms over time. Studies typically quantify change via increases or decreases in total scores on self-report measures; however, this assumes that the total score quantifies symptoms the same way at each time point. For example, change score approaches assume that a score of 13 at baseline is comparable to a score of 13 at post-treatment (i.e., is largely compromised by the same symptom profile and identical factor structure). The ability for a measure to quantify the same construct, the same way, across different times points is referred to as *longitudinal measurement invariance*.

The three most commonly assessed types of measurement invariance are configural, metric (i.e., "weak"), and scalar (i.e., "strong"; for a more detailed review on measurement invariance, see Putnick & Bornstein, 2016). We provide a brief conceptual overview here, with more technical information available in the Methods section. Configural invariance refers to the equivalence of model form (i.e., which items load onto which latent constructs). Metric invariance refers to equality of factor loadings. Scalar invariance refers to equality of item intercepts (i.e., the average response to an item when the associated latent score is zero). For visualization of these measurement invariance types see Figure 1. Without measurement invariance, a questionnaire does not measure a construct the same way across different time points, precluding mean comparison of scores to evaluate change in the underlying construct. A non-psychological example would be to consider if you weighed yourself on a scale at home today and re-weighed yourself using the same scale from the moon tomorrow. The subject and measurement tool are constant, but the underlying measurement properties change overtime in a way that invalidates direct comparison of the two measurements.

Figure 1.

Visualization of Measurement Invariance Types Illustrated by the Measurement Properties of Item C



Note. Focal differences associated with the specified type of invariance are highlighted by a bolded and underlined statement.

This concern is amplified when different measures are used to assess the same construct at different time points in a longitudinal study. For example, the TRacking Adolescents' Individual Lives Survey (TRAILS) is a large prospective cohort following 11-year-olds and re-assessing them every 2-3 years. At the onset of TRAILS, participants completed the Youth Self-Report (YSR; Achenbach, 1991) from the Achenbach System of Empirically Based Assessment

(ASEBA, a comprehensive set of assessments designed to assess adaptive and maladaptive functioning) to assess youth psychological health. The YSR includes 112 items and has been disaggregated into several different factor structures based on researcher/clinician needs. The TRAILS data documentation provides two strategies: syndromes (comprised of 11 scales) or DSM-oriented scales (comprised of 6 scales). Given that TRAILS waves have been completed during childhood, adolescence, and adulthood, all original participants "aged out" of the YSR by Wave 4 and shifted to completing the more developmentally appropriate Adult Self-Report (ASR; Achebach & Rescorla, 2003) of the ASEBA system. However, this change in measures, in addition to potential developmental changes in both symptom-reporting by increasingly mature individuals and age-related differences in the latent construct, could be a source of measurement *non*-invariance.

Given the size of the ASR/YSR, the few investigations into their longitudinal measurement invariance have tested select subscales to maintain computational brevity. For example, Barzeva, Meeus, & Oldehinkel (2019) found that the social withdrawal scales were measurement invariant in people measured four times in the TRAILS study using both the YSR and ASR. Research from the Netherlands Twin Registry (Abdellaoui et al., 2012) found that the ASR Thought Problems Subscale was measurement invariant across three age groups (12–18, 19–27, and 28–59 years). However, only one time point per participant was used in this analysis so longitudinal measurement invariance *within people* was not tested, just *between different age groups*, thus these results only measure between-person differences (e.g., potentially cohort effects) instead of testing measurement invariance across time within individuals. The only study we found testing age-related measurement invariance of the entire eight factor model used a similar age-stratification technique—supporting measurement invariance of the ASR between

two age groups (18-35 vs. 36-59 years; Guerrero et al., 2020). Thus, while preliminary evidence supports longitudinal measurement invariance of the YSR/ASR to the extent they have been investigated, much more work is needed. Specifically, 1) additional subscales of the YSR/ASR must be examined (ideally in the same sample), 2) longitudinal measurement noninvariance (vs. age group measurement invariance) must be evaluated to facilitate change-in-symptom research, and 3) measurement invariance of symptoms in individuals transitioning between the YSR and ASR should be investigated to determine the appropriateness of using both in longitudinal research.

The theoretical and clinical utility of a large, longitudinal dataset such as TRAILS for garnering developmental insight through adolescence and across the transition to adulthood is immense, if the foundational psychometric work is done to inform future longitudinal modeling and data collection. Further strengthening the utility of this dataset for this purpose, TRAILS includes two subsamples: one population-based sample and a second clinical sample featuring youth referred to an outpatient clinic before age 11 years—facilitating comparison of measurement properties in both community and clinical samples. Given its widespread use in TRAILS and other studies, the present investigation evaluated the longitudinal measurement invariance of DSM-IV subscales that are shared between the YSR and the ASR (i.e., Depression/Affective Problems, Anxiety, Attention Deficit Hyperactivity Disorder (ADHD), and Somatic) using six waves of TRAILS Data. Further, this study also tested the longitudinal measurement invariance of two constituent subscales of the Depression Scale/Affective Problems (cognitive and somatic symptoms) previously identified using TRAILS data (Bosch et al., 2009).

Methods

Participants

Data were drawn from the TRacking Adolescents' Individual Lives Survey (TRAILS), a prospective cohort study examining psychosocial development and mental health in youth. Adolescents aged 11 years were recruited and invited to attend regular follow-up assessments every 2-3 years. Two separate cohorts were followed by TRAILS—one population-based and another clinic-based (Huisman et al., 2008; Oldehinkel et al., 2015). Adolescents in the population-based cohort were recruited from 135 schools in five municipalities in the north of The Netherlands, including both urban and rural areas. Eligible participants were required to be enrolled in primary school, and of 2,935 youth who met this criterion, 2,230 (76%) provided informed consent from both parent and child to participate. The clinic-based cohort consisted of children referred to a psychiatric outpatient clinic before the age of 11 for a variety of psychiatric and behavioral problems. The current study utilized data from 1,309 participants (N=1,090 population-based, N=219 clinic-based) from Waves 1 – 6 (see Table 1 for descriptives and below for data cleaning details).

