

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/397061705>

Is mental health problems measurement equivalent across disability, gender, age, and cohort in children? Findings from a Swedish national survey

Article in *Scandinavian Journal of Public Health* · October 2025

DOI: 10.1177/14034948251380635

CITATIONS

0

READS

11

3 authors:



Magnus Ivarsson

Linköping University

11 PUBLICATIONS 97 CITATIONS

SEE PROFILE



Lina Homman

Linköping University

17 PUBLICATIONS 351 CITATIONS

SEE PROFILE



Henrik Danielsson

Linköping University

130 PUBLICATIONS 3,524 CITATIONS

SEE PROFILE

ORIGINAL ARTICLE

Is mental health problems measurement equivalent across disability, gender, age, and cohort in children? Findings from a Swedish national survey

MAGNUS IVARSSON^{ID}, LINA HOMMAN & HENRIK DANIELSSON*Department of Behavioural Sciences and Learning, Linköping University, Sweden***Abstract**

Aims: To investigate whether mental health problems can be described as a latent factor indicated by nine items in the Swedish Survey of Children's Living Conditions (Barn-ULF), and to explore whether this structure varies across disability, age, gender and cohorts. **Methods:** Cross-sectional survey data collected yearly from 2013 to 2019 regarding 3676 children (aged 10–18 years) was retrieved from two linked national registers: Barn-ULF (child-reported mental health problems data) and ULF/SILC (parent-reported data on child characteristics). The structural validity of the one-factor model was assessed using confirmatory factor analysis (CFA) and threshold invariance concerning disability, gender, age and cohort was tested using multigroup CFAs. The practical significance of invariance was assessed by comparing weighted scores from the single-group model and the multigroup models. **Results:** A substantially modified one-factor model demonstrated adequate fit ($\chi^2/df = 1.94$, comparative fit index = 1.00, Tucker–Lewis index = 0.99, root mean square error of approximation = 0.02). Significant measurement non-invariance was found across all groups; however, its impact on the latent variable mean was minimal (<0.03 in Cohen's d). **Conclusions:** Barn-ULF might be a structurally valid tool for monitoring Swedish children's mental health problems. However, differences in model performance across child characteristics and time – even if small – indicate that group comparisons must be conducted cautiously.

Keywords: Surveys and questionnaire, mental health, Sweden, psychometrics, disabled persons, child, factor analysis, sex factors, age factors

Background

Adolescent mental health has been intensely discussed over the past few decades. In this context, data from national surveys is often used to analyse secular trends and the influence of different factors on child and adolescent health outcomes (e.g. Trethewey et al. [1], Eriksson and Stattin [2]). For group comparisons to be valid, surveys must be equally applicable to all subgroups of interest. If questions are interpreted differently or responded to inconsistently across groups for reasons unrelated to the health status, there is a risk of biased comparisons. In the current study, we assess the structural

validity and the equivalence (threshold measurement invariance) of nine questions measuring mental health problems from the Swedish National Survey of Children's Living Conditions (Barn-ULF).

Mental health in children and adolescents

Globally, mental health problems are a major concern for children and adolescents [3]. However, there is ongoing debate about temporal trends for specific diagnoses and aspects of mental health problems. Differentiating psychiatric disorders and mental ill-health from the broader construct of mental health problems is key to this discussion [4]. For instance,

Correspondence: Magnus Ivarsson, Department of Behavioural Sciences and Learning, Linköping University, S-581 83 Linköping, Sweden.
E-mail: magiv10@liu.se

Date received 3 October 2024; reviewed 1 September 2025; accepted 5 September 2025

© Author(s) 2025

Article reuse guidelines: sagepub.com/journals-permissions

DOI: 10.1177/14034948251380635

journals.sagepub.com/home/sjp



evidence collected mainly from the Global North indicates that while mental health issues in younger children have remained stable, internalizing problems – but not externalizing – have increased in adolescent girls over recent decades [5]. Recent studies from Sweden [6], Finland [7] and Norway [8] largely confirm these findings. In addition to gender and age, evidence suggests an increased risk for mental health problems in children with specific disabilities and chronic health conditions – such as asthma [9], autism [10] and cerebral palsy [11].

