Development of the Gothenburg Well-Being Scale in Late Adolescence: The Grow Up 1990 Gothenburg Study

Sarah Hitz¹ · Ebba Brann² · Kerstin Albertsson-Wikland³ · Zita Schillmöller¹ · John Eric Chaplin⁴

Published online: 26 February 2019

© The Author(s) 2019

Abstract

The objective was to update the factor structure of the Gothenburg Well-Being scale (GWB) for use in older adolescents. A cross-sectional population of final-year high school students (17–20 years; $N=5395$) completed three questionnaires with classroom administration as part of a larger investigation into health and lifestyle. Students completed the GWB, which is composed of 43 polar-opposite adjective VAS scales, the RS-11 resilience questionnaire and an 8-item Body Image scale. Model fit to the previously data-derived six-factor model (GWB Child) was tested. The data were then randomly divided into training and test datasets and a theoretically hypothesised model was tested and revised. Model 1 did not have adequate model fit. The theoretically hypothesised Model 2 had better fit; however, a modified model (Model 3 GWB Adolescent), using nine questions loading on factors of Mood, Stress Balance and Activation with one general factor of Well-being, was found to meet all model fit criteria (GFI .978; TLI .970, CFI .980; RMSEA .059). Measurement invariance was attested across datasets and gender. Internal reliability was satisfactory (Cronbach's alpha 0.59). Convergent validity was demonstrated by correlation with resilience ($r=.42$) and body image ($r=.35$). Discriminant known-groups analysis gave results in the predicted direction for gender. The findings contribute to the validity of the GWB Adolescent, which explores relevant elements of well-being in late adolescence and can be used for group comparisons. Further testing is required to identify relationships with independent aspects of late adolescent life and clinical variables.

Keywords Adolescence · Well-being · Factor analysis · Resilience · Body image · Short stature

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s41543-019-00016-x>) contains supplementary material, which is available to authorized users.

✉ John Eric Chaplin
john.chaplin@gu.se

Extended author information available on the last page of the article

1 Introduction

The measurement of subjective well-being (SWB) has become an area of growing prominence for academics, healthcare professionals and policymakers (Krueger and Stone 2014), who agree that it is vital to be able to monitor well-being adequately in order to develop health policies and strategies (WHO 2011). Two general approaches to understanding SWB have been proposed (Deci and Ryan 2008). The first approach, based on hedonic theories, has generated possibly the most studied models of well-being (Gallagher et al. 2009). Diener, who has published extensively on SWB (Diener 1984; Diener 2009), defines it as the experience of high levels of pleasant emotions and moods, low levels of negative emotions and moods and high life satisfaction. The alternative approach is that of the eudaimonic tradition (Waterman 1993); here, SWB is considered to be the outcome of positive goal pursuits (Ryan et al. 2008). Ryff (1989) presented a model of eudaimonic well-being built on the assumption that individuals strive to function fully and realise their unique talents. Recent research has also focused on the conceptual structure of 'being, belonging and becoming' (Chaplin et al. 2017), which can be seen as related to the assessment of positive affect, life satisfaction (hedonic) and personal growth (eudaimonic).

SWB is characterised as a multi-domain concept (Dodge et al. 2012). A recent review of adult measures by Linton et al. (2016) found 99 instruments which included 196 different dimensions of well-being. These were clustered around six key themes: psychological well-being, social well-being, physical well-being, spiritual well-being, activities and functioning, and personal circumstances, and these themes were measured with between 1 and 317 items (Linton et al. 2016). From an examination of the dimensions, it was found that they included both strength-based issues, such as joy and social acceptance, and deficit-based issues, such as anxiety and pain. This observation is also reported in reviews of paediatric SWB (Pollard and Lee 2003). Strength-based or deficit-based approaches can be used as a method to conceptualise well-being measurement. Despite extensive literature and the development of a large number of measures (Tsang et al. 2012; Amerijckx and Humblet 2014) there is little consensus on the theoretical grounding and dimensions of well-being (Ryff 1989). This has led to SWB often being considered synonymous with the concept of health-related quality-of-life (HrQoL) (Taillefer et al. 2003). SWB can however, be characterised as a strength-based concept (Tsang et al. 2012) with the purpose of quantifying continued growth and development, and HrQoL as a deficit-based concept (Schipper et al. 1996) with the purpose of quantifying the degree to which a medical condition or its treatment impacts the individual's life (Guyatt et al. 1993).

It is useful to think of SWB as a strength-based concept within health research, as this refers back to the measurement of health as originally conceived by the WHO (1948), which requires not just the absence of clinically defined symptoms but also the positive aspects of good health. Recent patient-centred initiatives (Epstein 1999) and a shift towards a focus on health promotion (Coverdale and Long 2015) have also encouraged the documentation of an individual's strengths and resources as measures of well-being.

Another clear trend in well-being research has been a focus on the relationship of physical activity to well-being especially in young people (Calfas and Taylor 1994; Nieman 2002; Kanning and Schlicht 2010). Schools-based research has illustrated a strong association between well-being, physical activity, cognition and day-to-day mood (Huppert 2009), although this association is not always found (Konijnenberg and Fredriksen 2018). A

combination of these approaches suggests a model of well-being that may be more appropriate for a youth population, bringing together emotional reactions, sense of achievement and physical health or activity.

The methods used to measure SWB have also varied over time. Linton and colleagues (Linton et al. 2016) found that the majority of instruments use the Likert-type scale. Visual analogue scales (VAS) tend to be used for global well-being measurement, as in the EQ5D and SF-36, but there is wider scope for their application (Brazier et al. 1999). VAS scales have the advantage of being able to measure a continuum across the full range of the concept from negative to positive and therefore have important advantages for the measurement of well-being (Parkin and Devlin 2006).

In this research, we turn our attention to the measurement of child well-being. Age-sensitive models of well-being are important because the relevant dimensions change with age, cognitive development and biological maturity. A recent systematic review of well-being instruments for adolescents (Rose et al. 2017) has identified eleven instruments that meet the criteria of following a strength approach and covering both affective and functional aspects, although it is noted that measurement instruments for children tend to be more subjective than objective measures (Pollard and Lee 2003). Tsang and colleagues (Tsang et al. 2012) focused on SWB in children and described 17 measures categorised into indicators based on deficits or strengths, or both. Deficit-oriented indicators included poor emotional awareness, negative affect, inhibition and social anxiety, whereas strength-oriented indicators were fewer and included positive affect, emotional regulation, social and family involvement and personal adaptation in terms of resilience or coping.