Procedures

In this study, symptoms were measured at each assessment (Waves 1 – 6) using either the Youth Self-Report or the Adult Self-Report (determined by participant age at the time of assessment). Children started the study with the Youth Self-Report at approximately 11 years old and shifted to the Adult Self-Report when they turned 16 years old (Wave 4).

144

145

146

147

148 **Table 1. Descriptive Statistics for Variables of Interest in Combined Population + Clinic-**
149 **based Cohorts (N=1309)**

	Mean	*SD*	Range
Wave 1			
% Female	56.7%		
Age (years)	11.09	.55	10.01—12.54
SES (z)	.16	.75	-1.73—1.73
Depression/Affective Problems	4.06	3.21	0—18
Somatic Subscale	2.29	1.94	0—11
Cognitive Subscale	1.77	1.85	0—11
Anxiety	2.28	1.88	0—10
ADHD	4.36	2.56	0—13
Somatic	3.33	2.32	0—11
Wave 2			
Age (years)	13.39	.59	11.58—14.93
Depression/Affective Problems	3.80	3.38	0—24
Somatic Subscale	2.24	2.00	0—10
Cognitive Subscale	1.56	1.93	0—14
Anxiety	2.40	1.93	0—10
ADHD	4.77	2.67	0—14
Somatic	2.23	2.02	0—9
Wave 3			
Age (years)	16.15	.66	14.42—18.33
Depression/Affective Problems	3.97	3.58	0—24
Somatic Subscale	2.55	2.18	0—12
Cognitive Subscale	1.42	1.94	0—14
Anxiety	2.13	1.90	0—11
ADHD	4.88	2.77	0—14
Somatic	1.82	1.94	0—10
Wave 4			
Age (years)	18.97	.59	17.98—21.06
Depression/Affective Problems	4.40	4.40	0—24
Somatic Subscale	2.10	2.13	0—10
Cognitive Subscale	1.37	1.89	0—12
Anxiety	2.83	2.54	0—14
ADHD	5.91	4.45	0—21
Somatic	1.13	2.09	0—14
Wave 5			
Age (years)	22.13	.66	20.74—24.10
Depression/Affective Problems	4.63	4.37	0—26
Somatic Subscale	2.33	2.10	0—9
Cognitive Subscale	1.41	1.83	0—13
Anxiety	2.94	2.59	0—13
ADHD	5.53	4.27	0—22
Somatic	1.89	2.19	0—16
Wave 6			
Age (years)	25.66	.63	24.35—27.82
Depression/Affective Problems	5.60	4.97	0—26
Somatic Subscale	2.69	2.27	0—10
Cognitive Subscale	1.64	2.08	0—12
Anxiety	3.61	2.94	0—14
ADHD	5.79	4.31	0—25
Somatic	2.19	2.41	0—15

150 Note: z= z-standardized on whole sample (not analytic sample). ADHD = Attention Deficit Hyperactivity Disorder.

Missing Data Analyses

Participants were removed if they were missing 100% of symptom data at any time point (removing these participants solved some issues with model convergence). Individual analytic datasets were created to maximize sample size by only removing participants missing data on a given DSM-IV scale. This resulted in identical analytic datasets (N=1,090 population-based, 219 clinic-based) except for the Somatic Scale datasets, which were slightly smaller (N=1,074 population-based, 217 clinic-based). Because of the negligible difference in samples, descriptive statistics for the analytic dataset corresponding to the non-Somatic Scales.

T-tests and chi-squared tests examined whether the analytic sample (N=1,309, 83% population cohort, 17% clinic-based cohort) differed significantly from the entire baseline sample (N=2772, 80% population cohort, 20% clinic-based cohort) based on reported age, gender, socioeconomic status, and depression symptoms. The mean level of socioeconomic status (indexed by a composite of z-standardized variables, see below for more information) in the analytic sample was higher than in the excluded sample [$t(2762) = 12.88$, $p < .001$; mean difference = .49 standard deviations]. Further, the analytic sample was younger [$t(2770) = 2.09$, $p = .035$; mean difference = .04 years], had higher anxiety symptoms [$t(2675.44) = 3.598$, $p < .001$; standardized mean difference = .14], and higher somatic symptoms [$t(2692) = 2.687$, $p = .007$; mean difference = .10]. The analytic sample also differed from the excluded sample in the proportion of females that were retained in the sample, $\chi^2(1, 2772) = 84.36,\ p < .001$, with fewer males present in the analytic sample (Standardized Residual = -4.6). No differences between the analytic and entire sample in baseline depressive symptoms were reported for the Depression/Affective Problems Scale [$t(2716) = 1.453$, $p = .146$; standardized mean difference = .06], Cognitive Symptoms Subscale [$t(2715) = 1.926$, $p = .054$; standardized mean difference =

.07], Somatic Symptoms Subscale [$t(2714) = .584$, $p = .559$; standardized mean difference = .02], or ADHD Scale [$t(2717) = .278$, $p = .781$; standardized mean difference = .01] of the YSR (note degrees of freedom for symptom measures are slightly different due to different degrees of item-level missingness relative to the size of the scale in question).

Measures

Symptoms

During Waves 1 – 3, symptoms were measured using the Youth Self-Report (YSR; Achenbach, 1991). During Waves 4 – 6, symptoms were measured using the Adult Self-Report (ASR) were used during (Achenbach & Rescorla, 2003). Item wording can be found in Table 2. All items were answered using a 3-point Likert scale (0-2), with higher endorsements indicating more severe symptoms. For the descriptive (Table 1) and missing data analyses (described above) involving symptom summary statistics, scores were determined by taking the average of the items responded to in the scale in interest and then multiplying by the total number of items in the scale. Symptom summary scores were not calculated for observations with < 80% of item-level data.

Depression/Affective Problems

The YSR Depression/Affective Problems Scale had 13 items (split into a seven item Cognitive Subscale and a six item Somatic Subscale based on item content in previous TRAILS studies (Bosch et al., 2009)). The ASR Depression/Affective Problems Scale had 14 items (split into a seven item Cognitive Subscale and a five item Somatic Subscale based on item content in previous TRAILS studies (Bosch et al., 2009), refer to Table 2 to compare items and wording between measures). The Ω reliability coefficient at Waves 1 and 4 (first wave using the ASR) were .74 and .86 (respectively) for the Depression/Affective Problems Scale, .58 and .72

(respectively) for the Somatic Symptoms Subscale, and .69 and .80 (respectively) for the Cognitive Symptoms Subscale.