The Swedish survey of children's living conditions

In Sweden, Barn-ULF is one of the key official data sources on adolescent health [12]. Barn-ULF is collected by Statistics Sweden (Statistiska centralbyrån (SCB)) following a directive from the Swedish government to gather data on the living conditions of Swedish citizens. It is one of few national surveys that capture the subjective experiences of children and adolescents on topics covering the most important areas of their lives such as school, leisure time, health, friends, family and other relationships. Data from Barn-ULF is used by government officials to guide decisions, researchers (e.g. Låftman et al. [13], Olsson [14]) and the public. A specific application of the data, as noted by SCB [12], is to compare the living conditions of different subgroups.

Measurement invariance

However, for group comparisons to be valid, it is essential that the data collection method does not introduce bias between groups. A key quality of a survey is measurement invariance, meaning that it remains indifferent to respondent characteristics. Measurement invariance testing often takes place within a structural equation modelling framework using confirmatory factor analysis (CFA). CFA is used to study how observed variables (e.g. items in a questionnaire) represent one or more underlying latent constructs, such as mental health problems [15]. Invariance testing typically follows a stepwise approach, involving the assessment of different aspects of invariance: i) configural invariance, which assesses whether the basic structure of a model is equivalent across groups, ii) metric invariance, which tests whether the factor loadings of the indicators are equivalent, iii) scalar invariance, which evaluates the equivalence of item intercepts, and iv) residual invariance, which examines the equivalence of item residual variances. For categorical or ordinal level data, the number of steps is often reduced to three as

metric and scalar invariance are combined in one step (threshold invariance test) when using specific estimators. Each step imposes increasing constraints on the model, with invariance at lower levels being a prerequisite for testing at higher levels [15].

The measurement invariance for mental health problems outcomes – the Strengths and Difficulties Questionnaire (SDQ) [16] in the following examples – have repeatedly been investigated in relation to participant gender (e.g. Bøe et al. [17], Murray et al. [18]), age (e.g. Stone [19]) and the time of data collection (e.g. Speyer et al. [20]). In recent years, increasing attention has been paid to measurement invariance across various disabilities [21]. Disability-related measurement non-invariance has been identified in common mental health screening instruments such as the Child Behavior Checklist (CBCL) [22] when making a comparison between children with autism and children with both autism and intellectual disability [23], and the SDQ when comparing autistic adolescents with typically developing peers [24]. Nevertheless, indices from these instruments are frequently used in studies of mental health-related outcomes in individuals with disabilities [25]. For a recurring national survey aimed at comparing groups, ensuring measurement invariance across disability, gender, age and cohorts is essential. To date, no investigations of the structural validity or the measurement invariance of constructs in Barn-ULF have been conducted.

Aim of the study

This study aims to investigate the structural validity of a one-factor mental health problem model indicated by nine items in Barn-ULF and whether this model displays threshold measurement invariance across disability, gender, age and cohorts.

Methods

The methods used in this study were determined before the authors had access to the data and were preregistered at <https://osf.io/czrx8>. The study protocol was reviewed and approved by the Swedish Ethical Review Authority (2021-04388). The code used for the analyses is available at <https://osf.io/x9sn8/>. The registry data used in this study cannot be shared as a public dataset.

Data

The study utilizes cross-sectional survey data from two linked registers administered by SCB: (i) ULF/SILC, which contains information about randomly

Table I. Participant characteristics.

	Gender		Total
	Boys	Girls	
Age, <i>M</i> (<i>SD</i>)	14.47 (2.27)	14.36 (2.31)	14.41 (2.29)
<i>N</i>			
2013–2014	655	746	1401
2015–2016	507	592	1099
2018–2019	569	607	1176
Total	1731	1945	3676
Disability/health condition, % (<i>n</i>)			
Asthma/allergy	12.56 (217)	9.26 (180)	10.81 (397)
ADHD/ASD	5.52 (95)	2.27 (44)	3.80 (139)
Dyslexia/dyscalculia	7.20 (124)	4.87 (94)	5.97 (218)
Hearing	1.22 (21)	1.24 (24)	1.23 (45)
Vision	1.39 (24)	1.18 (23)	1.28 (47)
Mobility	1.10 (19)	1.23 (24)	1.17 (43)
Other	2.60 (45)	3.81 (74)	3.24 (119)
Any	25.19 (430)	20.14 (387)	22.51 (817)

ADHD: attention deficit hyperactivity disorder; ASD: autism spectrum disorder

Table II Barn-ULF items included as mental health problem indicators.

