Research Square

Preprints are preliminary reports that have not undergone peer review. They should not be considered conclusive, used to inform clinical practice, or referenced by the media as validated information.

Single- and Multi-Item Life Satisfaction in Adolescence: Measurement Properties, Convergence, and Domain Associations

Jose Marquez

jose.marquez@hmc.ox.ac.uk

Consejo Superior de Investigaciones Científicas

Wanying Zhou

University of Oxford

Leoni Boyle

University of Oxford

Laura Taylor

University of Oxford

Research Article

Keywords: adolescent wellbeing, life satisfaction, Satisfaction With Life Scale, single-item measures, measurement invariance, domain satisfaction

Posted Date: October 22nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11265618/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.

Abstract

This study examined the measurement properties of the Satisfaction With Life Scale (SWLS), its convergence with single-item Overall Life Satisfaction (OLS), and their associations with life-domain satisfaction. Participants were 6,445 students aged 11–18 from 38 International Baccalaureate schools across 24 countries, completing an English, Spanish or French survey. Analyses combined confirmatory factor analysis, measurement invariance tests, and paired correlation comparisons with school-bootstrap inference and multiplicity adjustment. The preferred five-item SWLS model showed good approximate fit and strong internal consistency. Models treating students as independent supported approximate metric and scalar invariance across age and gender groups. Across the observed survey-language groups, metric evidence was mixed and scalar invariance was not supported. SWLS–OLS convergence was substantial (r = .713), even after excluding the similarly worded SWLS item (r = .680). Estimated age and gender differences in convergence were small and did not survive multiplicity adjustment. Domain associations were broadly similar, but OLS correlated more strongly with student-life satisfaction, whereas SWLS correlated more strongly with satisfaction with material possessions and freedom. These small differences persisted after removing school mean differences and in English-only analyses. OLS serves as a brief concurrent indicator of adolescent life satisfaction, while SWLS supports reliability estimation and latent-variable modelling; the choice should follow from the intended application rather than from assumed equivalence.

Introduction

Life satisfaction is one of the most widely used indicators of adolescent subjective wellbeing (Li et al., 2026). It captures the cognitive or evaluative component of wellbeing: how young people judge the quality of their lives as a whole. Life satisfaction is an outcome of interest in itself. The OECD (2026b) summarises associations with engagement in learning, academic performance, social relationships and subsequent life satisfaction, while noting that these associations are not uniform. For schools and policy systems, valid and efficient measures of adolescent life satisfaction are therefore important for monitoring wellbeing, identifying groups at risk, and evaluating whether interventions are improving students' lives (OECD, 2026a, 2026b).

A recent scoping review identified 52 subjective wellbeing measures used in studies of adolescents aged 10–24, of which 19 had been specifically developed or adapted for adolescents, highlighting the diversity of available measurement approaches (Li et al., 2026). The Satisfaction with Life Scale (SWLS) was the most frequently used of these, appearing in 79 of the 315 reviewed studies, while a single-item overall life satisfaction (OLS) question was among the most common brief measures, appearing in 27. Both instruments are therefore well established in adolescent research, yet their correspondence within the same sample has rarely been examined directly.

This matters because international surveys have not converged on a common approach. The OECD (2026b) contrasts the multi-item instruments used with younger children in Children's Worlds (ages 8, 10

and 12) with the single-item approaches used with older adolescents in the Programme for International Student Assessment (PISA; age 15) and Health Behaviour in School-aged Children (HBSC; ages 11, 13 and 15). Even the two adolescent surveys differ: PISA asks directly about life satisfaction, as our OLS item does, whereas HBSC uses an adapted Cantril Ladder (Cantril, 1965). Evidence on convergence can therefore inform instrument selection, though comparability across surveys also depends on their populations, administration and specific measures. Here, we examine the narrower question of SWLS–OLS convergence within an adolescent sample; it does not establish equivalence with the child-specific multi-item instruments used in younger-child surveys (Children's Worlds, ages 8–12).

The SWLS (Diener et al., 1985) has an established adolescent validation literature. In Serbian adolescents aged 16–18, Jovanović (2016) supported a one-factor structure and strong measurement invariance across gender. Jovanović et al. (2022) subsequently examined 22,710 adolescents aged 13–19 across 24 countries and regions. A one-factor model allowing residual covariance between the two past-oriented items fitted well in 19 countries and regions. Partial metric invariance across all 24 required additional country-specific model modifications, while partial scalar invariance was supported in a subset of nine. Their pooled analyses supported scalar invariance across age groups and gender, although country-specific gender analyses and continuous-age analyses revealed qualifications. They advised caution in cross-cultural applications.

Using item response theory with 2,552 ninth-grade students in Los Angeles, Donaldson et al. (2025) found adequate discrimination for all five items, with the scale providing most information at low to average levels of life satisfaction. In adults, Swami et al. (2025) examined 56,968 respondents across 65 nations and 40 languages. They supported configural and metric invariance across the grouping variables examined and full scalar invariance across gender identities and age groups; cross-national and cross-language scalar comparability required partial-invariance approaches informed by alignment optimisation. Building on this work, the present study combines SWLS measurement assessment with comparisons against OLS and individual life-domain satisfaction items in a multilingual adolescent survey, including students younger than the lower age limit of the SWLS studies cited above.

The inclusion of adolescents younger than those in Jovanović et al. (2022) makes the functioning of the two past-oriented SWLS items particularly relevant. Items 4 and 5 concern what respondents have obtained from life and whether they would change their life if they could live it again. However, the present comparison of broad age groups does not isolate the functioning of these items specifically among 11- and 12-year-olds; the item-subset analyses should therefore be understood as sensitivity analyses rather than direct tests of developmental appropriateness. Kjell and Diener (2021) evaluated an abbreviated SWLS retaining Items 1–3, which corresponds to the SWLS-3 subset we examine below, and reported broadly comparable psychometric properties to the full scale in their samples. Their research included a preregistered adult study administering the abbreviated SWLS alongside the abbreviated Harmony in Life Scale, rather than extracting the short score from a five-item administration. Examining these items within the present study therefore offers a useful sensitivity check on convergence after excluding the two more past-oriented items. A separate question is whether convergence with OLS

depends on Item 3, whose wording closely resembles the OLS question. Removing that item permits an assessment of the association that remains, although it also changes scale content and reliability.

