Most people's life satisfaction matches their personality traits: True correlations in multi-trait, multi-rater, multi-sample data

René Mõttus, Anu Realo, Jüri Allik, Liisi Ausmees, Samuel Henry, Robert R. McCrae, Uku Vainik

Abstract

Despite numerous meta-analyses, the true extent to which life satisfaction reflects personality traits has remained unclear due to over-reliance on a single method to assess both and insufficient attention to construct overlaps. Using data from three samples tested in different languages (Estonian, N = 20,886; Russian, N = 768; English, N = 600), we combined self- and informant-reports to estimate personality domains' and nuances' true correlations (r_{true}) with general life satisfaction (LS) and satisfactions with eight life domains (DSs), while controlling for single-method and occasion-specific biases and random error, and avoiding direct construct overlaps. The associations replicated well across samples. The Big Five domains and nuances allowed predicting LS with accuracies up to r_{true} ≈ .80 to .90 in independent (sub)samples. Emotional stability, extraversion, and conscientiousness correlated r_{true} ≈ .30 to .50 with LS, while its correlations with openness and agreeableness were small. At the nuances level, low LS was most strongly associated with feeling misunderstood, unexcited, indecisive, envious, bored, used, unable, and unrewarded (r_{true} ≈ .40 to .70). Supporting LS's construct validity, DSs had similar personality correlates among themselves and with LS, and an aggregated DS correlated r_{true} ≈ .90 with LS. LS's approximately 10-year stability was r_{true} = .70 and its longitudinal associations with personality traits mirrored cross-sectional ones. We conclude that without measurement limitations, most people's life satisfaction is highly consistent with their personality traits, even across many years. So, satisfaction is usually shaped by these same relatively stable factors that shape personality traits more broadly.

Keywords: personality traits; life-satisfaction: well-being; multi-rater

General life satisfaction (LS) – an evaluative assessment of the overall degree of being satisfied with one's life (Heller et al., 2004) – is among the most desirable psychological outcomes and often an end unto itself, at least in the Western world. Historically the purview of religion and philosophy, studying LS's causes and psychological background now involves scientists from numerous fields working worldwide (Diener et al., 2018). Much of this work has focused on LS's degree of reflecting a broader range of relatively stable psychological characteristics, besides being directly influenced by short-term situational influences and more enduring life circumstances like culture, societal and economic processes, income, health, career, relationships and how people interpret these (e.g., Diener et al., 2018; Heller et al., 2004; Jagodzinski, 2010; Luhmann et al., 2012).

Many of the psychological characteristics are summarized with the Big Five or Five-Factor Model (FFM; Costa & McCrae, 1992) or HEXACO (Ashton & Lee, 2020) personality domains (Bainbridge et al., 2022). In the Big Five, neuroticism tends to have the strongest (r ≈ .40) and openness the weakest correlation (r ≈ .10) with LS. In the HEXACO, LS correlates the strongest with extraversion (r ≈ .40) and the weakest with emotionality, openness, and honesty-humility (r ≈ .10). These "Big Few" domains collectively account for about 30% of LS's variance (Anglim et al., 2020; Busseri & Erb, in press). Insomuch as the domains represent relatively – albeit far from fully – stable individual differences, LS's correlations with them are consistent with other evidence of its trait-like nature, such as moderate long-term stability (e.g., Lucas et al., 2018), similarity among genetically related people (Weiss et al., 2008) and visibility to others (e.g., Dobewall et al., 2013; Schneider & Schimmack, 2009).

Distinct but entangled

Regardless of its empirical correlations with personality traits, LS can remain conceptually distinct from them. On the one hand, people's differences in LS could mirror their personality traits in normal circumstances that allow them to shape and evaluate their lives according to their psychological and other traits.[1] In this case, LS can appear like any other

[1] Here, unusual circumstances would be those that are unrelated to people's own traits and that impose extreme constraints on people's freedom to live and/or assess their life according to their characteristics; examples could include active war-zones, extreme societal poverty or crime, or strict pandemic lock-downs. It is likely that most of our participants did not experience such unusual circumstances, although some may have experienced acute stress stemming from sources unrelated to their own characteristics (e.g., the death of a loved one). Data from Estonian- and Russian-speakers were collected during mild anti-pandemic measures that did not restrict most individuals' freedoms. Only our English-speaking participants were tested during a stricter pandemic lock-down.

trait – relatively stable, observable by others, and partly tracking individuals' genetic differences that provide distal backgrounds for any developmental aspect (Avinun, 2020; Bouchard, 2016; Johnson, 2010; Turkheimer et al., 2014). On the other hand, at least hypothetically, it may be possible to imagine these same people living in such dreadful circumstances (e.g., in an active war-zone or concentration camp) that the majority are unhappy with most aspects of their current lives, despite still differing in some personality traits that could otherwise track LS (e.g., self-discipline; Anglim et al., 2020). LS's empirical associations with personality traits would then be weakened, primarily due to the reduced variance in LS.

When present, empirical associations may imply personality traits' involvement in LS. For example, they are weakly but pervasively linked with many life outcomes that may contribute to LS, as well as with people's interpretations of their circumstances (Beck & Jackson, 2022; Soto, 2019). But this does not necessarily mean that these traits are LS's *directly* interpretable causes. Aspects of people's differences can become *indirectly* entangled over time as individuals strive towards circumstances that match their traits and possibly adapt their traits to the circumstances (Caspi et al., 2005; Johnson, 2010). For example, multiple personality traits are often linked with academic, occupational, relationship, lifestyle, and other outcomes (Seeboth & Mõttus, 2018; Soto, 2019; Stewart et al., 2022). Over time, these can further contribute to other outcomes and traits, including LS and the personality traits that influenced the outcomes in the first place (Caspi et al., 2005). Such reciprocal, crisscrossing interplay among traits and outcomes can lead to correlation patterns without easily discernible one-to-one causal relationships (Avinun, 2020). If this is the case, the *overall predictability* of LS from personality traits – the degree to which LS typically becomes aligned with personality – might be an equally meaningful research question compared to identifying which *specific traits* most closely track LS.

Here, we assumed that, in normal circumstances, LS is a relatively stable trait that both people themselves and others who know them well can evaluate with some degree of accuracy. Given this, our unique multi-trait, multi-rater design allowed us to ask *exactly how strongly* LS reflects numerous other traits when controlling for previously unresolved methodological issues such as single-method and occasion-specific biases, random error, and construct overlaps. In other words, with common measurement issues eliminated, can individuals' LS be accurately predicted from their personality traits, suggesting that it is usually shaped by these same factors that shape personality traits? And if so, which traits become particularly strongly linked with LS? Or is most of LS's variance unshared with personality traits, implying that it is largely shaped by social, cultural, situational, cognitive, and other factors that have little to do with personality more broadly? These are among LS research's most fundamental questions, and accurate answers to them will necessarily constrain theorizing on LS's nature and origin (Diener et al., 2018; Heller et al., 2004). Currently, however, these answers are inconclusive despite hundreds of studies and multiple meta-analyses (DeNeve & Cooper, 1998; Steel et al., 2008; Anglim et al., 2020).

Need to move beyond single-method studies

The typically reported correlations between LS and personality traits may misrepresent their overlaps. This is because most studies have relied on self-ratings to assess both, likely overestimating their associations due to shared single-method effects like biased self-perception or characteristic response styles (Paulhus & Vazire, 2007) that can make up much of trait score variance (McCrae & Mõttus, 2019). In cross-sectional data, correlations may also be inflated due to occasion-specific short-term effects, such as mood fluctuations or recent events. Conversely, random measurement error and raters' idiosyncratic interpretations of each construct's measures can attenuate the correlations. So, observed correlations like .10 or .40 may be either inflated estimates of much weaker or even non-existent "true" associations, or attenuated estimates of much stronger true associations. Substantial overlaps among personality traits can further distort the correlations (Busseri & Erb, in press).

Combining self-reports with other information sources can help better approximate the correlations' true magnitude (Schimmack, 2010). Ratings by informants like partners, friends, or relatives provide one such source (Vazire, 2006) and show at least moderate and comparable agreement with self-reports for both LS (Schneider & Schimmack, 2009) and personality traits (Connelly & Ones, 2010). Despite numerous calls for multi-rater designs (Anglim et al., 2020; Diener et al., 2018), they remain rare (Dobewall et al., 2013; Schimmack et al., 2004), especially in large multi-sample studies that are most likely to provide robust estimates.

Disattenuating for invalidity

Self-reported LS's correlations with informant-reported personality traits, and the other way around, are not inflated by shared single-method or (measurement) occasion-specific biases. However, they are attenuated by imperfect cross-rater agreement on both constructs – for example, due to different access to trait-relevant information or each rater's idiosyncratic interpretations of personality trait and/or LS measures –, occasion-specific effects, and random error. But these factors also attenuate raters' same-trait correlations, so ratios of average (across the two directions) cross-rater, *cross*-trait correlations to the average of the two cross-rater, *same*-trait correlations approximate traits' true associations, free of single-method and occasion-specific biases and random error.

This approach exactly parallels the familiar method of disattenuating monomethod correlations for unreliability, in which two variables' (*x* and *y*) raw correlation is divided by the square root of the product of the reliabilities of the two variables:

$$\frac{r_{xy}}{\sqrt{r_{xx} * r_{yy}}}.$$

This provides an estimate of the correlation that would be observed if both measures were perfectly reliable. In the present study, we divide the cross-method, cross-variable correlation by the square root of the product of the cross-method validities, such as:

$$\frac{r_{x(self)\,y(informant)}}{\sqrt{r_{x(self)\,x(informant)}\,r_{y(self)\,y(informant)}}}.$$

A second estimate of this value is given by disattenuating the complementary cross-method correlation, $r_{x(informant)y(self)}$, and we define true correlations, r_{true}, as the geometric mean of these two, so:

$$r_{true} = \sqrt{\frac{r_{x(self)\,y(informant)}\,r_{x(informant)\,y(self)}}{r_{x(self)\,x(informant)}\,r_{y(self)\,y(informant)}}}.$$

This is the correlation that would be observed if both measures were perfectly reliable and valid.

Need to move beyond broad trait domains

Because traits are hierarchically organized, broad domains may misrepresent LS's relations with personality traits. Domains can be subdivided into a few dozens of narrower traits, facets, and these further into many dozens of yet narrower traits, nuances, that also demonstrate the essential properties of traits such as relative stability over time, cross-method correlations, and partially unique etiologies (McCrae, 2015; Mõttus et al., 2019). Facets and nuances often hold unique information about life outcomes and other traits (e.g., Revelle et al., 2021; Seeboth & Mõttus, 2018; Stewart et al., 2022). LS is likely no exception, attested by its different correlations with supposedly parallel domain and facet scales that combine different nuances (Anglim et al., 2020), such as those considered similar in the Big Five and HEXACO (Thielmann et al., 2021) or assessed with different Big Five questionnaires.

LS's evaluations may directly overlap with some personality facets and nuances, hence trivially inflating their domains' correlations with LS (Steel et al., 2008; Wood & Harms, 2016). For example, LS correlates most strongly with scales asking people about self-esteem, happiness, and optimism and not feeling depressed, hopeless, and inferior to others (Anglim et al., 2020). These traits – hidden behind facet labels like depression and positive emotions – could be among life quality's definitional characteristics for many people, which would be evidenced by their r_{true}s with LS items being nearly equal to or even higher than LS items' r_{true}s among themselves (Campbell & Fiske, 1959). Net of such directly overlapping facets and/or nuances within them, LS may reflect personality traits to a lesser degree than typical estimates show (Mõttus, 2016).

But many domains' constituent traits could also have meaningful links with LS such as agreeableness' trust facet or conscientiousness' achievement-striving and self-discipline facets, or nuances within these facets (Anglim et al., 2020). For example, a sociability facet's nuance about enjoying others' company might be more strongly linked with LS than its talkativeness nuance. Moreover, LS can be linked with specific personality traits not yet covered by most Big Five and HEXACO measures. For instance, given LS's link with relative as well as absolute income (Boyce et al., 2010; Cheung & Lucas, 2016), envy may be one narrow trait tracking low LS (Rentzsch & Gross, 2015). Or, given LS's links with having strong relationships (e.g., Diener & Seligman, 2002), low LS may have a distinct association with a tendency to feel mistreated/alienated. In this case, using only domains or even their commonly assessed facets may underestimate the overall extent to which LS reflects personality traits, let alone the associations' details.

A systematic description of LS's correlations with a range of personality nuances is currently lacking, but it would help to better understand LS's broader psychological background. In particular, combining self-ratings with informant ratings to approximate true associations makes nuances' and broader traits' degrees of reflecting LS directly comparable, allowing nuances' distinct associations with LS to emerge more clearly, should they exist. For example, not only LS but many other desirable life outcomes tend to go with desirable levels of (nearly) all Big Five domains, whereas high LS may correspond to a more distinctive nuance-level profile (Stewart et al., 2022).