Item	Indicators	Abbreviation
1	How often do you feel sad or down?	Sad/down
2	How often do you have difficulties to sit still and focus?	Difficulties sitting still/focusing
3	How often do you feel tense or nervous?	Tense/nervous
4	How often do you feel stressed?	Stress
5	How often do you have sleeping problems at night?	Sleep problems
6	How often do you have difficulties falling asleep?	Difficulties falling asleep
7	How often do you feel tired during school time?	Tired
8	How often do you have a headache?	Headache
9	How often do you have a stomach ache?	Stomach ache

Barn-ULF: Swedish National Survey of Children's Living Conditions

selected Swedish adults and young people (aged 16 years and older), and (ii) Barn-ULF, which includes data from children aged 12–18 years (10–18 before 2014) whose parents took part in ULF/SILC [12]. Data in both ULF/SILC and Barn-ULF is collected by SCB through structured telephone interviews. For the present study, data collected between 2013 and 2019 was requested owing to the introduction of disability-related variables to the ULF database in 2013. Most of the data in the present study was child-rated (part of Barn-ULF). However, information on the child's disability and age was parent reported (ULF/SILC) as these items were not directed to children. The data is cross-sectional, with different participants sampled each year. Across 2013–2019, Barn-ULF response rates ranged from 51% to 60% [12]. Owing to systematic differences between responders and non-responders, SCB applies weights to reduce bias in the results.

Participants

Barn-ULF contains data for a total of 4305 children across the years 2013–2019. However, as no data was

collected on disability-related variables in 2017, this year was excluded from the analyses, resulting in a total of 3676 participants aged 10–18 years. Sample characteristics are displayed in Table I.

Material

Indicators

All items that were identified as possible indicators of mental health problems in the Barn-ULF survey were included in the analyses (see Table II). Specifically, Barn-ULF items with thematic resemblance to items in well-established mental health problems screening scales, such as SDQ and CBCL, were identified and selected in this process. In Barn-ULF, slight changes in item wording and the number of response levels have been made over the years (mainly from 2016), which meant that items and/or response levels had to be collapsed to enable the merge of data from 2013–2015 and 2016–2019 for some indicators (see Supplementary material Table I online for a full list of items, response levels and information on how variables were merged). Depending on the item and the year the data was collected, there were 6–7 response

options on the Likert-style response scales. However, after re-coding 'Do not know' and 'Refuse to answer' to missing and merging levels when necessary, 4–5 levels remained. For three indicators (sad/down, difficulties sitting still/focusing, and tense/nervous), the scoring was as follows: 'Most of the time' (3), 'Pretty often' (2), 'Not that often' (1) and 'Almost never' (0). Four indicators (sleep problems, difficulties falling asleep, headache, and stomach ache) were scored as: 'More than once a week' (3), 'Once a week' (2), 'Once a month or so' (1) and 'Less often or never' (0). Finally, two indicators (stress and tired) had the following scoring: 'Every day' (4), 'A few times per week' (3), 'Once a week' (2), 'Once a month or so' (1) and 'Less often or never' (0). The data transformation followed the procedure outlined in the preregistration with one exception. The preregistration stated that six variables derived from pairs of yes/no questions (presence of symptom or not) and linked follow-up questions (frequency of those problems) provided by SCB would be used as indicators. Instead, the original yes/no and frequency items were used and combined into six variables as part of the analysis to enhance transparency.

Grouping variables

Information on seven disabilities/health conditions (henceforth referred to as disabilities) was collected in the form of binary ('Yes' (1)/'No' (2)) questions answered by the caregiver of the child. For asthma/allergy, visual impairment, and other disabilities, the item contained the additional description 'that hinders his/her daily life'. Based on the responses to the specific items, an overall binary disability condition variable was created ('Any disability' (1)/'No disability' (2)). Information about the child's age (in years), gender ('Boy' (1)/'Girl' (2)) and the year the data was collected (cohort) was also retrieved. For age, two groups were defined: i) children 10–14 years of age and ii) children 15–18 years of age.

