Testing the factor structure of the Warwick-Edinburgh Mental Well-Being Scale in adolescents: A bi-factor modelling methodology

Title: Testing the Factor Structure of the Warwick-Edinburgh Mental Well-Being Scale in Adolescents: a Bi-Factor Modelling Methodology.

Dr. Stephen Shannon*[1,2], Dr. Gavin Breslin[1,2], Dr. Garry Prentice[3] and Professor Gerard Leavey[2].

[1]Sport and Exercise Sciences Research Institute, Ulster University Jordanstown Campus, Shore Road, Newtownabbey, Northern Ireland, BT370QB

[2]Bamford Centre for Mental Health and Well-being Ulster University, Magee Campus, Derry, Northern Ireland, BT487JL.

[3] Dublin Business School, 13 Aungier Street, Dublin 2.

Corresponding Author: *Dr Stephen Shannon, email: s.shannon@ulster.ac.uk

Conflict of interest

All authors have no conflict of interest to declare.

Ethical approval

Ethical approval was granted by Ulster University's ethical committee (approval number= REC/12/0201)

Informed consent

Informed consent was provided by all participants included in the study.

Abstract

Aim: Despite extensive use in mental health research and practice, limited evidence exists for the hypothesised unidimensional model of the Warwick-Edinburgh Mental Well-Being Scale in adolescents. Few studies have assessed competing Confirmatory Factor Analysis (CFA) models, and the instrument has yet to be assessed in younger adolescents in Northern Ireland, a jurisdiction characterised by high rates of mental illness.

Subject and Methods: School pupils (*n*=1,673) aged 13-18 years (M = 14.87, SD = 1.16), including 1,036 females, 997 urban children, and 312 from lower socio-economic status, completed psychometric tests. Seven CFA models based on extant research were tested, including unidimensional, bi-factor, higher-order and clustered.

Results: Several models, including the default unidimensional model, did not achieve recommended CFA fit thresholds. Model 6 comprising one strong 'general well-being' factor and three residual factors (i.e., figuratively labelled: 'Affective', 'Psychological Functioning' and 'Social Relationships') was confirmed as the superior model. Most item variance was explained by the general factor, relative to residual factors.

Conclusions: Adolescents predominantly conceptualise well-being as a unitary construct that coexists with relatively weak affective, psychological and social relationship domains. Researchers and practitioners should foremost calculate a composite score of well-being, and explore sub-domains to supplement understanding of adolescent well-being.

Keywords: Well-being; mental health; psychology; confirmatory factor analysis; validity;

Abbreviations

CFA= confirmatory factor analysis;

CFI= Comparative Fit Index

RMSEA = root mean square error of approximation

SDQ=Strengths & Difficulties Questionnaire;

STROBE =Strengthening the Reporting of Observational Studies in Epidemiology;

TLI = Tucker-Lewis Index

$\chi 2$= Chi-Square Value

WEMWBS = Warwick-Edinburgh Mental Well-Being Scale;

WRMSR = Weighted Root Mean Square Residual

Introduction

Well-being is defined as a state of optimal functioning (Ryan & Deci, 2017) and represents one dimension of a two continua model of mental health, wherein mental ill-being and well-being coexist as two distinct, but correlated factors (Keyes, 2005). Within the well-being construct, Keyes' (2002) theory outlines three factors of hedonic (i.e., positive affective states), eudemonic (e.g., psychological functioning, sense of purpose), and social (i.e., relationships, integration) well-being. Individual well-being is considered a measurable marker of societal health beyond economic indicators such as Gross Domestic Product (GDP) per-capita (Huppert, 2009), with educational success, living in a safe neighbourhood, family support, and economic prosperity all correlates of self-reported well-being (United Nations, 2015). Further, when high in well-being, one is at reduced likelihood of mental illness, and more likely to flourish in society (Keyes, 2007; Ryan & Deci, 2017). Most mental health disorders have their onset during early adolescence (i.e., age 11-14) (Jones, 2013), and relatedly, the importance of young people's well-being development is acknowledged as a human right in Article 12 of the United Nations Convention of the Rights of the Child (UNICEF, 1989). As such, the United Kingdom (UK) and Irish Governments aim is to continually assess and improve young people's well-being (Coulter, 2017), delineating the need for valid and reliable measurement instruments.

Adolescents conceptualisation of well-being may be diverse, wherein cultural contexts, socio-economic circumstances, and age-related developmental stages (e.g., puberty) likely exerting a role (Witten, Savahl & Adams, 2019). As such, competing psychometric measurement models require assessment among diverse population groups (Park, Han & Cho, 2011), and the accurate reporting of psychometric measures of well-being affects public health to the extent that an instrument's validity evidence informs clinical practice, research, economic and policy decisions (Doran, Wallace & Woods, 2014). Fried (2017) outlined that when observed items on a questionnaire correlate with a latent variable(s) (e.g., general well-being factor), it is more likely that a psychometric instrument is measuring its underlying 'true' or 'natural' construct. Confirmatory Factor Analysis (CFA) is a robust analytic method that can explain and establish construct validity among observed and latent variables (Hagger & Chatzsarantis, 2009). Specifically, researchers can assess competing CFA models through inspection of fit indices, helping determine whether items, structural pathways and/or factors are acceptable or require modifications (e.g., removal of items/constructs) (Schreiber et al., 2006). A number of potential CFA models can be specified, including; unidimensional (i.e., one underlying construct), higher order (i.e., general dimension as well as specific sub-dimensions), and clustered (i.e., multiple correlated dimensions with no general dimension) (Jackson et al., 2009). Recent bi-factor approaches to CFA also permit items to correlate with a strong general factor alongside residual sub-factors (Reise, 2012), wherein sub-factors and general factors are orthogonal, and compete for explaining item covariance (Beaujean, 2015).