Clinical-based outcome measures are usually deficit-oriented indicators related to symptoms, for example social exclusion, depression and physical limitations. Although the documentation of deficits is essential for eligibility requirements of clinical services, it can be questioned whether this is the best way to evaluate well-being when compared to the general population. Reports indicate that there is a reduced quality of life associated with medical conditions (Olsson et al. 2013); however, the use of deficit-oriented indicators and a focus on a clinical population may lead to the over-estimation of psychological or social problems in certain populations due to a focusing effect (Brazier and Tsuchiya 2010). One good example of this is in relation to the psychosocial effects of short stature (Erling 2004; Chaplin et al. 2012), where deficit-based measures designed for clinical populations have been claimed to exaggerate problems when applied to the general population (Sandberg et al. 1994). Well-being measurement as an outcome of treatment can therefore contribute to the ongoing debate in relation to the psychosocial outcome of growth hormone treatment for short stature (Bullinger et al. 2009; Chaplin et al. 2011; Chaplin et al. 2012; Quitmann et al. 2016; Sandberg and Voss 2002; Voss 2001; Voss and Sandberg 2004). In order to explore this issue and compare clinical and population samples, an instrument with questions and indicators that can identify positive strengths and abilities in both populations is needed.

To meet this need, the Gothenburg Well-Being scale for children (GWB Child) (Erling et al. 2002) was developed. The scale was designed as a multi-dimensional questionnaire to measure well-being with reference to a general population (Wiklund et al. 1994). The questionnaire uses a series of VAS scales in order to measure the continuum between negative and positive aspects of well-being with the aim that the instrument is applicable to both clinical and non-clinical populations. It has been used on an annual basis to follow up the progress of children with short stature in clinical trials of growth hormone treatment and in a number of research studies (Erling et al. 1994; Wirén et al. 1998; Wirén et al. 2001; Erling et al. 2002; Erling 2004; Chaplin et al. 2011).

GWB Child has shown to have good internal consistency ($ICC > 0.8$) and has been used in both child and young adolescent populations (Erling et al. 1994), having been validated previously on children reaching puberty aged 9–13 years. However, knowledge of the psychometric properties of the instrument in an older adolescent population (up to 20 years) is required for clinical follow-up. This study aimed to test the factorial structure of GWB Child (six factors) in an older adolescent population. The questionnaire was therefore included in a study of students in the final year of high school (age 17–20 years). This also provided the opportunity to examine measurement invariance across gender in a large population. In healthcare research, a common approach to well-being measurement has been to develop a new instrument for each population and context; however, this takes time and resources and potentially makes it difficult to compare results across conditions and populations; furthermore, it may also waste useful historical data. A better approach may be to adapt, update or incorporate the existing models that have already been developed for use in similar contexts or populations.

1.1 Hypotheses

Our first hypothesis (H1) was that the factor structure developed for GWB Child (Wiklund et al. 1994) would also show good model fit in the data from the adolescent population.

Our second hypothesis (H2) was that an improved item pool which included items previously excluded due to poor fit in the younger age group can be found to have better model fit. The theoretical model incorporates emotional reactions, sense of achievement and physical activity. The robustness of the model fit was tested in a second population sample and assessed for measurement invariance across datasets and gender.

Our third hypothesis (H3) was that convergent and discriminant validity can be demonstrated in the final model. Convergent validity via a positive correlation between well-being and resilience (Stewart-Knox et al. 2012; Gomez et al. 2013) and between well-being and body image as an important aspect of adolescent life (Frisén 2007; Mond et al. 2010; Delfabbro et al. 2011; Erling 1999; Frisén 2007). Although these concepts are related to well-being, they are not equivalent (Avalos et al. 2005; Holmqvist and Frisen 2012); therefore, a moderate correlation can be predicted. The well-being scores are expected to differ between boys and girls; therefore, the scale should discriminate across gender, with boys having significantly higher well-being scores (Moksnes and Espnes 2013).

2 Methods

2.1 Ethics

Ethical approval was obtained from the Regional Ethics Committee (approval no. 444–08). Informed consent was obtained from all participants prior to investigation which included a comprehensive questionnaire and anthropometric measurements.

2.2 Participants and Test Procedure

The study was embedded within a larger study of linear growth, weight status, health and related behaviours in a population of 18-year-olds who had been followed from birth (Sjöberg et al. 2012). This study gave the opportunity for the validity to be assessed but, due to the

number of procedures and questionnaires included in the main study, there was no opportunity for the inclusion of additional well-being measures which might have been used to further establish the convergent validity. Forty high schools out of a total of 47 in Gothenburg and surrounding areas participated. These ranged in size and contributed between 14 and 395 pupils to the survey. Participants completed a questionnaire as part of a larger survey examining growth, health, lifestyle and body image (Sjöberg et al. 2012). A study team visited the schools during normal lesson times to administer the questionnaire and take anthropometric measurements. Students sat together in their classroom to complete the questionnaire.

2.3 Instruments

Sociodemographic information was collected using a self-administered questionnaire which included age, gender and lifestyle questions.

2.3.1 The GWB Scale

The item pool generated for the GWB instrument was composed of a series of 43 polar-opposite adjective pairs (e.g. sad–happy or tense–relaxed) aimed at describing behavioural attributes, feelings and important concepts of well-being identified through discussions with children and teachers (Wiklund et al. 1994). The GWB Child six-factor model was developed in three gender-balanced populations of 9-year-olds ($N = 117$), 11-year-olds ($N = 116$) and 13-year-olds ($N = 109$). Thirty-four of the 43 questions were retained from the exploratory factor analysis (EFA) (Wiklund et al. 1994). Convergent validity of the GWB Child was assessed by correlation with the ‘I Think I Am’ self-perception scale (Ouvinen-Birgerstam 1985; $r = .64$; $p < .001$). The internal consistency of this scale was shown to be excellent (Cronbach’s $\alpha = .88$). The factor structure was identified by the test creators (Erling et al. 1994) using an EFA alpha factoring and varimax rotation. Thirty-four items with a factor loading of ≥ 0.40 were combined to form six factors: Alertness (11 items, e.g. contented/not contented), Self-esteem (seven items, e.g. shy/brave), Mood (four items, e.g. happy/sad), Elation (five items, e.g. talkative/quiet), Stability (four items, e.g. relaxed/tense) and Vitality (three items, e.g. quick/slow). No other information concerning the factor analysis is provided in the published works.

2.3.2 Resilience Scale (RS-11)

Resilience was measured using the RS-11 scale (Schumacher et al. 2005; von Eisenhart Rothe et al. 2013; Kocalevent et al. 2015), which is based on an earlier 25-item scale (Wagnild and Young 1993; Wagnild and Collins 2009). Construct validity of the RS-11 was assessed by correlation with the general self-efficacy (GSE) scale (Schwarzer and Jerusalem 1995; $r = .70$; $p < .001$). The internal consistency of this scale was shown to be excellent, with Cronbach’s $\alpha = .91$ (Schumacher et al. 2005). The RS-11 uses an equally weighted summed index composed of all eleven questions. The resulting sum score was then transformed into a 0–100 scale, where higher scores represent greater resilience.