Procedure

Data analysis

CFA using the *lavaan* [26] R package was performed to evaluate the fit of a model with the nine variables listed in Table II as indicators of a latent mental health problems variable. The model was optimized in a step-by-step approach by iteratively adding residual correlations between pairs of indicators with the largest impact on model fit. This procedure was repeated until there were no more such modifications with the potential of making a meaningful improvement to model fit (i.e. modification index above 4).

The degree of fit of the model was evaluated by a combination of different measures (and commonly applied cut-offs for acceptable fit): the comparative fit index (>0.95), the Tucker–Lewis index (>0.95), the standardized root mean squared residual ((SRMR) ≤ 0.08 ; not included in the preregistration), the root mean square error of approximation ((RMSEA) ≤ 0.06 or 90% confidence interval ≤ 0.08) and χ^2 / df (< 2).

In the second step, the measurement invariance of the optimized one-factor model was tested, where metric and scalar (threshold) invariances were assessed simultaneously, given that the indicators were ordinal. Four multigroup analyses were performed to assess whether the factor loadings and item thresholds specified in the original model differed according to disability, gender, age and cohort. For each analysis, a constrained model (with loadings and thresholds held equal across groups, scale factors fixed at 1 in the first group and freed in the other, and the factor mean fixed at 0 in one group and freed in the other) was compared with a fully freed ('configural') model (where factor loadings and thresholds were freed across groups, scale factors fixed at 1, and factor means fixed at 0).

For multigroup analyses with a significant difference in the fit to data (χ^2) between the constrained and configural models, the possible group effect of each item's thresholds and factor loadings was examined. This was achieved by comparing the configural model with a series of models, each releasing the loading and threshold of a different single indicator, so that across models, all indicators were tested. All multigroup analyses were conducted in Mplus [27] owing to the lack of methods to compare robust test statistics of nested models in *lavaan*. Outputs from Mplus were imported to RStudio with *MplusAutomation* [28].

For all analyses, the WLSMV estimator was used to account for the ordered nature of the indicators. The fully constrained models and the models in which one item was freed at a time were compared with the configural models with likelihood tests based on χ^2 differences between models. For each comparison, the configural model was preferred if the models had a significantly different fit ($p < 0.05$). A significant difference in the fit of the models was interpreted as an indication of variation in the structure of mental health problems based on disability, gender, age and cohort.

In addition to the procedure described in the preregistration, the practical consequences of identified model non-invariance were assessed by comparing approximated factor scores from the single-group

Table III. Correlations between the indicator variables.

	1	2	3	4	5	6	7	8
1. Headache								
2. Stomach ache	0.29							
3. Sleep problems	0.24	0.21						
4. Difficulties falling asleep	0.18	0.20	0.45					
5. Tired	0.24	0.23	0.32	0.29				
6. Stress	0.25	0.25	0.29	0.26	0.36			
7. Sad/down	0.23	0.20	0.20	0.19	0.23	0.28		
8. Difficulties sitting still/focusing	0.10	0.14	0.17	0.17	0.22	0.14	0.17	
9. Tense/nervous	0.16	0.19	0.16	0.16	0.20	0.30	0.34	0.18

The reported values are Kendall's τ coefficient between the median values across the five imputed datasets. All correlations are significant at the $p < 0.001$ level.

CFAs with those from the corresponding multigroup CFAs, using paired t -tests and Cohen's d . These comparisons were conducted in cases where threshold invariance was not established. Factor scores were approximated by computing weighted sums of item responses, using standardized factor loadings from the Mplus models.

Missing data

Six participants were removed owing to not having provided answers to any of the mental health problem-related items. For the resulting data set, missing data was considered missing at random (missing data was 0.23%), and multivariate imputation by chained equations was conducted with the *mice* [29] R package. Indicators and disability-related variables were imputed separately in one step (five imputations each).