Beyond CFA, a sound measure should display nomological validity, which refers to how the proposed construct correlates with constructs derived from the same theory (Hagger & Chatzsarantis, 2009). Lastly, to be confident in an instrument's responsiveness to change over time (e.g., in response to intervention), Terwee et al. (2007) indicate that a measure should not display floor or ceiling effects, wherein 15% of a sample score at the extreme ends of the scale.

The Warwick-Edinburgh Mental Well-Being Scale

The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) is a 14-item questionnaire, developed in the UK to standardise the measurement of the population's well-being (Tennant et al., 2007) and captures the positive well-being dimension of mental health. Respondents self-report their experiences using a 5-point Likert scale, and items are theorised to asses

Deleted: by capturing

hedonic or eudemonic well-being components (Tennant et al., 2007). The instrument development phase concluded a hypothesised unidimensional general well-being factor (Tennant et al., 2007; Stewart-Brown et al., 2009a). Psychometric support for the original hypothesised unitary WEMWBS measurement model has been found among diverse UK population groups, although relatively fewer studies exist among adolescents compared to adults (Stewart-Brown et al., 2009a; Stewart-Brown et al., 2009b; Maheswaran, Weich, Powell & Stewart-Brown, 2012; Melendez-Torres et al., 2019). Further evidence has shown that the unidimensional factor structure has not been robustly supported by recommended CFA fit thresholds (e.g., Kim et al., 2014; Lopez et al., 2015), including a study among adolescents showing an unacceptable fit (Hunter, Houghton & Wood, 2015).

Some authors have re-specified the model with item residual error covariances (e.g., Hunter, Houghton & Wood, 2015; Haver et al., 2015; Clarke et al., 2011), which may be unintentionally omitting theoretically meaningful variable(s) (Hermida, 2015). For example, in a community-based sample of adults in Austria, Lang and Bachinger (2017) found that the unidimensional model was not supported, but a bi-factor model comprised one strong general well-being factor alongside three residual domain factors (i.e., positive affect, psychological functioning and social relationships) was. That said, the retained model included item residual error covariances across different residual factors, which given such limited semantic overlap (e.g., 'I've been feeling optimistic about the future' with 'I've been feeling useful'), provides insufficient theoretical explanation (Reise, 2012). Furthermore, given bi-factor modelling restricts correlations between residual domain factors, the research may have benefitted from testing a higher order model, in which well-being may hypothetically subsume the shared variance between these domain factors (Beaujean, 2015). In another study with Argentine elders (Azcurra, 2015) a lack of fit was found for the unidimensional structure, but a clustered model comprising two covaried factors labelled as emotional and psychological well-being factors was retained. However, the authors did not test competing bi-factor or higher order models in their research.

Most notable for the rationale for the present study, in the few adolescent studies that have been conducted on the WEMWBS, there has yet to be a comprehensive assessment of competing CFA measurement models, with most adopting the traditional default unidimensional model (Reise et al., 2013). Such incomplete and contrasting statistical evidence is significant, as Fried (2017) has suggested that model misfit issues should be identified for improved construct clarity. Hence, iterations of Tennant et al.'s (2007), Lang and Bachinger (2017) and Azcurra's (2015) models require further assessment among adolescents. In addition to the theoretical underpinning of the structure of the instrument, the demography of samples included in WEMWBS psychometric validation studies in Northern Ireland (NI) (i.e., Lloyd & Devine, 2012; McKay & Andretta, 2017) was restricted to individuals aged $\geq$16. NI has been identified as having the poorest adult mental health rates across the UK and Republic of Ireland (Leavey, Galway, Rondon, & Logan, 2009). Given the aforesaid point that most mental health disorders have their onset in early adolescence (i.e., aged 11-14) (Jones, 2013), it is important to include a broader spectrum of adolescents (i.e., aged 12-18) for improved model accuracy in research (Wang et al., 2007; Park, Han & Cho, 2011).

Aims and Objectives

Given the extensive application of the WEMWBS in mental health research and practice, the current study responded to the contrasting statistical evidence on the WEMBWS factor structure in adolescents. Hence, the objective was to assess competing measurement models of the WEMWBS, and its psychometric properties (i.e., nomological validity, floor and ceiling effects) across a large, understudied, sample of adolescents who took part in the Northern Ireland Study of Adolescent Well-Being (NISAW; Corry & Leavey, 2017).

Materials and Methods

Study Design, recruitment and participants

This study adhered to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement. This was a cross-sectional study of adolescent school pupils in NI aged 13-18 years, collected as part of the NISAW (Corry & Leavey, 2017; Leavey et al., 2019; Leavey et al., 2020). Inclusion criteria was based on being a registered adolescent (aged 11-18) school pupil in NI. Individuals were excluded if they were under or above adolescent years, and resided outside the jurisdiction of NI. To achieve a representative sample, we randomly selected eight post-primary schools from Education and Library Board databases, which were stratified by markers of urbanicity, school type (i.e., Grammar/Secondary Modern), and socio-economic index (e.g., average income, lone parenthood). The research was ethically approved by Ulster University Research Ethics Committee. Further methodological information on the data collection procedures can be found in Leavey et al. (2019).