Anxiety

The YSR Anxiety Scale had 6 items and the ASR Anxiety Scale had 7 items (refer to Table 2 to compare items and wording between measures). The Ω reliability coefficient at Waves 1 and 4 (first wave using the ASR) were .63 and .78, respectively.

Attention Deficit Hyperactivity Disorder (ADHD)

The YSR ADHD Scale had 7 items and the ASR ADHD Scale had 13 items (refer to Table 2 to compare items and wording between measures). The Ω reliability coefficient at Waves 1 and 4 (first wave using the ASR) were .71 and .85, respectively.

Somatic

The YSR Somatic Scale had 7 items and the ASR Somatic Scale had 9 items (refer to Table 2 to compare items and wording between measures). The Ω reliability coefficient at Waves 1 and 4 (first wave using the ASR) were .71 and .83, respectively.

Sociodemographic variables

Participant sex was assessed at Wave 1, when participants could respond that they identified as 'Female', which was scored as '0' or 'Male', which was scored as '1'. Age was assessed at all assessments. Socio-economic status (SES) was measured at Wave 1 and Wave 4. SES was estimated using five indicators: family income, maternal educational level, paternal educational level, maternal occupational level and paternal occupational level using the International Standard Classification of Occupations (Ganzeboom & Treiman, 1996). A composite measure of SES was calculated for the TRAILS cohort based on five z-transformed

indicators (which has been consistently used in TRAILS), with higher values representing higher SES and a one unit difference representing one standard deviation in difference (Jonker et al., 2017). The composite measure of SES was assessed at Wave 1 and Wave 4 and were highly correlated (r = .86) with one another, as previously reported (Mac Giollabhui & Hartman, 2022).

Table 2. The Cognitive and Somatic Subscale Item Wording

Item		Scale
English	Dutch	
There is very little that I like	Er is heel weinig wat ik leuk vind	Depression/Affective Problems: Cognitive
I cry a lot	Ik huil veel	Depression/Affective Problems: Cognitive
I intentionally try to injure myself or attempt suicide	Ik probeer mezelf opzettelijk te verwonden of doe zelfmoordpogingen	Depression/Affective Problems: Cognitive
I don't eat as well as I should	Ik eet niet zo goed als zou moeten	Depression/Affective Problems: Somatic
I feel worthless or inferior	Ik voel me waardeloos of minderwaardig	Depression/Affective Problems: Cognitive
I feel too much guilt	Ik heb te veel last van schuldgevoel	Depression/Affective Problems: Cognitive
I feel overtired for no apparent reason	Ik voel me oververmoeid zonder duidelijke reden	Depression/Affective Problems: Somatic
I sleep more than most of my peers during the day and/or at night	Ik slaap meer dan de meeste van mijn leeftijdgenoten overdag en/of 's nachts (geef aan):	Depression/Affective Problems: Somatic
I'm thinking about ending my life	Ik denk erover een eind aan mijn leven te maken	Depression/Affective Problems: Cognitive
I have trouble sleeping	Ik heb problemen met slapen (geef aan)	Depression/Affective Problems: Somatic
I don't have much energy	Ik heb niet veel energie	Depression/Affective Problems: Somatic
I am unhappy, sad or depressed	Ik ben ongelukkig, verdrietig of gedeprimeerd	Depression/Affective Problems: Cognitive
*I sleep less than most of my peers	*Ik slaap minder dan de meeste van mijn leeftijdgenoten	*Depression/Affective Problems: Somatic
⸸I have trouble making decisions	⸸Ik heb moeite om beslissingen te nemen	⸸Depression/Affective Problems
⸸I feel like I can't succeed	⸸Ik heb het gevoel dat ik niet kan slagen	⸸Depression/Affective Problems
I am afraid of certain animals, situations or places	Ik ben bang voor bepaalde dieren, situaties of plaatsen	Anxiety
I am nervous or tense	Ik ben nerveus of gespannen	Anxiety
I'm too anxious or scared	Ik ben te angstig of bang	Anxiety
I often worry	Ik maak me vaak zorgen	Anxiety
*I'm too dependent on adults	*Ik ben te afhankelijk van volwassenen	*Anxiety
*I'm afraid to go to school	Ik ben bang om naar school te gaan	*Anxiety
⸸Palpitations	⸸Hartkloppingen	⸸Anxiety
⸸I'm worried about my family or relatives	⸸Ik maak me zorgen over mijn familie of gezin	⸸Anxiety
⸸I worry about my future	⸸Ik maak me zorgen over mijn toekomst	⸸Anxiety
Pains (no stomachache or headache)	Pijnen (geen buikpijn of hoofdpijn)	Somatic
Headache	Hoofdpijn	Somatic
Nausea	Misselijkheid	Somatic
Eye problems (for which glasses or lenses do not help)	Oogproblemen (waarvoor een bril of lenzen niet helpen)	Somatic
Rash or other skin problems	Huiduitslag of andere huidproblemen	Somatic
Stomach ache	Buikpijn	Somatic
Vomit	Overgeven	Somatic
⸸I feel dizzy or light-headed	⸸Ik voel me duizelig of licht in mijn hoofd	⸸Somatic
⸸Dead feeling or tingling in body parts	⸸Dood gevoel of tintelingen in lichaamsdelen	⸸Somatic
I don't finish things I need to do	Ik maak dingen die ik moet doen niet af	ADHD
I have difficulty concentrating or keeping my attention on something for long periods of time	Ik heb moeite om me te concentreren, of om lang mijn aandacht ergens bij te houden	ADHD
I am impulsive or do things without thinking	Ik ben impulsief of doe dingen zonder er bij na te denken	ADHD
I have trouble sitting still	Ik heb moeite om stil te zitten	ADHD
*I am inattentive or easily distracted	*Ik ben onoplettend of makkelijk afgeleid	*ADHD
*I talk too much	*Ik praat te veel	*ADHD
*I make more noise than other boys or girls	*Ik maak meer lawaai dan andere jongens of meisjes	*ADHD
⸸I'm too forgetful	⸸Ik ben te vergeetachtig	⸸ADHD
⸸I often accidentally hurt myself, often get injured accidentally	⸸Ik bezeer me vaak per ongeluk, raak vaak per ongeluk gewond	⸸ADHD
⸸I'm not doing well at my job	⸸Ik doe het niet goed op mijn werk	⸸ADHD
⸸I throw myself into things without thinking about the risks	⸸Ik stort mij in dingen zonder over de risicos na te denken	⸸ADHD
⸸People think I'm chaotic	⸸Mensen denken dat ik chaotisch ben	⸸ADHD
⸸I often lose things	⸸Ik ben vaak dingen kwijt	⸸ADHD
⸸I feel restless	⸸Ik voel me rusteloos	⸸ADHD
⸸I'm too impatient	⸸Ik ben te ongeduldig	⸸ADHD
⸸I don't pay much attention to details	⸸Ik let niet goed op details	⸸ADHD

