RESEARCH

Open Access

Four scales measuring mental wellbeing in the Nordic countries: do they tell the same story?

Leif Edvard Aarø[1*], Otto Robert Smith[1,2,3], Mogens Trab Damsgaard[4], Anne-Siri Fismen[5], Marit Knapstad[1], Nelli Lyyra[6,7], Oddrun Samdal[8], Einar Baldvin Thorsteinsson[9] and Charli Eriksson[10]

Abstract

Background Mental wellbeing is an important focus in surveys among adolescents. Several relevant instruments are available. In the Nordic part of the Health Behaviour in School-aged Children (HBSC) study 2022, four different scales for the measurement of wellbeing, were employed: Cantril's Ladder, the WHO-5 Wellbeing Index, the seven-item Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS), and the HBSC Health Complaints Scale. This study aims to examine statistically to what extent these scales overlap or measure distinctly different aspects of mental wellbeing.

Methods Data stem from the Nordic part of the HBSC 2022 study ($n = 28\ 189$). In all statistical analyses, data are weighted to ensure equal representation of genders, age groups (ages 11, 13, and 15 years), and countries (Denmark, Finland, Iceland, Norway, Sweden). Adjustments were made for cluster effects (school classes). The statistical analyses included factor analysis, general linear modeling, variants of latent variable analysis, and structural equation modeling including bifactor modeling.

Results Exploratory factor analysis produced three factors corresponding well to the three multi-item instruments, with the single item Cantril's ladder loading on the factor defined by the WHO-5 Wellbeing Index. Confirmatory factor analysis produced good fit for a model with one factor consisting of the three positively worded scales and a separate factor for health complaints, but with a high negative correlation between the two factors. Analyses of each of the four scales against gender, age, and 16 other covariates, showed strikingly similar patterns of associations. In an analysis based on a hierarchical model, adjustments for the general mental wellbeing (second-order) factor reduced associations between the first-order factors (one for each scale) and covariates substantially. Latent variable and bifactor modeling confirmed that most of the covariance among all items from all scales combined was captured by one general dimension. Information curve analysis showed that for all scales, the most reliable scores were obtained for participants with below average latent scores.

Conclusion The study indicates that the four scales essentially reflect one underlying dimension. In studies such as HBSC, efforts should be made to use instruments that cover distinctly different aspects of mental health and wellbeing.

*Correspondence:
Leif Edvard Aarø
leea@fhi.no
Full list of author information is available at the end of the article

Aarø et al. Health and Quality of Life Outcomes   (2025) 23:23

Page 2 of 23

Introduction

According to the Constitution of the World Health Organization (WHO) first formulated in 1946, health is "a state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity" [1]. The concept of wellbeing is also at the core of the WHO conceptualization of mental health, which is defined as "a state of mental wellbeing that enables people to cope with the stresses of life, realize their abilities, learn well and work well, and contribute to their community" [2]. The concept of wellbeing is in other words used to define "health" and "mental health".

Despite this conceptual proximity, the relationship between the concepts of health and wellbeing has been described as diverse, complex, and fuzzy [3]. In the context of the present publication, we will not dive deeply into these complexities but keep the use of concepts fairly simple.

Distinctions have been made between subjective and objective wellbeing. Objective wellbeing includes material resources and social attributes [4]. The current study will focus on subjective wellbeing.

An important topic of discussion in the scientific literature is the distinction between a bipolar versus a dual factor model of mental health. The bipolar model describes being mentally ill and mentally healthy as opposite ends of a single continuum. The dual factor model postulates that mental ill health and positive mental health constitute separate factors. This means that individuals can experience high levels of positive mental health even when they are diagnosed with mental illness. Some studies have provided support for the dual factor model and concluded that mental illness and positive mental health are two distinct, but interrelated domains of mental health [5, 6]. Studies on mental distress or depression and mental wellbeing have come to the same conclusion [7, 8].

Keyes, with a focus on adults, has suggested a distinction between languishing and flourishing [9]. To be flourishing means to be filled with positive emotions and to function well socially as well as psychologically. Individuals who are languishing are in a state of incomplete mental health and may typically experience emptiness and stagnation. A third group are those who are moderately mentally healthy, those who are neither languishing nor flourishing. Languishing does not necessarily imply presence of mental illness [9].

In a study of the structure of wellbeing, Gallagher and associates, based on fourteen indicators, identified three interrelated aspects of wellbeing. Hedonic wellbeing included positive affect, negative affect, and life satisfaction. Eudaimonic wellbeing included autonomy, environmental mastery, personal growth, purpose in life, and self-actualization. Social wellbeing included social acceptance, social actualization, social coherence, social contribution, social integration, and positive relations with others [10]. Three different models (one, two, and three factors) were tested with confirmatory factor analysis in two samples. In the youngest sample (mean age 19 years), which is the most relevant one in the context of the present study, the three-factor model obtained only marginally better fit than the one-factor model (RMSEA = 0.061 versus 0.065; CFI = 0.978 versus 0.975) [10].

For the World Health Organization (WHO), the promotion of adolescent wellbeing is a global priority [11]. As in all fields of public health, policies, programs, and practices must be based on evidence. Survey-based research among adolescents represents a major source of knowledge. To produce relevant and valid evidence, survey-based research on adolescents' mental wellbeing must utilize high-quality scales which measure relevant aspects of wellbeing. The use of mental wellbeing scales may serve various purposes, including estimating the overall level of wellbeing, identifying differences between population segments, tracking changes over time, identifying determinants, or evaluating policies and interventions.

Challenges related to scale redundancy in areas like wellbeing research have been discussed in the scientific literature [12]. As Fiske pointed out in a 1982 publication, the risk of measurement overlap is higher for instruments measuring broad constructs than for those measuring more narrow constructs [13]. For scales measuring broad concepts like mental wellbeing, it is particularly important to examine their discriminant validity [12]. While considerable overlap between instruments measuring wellbeing and aspects of wellbeing has been shown in studies among adults [14–16], fewer such studies have been conducted with data collected among adolescents.

In health-related surveys among adolescents, comprehensive questionnaires pose a threat to participation rates and data quality [17]. To keep questionnaires reasonably short, each included scale should measure distinctly different aspects of health, wellbeing, and their determinants. Scale redundancy should be avoided.

A classic approach to examine the distinctiveness and overlap among survey instruments was the multitrait-multimethod matrix, developed by Campbell and Fiske [18]. Currently, the framework of latent variable and structural equation statistical analyses, including the bifactor modeling approach, offer new and promising opportunities for examining scale uniqueness and overlap including discriminant validity [19].

Four scales from the HBSC Study

Several scales and instruments pertinent to mental wellbeing were used in the 2022 HBSC data collection. In the Nordic part of the study, four such scales were employed: Cantril's ladder [20], The WHO-5 Mental Wellbeing Index [21], The Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS) [22], and The HBSC Health Complaints Scale [23, 24]. It may be argued that one of the scales, the HBSC Health Complaints Scale, extends beyond the domain of the wellbeing concept with its negatively worded items. In the context of the present publication, we have chosen to include this scale in our analyses and to regard its items as indicators of low levels of wellbeing. This is consistent with defining negative affect as one of the domains to cover when measuring wellbeing [10].

A number of publications emerging from the HBSC Study describes associations between mental wellbeing outcomes and predictors such as demographic variables. However, these studies typically analyze each scale in isolation [25–28]. Consequently, the extent to which these scales may overlap is therefore rarely investigated. Furthermore, their psychometric properties are usually examined in separate publications within the HBSC network [29–31] as well as beyond [32]. Finally, in studies presenting results based on multiple wellbeing and subjective health indicators, systematic attempts to examine scale redundancy are generally lacking [33].

Purpose of the present study

The aim of the present study is to examine the extent to which the four scales included in the Nordic part of the HBSC 2022 study measure distinctly different aspects of mental wellbeing.

More specifically we will

1. Describe the factor structure of the 21 items constituting all four scales.
2. Based on meanscores (sumscores divided by number of items), examine to what extent the associations of the four scales with gender and age differ across scales.
3. Examine the consistency of the associations of the four scales with a number of relevant covariates beyond age and gender. This includes for instance family affluence, self-esteem, self-efficacy, being bullied, subjective stress, loneliness, and indicators of social support.
4. Estimate a hierarchical second-order factor model with the four scales forming the first-order factors and a second-order factor that explains the correlations between the first-order factors.
5. Examine associations between the first-order factors and covariates after partialling out the covariance of the second-order factor.
6. Test a bifactor model to calculate a series of omega-related and other coefficients which may throw light on the assumption of unidimensionality across all four scales.
7. Estimate test information functions for all items combined and for each scale separately in order to examine to what extent the functions differ across scales.

Methods

Instruments

Scales for the measurement of mental wellbeing included:

(a) The single-item Cantril's Ladder [20]
(b) 5-item WHO Wellbeing Index [21, 34]
(c) 7-item Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS) [22],
(d) 8-item HBSC Health Complaints Scale [23, 24]

The single-item Cantril's ladder is a global measure of life satisfaction. The question goes like this: "Here is a picture of a ladder. The top of the ladder "10" is the best possible life for you and the bottom "0" is the worst possible life for you. In general, where on the ladder do you feel you stand at the moment?" The response scale goes from 0 (worst possible life) to 10 (best possible life) [20].

The WHO Wellbeing Index covers affect, vitality, and taking interest in things [21]. The five items are shown in Table 1. The same set of six response categories that span from "At no time" to "All of the time" are used for all items (coded 0–5). Previous studies have confirmed the unidimensionality of the WHO-5 Wellbeing Index [35–38].

The SWEMWBS covers affective-, functioning-, and social (one item only) aspects of wellbeing. The seven items are shown in Table 1. The response categories span from "Never" to "All of the time" (coded 1–5). Several studies have confirmed that the SWEMWBS is unidimensional [22, 39–46]. Other studies have shown a strong general factor and weak sub-factors [47–49].

The HBSC Health Complaints Scale contains items on psychological and somatic complaints. The eight items are shown in Table 1. The five response categories span from "Rarely or never" to "About every day" (coded 1–5). There are studies which have suggested that the HBSC Health Complaints Scale consists of two, highly correlated dimensions, one somatic and one psychological [23, 50, 51]. Other studies have supported unidimensionality of the HBSC Scale [52, 53]. In one study, unidimensionality was confirmed for 16 out of 46 countries involved

Table 1 Mental wellbeing items – factor loadings from exploratory factor analysis (Principal axis factoring, oblique rotation*, pairwise deletion of cases). Weighted data

Scale	Item	Factor number		
		1	2	3
Life satisfaction	1. Worst to best possible life	.42	-.20	.19
WHO Wellbeing Index	1. Cheerful and in good spirits	.74	-.06	.04
	2. Calm and relaxed	.59	-.06	.11
	3. Active and vigorous	.75	.03	.01
	4. Wake up feeling fresh and rested	.62	-.10	-.02
	5. Daily life filled with things that interest me	.70	.06	.07
HBSC Health Complaints Scale	1. Headache	-.05	-.70	-.03
	2. Stomachache	-.08	-.70	.04
	3. Backache	-.03	-.54	.00
	4. Feeling low	.24	-.57	.05
	5. Irritable	.17	-.61	-.02
	6. Nervous	.07	-.62	.04
	7. Sleeping difficulties	.16	-.47	.00
	8. Dizziness	-.06	-.71	03
Short Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS)	1. Optimistic about the future	.06	.03	.71
	2. Feeling useful	.13	-.01	.72
	3. Feeling relaxed	.20	-.13	.52
	4. Dealing with problems well	.01	-.04	.77
	5. Thinking clearly	-.01	-.06	.78
	6. Feeling close to other people	.10	.03	.63
	7. Able to make up my mind	-.13	.01	.79

* Correlations between factors:

Factors 1 – 2: -.58

Factors 1 – 3: .70

Factors 2 – 3: -.55

Kaiser–Meyer–Olkin Measure of Sampling Adequacy: .956

Bartletts Test of Sphericity: Chi-square = 166 814.007; df = 210; $p < .001$

in the HBSC study. Furthermore, deviations from a one-dimensional structure of HBSC were found to be negligible in most countries [31].