A further question is whether SWLS and OLS relate similarly to particular areas of adolescents' lives. Previous studies have compared the external associations of single- and multi-item life satisfaction measures. Among Serbian adolescents aged 16–18, Jovanović (2016) found strong convergence between the measures and generally similar associations with other wellbeing indicators. One exception was stress in Study 3, for which the single-item measure had a stronger negative association; this difference was not found in Study 2. Cheung and Lucas (2014) also found broadly similar patterns in three large adult samples, including comparisons with domain satisfaction. Likewise, Jovanović and Lazić (2020) reported strong convergence and generally similar external associations across six Serbian samples, although some differences depended on the measure and sample. We extend this work by comparing the correlations of OLS and the five-item SWLS with 14 life-domain satisfaction items. Given the cross-sectional design, these comparisons do not establish whether satisfaction with particular domains causes global life satisfaction or vice versa.

Previous analyses of the same sample compared OLS with the Cantril Ladder (Marquez et al., 2026) and examined a broader network of wellbeing indicators (Zhou et al., 2026). Marquez et al. found that the two single-item measures loaded on a common life evaluation factor but were not psychometrically equivalent: OLS showed the higher loading and predicted other wellbeing measures more strongly, including the SWLS. Zhou et al. identified OLS and SWLS Item 3 ("I am satisfied with my life") as highly central nodes across age groups, with student-life satisfaction also prominent, particularly among younger adolescents. Item 3 is also the SWLS item most closely resembling the OLS question. The present study addresses SWLS measurement properties, convergence with OLS, and symmetric comparisons of their associations with individual domains. The domain comparisons assess observed-score associations rather than network centrality or the relative explanatory contributions of an intercorrelated set of domains. These analyses extend the use of the existing sample; they do not provide independent replication of the companion papers.

Using data from 6,445 students aged 11–18 in 38 International Baccalaureate schools across 24 countries, we address three questions: (1) How well does the SWLS function psychometrically in this sample?; (2) How closely does it converge with single-item OLS?; (3) How similar are the two measures' associations with life-domain satisfaction? Age and gender comparisons concern SWLS measurement invariance and differences in SWLS–OLS convergence. Language analyses address measurement comparability, with English-only sensitivity checks for convergence and domain associations. The domain comparisons are exploratory, and the interpretation of all findings recognises the limits of this non-representative school sample.

The rest of the manuscript is organised as follows: the Method section describes the sample, measures and analytical strategy; the Results address SWLS measurement properties, convergence with OLS, and

comparisons of domain associations in turn; the Discussion considers the implications for selecting life satisfaction measures in adolescent surveys.

Method

Sample and procedure

Data came from an online wellbeing survey administered in May 2024 using Qualtrics. Schools were recruited through the International Baccalaureate school system, which enabled participation from multiple world regions but did not generate nationally representative country samples. The study received ethical approval from the University of [XXX]. A passive parental consent/opt-out procedure was used, and students provided active assent before starting the questionnaire. Students completed the questionnaire during school lesson time.

The final analytic sample included 6,445 students aged 11–18 from 38 schools in 24 countries. Of these, 3,343 were aged 11–14 and 3,102 were aged 15–18; 2,913 reported being boys, 3,335 girls and 197 preferred not to say. The survey was completed in English by 5,369 students, Spanish by 396 and French by 680. Fourteen countries contributed one school each. Both age groups were represented in 29 schools; the remaining nine schools contributed only older respondents. Survey language was concentrated within particular schools: 639 of the 680 French responses came from one Canadian school, and 358 of the 396 Spanish responses came from three schools in Argentina, Ecuador and Mexico. Table 1 summarises the sample and scores; Table A3 in Appendix 1 provides country composition.

The English questionnaire was adapted into French and Spanish with input from native-speaking collaborators. The translations were reviewed jointly by a native-speaking subject specialist with postdoctoral training, a professional translator, and a researcher. Official translations of questionnaire items were used where available. ChatGPT-4 was used to back-translate the French and Spanish versions into English, and these back-translations were compared with the source questionnaire to assess semantic consistency. The survey versions were finalised when the back-translations closely corresponded to the English wording.

Table 1
Sample characteristics and descriptive statistics

Variable	n / Mean	% / SD	Range
Age 11–14	3,343	51.9%	—
Age 15–18	3,102	48.1%	—
English survey language	5,369	83.3%	—
Spanish survey language	396	6.1%	—
French survey language	680	10.6%	—
Boys	2,913	45.2%	—
Girls	3,335	51.7%	—
Prefer not to say	197	3.1%	—
SWLS sum	23.84	6.55	5–35
SWLS-3 sum	15.05	4.01	3–21
SWLS-4 sum	18.73	5.28	4–28
SWLS-2 sum	9.95	2.72	2–14
OLS	6.66	2.15	0–10

Note. *N = 6,445. Percentages may not sum to 100 due to rounding. SWLS = Satisfaction With Life Scale; OLS = Overall Life Satisfaction.*

Measures

Satisfaction With Life Scale

Life satisfaction was measured using the five-item Satisfaction With Life Scale (SWLS; Diener et al., 1985). Items were rated on a 1–7 scale from strongly disagree to strongly agree. The five items ask respondents to evaluate: (1) whether their life is close to their ideal; (2) whether the conditions of their life are excellent; (3) whether they are satisfied with their life; (4) whether they have obtained the important things they want in life; and (5) whether, if they could live their life again, they would change almost nothing. In addition to the five-item sum score, we analysed the sum of Items 1–3, corresponding to the abbreviated SWLS-3 (Kjell & Diener, 2021). This provided a sensitivity check excluding the two more past-oriented items. Because all respondents completed the five-item SWLS, these analyses concern a subset of responses obtained within the full scale and do not validate a separately administered three-item questionnaire. We also created a four-item version excluding Item 3 (SWLS-4) and a two-item version using Items 1 and 2 only (SWLS-2). These subsets were used to examine whether

substantial convergence with OLS remained after excluding Item 3; they cannot isolate a wording effect because item removal also changes scale content and reliability.