2.2 Outcome measures

Well-being

The WEMWBS (Tennant et al., 2007) was used to assess well-being. The instrument comprises 14-items designed to measure hedonic (e.g., happiness and life satisfaction), and eudemonic (i.e., relationships, self-actualisation) well-being components (Stewart-Brown et al., 2011). Each item is positively worded and scored on a 5-point Likert scale ranging from ‘none of the time’ (1), to ‘all of the time’ (5). The measure has undergone robust psychometrics (i.e., test-retest, convergent validity, discriminant validity) validated across adolescents in the UK, with a proposed unitary factor structure (Tennant et al., 2007; Stewart-Brown et al., 2009a). Total scores can range from 14 through to 70, with higher scores indicating higher well-being (Stranges et al., 2014; Fat et al., 2017).

Psychological difficulties

The Strengths & Difficulties Questionnaire (SDQ) (Goodman, 1997) was used to assess psychological difficulties such as emotional problems, conduct problems, hyperactivity and peer problems. The SDQ has been widely validated in youth samples (Theunissen, de Wolff, & Reijneveld, 2019). Due to construct overlap with the SDQ prosocial items (e.g., 'Other people my age generally like me') and social questions within the WEMWBS (e.g., 'I have been feeling close to other people'), we only computed the SDQ difficulties scale (McCrory & Layte, 2012). The SDQ difficulties scale showed an acceptable Cronbach's alpha value within the sample (α=.75). All items were scored on a 3-point scale (i.e., 0 =Not True, 1= Somewhat True, 2=Certainly True). Total difficulty scores can range between 0 to 40, with higher scores indicating more psychological difficulties.

Ill-being

A 3-item ill-being scale used in the NISAW study (Corry & Leavey, 2017) was used to measure common negative emotions reflective of psychological ill-being (Keyes, 2002). The items were scored on a 5-point Likert scale ranging from 1 (all to time) to 5 (none of the time), and included: "How often do you feel stressed?", "How often do you feel depressed?", and "How often do you feel like your emotions have become too much for you?". This scale demonstrated good internal consistency within the sample (α = .86) and higher scores reflect the presence of fewer negative emotions.

Home life

The stability and support in one's home environment was measured using a questionnaire from the Understanding Society UK Household Longitudinal Study (McFall & Garrington, 2011). The 12-item questionnaire measures aspects of home life including: sibling relationships; family activities and meals; and perceived support during times of difficulty. Items are scored on Likert scales ranging from 2 to 6-points, and negatively worded items (e.g., "In the past month, how many times have you stayed out after 9.00pm at night without your parents knowing where you are?") were reverse coded for analysis. Higher scores indicate a stable, supportive home environment (McFall & Garrington, 2011).

Parental relationships

An eight-item questionnaire from the Determinants of Adolescent Social Well-Being and Health study (DASH; Maynard & Harding, 2010) was used to measure the quality of parental relationships. This scale was internally consistent within the sample (Cronbach's $\alpha = .81$), and used a 4-point Likert scale (i.e., 1 = "Never" to 4 = "Always"). Items are scored so that higher scores are indicative of a more positive, healthy relationship with parents (Maynard & Harding, 2010), and therefore negatively worded items (e.g., "They treat me like a baby") were reverse coded for analysis.

Data Analysis

Detailed information on the data input, cleaning and preparation process can be found in Corry & Leavey (2017). Seven competing measurement models based on prior research were tested in Mplus (version 7) using Weighted Least Squares estimation. CFA was firstly conducted on the default hypothesised unidimensional structure (Model 1) (Tennant et al., 2007). Three model iterations based on the Azcurra (2015) study were then tested, including: a clustered model with two correlated factors (Model 2); a bi-factor model comprising a strong general factor and two orthogonal residual domain factors (Model 3), and; a higher order model comprising two correlated domain factors loading onto general well-being (Model 4). Furthermore, three iterations of Lang and Bachinger's (2017) model were tested, including: a clustered model (Model 5) with three correlated domain factors; the previously retained bi-factor model with one strong general factor, and three orthogonal residual factors of social relationships, affect and psychological functioning (Model 6), and; a higher order model with three correlated domain factors loading onto a general well-being factor (Model 7). Parameter estimates were fixed to 1, and cross-loadings on unintended factors were estimated at 0.

The adequacy of the competing models were assessed through inspection and comparison of multiple recommended goodness-of-fit indices (Hu & Bentler, 1999), with 5000 Bollen-Stine bootstraps to improve the accuracy of model parameters (Byrne, 2001). The Chi-Square [$\chi 2$] goodness-of-fit index was reported, however given that large sample sizes are sensitive to $\chi 2$ values (Bentler, 1990), we approached this value with caution. The Tucker-Lewis Index (TLI) and Comparative Fit Index (CFI) were reported with values of >.90 or >.95 considered as acceptable or good model fit, respectively; the root mean square error of approximation (RMSEA) with values of 0.08 or below considered acceptable, and; the Weighted Root Mean Square Residual (WRMSR) with values ~1 considered acceptable. We adopted the recommendations of Comrey and Lee (1992) for determining the perceived strength of factor loadings (i.e., <.30 = poor; >.45 = fair; >.55=good; >.63 = very good, and; >.71 = excellent).

After choosing a bifactor model as the superior model (see results below), factor weightings were entered in the Bi-factor Indices Calculator (Dueber, 2017). For items loading onto the general factor internal reliability was assessed through McDonald's omega hierarchical coefficient (ωH); McDonald's omega specific (ωS) was calculated to determine whether items loading onto their residual domains could reliability explain residual variances.

Further, total model variance was calculated through the explained common variance (ECV) value, that determines the ratio of variance explained by the general factor, divided by the variance explained by the general factor plus the residual domain factors. For interpretation, ωH values of > .80 and higher ECV values attributable to the general factor suggest the presence of a unitary construct (Reise, Schienes, Widaman & Haviland, 2013). A table comprising the fit statistics above was produced, and a second table for the retained model describing the items and corresponding factor loadings. A visual figure describing the retained model was also produced.