2.3.3 Body Image Scale

Body image was assessed via the ‘I Think I Am’ method (Ouvinen-Birgerstam 1985) which was developed in Sweden and used for studying various aspects of self-esteem. It had also

been used for validation in the original GWB Child (Wiklund et al. 1994). Eight statements relating to body image were used. These statements constitute an internally consistent scale (Cronbach's $\alpha = .83$) and summarise body image, body development and satisfaction with body image. For each statement, four response alternatives were available ('true', 'partly true', 'partly not true' and 'not true'). A summary variable was constructed, and the body image questions transformed to a score between 0 and 100, where 0 represents totally dissatisfied and 100 as the most positive perceived body image.

2.4 Data Analysis

Analysis was conducted using IBM SPSS Statistics for Windows, Version 21 (IBM Corp., Armonk, NY, USA). Descriptive statistics (means, standard deviations, medians, skewness and kurtosis) were used to compare groups. Item sensitivity was assessed through skewness and kurtosis for deviance from normal distribution (Kline 2011).

The population was divided into two equal-sized, randomly allocated sub-samples to be used as a training and a test dataset. The use of a second dataset allowed us to test whether the model would generalise to an independent population. Hypotheses H1 and H2 were tested using maximum-likelihood confirmatory factor analysis (CFA) using IBM SPSS AMOS 25. Model fitness was assessed using Goodness of Fit Index (GFI), Adjusted Goodness of Fit Index (AGFI), Normed Fit Index (NFI), Tucker-Lewis index (TLI) Comparative Fit Index (CFI), root mean square error of approximation (RMSEA) and standardised root mean square residual (SRMR). The model was considered to have good fit if GFI, AGFI, NFI and TLI were higher than 0.9 and CFI was higher than 0.95 (Kline 2011), and if RMSEA and SRMR are lower than 0.05 (Hooper et al. 2008). χ^2 was reported, but because it is extremely sensitive to sample size and rejects reasonable models if the sample is large (Andrich 1988), it was not relied upon for model fit (Bergh 2015); therefore, TLI, CFI and RMSEA, which are less affected than other indices by sample size and model complexity (Bollen and Long 1993), were preferred. Reliability of the scale was assessed using Cronbach's alpha.

Hypothesis 3 (H3) was that the model could differentiate between genders and demonstrate a relationship to the associated concepts of resilience and body image.

3 Results

A total of 5687 students completed the GWB scale (49 items); of these, 5% (291 cases) were excluded listwise: four had clearly spoiled responses to the VAS (reporting 0 for all questions following the first page); 77 had completely missing Body Image, Resilience and demographic data; 66 were missing all Body Image and Resilience items; 144 had greater than 40% item non-response, amounting to more than three pages of the questionnaire. Although this missing data could be considered to be missing at random, the large amount of missing data led us to conclude that these students had not fully participated in the survey, making the remaining responses untrustworthy. A binary logistic regression was used to assess whether it was possible to predict this group by demographic variables (age, gender, BMI or school). This revealed that both gender and school were predictors ($\chi^2 = 29.5$; $df = 4$; $p < .001$). Gender was another variable that had not always been completed by the students in this missing-items group, but of the 163 cases where gender was indicated in the sample with missing data, it was found that boys were twice as likely as girls to return questionnaires with missing data

($N=108/55$). Of the 40 schools surveyed, 33 had questionnaires with missing data, with a mean rate of 3%. The two schools with the highest rate of missing data (9% of cases) were examined but there was no obvious reason for these two medium-size city schools should be treated differently from the others. Therefore, listwise deletion of these 291 cases does not appear to introduce a bias nor lower the power to detect effects.

Of the remaining 5395 cases, 13% (685 cases) completed the three questionnaires with missing data, resulting in 885 missing data points (0.2%). Values were imputed using both multiple imputation and linear imputation in SPSS following reversal of negative items. The difference between these two techniques for the VAS was on average 0.0001 for the mean of the scales. The study population therefore comprised 5395 students: 2709 girls (50%) and 2686 boys (50%), mean age 19 (SD = 0.45; age range 17–20 years), see Table 1. Table 1 also describes the training ($N=2661$) and test datasets ($N=2734$). The datasets could not be predicted based on demographic variables of age, gender, BMI, or school ($\chi^2=3.88$; $df=4$; $p>.05$) and no significant differences between outcome variables was found (Table 3). This allowed the stability of the factor solution to be investigated in a second random subsample (test dataset). See Table 1 for a description of each of the datasets.

Table 2 presents the descriptive statistics of the results from the well-being (GWB), resilience (RS-11) and body image questionnaires. The total dataset was normally distributed with a mean skewness of -0.18 (SE = 0.033; range from -1.192 to 1.056) and mean kurtosis of -0.310 (SE = 0.067; range from -1.098 to 1.480). Normality was also maintained across gender, although differences between gender were found on all well-being dimensions and on Body Image. Boys had significantly higher well-being scores, although Stress Balance was better for girls. There was no gender difference in resilience (Table 3).

3.1 Confirmatory Factor Analysis

Table 4 presents the three models. The full list of factor loadings for all three models is presented in Appendix 1. Model 1, a data-derived, six-factor solution with 34 items (Gothenburg Well-Being scale for children, GWB Child), was found to have poor model fit in this population of older adolescents (GFI = .832, TLI = .770, CFI = .790, RMSEA = .073). Modification of the scale by co-variance and removing the more poorly performing items did not improve the model fit sufficiently without compromising the factor structure (data not reported here).

Therefore, Hypothesis H1 does not hold and the GWB Child model was rejected.

Model 2, a theoretically hypothesised, five-factor bifactor solution with 22 items (Gothenburg Well-Being scale for adolescents, GWB Adolescent), was found to fit the data fairly closely (GFI = .915, TLI = .888, CFI = .902, RMSEA = .066). Model 2 is shown in Fig. 1. This

Table 1 Sample characteristics of the adolescent population, divided by gender and datasets for training and testing

	Total $n=5395$ (girls 50%)	Girls $n=2709$	Boys $n=2686$	Training $n=2661$ (girls 50%)	Test $n=2734$ (girls 50%)
Mean age (SD)	18.63 (0.49)	18.67 (0.50)	18.58 (0.49)	18.62 (0.50)	18.63 (0.49)
Age range	17–20	17–20	17–20	17–20	17–20
BMI, kg/m^2 (SD)	22.39 (3.30)	22.02 (3.12)	22.76 (3.43)	22.37 (3.24)	22.42 (3.36)

Table 2 Descriptive statistics for outcome variables of well-being (GWBa), resilience (RS-11) and body image

	Total n = 5395 Mean (SD)	Skewness (Kurtosis)	Boys n = 2686 Mean (SD)	Skewness (Kurtosis)	Girls n = 2709 Mean (SD)	Skewness (Kurtosis)
Mood	68.53 (18.97)	-.78 (.25)	72.09 (17.72)	-.94 (.68)	64.99 (19.50)	-.63 (.01)
Stress	46.43 (23.84)	.10 (-.80)	40.16 (23.02)	.30 (-.69)	52.64 (23.00)	-.08 (-.71)
Balance						
Activation	58.51 (18.93)	-.29 (-.36)	62.46 (18.28)	-.42 (-.21)	54.60 (18.75)	-.18 (-.37)
GWBa	60.28 (11.44)	-.48 (.22)	61.79 (10.86)	-.63 (.69)	58.78 (11.81)	-.32 (-.03)
total						
Resilience	72.00 (13.78)	-.70 (.98)	71.87 (13.95)	-.72 (1.19)	72.12 (13.62)	-.69 (.74)
Body Image	28.41 (18.08)	.56 (.18)	23.94 (16.16)	.62 (.32)	32.84 (18.79)	.41 (.03)

GWBa - Gothenburg Well-Being scale for adolescents

model was modified by removing the more poorly performing items to produce the final model, Model 3 (GWB Adolescent), which was composed of three factors with one general factor.