Results

As seen in Table III, the correlations between indicators were weak overall (median 0.21, range 0.10–0.45). For the non-modified model (see Figure 1(a)), the CFA showed that all pooled (by calculating the mean across the imputed data sets) model parameters made significant contributions. Only one of the indicators ('Difficulties sitting still/focusing') had a factor loading below 0.4. However, the fit of the model to data was inadequate according to all but one SRMR) of the fit indices (see Table IV). The optimization of the model, according to the predefined principles, resulted in a model with nine residual correlations added (see Figure 1(b)). As with the original model, all pooled model parameters made significant contributions. Two items loaded below 0.4, but the fit to data was acceptable according to all selected fit indices (see Table IV).

Multigroup analyses

As indicated by the group equivalent (χ^2) tests reported in Table IV, there were significant

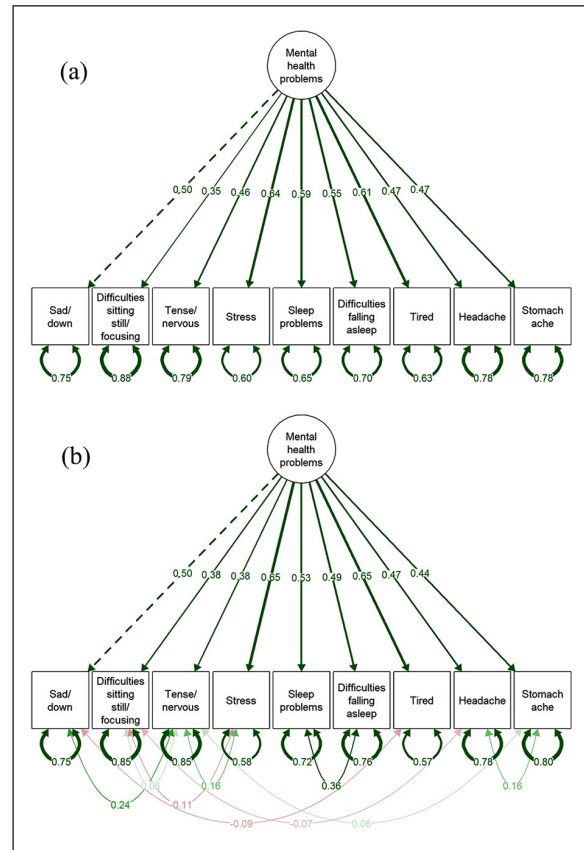


Figure 1. Graphical representation of the original (a) and optimized (b) mental health problems models, with latent variables shown as circles, indicators as squares, factor loadings (fully standardized) as single-headed arrows, residual correlations and error variances as double-headed arrows, and the loading fixed for model identification indicated by a dashed line.

differences in the structure across all four conditions investigated. In other words, factor loadings and thresholds differed when making comparisons between children with disability and those without, between boys and girls, and between children aged 10–14 years and those aged 15–18 years, and depending on whether data was collected in 2013–2014, 2015–2016, or 2018–2019. Thus, additional analyses were conducted for each group to investigate which

Table IV. Selected fit indices for single- and multigroup mental health problems models.

Model	χ^2	df	$p(\chi^2)$	χ^2/df	CFI	TLI	SRMR	RMSEA		Diff. test		
								Estimate	90% CI	χ^2	df	p
Single-group analyses												
Original	607.998	27	0.000	22.518	0.895	0.861	0.051	0.077	0.071–0.082			
Optimized	34.915	18	0.010	1.940	0.997	0.994	0.011	0.016	0.008–0.024			
Multigroup analyses												
Age group configural	64.864	37	0.003	1.753	0.997	0.995	0.053	0.020	0.012–0.028			
Age group constrained	294.264	63	0.000	4.671	0.977	0.974	0.056	0.045	0.040–0.050	193.107	26	0.000
Cohort configural	117.480	56	0.000	2.098	0.994	0.989	0.054	0.030	0.022–0.038			
Cohort constrained	274.542	108	0.000	2.542	0.985	0.985	0.054	0.035	0.030–0.041	152.747	52	0.000
Disability configural	65.588	37	0.003	1.773	0.997	0.995	0.050	0.021	0.012–0.028			
Disability constrained	152.717	63	0.000	2.424	0.991	0.990	0.051	0.028	0.022–0.033	80.206	26	0.000
Sex configural	65.713	37	0.002	1.776	0.997	0.995	0.049	0.021	0.012–0.029			
Sex constrained	422.596	63	0.000	6.708	0.965	0.960	0.053	0.056	0.051–0.061	294.230	26	0.000