We have not found any studies that examine the dimensionality across the four scales we used in the present study (Medline and APA PsycInfo). Our study may be the first one to examine to what extent all these four scales largely reflect a single, underlying dimension.

For each multi-item scale, simple meanscores (sum-scores divided by number of items) were produced. Gender and age were self-report measures. With regard to gender and age, study participants could only choose between "Boy" and "Girl", and they were asked to report their age in whole years.

Details of the measurement of the covariates are well described in the HBSC Protocol for the HBSC-2022 data collection [54]. All covariates are simple mean- or sum-scores based on the relevant items. An overview of all items of the covariates is shown in Appendix, Table 7.

Descriptives for all covariates are provided in Appendix, Table 8.

Cantril's Ladder and the HBSC Health Complaints Scale are standard HBSC instruments. The WHO Wellbeing Index was introduced as mandatory in 2022. The short form of the Warwick-Edinburgh Mental Wellbeing Scale (SWEMWBS) was only used in the data collections in the Nordic countries in 2022. Also, the Self-esteem scale, which is one of the covariates used in the present study, was only used in the Nordic countries.

Sampling and data collection

A standardized international research protocol was followed to ensure consistency in sampling, survey instruments, data collection, and data processing procedures.

The aim of the sampling procedures was to produce samples of three age groups, 11-, 13-, and 15-year-old school students. For practical reasons, since data collections were administered through schools, students were

Aarø *et al. Health and Quality of Life Outcomes*　　(2025) 23:23

Page 5 of 23

sampled with school classes or schools as the primary sampling unit and only students in grades which corresponded to the defined age groups were included in the study. The recommended national sample size per age group was minimum 1500, and the mean age should be as close as possible to 11.5, 13.5, and 15.5 years.

In Norway and Sweden, school classes were used as primary sampling units. In Finland, the first step was to draw a sample of schools. In the next step, classes were sampled within each relevant grade. In Denmark, the primary sampling unit was schools. Iceland invited all schools in the country and therefore had a larger sample compared to the other countries.

In the HBSC study the primary version of the questionnaire is in English language. Procedures for translation and adaptation to other languages have to be followed by all countries. This includes back-translations from national languages to English and piloting. Questionnaires were made available in the relevant languages. This includes two versions in Finland (Finnish and Swedish) and Norway ("Bokmål" and "Nynorsk"). The questionnaire was not administered in any of the immigrant group languages or the Sami language.

The data collections could take place any day of the week, from Monday to Friday, and during any school hour, but the day and hour had to be the same for all students within one school class. The time needed for responding to the questionnaires was approximately one school hour (45 min) in all countries.

In all countries the teachers followed procedures ensuring anonymity to the students. Oral and written information on the confidentiality of their responses were provided, and participation was confidential and voluntary. In Denmark, Finland, Iceland, and Norway, the students responded to the questionnaire on computers, tablets, or mobile devices in the classroom after receiving instructions from teacher. In Denmark, an instructional video was also shown. In Sweden the data collection was carried out at school (also administered by teachers) with either printed questionnaires or computers. Overall, in Sweden, 56 percent of the students answered the survey online and 44 percent on paper.

Participation rates of eligible students: Denmark, 70%, Finland 74% (estimated number), Iceland 83%, Norway 81%, and Sweden 80%. The proportion of schools or classes that accepted the invitation varied a lot across countries: Denmark 16%, Finland 21%, Iceland 77%, Norway 8%, and Sweden 55%. In Iceland, all schools in the country were invited to participate in the study. Therefore, the number of participating schools and the number of students in the Icelandic sample are disproportionately high.

The HBSC Data Management Centre, located at the University of Bergen, Norway, usually checks the quality of the data, performs appropriate cleaning of the data, and merges national data sets into an international data file. Detailed information about the study and data handling is available at http://www.hbsc.org/. The methodology of the study is described in the HBSC protocol for 2021/22, which prescribes sampling plans, survey instruments, and standards for data collection [54]. The version of the data used in the present study was based on a separate merging of data from five Nordic countries carried out within the context of a Nordic collaboration to which this study belongs [55].

Statistical analyses

Preparations for data analyses included the construction of a common cluster variable (school classes) across all countries. The Finnish data were weighted to have a correct representation of Swedish-speaking school students. In addition, data were weighted to have an equal number of students in each subgroup defined by gender, age, and country, while approximately preserving the total number of observations. The number of observations by gender, age, and country before weighting is shown in Appendix Table 9.

The analyses of data started with descriptives (means, standard deviations, correlations), Cronbach's alpha values, and exploratory factor analyses (with principal axis factoring, oblique rotation, and pairwise deletion of cases) done with SPSS (version 28.0.1.1). Consistent with conventions, factor loadings higher than 0.40 indicate that a specific variable belongs to a factor [30]. The next step was to use confirmatory factor analysis with the WLSMV estimator in Mplus (version 8) (with no restrictions on correlations between factors) in order to test unidimensionality.

The analyses of outcome variables by age, gender, and country as well as the estimation of associations between the four outcome variables and sixteen covariates were carried out with General Linear Modeling (GLM) in the SPSS Complex module. In order to estimate the degree of similarity of associations between mental wellbeing measures and the sixteen covariates across mental wellbeing measures, one correlation of correlations was calculated for each pair of mental wellbeing indicators. High correlations indicate similarity of patterns of associations.

Associations between mental wellbeing indicators and the sixteen covariates were estimated for the general factor in the hierarchical model and for each of the specific factors after adjustment for the general factor. Since all variables were standardized, we have chosen to interpret the size of the associations similar to effect sizes.

Coefficients around 0.20 are small, around 0.50 medium, and around 0.80 large [56].

Latent variable analyses, including estimation of a hierarchical factor model and bifactor modelling, were performed with Mplus. In the hierarchical factor model, first-order factors were created for each of the multi-item scales, and these three factors as well as Cantril's ladder were set to form a second-order Mental Wellbeing factor [57]. The bifactor model was constructed by allowing all 21 items to load on a single general factor and each of the four scales (including Cantril's ladder) to load on specific factors. Intercorrelations between all factors were restricted to zero. In both models, three correlated error terms were added to each of the models in order to improve model fit. Since the specific factor for Cantril's ladder was based on one item only, we used a reliability estimate from a previous study [58] to fix the residual variance to a specific value in order to identify the model (residual variance = (1-reliability)*sample variance).

All analyses were carried out on weighted data, and in all statistical testing and during calculations of confidence intervals, adjustments were made for cluster effects (school classes). The Weighted Least Squares Mean and Variance adjusted estimator (WLSMV) was used in the analyses which included use of latent variables, and standard fit indices (RMSEA, CFI, and TLI) were reported. Various criteria for what constitute good fit have been suggested. In this study good fit is demonstrated when RMSEA ≤ 0.06, CFI ≥ 0.95, or TLI ≥ 0.95 [59, 60].

Several psychometric indices were derived to answer the research questions: Explained Common Variance (ECV) [61], Global Omega (ω), Omega Subscale (ωS) [62], Omega Hierarchical (ωH), Omega Hierarchical Subscale (ωHS) [63], Relative Omega [64], H [65], Factor determinacy [66], Percent of Uncontaminated Correlations (PUC) [19], and Average Relative Parameter Bias (ARPB) [67]. All the omega-coefficients are estimates of reliability based on the factor loadings of factor analysis (common factoring), and are most useful in the context of analysis of latent variables [68].

Explained Common Variance (ECV) for the general factor is the proportion of all common variance explained by that factor. For specific factors, in our context, ECV shows the strength of a specific factor relative to all explained variance only of the items loading on that specific factor [61, 69].

Omega (global omega) is an estimate of reliability which includes the general factor as well as the specific factors combined. Omega S (subscale) includes subscale items only, but with their loadings on the general factor as well as the specific factor included [62]. Omega H (hierarchical) for the general factor is based on loadings on the general factor only. Omega HS (hierarchical specific) for specific factors is based on loadings on each subfactor separately without including loadings on the general factor [63].

Relative Omega is Omega H divided by Omega and applies both to the general factor and to specific factors. For the general factor, Relative Omega shows the proportion of the total reliable variance (general plus specific) that is covered by the general factor. For a specific factor Relative Omega is the proportion of the reliable variance in the subscale that is independent of the general factor [64].

H is a measure of construct replicability and represents the correlation between a factor and an optimally weighted item composite. High H values (H > 0.80) indicate a well-defined latent variable [65].

Factor Determinacy (FD) is the correlation between factor scores and the factors. It is recommended that factor score estimates should only be used when FD > 0.90. Percent Uncontaminated Correlations (PUC) represent the proportion of covariance which only reflects variance from the general dimension [70]. When PUC and ECV values are higher than 0.70, the common variance in the model can be regarded as essentially unidimensional [19].

Average Relative Parameter Bias (ARPB) – an indicator of bias if items are forced into a unidimensional structure – is based on the difference between an item's loading in the unidimensional solution and its general factor loading in the bifactor model, divided by the general factor loading in the bifactor model. An ARPB smaller than 10–15% is acceptable and represents no serious threat to the assumption of unidimensionality [67].

For more information about the coefficients described above, please see Dueber 2017 [68].

The final results presented are based on analyses of information curves. This analysis assumes that the items of all four specific factors reflect a single, underlying latent factor. The information curves describe to what extent each of the four scales as well as the four scales combined (the total information function) are able to distinguish reliably between scores along the whole range of values on the latent factor.

The different statistical techniques applied in the present study serve different purposes. The initial conventional exploratory factor analysis demonstrates the rather simple approach used to analysing dimensionality used in many studies. The confirmatory factor analysis that followed was used in order to specifically test unidimensionality versus multidimensionality. General Linear Modelling is used for the purpose of calculating associations between wellbeing scales and covariates. The hierarchical factor model was used to examine to what extent associations between specific (first-order)

factors and covariates are reduced when adjusting for a general (second-order) factor. Bifactor modeling with its variety of omega-related and other coefficients is ideal for examining dimensionality. And the analysis of information curves shows to what extent the various scales as well as all scales combined provides high levels of information values across all levels of a hypothesized underlying latent variable. Critical reviews of hierarchical and bifactor models have been provided by Markon [71] and DeMars [72].

Results

Initial factor analyses

An exploratory factor analysis of all 21 items in the four scales was conducted (Table 1). Kaiser–Meyer–Olkin Measure of Sampling Adequacy was as high as 0.956. Bartlett's Test of Sphericity obtained significance (Chi-square = 166 814.007; df = 210; $p < 0.001$). An eigenvalue greater than 1.00 was obtained for three factors. The three factors accounted for 58.3% of the variance in the full set of variables.