Need to move beyond a single way to assess satisfaction

If LS is defined as people's satisfaction with their *lives* rather than with *themselves*, its evaluation should reflect a broad combination of satisfactions with life's specific domains, such as work, financial and residential circumstances, and relationships (Payne & Schimmack, 2020). If so, LS should track a range of domain satisfactions (DSs) and especially their aggregate variance, and the DSs and LS should have similar correlation patterns with personality traits. Theoretically, this could show the extent that population variance in being satisfied reflects a general trait rather than many domain-specific evaluations, possibly because the same personality traits are similarly, if indirectly, linked with how people shape different aspects of their lives and evaluate these. From a methodological perspective, assessing DSs beside LS could mitigate the risk that correlations between personality traits and satisfaction are merely due to superficial overlaps in constructs or measurements: even if unspecific LS assessments (e.g., "Am happy with my life") may be directly based on behaviors, thoughts, and feelings also asked about to assess personality traits (e.g., "Am energetic"), this could be less likely for individual DSs (e.g., "Am happy with *my relationships*" or "Am happy with *where I live*"). Therefore, we operationalize satisfaction as both general LS and a combination of eight specific DSs, estimating their true associations among each other and with personality traits.

Personality traits might track with LS more strongly than with a broad combination of DSs. This may be because people assess their life quality based on their personal characteristics besides their life circumstances *per se* (Heller et al., 2004), the LS's links with personality traits are inflated by construct/measurement overlaps that researchers could not avoid, and/or researchers did not consider all relevant DSs. Therefore, we are skeptical that any given research design could fully disentangle the so-called "bottom-up" and "top-down" causal explanations (Heller et al., 2004; Payne & Schimmack, 2020) whereby, respectively, personality traits are linked with satisfaction via shaping different life domains and evaluations of these (personality traits —> DSs —> LS) *versus* primarily tracking general satisfaction that then influences satisfactions with different life domains (personality traits —> LS —> DSs). Besides, these explanations are not mutually exclusive (Heller et al., 2004). Here, we assessed both LS and DSs to study satisfaction's construct validity and the robustness of its links with personality traits to different ways of operationalizing it.

Need for multi-sample studies

LS's correlations with social and economic factors can vary across cultural and societal circumstances (Oishi et al., 1999; Suh et al., 1998), and so could its associations with personality traits. For example, although the domains of positive and negative emotionality, respectively resembling the extraversion and neuroticism domains, are linked with LS, these associations' strengths can vary, with the former being stronger in individualist countries and the latter in countries valuing self-expression over survival (Kööts-Ausmees et al., 2013; Kuppens et al., 2008).

So, research estimating LS's (true) associations with personality traits should examine the findings' robustness across samples with diverse backgrounds. It is possible, for example, that narrower traits' links with LS are less generalizable than those of broad personality domains because subtle cultural and societal effects may be diluted in the latter. Likewise, using a multi-rater design to control for methodological issues may either dampen or magnify cross-sample variations if single-method associations have been differentially biased in different samples. In any case, the degree of the links' robustness across samples speaks to the extent to which LS's variance reflects personality traits, besides being directly sensitive to circumstances that vary between samples and do not influence personality traits more broadly. Here, we examine the robustness of LS-personality trait links across three samples. While all samples are predominantly of European heritage, they differ significantly in historical-societal backgrounds and languages spoken: an Estonian-speaking majority sample of Estonian residents, a Russian-speaking minority sample of Estonian residents, and a mixed-background sample of mostly Western Europeans who were tested in English.

Personality and life satisfaction

This study

In this largest yet multi-trait, multi-rater, multi-sample study, we estimated LS's and DSs' true associations with each other and a range of broad and narrow personality traits, controlling for single-method biases, occasion-specific effects, and random error. We also avoided direct construct/measurement overlaps between personality traits themselves and with LS/DSs by ensuring that LS's indicators had higher convergent validity among themselves than discriminant validity with personality trait indicators. We additionally tested LS' true rank-order stability across several years and compared its cross-sectional true associations to longitudinal ones. Specifically, 20,886 Estonian adults provided self-reports and were rated by an informant using a diverse pool of 198 items. These items were carefully selected to cover LS and encompass a broader-than-usual range of personality traits, including the Big Five. Participants also rated their satisfaction with eight life domains: job, career choice, financial situation, residence, country, relationships, health, and appearance. In a sub-sample of 514 participants, personality traits and LS had also been rated by participants and their informants approximately ten years earlier. We tested the findings' robustness among Russian-speakers living in Estonia (N = 768) and English-speaking participants from various mostly European countries (N = 600). All this allowed us to estimate satisfaction's overall extent of reflecting personality traits and the associations' details with a level of precision and robustness rarely, if ever, attained yet.

We may already know that some personality traits' correlations with LS are greater than zero, at least in usual circumstances. However, a more important but not yet compellingly answered question is: *how much* greater? For example, it would be a two-fold difference if LS could be predicted from personality traits with an accuracy of .80 to .90 as opposed to an accuracy of .50 to .60,[2] and we should care about this as natural scientists care whether the speed of light is $1.5*10^8$ or $3*10^8$ m/s or whether the Earth's atmosphere contains about 21% or 11% of oxygen. Thus, while our empirical work is descriptive and predictive (Mõttus et al., 2020), the findings significantly contribute to our theoretical understanding of satisfaction's relatively stable psychological basis.

Methods

Transparency and Openness

Our sample sizes were determined by practical constraints rather than power calculations, but provide high power for any non-trivial effect sizes. We report all data exclusion criteria and variable manipulations. We make our data analytic (R) scripts publicly available, as well as data from one (English-speaking) sample (https://osf.io/yw7x3/?view_only=065344609f14482c9e2595fae9a51abd). Other data cannot be made publicly available due to being part of a large and ongoing biobank study, but researchers can apply for access (https://genomics.ut.ee/en/content/estonian-biobank). Data used in the Supplementary Analyses are also publicly available. All statistical analyses were carried out with R language, version 4.1.2 (R Core Team, 2021). The analyses were not preregistered.

Participants

The Estonian- and Russian-speakers were members ("gene donors") of the Estonian Biobank, a population sample of approximately 200,000 adults encompassing about 20% of Estonian adult residents or past residents currently living abroad (https://genomics.ut.ee/en/content/estonian-biobank). Data used for this study was collected through an online survey between November 2021 and April 2022 and participants could choose to participate in either Estonian or Russian, most likely depending on their native language. Because Estonia has a substantial Russian-speaking minority with a somewhat distinct cultural and historical background, we treated the Estonian- and Russian-speakers as separate samples. For example, although most Russian-speakers were likely born or had been living in Estonia for many years and were well integrated with the Estonian society, many Russian-speakers are geographically concentrated, follow different (often Russian) media and have distinct identities (Vihalemm et al., 2019); this is also a likely reason that Russian-speakers are underrepresented among the gene donors. Email invitations were sent to 182,405 gene donors, with up to two follow-up invitations as necessary. Participants who completed the survey were offered feedback on their Big Five personality trait scores. To encourage participation, the study was also advertised on national radio, television, newspapers and magazines, and on social media. Participants were optionally asked to provide an email of another person (informant) who could complete the third-person form of the personality items about them. After reading information about the study, both participants and their informants electronically signed a consent form. In total, N = 73,266 + 3,719 (Estonian- + Russian-speaking) participants completed the survey.

[2] Correlations have a non-linear scale. To make them comparable, they have to be first z-transformed.

Personality and life satisfaction

After removing participants who either did not invite an informant or whose informant did not submit their responses, and participants with more than ten missing responses in either self- or informant-report surveys, we were left with 20,886 participants who completed the survey in Estonian (sex assigned at birth: 14,228 women, 6,658 men; age: range from 18 to 93; M = 44.0, Mdn = 45.2, SD = 13.7) and 768 participants who completed the survey in Russian (sex assigned at birth: 533 women, 235 men; age: range from 18 to 88; M = 43.4, Mdn = 43.0, SD = 13.0). The included and excluded participants somewhat differed in their average personality traits and LS; for example, the 52,380 excluded Estonian-speaking participants were less open and life-satisfied than their 20,886 included peers (respectively, $d = -0.25$ and -0.14, $p < .001$), while differences in their other traits were negligible ($0.01 \leq |d| \leq 0.06$). The informants were usually partners or spouses, children/grand-children, friends, or parents/grandparents (56%/54%, 14%/15%, 14%/16% and 7%/8% of Estonian/Russian-speaking informants, respectively). Between 2008 and 2017, 514 of the Estonian-speakers (321 females; age: range from 18 to 79 years; M = 38.7, Mdn = 38.0, SD = 13.3) had completed another personality test and answered to an LS question (79% had participated by 2012, mostly from 2009 to 2010, and further 15% participated in 2013, so the re-testing interval was usually about or more than 10 years). These data collections were approved by the Estonian Committee on Bioethics and Human Research.

Between March and June 2020, 300 dyads completed the personality and LS items about themselves and the other dyad member in English (436 females, 7 preferred not to say; age: range 12 to 82 years; M = 28.5, Mdn = 23.0, SD = 12.9). People were recruited online so that the person who started the study was asked to identify another dyad member and provide their email, who was then invited to similarly participate. Although the study was intended for adults, six participants invited adolescent dyad members. Participants were offered feedback on their Big Five traits and most salient personality nuances, and how well they and their informant agreed regarding each other's traits. Some participants were also compensated monetarily. Most participants were British residents, but many resided in other Western countries or India. Initially, these data were collected for student projects exploring items' cross-rater correlations, approved by the University of XXXX institutional review board.

Measures

The 100 Nuances of Personality (100-NP) is a 198-item pool designed to cover personality traits comprehensively and with minimal redundancy. It captures trait content associated with most facets and domains assessed in standard Big Few measures as well as some individual differences measures beyond these (e.g., competition, envy, humor, sexuality, spirituality, and the "Dark Triad" traits). The items were iteratively selected from larger item pools such as the International Personality Item Pool (Goldberg, 1999) and Synthetic Aperture Personality Assessment (Condon & Revelle, 2016) for their content, and retained if they demonstrated 1) acceptable levels of empirical properties (e.g., test-retest reliability, variance, and cross-rater agreement) and 2) not excessive redundancy with other items, except a small amount of highly correlated items to generate a test of acquiescent responding or provide a pair for items of apparently less reliably assessable traits, such as impulsiveness. A full description of the 100-NP's development can be found in Henry and Mõttus (2022). The 100-NP was completed by people themselves and their informants. We selected four of the items to capture LS (Table 1) and three to capture DSs (about satisfaction with relationships, health and appearance: "Am satisfied with my relationships", "Consider myself healthy for my age", "Consider myself good-looking"). Other DSs items about satisfaction with job, choice of career, financial situation, residence, and country) were only completed by participants themselves due to the limited number of items that could be administered to informants. In English-speakers' data, six personality items and DS items were not administered. A full item list is in the Supplementary Material at Open Science Framework (OSF; https://osf.io/yw7x3/?view_only=065344609f14482c9e2595fae9a51abd). Items were responded to using a six-point Likert scale from "Completely inaccurate" to "Completely accurate". Missing responses were replaced with the median.

For the earlier (between 2008 and 2017) data collection, personality traits were measured with the Estonian version of the NEO Personality Inventory-3 (NEO-PI-3; McCrae et al., 2005), and LS with a single item: "All things considered, how happy are you with your life generally?", rated on a 10-point scale from "Not at all" to "Completely". The NEO-PI-3 scales were scored as sum-scores of their items, as per test manual.

Analyses to estimate true associations

True correlations (r_{true}s). To estimate variables', say x and y, r_{true}s, we correlated self-reported x with informant-reported y and vice versa, and calculated the geometric mean of these cross-rater, cross-variable correlations. We then correlated self-reported x with informant-reported x and the same for y, and calculated the geometric mean of these cross-rater, same-variable correlations. We treated the ratio of the former geometric mean to the latter as the r_{true} between x and y, free of single-method biases that are either specific to either x or y or shared among them, rating occasion-specific effects (e.g., mood) and random error because these four variance components would similarly affect both cross-rater, cross-item and cross-rater, same-item correlations and therefore cancel out in their ratio. The approach is based on the simplifying assumption that both variables' valid (true) variance is *at least partly* shared between raters — hence, partly independent of assessment method — whatever its fraction to total variance, and that rating biases and occasion-specific effects are not shared between raters. The degrees of rater- and occasion-specific effects and random error may differ across variables and raters, but as long as all four correlations are used, they are equally represented in both the numerator and denominator of the r_{true} calculation and hence cancel out in equal proportions. Among other things, this means that raters' asymmetrical information about the traits does not influence the model's estimates. An extended, algebraic formalization of the variance decomposition model underlying the r_{true} calculation is in Supplementary Material. The idea is similar to how Wood and colleagues (2022) used test-retest data to estimate items' semantic similarity, except that we used informant-ratings instead of retest scores, which allowed us to control for single-method effects.

True predictive accuracy. To estimate personality traits' true combined overlap with LS (unbiased "multiple R") in the Estonian-speaking sample, we created elastic net models tailored to maximize the traits' out-of-sample predictive accuracy for aggregate LS in one sample partition (67%) and calculated the correlation between LS and its values predicted from personality traits using this model in another sample partition (33%). The elastic net models with .50 alpha parameter were trained to minimize prediction error in 10-fold cross-validation within training samples. For true predictive accuracy, net of single method biases and random error, we "cross-predicted" self-reported LS from informant-reported personality traits and vice versa in 10 random training-validation sample splits and averaged the predictive accuracies within each direction, and divided the geometric means (across directions) of these cross-prediction accuracies by the geometric means of self-informant correlations for a) observed LS scores and b) their predicted-from-personality values. For replications in Russian- and English-speaking samples, we used models trained in the Estonian data, hence training and validating models in different languages.