Ceiling and floor effects were calculated and determined present if more than 15% scored at the highest (i.e., 70) or lowest (i.e., 14) end of the WEMWBS respectively (Terwee et al., 2007). Nomological validity assessments were based on Keyes (2002) theory of mental health, which as previously discussed, encompasses two distinct but correlated mental well-being and mental ill-being continuums, represented by psychological, social and emotional factors (Keyes, 2002). Specifically the WEMWBS composite score was correlated with psychological difficulties (SDQ; Goodman, 1997) and the ill-being scale (NISAW study; Corry & Leavey, 2017) to reflect indicators of emotional, social and psychological ill-being (Keyes, 2002). Positive home environment (McFall & Garrington, 2011), and parental relationships (Maynard & Harding, 2010) were correlated as proxy measures of social well-being. Pearson's Product-Moment Correlation (r) was used with alpha significance set at $p < 0.05$. A correlation matrix was calculated and inputted into a table, considering values of 0.0 - 0.3 as weak, 0.31 - 0.70 as moderate, and 0.71 and above as strong (Field, 2013). Based on Keyes (2002) theory we hypothesised that the mental ill-being (i.e., ill-being and psychological difficulties) scales would be strongly inversely correlated with well-being, whereas the social factors (i.e., home life, parental relationships) would be moderately correlated.

Results

Data was collected from 1,673 participants comprising 1,036 females (61.9%) and 623 males (37.2%), with 14 participants (0.8%) not disclosing gender. Participant ages ranged between 13-18 (M = 14.87, SD = 1.16), with 997 (59.80%) living in an urban and 670 (40.20%) rural environment. Ethnicity was collected from 1,649 participants (98.6%), with 1,562 participants (93.4%) reporting 'White European' and 87 (5.2%) as 'minority ethnic groups (e.g., 'Indian' =0.7%; 'Irish Traveller' = 0.5%; 'Black African' = 0.4%, and; 'Mixed Ethnicity'=1.6%). In the total sample, 312 participants (20.9%) claimed free school meals which is a proxy measure of lower socio-economic status for youths in NI (Corry & Leavey, 2017). Table 1 details a description of the mean scores for the total sample and demographic groupings for each study outcome.

Please insert Table 1

Fit indices for the competing CFA models are presented in Table 2. Given the $\chi 2$ value sensitivities with sample size, all values were significant and thus did not lead to rejection of the models. All factor loadings in Model 1 were statistically significant ($p < 0.05$) ranging from 0.36 (item 4) to 0.80 (item 8) as reflected in the original unidimensional WEMWBS structure. However, the RMSEA value was above the recommended threshold of equal or close to 0.08 (i.e., 0.100). However, other fit statistics were within the acceptable-to-good ranges.

Similarly, all factor loadings were statistically significant in iterations of Azcurra's (2015) Models 2 and 4 ($p < 0.05$), but both RMSEA values were unacceptable. Further, the covariation pathway was .97 in Model 2, suggesting little distinct variation between the two constructs beyond that of a general factor. The bi-factor version of Azcurra's (2015) model (Model 3) yielded a substantially better fit, and the RMSEA value was close to acceptance at

0.087. With that said, none of the items significantly loaded onto their residual factors ($p >$ 0.05), suggesting very little contribution of the residual factors to the explained variance beyond that of the unitary well-being construct.

Lang and Bachinger's (2017) Models 5 and 7 iterations displayed significant ($p <$ 0.05) item factor loadings on their latent variables, but again were above the recommended RMSEA threshold. However, the bi-factor Model 6 yielded a substantially better fit, and was subsequently confirmed as the superior model among the competing CFA models. Specifically, all values were within the recommended confirmatory fit thresholds, including an excellent CFI value of 0.97. Applying Lang and Bachinger's (2017) figurative labels to the factors, Model 6 comprised a strong 'general well-being' factor, and three comparatively weak residual factors labelled: 'Affect'(n.b., items 1, 3, 5, 8, 10, 14), 'Social Relationships' (n.b., items 2, 4, 9, 12) and 'Psychological Functioning' (n.b., items 6, 7, 11, 13). As visually depicted in Figure 1, paths between the items and GF symbol refer to loadings on to the general well-being unitary construct, whereas the item loadings onto SR (Social Relationships), AF (Affect), and PF (Psychological Functioning) represent the sub-domain factors.

As further outlined in Table 3, all factor loadings onto the general factor were significant ($p < 0.05$), and ranged from 0.44 - 0.83, thus deemed in the 'fair' to 'excellent' ranges. All but two item loadings (n.b., 5 and 9) on residual factors were statistically significant ($p < 0.05$), and ranged from 0.07 – 0.59 (n.b., item 9). By including the residual factor loadings alongside the general factor loadings, up to 93% of item variance was explained (item 7), to as low as 26% (item 4). However, inspection of the construct paths on each side of the Figure 1's items demonstrated that general factor item loadings were consistently larger than residual factor item loadings. Notably, one item on each residual factor loaded negatively, albeit 'poorly' (i.e., items 12 and 13 = -0.12 and -0.13, respectively) or ‘fairly’ (i.e., item 1 = -0.32). In these cases, the items appear to have an opposite contribution to that factor compared to the other items loading onto that factor. Therefore, when calculating residual factor subscale scores, items 1, 12 and 13 would require subtraction from the remaining items.