The variables were sorted into three broad categories with all the items from the WHO Wellbeing Index obtaining high loadings on the first factor. The loadings on this scale varied from 0.59 to 0.75. In addition, the single item on life satisfaction measured with Cantril's ladder obtained a loading higher than 0.40 (0.42) on this factor. All items from the HBSC Health Complaints Scale obtained high negative loadings (in the range −0.47 to −0.71) on the second factor. All items from the SWEM-WBS obtained high loadings (0.52 to 0.79) on the third factor. All other loadings were low, in the area between 0.00 and 0.24 (absolute numbers).

Because oblique rotation of factors was applied, intercorrelations between factors can be reported. The HBSC Health Complaints factor was negatively correlated with the SWEMWBS (r = −0.55) and with the WHO Wellbeing Index (which also contained the single item Cantril's Ladder) (r = −0.58). The correlation between the SWEM-WBS and the WHO Wellbeing Index was 0.70.

Since the correlations between the HBSC Health Complaints factor and the other factors were markedly lower than the intercorrelation between combined Cantril's ladder/WHO Wellbeing Index and the SWEMWBS, a confirmatory two-factor model with three latter scales forming one factor and the HBSC Complaints Scale constituting a second factor was tested. This two-factor model obtained good fit (RMSEA = 0.052; CFI = 0.942; TLI = 0.935). The correlation between the two factors was as high as minus 0.725. A model with all items from all four scales loading on one general factor did, however, not obtain good fit (RMSEA = 0.077; CFI = 0.874; TLI = 0.859). In both models one correlated error term

was included, between "Calm and relaxed" from the WHO Wellbeing Index and "Feeling relaxed" from the SWEMWBS.

Simple meanscores (sumscores divided by number of items) were constructed for each of the multi-item scales. Descriptives and Cronbach's alpha values for the four mental wellbeing indicators are shown for each country as well as for all countries combined in Appendix, Table 10. All scales had high alpha values, ranging from 0.826 to 0.912. Intercorrelations between sumscores ranged from 0.55 to 0.67 (absolute numbers) (Table 11 in Appendix). The correlations are sufficiently high to indicate some common, underlying dimension, but not sufficiently high to eliminate the possibility that they also measure distinctly different aspects of mental wellbeing.

Associations with age and gender

The analyses presented in this section were carried out with the General Linear Modeling (GLM) module in SPSS Complex. When analysing the scales which had data for all three age groups across all five countries against gender, age, and country, some common patterns emerged (full set of diagrams can be obtained from lead author). Mean complaints scores were distinctly higher for girls than boys in all countries (effect sizes 0.58 to 0.73). Mean scores on Cantril's Ladder were higher among boys than girls in all countries (effect sizes 0.32 to 0.43). This was also the case for the WHO Wellbeing Index (effect sizes 0.42 to 0.62) and the SWEMWBS (effect sizes 0.42 to 0.59).

Outcome variables by gender and age for all countries combined are shown in Fig. 1. Due to incomplete data across countries and gender, results for the SWEMWBS are not shown. For all three remaining outcomes, there were significant gender by age interactions. Significance tests shown under Fig. 1. Mean scores on the HBSC Health Complaints Scale increased more strongly with age among girls (E.S. $_{oldest\ vs.\ youngest}$ = 0.58) than among boys (E.S. $_{oldest\ vs.\ youngest}$ = 0.14). For the positively phrased mental wellbeing scales, the mean scores decreased over age groups, but more strongly among girls than boys. This was the case for the single item Cantril's Ladder scale (E.S. $_{oldest\ vs.\ youngest}$ = 0.45 for girls and E.S. $_{oldest\ vs.\ youngest}$ = 0.31 for boys) as well as for the WHO Wellbeing Index (E.S. $_{oldest\ vs.\ youngest}$ = 0.49 for girls and E.S. $_{oldest\ vs.\ youngest}$ = 0.18 for boys).

Among girls, changes over age were not linear. On the HBSC Health Complaints Scale, the mean score increased more strongly between age 11 and age 13 than between age 13 and age 15. Also, this finding was mirrored in the positive measures of mental wellbeing. The decreases in mean scores among girls were stronger between age 11 and age 13 than was the case between

13 and 15. Testing of deviations from linearity across age groups revealed significance among girls for all three outcomes ($p < 0.001$). For boys, deviations from linearity were less strong and significant at the $p < 0.001$ level for Cantril's ladder only.

The differences in mean scores across countries were distinctly different for the different scales. No consistent pattern could be observed (in Appendix, Fig. 5). This clearly contrasts the high level of consistency when examining differences across gender and age groups.

Not all the four wellbeing scales were administered to all age groups in all countries. An overview of coverage across age groups and countries is shown in Appendix, Fig. 6.

Associations of the four scales with selected covariates

To further describe the external consistency between the four scales measuring mental wellbeing, associations with sixteen covariates were estimated (Table 22). Adjustments were made for gender, age, country, and gender by age interactions. Since all variables are standardized, all coefficients can be interpreted as correlations.

All coefficients, except one which was significant at the $p < 0.01$ level, were significant at the $p < 0.001$ level.

Family Affluence (FAS) was only weakly associated with the four scales with coefficients varying from |.026| to |.112| (absolute values). Another indicator of socioeconomic status showed stronger associations with the four scales, with coefficients varying from |.209 to 0.299|.

The highest coefficients across all four scales were observed for self-esteem, subjective health, liking school, a number of support variables, loneliness, and high as well as low stress.

In the context of the present study, the most important observation is that all associations between the three indicators of positive mental wellbeing and covariates showed similar patterns of variation across covariates. This was also the case for the HBSC Health Complaints Scale, but all correlations had opposite directions when compared with the positive indicators.

Table 3 shows correlations between the columns in Table 2, in other words, correlations based on correlations. They vary between |.979| and |.997| (absolute numbers). This demonstrates that the patterns of associations

Table 2 Mental wellbeing with selected covariates, adjusted for gender, age, country and gender age interaction. Weighted data and with adjustments for cluster effects. All scale variables standardized

	Life satisfaction (Cantril's ladder)	WHO Wellbeing Index [a]	Mental wellbeing (SWEMWBS) [b]	HBSC Health Complaints
Family well off [c]	.299***	.273***	.248***	-.209***
Self-esteem [d]	.539***	.548***	.612***	-.459***
Self-efficacy	.340***	.381***	.590***	-.292***
Family affluence (FAS)	.099***	.100***	.112***	-.026**
Global subjective health	.478***	.453***	.434***	-.382***
Been bullied	-.235***	-.198***	-.230***	.243***
Likes school	.416***	.436***	.406***	-.357***
Schoolwork pressure	-.290***	-.319***	-.326***	.357***
Family support	.359***	.323***	.444***	-.288***
Classmate support	.340***	.371***	.364***	-.299***
Teacher support	.340***	.365***	.384***	-.306***
Peer support	.303***	.304***	.394***	-.230***
Stressful social media use	-.221***	-.253***	-.275***	.298***
Loneliness	-.467***	-.456***	-.483***	.458***
High stress [e]	-.399***	-.407***	-.380***	.420***
Low stress [e]	.407***	.407****	.565***	-.348***

** $p < .01$

*** $p < .001$

[a] Data not available for 11-year-olds in Denmark

[b] Data not available for Iceland nor for 11-year-olds in Denmark, Finland, Norway

[c] Data not available for Denmark

[d] Data not available for Iceland, for Norwegian 11- and 13-year-olds, nor for Swedish 11-year-olds

[e] Data not available for 11-year-olds in Denmark, Finland, Iceland

Table 3 Correlations between columns in Table 2

	Life satisfaction (Cantril's ladder)	WHO Wellbeing Index	Emotional Wellbeing (SWEMWBS)	HBSC Health Complaints Scale
Life satisfaction (Cantril's ladder)	1.00			
WHO Wellbeing Index	.997	1.00		
Emotional Wellbeing (SWEMWBS)	.979	.984	1.00	
HBSC Complaints	-.996	-.997	-.979	1.00

are remarkably similar across the four mental wellbeing scales.

Not all the sixteen covariate measures were administered to all age groups in all countries. An overview of coverage across age groups and countries is shown in Appendix, Fig. 7. Only covariates with incomplete coverage across subgroups defined by age and country are included in the figure.

Hierarchical second-order factor model

A confirmatory hierarchical model where single items were grouped into four first-order factors and the first-order factors were allowed to load on a second-order "Mental wellbeing" factor was estimated (Fig. 2). With three correlated error terms, the model obtained a good fit (RMSEA = 0.038; CFI = 0.969; TLI = 0.964). All four first-order factors obtained high loadings on the second-order factor varying from 0.74 to 0.90 (absolute values).

The correlated error terms occurred between WHO item 2 (calm and relaxed) and SWEMWBS item 3 (relaxed); between HBSC Item 1 (Headache) and HBSC items 2 (Stomachache) and 8 (Dizziness).

Associations between the first-order factors and the sixteen covariates after adjustments for the second-order factor are shown in Table 4. When adjusting for the second-order factor, the first-order factors loose most of their associations with covariates. Only two of these associations are stronger than 0.20. The association between self-efficacy and the SWEMWBS (coefficient = 0.314) is probably a product of the self-efficacy-related items

Table 4 General mental wellbeing factor and specific factors (adjusted for the general factor, one by one) on various predictors. Regression coefficients. All variables standardized. Based on the hierarchical factor model shown in Fig. 2

Predictor	General	Spes1 Cantril's Ladder	Spes 2 WHO Wellbeing Index	Spes 3 Mental Wellbeing (SWEMWBS)	Spes 4 HBSC Health Complaints Scale
	Bivariate	Adjusted for the general factor, one by one			
Family well off	.341***	.088***	.009 n.s	-.013 n.s	.033***
Self esteem	.756***	.010 n.s	-.088***	.130***	.055***
Self-efficacy	.533***	-.074***	-.119***	.314***	.131***
Family affluence (FAS)	.114***	-.003 n.s	.015 n.s	.066***	.066***
Global subjective health	.593***	.106***	-.016 n.s	-.024*	.011 n.s
Been bullied	-.272***	-.025***	.073***	-.004 n.s	.051***
Likes school	.518***	.042***	.008 n.s	-.045***	-.009 n.s
Schoolwork pressure	-.507***	.063***	.109***	.043***	.151***
Family support	.473***	.038***	-.108***	.145***	.042***
Classmate support	.498***	.000 n.s	.001 n.s	-.020 n.s	-.016*
Teacher support	.491***	.012*	-.027**	.021*	-.002 n.s
Peer support	.372***	.030***	-.032***	.131***	.086***
Stressful social media use	-.394***	.051***	.081***	.011 n.s	.099***
Loneliness	-.669***	-.005 n.s	.137***	.044***	.141***
High stress	-.579***	.003 n.s	.079***	.073***	.144***
Low stress	.608***	-.034***	-.173***	.247***	.097***

n.s. Not significant, * *p* < .05, ** *p* < .01, *** *p* < .001

included in the SWEMWBS: "Dealing with problems well", "Thinking clearly", and "Able to make up my mind". The adjusted association between "Low stress" and the SWEMWBS (coefficient = 0.247) can also be explained by item overlap between the scales, such as between "*Confident about ability to handle personal problems*" and "*Dealing with problems well*" for Low Stress and SWEMWBS, respectively. The remaining adjusted associations are too small that they would require any closer interpretations.