Domain satisfactions. Because most DSs were assessed with only self-reports, we approximated their r_{true}s with LS by correlating self-reported DS items with the informant-reported LS aggregate and then dividing these correlations by the geometric mean of a) average cross-rater correlation of three DS items for which cross-rater data was available (as a proxy for all DS items' cross-rater correlation; .44) and b) the LS aggregates' cross-rater correlation. Likewise, we approximated the DS aggregate's and LS aggregate's r_{true} by calculating self-reported DS aggregate's correlation with informant-reported LS aggregate and dividing this by the geometric mean of the LS aggregate's cross-rater correlation and the cross-rater correlation of the principal components of the three DS items for which cross-rater data were available (Table 2). Because not all cross-rater correlations were used in these calculations and ratings of different variables and/or by different raters could contain somewhat different degrees of biases and errors that were then not equally represented in the numerators and denominators of the r_{true}s approximations, the r_{true} estimates pertaining to DSs could be to some degree biased, unlike the LS-personality trait r_{true}s based on four correlations each.

Standard errors. Because most r_{true}s were based on four correlations each, the usual standard error formulas did not apply to them. To find a formula to estimate the standard errors, we relied on an iterative process of inductive reasoning and tinkering, comparing the results against the ground truth in simulated data until the formula results closely approximated the simulation results (see Supplementary Material). The main sample of Estonian speakers was so large that the standard errors were bound to be small, but they were larger for estimates in smaller Russian- and English-based samples.

Variable selection and aggregation

We describe these analyses based on the Estonian-based data, but Tables 1 and 2 also contain correlations for Russian- and English-based analyses. To proof-of-principle test whether the cross-informant design approximates variables' r_{true}s, we included some pairs of highly similar items (e.g., "Keep my promises" vs "Break my promises"). In the main, Estonian-based data, these items' $|r_{true}s|$ reached .97, providing support for the research design (Supplementary Table S1). We used r_{true}s for variable selection and aggregation, unless said otherwise, and relied on the general idea that items measuring the same construct should have stronger correlations than items measuring different constructs (Campbell & Fiske, 1959).

Personality and life satisfaction

Life satisfaction (LS). The $|r_{true}s|$ among three intended LS items ("Am happy with my life", "Feel that my life lacks direction", "Am pessimistic about the future") varied between .74 and .77 (Table 1). For comparison, their single-method absolute correlations varied between .48 and .52 in self- and informant-reports. The fourth item, "Life has been kind to me", had lower $|r_{true}s|$ with other items (.32 to .56), so we removed it from further analyses to retain LS high construct validity (its absolute single method correlations varied from .16 to .41). The cross-rater correlations of the three retained LS items were .42, .37, and .36. Separately in self- and informant-ratings, we used scores of the first principal components of the three LS items as aggregate LS scores (respectively explaining 66% and 68% of the items' variance; all loadings > |.80|; cross-rater correlation .48). This pattern replicated in English and Russian samples.

In Supplementary Analyses 1 (https://osf.io/yw7x3/?view_only=065344609f14482c9e2595fae9a51abd), we show that latent trait scores based on these three items correlated highly (r = .95, .90 and .80, respectively among Estonian-, Russian-, and English-speakers) with latent trait scores of a more widely used LS assessment, Satisfaction with Life Scale (SWLS; Diener et al., 1985), based on self-reported data collected in Estonian, Russian and English. This supports the validity of our aggregate LS scores.

Table 1. True correlations ($r_{true}s$) among the four items designed to measure LS in Estonian-/Russian-/English-based data.

	Am happy with my life	Feel that my life lacks direction	Have a dark outlook on the future	Life has been kind to me (dropped)
Am happy with my life	.42/.44/.53	.011/.052/.049	.011/.058/.044	.012/.082/.057
Feel that my life lacks direction	-.74/-.85/-.66	.37/.39/.45	.012/.064/.052	.014/.09/.070
Have a dark outlook on the future	-.77/-.77/-.82	.75/.69/.68	.36/.33/.48	.013/.097/.063
Life has been kind to me (dropped)	.56/.76/.64	-.32/-.55/-.33	-.46/-.63/-.50	.34/.19/.36

NOTE: Standard errors are above the diagonal. In the English version, the item "Life has been very kind to me" was worded as "Have been richly blessed in my life". The diagonal contains cross-rater correlations.

Domain satisfactions (DSs). Correlations among the eight self-report items selected to assess specific DSs were lower than those for LS items, varying from .08 to .62 (Mdn = .20, compared to the respective Mdn = .50 for the three LS items), but were mostly within the recommended range of a typical scale's inter-item correlations (Clark & Watson, 1995). So, we used the scores of their first principal component as an aggregate DS score (explaining 34% of items' variance; all loadings > .40). In Supplementary Analyses 1 (https://osf.io/yw7x3/?view_only=065344609f14482c9e2595fae9a51abd), we show that latent trait scores based on these eight DS items correlated highly (r = .96) with latent trait scores from the SWLS, based on separate self-reported data collected among Estonian-speakers. This supports the validity of the DS aggregate as a measure of LS.

Table 2. Correlations among self-reported DSs items in Estonian- and Russian-based data.

	1	2	3	4	5	6	7	8
1. Am happy with my job	-	.005/.027	.006/.032	.007/.034	.007/.035	.007/.034	.007/.035	.007/.035
2. Am happy with my choice of profession	.62/.67	-	.006/.033	.007/.034	.007/.036	.007/.035	.007/.036	.007/.036
3. Am happy with my financial situation	.43/.48	.37/.39	-	.006/.034	.007/.034	.007/.034	.007/.035	.007/.034
4. Am happy with my residence	.29/.37	.31/.34	.42/.36	-	.007/.034	.007/.035	.007/.035	.007/.036
5. Am happy with how things are organized in our country	.17/.19	.15/.10	.27/.33	.21/.30	-	.007/.036	.007/.035	.007/.036
6. Am satisfied with my relationships	.22/.31	.22/.28	.33/.32	.31/.29	.12/.12	.44 / .24	.007/.035	.007/.034
7. Consider myself healthy for my age	.19/.21	.16/.12	.22/.22	.16/.20	.13/.21	.16/.25	.50 / .46	.006/.034
8. Consider myself good-looking	.15/.22	.14/.17	.18/.31	.15/.15	.08/.15	.19/.36	.35/.37	.39 / .33

NOTE: Standard errors are above the diagonal. The diagonal contains cross-rater correlations for the three items with informant-reports available (Estonian / Russian).

Personality items. We dropped personality items that strongly overlapped among themselves (51 items; Supplementary Table S2) and with any LS item (one item: "Tend to feel very hopeless"), using $|r_{true}| \geq .75$ as the cut-off (given the $|r_{true}| \approx .75$ among LS items) and dropping weaker LS-correlate from each pair of highly correlating items (Supplementary Table S2). We also dropped three items ("Worry about my health", "Worry a lot about my looks", "Wear stylish clothing") that could semantically overlap with DSs about health and appearance. We treated the remaining 136 personality items as *possible markers of at least partly distinct personality nuances with discriminant validity for LS and DS*. Their cross-rater correlations ranged from .15 to .64 (Mdn = .30). We do not treat items as nuances *per se*, but as markers for both broad traits like personality domains and narrow traits like nuances. However, we do use item-level correlations to describe the nuanced-ness of LS' personality correlates, where evidence for it exists.

Personality domains. There is no universally agreed organization of nuances (or more concretely, items) into facets and domains (Condon et al., 2020). Therefore, we skipped the facet level and combined the 136 personality items into five domains by performing a principal component analysis on their r_{true}s (Supplementary Table S3). After varimax rotation, we retained 15 highest-loading items from each component to ensure roughly balanced and most relevant content representation for each domain and re-calculated the five components based on the 75 remaining items; these accounted for 60% of items' variance. After varimax rotation, these components' loadings clearly resembled the typical Big Five themes (Supplementary Table S4), and we used them to calculate domain scores in self- and informant-ratings, multiplying standardized item scores by the items' inverted correlation matrix and the principal component loadings. Had we chosen more items per component, the loading pattern would have started differing from what we considered more typical Big Five content and some loadings would have dipped below .40. This procedure ensured that domain scores were calculated similarly in self- and informant-reports. The domains' cross-rater correlations were .56 (emotional stability), .58 (extraversion), .57 (openness), .46 (agreeableness), and .50 (conscientiousness), which are comparable or higher than usual (Connelly & Ones, 2010). In single-method designs using common Big Five instruments, the domain scores can correlate as highly as .40s and .50s (van der Linden et al., 2010). Intentionally and desirably, our domain scores' correlations were lower, varying from 0 to .12 (Mdn = .04) in self-reports and from 0 to .21 (Mdn = .02) in informant-reports; $|r_{true}s|$ varied from .01 to .31 (Mdn = .07). This relative independence of domain scores ensured that the domain-LS correlations would be less inflated by shared variance than usual. Our longitudinal data allowed us to estimate the domains' empirical similarity to those of a widely-used Big Five assessment, NEO-PI-3 (Table 7).

Results

The following three sections describe results from the main, Estonian-speaking sample.

Correlations with personality items

We started with LS' correlations with the 136 items retained as possible markers of personality nuances with discriminant validity for LS and DS, calculating their r_{true}s with individual LS items and the aggregate of these. The three vectors containing the LS items' z-transformed r_{true}s with the personality items were highly similar ($r \geq .94$), supporting the LS aggregate's construct validity. For the LS aggregate, r_{true}s varied from -.69 to .60 ($|r_{true}|_{median}$ = .22; $|r_{true}|_{min}$ = .01). Figure 1 shows 70 items correlating with the aggregate LS at $|r_{true}| \geq .20$ ($|r_{true}|_{median}$ = .33), while all 136 r_{true}s are in Supplementary Table S5, alongside their underlying cross-variable, cross-rater correlations and same-variable, cross-rater correlations.[3,4] For comparison, single-method correlations for these 70 items with the LS aggregate (Table S5) varied from -.45 to .46 in self-reports ($|r|_{median}$ = .17) and from -.50 to .44 in informant-reports ($|r|_{median}$ = .20), so $|r_{true}s|$ tended to be stronger despite not being influenced by single-method biases.

For interpretation ease, we also highlight the LS aggregates' strongest *relatively unique* correlates, showing the 19 items not having $|r_{true}| > .50$ with any other personality item with a larger and darker font in Figure 1. Because these 19 items were comparatively less inter-correlated, they necessarily covered a broader range of traits than our Big Five domains (12 were not included among the 75 Big Five items). Low LS tracked with feeling misunderstood (-.69), lack of excitement (-.61), indecisiveness (-.51), envy (-.49), boredom (-.45) and feeling used (-.41), whereas high LS tracked confidence in ones' abilities (.44) and believing that effort is rewarded (.40). Less strongly ($.20 < |r_{true}| < .40$), high LS tended to be uniquely characterized by taking risks, finding it easy to apologize, feeling special commitment to one's family, being loyal, respecting authority, liking to visit new places, and working on self-improvement, whereas low LS tended to go with making enemies, telling lies, forgetting things, and crying easily.

[3] The .20 cutoff was chosen for presentation ease and because correlations of .20 and higher are said to heuristically represent "a medium effect that is of some explanatory and practical use even in the short run" (Funder & Ozer, 2019, p. 156). Even if significant, small correlations among psychological variables may sometimes reflect the pervasive "crud factor", thus not being meaningfully interpretable.

[4] The two directions of cross-variable, cross-rater correlations (self-rated LS's correlations with informant-rated personality traits and the other way around; Table S5) were highly similar, with the z-transformed correlation vectors from the two directions correlating .98/.99 for the 136/70 items. This suggests that self- and informant-rated items contain broadly similar degrees of information about the variables involved, including LS. This mitigates the possibility that LS's informant-reports are (more) biased (than self-reports).