In the configural models, all model parameters are freed. In the constrained models, factor loadings have been constrained across groups.

df: degrees of freedom; CFI: comparative fit index; TLI: Tucker–Lewis index; SRMR: standardized root mean squared residual; RMSEA: root mean square error of approximation; CI: confidence interval; Diff.: difference

specific parameters differed across groups. Models with one path freed at a time were compared with the configural model. The results from these analyses (presented in Supplementary Table II) showed that all investigated paths differed significantly when grouping on gender and that half or more of the paths differed when grouping on age group and cohort. However, for disability, only two paths (stress and difficulties sitting still/focusing) differed.

Factor loading differences impact on factor scores

There were no significant differences, with effect sizes (Cohen's d) less than 0.04, in mean factor scores between the optimized single-group CFA and the multigroup CFAs across all comparisons: for disability, $M_D=0.01$, 95% confidence interval (CI) (−0.12,0.14), $t(3,675)=0.21$, $p=.0835$, $d=0.00$; for age group, $M_D=0.07$, 95% CI (−0.07,0.20), $t(3,675)=0.96$, $p=0.336$, $d=-0.02$; for sex, $M_D=0.11$, 95% CI (−0.02,0.25), $t(3,675)=1.61$, $p=0.107$, $d=-0.03$; or for cohort, $M_D=0.00$, 95% CI (−0.02,0.01), $t(3,675)=-0.36$, $p=0.721$, $d=0.01$. These comparisons involved factor scores that were derived from the same set of items but calculated using different models with varying factor loadings.

Discussion

In this study, we aimed to investigate whether mental health problems can be described as a latent factor indicated by nine thematically linked items in Barn-ULF and whether this model was invariant to participant disability, gender, age and cohort. The results showed that the one-factor model had an acceptable

fit to the data after being substantially modified according to predefined principles. Measurement non-invariance was observed for all grouping variables. However, analyses of the differences between models taking group differences into account and those that do not raise questions about the functional impact of the observed lack of invariance.

The indicators' factor loadings provide further insights into the type of mental health problems construct being measured in Barn-ULF. Notably, difficulties sitting still/focusing and being tense/nervous contributed the least to explaining variability in the mental health problems factor, while being stressed and tired contributed the most. One possible explanation for the relatively smaller importance of the difficulties sitting still/focusing indicator is that it represents a functional impairment rather than an emotional state and is therefore more closely related to disability-related constructs [4]. This interpretation is supported by the fact that difficulties sitting still/focusing was one of two indicators that were not invariant to the presence of a disability, that is, for children with a disability; this question might have been interpreted differently from for children without. The rationale for including it as a possible indicator in the model was that similar items are typically included in scales screening broadly for emotional and behavioural problems in children. Further, difficulty concentrating is one of the most frequently reported symptoms across disorders in the *Diagnostic and statistical manual of mental disorders* [30].

The results of the multigroup analyses suggest that the one-factor model does not perform equally for children with disabilities and those without, for boys and girls, for younger and older children, and across different cohorts. At the item level, no indicator was

found to be invariant across all grouping conditions. Notably, stress was the only indicator that consistently showed non-invariance across all multigroup analyses. However, significant measurement non-invariance does not automatically imply substantial practical importance [21].

One way to approximate the practical consequences of measurement non-invariance is to estimate the difference in factor scores when using factor loadings from a model that does not account for group differences (a single-group analysis) compared with one that does (a multigroup analysis). The magnitude of this difference indicates whether the non-invariance would introduce practically relevant bias. In this study, no significant differences were observed between the estimated latent variable scores derived from models that accounted for factor loading differences between groups and those that did not. Cohen's d effect sizes were very small ($d < 0.04$) in all cases. Therefore, it is unlikely that the identified non-invariance would significantly distort conclusions about group differences in future studies using Barn-ULF data.