Bifactor model

A bifactor model with all 21 items loading on a general factor and each group of items (one group for each scale) loading on specific factors is shown in Table 5 and Fig. 3. Since all HBSC Health Complaints items were reversed, all loadings are positive. After allowing for three correlated error terms, the model obtained a good fit (RMSEA = 0.031; CFI = 0.982; TLI = 0.977). The highest correlation between error terms was the same as in the Hierarchical second-order factor model, WHO scale item 2 with SWEMWBS item 3 (0.047). The other two were between SWEMWBS items 1 and 3 (0.033) and HBSC items 4 and 5 (0.027).

Omega and other coefficients are shown in Table 6. In the context of the present study, the most important findings are a high Omega Hierarchical for the general factor (Omega H = 0.834) and relatively low Omega HS values for the specific factors (ranging from 0.106 to 0.369). Furthermore, PUC is 0.719 and ECV is 0.689. When Omega H is higher than 0.800, total scores can be considered essentially unidimensional [63]. As previously mentioned, when PUC and ECV values are higher than 0.700, the common variance in the model can be regarded as essentially unidimensional [19]. In our case, PUC is slightly above and ECV marginally under the critical value. ARPB (here calculated as the average of the absolute relative parameter bias) is 0.114. ARPB values in the interval 0.10–0.15 or lower are indicative of unidimensionality [67, 73].

Information curves

Results of the information function analyses are shown in Fig. 4. The total information function peaks between −1 and −2 standard deviation below the population mean. This is where all items together provide most of the information and the standard error of measurement is smallest. That is, more precise measurements are obtained for

Table 5 Bifactor modeling* of four scales for the measurement of mental wellbeing (*n* = 27 364). YX-standardized factor loadings. WLSMV estimator. RMSEA = .031; CFI = .982; TLI = .977. Weighted data. Intercorrelations between all factors (including Cantril's ladder) restricted to zero

		General factor	Single variable	Specific factor 2	Specific factor 3	Specific factor 4
Cantril's Ladder	1	.738	.326			
WHO Wellbeing Index	1	.788		.302		
	2	.691		.272		
	3	.659		.463		
	4	.641		.297		
	5	.636		.390		
Mental Wellbeing Scale (SWEMWBS)	1	.634			.400	
	2	.741			.384	
	3	.737			.298	
	4	.696			.498	
	5	.696			.494	
	6	.610			.391	
	7	.517			.543	
HBSC Health Complaints Scale	1	.438				.573
	2	.474				.549
	3	.375				.444
	4	.719				.330
	5	.615				.391
	6	.593				.424
	7	.535				.312
	8	.514				.577

* In order to obtain good fit, <u>three error terms</u> had to be included in the model: SWEMWBS item 1 with item2 (standardized coeff. = .326); SWEMWBS item3 with WHO item 2 (standardized coeff. = .466), and HBSC items 4 and 5 (standardized coeff. = .288)
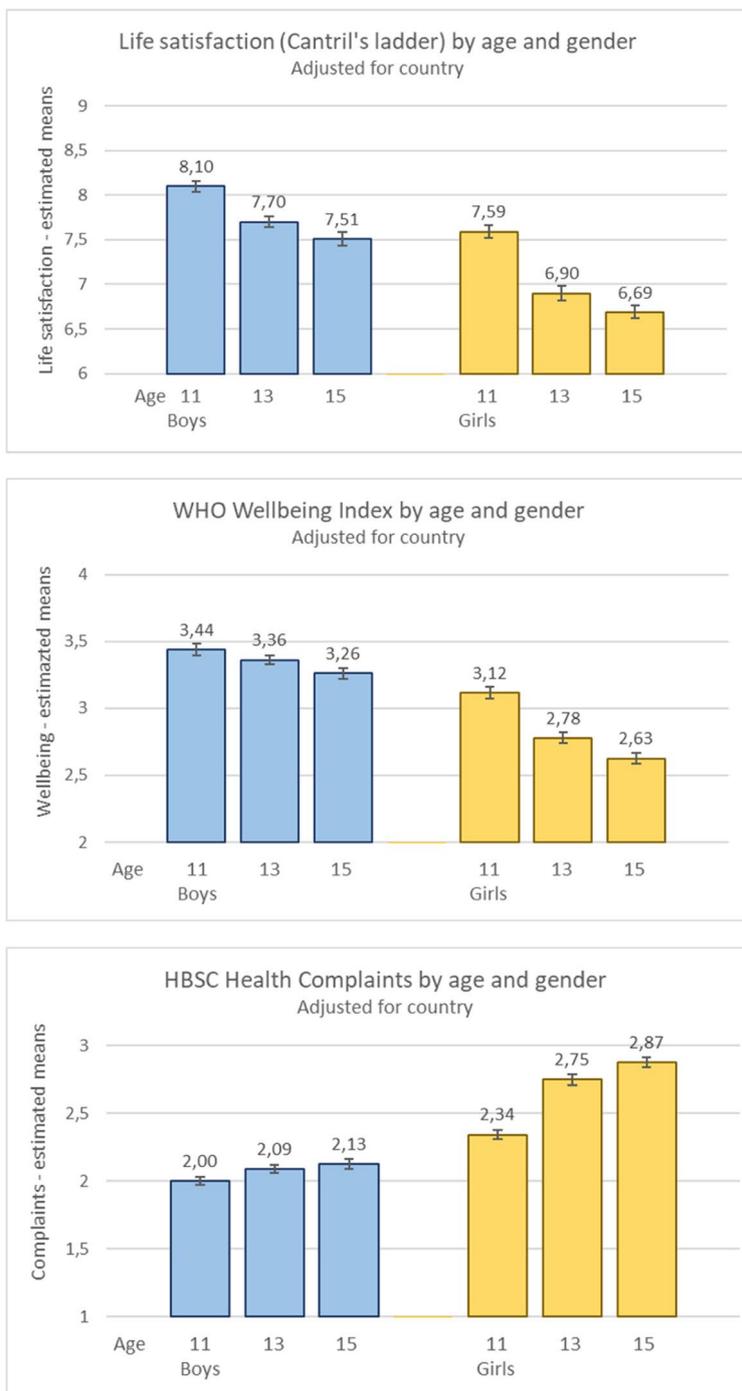
Fig. 1 Indicators of mental wellbeing by age and gender, all Nordic countries combined. Weighted data and adjustment for cluster effects. No data available on the WHO Wellbeing Index for 11-year-olds in Denmark. Gender by age interactions: Cantril's ladder: Wald F = 13.747; df1 = 2; df2 = 1324; $p < .001$. WHO Wellbeing Index: Wald F=32.671; df1=2; df2=1204; p<.001. HBSC Health Complaints Score: Wald F=100.614; df1=2; df2=1324; p<.001

those with below average mental wellbeing scores rather than above average mental wellbeing scores.

The partial information curves for the SWEMWBS and the WHO Wellbeing Index follow a pattern similar to the total curve, with the SWEMWBS providing somewhat more information than the WHO Index. This might be due to the SWEMWBS having two more items. The reliability of a scale tends to increase with increasing number
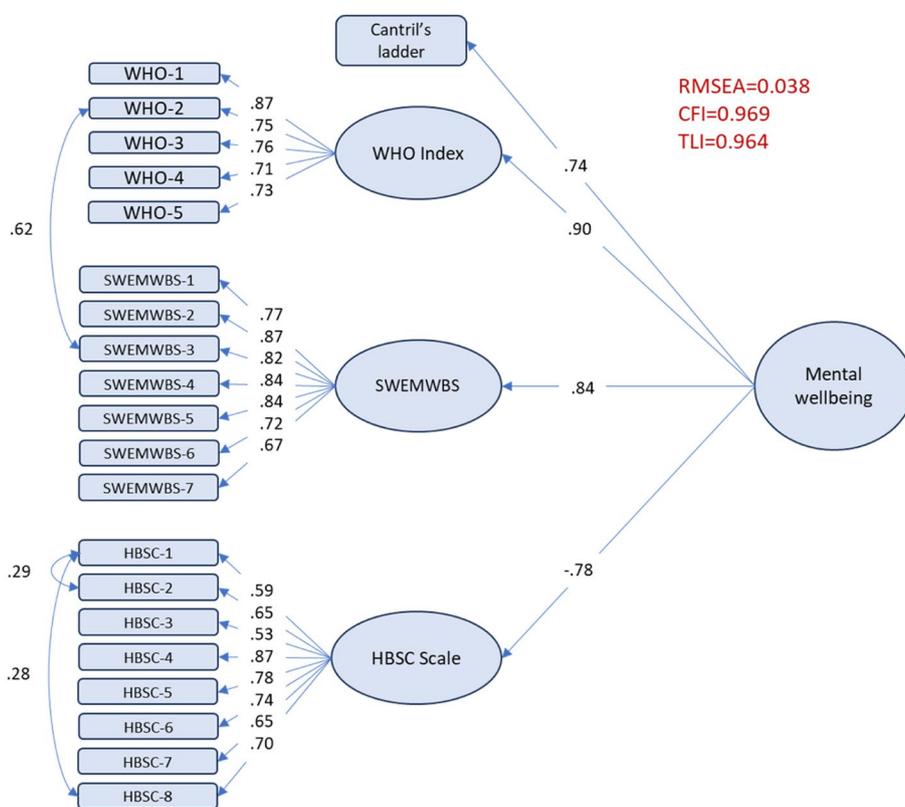
Fig. 2 Indicators of mental wellbeing in a hierarchical measurement model

of items [74]. The partial information curve of the HBSC Health Complaints Scale is both narrower and lower compared to the SWEMWBS and the WHO Wellbeing Index despite being the instrument with most items. As expected, since it only consists of a single item, Cantril's ladder provides the least information of all measures. Although this single measure also shared most of its reliable variance with the general factor and may as such be a valid measure of mental wellbeing, it may not necessarily obtain sufficient precision.

And finally, the two multi-item positive mental wellbeing scales (SWEMWBS and WHO-5) both show narrow peaks in the information curves at values around plus 1.5 standard deviations, similar to the total curve. The HBSC Health Complaints Scale has no such extra peak on the positive side of zero. This indicates that, after all, the multi-item wellbeing scales provide a small portion of extra precision on the positive side of the zero.

Discussion
Analyses of three scales (those with the most complete data for age groups) showed a high level of consistency in their associations with age and gender. The association of all four scales with country showed no similar

consistency. The inconsistent variation of mean scores on the mental wellbeing scales across countries may reflect problems in ensuring high comparability of scales across languages. Such inconsistencies may easily influence overall prevalences and means, but do not necessarily have noticeable impact on associations with demographic variables and covariates. Analyses of the three mental wellbeing meanscores and Cantril's Ladder against 16 covariates available in the HBSC Nordic data from 2022 showed a remarkable degree of consistency of associations across the scales.

Hierarchical modeling which included four first-order factors and one second-order factor, provided support for a strong general factor. Associations between the first-order factors and sixteen covariates were generally low after adjustments for the second-order factor.