Personality and life satisfaction

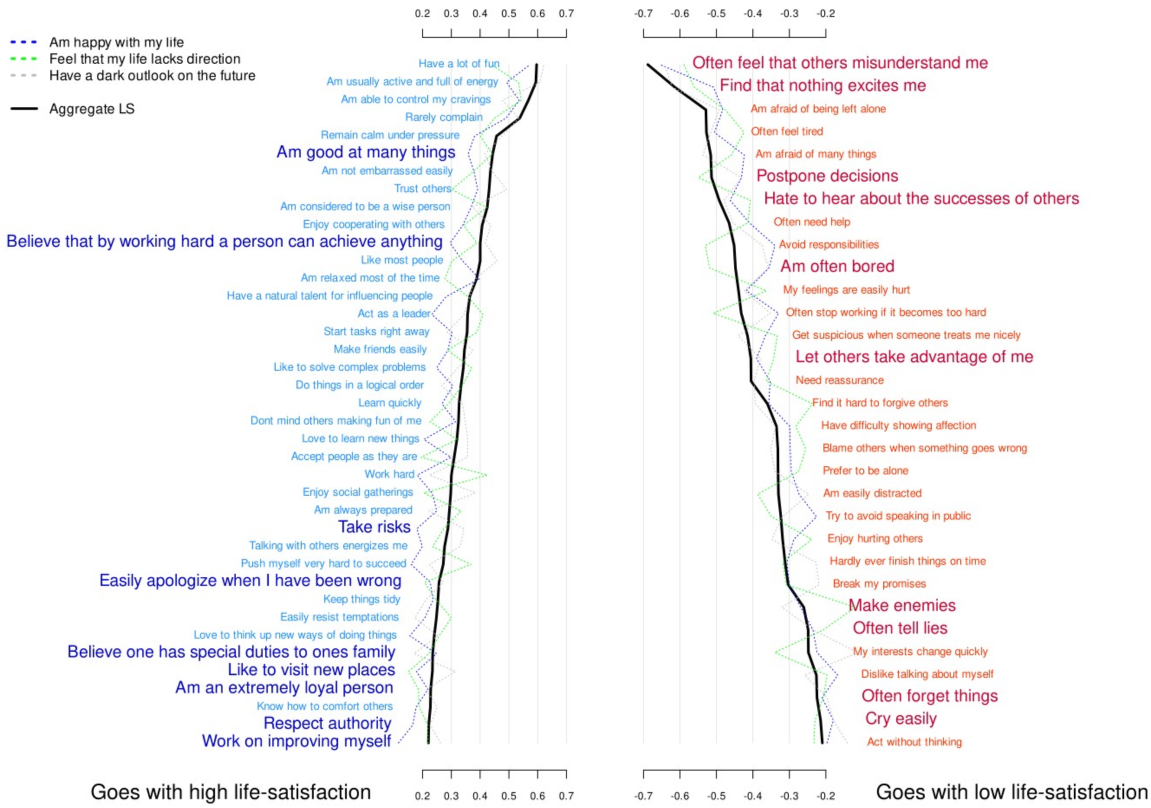

Figure 1. *Personality items' true correlations (r_{true}s) with the aggregate LS and individual LS items in the Estonian-speaking sample (N = 20,886). Items with greater and darker font size didn't have true correlations higher than .50 with any other personality items, thus reflecting relatively distinct personality nuances. (Small) standard errors are in Table S5.*

Correlations with personality domains

Next, we correlated LS with Big Five domains to represent the LS' personality correlates more parsimoniously and comparably with typical findings in the existing literature. The domains' r_{true}s with the LS aggregate (Table 3) ranged between .30 and .47 for conscientiousness, extraversion, and emotional stability, but remained below .05 for openness and agreeableness.[5] For comparison, single-method correlations of the five domains with the LS aggregate were .34, .36, .10, .11, and .28 in self-reports and .41, .34, .07, .09, and .27 in informant-reports, respectively for emotional stability, extraversion, openness, agreeableness, and conscientiousness (Supplementary Table S5). So, domains' true correlations with LS were comparable or higher than single-method correlations for emotional stability, extraversion and conscientiousness, despite not being influenced by variables' shared single-method biases. These correlations are comparable or even higher than those reported in other single-method studies (Anglim et al., 2020; Supplementary Table S5), despite our domain scores being less inter-correlated than those in studies using common Big Five scales. For openness and agreeableness, true correlations were lower than single-method correlations in this and other data (Supplementary Tables S5 and S6), possibly because of not being inflated by single-method biases and/or overlaps with other personality domains.

However, items primarily loaded on by the same domains often varied considerably in their correlations with the LS aggregate, such as "Enjoy hurting others" (r_{true} = -.32) and "Believe that I am always right" (r_{true} = 0.03) that were both negatively loaded on by the agreeableness domain (Supplementary Table S4). This partly explains why domains $|r_{true}s|$ were lower than those of several items within and beyond the domains. So, although domains provide a parsimonious representation of LS's associations with personality traits, they can also partly misrepresent these associations.

[5] The two directions of calculating the cross-variable, cross-rater correlations (self-rated LS's correlations with informant-rated personality domains and the other way around) yielded broadly similar, although not identical results. Informant-related LS's correlations with self-rated personality domains were .28, .20, .00, .03, and .14, while self-rated LS's correlations with informant-rated domains were .21, .25, .02, -.01, and .15, respectively for emotional stability, extraversion, openness, agreeableness, and conscientiousness.

Table 3. LS's true correlations with personality domains in Estonian-/Russian-/English-based data.

	Estonian-based data		Russian-based data		English-based data		Meta-analytically combined (Russian and English)		Past results (for reference)
	r_{true}	SE	r_{true}	SE	r_{true}	SE	r_{true}	SE	r (single method)
Emotional stability	.47	.008	.36	.053	.32	.053	.34	.038	.39
Extraversion	.43	.008	.50	.052	.45	.048	.47	.035	.32
Openness	.02	.011	-.05	.066	.00	.064	-.03	.046	.08
Agreeableness	.04	.012	.02	.072	.10	.067	.06	.049	.20
Conscientiousness	.30	.010	.47	.052	.26	.060	.38	.039	.27

NOTE: SE = standard error. For single-method and cross-method correlations, see Supplementary Table S5. Past results (for reference) = meta-analytic estimates from Anglim et al. (2020).

Life satisfaction's overall predictability

Next, we evaluated the overall degree to which life satisfaction aligns with individuals' personality traits. Predicting LS from the full set of 136 items and five domains, the respective true predictive accuracies were .91 and .79 (Table 4). Even though hypothetical estimates disattenuated for measurement issues, these represent unusually high correlations in psychological research (Funder & Ozer, 2019).[6] To see whether these high estimates were driven by numerous predictors – either by many items individually or by many items contributing towards domain scores – we also explored true predictive accuracies of smaller item sets such as the 70 items shown in Figure 1, the 19 relatively unique items among them (not having $|r_{true}s| > .50$ with any other items) and the three most strongly LS-related items among these 19 ("Often feel that others misunderstand me"; "Find that nothing excites me", and "Postpone decision"). These smaller item subsets provided true predictive accuracies between .81 and .88 (Table 4), showing that LS was highly predictable from even a few personality traits. For comparison, the predicted-observed LS correlations in single-method data were .75 and .64 in self-reports and .76 and .64, in informant-reports, respectively for models based on 136 items and five domains. For domains, this corresponds to $R^2 = .41$, which is higher than the $R^2 \approx .30$ usually found in single-method studies (Anglim et al., 2020; Busseri & Erb, in press).

Table 4. LS's true out-of-sample predictability from personality domains and items in Estonian-/Russian-/English-based data from models trained in Estonian data.

	Estonian-based data		Russian-based data		English-based data	
	r_{true}	SE	r_{true}	SE	r_{true}	SE
Five domains	.79	.008	.74	.046	.64	.049
136/134 items*	.91	.007	.90	.040	.84	.035
70/69 items*	.88	.008	.86	.050	.82	.037
19/18 items*	.86	.010	.88	.050	.82	.042
Three items	.81	.010	.82	.057	.82	.046

NOTE: r_{true} = true correlation between predicted and observed life satisfaction. SE = standard error. * Smaller item numbers apply to English-base data.

Domain-specific life satisfactions

Next, we cross-validated the LS and its personality correlations against individual DSs' and their aggregate, representing an alternative way of conceptualizing and assessing general satisfaction. The r_{true} between aggregate DS and LS was .87 (Table 5), suggesting that LS's assessments closely tracked how satisfied people were with several specific life domains combined. For reference, self-reported LS and DS aggregate correlated .67. Likewise, LS was linked with all DS items,

[6] The two directions of predicting LS from personality traits yielded similar results. For example, LS correlated with its values predicted from the 136 self-rated items .47, while self-rated LS's correlation with its values predicted from 136 informant-rated items was .45. So, there was no evidence that either informant- or self-rated LS would be more or less informative in relation to other traits.

especially those referring to satisfaction with relationships and financial situation (r_{true} > .65), career and residence (r_{true} ≈ .55 to .60), health and appearance (r_{true} ≈ .40 to .45); satisfaction with how things are organized in the country was less correlated with LS (r_{true} = .31). To show how the r_{true}s were calculated, we give all relevant correlations in Table 6. These findings provide strong evidence for the construct validity of both ways of assessing satisfaction, LS and the (aggregate) DS's.

Table 5. *DSs' correlations with LS.*

	Estonian-based data					Russian-based data				
	Cross-rater, cross-variable correlation	Cross-rater, same-variable correlation	LS' cross-rater correlation	r_{true}	r_{true} standard error	Cross-rater, cross-variable correlation	Cross-rater, same-variable correlation	LS' cross-rater correlation	r_{true}	r_{true} standard error
DS aggregate	.42	.49	.48	.87	.009	.43	.43	.50	.92	.052
Am happy with my financial situation	.31	.44	.48	.68	.009	.34	.35	.50	.81	.059
Am satisfied with my relationships	.30	.44	.48	.66	.009	.29	.35	.50	.70	.058
Am happy with my choice of profession	.27	.44	.48	.59	.009	.28	.35	.50	.68	.058
Am happy with my job	.27	.44	.48	.58	.009	.30	.35	.50	.73	.058
Am happy with my residence	.25	.44	.48	.55	.010	.25	.35	.50	.59	.058
Consider myself healthy for my age	.21	.44	.48	.45	.010	.21	.35	.50	.51	.058
Consider myself good-looking	.19	.44	.48	.42	.010	.25	.35	.50	.61	.058
Am happy with how things are organized in our country	.14	.44	.48	.31	.011	.12	.35	.50	.29	.063

NOTE: Cross-rater, cross-variable correlations were calculated between self-rated DSs and informant-reported LS. Cross-rater, same variable correlations are either averages of the cross-rater correlations for three DSs for which cross-rater data was available (this is why they are identical for the eight DSs) or the cross-rater correlations of these three items' principal components (for the DS aggregate). r_{tue} = (cross-rater, cross-variable correlation) / (geometric mean of cross-rater, same variable correlation and LS's cross-rater correlation). Because cross-rater, cross-variable correlations were only available in one direction and cross-rater, same-variable correlations were partly estimated based on other variables, the r_{tue} estimates may be somewhat biased.

Next, we assessed the extent to which different DSs shared personality correlates among themselves and with the LS, to cross-validate the findings and further assess general satisfaction's construct validity. Specifically, we calculated the 136/70 informant-reported personality items' correlations with individual self-reported DS items and the DS and LS aggregates. We did not estimate r_{true}s here because the purpose was comparing correlation patterns, net of single-method effects, not their absolute values. Most DS items' cross-rater correlations with personality items tended to be similar: vector correlations after z-transformation ranged from .36 to .97 for the 136 personality items retained for analysis (Mdn = .81; only the cross-rater correlation profiles for being satisfied with country and appearance had vector correlations below .58) and from .70 to .98 for the 70 more LS-correlated items (Mdn = .89; again, only the cross-rater correlations profiles for being satisfied with country and appearance had vector correlations below .79). Moreover, the personality items' cross-rater correlations with the DS aggregate tended to be nearly identical to their cross-rater correlations with the LS aggregate, with vector correlations (after z-transformation) of .98/.99 for 136/70 personality items (Figure 2). This suggests that different satisfaction kinds – various DSs, their aggregate and the LS – were similar in their broader psychological backgrounds assessed with an expansive pool of personality items by independent raters. This provides evidence for the robustness of the findings regarding different ways of operationalizing and assessing life satisfaction, and supports general satisfaction's construct validity.

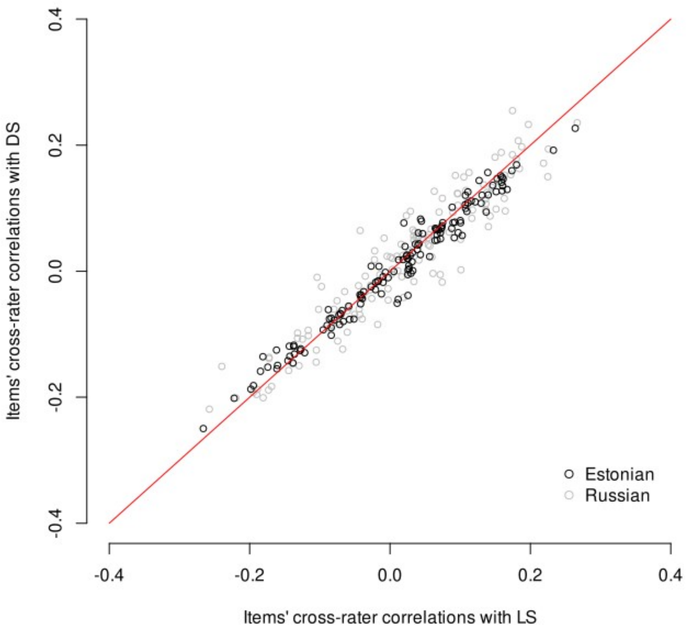

Figure 2. *Informant-rated personality items' correlations with self-rated DS and LS aggregates among Estonian- and Russian-speakers. The red line indicates a relation with identical values on both variables.*

To further cross-validate our findings, we then compared the extents to which different DSs and their aggregate could be predicted from personality traits to LS's predictability from these same traits. Using elastic net, the 136/70 informant-report personality items allowed predicting the DS aggregate with the accuracy of r_{true} = .42/.38, whereas the accuracy was .32 for five informant-reported domains; these estimates were only somewhat lower than the similarly estimated predictive accuracies for LS: .45/.42 and .37 (Table 6). In fact, the model trained to maximize the predictive accuracy for LS predicted the DS aggregate almost as well and also predicted individual DSs (Table 6). The predictive accuracies for the eight specific DSs from the 136/70 informant-reported personality items/domains varied from .27/.23./16 to .39/.31/.24 (Mdn = .34/.28/.20). Hence, there was little evidence that a combination of DSs would be substantially less linked with personality traits than LS, and the eight self-rated DSs' predictabilities from informant-rated personality traits mostly varied in a relatively narrow range. So, DSs overall extents and details of reflecting a broad range of personality traits were fairly similar among themselves and with LS, providing further evidence for the robustness of the findings across different ways of assessing satisfaction.