Note: Unless otherwise noted the Dutch wording reflects the ASR version of the items; wording for some YSR items is slightly different to be more developmentally appropriate; the English version of the items is translated from the Dutch version and may differ slightly from the wording in the original ASR or YSR. ADHD = Attention Deficit Hyperactivity Disorder. *This item was only available in the YSR (Waves 1-3). ⸸ This item was only available in the ASR (Waves 4-6)

Statistical Methods

All analyses were conducted in R Version 4.2.2 (R Core Team, 2013). Analyses were conducted in lavaan (Rosseel, 2012). Template code was adapted from https://longitudinalresearchinstitute.com/tutorials/item-factor-analysis-measurement-invariance-2nd-order-growth-model-ecls-k/. The analytic code and output is available as supplemental material (https://osf.io/hbafn/?view_only=65d1c791a5b74fe7ab71ee0eca56ecdc). Data are not publicly available due to privacy regulations but can be requested for replication, unconditionally and free-of-charge, from TRAILS at www.trails.nl.

All models were estimated with a theta parameterization, pairwise deletion for missing data, a combination of diagonally weighted least squares (DWLS; for parameters) and weighted least square mean and variance adjusted (WLSMV; for robust standard errors) estimation, and nonlinear minimization subject to box constraints (NMLINB) optimization. The first factor loading for each factor was constrained to 1 for identification. Variances and covariances were estimated freely. Latent variable means were constrained to zero and propensity variances for items were constrained to 1 unless otherwise specified below. Items that were in one version of a scale but not the other were still modeled to maximize fidelity to clinical use of this measure; however, items that only appeared in one version of the measure were only constrained to equality in different waves of that particular measure. For example, the item "I sleep less than most of my peers" was only assessed in the YSR (i.e., Waves 1-3). As such, the specific equality constraints for testing measurement invariance in this item were only specified in Waves 1-3.

As described in the introduction and shown in Figure 1, three types of measurement invariance were tested: configural, metric, and scalar (listed here with increasing stringency). The configural invariance model only imposes the constraint that each item loads onto its

specified factor at each time point. The metric invariance (i.e., "weak") model adds constraints that the factor loadings of an item on its factor are equivalent across timepoints. Finally, the scalar invariance (i.e., "strong") model incorporates the constraint that item intercepts (in this case thresholds between item-response options) be equivalent across timepoints while latent variable means are allowed to vary. Thus, while the first timepoint for each latent variable mean is set to zero (identical to configural, metric, and scalar invariance models for scaling reasons), latent variable means are estimated freely for later timepoints. There were no additional residual variances because item responses were modeled using thresholds (given ordinal rather than continuous response scales); thus, the scalar invariance model tests both strong and strict invariance. Items that had response options that were not all endorsed at one or more timepoints were dichotomized ("0" = "0" and "1-2" = "1") at all timepoints to facilitate comparison of item thresholds in that particular sample. Models with estimation/convergence issues applied the same dichotomization scheme to any items with low endorsement (i.e., <5% in any response option at any timepoint) in that particular sample.

Chi-square tests of fit are reported but were not heavily considered regarding conclusions due to over-sensitivity to negligible differences in large sample sizes Acceptable model fit criteria were a comparative fit index [CFI] ≥ .95, root mean square-error of approximation [RMSEA] ≤ .06, and standardized root-mean-square residual [SRMR] ≤ .08 (Hu et al., 1999). Metric invariance was evaluated based on the following cut-off criteria in change of model fit comparing the metric invariance model to the configural invariance model: -.010 change in CFI, .015 change in RMSEA, and .030 change in SRMR (Chen, 2007). Scalar invariance had identical criteria when comparing the scalar invariance to the metric invariance models except the cut-off for SRMR was reduced to .010 (per Chen, 2007). (Chen, 2007). It is worth noting that these cut-

offs were established using continuous data. To our knowledge cut-offs have not yet been established using ordinal data and some estimators for ordinal data (including the DWLS/WLSMV used here) have a tendency not to discover misfit (Xia & Yang, 2019). As such, results are preliminary and would benefit from re-analysis when appropriate cut-offs for ordinal data are established.

Results

Tables 3 and 4 include details about the fit of each model in the population-based sample and the clinic-based sample (respectively). All factor loadings and item thresholds for the models can be found in the supplemental material (https://osf.io/hbafn/?view_only=65d1c791a5b74fe7ab71ee0eca56ecdc).

Population-Based Sample

Depression Symptoms/Affective Problems Scale

The most severe response option for the self-injury item was not endorsed at all time points; therefore, the item was dichotomized. All three interpreted fit indices supported acceptable model fit for the configural invariance model (CFI = .962, RMSEA = .040, and SRMR = .079). Both CFI and RMSEA supported acceptable global model for the metric model (CFI = .952, RMSEA = .044, SRMR = .085). Only RMSEA supported acceptable global model for the scalar model (CFI = .941, RMSEA = .048, SRMR = .085). All comparisons of model fit supported metric invariance (ΔCFI = -.010, ΔRMSEA = .004, ΔSRMR = .006). Only two out of three comparisons of model fit, specifically ΔRMSEA and ΔSRMR, supported scalar invariance (ΔCFI = -.011, ΔRMSEA = .004, ΔSRMR = .000).