Further model reliability assessment through the Bi-Factor Indices Calculator (Dueber, 2017), revealed a high (i.e., 0.92) omega hierarchical coefficient (ωH) for items loading onto the general factor, indicting that the composite score reliability assesses a unitary dimension of well-being. However, Omega scores for specific residual factors (ωS) suggested that the subscales have a sound ability to explain residual variances when assessed in isolation, i.e, 0.83 (Affect), 0.87 (Psychological Functioning), and 0.66 (Social Relationships). Yet, and aligned with the larger item loadings onto the general factor relative to the residual factor item loadings, ECV values showed that the general factor accounted for the substantial proportion of model variance (86%). However, residual factors did explain a significant proportion of item variance, and contributed to the model’s overall parsimony; i.e., 6% for Psychological Functioning, 6% for Affect, and 2% for Social Relationships. Taken the results collectively, adolescents perceive well-being as a largely unidimensional construct, with comparatively weak residual subdomains.

Please insert Table 2, 3 and Figure 1 here

In the retained model, no floor or ceiling effects were present, such that 0.05% reported the highest possible 70, and 0.01% reported the lowest possible of 14. The correlation matrix for the retained Model 6 factors and study outcomes is detailed in Table 4. All correlations were statistically significant at $p < .001$, and moderate, with r ranging from 0.36 to 0.70. As hypothesised, the composite score representing the strong general well-being factor showed larger correlations with ‘Ill-being’ and ‘Psychological Difficulties’ than that of

'Parental Relationships' and 'Home Life'. Within the WEMWBS residual factors, correlations were moderate-to-strong.

Please insert Table 4 here: *Correlation matrix for the retained Model 6 and study outcomes*

Discussion

This study tested the factor structure and psychometric properties of the WEMWBS in a large and diverse sample of NI adolescents and extends existing theoretical and methodological assessments of the WEMWBS. CFA conducted on seven competing CFA models found that, consistent with some prior research (Kim et al., 2014; Lopez et al., 2015; Santos et al., 2015), the original unidimensional model was not supported by recommended fit thresholds in adolescents. However, and for the first time in adolescents, an adapted model proposed by Lang & Bachinger (2017) comprising one strong general well-being factor, and three relatively weak residual factors was tested and confirmed as the superior model. Inspection of fit indices, factor loadings and model reliability confirmed the adequacy of the retained model, which did not display floor or ceiling effects. Moderate nomological validity correlations were present with several mental health outcomes. Overall, findings provide a theoretical and methodological basis for an adapted WEMWBS measurement model in mental health research and practice in adolescents.

CFA conducted on the originally theorised unidimensional WEMWBS model (Tennant et al., 2007; Stewart-Brown et al., 2009a) displayed an unacceptable RMSEA fit threshold, despite the factor loadings all displaying statistical significance. Findings provide further evidence of suboptimal fit for the unidimensional model (e.g., Kim et al., 2014; Lopez et al., 2015; Hunter, Houghton & Wood, 2015; Azcurra, 2015; Santos et al., 2015; Lang & Bachinger, 2017). While some studies among adolescents (e.g., Clarke et al., 2011) have found good-to-excellent fit indices for the unidimensional model among urban adolescents, Hunter, Houghton and Wood's (2015) study among rural adolescents did not. Having included a close to 60/40% split of urban to rural children, our study supports the view that adolescents may indeed have a broader conceptualisation of well-being (Santos et al., 2015). Therefore, it was pertinent to test competing alternative models in order to improve construct clarity (Fried, 2017).

Clustered and higher order versions of Azcurra's (2015) adapted WEMBWS model comprising 'positive emotions' and 'positive functioning' factors did not display any substantial changes in model fit compared to the unidimensional model. While initial inspections of a bi-factor version of Azcurra's (2015) model showed promise, namely positive changes across all fit indices, item loadings onto residual factors were not significant. Within bi-factor models, residual sub-factors should account for, at least in some part, item variance (Reise, 2012). Therefore, despite the improved fit indices there is limited contribution of the addition of subfactors in this case (Beaujean, 2015).

Including three well-being subfactors in clustered and higher order adaptations of Lang & Bachinger's (2017) model showed improvements in model fit indices compared to the unidimensional model, specifically through WRMR and RMSEA values. However, the bi-factor Model 6 was chosen as the superior model as findings (see Table 2) revealed the most positive fit indices. Indeed, all fit values exceeded or were within acceptable ranges (Hu & Bentler, 1999). Inspection of the three residual factors and their corresponding significant item loadings provided a further level of confidence in the retained model (see Table 3). Corresponding figurative labels by Lang & Bachinger (2017) were thus applied: (i) an 'Affective' hedonic domain wherein 'feelings' are emphasised; (ii) a 'Social Relationships' domain that is dependent on closeness of others, and; (iii) a 'Psychological Functioning' domain emphasising higher-order cognitive processes (e.g., problem-solving, interests). Importantly, we did not allow item residual errors to correlate with each other within or

Deleted: 1

across domain factors, as we did not see a sematic or theoretical justification for doing so (Hermida, 2015).

Model 6 arguably taps somewhat into Keyes' (2002) definition of positive mental health that comprises hedonic, eudemonic and social well-being domains. However, further analysis of Keyes' (2002) model by Jovanovic´ (2015) have, likewise to the present study, found more convincing evidence that one should foremost calculate a composite score of general well-being, and approach residual domain factors with caution. Specifically, and as outlined in Table 2, item loadings on the general factor were consistently larger than their respective residual factor loadings. A high (i.e., 0.92) omega hierarchical coefficient (ωH) for the unitary well-being construct was also found, and ECV values showed that common item variation was largely explained by the general factor (86%), with relatively weak contribution of the residual factors ranging from 2% to 6%.