Table 6. Predictive accuracies for self-rated DSs and LS from informant-reported personality variables.

	Estonian data				Russian data (from models trained in Estonian data)			
	136 items	70 items	Five domains	136 items (LS model)	136 items	70 items	Five domains	136 items (LS model)
DS aggregate	.42	.38	.32	.39	.38	.38	.31	.41
Am happy with my job	.29	.27	.21	.25	.33	.28	.19	.29
Am happy with my choice of profession	.32	.30	.21	.25	.34	.30	.15	.26
Am happy with my financial situation	.34	.29	.23	.26	.29	.26	.22	.27
Am happy with my residence	.27	.24	.19	.23	.24	.22	.16	.23
Am happy with how things are organized in our country	.34	.23	.16	.15	.26	.17	.16	.14
Am satisfied with my relationships	.34	.31	.20	.24	.20	.20	.26	.31
Consider myself healthy for my age	.34	.31	.20	.23	.33	.31	.21	.24
Consider myself good-looking	.39	.31	.24	.23	.38	.27	.23	.26
LS aggregate (as reference)	*.45*	*.42*	*.37*		*.49*	*.42*	*.39*	

NOTE: LS model = DSs were predicted from the model trained to predict LS.

Personality and life satisfaction

Replications in Russian and English

The patterns of findings replicated well in smaller Russian- and English-speaking samples, allowing for some sampling variance in the comparatively smaller samples. At the construct validity level, the three LS items had $|r_{true}s|$ from .69 to .85 in Russian and from .66 to .82 in English (Table 1), further supporting the robustness of LS's construct validity. Likewise, the correlations among the self-reported DS items were similar in Estonian and Russian-based data (Table 2). Also, the vectors of LS aggregate's z-transformed correlations with the 136/70 items personality items were similar in the three samples (r = .89 to .94; Supplementary Table S5; for English-based data, only 134 and 69 items were used). We also meta-analytically combined the $|r_{true}s|$ in Russian- and English-based data for 69 items (one of the 70 items was not administered in English), using inverted standard errors as individual estimates' weights for the meta-analytic estimates. These meta-analytic estimates correlated .97 with the Estonian-based $r_{true}s$, even though the 99% confidence intervals of these $r_{true}s$ did not span the respective Estonian-based estimates for 12 items (Figure 3). For example, among those tested in Estonian, often complaining, not being able to control cravings and the fear of being left alone were more strongly linked with low LS than in the combined Russian- and English-based findings, while the reverse applied to lacking excitement, being suspicious and avoiding responsibilities, among others.

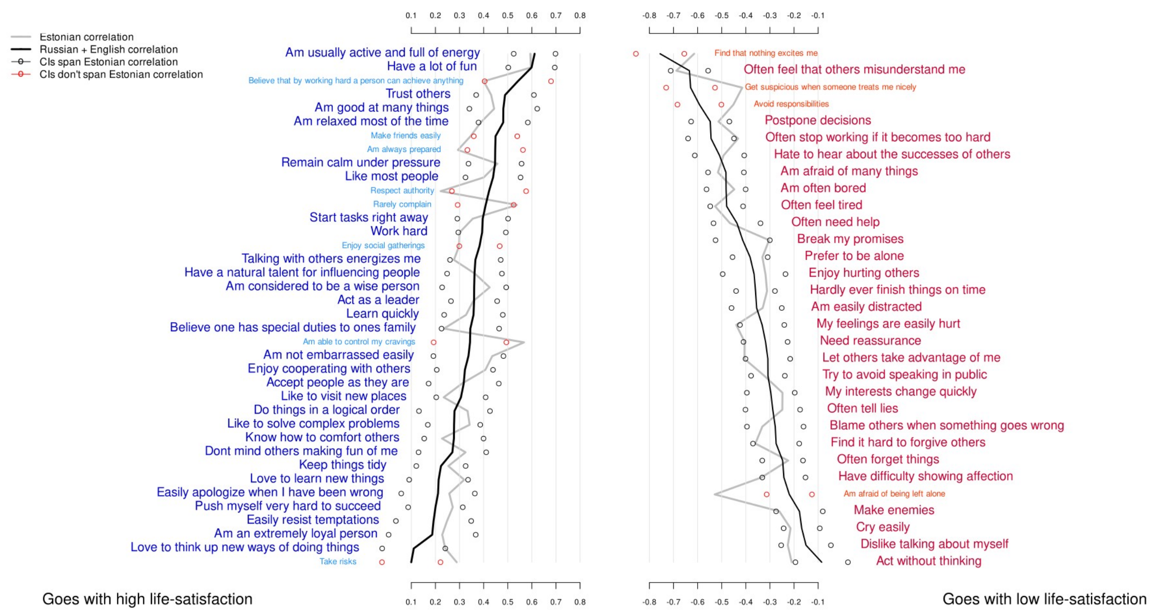

Figure 3. *Personality items' meta-analytic correlations with the aggregate LS in Russian- and English-based data (gray), and Estonian data (black). Meta-analytic 99% confidence intervals for items with greater and darker font span Estonian correlation. Standard errors are in Table S5.*

Loadings of the five principal components on the 75/73 personality items in the Russian-/English-based data were similar to those in Estonian-based data (two of the 75 items were not administered in English), with the total factor loading congruence estimates after Procrustes rotation (McCrae et al., 1996) to the Estonian loading matrix exceeding .95 (Supplementary Table S4). For direct comparability of the domains' correlations with LS, we therefore used loadings from the Estonian data to calculate the domain scores in Russian and English data and used these scores to estimate the domains' $r_{true}s$ with LS. As in Estonian-based data, openness and agreeableness were less correlated with LS than other domains, with $|r_{true}s|$ varying from .02 to .11. Emotional stability, extraversion and conscientiousness had $r_{true}s$ with the LS aggregate from .36 to .50 among Russian-speakers and from .26 to .45 among English-speakers, with their meta-analytic correlations being .34 (emotional stability), .47 (extraversion), and .38 (conscientiousness; Table 3). So, emotional stability was more strongly but extraversion and conscientiousness less strongly linked with LS among people tested in Estonian than among those tested in Russian or English, similarly to Realo (2006) and Kööts-Ausmees and colleagues (2013).

Personality and life satisfaction

We used elastic net models trained in the Estonian-based data to predict self- and informant-reported LS from the informant- and self-reported items and domains in Russian- and English-based data. The cross-sample true predictive accuracies were .90 and .84 for the 136/134 items, respectively, and .74 and .64 for the domains; other item combinations provided similarly good cross-sample predictions (Table 4). So, LS's true degrees of reflecting other personality traits, net of biases and random error, were strikingly robust across samples and languages even if some individual estimates varied in strength (Figure 2, Table 3).

The findings pertaining to DSs also replicated well among people tested in Russian; data for most DSs were not available in English, and hence we did not replicate DS-related analyses in those data. The individual DSs were correlated with LS similarly to the Estonian data (Table 5) and the LS-DS aggregates' r_{true} was .92 (Table 5). The median correlations among z-transformed cross-rater correlation vectors of the 136/70 informant-rated personality items with the eight DS items were .72 and .83, respectively, and the items' correlations with the LS and DS aggregates were also highly similar (r = .94 and .97, respectively; Figure 2). Using models trained in the Estonian data, the 136/70 informant-report personality items allowed for predicting the DS aggregate with the accuracy of .38/.38, compared to similarly estimated true prediction accuracy of .49/.42 for LS; for the five domains, the respective accuracies were .31 and .39; and for individual DSs, the predictive models trained in Estonian data were almost as predictive in Russian data (Table 6).

In sum, thus, the findings were remarkably robust across samples.

Longitudinal analyses

We used the longitudinal assessments in the Estonian-speaking sample to assess the stability of LS and its personality correlates over time. Should LS' cross-sectional and longitudinal r_{true}s with personality traits be similar and approach LS true stability, this would suggest that personality traits' systematic involvement in LS endures over time, irrespective of time-varying influences on either.

The two Big Five domains' assessments, separated by approximately ten years, had r_{true}s between .73 and .82, and the single-item LS had the r_{true} of .70 with LS 10 years later (Table 7).[7] The later Big Five scores correlated with the earlier and later LS similarly for all domains but conscientiousness, for which the cross-sectional correlation was r_{true} = .30, but longitudinal r_{true} = .12; the earlier Big Five scores had similar cross-sectional and longitudinal correlations with LS. At the item-level, later personality traits' correlations with earlier and later LS were highly similar, with the z-transformed correlation-vectors' correlation of .91 (Figure 4). At the facet-level, the earlier personality traits' correlations with LS were partly driven by four facets: N3 Depression, N6 Vulnerability, E6 Positive Emotions and C1 Competence (Supplementary Table S6). We also predicted the earlier LS from models trained to predict the later LS from the later-assessed 136 personality items and Big Five domains, omitting participants with earlier data from model training; the respective true predictive accuracies were .75 (SE = .051) and .61 (SE = .058). So, individual differences in both personality traits and LS as well as their correlations tended to endure over time, and it did not matter much whether r_{true}s were calculated, and LS predicted, cross-sectionally and longitudinally.

Table 7. Longitudinal correlations.

	Later LS		Earlier LS		Correlations with earlier domains	
	r_{true}	SE	r_{true}	SE	r_{true}	SE
Later domains (100-NP)						
Emotional stability	.47	.046	.40	.049	.73	.038
Extraversion	.43	.045	.38	.049	.74	.033
Openness	.02	.069	.07	.067	.82	.035
Agreeableness	.04	.074	.06	.075	.82	.061
Conscientiousness	.30	.054	.12	.066	.76	.047
Earlier domains (NEO-PI-3)						
Emotional stability	.53	.045	.59	.047		
Extraversion	.43	.042	.39	.045		
Openness	.13	.058	.08	.064		
Agreeableness	.03	.077	.06	.077		

[7] The earlier single-item LS correlated with the later scores of the most similar single item, "Am happy with my life", r_{true} = .72.

Personality and life satisfaction

Conscientiousness	.29	.053	.30	.055
Later LS			.70	.056

NOTE: LS = general life satisfaction; SE = standard error; earlier = measured from 2008 to 2017; later = measured from 2021 to 2022. For consistency, we reverse-keyed NEO-PI-3's neuroticism as emotional stability.

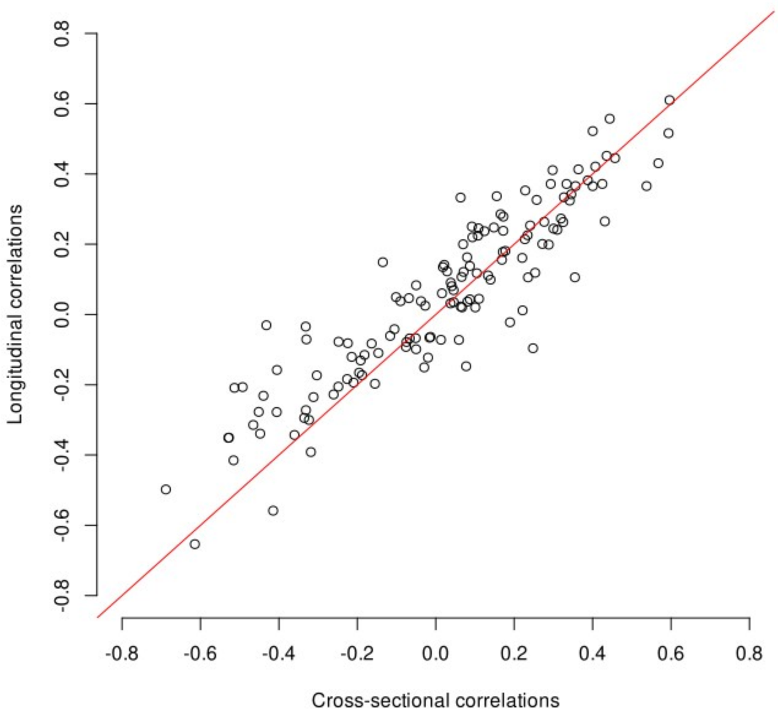

Figure 4. *136 personality items' true correlations with LS assessed cross-sectionally and approximately 10 years earlier. The red line indicates a relation with identical values on both variables.*

Further robustness analyses

To address concerns raised during the articles' review process, we carried out two more robustness checks that are fully described in the Supplementary Analyses 2 and 3 (https://osf.io/yw7x3/?view_only=065344609f14482c9e2595fae9a51abd). First, we reduced LS to a single item, "Am happy with my life", instead of being an aggregate of three items. This somewhat lowered the nuances' and domains' correlations with LS, but the overall patterns of findings remained similar in all three samples. For example, in the Estonian-speaking sample, emotional stability, extraversion, openness, agreeableness and conscientiousness correlated with the single-item LS at r_{true}s = .41, .34, -.01, .02 and .23, respectively, whereas the true predictive accuracies of the 136 items and five domains were .88 and .70, respectively. Likewise, the nuances' correlations with the three-item and single-item LS were similar, with the two z-transformed correlation vectors correlating at r = .99. Combined with the finding that all three LS items had highly similar correlations with personality traits while the items were not perfectly correlated among themselves, these results suggest that the LS aggregate captured a somewhat more broadly defined construct than the single highly face-valid LS item, "Am happy with my life", yet the broader construct and its personality correlates were well aligned with that item.