Cognitive Symptoms Subscale. The most severe response option for the self-injury item was not endorsed at all time points; therefore, the item was dichotomized. Two out of three fit indices (CFI and RMSEA) suggested acceptable model fit for all three invariance models (configural: CFI = .977, RMSEA = .037, SRMR = .084; metric: CFI = .973, RMSEA = .039, SRMR = .088; scalar: CFI = .967, RMSEA = .041, SRMR = .088). All three comparisons of model fit supported metric and scalar invariance of the Cognitive Symptoms Subscale (metric: ΔCFI = -.004, ΔRMSEA = .002, ΔSRMR = .004; scalar: ΔCFI = -.006, ΔRMSEA = .002, ΔSRMR = .000).

Somatic Symptoms Subscale. Two out of three fit indices (RMSEA and SRMR) suggested acceptable model fit for the configural invariance model (configural: CFI = .940, RMSEA = .056, SRMR = .080). All model fit indices indicated unacceptable model fit for the metric and scalar models (metric: CFI = .914, RMSEA = .066, SRMR = .093; scalar: CFI = .885, RMSEA = .073, SRMR = .094). Both ΔRMSEA and ΔSRMR supported metric and scalar invariance of the Somatic Symptoms Subscale (metric: ΔRMSEA = .010, ΔSRMR = .013; scalar: ΔRMSEA = .007, ΔSRMR = .001). ΔCFI was the only comparison of model fit that did not support metric and scalar invariance (metric ΔCFI = -.026, scalar ΔCFI = -.029).

Anxiety Scale

All three interpreted fit indices supported acceptable global model fit for all three invariance models (configural: CFI = .974, RMSEA = .040, and SRMR = .067; metric: CFI = .972, RMSEA = .041, and SRMR = .069; scalar: CFI = .959, RMSEA = .048, and SRMR = .069). All comparisons of model fit supported metric invariance (ΔCFI = -.002, ΔRMSEA = .001, ΔSRMR = .002). Only two out of three comparisons of model fit, specifically ΔRMSEA and ΔSRMR, supported scalar invariance (ΔCFI = -.013, ΔRMSEA = .007, ΔSRMR = .000).

ADHD Scale

Two of the three interpreted fit indices (RMSEA and SRMR) suggested acceptable global model fit for the configural, metric, and scalar invariance models (configural: CFI = .947, RMSEA = .053, SRMR = .073; metric: CFI = .939, RMSEA = .056, SRMR = .077; scalar: CFI = .932, RMSEA = .057, SRMR = .077). All three comparisons of model fit supported metric and scalar invariance (metric: ΔCFI = -.008, ΔRMSEA = .003, ΔSRMR = .004; scalar: ΔCFI = -.006, ΔRMSEA = .001, ΔSRMR = .000).

Somatic Scale

Two of the three interpreted fit indices (CFI and RMSEA) supported acceptable global model fit for the configural and metric invariance models (configural: CFI = .963, RMSEA = .039, and SRMR = .091; metric: CFI = .956, RMSEA = .041, and SRMR = .096). Only one of the three interpreted fit indices (RMSEA) supported acceptable global model fit for the scalar invariance models (CFI = .940, RMSEA = .047, and SRMR = .097). All comparisons of model fit supported metric invariance (ΔCFI = -.007, ΔRMSEA = .002, ΔSRMR = .005). Only two out of three comparisons of model fit, specifically ΔRMSEA and ΔSRMR, supported scalar invariance (ΔCFI = -.016, ΔRMSEA = .006, ΔSRMR = .001).

Table 3. Model Fit: Population-based Sample

	df	χ^2 *p*	CFI	ΔCFI	RMSEA	ΔRMSEA	90% CI RMSEA	SRMR	ΔSRMR
Depression/Affective Problems (N=1,090)									
Configural	3144	8660.334 *p* < .001	.962		.040		.039—.044	.079	
Metric	3205	10072.682 *p* < .001	.952	-.010	.044	.004	.043—.045	.085	.006
Scalar	3327	11821.990 *p* < .001	.941	-.011	.048	.004	.047—.049	.085	.000
Depression/Affective Problems: Cognitive Subscale (N=1,090)									
Configural	804	1978.987 *p* < .001	.977		.037		.035—.039	.084	
Metric	834	2201.443 *p* < .001	.973	-.004	.039	.002	.037—.041	.088	.004
Scalar	894	2542.368 *p* < .001	.967	-.006	.041	.002	.039—.043	.088	.000
Depression/Affective Problems: Somatic Subscale (N=1,090)									
Configural	480	2137.515 *p* < .001	.940		.056		.054—.059	.080	
Metric	502	2876.218 *p* < .001	.914	-.026	.066	.010	.064—.068	.093	.013
Scalar	551	3711.493 *p* < .001	.885	-.029	.073	.007	.070—.075	.094	.001
Anxiety (N=1,090)									
Configural	687	1883.328 *p* < .001	.974		.040		.038—.042	.067	
Metric	712	2009.606 *p* < .001	.972	-.002	.041	.001	.039—.043	.069	.002
Scalar	767	2672.602 *p* < .001	.959	-.013	.048	.007	.046—.050	.069	.000
Attention Deficit Hyperactivity Disorder (N=1,090)									
Configural	1695	6799.622 *p* < .001	.947		.053		.051—.054	.073	
Metric	1734	7599.285 *p* < .001	.939	-.008	.056	.003	.054—.057	.077	.004
Scalar	1816	8300.253 *p* < .001	.932	-.006	.057	.001	.056—.059	.077	.000
Somatic (N=1,074)									
Configural	1065	2765.502 *p* < .001	.963		.039		.037—.040	.091	
Metric	1099	3104.100 *p* < .001	.956	-.007	.041	.002	.040—.043	.096	.005
Scalar	1172	3908.835 *p* < .001	.940	-.016	.047	.006	.045—.048	.097	.001

Note: Δ = change between current model and previous model (i.e., change between configural and metric and change between metric and scalar). *df* = degrees of freedom, CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation, CI = Confidence Interval, SRMS = Standardized Root Mean Square Residual.