Limitations

The unequal distribution of participants between the groups with and without a disability might have introduced bias into the results, as some fit indices, such as χ^2 , are sensitive to sample size. For the other grouping variables, group sizes were more equal. While the χ^2 test is standard in measurement invariance testing, other fit indices offer supplementary information on aspects of model fit, especially with large samples. While the preregistered plan for this study focused on χ^2 -based inference, future studies would benefit from broadening the inference criteria. Further, it should be noted that most children with a disability reported conditions that might be related to fewer functional impairments and/or participation restrictions than others, for example, asthma/allergy and dyslexia/dyscalculia as compared with severe or profound intellectual and motor disabilities. It is possible that measurement invariance could differ between specific diagnostic categories and results may not generalize to the disability group as a whole and/or specific diagnostic groups. Unfortunately, most of the subgroups (e.g. hearing impairment, attention deficit hyperactivity disorder/autism spectrum disorder) were too small to allow for separate analyses. Finally, although the practice of adding residual covariances between indicators with design and content similarities is justifiable, the extent of modifications to the model limits the generalizability of the findings.

Future research

As Barn-ULF contains many more items measuring other aspects of children's living conditions, future studies will need to investigate structural validity and measurement for these items and factors. Future studies should also assess measurement invariance in relation to specific disabilities and/or children with different functional levels (regardless of diagnosis). As SCB has modified the recruitment strategy for Barn-ULF, future cohorts might increase in size, allowing for more fine-grained analyses.

Conclusions

Barn-ULF is an important tool for monitoring the living conditions of the general Swedish child population. This study demonstrates that it was possible to combine nine survey items related to mental health problems into a single factor after substantial modifications to the model. Further, these items function differently based on child disability, gender, age and cohort. Although there were indications that the effects of these group differences were very small, the lack of measurement invariance suggests that group comparisons should be made with caution.

Author contributions

The authors made the following contributions. Magnus Ivarsson: conceptualization, data curation, formal analysis, methodology, project administration, visualization, writing – original draft, writing – review and editing; Lina Homman: conceptualization, data curation, formal analysis, methodology, project administration, writing – review and editing; Henrik Danielsson: conceptualization, funding acquisition, methodology, supervision, writing – review and editing.

Declaration of conflicting interests

The authors have no conflicts of interest to declare.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the Swedish Research Council (grant number 2018-05824_VR).

ORCID iD

Magnus Ivarsson  <https://orcid.org/0000-0002-5456-1597>

Supplementary material

Supplementary material for this article is available online.