Overall Life Satisfaction

Overall Life Satisfaction (OLS) was assessed with a single 0–10 item asking students to rate their satisfaction with life as a whole, where higher scores indicate greater satisfaction. This item is closely aligned with life satisfaction wording used in large-scale subjective wellbeing measurement and with the single-item life evaluation tradition discussed in OECD guidance (OECD, 2026a, 2026b).

Domain satisfaction

Domain-specific satisfaction was assessed using 14 items rated from 0 to 10: satisfaction with the people students live with, student life, learning at school, classmates, friends, the local area, material possessions, time use, safety, freedom, body image, the future, health, and relationships with teachers. Each item was analysed separately when comparing its correlations with SWLS and OLS. Higher scores indicated greater satisfaction.

Grouping variables

Age was grouped as 11–14 and 15–18 years. Survey language was English, Spanish or French. The gender variable distinguished boys, girls, and respondents who preferred not to say. Gender-specific invariance and convergence comparisons were restricted to boys and girls (n = 6,248). Respondents who preferred not to say were retained in the overall, age and domain analyses when otherwise eligible. The response categories do not provide a detailed assessment of gender identity.

Analytical Approach

Analyses addressed SWLS measurement properties, convergence with OLS, and comparisons of life-domain associations. We first examined descriptive statistics and internal consistency. Standardised-item Cronbach's alpha was estimated for the five-item scale and its three-, four- and two-item subsets. McDonald's omega was calculated for SWLS-3 and for the five-item scale under the preferred CFA specification. For the five-item calculation, the residual correlation between Items 4 and 5 was converted to a residual covariance on the standardised-item scale before inclusion in the denominator. These reliability estimates refer to standardised-item composites; the observed-score analyses used unweighted raw sums.

The baseline CFA specified one latent SWLS factor. The preferred model additionally allowed a residual covariance between Items 4 and 5. Both ask respondents to evaluate their life to date rather than their life at present, and this shared retrospective framing is expected to produce covariation beyond the

general factor; the same specification improved fit in most of the adolescent samples examined by Jovanović et al. (2022). Approximate fit was evaluated using the comparative fit index (CFI), Tucker–Lewis index (TLI), root mean square error of approximation (RMSEA) and standardised root mean square residual (SRMR). Values around .90–.95 for CFI/TLI, .06–.08 for RMSEA, and .08 for SRMR were treated as descriptive guidelines rather than universal decision rules; fit indices were interpreted jointly.

Measurement invariance was examined using separate multi-group models for age, survey language and gender. Age and gender models pooled respondents across languages. Configural, metric and scalar specifications were compared using changes in CFI, RMSEA and SRMR, alongside overall fit (Chen, 2007). The scalar specification constrained loadings and intercepts to equality while allowing non-reference-group latent means to be estimated. Where approximate scalar invariance was supported, latent mean differences were reported as secondary model-based estimates. Pooled age and gender invariance does not establish invariance within each language.

Convergence was described using Pearson correlations between OLS and the five-item SWLS sum, including subgroup estimates. Correlations with SWLS-3, SWLS-4 and SWLS-2 assessed sensitivity to item selection. Models relating a latent SWLS factor to observed OLS provided additional convergence estimates for the five-, three- and four-item versions. These models account for error in SWLS indicators but do not identify OLS reliability; the latent coefficients therefore do not represent correlations between two error-free constructs.

Age and gender differences in five-item SWLS–OLS correlations were tested directly. The age contrast was the correlation among students aged 15–18 minus that among students aged 11–14; the gender contrast was the correlation among girls minus that among boys. The two contrasts formed one multiplicity-adjustment family. Sensitivity checks repeated the age comparison in schools containing both age groups and repeated age and gender comparisons among English respondents; these three contrasts formed a separate family. Shared-school restriction addresses differences in school coverage, but does not equalise the groups' composition within those schools.

For each domain, we compared its Pearson correlation with observed OLS against its correlation with the five-item SWLS sum. The paired contrast was $\Delta r = r(\text{OLS, domain}) - r(\text{SWLS, domain})$. Positive values indicate a more positive correlation with OLS. All 14 contrasts used the same 6,445 respondents with complete values on the five SWLS items, OLS and the domain items. No common domain factor was imposed, and neither global measure was conditioned on the other. These are comparisons of observed-score associations and were not corrected for measurement error.

A within-school sensitivity analysis subtracted the school mean from each global score and each domain item before repeating the paired comparisons. This estimates correlations between individual deviations from school means, not an average of school-specific correlations. Whole schools were resampled, so each resampled school copy retained the same members and centring values. Pooled and within-school domain comparisons were also repeated in the English-only sample (n = 5,369), which

retained all 38 schools. English-only restriction is a sample sensitivity check, not an isolated test of translation effects.

For direct group contrasts and paired domain contrasts, uncertainty was estimated with 4,999 nonparametric whole-school bootstrap replications, jointly resampling the scores and grouping variables. The random-number seed was set to 20260909 before each bootstrap analysis. Two-sided normal-approximation *p* values based on bootstrap standard errors were adjusted using Holm's procedure (Holm, 1979). Domain contrasts formed a family of 14 tests within each specification; the all-language pooled analysis was primary and the remaining specifications were sensitivity checks. Pointwise normal-approximation 95% confidence intervals are reported, and percentile-bootstrap intervals were examined as a sensitivity check on interval estimation. These intervals are not simultaneous confidence intervals across the tested domains. Schools were treated as independent resampling units; country-level dependence was not separately modelled. Inference remains approximate with 38 unequal schools.