Clinic-based Sample

Depression Symptoms/Affective Problems Scale

The most severe response option for both the self-injury and suicidal ideation items were not endorsed at all time points; therefore, these items were dichotomized.. The initial estimation of this model resulted in a nonpositive definite covariance matrix. Estimation was successfully re-attempted after dichotomizing the items featuring at least one time point where a response option was endorsed at <5% (7 items, see Supplemental code for specific items). Two of the three fit indices (CFI and RMSEA) suggested acceptable model fit for the configural invariance model (CFI = .960, RMSEA = .035, and SRMR = .129). Only RMSEA suggested acceptable model fit for metric and scalar models (metric: CFI = .934, RMSEA = .045, SRMR = .137; scalar CFI = .919, RMSEA = .049, SRMR = .137). ΔCFI was the only comparison of model fit that did not support metric and scalar invariance (metric ΔCFI = -.026, scalar ΔCFI = -.015). Both ΔRMSEA and ΔSRMR supported metric and scalar invariance of the Somatic Symptoms Subscale (metric: ΔRMSEA = .010, ΔSRMR = .008; scalar: ΔRMSEA = .004, ΔSRMR = .000).

Cognitive Symptoms Subscale. The most severe response option for the self-injury and suicidal ideation items were not endorsed at all time points; therefore, the item was dichotomized. The CFI and RMSEA suggested acceptable global model fit for the configural, metric, and scalar invariance models (configural: CFI = .973, RMSEA = .038; metric: CFI = .958, RMSEA = .046; scalar: CFI = .952, RMSEA = .049). SRMR indicated unacceptable model for the configural, metric, and scalar invariance models (configural SRMR = .141, metric SRMR = .149, scalar SRMR = .149). ΔCFI was the only comparison of model fit that did not support metric invariance (ΔCFI = -.015, ΔRMSEA = .008, ΔSRMR = .008). All comparisons of model fit supported scalar invariance (ΔCFI = -.006, ΔRMSEA = .003, ΔSRMR = .000).

Somatic Symptoms Subscale. Only one out of three fit indices (RMSEA) suggested acceptable model fit for the configural invariance model (CFI = .923, RMSEA = .059, and SRMR = .109). All model fit indices indicated unacceptable model fit for the metric and scalar models (metric: CFI = .874, RMSEA = .074, SRMR = .125; scalar CFI = .839, RMSEA = .080, SRMR = .125). ΔCFI was the only comparison of model fit that did not support metric and scalar invariance (metric ΔCFI = -.049, scalar ΔCFI = -.035). Both ΔRMSEA and ΔSRMR supported metric and scalar invariance of the Somatic Symptoms Subscale (metric: ΔRMSEA = .015, ΔSRMR = .016; scalar: ΔRMSEA = .006, ΔSRMR = .000).

Anxiety Scale

The most severe response option for the item assessing fear of going to schools was not endorsed at all timepoints; therefore, the item was dichotomized. The CFI and RMSEA suggested acceptable global model fit for the configural, metric, and scalar invariance models (configural: CFI = .969, RMSEA = .044; metric: CFI = .959, RMSEA = .050; scalar: CFI = .950, RMSEA = .053). SRMR indicated unacceptable model for the configural, metric, and scalar invariance models (configural SRMR = .105, metric SRMR = .112, scalar SRMR = .112). All comparisons of model fit supported both metric and scalar invariance (metric: ΔCFI = -.010, ΔRMSEA = .006, ΔSRMR = .007; scalar: ΔCFI = -.009, ΔRMSEA = .003, ΔSRMR = .000).

ADHD Scale

The CFI and RMSEA suggested acceptable global model fit for the configural, metric, and scalar invariance models (configural: CFI = .968, RMSEA = .044; metric: CFI = .954, RMSEA = .052; scalar: CFI = .950, RMSEA = .053). SRMR indicated unacceptable model for the configural, metric, and scalar invariance models (configural SRMR = .097, metric SRMR =

384 .102, scalar SRMR = .102). ΔCFI was the only comparison of model fit that did not support
385 metric invariance (ΔCFI = -.014, ΔRMSEA = .008, ΔSRMR = .005). All comparisons of model
386 fit supported scalar invariance (ΔCFI = -.004, ΔRMSEA = .001, ΔSRMR = .000).

387 ***Somatic Scale***

388 The most severe response option for the items assessing both eye problems and vomiting
389 were not endorsed at all timepoints; therefore, the items were dichotomized. The CFI and
390 RMSEA, but not SRMR, suggested acceptable global model fit for the configural and metric
391 invariance models (configural: CFI = .964, RMSEA = .035, SRMR = .158; metric: CFI = .954,
392 RMSEA = .040, SRMR = .163). RMSEA was the only index of model fit that suggested
393 acceptable for the scalar invariance model (CFI = .940, RMSEA = .044, SRMR = .164). All
394 comparisons of model fit supported metric invariance (ΔCFI = -.010, ΔRMSEA = .005, ΔSRMR
395 = .005). Only ΔRMSEA and ΔSRMR supported scalar invariance (ΔCFI = -.014, ΔRMSEA =
396 .004, ΔSRMR = .001).