Second, we addressed the possibility that LS could only be validly assessed using self-reports, and that the LS's cross-rater correlations only arose because informants had observed the targets' personality traits, which were correlated with the targets' otherwise private LS. Specifically, we removed LS's informant-reports from our calculations of true associations and predictive accuracies. Thus, r_{true}s were calculated as $r_{LS(self)x(informant)}$ / $r_{x(self)x(informant)}$, where x represents the personality trait in question. The more closely LS is related to a trait, the more similar its correlations with observer ratings of the trait will be to the cross-rater agreement on that trait, so the ratio estimates their r_{true}. Again, this change in calculating r_{true}s somewhat changed the results, but the overall patterns of findings remained similar in all three samples. For example, in the Estonian-speaking sample, emotional stability, extraversion, openness, agreeableness and conscientiousness now correlated with LS at r_{true}s = .38, .44, .04, -.03 and .30, respectively, whereas the true predictive accuracies of 136 items and five domains were .83 and .75, respectively. Likewise, the items' r_{true}s with LS were fairy similar regardless of whether LS's informant-reports were used or not, with the two z-transformed correlation vectors correlating at r = .98. This is consistent with the finding that correlating self-reported personality traits with informant-reported LS yielded similar results to correlating informant-reported personality traits with self-reported LS (footnotes 4 to 6). There is little evidence that LS is any more private than most other traits.

Discussion

In one of the most comprehensive studies on this topic yet, we analyzed data from and across three samples where a range of personality traits and general life satisfaction (LS) were rated by participants themselves and their informants. This allowed us to estimate LS's and personality traits' true associations free of single-method biases, occasion-specific effects, and random error. Besides avoiding direct construct overlaps at the item level, we cross-validated the findings with a different way of assessing satisfaction: an aggregate of satisfactions with eight specific life domains (DSs). Our findings suggest that in a world without common yet usually unaddressed measurement limitations, it would be possible to fairly accurately predict someone's satisfaction from a handful of personality traits. Specifically, correlations between actual LS and its values predicted from the Big Five personality domains or nuances could reach .90, even when the predictions were based on associations in independent samples tested in different languages. Strikingly, even just three personality items allowed us to predict LS with .80 accuracy. Moreover, LS could be predicted with around .70 accuracy over approximately ten years, similarly to LS's own stability.

We had no reason to *a priori* expect such findings. Because associations observed in typical single-method studies are likely inflated by shared method biases and at least sometimes by trivial construct overlaps, we could have found that LS's true predictability is lower than is usually observed.[8] Yet, the predictability of LS turned out to be considerably higher. It is unsurprising that LS overlaps with other personality traits to some degree in normal circumstances, where people can shape and evaluate their life according to their traits. But our estimates of this overlap's true extent are strikingly high, suggesting that how satisfied people are with their lives is usually quite close to what one could expect from their personality traits (Costa & McCrae, 1980). So, most life circumstances and other influences that are relevant for LS are those that also shape personality traits more broadly and endure over time.

How much higher than the usual estimates?

Depending on the questionnaire, the Big Five domains have explained about 30% of LS's variance in self-report studies (Anglim et al., 2020; Busseri & Erb, in press). This translates to a maximum out-of-sample predictive accuracy of .55, assuming no over-fitting (Yarkoni & Westfall, 2017). Comprehensive facet sets such as those of the NEO Personality Inventory (Costa & McCrae, 1992) may explain up to about 40% of LS's variance (Anglim et al., 2020), translating to a maximum out-of-sample prediction of about .65. But these facets' assessments often directly ask about hopelessness, worthlessness, happiness, and optimism, and may therefore suffer from construct overlaps with LS, potentially leading to its over-estimated predictability. Indeed, less expansive facet sets explain less LS variance. For example, in a large sample tested with the Big Five Inventory (Soto & John, 2017), Stewart and colleagues (2022) found out-of-sample predictive accuracies of .48 and .50 for LS, respectively, for the Big Five domains and facets. In our single-method data, the Big Five domains provided about .65 out-of-sample predictive accuracy for LS. This suggests that our Big Five scales inherently captured more LS-related variance than many other Big Five scales, despite avoiding direct item overlaps. One plausible reason is that our Big Five scores were nearly orthogonal, thus capturing more personality variance in aggregate.

Generously putting the usual estimates' higher bound at .65, this is less than two-thirds of the .80 *true* predictive accuracy we observed for the Big Five domains and less than half of the .90 *true* predictive accuracy for items that capture personality nuances besides domains (after having z-transformed the correlations to make such comparisons meaningful). Therefore, the extent to which LS reflects personality traits may be underestimated by a factor of two or more in typical single-method Big Five studies, including various meta-analyses (e.g., Anglim et al., 2020). But again, our

[8] Our informal conversations with personality/LS researchers have reinforced that expectation.

findings would have been equally meaningful even if the findings did resemble typical estimates of single-method studies because these could have been biased upward – this could not have been known *a priori*.

Even regardless of how individual researchers prefer to theorize on the personality trait-LS overlap, the mere fact that this overlap may be about twice as strong as typical findings show is highly important in and of itself and must constrain any theorizing on LS's origins. That researchers care about this overlap's degree is evidenced by the thousands of citations to previous meta-analyses such as DeNeve and Cooper (1998) and Steel et al. (2008) and the hundreds of citations already attracted by Anglim et al. (2020), despite the results of these meta-analyses being likely distorted due to unaddressed measurement issues.

Not just semantically overlapping evaluations

It is possible that people's general evaluations of their lives (e.g., "Am happy with my life") sometimes overlap with their personality trait evaluations (e.g., "Am energetic", "Often feel misunderstood") for reasons that are trivial or make LS's assessments inconsistent with its definition. For example, the items may appear semantically overlapping, or people may think about their personality rather than their life *per se* when assessing their LS. However, assuming that people's evaluations of various specific life domains such as their job, career choice, relationships, financial situation, health, appearance, home, and country are less likely to overlap with these same personality traits for these same reasons, our results circumvent the possibility that LS's associations with personality traits are trivial.[9] This is because LS was highly correlated with a combination of eight DSs, and the different DSs largely shared (informant-reported) personality correlates among themselves and with LS. Moreover, the model trained to predict LS allowed predicting the DS aggregate almost as accurately as LS itself both within and across languages, besides predicting individual DSs. Also, we ensured that no personality item correlated with LS items more strongly than LS items correlated among themselves, supporting the traits' discriminant validity.

Robustness across samples and languages

It is also reasonable to think that LS's meaning and correlations with personality traits may be sensitive to context and/or assessment language, thus not necessarily replicating across diverse samples. Also, other factors may influence LS to different degrees across samples, leaving more or less room for personality-related variance. If so, for example, even findings based on the whole Estonian population would have limited relevance for the French, Americans, Angolans, or Vietnamese. Indeed, there is already evidence that LS's correlations with positive and negative affect can systematically vary across countries (Kööts-Ausmees et al., 2013; Kuppens et al., 2008).

In our data, some r_{true}s did vary across samples tested in different languages. For example, emotional stability's correlation with LS was considerably higher among Estonian speakers (r_{true} = .47) than among those tested in Russian/English (r_{true} = .34), whereas the correlation with conscientiousness was lower (r_{true} = .30 vs r_{true} = .38). Several individual items, reflecting personality nuances within and beyond the Big Five domains, also had somewhat different correlations among Estonian speakers than in other samples. However, although these cross-sample differences could speak to important questions about LS's context-sensitivity (besides translation differences), here we focus on the big picture according to our data: the patterns of how LS and DSs were related to one another and a range of personality traits remained highly replicable across several Western samples tested in different languages. This is best illustrated by our finding that models trained to predict LS in the Estonian-speaking sample tended to almost as accurately predict LS among people tested in Russian (living in Estonia) and English (living in Western Europe). This would not have been possible if LS's personality correlates were highly contextual. Such cross-sample predictive accuracy also has methodological implications, making it unlikely that the models were overfitted to data and that more complex models' predictive advantages reflected model complexity (Mõttus et al., 2020).

This does not mean LS's true associations with DSs and personality traits could not vary across more diverse samples, such as those with non-European backgrounds or living in vastly different socioeconomic circumstances.

LS's stable variance is largely shared with personality traits

LS is far from perfectly stable over time. But personality traits' involvement in it endures over time because their longitudinal associations over several years were about as strong as cross-sectional ones in both directions. In fact, people's LS about ten years earlier could be predicted from their later personality traits at least as accurately as it could

[9] For example, even if people (partly) base their rating to the item "Am happy with my life" on how well they think others understand them (or these items are semantically overlapping), this seems less likely for items asking about satisfaction with health, appearance, and residency.

be predicted from LS itself. So, LS may fluctuate spontaneously or respond to variable circumstances, but its stable variance is largely shared with personality traits' stable variance. This finding also mitigates the concern that our findings may be specific to circumstances concurrent to our main data collection, such as the pandemic or the looming Russian invasion of Ukraine.

Is satisfaction just a reflection of other traits?

Although different ways of assessing satisfaction – LS and DSs – strongly overlap with personality traits in usual circumstances, being satisfied (with life) can remain conceptually distinct from personality more broadly. In some hypothetical circumstances, the associations of LS and DSs with some personality traits, such as feeling understood by others, might be weakened because the satisfactions are primarily shaped by strong external influences beyond an individual's control, while the personality traits remain less influenced. As one possibility, thus, the strength of the overlap between LS and personality traits can be seen as a measure of the extent to which individuals can influence and assess their lives according to their traits. The more satisfaction appears as a stable, observable, and partly heritable trait that similarly manifests across different life domains and is entangled with other traits, the more it could reflect people's own choices, aspirations, behaviors, skills, and emotional and cognitive processes, rather than external circumstances imposed on people without their own involvement. While here this remains an untested hypothesis meant to illustrate the conceptual distinction between LS and personality traits, it could be tested by studying personality trait-LS associations in highly unusual, uncontrollable, and restrictive circumstances such as living in a war zone (for relevant studies, see Cheung et al., 2020 and Coupe & Obrizan, 2016).

It is unnecessary to assume that particular personality traits are LS's directly-interpretable causes, though, even when they strongly correlate with LS. People differ in many traits, and each of these can contribute to and be further shaped by multiple traits and outcomes, including LS. This means that causal contributions can crisscross multiple traits and outcomes in any number of ways (Avinun, 2020), making them correlated over time but potentially leaving some or many of the individual causal pathways too complex to be meaningfully interpretable on their own (Brown & Rohrer, 2020; Mõttus et al., 2020). If so, the overall correlatedness among personality traits and variables like LS might often provide as much insight as their individual associations.

Still room for other influences

Although correlations as high as .90 are uncommon in psychology, even when corrected for measurement error, they must not be over-interpreted. Even such strong population trends leave considerable room for individuals to deviate from them, especially for those with the variables' medium levels (Mõttus, 2022). For example, if we trisected both predicted and observed LS, their .90 correlation would mean that the predicted and actual LS levels are different for every fourth individual. Specifically, every fifth individual predicted to have a high or low LS would actually have a different LS level, whereas among those predicted to have a medium LS, nearly two out of five would defy the prediction (Figure 5). Put differently, as the typical difference between two normally distributed measurements correlating at .90 is approximately a third of a standard deviation, most individuals' observed LS differs from its predicted-from-personality value by about the influence one would expect from a consequential life event (e.g., Denissen et al., 2019; Luhmann et al., 2012). This means there is still room for factors beyond those also captured in personality traits to explain why some people's LS is higher or lower than expected from their personality traits. However, the factors that are also captured in personality traits – enduring life circumstances, idiosyncratic experiences, or genetics – matter more for most people, most of the time.