The CFA results from the training dataset (A) and the test dataset (B) are shown in Table 4. In the training dataset (A), Model 3 was found to have achieved good fit (GFI = .978, TLI = .970, CFI = .980, RMSEA = .059). The graphical presentation of Model 3 CFA is shown in Fig. 2. Next, we tested for measurement fit of Model 3 in the test dataset (B) using multi-group CFA (Table 5). This also indicated a good fit (GFI = .985, TLI = .979, CFI = .986, RMSEA = .049) (Fig. 3). Measurement invariance was tested between groups (Table 6) and between genders (Figs. 4 and 5, Table 7), based on delta CFI (Cheung and Rensvold 2016). Taking the practical approach of examining the difference in CFI using the threshold of < .002 (Meade et al. 2008), it is possible to draw the conclusion that Model 3 is invariant across the two groups. It can also be concluded, based on the CFI difference, that the model is invariant across gender (Table 7).

An alternative approach provided by AMOS is to use a multigroup model (Table 6.2 and 6.3). This uses the traditional invariance-testing approach based on the χ^2 difference test. Using this method, the delta χ^2 value between the unconstrained and configural models was

Table 3 Gender differences across the outcome variables

	Training / Test			Boy / Girl		
	Mean difference	t-test (df = 5394)	p value	Mean difference	t-test (df = 5393)	p value
Mood	-0.21	-0.40	NS	7.10	13.99	<.001
Stress Balance	0.29	0.45	NS	-12.48	-19.92	<.001
Activation	0.35	0.67	NS	7.86	15.58	<.001
Total Well-being (GWBa)	0.09	0.28	NS	3.00	9.72	<.001
Resilience	-0.42	-1.11	NS	-0.26	-0.68	NS
Body Image	0.55	1.12	NS	-8.89	-18.63	<.001

GWBa - Gothenburg Well-Being scale for adolescents

Table 4 Goodness-of-fit statistics for original, hypothesized and, and final model across training/test dataset and gender

	Model 1 6 factors, 34 items N = 5395	Model 2 5 factors, 22 items N = 2661	Model 3 (GWBa) 3 factors, 1 general factor, 9 items			
			Training N = 2661 (A)	Test N = 2734 (B)	Male N = 2686 (A + B)	Female N = 2709 (A + B)
χ^2 (df)	7685.16 (512)	2580.93 (203)	245.95 (24)	182.14 (24)	208.90 (24)	183.43 (24)
χ^2 /df	15.01	12.71	10.25	7.59	8.71	7.64
GFI	.832	.915	.978	.985	.982	.984
AGFI	.805	.894	.959	.972	.966	.971
NFI	.778	.894	.978	.984	.982	.982
TLI	.770	.888	.970	.979	.976	.977
CFI	.790	.902	.980	.986	.984	.985
RMSEA	.073	.066	.059	.049	.054	.050
p-close	<.001	<.001	.013	.574	.180	.533
SRMR	.066	.054	.026	.021	.019	.026

Goodness-of-fit statistics for three models: Model 1, derived from the original exploratory factor analysis (1994); Model 2, hypothesized; Model 3 (GWBa), the final model divided by training/test dataset and by gender

GWBa - Gothenburg well-being scale for adolescents, *GFI* - Goodness of Fit Index, *AGFI* - Adjusted Goodness of Fit Index, *NFI* - Normed Fit Index, *TLI* - Tucker Lewis Index, *CFI* - Comparative Fit Index, *RMSEA* - Root Mean Square Error of Approximation, *p-close* - Probability of close fit, *SRMR* - Standardised Root Mean Square Residual

found to be 13.87 with 6 degrees of freedom. This χ^2 difference value is statistically significant at $p < .05$. Based on this result, it could be concluded that one or more of the factor loadings are not operating equivalently across the two groups and that Model 3 is completely non-equivalent across the training and test datasets. Although the results are presented here for comparison, it was concluded by the authors that, given that the datasets were randomly assigned and could not be predicted based on demographic variables (as described above), the results from the CFI analysis were the most trustworthy. This was also true for the gender analysis.

3.2 Reliability and Validity

Table 5 presents reliability and convergent validation results. Reliability of the total scale was indicated by a Cronbach's alpha of 0.59. Cronbach's alpha for the internal scales ranged from .70 to .89. The correlation between the four well-being factors shows a good correlation between Mood and QWB adolescent (GWBa) total and between Activation and GWBa total but a poor relationship between Stress Balance and GWBa total. Table 5 also shows the correlation between the well-being factors and the theoretically linked Resilience and Body Image variables. GWBa total explains 18% of the variance in Resilience ($r = .42$) and 12% of Body Image ($r = .35$). This is in line with the expectations of a modest agreement between the concepts that these measures represent. Discriminant validity scores for gender differed as predicted. Table 2 shows that boys recorded a total mean score on GWBa total of 61.79 (1 SD 0.86) as opposed to the lower GWBa total score for girls of 58.78 (11.81 SD). This was statistically significant ($p < .001$) in the predicted direction. Boys scored significantly higher than girls on the

Fig. 1 A theoretically hypothesised, five-factor model with 22 items (Model 2)

The loading for each item is shown above the arrow on the left side. The correlation coefficients between the first-order and second-order factor are shown beside the lines between the factors. Standardised estimates are shown. Chi-square test ($\chi^2/df = 12.71$), Goodness of Fit Index (GFI = .915), Adjusted Goodness of Fit Index (AGFI = .894), Normed Fit Index (NFI = .894), Tucker-Lewis Index (TLI = .888) Comparative Fit Index (CFI = .902), Root Mean Square Error of Approximation (RMSEA = .066) and Standardised Root Mean Square Residual (SRMR = .054)

Fig. 2 Model 3, training dataset ($N = 2661$)

GWBa = Gothenburg Well-Being scale for adolescents. The loading for each item is shown above the arrow on the left side. The correlation coefficients between the first-order and second-order factor are shown beside the lines between the factors. Standardised estimates are shown. Chi-square test ($\chi^2/df = 10.25$), Goodness of Fit Index (GFI = .978), Adjusted Goodness of Fit Index (AGFI = .959), Normed Fit Index (NFI = .978), Tucker-Lewis Index (TLI = .970) Comparative Fit Index (CFI = .980), Root Mean Square Error of Approximation (RMSEA = .059) and Standardised Root Mean Square Residual (SRMR = .026)

factors Mood, Activation and GWBa total but lower than girls on the factor Stress Balance.