Despite these findings, some residual factor item loadings were as high as 0.59 (i.e., item 9: 'I've been thinking clearly'), wherein the collective contribution of the general and residual factor resulted in 93% of item variance explained. Therefore, residual factors did contribute in small part to the model's overall parsimony, and subscales may be useful alongside the composite score in clinical practice, research, or public health provision (Doran, Wallace & Woods, 2014). For example, researchers and practitioners should foremost calculate a composite score of adolescent well-being, whilst supplementing their understanding of adolescent mental health through using specific residual subscales for further interpretation (Hajian-Tilaki, 2013). For example, the effect of school-based social support interventions can be assessed from the perspective of total and domain factors of adolescent well-being. However, given items 1, 12 and 13 loaded negatively, they appear to have an opposite contribution to that factor compared to the other items (DiStefano, Zhu & Mindrila, 2009), and would require subtraction from their remaining items.

An additional feature of our results was that the WEMWBS did not display ceiling or floor effects, meaning the WEMWBS will likely have sound responsiveness to change over time in intervention studies (Terwee et al., 2007). Moreover, moderate correlations were found between the WEMWBS and several additional mental health outcomes, including fewer psychological difficulties as assessed by the SDQ (Goodman, 1997), and reduced symptoms of ill-being. Consistent with Keyes' (2002) model, findings suggest that well-being and ill-being are linked, but likely distinct dimensions of the mental health construct.

Notwithstanding the study contributions, a limitation of our research was the cross-sectional design implemented. A longitudinal design would have provided the opportunity for test-retest reliability assessments, and if performed over a longer period (i.e.., transition from adolescence to adulthood) researchers could examine if individuals' well-being conceptions change over time. Furthermore, an additional measure of well-being such as the mental health continuum short-form (Lamers et al., 2011) and inclusion of a social desirability scale, would have helped provide a more comprehensive validity assessment.

Conclusion

Overall, our study suggest that adolescents' well-being is predominantly represented by an overarching strong general well-being factor, that coexists with relatively weak affective, psychological and social relationship domains. As such, practitioners, researchers and public health providers should foremost calculate a composite score of well-being in the knowledge that it is explaining the vast majority of the WEMWBS item variance. However, given some contribution of the residual factors to item and model variance, and sound subscale internal reliability, further analysis may calculate subdomain scores with the aforesaid recommendations - albeit strictly to supplement the composite score. The advantages of adopting these recommendations is that following analysis of the composite well-being score, domains (e.g., social relationships) relevant to the effect of mental health interventions (e.g., anti-bullying campaigns) could be explored. Limitations of this cross-sectional study are that the test-retest reliability, and concurrent and divergent validity of the WEMWBS remain unassessed in this population. Further longitudinal research study designs examining the WEMWBS with additional instruments can provide a more comprehensive psychometric assessment.

Acknowledgements

The authors would like to thank the school pupils and pastoral care staff for their help and participation in the study.

Study funding: The Research & Development Division of the Northern Ireland Public Health Agency through the Translational Research Group – Mental Health.

Author contributions: SS and GL lead on the design of the study. SS and GP performed the analysis. SS, GL, GP and GB contributed to the drafting of the article, and approval of the final version.

References

Azcurra, D. (2015). Translation, Spanish adaptation and validation of the Warwick-Edinburgh Well-being Scale in a sample of Argentine older adults. *Acta Colombiana de Psicología, 18*(1), 79-93.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246.

Beaujean, A. A. (2015). John Carroll's views on intelligence: Bi-factor vs. higher-order models. *Journal of Intelligence, 3*(4), 121-136.

Böhning, D., Böhning, W., & Holling, H. (2008). Revisiting Youden's index as a useful measure of the misclassification error in meta-analysis of diagnostic studies. *Statistical Methods in Medical Research, 17*(6), 543-554.

Byrne, B. M. (2001). Structural equation modeling with AMOS, EQS, and LISREL: Comparative approaches to testing for the factorial validity of a measuring instrument. *International Journal of Testing, 1*(1), 55-86.

Campbell, S., Nuevo-Chiquero, A, Popli, G., and Ratcliffe, A (2019). Parental Ethnic Identity and Child Development. IZA Discussion Paper No. 12104. Available at SSRN: https://ssrn.com/abstract=3390092

Comrey, A. L., & Lee, H. B. (1992). A First Course in Factor Analysis, 2nd Edn. Hillsdale, NJ: L.

Corry, D. A. S., & Leavey, G. (2017). Adolescent trust and primary care: Help-seeking for emotional and psychological difficulties. *Journal of adolescence, 54,* 1-8.

Coulter, A. (2017). Measuring what matters to patients. BMJ; 356-603 doi: https://doi.org/10.1136/bmj.j816

Clarke, A., Friede, T., Putz, R., Ashdown, J., Martin, S., Blake, A., & Stewart-Brown, S. (2011). Warwick-Edinburgh Mental Well-being Scale (WEMWBS): validated for teenage school students in England and Scotland. A mixed methods assessment. *BMC Public Health, 11*(1), 487.

DiStefano, C., Zhu, M., & Mindrila, D. (2009). Understanding and using factor scores: Considerations for the applied researcher. *Practical Assessment, Research, and Evaluation, 14*(1), 20.

Doran, P., Wallace, J., & Woods, J. (2014). Measuring Wellbeing in Northern Ireland: A new conversation for new times. Dunfermline: Carnegie UK Trust.

Dueber, D. M. (2017). Bifactor Indices Calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models. https://dx.doi.org/10.13023/edp.tool.01 [Accessed 09/01/2020: available at http://sites.education.uky.edu/apslab/resources/]

Fat, L. N., Scholes, S., Boniface, S., Mindell, J., & Stewart-Brown, S. (2017). Evaluating and establishing national norms for mental wellbeing using the short Warwick–Edinburgh Mental Well-being Scale (SWEMWBS): findings from the Health Survey for England. *Quality of Life Research, 26*(5), 1129-1144.