Coefficients derived from the bifactor model analysis to a high extent support the hypothesis that the 21 items of the four scales reflect one underlying dimension. However, from a theoretical and conceptual perspective, the items are supposed to measure different aspects of mental wellbeing. Cantril's ladder is meant to measure global life satisfaction. The WHO Wellbeing Index items cover positive mood, vitality, and taking an interest in things. The items of the SWEMWBS can be described
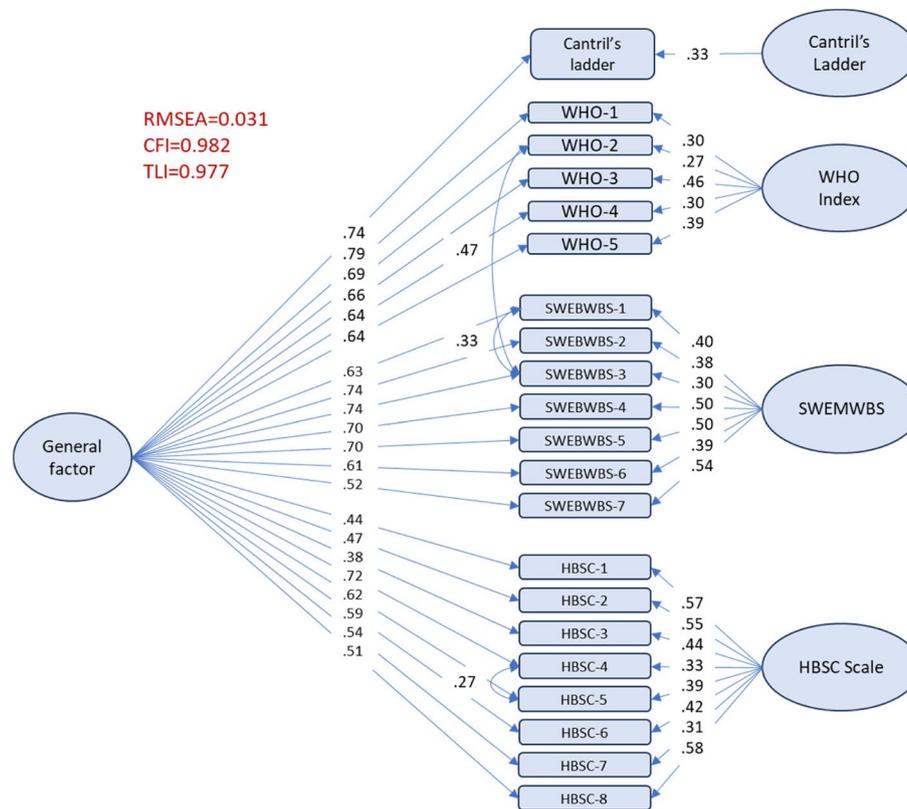
Fig. 3 Indicators of mental wellbeing in a bifactor measurement model

Table 6 Omega and related coefficients for the model presented in Table 5. Explanations in footnotes based on Dueber, 2017 [68]

	General factor	S1: Cantril's ladder	S2: WHO Wellbeing Index	S3: Mental Wellbeing (SWEMWBS)	S4: HBSC Health Complaints Scale
ECV[1]	0.689	0.009	0.051	0.110	0.140
ECV (new)[2]		0.163	0.209	0.301	0.419
Omega/Omega S[3]	0.956	0.651	0.878	0.923	0.887
Omega H/Omega HS[4]	0.834	0.106	0.178	0.274	0.369
Relative Omega[5]	0.872	0.163	0.203	0.297	0.416
H[6]	0.939	0.106	0.422	0.631	0.696
FD[7]	0.941	0.449	0.679	0.798	0.834
PUC[8]	0.719				
ARPB[9]	0.114				

[1] ECV – Explained Common Variance

[2] ECV (new) – strength of a specific factor relative to all explained variance based on items loading on that factor

[3] Omega/Omega S – Internal reliability of the multidimensional composite

[4] Omega H/Omega HS – Omega for the general factor only and for subscales after partitioning out variability explained by the general factor

[5] Relative Omega—Omega H/Omega and Omega HS/Omega S

[6] H – Correlation between a factor and an optimally weighted item composite

[7] FD – Factor determinacy – correlation between factor scores and the factors

[8] PUC – Percent Uncontaminated Correlations represents the percentage of covariance terms which only reflects variance from the general dimension

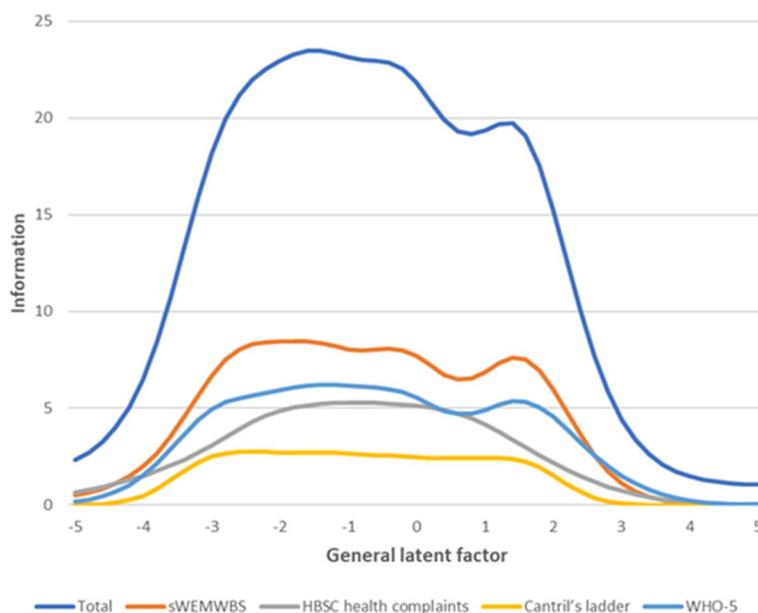[9] ARPB – Average Relative Parameter Bias

Fig. 4 Test information function for all items combined and partial test information functions for the items of the SWEMWBS, HBSC Health Complaints Scale, Cantril's ladder, and WHO Wellbeing Index. This analysis is based on the bifactor model shown in Fig. 3. The general latent factor is standardized with mean = 0 and SD = 1

as reflections of positive affect, positive functioning, and satisfying interpersonal relationships.

The HBSC Health Complaints Scale is different from the other three scales by containing items with negative content. Furthermore, this scale includes four items covering somatic symptoms. It may be argued that somatic pain clearly has negative emotional aspects. Still these items represent a domain beyond what is usually meant by measures of wellbeing. The other four items in the HBSC Health Complaints Scale are not about somatic complaints but cover psychological distress (including aspects of depression and anxiety). Studies among adults indicate that scales for the measurement of wellbeing or subjective quality of life to a large extent measure the same underlying factor as scales measuring depression or distress [75, 76].

Empirical overlap between scales can be explained in different ways. Firstly, there may be a conceptual overlap, and secondly there can be problems with operationalization including questionnaire construction. Since the scales measure closely related concepts such as positive emotional mood, positive affect, positive functioning and (absence of) psychological health complaints, conceptual overlap may to some extent explain the empirical overlap between the scales.

The lack of discriminant validity shown in the present study may also be related to questionnaire construction. The two scales that show tendencies to deviate from unidimensionality, the SWEMWBS and the HBSC Health

Complaints Scale, contain mixtures of affective and other items. In the HBSC Health Complaints Scale, clearly affective items on psychological complaints are combined with items on somatic complaints. In the SWEMWBS, affective items are combined with items on cognitive functioning and closeness to others. These combinations of items within scales may lead to artificially high intercorrelations between scales [77, 78]. Within each scale, the affective items may influence responses to the other items. If the non-affective items were removed and presented in separate scales, the non-affective items might have proven to constitute separate factors more clearly.

Lack of discriminant validity may occur when relying on measurement methods (i.e. questionnaires) that may not be suitable to capture the theoretically distinct aspects of mental wellbeing. Participants tend to complete survey items quickly and intuitively, without much cognitive consideration. That is, participant responses to these types of subjective items may largely be informed by affective processes rather than based on thorough cognitive reflections. As such, items on life satisfaction, feeling useful or being calm and relaxed may activate a similar intuitive response driven by a persons' current and/or recent affective state. So even though these items may semantically and conceptually be different, they largely appeal to the same affective state and serve as such as interchangeable indicators of mental wellbeing, rather than separate constructs that can reliably be distinguished from each other by means of questionnaires.

Some of the problems with scale redundancy may stem from unclear and multi-aspect definitions of mental health and wellbeing. This fosters composite indices like the fourteen items version of the WEMWBS, which tries to cover multiple dimensions of mental wellbeing. Even if such an index appears unidimensional, it may in fact consist of different aspects that are causally related. In that way composite indices may blur instead of clarifying the relationships between different aspects of mental health and wellbeing.

Cantril's ladder, as a measure of global life satisfaction, stands out as conceptually different from the other three scales, and the way it is worded appears adequate and different from the wording of items in all the other scales. Still, it does not come out as a separate factor in the initial factor analysis but loads on the WHO-5 Wellbeing Index factor. In the hierarchical factor analyses, it loses almost all its correlation with covariates when adjustments are made for the general Mental Wellbeing factor. And in the bifactor modeling it loses most of its reliability when adjustments are made for the general factor. In this case, conceptual clarity and seemingly adequate operationalizations appear to be combined with strong empirical overlap, but as outlined above this may be explained by the inherent limitations of questionnaires as a measurement method. It would be interesting to examine whether changes in measurement situation and/or changes in the questionnaire itself could make the included instruments more distinguishable and less redundant. In the case of Cantril's ladder, one could for example try to prime participants to consider more specific aspects of their life such as work, income, and relationships before asking them to rate their overall satisfaction with life.

Despite the rather clear confirmation of one single underlying dimension across all 21 mental wellbeing items, there are some indications of minor deviations from unidimensionality. Omega HS is generally low, but slightly higher for the HBSC Health Complaints Scale (0.369). The highest loadings on this specific factor are found on the somatic items. These small deviations from unidimensionality indicate that the somatic complaints items capture something more than just mental wellbeing.

Even in the absence of discriminant validity, applying more than one scale for measuring mental wellbeing may prove to make sense. This would be the case if the information functions had markedly different shapes. If wellbeing scales demonstrated higher precision in intervals on the positive side of the latent wellbeing scale, while distress scales functioned better on the negative side, combining these scales might provide adequate precision of measurement on a broader range of values. This is, however, not confirmed. The two positively worded multi-item scales obtain higher information values than the HBSC Health Complaints scale across almost all values of the underlying latent variable. And they obtain the highest information values below the mean of the latent general wellbeing factor. The latter finding challenges the idea that positively phrased wellbeing scales measure something different from negatively worded "deficit-oriented" scales. If they did, we would expect their highest information values to appear on higher values of the latent wellbeing dimension.

As shown in this study, exploratory factor analysis carried out with conventional statistical tools results in a three-factor model with highly correlated factors. Most researchers would be satisfied with this analysis and proceed with analysis of data on the premise that a clear distinction can be made between these factors. High correlations between factors would probably not change the way the data were analyzed and reported. Our conclusion is that the underlying general factor is so strong that the advantages of analyzing each of the three positively phrased mental wellbeing scales separately are limited. There is also considerable overlap between these three scales and the HBSC Health Complaints Scale.