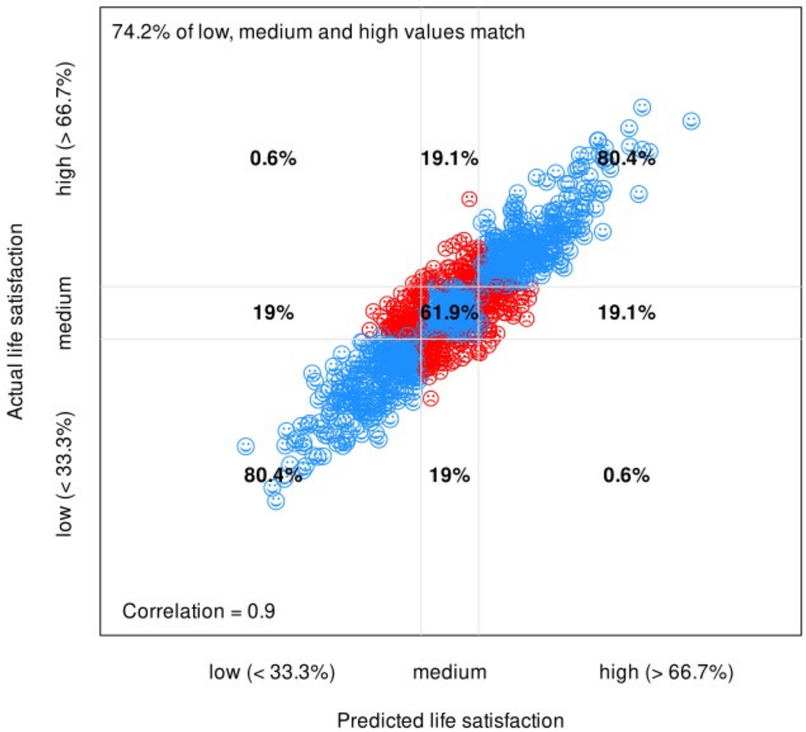

Figure 5. Predicted-from-personality and actual satisfaction levels, overlapping for blue individuals (74%) but being incongruent for red individuals (26%).

Which traits are most strongly linked with LS?

For the Big Five domains, LS had .30 to .50 r_{true}s with emotional stability, extraversion, and conscientiousness. Although somewhat higher, these correlations are consistent with those usually observed in single-method studies (Anglim et al., 2020; Supplementary Table S5), despite not being influenced by several common methodological limitations of these studies. However, agreeableness and openness had small true correlations with LS, while agreeableness usually has stronger correlations with LS in single-method studies (see Anglim et al., 2020). Given our single-method correlations, this difference from previous research could partly result from our use of r_{true}s or the near-orthogonality of our Big Five domains; the domains are usually more inter-correlated in other Big Five measures, contributing to spurious correlations with other variables (Busseri & Erb, in press; Stewart et al., 2022). Previous work with orthogonal Big Five scores has also resulted in somewhat weaker agreeableness-LS correlations, at least in self-reports (Busseri & Erb, in press; McCrae & Costa, 1991). Another part of the explanation may lie with cultural differences because the NEO-PI-3 agreeableness domain also had relatively small true and single-method correlations with LS in the Estonian-speaking data (Supplementary Table S6); the r_{true} was also slightly higher among our English-speaking participants.

However, the Big Five items often differed in their correlations with LS, as is common for many other outcomes (e.g., Revelle et al., 2021; Seeboth & Mõttus, 2018; Stewart et al., 2022). Many items also had stronger correlations with LS than any domain, including items not included in the domains. As for LS's strongest and relatively distinct correlates, its low levels were associated with feeling misunderstood, unexcited, indecisive, envious, bored, and used by others, whereas high LS tended to go with confidence in one's abilities and believing that efforts are rewarded.

In fact, even three items (feeling misunderstood, lack of excitement, and being indecisive) provided (out-of-sample) prediction of LS with true accuracy of .80, suggesting that most people with low LS could be recognized from just a few personality nuances. This is comparable to the predictive accuracy provided by the Big Five domains that encompass a broad range of traits, some of which are more and some less correlated with LS, making the domain-level results more ambiguous (Mõttus, 2016) and conducive to "just so" stories. For example, if we were only told that many people with low LS are low on emotional stability, extraversion, and conscientiousness, we could explain low LS by referring to any number of traits subsumed under these broad domains, only a few of which might actually correlate with LS. So, without further specifics, we could easily indulge in baseless speculations. Instead, now knowing that most people with low LS tend to feel misunderstood, lack excitement, and struggle with making decisions, our degrees of freedom in explaining low LS become smaller. Also, low emotional stability, low extraversion, and low conscientiousness tend to go with many undesirable outcomes, offering limited discriminant validity in explaining these outcomes. It remains to be seen if the LS's more nuanced personality correlates are specific to this outcome, offering greater discriminant validity and suggesting that there are factors that shape LS specifically rather than a desirable life more generally.

In conclusion, although the Big Few domains will continue to provide a parsimonious representation of LS's associations with personality traits, supplementing domain-level analyses with nuances offers a richer and more accurate picture of how LS intersects with psychological traits more broadly, besides providing greater predictive accuracy. The ability to estimate error-free associations with multi-rater or multi-timepoint data (Wood et al., 2022) is particularly useful for this research because it makes nuance-level associations directly comparable to those of aggregate personality traits.

The self is not privileged to evaluate LS

One may think that people's LS levels are private. However, if we accept that people's personality traits are to some extent observable to others (Connelly & Ones, 2010), then we have to accept the same for LS because its cross-rater agreement is similar to that of personality traits (Dobewall et al., 2013; Schneider & Schimmack, 2009). Moreover, we found that self-reported personality traits' correlations with informant-reported LS were similar to informant-reported personality traits' correlations with self-reported LS (footnotes 4 and 6), suggesting that self- and informant-reports of LS contained comparable degrees of information about personality traits. Besides, we calculated domains' and nuances' r_{true}s with LS and their true predictive accuracies for LS based on just self-reported LS (Supplementary Analyses), and none of the findings were different enough to change our conclusions, further alleviating the concerns that our findings could have been an artefact of combining self-reported LS with informant-reported LS. Further, our DS-related analyses did not include informant-reports, yet the patterns of findings were similar to those of LS-related analyses that did include informant-reports.

Of course, self-informant agreement is high for neither LS nor personality traits. For example, if we trisected self- and informant-report scores that correlate about .50, the targets' scores would be similar in only about half of the self-informant pairs (Mõttus, 2022). However, our method only required that there was *some* agreement and, ideally, that self- and informant-reports were available for both variables being correlated; the imperfect agreement would then cancel out because it would similarly influence both the numerator and denominator in r_{true} calculations (see Supplementary Material for the algebraic proof). For our analyses involving DSs, informant-reports were unavailable, so their r_{true}s could have been somewhat distorted. However, given that there was a substantial level of cross-rater agreement for all items for which both self- and informant-reports were available – personality items, LS items, and three DSs items – and DS-related and LS-related findings were similar, it is unlikely that even the DS-related r_{true}s were distorted enough to bias our conclusions.

In short, we found no compelling evidence that our use of informant-reported LS, in addition to self-reports, caused the observed pattern of findings.

LS's construct validity

Desirably, the three LS items assessed slightly different aspects of the construct because their r_{true}s were around .75 in the Estonian-speaking sample, unlike the near-unity r_{true}s among semantically nearly identical personality items before we removed redundant items. In the smaller samples tested in Russian and English, the three items had somewhat more variable but still high r_{true}s among themselves, with the variability likely due to their higher sampling variance (all correlations were within +/- two standard errors from the Estonian estimates). Substantially higher true correlations among the LS items would have been undesirable, narrowing the construct's scope (Clark & Watson, 1995). Supporting LS's construct validity, its items correlated among themselves more strongly than they correlated with personality items, and they had highly similar correlation profiles with personality items, showing similar broader psychological backgrounds. In the Supplementary Analyses, we also showed that LS's r_{true}s with domains and nuances, as well as personality traits' true predictive accuracy for LS, would have been quite similar – although generally somewhat lower due to LS being more narrowly defined – if we had assessed LS with only one single item, "Am happy with my life".

Personality and life satisfaction

Moreover, in Supplementary Analyses, we also showed that our LS assessments correlated very highly – r = .80 (in Russian), .90 (in English), and .95 (in Estonian) – with the widely used SWLS scale (Diener et al., 1985). Finally, LS correlated very highly with the aggregate of a range of DSs, which in turn correlated extremely highly with the SWLS, further aligning the LS's assessment with its definition.

Our three LS items covered a general life satisfaction assessment ("Am happy with my life"), purpose in life ("Feel that my life lacks direction"), and perspective on the future ("Have a dark outlook on the future"). Arguably, thus, our LS assessment had a broader scope than the SWLS (Diener et al., 1985) despite their very high empirical overlap. For example, our LS assessment also covered an aspect of the eudaimonic well-being (purpose) besides the hedonic well-being aspects usually associated with LS (Ryff et al., 2021).[10] Given this, it is not surprising that the LS's strongest correlates included items beyond those directly referring to emotional well-being. In particular, our LS assessment and many of its correlates fit with the components of Ryff's (1989) model of psychological well-being, which includes: positive relations (e.g., items about feeling understood, trust, liking others, and enjoying cooperation), autonomy, environmental mastery and personal growth (e.g., items about self-competence, learning quickly, solving complex problems, leadership and influencing others, believing in hard work, taking risks, learning new things and visiting new places, and self-improvement), purpose in life (e.g., items about lack of excitement, indecisiveness, avoiding responsibilities, and boredom) and self-acceptance (e.g., an item about wisdom). Thus, the hedonic and eudaimonic well-being aspects may overlap more than often thought, both empirically and in their broader psychological correlates.

In conclusion, we believe that our findings provide strong evidence for the validity of a broad LS construct in general and our LS assessments in particular.

Limitations

Our study did not have the usual limitations of personality research, such as relying on a monocultural sample, self-report-only measures, brief questionnaires, broad trait domains, or a single operationalization of the target construct, nor did it suffer from limited statistical power. However, although our personality item pool was intentionally expansive, it almost certainly did not cover all possible personality nuances, hence likely missing some LS-relevant personality traits. If so, we could underestimate personality traits' predictive accuracy for LS. However, given that our estimated true predictive accuracy was as high as .90, the completely missed personality content could not have been extensive. Likewise, our list of eight DSs likely missed some life domains that may be particularly relevant to some people's well-being. Also, the DSs were only assessed with self-reports, introducing possible biases to their r_{true}s with LS and personality traits. Further, our samples were convenience samples with a high percentage of females and high levels of education, possibly leading to underestimated correlations due to reduced variance. Finally, future studies should aim to generalize our findings to more diverse populations.

Conclusion

When addressing common methodological limitations, most people's LS levels are accurately predictable from their personality traits, even when avoiding direct construct overlaps. This does not mean LS is inherently and irrevocably reducible to personality traits. Instead, the degree to which it reflects personality traits may be seen as a measure of people's freedom to shape and assess their lives according to their traits. At least hypothetically, there could be circumstances where LS is less aligned with personality traits. But in usual circumstances, there does not seem to be much reason to think that LS is shaped by circumstances unrelated to personality more broadly – for most people, and most of the time, their satisfaction level is just about what we would expect from their other traits. Personality traits can be shaped by any number of factors, but usually, these same factors also shape LS, through personality or otherwise.

[10] We dropped the item referring to life having been kind to the person because it was less consistent with other LS items in all three languages. However, the SWLS has a parallel item: "So far I have gotten the important things I want in life"

References

Anglim, J., Horwood, S., Smillie, L. D., Marrero, R. J., & Wood, J. K. (2020). Predicting psychological and subjective well-being from personality: A meta-analysis. *Psychological Bulletin*, *146*, 279–323. https://doi.org/10.1037/bul0000226

Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO Model of Personality Structure—And why those Objections Fail. *European Journal of Personality*, *34*(4), 492–510. https://doi.org/10.1002/per.2242

Avinun, R. (2020). The E Is in the G: Gene–Environment–Trait Correlations and Findings From Genome-Wide Association Studies. *Perspectives on Psychological Science*, *15*(1), 81–89. https://doi.org/10.1177/1745691619867107

Bainbridge, T. F., Ludeke, S. G., & Smillie, L. D. (2022). Evaluating the Big Five as an organizing framework for commonly used psychological trait scales. *Journal of Personality and Social Psychology*, *122*, 749–777. https://doi.org/10.1037/pspp0000395

Beck, E. D., & Jackson, J. J. (2022). A mega-analysis of personality prediction: Robustness and boundary conditions. *Journal of Personality and Social Psychology*, *122*, 523–553. https://doi.org/10.1037/pspp0000386

Bouchard, T. J. (2016). Experience producing drive theory: Personality "writ large". *Personality and Individual Differences*, *90*, 302–314. https://doi.org/10.1016/j.paid.2015.11.007

Boyce, C. J., Brown, G. D. A., & Moore, S. C. (2010). Money and Happiness: Rank of Income, Not Income, Affects Life Satisfaction. *Psychological Science*, *21*(4), 471–475. https://doi.org/10.1177/0956797610362671

Brown, N. J. L., & Rohrer, J. M. (2020). Easy as (Happiness) Pie? A Critical Evaluation of a Popular Model of the Determinants of Well-Being. *Journal of Happiness Studies*, *21*(4), 1285–1301. https://doi.org/10.1007/s10902-019-00128-4

Busseri, M. A., & Erb, E. M. (in press). The happy personality revisited: Re-examining associations between Big Five personality traits and subjective well-being using meta-analytic structural equation modeling. *Journal of Personality.* https://doi.org/10.1111/jopy.12862

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81–105. https://doi.org/10.1037/h0046016

Caspi, A., Roberts, B. W., & Shiner, R. L. (2005). Personality Development: Stability and Change. *Annual Review of Psychology*, *56*, 453–484. https://doi.org/10.1146/annurev.psych.55.090902.141913

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Cheung, F., Kube, A., Tay, L., Diener, E., Jackson, J. J., Lucas, R. E., Ni, M. Y., & Leung, G. M. (2020). The impact of the Syrian conflict on population well-being. *Nature Communications, 11*(1), 3899. https://doi.org/10.1038/s41467-020-17369-0

Cheung, F., & Lucas, R. E. (2016). Income inequality is associated with stronger social comparison effects: The effect of relative income on life satisfaction. *Journal of Personality and Social Psychology, 110*(2), 332–341. https://doi.org/10.1037/pspp0000059

Condon, D. M., & Revelle, W. (2016). Selected ICAR Data from the SAPA-Project: Development and Initial Validation of a Public-Domain Measure. *Journal of Open Psychology Data*, *4*(1), e1. https://doi.org/10.5334/jopd.25

Condon, D. M., Wood, D., Mõttus, R., Booth, T., Costantini, G., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A. G. C., Ziegler, M., & Zimmermann, J. (2020). Bottom Up Construction of a Personality Taxonomy. *European Journal of Psychological Assessment*, *36*(6), 923–934. https://doi.org/10.1027/1015-5759/a000626

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*, *136*(6), 1092–1122. https://doi.org/10.1037/a0021212

Costa, P. T., Jr., & McCrae, R. R. (1980). Influence of Extraversion and Neuroticism on subjective well-being: Happy and unhappy people. *Journal of Personality and Social Psychology, 38,* 668-678.