Field, A. (2013). Discovering statistics using IBM SPSS statistics: And sex and drugs and rock 'n' roll (4th ed.). London: SAGE.

Fried, E. I. (2017). What are psychological constructs? On the nature and statistical modelling of emotions, intelligence, personality traits and mental disorders. *Health Psychology Review*, *11*(2), 130-134.

Goodman, R. (1997). The strengths and difficulties questionnaire: a research note. *Journal of Child Psychology and Psychiatry, and Allied Disciplines, 38,* 581-586.

Goswami (2014). Children’s subjective well-being: socio-demographic characteristics and personality. *Child Indicators Research, 7*(1), 119-140.).

Gruebner, O., Rapp, M. A., Adli, M., Kluge, U., Galea, S., & Heinz, A. (2017). Cities and mental health. *Deutsches Ärzteblatt International, 114*(8), 121.

Hagger, M. S., & Chatzisarantis, N. L. (2009). Assumptions in research in sport and exercise psychology. *Psychology of Sport and Exercise, 10*(5), 511-519.

Hajian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian Journal of Internal Medicine, 4*(2), 627.

Haver, A., Akerjordet, K., Caputi, P., Furunes, T., & Magee, C. (2015). Measuring mental well-being: A validation of the short Warwick–Edinburgh mental well-being scale in Norwegian and Swedish. *Scandinavian Journal of Public Health, 43*(7), 721-727.

Hermida, R. (2015). The problem of allowing correlated errors in structural equation modeling: concerns and considerations. *Computational Methods in Social Sciences, 3*(1), 5-17.

Hoofs, H., Jansen, N. W. H., Mohren, D. C. L., Jansen, M. W. J., & Kant, I. J. (2015). The context dependency of the self-report version of the Strength and Difficulties Questionnaire (SDQ): A cross-sectional study between two administration settings. *PloS one, 10*(4), e0120930.

Huppert, F. A. (2009). Psychological well-being: Evidence regarding its causes and consequences. *Applied Psychology: Health and Well-Being, 1*(2), 137-164.

Hunter, S. C., Houghton, S., & Wood, L. (2015). Positive mental well-being in australian adolescents: Evaluating the Warwick-Edinburgh mental well-being scale. *The Educational and Developmental Psychologist, 32*(2), 93-104.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55.

Jackson, D. L., Gillaspy Jr, J. A., & Purc-Stephenson, R. (2009). Reporting practices in confirmatory factor analysis: An overview and some recommendations. *Psychological Methods, 14*(1), 6.

Jones, P. B. (2013). Adult mental health disorders and their age at onset. *The British Journal of Psychiatry, 202*(s54), s5-s10.

Jovanović, V. (2015). Structural validity of the Mental Health Continuum-Short Form: The bifactor model of emotional, social and psychological well-being. *Personality and Individual Differences, 75,* 154-159.

Keyes, C. L. (2005). Mental illness and/or mental health? Investigating axioms of the complete state model of health. *Journal of Consulting and Clinical Psychology, 73*(3), 539.

Keyes, C. L. (2007). Promoting and protecting mental health as flourishing: A complementary strategy for improving national mental health. *American Psychologist, 62*(2), 95.

Kim, S., Han-Yong, J., Na K-S., Lee, S,I., Kim, S-G, Lee, A, R., Cho, J,T., (2014). A validation of the Korean version of the Warwick-Edinburgh Mental Well-Being Scale. *Journal of Korean Neuropsychiatry Association, 53,* 237–245. doi:10.4306/jknpa.2014.53.4.237

Lamers, S. M., Westerhof, G. J., Bohlmeijer, E. T., ten Klooster, P. M., & Keyes, C. L. (2011). Evaluating the psychometric properties of the mental health continuum-short form (MHC-SF). Journal of clinical psychology, 67(1), 99-110

Lang, G., & Bachinger, A. (2017). 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. *Journal of Public Health, 25*(2), 135-146.

Leavey, G., Galway, K., Rondon, J., & Logan, G. (2009). A Flourishing Society: Aspirations for emotional health and wellbeing in Northern Ireland. Belfast, Northern Ireland: Northern Ireland Association for Mental Health.

Leavey, G., Corry, D., Divin, N., De Cock, P., & McConnell, G (2019). NISAW – Northern Ireland Study of Adolescent Wellbeing. Accessed (08/2019), available from: https://www.ulster.ac.uk/bamfordcentre/research/nisaw.

Leavey, G., Rosato, M., Harding, S., Corry, D., Divin, N., & Breslin, G. (2020). Adolescent mental health problems, suicidality and seeking help from General Practice: a cross-sectional study (Northern Ireland Schools and Wellbeing study). *Journal of Affective Disorders. 274,* 535-544.

López, M. A., Gabilondo, A., Codony, M., García-Forero, C., Vilagut, G., Castellví, P., & Alonso, J. (2013). Adaptation into Spanish of the Warwick–Edinburgh Mental Well-being Scale (WEMWBS) and preliminary validation in a student sample. *Quality of Life Research, 22*(5), 1099-1104.

Maheswaran, H., Weich, S., Powell, J., & Stewart-Brown, S. (2012). Evaluating the responsiveness of the Warwick Edinburgh Mental Well-Being Scale (WEMWBS): Group and individual level analysis. *Health and Quality of Life Outcomes, 10*(1), 156.

Melendez-Torres, G. J., Hewitt, G., Hallingberg, B., Anthony, R., Collishaw, S., Hall, J., & Moore, G. (2019). Measurement invariance properties and external construct validity of the short Warwick-Edinburgh mental wellbeing scale in a large national sample of secondary school students in Wales. *Health and Quality of Life Outcomes, 17*(1), 139.