Analyses were conducted in Stata 15. Single-group CFA and latent convergence models used maximum likelihood with missing values; multi-group invariance models used standard maximum likelihood. These latent-variable models did not adjust for school clustering, unlike the bootstrap analyses. Direct group contrasts and domain comparisons were developed exploratorily during manuscript revision. Nonsignificant contrasts were not interpreted as evidence of equivalence.

Results

Reliability and factor structure of the SWLS

The five-item SWLS showed strong internal consistency: standardised-item alpha was .860 and omega was .855 under the preferred CFA specification. Item-rest correlations ranged from .596 for Item 5 to .770 for Item 3. The three-item subset also showed strong internal consistency (alpha = .840; omega = .841). Standardised-item alpha was .746 for SWLS-2.

The baseline one-factor CFA had CFI = .993, TLI = .985, RMSEA = .056 and SRMR = .016. Allowing a residual covariance between Items 4 and 5 improved fit, $\Delta\chi^2(1) = 46.17$, $p < .001$. The preferred model had CFI = .996, TLI = .990, RMSEA = .047 and SRMR = .011 (Table 2). Standardised loadings were .784, .748, .860, .683 and .628 for Items 1–5, respectively (Fig. 1). The residual correlation between Items 4 and 5 was .097, corresponding to a covariance of .055 on the standardised-item scale. This association is consistent with shared features beyond the general factor, but does not establish their source. The three-item CFA was just-identified, so its fit could not independently test the measurement structure.

Table 2
Reliability and confirmatory factor analysis model fit

Model	χ^2(df)	CFI	TLI	RMSEA	SRMR	Reliability / key estimate
SWLS-5 baseline one-factor CFA	107.62 (5)	.993	.985	.056	.016	α = .860
SWLS-5 preferred CFA with Item 4–5 residual covariance	61.45 (4)	.996	.990	.047	.011	α = .860; ω = .855; residual r = .097
SWLS-3 CFA	0.00 (0)	—	—	—	—	α = .840; ω = .841; just-identified
Latent SWLS-5 associated with OLS	143.55 (8)	.993	.987	.051	.014	β = .778
Latent SWLS-3 associated with OLS	40.82 (2)	.997	.991	.055	.010	β = .781
Latent SWLS-4 associated with OLS (Item 3 excluded)	49.16 (4)	.996	.991	.042	.011	α = .807; β = .760

Note. *SWLS = Satisfaction With Life Scale; OLS = Overall Life Satisfaction. Alpha and omega refer to standardised-item composites. Five-item omega was calculated under the preferred CFA, including the residual covariance between Items 4 and 5 on the standardised-item scale (.055); the corresponding residual correlation is .097. SWLS-3 is just-identified, so its fit does not test the measurement structure. Latent SWLS–OLS coefficients relate latent SWLS to observed OLS and do not correct for OLS measurement error. CFA and latent convergence models did not adjust for school clustering.*

Measurement invariance across age, language, and gender

Table 3 reports the separate multi-group analyses. Across age groups, the metric step changed CFI by − .002 and RMSEA by .000; the scalar step changed CFI by − .004 and RMSEA by + .005. These approximate-fit comparisons supported metric and scalar invariance between ages 11–14 and 15–18 in the pooled sample. In the scalar model, the older group had a lower latent mean (unstandardised difference = − .386, model-based SE = .032). This is a cross-sectional group difference, not evidence of within-person developmental change.

Across survey languages, metric evidence was mixed. CFI decreased by .005 and RMSEA increased by .004, but SRMR increased by .054. Although the combined CFI-based guideline did not indicate rejection at the metric step, the discordant SRMR change warrants caution about loading comparability across survey-language groups. Scalar invariance was not supported (ΔCFI = − .018; ΔRMSEA = + .024). Single-group model fit was less favourable in Spanish (CFI = .976; TLI = .940; RMSEA = .104; SRMR = .031) than in English or French (Table A1 in Appendix 1). These findings concern the language groups represented in this sample, whose school composition differed substantially.

Across boys and girls, the metric step produced ΔCFI = .000 and ΔRMSEA = − .009, and the scalar step produced ΔCFI = − .002 and ΔRMSEA = + .004, supporting approximate metric and scalar invariance. Girls had a lower latent mean in the scalar model (unstandardised difference = − .164, model-based SE = .032). Both latent mean comparisons use standard errors from models that treat observations as independent; they are secondary to the measurement comparisons and should be interpreted with the school-dependence limitation in mind.

Table 3
Measurement invariance by age group, survey language, and gender

Grouping	Model	CFI	RMSEA	SRMR	ΔCFI	ΔRMSEA	ΔSRMR	LR test vs previous
Age group	Configural	.996	.046	.011	—	—	—	—
Age group	Metric	.994	.046	.028	−.002	.000	+ .017	$\chi^2(4) = 30.15$, $p<.001$
Age group	Scalar	.990	.051	.028	−.004	+ .005	.000	$\chi^2(4) = 55.68$, $p<.001$
Survey language	Configural	.995	.052	.021	—	—	—	—
Survey language	Metric	.990	.056	.075	−.005	+ .004	+ .054	$\chi^2(8) = 73.42$, $p<.001$
Survey language	Scalar	.972	.080	.085	−.018	+ .024	+ .010	$\chi^2(8) = 261.31$, $p<.001$
Gender	Configural	.996	.045	.011	—	—	—	—
Gender	Metric	.996	.036	.014	.000	−.009	+ .003	$\chi^2(4) = 3.85$, $p=.427$
Gender	Scalar	.994	.040	.015	−.002	+ .004	+ .001	$\chi^2(4) = 36.48$, $p<.001$

Note. *Models use five SWLS items and allow residual covariance between Items 4 and 5. Age and gender models pool languages; they do not establish invariance within each language. Scalar models constrain loadings and intercepts, fix the reference-group latent mean at zero, and estimate the remaining latent means. Gender models include boys and girls only (n = 6,248); age and language models include 6,445 respondents. Approximate-fit changes were interpreted jointly with overall fit (Chen, 2007). Δ values are differences between the rounded indices displayed and may differ by .001 from differences based on unrounded estimates. These models did not adjust for school clustering.*

Convergence between SWLS and OLS

The five-item SWLS sum correlated substantially with OLS in the full sample (r = .713). Convergence remained substantial for SWLS-3 (r = .711), SWLS-4 excluding Item 3 (r = .680), and SWLS-2 using Items 1 and 2 (r = .652). The latent SWLS–OLS coefficients were .778, .781 and .760 for the five-, three- and four-item versions, respectively (Table 2). Thus, substantial convergence remained after excluding the item most similar to OLS, although these comparisons do not isolate the contribution of shared wording. Correlations by age and language are provided in Table A2 in Appendix 1.