397

Table 4 Model Fit: Clinic-based Sample

	df	χ^2 *p*	CFI	ΔCFI	RMSEA	ΔRMSEA	90% CI RMSEA	SRMR	ΔSRMR
Depression/Affective Problems (N=219)									
Configural	3144	3980.896 *p* < .001	.960		.035		.031—.038	.129	
Metric	3205	4602.924 *p* < .001	.934	-.026	.045	.010	.042—.048	.137	.008
Scalar	3287	4984.286 *p* < .001	.919	-.015	.049	.004	.046—.051	.137	.000
Depression/Affective Problems: Cognitive Subscale (N=219)									
Configural	804	1054.672 *p* < .001	.973		.038		.031—.044	.141	
Metric	834	1221.855 *p* < .001	.958	-.015	.046	.008	.041—.052	.149	.008
Scalar	869	1316.430 *p* < .001	.952	-.006	.049	.003	.043—.054	.149	.000
Depression/Affective Problems: Somatic Subscale (N=219)									
Configural	480	847.769 *p* < .001	.923		.059		.053—.066	.109	
Metric	502	1105.179 *p* < .001	.874	-.049	.074	.015	.068—.080	.125	.016
Scalar	551	1321.866 *p* < .001	.839	-.035	.080	.006	.075—.086	.125	.000
Anxiety (N=219)									
Configural	687	975.182 *p* < .001	.969		.044		.037—.050	.105	
Metric	712	1098.660 *p* < .001	.959	-.010	.050	.006	.044—.056	.112	.007
Scalar	765	1228.547 *p* < .001	.950	-.009	.053	.003	.047—.058	.112	.000
Attention Deficit Hyperactivity Disorder (N=219)									
Configural	1695	2418.379 *p* < .001	.968		.044		.040—.048	.097	
Metric	1734	2758.397 *p* < .001	.954	-.014	.052	.008	.048—.056	.102	.005
Scalar	1816	2927.656 *p* < .001	.950	-.004	.053	.001	.049—.056	.102	.000
Somatic (N=217)									
Configural	1065	1353.662 *p* < .001	.964		.035		.029—.041	.158	
Metric	1099	1470.543 *p* < .001	.954	-.010	.040	.005	.034—.045	.163	.005
Scalar	1157	1641.944 *p* < .001	.940	-.014	.044	.004	.039—.049	.164	.001

Note: Δ = change between current model and previous model (i.e., change between configural and metric and change between metric and scalar). *df* = degrees of freedom, CFI = Comparative Fit Index, RMSEA = Root Mean Square Error of Approximation, CI = Confidence Interval, SRMS = Standardized Root Mean Square Residual.

Discussion

The YSR and ASR are widely used self-report measures of psychological symptoms and well-being. To facilitate their use across developmental periods in research and clinical practice, the YSR and ASR were designed to be comparable measures designed to be developmentally appropriate for youth and adults, respectively. However, use of these measures to quantify change in symptoms for the same individual requires that they assess psychopathology in the same way across time and across measure forms despite different item wordings to complement intended developmental stages (i.e., YSR→ASR)—otherwise known as longitudinal measurement invariance.

To date, no study has investigated the longitudinal measurement invariance of the Depression/Affective Problems Scale (or its constituent Cognitive and Somatic Subscales), Anxiety Scale, ADHD Scale, and Somatic Scale of the YSR and ASR in a sample where participants completed both measures. The present study finds differential support for each of these measures, underscoring the value in separately considering the psychometric properties of multidimensional scales. Results will be discussed in the order they were presented in the Results section (i.e., Depression/Affective Problems, Anxiety, ADHD, and Somatic).

Out of the Depression/Affective Problems Scale, the strongest support was for the Cognitive Symptoms Subscale, which featured consistently good model fit in both the population and clinical cohort (except the SRMR which was consistently above the cut-off in both samples) and all comparisons of model fit indices across both samples (except ΔCFI for metric invariance in the clinical cohort) supported all tested levels of invariance. There was slightly less support for the broader Depression/Affective Problems Scale. Specifically, while all three change indices in the population cohort supported metric invariance, ΔCFI did not support scalar invariance.

Similar results were found in the clinical cohort, expect that ΔCFI did not support either metric or scalar invariance. While the change in model fit statistics is the focal measurement of interest in invariance testing (because it focuses on how model fit reacts to the constraints that define the invariance), it is worth considering that all three absolute model fit statistics (CFI, RMSEA, SRMR) only indicated adequate fit for the configural model in the population cohort. In this cohort, SRMR did not support adequate global fit of the metric invariance model, and neither CFI nor SRMR supported adequate fit of scalar invariance models. There was worse absolute fit in the clinical cohort. With SRMR not indicating acceptable model fit for any of the models tested and CFI only supporting the configural model. Across both the population and clinical cohorts, there was less support for the global model fit, and longitudinal measurement invariance of, the Somatic Subscale of the Depression/Affective Problems Scale. Nearly identical patterns of results were found in both cohorts. Two relative change metrics, ΔRMSEA and ΔSRMR, supported the metric and scalar invariance of this scale; however, ΔCFI did not support metric or scalar invariance. In fact, the magnitude of the changes in CFI were quite notable (2.6-4.9x the acceptable cut-off). With respect to global model fit, only two indices (RMSEA and SRMR) supported acceptable model fit for configural invariance in the population cohort and only RMSEA supported acceptable model fit for configural invariance in the clinical cohort. No global model fit indices supported the metric or scalar invariance models. Consequently, combination use of the Depression/Affective Problems Scale and Cognitive Symptom Subscale of the YSR and ASR are likely suitable for clinical work or research in adolescent and/or adult populations when depression symptoms are of interest; however, the Somatic Symptoms Subscale be used with caution or with adjustments to account for measurement noninvariance (Putnick & Bornstein, 2016).

The Anxiety Scale showed the strongest support for both global model fit and longitudinal measurement invariance across the tested scales, as all metrics supported its psychometric properties except ΔCFI for the population scalar invariance model and the SRMR statistics in the clinical cohort (which were, notably, above the suggested cut-off in all models in the clinical cohort). The ADHD Scale, which had the greatest item-level differences between the YSR and ASR had strong support for longitudinal measurement invariance in the population-based sample (although the CFI was below the cutoff for acceptable model fit in all models estimated). There was slightly less support for the longitudinal measurement invariance of this scale in the clinical cohort, with ΔCFI below the acceptable cutoff for the metric invariance model and SRMR above the acceptable cutoffs for all models estimated. Finally, the Somatic Scale had identical patterns of psychometric support across the population and clinical cohorts. Specifically, all three relative change metrics supported metric invariance and two of the three (ΔRMSEA and ΔSRMR) supported scalar invariance. However, it is worth noting that the SRMR was above the cutoff for acceptable model fit in all three models, and both CFI and ΔCFI were below the acceptable cutoffs in the scalar invariance model.