McCrory, C., & Layte, R. (2012). Testing competing models of the Strengths and Difficulties Questionnaire's (SDQ's) factor structure for the parent-informant instrument. *Personality and Individual Differences, 52*(8), 882-887.

McDonald, R. P. (1999). *Test theory: A unified approach.* Mahwah, NJ: Erlbaum

McKay, M. T., & Andretta, J. R. (2017). Evidence for the psychometric validity, internal consistency and measurement invariance of Warwick Edinburgh mental well-being scale scores in Scottish and Irish adolescents. *Psychiatry Research, 255,* 382-386.

McKay, M. T., Dempster, M., & Byrne, D. G. (2014). An examination of the relationship between self-efficacy and stress in adolescents: the role of gender and self-esteem. *Journal of Youth Studies, 17*(9), 1131-1151

McMahon, E. M., Corcoran, P., O'Regan, G., Keeley, H., Cannon, M., Carli, V., & Balazs, J. (2017). Physical activity in European adolescents and associations with anxiety, depression and well-being. *European Child & Adolescent Psychiatry, 26*(1), 111-122.

Moriwaki, A., & Kamio, Y. (2014). Normative data and psychometric properties of the strengths and difficulties questionnaire among Japanese school-aged children. *Child and Adolescent Psychiatry and Mental Health, 8*(1), 1.

Park, J. M., Han, A. K., & Cho, Y. H. (2011). Construct equivalence and latent means analysis of health behaviors between male and female middle school students. *Asian Nursing Research, 5*(4), 216-221.

Pitchforth, J., Fahy, K., Ford, T., Wolpert, M., Viner, R. M., & Hargreaves, D. S. (2019). Mental health and well-being trends among children and young people in the UK, 1995–2014: analysis of repeated cross-sectional national health surveys. *Psychological Medicine, 49*(8), 1275-1285.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667-696.

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement, 73*(1), 5-26.

Ryan, R. M., & Deci, E. L. (2017). Self-determination theory: Basic psychological needs in motivation, development, and wellness. Guilford Publications.

Santos, J. J. A. D., Costa, T. A. D., Guilherme, J. H., Silva, W. C. D., Abentroth, L. R. L., Krebs, J. A., & Sotoriva, P. (2015). Adaptation and cross-cultural validation of the Brazilian version of the Warwick-Edinburgh mental well-being scale. *Revista da Associação Médica Brasileira, 61*(3), 209-214.

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323-338

Stewart-Brown, S., Tennant, A., Tennant, R., Platt, S., Parkinson, J., & Weich, S. (2009). 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 and Quality of Life Outcomes, 7*(1), 15.

Stewart-Brown, S. L., Platt, S., Tennant, A., Maheswaran, H., Parkinson, J., Weich, S., & Clarke, A. (2011). The Warwick-Edinburgh Mental Well-being Scale (WEMWBS):

a valid and reliable tool for measuring mental well-being in diverse populations and projects. *Journal of Epidemiology and Community Health, 65*(Suppl 2), A38-A39.

Stranges, S., Samaraweera, P. C., Taggart, F., Kandala, N. B., & Stewart-Brown, S. (2014). Major health-related behaviours and mental well-being in the general population: The Health Survey for England. *BMJ open, 4*(9), e005878.

Streiner, D. L., & Cairney, J. (2007). What's under the ROC? An introduction to receiver operating characteristics curves. *The Canadian Journal of Psychiatry, 52*(2), 121-128.

Taggart, F., Friede, T., Weich, S., Clarke, A., Johnson, M., & Stewart-Brown, S. (2013). Cross cultural evaluation of the Warwick-Edinburgh mental well-being scale (WEMWBS)-a mixed methods study. *Health and Quality of Life Outcomes, 11*(1), 27.

Tennant, R., Hiller, L., Fishwick, R., Platt, S., Joseph, S., Weich, S., & Stewart-Brown, S. (2007). The Warwick-Edinburgh mental well-being scale (WEMWBS): development and UK validation. *Health and Quality of life Outcomes, 5*(1), 63.

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, Daniëlle AWM, Knol, D. L., Dekker, J., de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34-42.

Theunissen, M. H., de Wolff, M. S., & Reijneveld, S. A. (2019). The strengths and difficulties questionnaire self-report: a valid instrument for the identification of emotional and behavioral problems. *Academic Pediatrics, 19*(4), 471-476.

Wang, P. S., Angermeyer, M., Borges, G., Bruffaerts, R., Chiu, W. T., De Girolamo, G., & Kessler, R. C. (2007). Delay and failure in treatment seeking after first onset of mental disorders in the World Health Organization's World Mental Health Survey Initiative. *World Psychiatry*, *6*(3), 177.

Witten, H., Savahl, S., & Adams, S. (2019). Adolescent flourishing: A systematic review. *Cogent Psychology*, *6*(1), 1640341.

Youngstrom, E. A. (2013). A primer on receiver operating characteristic analysis and diagnostic efficiency statistics for pediatric psychology: we are ready to ROC. *Journal of Pediatric Psychology*, *39*(2), 204-221.

Unicef. (1989). Convention on the Rights of the Child. Accessed online (07/2019) from: https://digitalcommons.ilr.cornell.edu/cgi/viewcontent.cgi?article=1007&context=child.

United Nations Department of Economic and Social Affairs Population Division. (2015). World population, ageing. In World population ageing. New York: United Nations.

Žukauskienė, R. (2014). Adolescence and Well-being. Handbook of child well-being: Theories, methods and policies in global perspective, 1713-1738.