The SWLS–OLS correlation was .718 among students aged 11–14 and .691 among students aged 15–18. The direct difference was − .028 (95% CI [− .059, .004]; Holm-adjusted p = .085). Correlations were .696 among boys and .724 among girls, yielding a difference of .027 (95% CI [.001, .053]; Holm-adjusted p = .083). The estimated differences were small and neither survived adjustment across the two primary contrasts (Table 4). The gender contrast's pointwise interval excludes zero, but that interval is not adjusted for multiplicity.

Restricting the age comparison to the 29 schools containing both age groups retained 3,343 younger and 2,385 older respondents. The older-group correlation increased to .703, yielding a smaller difference of − .015 (95% CI [− .049, .018]; adjusted p = .368). This indicates sensitivity of the estimated age contrast to school coverage, but does not demonstrate that school composition fully explains it.

In English-only analyses, correlations were .714 and .676 in the younger and older groups, respectively (difference = − .039; 95% CI [− .071, − .006]; adjusted p = .060). Correlations were .686 among boys and .715 among girls (difference = .029; 95% CI [.00004, .058]; adjusted p = .099). Neither contrast survived adjustment within the three-test sensitivity family. The findings support substantial convergence in each group, while leaving small group differences uncertain; they do not establish equivalence.

Table 4
Direct comparisons of SWLS-OLS correlations

Comparison	r reference	r comparison	Δr	95% CI for Δr	p Holm
Age: older − younger	.718	.691	−.028	[− .059, .004]	.085
Gender: girls − boys	.696	.724	.027	[.001, .053]	.083
Age: shared schools	.718	.703	−.015	[− .049, .018]	.368
Age: English only	.714	.676	−.039	[− .071, − .006]	.060
Gender: English only	.686	.715	.029	[.00004, .058]	.099

Note. *Correlations use the five-item SWLS sum and observed OLS. Reference/comparison groups are ages 11–14/15–18 for age and boys/girls for gender. Δr is the comparison-group correlation minus the reference-group correlation. Confidence intervals are pointwise normal-approximation intervals using standard errors from 4,999 whole-school bootstrap replications. Holm adjustment covers the first two rows as one primary family and the last three rows as a separate sensitivity family. Sample sizes (reference/comparison) are 3,343/3,102; 2,913/3,335; 3,343/2,385; 2,688/2,681; and 2,443/2,746, respectively. Shared-school analyses use 29 schools; other analyses use 38 schools in total. Confidence intervals are not multiplicity-adjusted. The English-only gender interval's lower bound is displayed to five decimal places because it is very close to zero.*

Comparisons of life-domain associations

All 14 domains were positively correlated with both global measures (Table 5). Correlations ranged from .364 to .596 for OLS and from .341 to .560 for SWLS. Paired differences were generally small, ranging from − .054 to .036. Across the 14 domains, the unweighted mean signed difference was .002 and the median was .007; nine differences were positive and five negative. These descriptive summaries indicate little average directional difference across the selected domains, although positive and negative differences can offset one another. Three differences survived Holm adjustment across the 14 domains. Student-life satisfaction correlated more strongly with OLS (r = .596) than with SWLS (r = .560), Δr = .036 (95% CI [.017, .054]; adjusted p = .002). Satisfaction with material possessions correlated more strongly with SWLS (r = .472) than with OLS (r = .418), Δr = − .054 (95% CI [− .068, − .040]; adjusted p < .001). Satisfaction with freedom also correlated more strongly with SWLS (r = .510) than with OLS (r = .481), Δr = − .029 (95% CI [− .046, − .013]; adjusted p = .004). The other 11 contrasts did not survive adjustment; this does not establish equal associations.

After subtracting school means, the same three differences survived adjustment: student life, Δr = .040; material possessions, Δr = − .052; and freedom, Δr = − .029 (Table A4 in Appendix 1). None of the 14 contrasts changed by more than .007 relative to the pooled analysis. The principal pattern therefore persisted when associations were estimated from individual deviations from school means.

English-only analyses retained 5,369 respondents from all 38 schools and again supported these three contrasts, both pooled and within school (Table A5 in Appendix 1). The pooled English differences were .043 for student life, − .057 for material possessions and − .034 for freedom. Learning at school

additionally survived adjustment in the English within-school specification (Δr = .034; 95% CI [.013, .056]; adjusted p = .020), but not in the primary analysis or the other sensitivity specifications. It is therefore treated as a less consistent sensitivity finding. Each bootstrap analysis completed 4,999 valid replications.