The purpose of this study was rather pragmatic, to examine scale distinctiveness and redundancy. The results reported may still have some relevance to the discussion on single versus dual factor models of mental health. Although confirmatory factor analyses obtained good fit for a two-factor model with the items of the three positively worded wellbeing scales loading on one factor and the HBSC Health Complaints Scale items loading on a second factor, the two factors were highly (negatively) correlated. And subsequent analysis using hierarchical and bifactor modeling approaches generally provided support for unidimensionality. These findings, supporting unidimensionality, are in line with the languishing-flourishing dimension described by Keyes [9] and consistent with results from the undergraduate student part of a study in the United States which showed good fit for a single general wellbeing factor [10].

There are several possible explanations for the discrepancy between studies supporting the dual factor model [5–8] and the present study. Firstly, while most previous studies are based on studies among adults, our study is based on data collected among adolescents. Distinctions between aspects of mental health and wellbeing may become more pronounced with age. Secondly, in the present study, only one scale contained negatively worded items. Other scales for the measurement of ill mental health may have produced different results. Thirdly, we have applied statistical approaches which are less frequently used, such as bifactor modeling with related coefficients. There is, after all, a discrepancy between the

initial, conventional exploratory factor analysis which appeared to produce three distinct factors as well as the confirmatory factor analysis which supported two factors, and the subsequent analyses. And finally, in our study we had no access to diagnose-based measurement of psychiatric illness. If such measurements had been available, we might have been able to distinguish between a flourishing-languishing dimension and a separate psychiatric disease dimension similar to what was done by Keyes [9].

Conclusion

The present study provides some evidence for limited discriminant validity among four scales intended to measure different aspects of mental wellbeing and distress used in the Nordic data collections of the HBSC study for 2022. Three of the scales, Cantril's Ladder, the HBSC Health Complaints Scale, and the WHO Wellbeing Index, are part of the standard set of scales now used by all countries in the HBSC. One of the scales, The Short Warwick-Edinburgh Mental Wellbeing Scale, was used in the Nordic countries only. Therefore, the findings from the present study are particularly relevant for the Nordic countries, but also relevant for any research project planning to use more than one scale for measuring mental wellbeing and distress among adolescents.

The strong overlap between the three positively worded scales for measuring mental wellbeing is well documented in the present study. The results with regard to the HBSC Health Complaints scale are less consistent. The initial factor analyses indicate that this scale stands out as somewhat different from the other scales. Subsequent analyses, however, indicate considerable overlap with the other scales.

It is beyond the scope of this paper to provide specific recommendations with regard to which scales should be used in future HBSC data collections. Considerations beyond those addressed in this paper need to be taken. A long history of use of a specific scale (like the HBSC Health Complaints Scale) is a good reason to continue using the scale. This in order to permit future analysis of long-time trends. Problems with scales beyond those described in this paper may also need to be considered, for instance high frequencies of straightliners, which have been documented for the SWEMWBS [17].

Within HBSC, there is obviously a need to examine discriminant validity and redundancy of scales beyond the five Nordic countries. Decisions about revision of scales and selection of scales cannot be based on results from one single study carried out with data from a few countries only.

Appendix

Table 7 Covariates - overview of questionnaire items

Covariate	Items	Response categories
Family well off	How well off do you think your family is?	Very well off; Quite well off; Average; Not so well off; Not at all well off
Self-esteem	I like myself	Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree
	I am very happy being the way I am	
	Other people my age generally like me	
Self-efficacy	How often do you find a solution to a problem if you try hard enough?	Never; Rarely; Sometimes; Most of the time; Always
	How often do you manage to do the things that you decide to do?	
Family Affluence Scale (FAS III)	Does your family own a car, van or truck?	No; Yes, one; Yes, two or more
	Do you have your own bedroom for yourself?	No; Yes
	How many computers do your family own (including laptops and tablets, not including game consoles and smartphones)?	None; One; Two; More than two
	How many bathrooms (rooms with a bath/shower or both) are in your home?	None; One; Two; More than two
	Does your family have a dishwasher at home?	No; Yes
	How many times did you and your family travel out of [insert country here] for a holiday/vacation last year?	None; Once; Twice; More than twice
Global subjective health	Would you say your health is …	Excellent; Very good; Good; Fair; Poor
Been bullied	How often have you been bullied at school in the past couple of months?	I have not been bullied at school in the past couple of months; It has happened once or twice; 2 or 3 times a month; About once a week; Several times a week
Likes school	How do you feel about school at present?	I like it a lot; I like it a bit; I don't like it very much; I don't like it at all
Schoolwork pressure	How pressured do you feel by the schoolwork you have to do?	Not at all; A little; Some; A lot
Family support	My family really tries to help me	*At the ends of a series of categories numbered from 1 to 7:* Very strongly disagree ….. Very strongly agree
	I get the emotional help and support I need from my family	
	I can talk about my problems with my family	
	My family is willing to help me make decisions	

Covariate	Items	Response categories
Classmate support	The students in my class(es) enjoy being together	Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree
	Most of the students in my class(es) are kind and helpful	
	Other students accept me as I am	
Teacher support	I feel that my teachers accept me as I am	Strongly agree; Agree; Neither agree nor disagree; Disagree; Strongly disagree
	I feel that my teachers care about me as a person	
	I feel a lot of trust in my teachers	
Peer support	My friends really try to help me	*At the ends of a series of categories numbered from 1 to 7:* Very strongly disagree ….. Very strongly agree
	I can count on my friends when things go wrong	
	I have friends with whom I can share my joys and sorrows	
	I can talk about my problems with my friends	
Stressful social media use	During the past year, have you …	
	… regularly found that you can't think of anything else but the moment that you will be able to use social media again?	No; Yes
	… regularly felt dissatisfied because you wanted to spend more time on social media?	
	… often felt bad when you could not use social media?	
	… tried to spend less time on social media, but failed?	
	… regularly neglected other activities (e.g. hobbies, sport) because you wanted to use social media?	
	… regularly had arguments with others because of your social media use?	
	… regularly lied to your parents or friends about the amount of time you spend on social media?	
	… often used social media to escape from negative feelings?	
	… had serious conflict with your parents, brother(s) or sister(s) because of your social media use?	
Loneliness	During the past 12 months, how often have you felt lonely?	Never; Rarely; Sometimes; Most of the time; Always.

Covariate	Items	Response categories
High stress	How often have you felt that you were unable to control the important things in your life?	Never; Almost never; Sometimes; Fairly often; Very often
	How often have you felt difficulties were piling up so high that you could not overcome them?	
Low stress	How often have you felt confident about your ability to handle your personal problems?	Never; Almost never; Sometimes; Fairly often; Very often
	How often have you felt that things were going your way?	

Table 8 Covariate descriptives. Weighted data

	N of items	Scale range	Mean	S.D.	Alpha	N (weighted)
Family well off [a]	1	1-5	1.84	0.81	n.a.	21 203
Self-esteem [b]	3	1-5	2.29	0.97	.87	15 877
Self-efficacy	2	1-5	3.71	0.76	.72	26 530
Family affluence (FAS)	6	0-13	8.79	1.88	n.a.	26 254
Global subjective health	1	1-4	1.83	0.72	n.a.	27 290
Been bullied	1	1-5	1.36	0.86	n.a.	26 192
Likes school	1	1-4	2.13	0.83	n.a.	26 695
School-work pressure	1	1-4	2.48	0.93	n.a.	26 653
Family support	4	1-7	5.87	1.49	.94	25 955
Classmate support	3	1-5	2.17	0.82	.83	26 429
Teacher support	3	1-5	2.04	0.91	.89	26 384
Peer support	4	1-7	5.44	1.59	.94	25 866
Stressful social media use	9	1-2	1.22	0.25	.79	23 151
Loneliness	1	1-4	2.34	1.01	n.a.	26 967
High stress [c]	2	1-5	2.58	0.95	0.64	21 410
Low stress [c]	2	1-5	3.44	0.88	0.66	21 397

n.a. – not applicable (single item or not a reflective measurement model)

[a] Data not available for Denmark

[b] Data not available for Iceland, for Norwegian 11- and 13-year-olds, nor for Swedish 11-year-olds

[c] Data not available for Iceland, for Norwegian 11- and 13-year-olds, nor for Swedish 11-year-olds

Table 9 Number of observations by gender, age, and country before weighting of data. Those with missing on gender (2.3%) are excluded. After weighting of the data, *n*=918 in each cell

Country	Age	Gender	
		Boys	**Girls**
Denmark	11	1034	1057
	13	1062	1028
	15	812	830
Finland	11	703	708
	13	693	698
	15	472	509
Iceland	11	1946	1826
	13	1697	1520
	15	1465	1392
Norway	11	764	703
	13	622	615
	15	497	477
Sweden	11	766	742
	13	746	778
	15	712	671

Table 10 Descriptives of meanscores (sumscores divided by number of items) and the single item Cantril's Ladder by country. Weighted data

	Country	Scale range	Mean	Standard deviation	Cronbach's alpha	Nª (weighted)
Life satisfaction (Cantril's ladder)	Denmark		7.39	1.83		5454
	Finland		7.73	1.82		5470
	Iceland		7.45	1.89		5445
	Norway		7.31	1.84		5466
	Sweden		7.20	1.96		5396
	All	0-10	7.41	1.88	---	27 231
WHO Wellbeing Index[b]	Denmark		3.12	0.90	0.826	3626
	Finland		3.06	1.02	0.862	5416
	Iceland		3.06	1.02	0.850	5260
	Norway		3.10	0.98	0.826	5424
	Sweden		3.05	1.09	0.843	5400
	All	0-5	3.07	1.01	0.842	25 125
Mental wellbeing (SWEMWBS)[c]	Denmark		3.80	0.66	0.883	3463
	Finland		3.47	0.80	0.912	3590
	Iceland		---	---	---	---
	Norway		3.43	0.81	0.902	3565
	Sweden		3.65	0.78	0.891	5354
	All	1-5	3.59	0.78	0.901	15 973
HBSC Health Complaints Scale	Denmark		2.21	0.88	0.843	5441
	Finland		2.33	0.93	0.876	5440
	Iceland		2.52	0.95	0.854	5244
	Norway		2.26	0.87	0.843	5421
	Sweden		2.52	0.90	0.847	5448
	All	1-5	2.37	0.92	0.855	26 994

Correlations between factors:

[a] The sum of number of observations (weighted) across countries may, due to rounding errors, not always add up to the number shown as Total

[b] The WHO Wellbeing Index was not administered to students in the lowest age group (11-year-olds) in Denmark

[c] The SWEMWBS scale was administered to all age groups in Sweden, and to the two oldest age groups in Denmark, Finland, and Norway. The SWEMWBS scale was not used in HBSC Iceland 2022

Table 11 Indicators of mental wellbeing: Intercorrelations between sumscores*. Weighted data and adjustment for cluster effects

	Life satisfaction (Cantril's ladder)	WHO Wellbeing Index	Mental wellbeing (SWEMWBS)	HBSC Health Complaints
Life satisfaction (Cantril's ladder)	1.00			
WHO Wellbeing Index	.61	1.00		
Mental wellbeing (SWEMWBS)	.59	.67	1.00	
HBSC Health Complaints	-.55	-.57	-.55	1.00