One of the key strengths of this study is the inclusion of both the YSR and ASR across multiple time points. Thus, instead of solely testing the longitudinal measurement invariance of one of these measures or using the YSR and ASR in different groups, we were able to evaluate the appropriateness of transitioning from the YSR to the ASR for the same participant or client, as most appropriate for their age. Additionally, the sample was large enough to bolster confidence in the generalizability of these findings. Generalizability is further amplified by the fact that this is not an exclusively clinical sample; thus, there are less concerns regarding restriction of range than if this study were conducted in a strictly clinical or nonclinical sample.

This study is also the first we are aware of that tested the longitudinal measurement invariance of all of the DSM-IV scales shared between the YSR and ASR. Finally, the ability to separately test some of these models in a population-based vs. a clinic-based cohort supports the relevance of these results for both community-based and clinical populations.

However, this study should also be considered in light of its limitations. First, several items had to be dichotomized due to lack of participants selecting the most severe option at some waves. Although this is not surprising given the item content relative to the ages of assessment, these are still modeling deviations from standard scoring of the YSR and ASR. Second, as would be expected of most psychiatric symptom data, responses were largely skewed toward less severe responses. Third, likely due to a small sample size (relative to parameters estimated), there were estimation issues in the clinic-based cohorts that required items to be dichotomized to address low endorsement rates of certain response items, resulting in some discrepancies between the modeling of identical subscales between cohorts.

Conclusion

In conclusion, the present study supports the longitudinal measurement invariance of the YSR and ASR Depression/Affective Problems Scale, Cognitive Symptom Subscale of the Depression/Affective Problems Scale, Anxiety Scale, ADHD Scale, and Somatic Scale. The greatest concerns for longitudinal measurement invariance were for the Somatic Symptoms Subscale of the Depression/Affective Problems Scale. Notably, the degree of psychometric support was fairly comparable between cohorts (with the caveat that SRMR was above acceptable cutoffs in every single model tested in the clinical cohort) with slightly stronger support in the population-based sample. Consequently, clinicians and researchers should carefully consider which items to use, and how to aggregate them, when considering the

YSR/ASR as a potential measure to track mental health symptoms overtime and across developmental stages. However, additional work is needed to replicate this study in other samples (e.g., in active episodes of poor mental health, larger clinical samples that would result in more successful model estimation) and with different durations between assessments.

References

Abdellaoui, A., de Moor, M. H. M., Geels, L. M., van Beek, J. H. D. A., Willemsen, G., & Boomsma, D. I. (2012). Thought Problems from Adolescence to Adulthood: Measurement Invariance and Longitudinal Heritability. *Behavior Genetics*, *42*(1), 19–29. https://doi.org/10.1007/s10519-011-9478-x

Achebach, T. M. (1991). *Manual for the youth self-report and 1991 profile*.

Achebach, T. M., & Rescorla, L. A. (2003). *Manual for the ASEBA adult forms & profiles*.

Achenbach, T. M. (1991). *Manual for the child behavior checklist/4-18 and 1991 profile*. Burlington, VT : Dept. of Psychiatry, University of Vermont. http://archive.org/details/manualforchildbe0000ache_g2r5

Achenbach, T. M., & Rescorla, L. (2003). *Manual for the ASEBA Adult Forms & Profiles: For Ages 18-59 : Adult Self-report and Adult Behavior Checklist*. ASEBA.

Barzeva, S. A., Meeus, W. H. J., & Oldehinkel, A. J. (2019). Social Withdrawal in Adolescence and Early Adulthood: Measurement Issues, Normative Development, and Distinct Trajectories. *Journal of Abnormal Child Psychology*, *47*(5), 865–879. https://doi.org/10.1007/s10802-018-0497-4

Bosch, N. M., Riese, H., Dietrich, A., Ormel, J., Verhulst, F. C., & Oldehinkel, A. J. (2009). Preadolescents' Somatic and Cognitive-Affective Depressive Symptoms Are Differentially Related to Cardiac Autonomic Function and Cortisol: The TRAILS Study. *Psychosomatic Medicine*, *71*(9), 944. https://doi.org/10.1097/PSY.0b013e3181bc756b

Chen, F. F. (2007). Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Ganzeboom, H. B. G., & Treiman, D. J. (1996). Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research*, *25*(3), 201–239. https://doi.org/10.1006/ssre.1996.0010

Guerrero, M., Hoffmann, M., & Pulkki-Råback, L. (2020). Psychometric Properties of the Adult Self-Report: Data from over 11,000 American Adults. *Stats*, *3*(4), 465–474. https://doi.org/10.3390/stats3040029

Hu, L., Bentler, P. M., & Hu, L. (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. https://doi.org/10.1080/10705519909540118

Huisman, M., Oldehinkel, A. J., de Winter, A., Minderaa, R. B., de Bildt, A., Huizink, A. C., Verhulst, F. C., & Ormel, J. (2008). Cohort profile: The Dutch "TRacking Adolescents" Individual Lives' Survey'; TRAILS. *International Journal of Epidemiology*, *37*(6), 1227–1235. https://doi.org/10.1093/ije/dym273

Jonker, I., Rosmalen, J. G. M., & Schoevers, R. A. (2017). Childhood life events, immune activation and the development of mood and anxiety disorders: The TRAILS study. *Translational Psychiatry*, *7*(5), e1112. https://doi.org/10.1038/tp.2017.62

Mac Giollabhui, N., & Hartman, C. A. (2022). Examining inflammation, health, stress and lifestyle variables linking low socioeconomic status with poorer cognitive functioning during adolescence. *Brain, Behavior, and Immunity*, *104*, 1–5. https://doi.org/10.1016/j.bbi.2022.04.020

Oldehinkel, A. J., Rosmalen, J. G., Buitelaar, J. K., Hoek, H. W., Ormel, J., Raven, D., Reijneveld, S. A., Veenstra, R., Verhulst, F. C., Vollebergh, W. A., & Hartman, C. A. (2015). Cohort Profile Update: The TRacking Adolescents' Individual Lives Survey

(TRAILS). *International Journal of Epidemiology*, *44*(1), 76–76n. https://doi.org/10.1093/ije/dyu225

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004.Measurement

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2), 1–36.

Team, R. C. (2013). *R: A language and environment for statistical computing. R Foundation for Statistical Computing* [Computer software]. http://www.r-project.org

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*(1), 409–428. https://doi.org/10.3758/s13428-018-1055-2