Evidence of convergent and discriminant validity was found; therefore, the third hypothesis (H3) was accepted. Model 3 does have divergent and convergent validity.

Table 5 Cronbach's alpha coefficient for the GWBa scale and subscales and correlation with outcome variables

	Cronbach alpha	Correlations				Resilience	Body Image
		1	2	3	4		
1. Mood	.89	—	-.53	.60	.82	.39	.41
2. Stress Balance	.81		—	-.36	-.13	-.22	.31
3. Activation	.70			—	.82	.42	.34
4. GWBa total	.59	.82	-.13	.82	—	.42	.35

Item correlation based on the total population $N = 5395$. All correlations are significant at the 0.01 level (2-tailed), Pearson's coefficients (r) are shown for the total score of the GWBa, the three domains and outcome variables of Resilience and Body Image, GWBa - Gothenburg Well-Being scale for adolescents

Fig. 3 Model 3, test dataset (N = 2734)

GWBa = Gothenburg Well-Being scale for adolescents. The loading for each item is shown above the arrow on the left side. The correlation coefficients between the first-order and second-order factor are shown beside the lines between the factors. Standardised estimates are shown. Chi-square test ($\chi^2/df = 7.59$), Goodness of Fit Index (GFI = .985), Adjusted Goodness of Fit Index (AGFI = .972), Normed Fit Index (NFI = .984), Tucker-Lewis Index (TLI = .979) Comparative Fit Index (CFI = .986), Root Mean Square Error of Approximation (RMSEA = .049) and Standardised Root Mean Square Residual (SRMR = .021)

4 Discussion

The results suggest that the nine-item model (Model 3) had acceptable model fit to older adolescents, good validity and acceptable reliability and can therefore be used in the measurement of well-being in this age group. The new factor structure provides a better model fit to the data with fewer questions than the GWB Child scale. The GWB Adolescent scale is composed of nine VAS bipolar adjective pairs and has three factors and one general well-being factor. These factors cover emotional well-being, activation and stress balance. Measurement invariance was found, for Model 3, between the two datasets and between genders, despite significant differences between the gender scores on the well-being dimensions, suggesting that the model can be used for group comparisons and that the model equally applies across gender. The GWB Adolescent scale has some unique qualities: it is based on an approach grounded on the experience of a paediatric population and the results were analysed on the basis of a theoretically hypothesised factor-analytic model that has been validated in a school-based population.

The relationship between well-being and resilience (Sagone and de Caroli 2014) and between well-being and body image, especially in this age group (Delfabbro et al. 2011), is

Table 6 Model fit summary: measurement invariance across training and test datasets**Table 6.1. Testing measurement invariance across training and test groups using multi-group confirmatory factor analysis**

Training / Test dataset	df	χ^2 goodness of fit	P value	NFI	IFI	CFI	Δ CFI
Unconstrained	48	427.990	.000	.981	.983	.983	
Measurement weights	54	441.859	.000	.981	.983	.983	.001
Structural weights	56	457.110	.000	.980	.982	.982	.001
Structural covariances	57	457.111	.000	.980	.982	.982	.000
Structural residuals	60	458.116	.000	.980	.982	.982	.000
Measurement residuals	69	503.508	.000	.978	.981	.981	.001

Table 6.2. Model fit summary: testing measurement invariance across training and test groups, assuming that the unconstrained model is correct

Training / Test dataset	df	χ^2 goodness of fit	P value	Δ NFI	Δ IFI	RFI	TLI
Measurement weights	6	13.870	.031	.001	.001	-.002	-.002
Structural weights	8	29.120	.000	.001	.001	-.002	-.002
Structural covariances	9	29.121	.001	.001	.001	-.003	-.003
Structural residuals	12	30.126	.003	.001	.001	-.004	-.004
Measurement residuals	21	75.518	.000	.003	.003	-.005	-.005

Table 6.3. Model fit summary: testing measurement invariance across training and test groups, assuming that model measurement weights are correct

Training / Test dataset	df	χ^2 goodness of fit	P value	Δ NFI	Δ IFI	RFI	TLI
Structural weights	2	15.250	.000	.001	.001	.000	.000
Structural covariances	3	15.251	.002	.001	.001	-.001	-.001
Structural residuals	6	16.256	.012	.001	.001	-.002	-.002
Measurement residuals	15	61.648	.000	.003	.003	-.003	-.003

df = Degrees of freedom, χ^2 = Chi squared, NFI = Normative Fit Index, IFI = Incremental Fit Index, CFI = Comparative Fit Index, Δ CFI = Change in Comparative Fit Index, Δ NFI = Change in Normative Fit Index, Δ IFI = Change in Incremental Fit Index, RFI = Relative Fit Index, TLI = Tucker-Lewis Index

well known. The correlation between these variables, as found in this study, contributes to the evidence for convergent validity of the instrument.

One unexpected result is the poor correlation between Stress Balance and Resilience ($r=0.22$). One explanation might be that well-being captures a psychological state at a single point in time; resilience, on the other hand, is a dynamic concept, taking into account how someone has reacted to stressors in the past. The poor correlation between Stress Balance and Resilience may indicate that these are non-related concepts. The possibility of high well-being and yet low resilience has been noted before and is referred to as the 'resilience paradox' (Mguni et al. 2012).

There are some limitations to this study which should be considered. In the collection of the items to be included in the GWB Adolescent scale we have referred back to the original GWB study (Wiklund et al. 1994) and taken the complete set of bipolar adjective pairs identified from discussions with children, young adults and teachers. These included issues that were relevant to older adolescents such as sexual interest; unfortunately, as in the original research, it was not possible to validate all of these items. This suggests that there is still work to be done to include this aspect of an adolescent's experience in the concept of well-being. An important limitation is that the data is drawn from a narrow age category (17–20 years) due to a focus on young people in the last year of school. However, this is an interesting stage of life, on the cusp of adult life, and therefore provides a unique dataset.

Fig. 4 Model 3, boys in the combined training and test datasets ($N = 2686$)

GWBa = Gothenburg Well-Being scale for adolescents. The loading for each item is shown above the arrow on the left side. The correlation coefficients between the first-order and second-order factor are shown beside the lines between the factors. Standardised estimates are shown. Chi-square test ($\chi^2/df = 8.71$), Goodness of Fit Index (GFI = .982), Adjusted Goodness of Fit Index (AGFI = .966), Normed Fit Index (NFI = .982), Tucker-Lewis Index (TLI = .976) Comparative Fit Index (CFI = .984), Root Mean Square Error of Approximation (RMSEA = .054) and Standardised Root Mean Square Residual (SRMR = .019)

The low Cronbach's alpha for the total scale is disappointing. Other combinations of items that prioritised the alpha were possible but here we have focused on testing the model and applying model fit criteria. The intention with the GWB Adolescent scale is to capture a broad concept of well-being and, given that this model is theory based, an alpha of 0.6 can be considered acceptable for a new measure in a new population (Nunnally 1988). A two-item scale (Stress Balance) might not be considered to be sufficient to provide adequate identification for the construct, however, there were no other items that improved model fit theoretically and loaded on this construct. The Cronbach's alpha for Mood and Stress Balance was high (> 0.80), which is surprising, given the small number of items in these scales and the fact that Cronbach's alpha tends to underestimate the reliability of a scale when there are only a few items. This may indicate that some of these items are basically the same item repeated (Cortina 1993). Further cognitive debriefing for these items may therefore be indicated.