* All correlations are significant $p < .001$
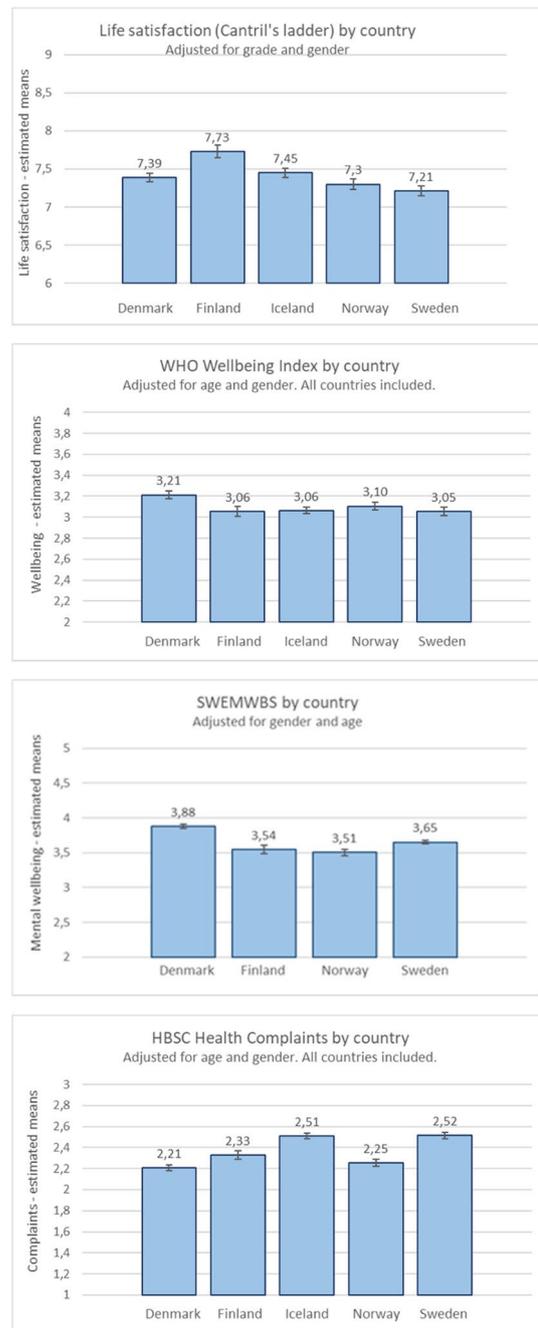
Fig. 5 Indicators of mental wellbeing by country, adjusted for gender and age. Weighted data and adjustments for cluster effects. No data available on the WHO Wellbeing Index for 11-year-olds in Denmark, Finland, and Norway. The SWEMWBS scale was not administered in the data collection in Iceland

Country	Age	Scale:			
		Cantril's ladder	WHO Wellbeing Index	SWEMWBS	HBSC Health Complaints Scale
Denmark	11	🟩	🟧	🟧	🟩
	13	🟩	🟩	🟩	🟩
	15	🟩	🟩	🟩	🟩
Finland	11	🟩	🟩	🟧	🟩
	13	🟩	🟩	🟩	🟩
	15	🟩	🟩	🟩	🟩
Iceland	11	🟩	🟩	🟧	🟩
	13	🟩	🟩	🟧	🟩
	15	🟩	🟩	🟧	🟩
Norway	11	🟩	🟩	🟩	🟩
	13	🟩	🟩	🟩	🟩
	15	🟩	🟩	🟩	🟩
Sweden	11	🟩	🟩	🟩	🟩
	13	🟩	🟩	🟩	🟩
	15	🟩	🟩	🟩	🟩

Fig. 6 Mental wellbeing scales included (green) and not included (orange) in data collections by country and age group

Country	Age	Covariate:		
		Family well off	Self-esteem	High stress/ Low stress
Denmark	11	🟧	🟩	🟧
	13	🟧	🟩	🟩
	15	🟧	🟩	🟩
Finland	11	🟩	🟩	🟧
	13	🟩	🟩	🟩
	15	🟩	🟩	🟩
Iceland	11	🟩	🟧	🟧
	13	🟩	🟧	🟩
	15	🟩	🟧	🟩
Norway	11	🟩	🟧	🟩
	13	🟩	🟧	🟩
	15	🟩	🟩	🟩
Sweden	11	🟩	🟧	🟩
	13	🟩	🟩	🟩
	15	🟩	🟩	🟩

Fig. 7 Covariate scales included (green) and not included (orange) in data collections by country and age group. Only covariates with incomplete coverage across subgroups defined by age and country are included

Abbreviations

ARPB	Average Relative Parameter Bias
CFI	Comparative Fit Index
ECV	Explained Common Variance
E.S.	Effect size – differences between means divided by standard deviation
FD	Factor Determinacy
H	Correlation between a factor and an optimally weighted composite
HBSC	Health Behaviour in School-aged Children (A WHO collaborative study)
PUC	Percent Uncontaminated Correlations
RMSEA	Root Mean Square Error of Approximation
SWEMWBS	The Short Warwick-Edinburgh Mental Wellbeing Scale
TLI	Tucker–Lewis Index
WLSMV	Weighted Least Squares Mean and Variance-adjusted (Estimator)

Authors' contributions

LEAa carried out the statistical analyses and drafted the first version as well as all consecutive revisions of the manuscript. ORS provided advice on statistical analyses. All co-authors contributed to the development of research questions and provided thorough feedback on all aspects and parts of this publication. MTD, ASF, NL, OS, and CE are members of the National HBSC teams in their respective countries. All the authors critically reviewed and approved the manuscript.

Data availability

The data file used for this study cannot be deposited online. The HBSC Data Management Centre organizes and distributes international HBSC data. The data file (HBSC 2022) is open access, and readers can apply for access and permission to analyze data. Application form and general information are available at www.hbsc.org. Data analysis syntax is available upon reasonable request to lead author.

Declarations

Ethics approval and consent to participate

The study adhered to the Guidelines of the Declaration of Helsinki. In Denmark, according to the Danish Scientific Ethical Committees Act, no ethics approval are needed for population-based questionnaire surveys. Approval was received from the school principal, the school board representing the parents, and the board representing the schoolchildren in every participating school. The participants were informed orally and in writing about the purpose and confidentiality of the study and that participation was voluntary. The parents also received written information and an electronic link to a short information video as well as an electronic link by which they could reject their child's participation in the study. In Finland, the University of Jyväskylä Ethics Committee reviewed the study and approved ethical clearance. The committee considered that the study was not a medical research project as defined in the Act on Medical Research and therefore did not require ethical review. Approval was collected from participating municipalities before approaching the schools and principals. At school-level, principals permitted school classes to participate in the study. In Iceland, the study received approval from the National Data Protection Agency and the University of Iceland Ethics Committee (Ethics approval number S7522). Additionally, permissions were obtained from educational authorities in each municipality as well as from the principals of all participating schools. In Norway, the Privacy Ombudsman at the University of Bergen verified that the study conformed to privacy and confidentiality standards, and the Regional Committee for Medical Research Ethics of South East Norway provided ethical approval (Ethics approval number 346737) including permission of applying informed (passive) consent. In Sweden, the study was reviewed by the Regional Ethical Review Board in Stockholm and classified as exempt from human subjects research (2023–05117-01). Participants were provided with both oral and written information regarding the confidentiality of their responses, ensuring that participation remained confidential and voluntary. All countries employed a passive consent approach.

Competing interests

The authors declare no competing interests.

Author details

[1]Department of Health Promotion, Norwegian Institute of Public Health, Zander Kaaes Gt. 7, 5015 Bergen, Norway. [2]Centre for Evaluation of Public Health Measures, Norwegian Institute of Public Health, PO Box 4404, Nydalen, N-0403 Oslo, Norway. [3]Department of Teacher Education, NLA University College, PO Box 74 Sandviken, 5812 Bergen, Norway. [4]Child and Adolescent Health, National Institute of Public Health, University of Southern Denmark, Studiestræde 6, 1455 Copenhagen, Denmark. [5]Department of Health and Caring Science, Western Norway University of Applied Science, Inndalsveien 28, 5063 Bergen, Norway. [6]Department of Education, University of Turku, 20014 Turun Yliopisto, Finland. [7]Faculty of Sport and Health Sciences, University of Jyväskylä, 40014 Jyväskylän Yliopisto, Finland. [8]Department of Health Promotion and Development, University of Bergen, Årstadveien 17, 5009 Bergen, Norway. [9]School of Psychology, University of New England, Armidale, NSW 2351, Australia. [10]Department of Learning, Informatics, Management, and Ethics, Karolinska Institute, 171 77 Stockholm, Sweden.

References

1. *Constitution of the World Health Organization*. 2024. Available from: https://apps.who.int/gb/bd/pdf/bd47/en/constitution-en.pdf. Cited 2024 10. December 2024
2. Mental health. 2024. Available from: https://www.who.int/health-topics/mental-health#tab=tab_1. Cited 2024 10. December 2024.
3. Pelters P. Right by your side? – the relational scope of health and wellbeing as congruence, complement and coincidence. Int J Qual Stud Health Well Being. 2021;16(1):1927482.
4. Ross DA, et al. Adolescent Well-Being: A Definition and Conceptual Framework. J Adolesc Health. 2020;67(4):472–6.
5. Iasiello, M., J. van Agteren, and E. Muir-Cochrane, Mental Health and/or Mental Illness: A Scoping Review of the Evidence and Implications of the Dual-Continua Model of Mental Health. Evidence Base, 2020;10(1).
6. Westerhof GJ, Keyes CL. Mental Illness and Mental Health: The Two Continua Model Across the Lifespan. J Adult Dev. 2010;17(2):110–9.
7. Kraiss JT, Kohlhoff M, ten Klooster PM. Disentangling between- and within-person associations of psychological distress and mental wellbeing: An experience sampling study examining the dual continua model of mental health among university students. Curr Psychol. 2023;42(20):16789–800.
8. Franken K, et al. Validation of the Mental Health Continuum-Short Form and the dual continua model of well-being and psychopathology in an adult mental health setting. J Clin Psychol. 2018;74(12):2187–202.
9. Keyes CL. The mental health continuum: from languishing to flourishing in life. J Health Soc Behav. 2002;43(2):207–22.
10. Gallagher MW, Lopez SJ, Preacher KJ. The hierarchical structure of well-being. J Pers. 2009;77(4):1025–50.
11. WHO. *Promoting adolescent wellbeing*. 2024. Available from: https://www.who.int/activities/promoting-adolescent-well-being. Cited 2024 27. September. A
12. Lucas RE, Diener E, Suh E. Discriminant validity of well-being measures. J Pers Soc Psychol. 1996;71(3):616–28.

13. Fiske DW. Convergent-discriminant validation in measurements and research strategies, in Forms of validity in research: New directions for methdology in social and behavioral science., D. Brinberg and L.H. Kidder, Editors. 1982, Jossey-Bass: San Francisco. p. 77–92.

14. Goodman FR, et al. Measuring well-being: A comparison of subjective well-being and PERMA. J Posit Psychol. 2018;13(4):321–32.

15. Longo Y, et al. Support for a general factor of well-being. Personality Individ Differ. 2016;100:68–72.

16. Disabato D, et al. Different Types of Well-Being? A Cross-Cultural Examination of Hedonic and Eudaimonic Well-Being. Psychol Assess. 2016;28(5):471–82.

17. Aarø LE, et al. Nordic adolescents responding to demanding survey scales in boring contexts: Examining straightlining. J Adolesc. 2022;94(6):829–43.