Table 5
Paired comparisons of life-domain correlations with OLS and SWLS

Domain satisfaction	r with OLS	r with SWLS	Δr	95% CI for Δr	p Holm
People lived with	.472	.481	−.009	[− .027, .008]	1.000
Student life	.596	.560	.036	[.017, .054]	.002
Learning at school	.435	.414	.021	[− .0004, .042]	.582
Classmates	.424	.406	.018	[− .001, .037]	.601
Friends	.364	.341	.022	[− .0003, .044]	.582
Local area	.383	.394	−.011	[− .031, .009]	1.000
Material possessions	.418	.472	−.054	[− .068, − .040]	< .001
Time use	.519	.517	.002	[− .016, .020]	1.000
Safety	.374	.385	−.011	[− .032, .011]	1.000
Freedom	.481	.510	−.029	[− .046, − .013]	.004
Body image	.516	.502	.014	[− .007, .035]	1.000
Future	.558	.543	.016	[− .003, .035]	.840
Health	.504	.495	.009	[− .011, .029]	1.000
Teacher relationships	.391	.386	.005	[− .016, .026]	1.000

Note. *Δr = r(OLS, domain) − r(SWLS-5 sum, domain). Positive values indicate a more positive correlation with OLS. Confidence intervals are pointwise normal-approximation intervals based on 4,999 whole-school bootstrap replications; they are not simultaneous intervals. Two-sided normal-approximation p values were adjusted using Holm's procedure across the 14 domains within this analysis. All comparisons use the same respondents. No correction for measurement error was applied. N = 6,445 across 38 schools. Student life, material possessions and freedom survive Holm adjustment.*

Discussion

The results provide a generally favourable assessment of the SWLS in this sample. The preferred one-factor model showed good approximate fit and strong internal consistency, and the pooled analyses supported approximate invariance across the two age groups and between boys and girls. SWLS and OLS were also strongly correlated, including when particular SWLS items were excluded. We found little clear evidence that this correlation differed by age or gender. Finally, most domain associations were similar, although small and reasonably consistent differences appeared for student life, material

possessions and freedom. These findings need to be interpreted in light of the multilingual but non-representative school sample.

The CFA results are broadly consistent with previous adolescent research. As in Jovanović et al. (2022), the preferred model allowed the residuals of the two past-oriented items to correlate. This does not tell us why these items share variance beyond the general factor. Donaldson et al. (2025) also found that Item 5 discriminated less strongly than some of the other items, but all five items functioned adequately in their item-response analysis. We therefore do not interpret the weaker loadings of Items 4 and 5 as evidence that they are unsuitable for adolescents. Similarly, our analysis of Items 1–3 is only a sensitivity analysis. Since these items were answered as part of the full SWLS, it does not validate the three-item scale as a separately administered questionnaire.

Approximate metric and scalar invariance across the two age groups and between boys and girls supports cautious comparisons under the fitted specifications. This broadly accords with Jovanović et al. (2022), whose pooled analyses supported scalar invariance across gender and seven age groups spanning ages 13–19. Their more detailed analyses nevertheless identified weak continuous-age moderation of the intercept and loading of Item 2, and gender invariance varied across countries. Similarly, our pooled results do not demonstrate identical cognitive interpretations at every age or within every language. The broad age bands may conceal more local differences, and the lower latent means among older respondents and girls are secondary, cross-sectional estimates from models without school-cluster adjustment. They should not be interpreted as within-person changes or nationally representative group differences.

The findings for the three survey-language groups were less clear. Changes in CFI and RMSEA were small at the metric step, but SRMR increased substantially, and scalar invariance was not supported. This differs from Swami et al. (2025), who supported metric invariance across 40 language groups in adults and obtained partial scalar invariance after allowing some parameters to vary. However, the studies differ considerably in their populations, language composition and methods. In our sample, most French responses came from one school and most Spanish responses from three schools. We therefore cannot determine whether the observed differences reflect language or translation, the participating schools and countries, or a combination of these factors. The English-only analyses show whether the main findings depend on including the other survey versions, but they do not isolate a language effect.

The full-sample SWLS–OLS correlation of .713 falls between the two estimates reported by Jovanović (2016), .668 and .734, and is close to the mean of .70 reported across six samples by Jovanović and Lazić (2020). These provide descriptive benchmarks rather than direct replication tests, because the samples and single-item wording or response formats differed. The association remaining after excluding Item 3 shows that convergence was not wholly dependent on that item, although the contribution of wording overlap cannot be separated from changes in content and reliability. The direct age and gender contrasts provided inconclusive evidence of small differences after multiplicity adjustment. The reduced age contrast in shared schools further cautions against a developmental

interpretation of the pooled difference. Non-significance should not be interpreted as proof that the correlations are equivalent.

The domain findings are consistent with earlier evidence of broadly similar external associations alongside occasional measure-specific differences (Cheung & Lucas, 2014; Jovanović, 2016; Jovanović & Lazić, 2020). Such differences were already present in the earlier comparisons. Our study adds estimates across a specified set of adolescent life domains and examines their consistency across school-centred and English-only analyses. OLS had a somewhat stronger association with student-life satisfaction, whereas SWLS had somewhat stronger associations with satisfaction with material possessions and freedom. These differences persisted after removing school means and in English-only analyses. Neither measure consistently produced stronger domain correlations. The differences are small, and we did not test them against any threshold of practical equivalence. What gives us some confidence in them is their persistence: the same three domains stand out after school means are removed, and again in the English-only sample. Whether differences of this size are practically important depends on the intended use. They warrant attention when comparing domain associations across studies, but our analyses do not establish whether they should influence instrument selection for routine monitoring.

Why these particular domains should differ is not something our design can answer. Item content, response format and reliability are all plausible contributors, and they are hard to separate here: OLS shares its 0–10 format with the domain items, and we did not estimate the reliability of OLS or of the individual domain items. Our comparisons are symmetric in that both global measures are correlated with the same domain in the same respondents, but they compare observed scores rather than error-free constructs, and they say nothing about which domains matter most or which might repay intervention.

The study complements the OLS–Cantril comparison (Marquez et al., 2026) and the broader wellbeing network analysis (Zhou et al., 2026) conducted with the same sample. The present contribution concerns SWLS measurement and its comparison with OLS, including direct tests of domain-correlation differences. These analyses use the same respondents and therefore do not provide independent corroboration of the companion papers. Network centrality and differences between domain correlations address different statistical questions; the present student-life finding should not be interpreted as evidence of a causal mechanism.