The difficulty of updating a questionnaire and factor structure have been explored in this study. Questionnaires must be allowed to evolve over time – either by renewing or updating the questions and analytic model at discrete time points or by allowing the questionnaire to operate

Fig. 5 Model 3, girls in the combined training and test datasets ($N = 2709$)

GWBa = Gothenburg Well-Being scale for adolescents. The loading for each item is shown above the arrow on the left side. The correlation coefficients between the first-order and second-order factor are shown beside the lines between the factors. Standardised estimates are shown. Chi-square test ($\chi^2/df = 7.64$), Goodness of Fit Index (GFI = .984), Adjusted Goodness of Fit Index (AGFI = .971), Normed Fit Index (NFI = .982), Tucker-Lewis Index (TLI = .977), Comparative Fit Index (CFI = .985), Root Mean Square Error of Approximation (RMSEA = .050) and Standardised Root Mean Square Residual (SRMR = .026)

dynamically, updating continuously in an online fashion. This process is made considerably easier where the measure is unidimensional and developed using item-response techniques (IRT). In an IRT framework, new questions could be added into existing measures without needing to re-validate the entire item bank. In this study we were aware that the original questionnaire was far too long and that a short form was required; the resulting 9-item questionnaire can be seen as a substantial improvement in terms of reduced burden on the respondent.

The importance of transparent reporting of the development, monitoring and validation of instruments should also be addressed here. We were in the privileged position of having access to some of the original focus group findings used in the development of the original questionnaire. However, this information was far from complete and there was no opportunity to complement it. Nonetheless, even this put us in a better position than other researchers wanting to update questionnaires. It should be a requirement of test construction that a repository for this information be created for the purpose of potential future revision.

This study has shown that the GWB Adolescent scale is appropriate for this age group, that it has good construct validity and is acceptable for use with a community of young people. The

Table 7 Model fit summary: measurement invariance across gender**Table 7.1. Model fit summary: testing measurement invariance across groups using multi-group confirmatory factor analysis**

	df	χ^2 goodness of fit	P value	NFI	IFI	CFI	Δ CFI
Male / Female							
Unconstrained	42	392.183	.000	.982	.984	.984	
Measurement weights	36	447.719	.000	.980	.982	.982	.002
Structural weights	34	452.819	.000	.979	.982	.982	.000
Structural covariances	33	456.488	.000	.979	.982	.982	.000
Structural residuals	30	475.319	.000	.978	.981	.981	.001
Measurement residuals	21	779.808	.000	.964	.967	.967	.014

Table 7.2. Model fit summary: testing measurement invariance across groups assuming that the unconstrained model is correct

	df	χ^2 goodness of fit	P value	Δ NFI	Δ IFI	RFI	TLI
Male / Female							
Measurement weights	6	55.535	.000	.003	.003	.000	.000
Structural weights	8	60.635	.000	.003	.003	.000	.000
Structural covariances	9	64.305	.000	.003	.003	-.001	-.001
Structural residuals	12	83.136	.000	.004	.004	-.001	-.001
Measurement residuals	21	387.625	.000	.018	.018	.010	.010

Table 7.3. Model fit summary: testing measurement invariance across groups assuming that model measurement weights are correct

	df	χ^2 goodness of fit	P value	Δ NFI	Δ IFI	RFI	TLI
Male / Female							
Structural weights	2	5.100	.078	.000	.000	-.001	-.001
Structural covariances	3	8.770	.033	.000	.000	-.001	-.001
Structural residuals	6	27.601	.000	.001	.001	-.001	-.001
Measurement residuals	15	332.090	.000	.015	.015	.010	.010

df = Degrees of freedom, χ^2 = Chi squared, NFI = Normative Fit Index, IFI = Incremental Fit Index, CFI = Comparative Fit Index, Δ CFI = Change in Comparative Fit Index, Δ NFI = Change in Normative Fit Index, Δ IFI = Change in Incremental Fit Index, RFI = Relative Fit Index, TLI = Tucker-Lewis Index

bidirectional nature of the questions means that positive affect and strength-based approaches are used, which is also in-line with the modern approach to well-being measurement. Future research is needed into the psychometric properties of the GWB Adolescent scale and how the questionnaire operates in different populations of young people. The relationship to resilience should be further investigated to determine the boundaries of these two concepts.

In conclusion, the nine-item, three-factor, strength-based GWB Adolescent scale offers reliable psychometric properties which will extend our understanding of well-being in older adolescents.

Acknowledgments The authors are thankful to all participants. Financial support was obtained by the Foundation Växthuset for Children, The Swedish Research Council (7509, 2006-7777), The Swedish Council for Working Life and Social Research (2006-1624, 2006-1506), The Swedish University Hospital (ALF) grants and Region Västra Götaland.

A translation of the Swedish GWB Adolescent and scoring key are available from the corresponding author.

Compliance with Ethical Standards

Conflict of Interest None of the authors have conflicts of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and

reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

- Amerijckx, G., & Humblet, P. C. (2014). Child well-being: What does it mean? *Children & Society*, 28(5), 404–415.
- Andrich, D. (1988). *Rasch models for measurement*. Newbury Park: Sage Publications.
- Avalos, L., Tylka, T. L., & Wood-Barcalow, N. (2005). The body appreciation scale: Development and psychometric evaluation. *Body Image*, 2(3), 285–297.
- Bergh, D. (2015). Sample size and chi-squared test of fit - a comparison between a random sample approach and a Chi-Square value adjustment method using Swedish adolescent data. In Q. Zhang & H. Yang (Eds.), *Pacific rim objective measurement symposium (PROMS) 2014 conference proceedings*. Berlin: Springer.
- Bollen, K. A., & Long, J. S. (Eds.). (1993). *Testing structural equation models*. Newbury Park: Sage.
- Brazier, J. E., & Tsuchiya, A. (2010). Preference-based condition-specific measures of health: What happens to cross programme comparability? *Health Economics*, 19, 125–129.
- Brazier, J. E., Deverill, M., Green, C., Harper, R., & Booth, A. (1999). A review of the use of health status measures in economic evaluation. *NHS R&D HTA programme, Health Technology Assessment*, 3(9).
- Bullinger, M., Kołtowska-Häggström, M., Sandberg, D., Chaplin, J. E., Wollmann, H., Noeker, M., & Brütt, A. L. (2009). Health-related quality of life of children and adolescents with growth hormone deficiency or idiopathic short stature - part 2: Available results and future directions. *Hormone Research*, 72(2), 74–81.
- Calfas, K., & Taylor, W. C. (1994). Effects of physical activity on psychological variables in adolescents. *Pediatric Exercise Science*, 6, 406–423.
- Chaplin, J. E., Kriström, B., Jonsson, B., Hägglöf, B., Tuvemo, T., Aronson, A. S., Dahlgren, J., & Albertsson-Wikland, K. (2011). Improvements in behaviour and self-esteem following growth hormone treatment in short prepubertal children. [research support, non-U.S. Gov't]. *Hormone Research in Paediatrics*, 75(4), 291–303.
- Chaplin, J. E., Kriström, B., Jonsson, B., Halldin Stenlid, M., Aronson, A. S., Dahlgren, J., & Albertsson-Wikland, K. (2012). When do short children realize they are short? Prepubertal short children's perception of height during 24 months of catch-up growth hormone treatment. *Hormone Research in Paediatrics*, 77(4), 241–249.
- Chaplin, J. E., Szakács, A., Hallböök, T., & Darin, N. (2017). The development of a health-related quality-of-life instrument for young people with narcolepsy: NARQoL-21. *Health and Quality of Life Outcomes*, 15(1), 135–135.
- Cheung, W. W., & Rensvold, R. B. (2016). Application 5: Testing factorial invariance of scales from a measurement scale (first-order CFA model). In B. M. Byrne (Ed.), *Structural equation modelling with AMOS: Basic concepts, applications, and programming* (3rd ed.). New York: Routledge.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104.
- Coverdale, G. E., & Long, A. F. (2015). Emotional wellbeing and mental health: An exploration into health promotion in young people and families. *Perspectives in Public Health*, 135(1), 27–36.
- Deci, E. L., & Ryan, R. M. (2008). Hedonia, eudaimonia and well-being: An introduction. *Journal of Happiness Studies*, 9(1), 1–11.
- Delfabbro, P. H., Winefield, A. H., Anderson, S., Hammarstrom, A., & Winefield, H. (2011). Body image and psychological well-being in adolescents: The relationship between gender and school type. *The Journal of Genetic Psychology*, 172(1), 67–83.
- Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, 95, 542–575.
- Diener, E. (2009). *Assessing well-being: The collected works of Ed Diener. Social indicators research series, volume 39*. Dordrecht: Springer.
- Dodge, R., Daly, A., Huyton, J., & Sanders, L. (2012). The challenge of defining wellbeing. *International Journal of Well-being*, 2(3), 222–235.
- Epstein, N. (1999). On being a patient: To dream. *Annals of Internal Medicine*, 131(7), 540–541.
- Erling, A. (1999). *Psychological development in children of short stature*. Unpublished doctoral dissertation, University of Göteborg.

- Erling, A. (2004). Why do some children of short stature develop psychologically well while others have problems? *European Journal of Endocrinology*, 151(Suppl 1), S35–S39.
- Erling, A., Wiklund, I., & Albertsson-Wikland, K. (1994). Prepubertal children with short stature have a different perception of their well-being and stature than their parents. *Quality of Life Research*, 3(6), 425–429.
- Erling, A., Wiklund, I., & Albertsson-Wikland, K. (2002). Psychological functioning in boys of short stature: Effects of different levels of growth hormone secretion. *Acta Paediatrica*, 91(9), 966–971.
- Frisén, A. (2007). Measuring health-related quality of life in adolescence. *Acta Paediatrica*, 96(7), 963–968.
- Gallagher, M. W., Lopez, S. J., & Preacher, K. J. (2009). The hierarchical structure of well-being. *Journal of Personality*, 77(4), 1025–1050.
- Gomez, M., Vincent, A., & Toussaint, L. L. (2013). Correlates of resilience in adolescents and adults. *International Journal of Clinical Psychiatry and Mental Health*, 1, 18–24.
- Guyatt, G. H., Feeny, D. H., & Patrick, D. L. (1993). Measuring health-related quality of life. *Annals of Internal Medicine*, 118, 622–628.
- Holmqvist, K., & Frisen, A. (2012). "I bet they aren't that perfect in reality": Appearance ideals viewed from the perspective of adolescents with a positive body image. *Body Image*, 9(3), 388–395.
- Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods*, 6(1), 53–60.
- Huppert, F. A. (2009). Psychological well-being: Evidence regarding its causes and consequences. *Applied Psychology: Health and Well-Being*, 1(2), 137–164.
- Kanning, M., & Schlicht, W. (2010). Be active and become happy: An ecological momentary assessment of physical activity and mood. *Journal of Sport & Exercise Psychology*, 32(2), 253–261.
- Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.
- Kocalevent, R. D., Zenger, M., Heinen, I., Dwinger, S., Decker, O., & Brähler, E. (2015). Resilience in the general population: Standardization of the resilience scale (RS-11). *PLoS One*, 10(11), e0140322.
- Konijnenberg, C., & Fredriksen, P. M. (2018). The effects of a school-based physical activity intervention programme on children's executive control: The health oriented pedagogical project (HOPP). *Scandinavian Journal of Public Health*, 46(21 suppl), 82–91.
- Krueger, A. B., & Stone, A. A. (2014). Psychology and economics. Progress in measuring subjective well-being. *Science*, 346(6205), 42–43.
- Linton, M. J., Dieppe, P., & Medina-Lara, A. (2016). Review of 99 self-report measures for assessing well-being in adults: Exploring dimensions of well-being and developments over time. *BMJ Open*, 6(7), 1–16.
- Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, 93, 568–592.
- Mguni, M., Bacon, N., & Brown, J. F. (2012). *The wellbeing and resilience paradox*. London: Young Foundation <https://youngfoundation.org/wp-content/uploads/2012/10/The-Wellbeing-and-Resilience-Paradox.pdf>. Accessed 27 Jan 2019.
- Moksnes, U. K., & Espnes, G. A. (2013). Self-esteem and life satisfaction in adolescents - gender and age as potential moderators. *Quality of Life Research: an International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation*, 22(10), 2921–2928.
- Mond, J., van den Berg, P., Boutelle, K., Hannan, P., & Neumark-Sztainer, D. (2010). Obesity, body dissatisfaction, and emotional well-being in early and late adolescence: Findings from the project EAT study. *Journal of Adolescent Health*, 48, 373–378.
- Nieman, P. (2002). Psychosocial aspects of physical activity. *Paediatrics and Child Health*, 7(5), 309–312.
- Nunnally, J. C. (1988). *Psychometric theory*. New Jersey: McGraw-Hill, Englewood Cliffs.
- Olsson, A., Hasselgren, M., Hagquist, C., & Janson, S. (2013). The association between medical conditions and gender, well-being, psychosomatic complaints as well as school adaptability. *Acta Paediatrica*, 102(5), 550–555.
- Ouvinen-Birgerstam, P. (1985). *I think I am: Assessment of children's and young people's self-esteem (jag tycker jag är: Bedömning av barns och ungdomars självkänsla. Manual)*. Stockholm: Psychology Society, Hogrefe & Huber Publishers.
- Parkin, D., & Devlin, N. (2006). Is there a case for using visual analogue scale valuations in cost-utility analysis? *Health Economics*, 15(7), 653–664.
- Pollard, E. L., & Lee, P. D. (2003). Child well-being: A systematic review of the literature. *Social Indicators Research*, 61(1), 59–78.
- Quitmann, J. H., Bullinger, M., Sommer, R., Rohenkohl, A. C., & Bernardino Da Silva, N. M. (2016). Associations between psychological problems and quality of life in pediatric short stature from patients' and parents' perspectives. *PLoS One*, 11(4), e0153953.
- Rose, T., Joe, S., Williams, A., Harris, R., Betz, G., & Stewart-Brown, S. (2017). Measuring mental wellbeing among adolescents: A systematic review of instruments. *Journal of Child and Family Studies*, 26(9), 2349–2362.