18. Campbell D, Fiske D. Convergent and Discriminant Validation By the Multitrait-Multimethod Matrix. Psychol Bull. 1959;56:81–105.

19. Rodriguez A, Reise SP, Haviland MG. Applying Bifactor Statistical Indices in the Evaluation of Psychological Measures. J Pers Assess. 2016;98(3):223–37.

20. Cantril H, The pattern of human concerns. New Brunswick. New Jersey: Rutgers University Press; 1966.

21. Wellbeing measures in primary health care: The Depcare project. Report on a WHO meeting, Stockholm, Sweden, 12–13 February 1998. 1998, World Health Organization, Regional Office for Europe: Copenhagen.

22. Stewart-Brown S, et al. Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): a Rasch analysis using data from the Scottish Health Education Population Survey. Health Qual Life Outcomes. 2009;7(1):15.

23. Haugland S, et al. Subjective health complaints in adolescence. A cross-national comparison of prevalence and dimensionality. Eur J Public Health. 2001;11(1):4–10.

24. Rimpelä M, Rimpelä A, Paronen O. Koetut oireet 12–18 -vuotiailla suomalaisilla. (Perceived health symptoms among 12–18 year-old Finns.). Sosiaalilääketieteellinen Aikakauslehti. 1982;4:219–233.

25. Corell M, et al. Socioeconomic inequalities in adolescent mental health in the Nordic countries in the 2000s - A study using cross-sectional data from the Health Behaviour in School-aged Children study. Archives of Public Health. 2024;82(1):20.

26. Lovis-Schmidt A, et al. Physical health complaints in adolescents: Findings from the 2018 Brandenburg HBSC study. European Journal of Health Psychology. 2022;29(3):121–33.

27. Lyyra N, Välimaa R, Tynjälä J. Loneliness and subjective health complaints among school-aged children. Scand J Public Health. 2018;46(20_suppl):87–93.

28. Låftman SB et al., Psychosocial School Conditions and Mental Wellbeing Among Mid-adolescents: Findings From the 2017/18 Swedish HBSC Study. Int J Public Health. 2023;67:Article 1605167.

29. Hagquist C, et al. Differential Item Functioning in Trend Analyses of Adolescent Mental Health – Illustrative Examples Using HBSC-Data from Finland. Child Indic Res. 2017;10(3):673–91.

30. Hagquist C, et al. Cross-country comparisons of trends in adolescent psychosomatic symptoms – a Rasch analysis of HBSC data from four Nordic countries. Health Qual Life Outcomes. 2019;17(1):27.

31. Heinz A, et al. Item response theory and differential test functioning analysis of the HBSC-Symptom-Checklist across 46 countries. BMC Med Res Methodol. 2022;22(1):253.

32. Hanzlová R, Lynn P. Item response theory-based psychometric analysis of the Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS) among adolescents in the UK. Health Qual Life Outcomes. 2023;21(1):108.

33. Inchley J, et al, Findings from the HBSC 2018 survey in Scotland. HBSC National report. . 2020, Social and Public Health Sciences Unit, University of Glasgow.: Glasgow.

34. WHO-EURO, Wellbeing measures in primary health care: The Depcare Project. Report of a WHO meeting. World Health Organization. Copenhagen: Regional Office for Europe; 1998.

35. Chan L, et al. Validation of the World Health Organization Well-Being Index (WHO-5) among medical educators in Hong Kong: a confirmatory factor analysis. Med Educ Online. 2022;27(1):2044635.

36. Faruk MO, et al. Validation of the Bangla WHO-5 Well-being Index. Global Mental Health. 2021;8: e26.

37. de Wit M, et al. Validation of the WHO-5 Well-Being Index in Adolescents With Type 1 Diabetes. Diabetes Care. 2007;30(8):2003–6.

38. Pattnaik P. Validation of the World Health Organisation 5-Item Well- Being Index (WHO-5) among the Adult Population Living in a Chronically Arsenic Affected Area of Rural West Bengal in India. Indian Journal of Public Health Research & Development. 2020;11(3):726–31.

39. Vaingankar JA, et al. Psychometric properties of the short Warwick Edinburgh mental well-being scale (SWEMWBS) in service users with schizophrenia, depression and anxiety spectrum disorders. Health Qual Life Outcomes. 2017;15(1):153.

40. Bartram DJ, Sinclair JM, Baldwin DS. Further validation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in the UK veterinary profession: Rasch analysis. Qual Life Res. 2013;22(2):379–91.

41. Ng SS, et al. Translation and validation of the Chinese version of the short Warwick-Edinburgh Mental Well-being Scale for patients with mental illness in Hong Kong. East Asian Arch Psychiatry. 2014;24(1):3–9.

42. Haver A, et al. Measuring mental well-being: A validation of the Short Warwick-Edinburgh Mental Well-Being Scale in Norwegian and Swedish. Scandinavian Journal of Public Health. 2015;43(7):721–7.

43. Fung SF. Psychometric evaluation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) with Chinese University Students. Health Qual Life Outcomes. 2019;17(1):46.

44. Smith ORF, et al. Measuring mental well-being in Norway: validation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS). BMC Psychiatry. 2017;17(1):182.

45. Ringdal R, et al. Validation of two versions of the Warwick-Edinburgh Mental Well-Being Scale among Norwegian adolescents. Scandinavian Journal of Public Health. 2017;46:140349481773539.

46. Black L, et al. Mental Health and Well-being Measures for Mean Comparison and Screening in Adolescents: An Assessment of Unidimensionality and Sex and Age Measurement Invariance. Assessment. 2024;31(2):219–36.

47. Shannon S, et al. Testing the factor structure of the Warwick-Edinburgh Mental Well-Being Scale in adolescents: A bi-factor modelling methodology. Psychiatry Res. 2020;293: 113393.

48. Sarasjärvi KK, et al. Exploring the structure and psychometric properties of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) in a representative adult population sample. Psychiatry Res. 2023;328: 115465.

49. Lang G, Bachinger A. Validation of the German Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) in a community-based sample of adults in Austria: a bi-factor modelling approach. J Public Health. 2017;25:135–46.

50. Hetland J, Torsheim T, Aarø LE. Subjective health complaints in adolescence: dimensional structure and variation across gender and age. Scandinavian Journal of Public Health. 2002;30(3):223–30.

51. Dey M, Jorm AF, Mackinnon AJ. Cross-sectional time trends in psychological and somatic health complaints among adolescents: a structural equation modelling analysis of 'Health Behaviour in School-aged Children' data from Switzerland. Soc Psychiatry Psychiatr Epidemiol. 2015;50(8):1189–98.

52. Ravens-Sieberer U, et al. An international scoring system for self-reported health complaints in adolescents. Eur J Pub Health. 2008;18(3):294–9.

53. Catunda C, Heinz A, Willems H. Subjective health complaints in adolescence – Validity of the HBSC symptom checklist in Luxembourg., in 32nd Annual Conference of the European Health Psychology Society. 2018: Galway, Ireland (poster).

54. Inchley, J., et al., *Health Behaviour in School-aged Children (HBSC) Study Protocol: Background, Methodology, mandatory questions and optional packages for the 2021/22 survey.* 2021, MRC/CSO Social and Public Health Sciences Unit, The University of Glasgow.: Glasgow.

55. Eriksson C. et al., Building knowledge of adolescent mental health in the Nordic countries. Nordisk välfärdsforskning | Nordic Welfare Research. 2019;4(2):43–53.

56. Cohen J, Statistical power analysis for the behavioural sciences (Second edition). Hillsdale. New Jersey: Lawrence Erlbaum; 1988.

57. Xu L. Scale development and factor analysis., in Scholarly Publishing and Research Methods Across Disciplines. 2019;159–183. IGI Global.

58. Levin KA, Currie C. Reliability and Validity of an Adapted Version of the Cantril Ladder for Use with Adolescent Samples. Soc Indic Res. 2014;119(2):1047–63.

59. Brown TA. Confirmatory factor analysis for applied research (2nd edition). 2nd ed. New York: The Guilford Press; 2015.

60. Hu L-T, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equ Model. 1999;6(1):1–55.
61. Bonifay W, et al. When Are Multidimensional Data Unidimensional Enough for Structural Equation Modeling? An Evaluation of the DETECT Multidimensionality Index. Struct Equ Modeling. 2015;22:1–13.
62. Dunn T, Baguley T, Brunsden V. From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. Brit J Psychol. 2013;105:399–412.
63. Reise SP, Bonifay WE, Haviland MG. Scoring and modeling psychological measures in the presence of multidimensionality. J Pers Assess. 2013;95(2):129–40.
64. Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: exploring the extent to which multidimensional data yield univocal scale scores. J Pers Assess. 2010;92(6):544–59.
65. Hancock GR, Mueller RO. *Rethinking construct reliability within latent variable systems* in *Structural equation modeling: Present and future - A festschrift in honor of Karl Jöreskog.*, Cudeck R, du Toit S, and Sørbom D, Editors. 2001, Scientific software international: Lincoldwood, Illinois. p. 195–216.
66. Gorsuch RL, Factor analysis. 2, editors. Hillsdale. New Jersey: Lawrence Erlbaum; 1983.
67. Rodriguez A, Reise SP, Haviland MG. Evaluating bifactor models: Calculating and interpreting statistical indices. Psychol Methods. 2016;21(2):137–50.
68. Dueber DM. *Bifactor Indices Calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA-models.* 2017; Available from: http://sites.education.uky.edu/apslab/resources/.
69. Stucky BD, Edelen MO. *Using hierarchical IRT models to create unidimensional measures from multidimensional data.*, in *Handbook of item response theory modeling: Applications to typical performance assessment.*, S.P. Reise and D.A. Revicki, Editors. 2014, Routledge. p. 183–206.
70. Reise SP, et al. Multidimensionality and Structural Coefficient Bias in Structural Equation Modeling: A Bifactor Perspective. Educ Psychol Measur. 2012;73(1):5–26.
71. Markon KE. *Bifactor and Hierarchical Models: Specification, Inference, and Interpretation.* Annual Review of Clinical Psychology, 2019. 15(Volume 15, 2019): p. 51–69.
72. DeMars CE. A Tutorial on Interpreting Bifactor Model Scores. Int J Test. 2013;13(4):354–78.
73. Muthén B, Kaplan D, Hollis M. On structural equation modeling with data that are not missing completely at random. Psychometrika. 1987;52(3):431–62.
74. Embretson SE, Reise SP, Item response theory for psychologists. Mahwah. New Jersey: Lawrence Erlbaum; 2000.
75. Böhnke JR, Croudace TJ. Calibrating well-being, quality of life and common mental disorder items: psychometric epidemiology in public mental health research. Br J Psychiatry. 2016;209(2):162–8.
76. Krieger T, et al. Measuring depression with a well-being index: further evidence for the validity of the WHO Well-Being Index (WHO-5) as a measure of the severity of depression. J Affect Disord. 2014;156:240–4.
77. Tourangeau R, Couper M, Conrad F. Spacing, Position, and Order: Interpretive Heuristics for Visual Features of Survey Questions. Public Opinion Quarterly - PUBLIC OPIN QUART. 2004;68:368–93.
78. Silber H, Roßmann J, Gummer T. When near means related: evidence from three web survey experiments on inter-item correlations in grid questions. Int J Soc Res Methodol. 2018;21(3):275–88.

Publisher's Note