Implications

OLS may be useful when researchers need a brief concurrent indicator of adolescent life satisfaction. The SWLS requires more questionnaire space but provides several indicators, allowing researchers to estimate internal consistency and model measurement error. The most appropriate choice therefore depends on the purpose of the study. In a subsample of 79 university students, Jovanović and Lazić (2020) reported higher 10-month test–retest reliability for SWLS than for their single-item measure, although the uncertainty around both estimates was considerable. Our cross-sectional results cannot

determine whether the measures differ in temporal reliability, sensitivity to change, individual-level classification or school-level comparisons.

Researchers comparing domain associations across studies should record which global life satisfaction measure was used and avoid assuming that correlated instruments necessarily yield identical associations. At the same time, the modest differences observed here do not justify describing one instrument as generally superior. The study addresses the SWLS–OLS comparison within adolescence; it does not establish equivalence between the child-specific scales used in younger-child surveys and the single-item measures used in adolescent surveys. That broader measurement question requires the relevant instruments to be assessed directly in overlapping populations (OECD, 2026a, 2026b).

Limitations

The sample was drawn from International Baccalaureate schools and is not nationally representative. Fourteen countries contributed only one school, and respondent numbers and age composition varied substantially across schools. Country was not treated as a basis for nationally representative comparisons. The shared-school and within-school sensitivity analyses address specific aspects of composition, but do not remove selection effects or make the sample representative.

Language was highly concentrated within schools and countries. Although French responses came from 16 schools, 94.0% came from one school; 90.4% of Spanish responses came from three schools. Thus, the nominal number of schools with any response in a language overstates the breadth of substantial language representation. Cross-language measurement findings may reflect differences in the participating populations as well as translation or administration, and these explanations cannot be separated here.

The data were cross-sectional and all measures were self-reported in the same survey. Correlation differences do not establish causal direction, differential responsiveness to interventions, or distinct cognitive processes. The observed-score comparisons also cannot separate reliability differences from content and response-format effects. No equivalence margins were specified, so nonsignificant contrasts do not establish that the measures have interchangeable associations.

The CFA, invariance and latent convergence models treated observations as independent and SWLS items as continuous. School dependence was addressed in the direct correlation comparisons through whole-school bootstrapping, but not in the latent models. Bootstrap inference remains approximate with 38 unequal schools, and dependence between schools within countries was not separately modelled. Accordingly, the model-based latent mean standard errors and fit comparisons warrant caution. An ordinal treatment of SWLS responses could also yield different estimates.

The SWLS was developed for adults. Although its measurement properties were favourable here, analyses of item subsets administered within the full scale do not validate independently administered short questionnaires or establish their developmental suitability. Excluding Item 3 changed content and

reliability as well as wording overlap. Direct comparisons with child- and adolescent-specific multi-item instruments remain necessary, particularly for younger respondents.

Gender comparisons were limited to boys and girls. Respondents who preferred not to say were not a separately analysed gender group, and that response cannot be interpreted as a specific gender identity. The available categories therefore limit conclusions about gender diversity. Finally, the domain comparisons and direct group contrasts were exploratory developments during revision. Multiplicity adjustment limits false-positive risk within the defined test families, but does not replace independent replication of these findings.

Conclusions

The SWLS showed strong internal consistency and good approximate fit in this adolescent school sample under the preferred one-factor specification with correlated residuals for Items 4 and 5. Separate pooled analyses supported approximate age and gender invariance. Comparability across the observed survey-language groups was less secure, and the study could not separate language or translation effects from differences in the participating schools and countries. SWLS–OLS convergence was substantial, including after excluding the item most similar to OLS, and direct age and gender differences in convergence remained uncertain after multiplicity adjustment. Observed-score domain associations were broadly similar, with modest, consistent differences for student life, material possessions and freedom. These are descriptive comparisons and do not establish thresholds for choosing between instruments in a particular application. These findings support considering OLS as a brief concurrent indicator and SWLS when multiple indicators are useful, without establishing interchangeability across measurement purposes.

Declarations

Author Contribution

Jose Marquez designed the study, conducted the analyses, and drafted the manuscript. Wanying Zhou contributed to the study design, manuscript review and drafting. Leoni Boyle and Laura Taylor contributed to manuscript review and drafting.

Acknowledgement

The authors gratefully acknowledge the efforts of participating young people and their schools in this research.

Data Availability

This manuscript analyses the same dataset used in Marquez et al. (2026; https://doi.org/10.1007/s12187-026-10338-x) and Zhou et al. (2026; https://doi.org/10.1007/s10964-026-02330-z). It addresses distinct research questions concerning SWLS measurement properties, convergence with single-item life satisfaction, and comparisons of domain associations. Both publications and the shared sample are explicitly acknowledged in the manuscript.

References

1. Cantril, H. (1965). The pattern of human concern. Rutgers University Press.
2. Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 14(3)*, 464–504. https://doi.org/10.1080/10705510701301834
3. Cheung, F., & Lucas, R. E. (2014). Assessing the validity of single-item life satisfaction measures: Results from three large samples. *Quality of Life Research, 23(10)*, 2809–2818. https://doi.org/10.1007/s11136-014-0726-4
4. Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment, 49(1)*, 71–75. https://doi.org/10.1207/s15327752jpa4901_13
5. Donaldson, S. I., Beard, T. A., Soto, D., Lee, R., Leventhal, A. M., & Unger, J. B. (2025). Validating the Satisfaction With Life Scale among early adolescents: Psychometric assessment using item response theory. *Current Psychology, 44(7)*, 5821–5829. https://doi.org/10.1007/s12144-025-07518-3
6. Holm, S. (1979). A simple sequentially rejective multiple test procedure. Scandinavian *Journal of Statistics*, *6(2)*, 65–70. https://www.jstor.org/stable/4615733
7. Jovanović, V. (2016). The validity of the Satisfaction With Life Scale in adolescents and a comparison with single-item life satisfaction measures: A preliminary study. *Quality of Life Research, 25(12)*, 3173–3180. https://doi.org/10.1007/s11136-016-1331-5
8. Jovanović, V., & Lazić, M. (2020). Is longer always better? A comparison of the validity of single-item versus multiple-item measures of life satisfaction. *Applied Research in Quality of Life, 15*, 675–692. https://doi.org/10.1007/s11482-018-9680-6
9. Jovanović, V., Rudnev, M., Arslan, G., Buzea, C., Dimitrova, R., Góngora, V., Guse, T., Ho, R. T. H., Iqbal, N., Jámbori, S., Jhang, F.-H., Kaniušonytė, G., Li, J., Lim, Y.-J., Lodi, E., Mannerström, R., Marcionetti, J., Neto, F., Osin, E., … Žukauskienė, R. (2022). The Satisfaction with Life Scale in adolescent samples: Measurement invariance across 24 countries and regions, age, and gender. *Applied Research in Quality of Life, 17(4)*, 2139–2161. https://doi.org/10.1007/s11482-021-10024-w
10. Kjell, O. N. E., & Diener, E. (2021). Abbreviated three-item versions of the Satisfaction with Life Scale and the Harmony in Life Scale yield as strong psychometric properties as the original scales. *Journal of Personality Assessment, 103(2)*, 183–194. https://doi.org/10.1080/00223891.2020.1737093