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Psychological Assessment Resources.

Coupe, T., & Obrizan, M. (2016). The impact of war on happiness: The case of Ukraine. *Journal of Economic Behavior & Organization*, *132*, 228–242. https://doi.org/10.1016/j.jebo.2016.09.017

DeNeve, K. M., & Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin, 124*(2), 197–229. https://doi.org/10.1037/0033-2909.124.2.197

Denissen, J. J. A., Luhmann, M., Chung, J. M., & Bleidorn, W. (2019). Transactions between life events and personality traits across the adult lifespan. *Journal of Personality and Social Psychology, 116*(4), 612–633. https://doi.org/10.1037/pspp0000196

Diener, E., Emmons, R. A., Larsen, R. J., & Griffin, S. (1985). The Satisfaction With Life Scale. *Journal of Personality Assessment*, *49*, 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Diener, E., Oishi, S., & Tay, L. (2018). Advances in subjective well-being research. *Nature Human Behaviour*, *2*(4), Article 4. https://doi.org/10.1038/s41562-018-0307-6

Diener, E., & Seligman, M. E. P. (2002). Very Happy People. *Psychological Science, 13*(1), 81–84. https://doi.org/10.1111/1467-9280.00415

Dobewall, H., Realo, A., Allik, J., Esko, T., & Metspalu, A. (2013). Self-Other Agreement in Happiness and Life-Satisfaction: The Role of Personality Traits. *Social Indicators Research*, *114*(2), 479–492. https://doi.org/10.1007/s11205-012-0157-y

Funder, D. C., & Ozer, D. J. (2019). Evaluating Effect Size in Psychological Research: Sense and Nonsense. *Advances in Methods and Practices in Psychological Science*, *2*(2), 156–168. https://doi.org/10.1177/2515245919847202

Heller, D., Watson, D., & Ilies, R. (2004). The Role of Person Versus Situation in Life Satisfaction: A Critical Examination. *Psychological Bulletin*, *130*, 574–600. https://doi.org/10.1037/0033-2909.130.4.574

Jagodzinski, W. (2010). Economic, Social, and Cultural Determinants of Life Satisfaction: Are there Differences Between Asia and Europe? *Social Indicators Research*, *97*(1), 85–104. https://doi.org/10.1007/s11205-009-9555-1

Johnson, W. (2010). Extending and testing Tom Bouchard's Experience Producing Drive Theory. *Personality and Individual Differences*, *49*(4), 296–301. https://doi.org/10.1016/j.paid.2009.11.022

Kööts-Ausmees, L., Realo, A., & Allik, J. (2013). The Relationship Between Life Satisfaction and Emotional Experience in 21 European Countries. *Journal of Cross-Cultural Psychology*, *44*(2), 223–244. https://doi.org/10.1177/0022022112451054

Kuppens, P., Realo, A., & Diener, E. (2008). The role of positive and negative emotions in life satisfaction judgment across nations. *Journal of Personality and Social Psychology*, *95*, 66–75. https://doi.org/10.1037/0022-3514.95.1.66

Lucas, R. E., Freedman, V. A., & Cornman, J. C. (2018). The short-term stability of life satisfaction judgments. *Emotion*, *18*, 1024–1031. https://doi.org/10.1037/emo0000357

Luhmann, M., Hofmann, W., Eid, M., & Lucas, R. E. (2012). Subjective well-being and adaptation to life events: A meta-analysis. *Journal of Personality and Social Psychology, 102,* 592–615. https://doi.org/10.1037/a0025948

McCrae, R. R. (2015). A More Nuanced View of Reliability: Specificity in the Trait Hierarchy. *Personality and Social Psychology Review*, *19*, 97–112. https://doi.org/10.1177/1088868314541857

McCrae, R. R., & Costa, P. T. (1991). Adding Liebe und Arbeit: The Full Five-Factor Model and Well-Being. *Personality and Social Psychology Bulletin, 17*(2), 227–232. https://doi.org/10.1177/014616729101700217

McCrae, R. R., Costa, P. T., & Martin, T. A. (2005). The NEO-PI-3: A More Readable Revised Neo Personality Inventory. *Journal of Personality Assessment*, *84*(3), 261–270. https://doi.org/10.1207/s15327752jpa8403_05

McCrae, R. R., & Mõttus, R. (2019). What Personality Scales Measure: A New Psychometrics and Its Implications for Theory and Assessment. *Current Directions in Psychological Science*, *28*(4), 415–420. https://doi.org/10.1177/0963721419849559

Mõttus, R. (2016). Towards more rigorous personality trait-outcome research. *European Journal of Personality*, *30*(4), 292–303. https://doi.org/10.1002/per.2041

Mõttus, R. (2022). What Correlations Mean for Individual People: A Tutorial for Researchers, Students and the Public. *Personality Science*, *xxx*, xxx.

Mõttus, R., Sinick, J., Terracciano, A., Hrebickova, M., Kandler, C., Ando, J., Mortensen, E. L., Colodro-Conde, L., & Jang, K. (2019). Personality characteristics below facets: A replication and meta-analysis of cross-rater agreement, rank-order stability, heritability and utility of personality nuances. *Journal of Personality and Social Psychology*, *117*, e35–e50.

Mõttus, R., Wood, D., Condon, D. M., Back, M. D., Baumert, A., Costantini, G., Epskamp, S., Greiff, S., Johnson, W., Lukaszewski, A., Murray, A., Revelle, W., Wright, A. G. C., Yarkoni, T., Ziegler, M., & Zimmermann, J. (2020). Descriptive, Predictive and Explanatory Personality Research: Different Goals, Different Approaches, but a Shared Need to Move beyond the Big Few Traits. *European Journal of Personality*, *34*(6), 1175–1201. https://doi.org/10.1002/per.2311

Oishi, S., Diener, E. F., Lucas, R. E., & Suh, E. M. (1999). Cross-Cultural Variations in Predictors of Life Satisfaction: Perspectives from Needs and Values. *Personality and Social Psychology Bulletin*, *25*(8), 980–990. https://doi.org/10.1177/01461672992511006

Paulhus, D. L., & Vazire, S. (2007). The self-report method. In *Handbook of research methods in personality psychology* (pp. 224–239). The Guilford Press.

Payne, J. W., & Schimmack, U. (2020). Construct validity of global life-satisfaction judgments: A look into the black box of self–informant agreement. *Journal of Research in Personality*, *89*, 104041. https://doi.org/10.1016/j.jrp.2020.104041

R Core Team. (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*.

Realo, A. (2006). Mis paneb eestlase elust rõõmu tundma? [What makes Estonians happy?]. *Horisont*, *2*, 24–28.

Rentzsch, K., & Gross, J. J. (2015). Who Turns Green with Envy? Conceptual and Empirical Perspectives on Dispositional Envy. *European Journal of Personality*, *29*(5), 530–547. https://doi.org/10.1002/per.2012

Revelle, W., Dworak, E. M., & Condon, D. M. (2021). Exploring the persome: The power of the item in understanding personality structure. *Personality and Individual Differences*, *169*, 109905. https://doi.org/10.1016/j.paid.2020.109905

Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of Personality and Social Psychology*, *57*, 1069–1081. https://doi.org/10.1037/0022-3514.57.6.1069

Ryff, C. D., Boylan, J. M., & Kirsch, J. A. (2021). Eudaimonic and hedonic well-being. Measuring well-being. In M T. Lee, L. D. Kubzansky, & T. J. Van der Weele (Eds.), *Measuring Well-Being* (pp. 92 - 135). Oxford University Press.

Schimmack, U. (2010). What multi-method data tell us about construct validity. *European Journal of Personality*, *24*(3), 241–257. https://doi.org/10.1002/per.771

Schimmack, U., Oishi, S., Furr, R. M., & Funder, D. C. (2004). Personality and Life Satisfaction: A Facet-Level Analysis. *Personality and Social Psychology Bulletin*, *30*(8), 1062–1075. https://doi.org/10.1177/0146167204264292

Schneider, L., & Schimmack, U. (2009). Self-Informant Agreement in Well-Being Ratings: A Meta-Analysis. *Social Indicators Research*, *94*(3), 363. https://doi.org/10.1007/s11205-009-9440-y

Seeboth, A., & Mõttus, R. (2018). Successful Explanations Start with Accurate Descriptions: Questionnaire Items as Personality Markers for More Accurate Predictions. *European Journal of Personality*, *32*(3), 186–201. https://doi.org/10.1002/per.2147

Soto, C. J. (2019). How Replicable Are Links Between Personality Traits and Consequential Life Outcomes? The Life Outcomes of Personality Replication Project. *Psychological Science*, *30*(5), 711–727. https://doi.org/10.1177/0956797619831612

Soto, C. J., & John, O. P. (2017). The Next Big Five Inventory (BFI-2): Developing and Assessing a Hierarchical Model With 15 Facets to Enhance Bandwidth, Fidelity, and Predictive Power. *Journal of Personality and Social Psychology*, *113*, 117–143. https://doi.org/10.1037/pspp0000096

Steel, P., Schmidt, J., & Shultz, J. (2008). Refining the relationship between personality and subjective well-being. *Psychological Bulletin, 134*, 138–161. https://doi.org/10.1037/0033-2909.134.1.138

Stewart, R. D., Mõttus, R., Seeboth, A., Soto, C. J., & Johnson, W. (2022). The finer details? The predictability of life outcomes from Big Five domains, facets, and nuances. *Journal of Personality*, *90*(2), 167–182. https://doi.org/10.1111/jopy.12660

Suh, E., Diener, E., Oishi, S., & Triandis, H. C. (1998). The shifting basis of life satisfaction judgments across cultures: Emotions versus norms. *Journal of Personality and Social Psychology*, *74*, 482–493. https://doi.org/10.1037/0022-3514.74.2.482

Thielmann, I., Moshagen, M., Hilbig, BenjaminE., & Zettler, I. (2021). On the Comparability of Basic Personality Models: Meta-Analytic Correspondence, Scope, and Orthogonality of the Big Five and HEXACO Dimensions. *European Journal of Personality*, 08902070211026793. https://doi.org/10.1177/08902070211026793

Turkheimer, E., Pettersson, E., & Horn, E. E. (2014). A phenotypic null hypothesis for the genetics of personality. *Annual Review of Psychology*, *65*(1), 515–540. https://doi.org/10.1146/annurev-psych-113011-143752

van der Linden, D., te Nijenhuis, J., & Bakker, A. B. (2010). The General Factor of Personality: A meta-analysis of Big Five intercorrelations and a criterion-related validity study. *Journal of Research in Personality*, *44*(3), 315–327. https://doi.org/10.1016/j.jrp.2010.03.003

Vazire, S. (2006). Informant reports: A cheap, fast, and easy method for personality assessment. *Journal of Research in Personality*, *40*(5), 472–481. https://doi.org/10.1016/j.jrp.2005.03.003

Vihalemm, T., Juzefovičs, J., & Leppik, M. (2019). Identity and Media-use Strategies of the Estonian and Latvian Russian-speaking Populations Amid Political Crisis. *Europe-Asia Studies*, *71*(1), 48–70. https://doi.org/10.1080/09668136.2018.1533916

Weiss, A., Bates, T. C., & Luciano, M. (2008). Happiness is a personal(ity) thing: The genetics of personality and well-being in a representative sample. *Psychological Science*, *19*(3), 205–210. psyh. https://doi.org/10.1111/j.1467-9280.2008.02068.x

Wood, D., Lowman, G. H., Armstrong III, B. F., & Harms, P. D. (2022). Using retest-adjusted correlations as indicators of the semantic similarity of items. *Journal of Personality and Social Psychology*, No Pagination Specified-No Pagination Specified. https://doi.org/10.1037/pspp0000441

Yarkoni, T., & Westfall, J. (2017). Choosing Prediction Over Explanation in Psychology: Lessons From Machine Learning. *Perspectives on Psychological Science*, *12*(6), 1100–1122. https://doi.org/10.1177/1745691617693393