- Ryan, R., Huta, V., & Deci, E. L. (2008). Living well: A self-determination theory perspective on eudaimonia. *Journal of Happiness Studies*, 9(1), 139–170.
- Ryff, C. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, 57(6), 1091–1081.
- Sagone, E., & de Caroli, M. E. (2014). Relationships between psychological well-being and resilience in middle and late adolescents. *Procedia - Social and Behavioral Sciences*, 141, 881–887.
- Sandberg, D. E., & Voss, L. D. (2002). The psychosocial consequences of short stature: A review of the evidence. *Best Practice & Research. Clinical Endocrinology & Metabolism*, 16(3), 449–463.
- Sandberg, D. E., Brook, A. E., & Campos, S. P. (1994). Short stature: A psychosocial burden requiring growth hormone therapy? *Pediatrics*, 94(6 Pt 1), 832–840.
- Schipper, H. C., Clinch, J. J., & Olweny, C. L. M. (1996). Quality of life studies: Definitions and conceptual issues. In B. Spilker (Ed.), *Quality of life and Pharmacoeconomics in clinical trials* (pp. 11–23). Philadelphia: Lippincott-Raven Publishers.
- Schumacher, J., Leppert, K., Gunzelmann, T., Strauß, B., & Brähler, E. (2005). Die Resilienzskala - Ein Fragebogen zur Erfassung der psychischen Widerstandsfähigkeit als Persönlichkeitsmerkmal. *Zeitschrift für Klinische Psychologie, Psychiatrie und Psychotherapie*, (53), 16–39. www.mentalhealthpromotion.net/resources/resilienzskala2.pdf. Accessed 17 July 2018.
- Schwarzer, R., & Jerusalem, M. (1995). Generalized self-efficacy scale. In S. W. J. Weinman & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35–37). Windsor: NFER-NELSON.
- Sjöberg, A., Barrenas, M. L., Brann, E., Chaplin, J. E., Dahlgren, J., Marild, S., & Albertsson-Wikland, K. (2012). Body size and lifestyle in an urban population entering adulthood: The 'Grow up Gothenburg' study. *Acta Paediatrica*, 101(9), 964–972.
- Stewart-Knox, B., Duffy, M. E., Bunting, B., Parr, H., Vas de Almeida, M. D., & Gibney, M. (2012). Associations between obesity (BMI and waist circumference) and socio-demographic factors, physical activity, dietary habits, life events, resilience, mood, perceived stress and hopelessness in healthy older Europeans. *BMC Public Health*, 12, 424.
- Taillefer, M. C., Dupuis, G., Roberge, M. A., & LeMay, S. (2003). Health-related quality of life models: Systematic review of the literature. *Social Indicators Research*, 64(2), 293–323.
- Tsang, K. L., Wong, P. Y., & Lo, S. K. (2012). Assessing psychosocial well-being of adolescents: A systematic review of measuring instruments. *Child: Care, Health and Development*, 38(5), 629–646.
- von Eisenhart Rothe, A., Zenger, M., Lacruz, M. E., Emeny, R., Baumert, J., Haefner, S., & Ladwig, K.-H. (2013). Validation and development of a shorter version of the resilience scale RS-11: Results from the population-based KORA—Age study. *BMC Psychology*, 1(1), 1–7.
- Voss, L. D. (2001). Short normal stature and psychosocial disadvantage: A critical review of the evidence. *Journal of Pediatric Endocrinology and Metabolism*, 14(6), 701–711.
- Voss, L. D., & Sandberg, D. E. (2004). The psychological burden of short stature: Evidence against. *European Journal of Endocrinology*, 151(Suppl 1), S29–S33.
- Wagnild, G. M., & Collins, J. A. (2009). Assessing resilience. *Journal of Psychosocial Nursing and Mental Health Services*, 47, 28–33.
- Wagnild, G. M., & Young, H. M. (1993). Development and psychometric evaluation of the resilience scale. *Journal of Nursing Measurement*, 1(2), 165–178.
- Waterman, A. S. (1993). Two conceptions of happiness: Contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment. *Journal of Personality and Social Psychology*, 64, 678–691.
- WHO. (1948). WHO definition of Health. Available at: http://www.who.int/governance/eb/who_constitution_en.pdf. Accessed 27 May 2018.
- WHO. (2011). *Sixty-fourth world health assembly. Resolution WHA 64.28: Youth and health risks*. Geneva: WHO.
- Wiklund, I., Wiren, L., Erling, A., Karlberg, J., & Albertsson-Wikland, K. (1994). A new self-assessment questionnaire to measure well-being in children, particularly those of short stature. *Quality of Life Research*, 3(6), 449–455.
- Wirén, L., Bengtsson, B. A., & Johannsson, G. (1998). Beneficial effects of long-term GH replacement therapy on quality of life in adults with GH deficiency. *Clinical Endocrinology*, 48(5), 613–620.
- Wirén, L., Johannsson, G., & Bengtsson, B. A. (2001). A prospective investigation of quality of life and psychological well-being after the discontinuation of GH treatment in adolescent patients who had GH deficiency during childhood. *Journal of Clinical Endocrinology and Metabolism*, 86(8), 3494–3498.

Affiliations

Sarah Hitz¹ · Ebba Brann² · Kerstin Albertsson-Wikland³ · Zita Schillmöller¹ · John Eric Chaplin⁴

¹ Department of Health Science, Faculty of Life Sciences, Hamburg University of Applied Sciences, Hamburg, Germany

² Department of Public Health and Community Medicine, Institute of Medicine, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

³ Department of Physiology/Endocrinology, Institute of Neuroscience & Physiology, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden

⁴ Department of Paediatrics, Institute of Clinical Sciences, Sahlgrenska Academy at University of Gothenburg, Gothenburg, Sweden