11. Li, L., Sawyer, S. M., Newby, H., Dunstan, L., & Azzopardi, P. S. (2026). Quantitative measures of subjective wellbeing for adolescents: A scoping review and synthesis of comparable data. *The Lancet Child & Adolescent Health, 10(2)*, 135–146. https://doi.org/10.1016/S2352-4642(25)00282-2
12. Marquez, J., Zhou, W., Boyle, L., & Taylor, L. (2026). Cantril Ladder or Overall Life Satisfaction? Comparing 0–10 summary indicators of adolescent wellbeing. *Child Indicators Research, 19(3)*, 1111–1131. https://doi.org/10.1007/s12187-026-10338-x
13. OECD. (2026a). Measuring children's subjective well-being: Rationale, empirical approaches and future directions. OECD Publishing. https://doi.org/10.1787/a9b57ebc-en
14. OECD. (2026b). Why and how to measure children's subjective well-being. OECD Policy Insights on Well-being, Inclusion and Equal Opportunity, No. 25. OECD Publishing. https://doi.org/10.1787/78eb787c-en
15. Swami, V., Stieger, S., Voracek, M., Aavik, T., Abdollahpour Ranjbar, H., Adebayo, S. O., Afhami, R., Ahmed, O., Aimé, A., Akel, M., Al Halbusi, H., Alexias, G., Ali, K. F., Alp-Dal, N., Alsalhani, A. B., Álvarez-Solas, S., Amaral, A. C. S., Andrianto, S., Aspden, T., … Tran, U. S. (2025). Life satisfaction around the world: Measurement invariance of the Satisfaction With Life Scale (SWLS) across 65 nations, 40 languages, gender identities, and age groups. *PLOS ONE, 20(1)*, e0313107. https://doi.org/10.1371/journal.pone.0313107
16. Zhou, W., Marquez, J., Boyle, L., & Taylor, L. (2026). Mapping adolescent wellbeing: Developmental network differences between early to middle adolescence in 24 countries. *Journal of Youth and Adolescence, 55(6)*, 1489–1505. https://doi.org/10.1007/s10964-026-02330-z

Figures

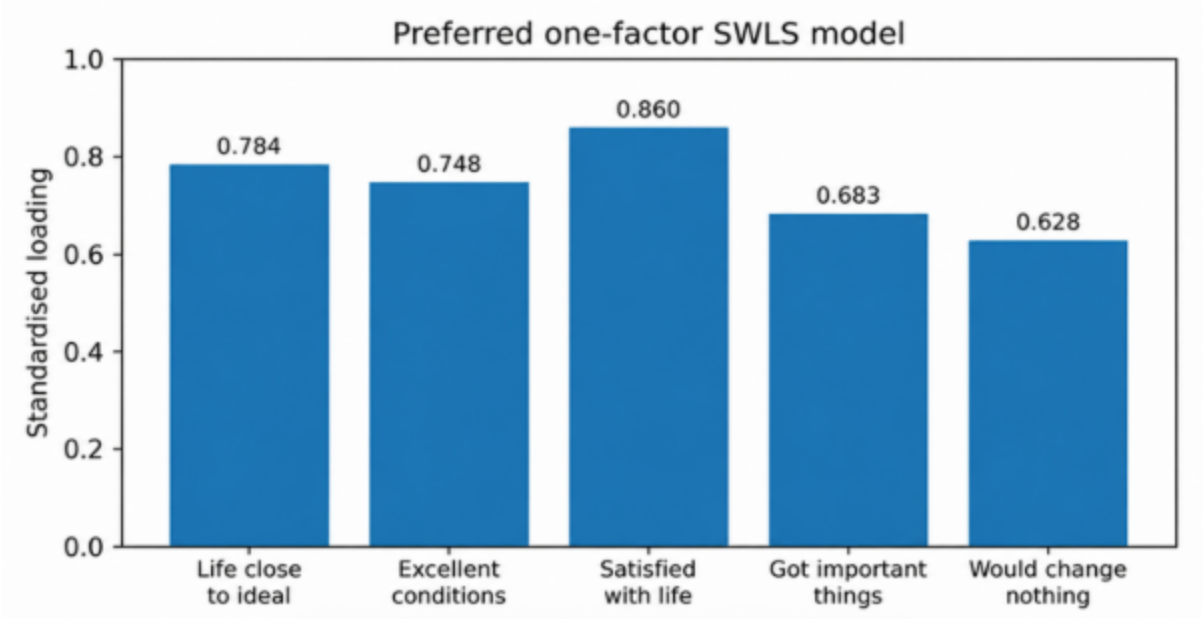

Figure 1

Standardised item loadings from the preferred one-factor SWLS model, allowing residual covariance between Items 4 and 5. The residual correlation is .097, corresponding to a covariance of .055 on the standardised-item scale. N = 6,445.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Appendix1.docx