References

- [1] Trethowey SP, Mathews F, Russell A, et al. Mental health of children and young people aged 5–16 in England: Socio-demographic and clinical characteristics associated with support and service contact. *Eur Psychiatry* 2023; 66 (S1):S582–S582.
- [2] Eriksson C and Stattin H. Secular trends in mental health profiles among 15-year-olds in Sweden between 2002 and 2018. *Front Public Health* 2023; 11: 1015509.
- [3] Kieling C, Baker-Henningham H, Belfer M, et al. Child and adolescent mental health worldwide: Evidence for action. *Lancet* 2011;378:1515–25.
- [4] Granlund M, Imms C, King G, et al. Definitions and operationalization of mental health problems, wellbeing and participation constructs in children with NDD: Distinctions and clarifications. *Int J Environ Res Public Health* 2021;18:1–19.
- [5] Bor W, Dean AJ, Najman J, et al. Are child and adolescent mental health problems increasing in the 21st century? A systematic review. *Aust N Z J Psychiatry* 2014;48:606–16.
- [6] Blomqvist I, Henje Blom E, Hägglöf B, et al. Increase of internalized mental health symptoms among adolescents during the last three decades. *Eur J Public Health* 2019;29:925–31.
- [7] Mishina K, Tiiri E, Lempinen L, et al. Time trends of Finnish adolescents' mental health and use of alcohol and cigarettes from 1998 to 2014. *Eur Child Adolesc Psychiatry* 2018;27:1633–43.
- [8] Potrebny T, Nilsen SA, Bakken A, et al. Secular trends in mental health problems among young people in Norway: A review and meta-analysis. *Eur Child Adolesc Psychiatry*.
- [9] Goodwin RD, Bandiera FC, Steinberg D, et al. Asthma and mental health among youth: Etiology, current knowledge and future directions. *Expert Rev Respir Med* 2012;6:397–406.
- [10] Van Steensel FJA, Bögels SM and Perrin S. Anxiety disorders in children and adolescents with autistic spectrum disorders: A meta-analysis. *Clin Child Fam Psychol Rev* 2011;14:302.
- [11] Downs J, Blackmore AM, Epstein A, et al. The prevalence of mental health disorders and symptoms in children and adolescents with cerebral palsy: A systematic review and meta-analysis. *Dev Med Child Neurol* 2018;60:30–8.
- [12] Statistikmyndigheten SCB. Undersökningarna av barns levnadsförhållanden, <https://www.scb.se/hitta-statistik/statistik-efter-amne/befolkning-och-levnadsforhallanden/levnadsforhallanden/undersokningarna-av-barns-levnadsforhallanden/> (2024, accessed 19 June 2025).
- [13] Låftman S, Fransson E and Östberg V. Sociodemographic inequalities in adolescents' health-related behaviours: The case of Sweden. *Eur J Public Health* 2016;26:ckw17 4.214.
- [14] Olsson E. The role of relations: Do disadvantaged adolescents benefit more from high-quality social relations? *Acta Sociol* 2009;52:263–86.
- [15] Putnick DL and Bornstein MH. Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Dev Rev* 2016;41:71–90.
- [16] Goodman R. The Strengths and Difficulties Questionnaire: A research note. *J Child Psychol Psychiatry* 1997;38:581–6.
- [17] Bøe T, Hysing M, Skogen JC, et al. The Strengths and Difficulties Questionnaire (SDQ): Factor structure and gender equivalence in Norwegian adolescents. *PLoS One* 2016;11:e0152202.
- [18] Murray AL, Speyer LG, Hall HA, et al. A longitudinal and gender invariance analysis of the Strengths and Difficulties Questionnaire across ages 3, 5, 7, 11, 14, and 17 in a large U.K.-representative sample. *Assessment* 2022;29:1248–61.
- [19] Stone LL, Janssens JMAM, Vermulst AA, et al. The Strengths and Difficulties Questionnaire: Psychometric properties of the parent and teacher version in children aged 4–7. *BMC Psychol* 2015;3:4.
- [20] Speyer LG, Auyeung B and Murray AL. Longitudinal invariance of the Strengths and Difficulties Questionnaire across ages 4 to 16 in the ALSPAC sample. *Assessment* 2023;30:1884–94.
- [21] Farmer C, Kaat AJ, Edwards MC, et al. Measurement invariance in intellectual and developmental disability research. *Am J Intellect Dev Disabil* 2024;129:191–8.
- [22] Achenbach TM and Ruffle TM. The Child Behavior Checklist and related forms for assessing behavioral/emotional problems and competencies. *Pediatr Rev* 2000;21:265–71.
- [23] Dovgan K, Mazurek MO and Hansen J. Measurement invariance of the child behavior checklist in children with autism spectrum disorder with and without intellectual disability: Follow-up study. *Res Autism Spectr Disord* 2019;58:19–29.
- [24] Turcan C, Delamain H, Loke A, et al. Measurement invariance of the parent-reported Strengths and Difficulties Questionnaire in autistic adolescents. *Autism* 2024; 28:2623–2636.
- [25] Danielsson H, Imms C, Ivarsson M, et al. A systematic review of longitudinal trajectories of mental health problems in children with neurodevelopmental disabilities. *J Dev Phys Disabil* 2024;36:203–42.
- [26] Rosseel Y. Lavaan: An R package for structural equation modeling. *J Stat Softw* 2012;48:1–36.
- [27] Muthén LK and Muthén BO. *Mplus user's guide*. 8th ed. Los Angeles, CA: Muthén & Muthén, 2017.
- [28] Hallquist MN and Wiley JF. *MplusAutomation: An R package for facilitating large-scale latent variable analyses in Mplus*. *Struct Equ Modeling* 2018;621–38.
- [29] Van Buuren S and Groothuis-Oudshoorn K. Mice: Multi-variate Imputation by Chained Equations in R. *J Stat Softw* 2011;45:1–67.
- [30] Forbes MK, Neo B, Nezami OM, et al. Elemental psychopathology: Distilling constituent symptoms and patterns of repetition in the diagnostic criteria of the DSM-5. *Psychol Med* 2024;54:886–94.