Intelligence 113 (2025) 101963

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

Research Paper

Lack of measurement invariance in mental health assessment across intelligence levels: Investigation into nonlinearity reveals a broader issue

Stanisław K. Czerwiński *, Roman Konarski, Paweł A. Atroszko

University of Gdańsk, Bażyńskiego 4, 80-309 Gdańsk, Poland

A R T I C L E I N F O

Keywords:
Cognitive ability
Gender
Giftedness
Intelligence
Mental health
Nonlinear relationship

A B S T R A C T

The notion of individuals with extremely high intelligence experiencing specific psychological difficulties has been brought up many times by researchers, with no consensus being reached on this matter. Modeling the relationship between intelligence and mental health nonlinearly allows for revealing the possibility for extremely intelligent individuals to experience specific difficulties affecting their mental health while still having better mental health on average. Moreover, with gender roles ingrained in the social world, men and women are likely to have different experiences with how their social environments react to their high intelligence. This study used data from two cohort studies: the National Longitudinal Survey of Youth 1979 (NLSY79; $n = 8474$) and 1997 (NLSY97; $n = 6472$). Although, the results of polynomial and piecewise regression suggested that the relationship of intelligence with depression and distress is nonlinear, the results of local structural equation models showed the lack of measurement invariance of mental health measures across intelligence values for both men and women. The factor loadings for the measures decreased, as the intelligence values increased. As such, values of mental health are not comparable across different levels of intelligence. If this effect is also present on a larger number of psychometric instruments, the implications on psychological research could be substantial. The lack of measurement invariance of psychometric instruments across intelligence can compromise the results of studies, not only those testing for nonlinear relationships of intelligence, but also those testing for linear ones.

1. Introduction

A number of high-quality studies show a generally positive relationship between intelligence and mental health (e.g., Aichele et al., 2018; Gale et al., 2010; Keyes et al., 2017). However, more nuanced yet important effects are typically overlooked. Researchers, along with the popular press, persistently raise concerns about the potential vulnerability of individuals with high intelligence issues related to mental health (Brown et al., 2021; Karpinski et al., 2018). Studies that rely on group comparisons of mental health between gifted and non-gifted individuals generally show either no mean differences in mental health between the two groups or higher mean mental health in the gifted group (Martin et al., 2010; Williams et al., 2022; Zeidner, 2021). The Study of Mathematically Precocious Youth (SMPY), one of the most comprehensive investigations into intellectual exceptionality, tracks cohorts with exceptional mathematical reasoning abilities. It has produced a plethora of results showing that individuals with extremely high cognitive abilities obtain high educational, as well as occupational achievements, and show signs of good psychological adjustment and well-being (Lubinski et al., 2023; Lubinski & Benbow, 2006).

However, several studies conducted on large, representative samples that use nonlinear modeling provided evidence suggesting that the relationship between mental health and intelligence might be nonlinear in such a way that at high values of intelligence, it turns from positive to negative. Horwood et al. (2008) showed this effect for psychotic symptoms in a sample of over 6000 children from the UK, Gale et al. (2012) for bipolar disorder in a sample of over 1000,000 Swedish men, and Czerwiński et al. (2025) for a number of mental health measures in a sample of almost 8000 British adults. A closer examination of the methodological approaches used to study this relationship reveals important nuances that can explain the discrepancies between the results of different studies.

1.1. Methodological nuances in interpretation of nonlinear effects

In their analysis of data from four large cohort studies, Brown et al. (2021) also reported a curvilinear relationship between intelligence and depression, as well as between intelligence and subjective well-being.

* Corresponding author.
E-mail addresses: stanislaw.czerwinski@ug.edu.pl (S.K. Czerwiński), roman.konarski@ug.edu.pl (R. Konarski), pawel.atroszko@ug.edu.pl (P.A. Atroszko).

https://doi.org/10.1016/j.intell.2025.101963
Received 3 May 2025; Received in revised form 10 September 2025; Accepted 19 September 2025
Available online 24 September 2025
0160-2896/

The authors disregarded these results, interpreting the effect size as too small. This conclusion was drawn based on the recommended interpretation of the correlation coefficient's effect size (Funder & Ozer, 2019; Gignac & Szodorai, 2016) transformed into the explained variance added by introducing the quadratic term (ΔR2). They proposed an $\Delta R^2 = 0.01$ (mathematically equivalent to an $r = 0.10$) as a threshold for a meaningful effect. Although the R^2 and the correlation coefficient are very similar on a mathematical level, the context in which they are used is different, and they cannot always be treated interchangeably.

A review by Aguinis et al. (2005) demonstrated that the unique variance explained by an interaction term is very low in most empirical studies. One of the reasons for this is that the interaction term is highly correlated to the two variables that compose it, and as such, they share a lot of their variance. This effect is likely even more pronounced in the case of nonlinear relationships, as the quadratic term of the predictor is essentially an interaction of a predictor with itself. Liu and Yuan (2021) argue that the ΔR^2 is not an adequate effect size measure for interaction effects and propose the proportion of the unique variance explained by the interaction term to the variance explained by the predictor and the interaction term jointly (ΔR^2_{mo}) as a more informative alternative.

Regardless of the chosen effect size measure, it is essential to consider the context and implications of every statistical result and avoid excessive reliance on rigid cut-off points (see Funder & Ozer, 2019 for an in-depth discussion on the subject). For the case of the nonlinear relationship between intelligence and mental health, it is assumed (and corroborated by previous research) that the inflection point is located at the high values of the predictor. Consequently, the nonlinearity only affects a certain percentage of the sample and, by extension, only a certain percentage of the statistical model. While this may result in reduced effect sizes, it does not diminish the meaningful information the effect provides.

Another methodological problem recurring in many previous studies regarding the experiences of individuals with high intelligence and the way their results are interpreted is that by using group comparisons, they are dichotomizing a continuous variable of intelligence. Categorization of continuous variables is known to have many statistical disadvantages (Altman & Royston, 2006; Naggara et al., 2011). As such, it is imperative to emphesize that findings from studies indicating higher mental health among individuals with high intelligence and those revealing a nonlinear relationship between intelligence and mental health are not contradictory to one another. When the inflection point of a nonlinear relationship occurs at high values of a predictor, the mean value of the dependent variable may still be higher for individuals above the inflection point compared to those below it due to starting at a higher average point on the dependent variable. As such, extremely intelligent individuals can experience specific difficulties affecting their mental health while still having better mental health on average (see Fig. 1). This means that some of them may experience considerable distress and associated mental health problems even though, on average, the group shows better mental health. Currently, there is a paucity of research devoted to understanding the factors influencing this phenomenon.

1.2. The nonlinear relationship between intelligence and mental health

A phenomenon of a generally desirable trait having potentially negative consequences at high enough values is referred to as the too-much-of-a-good-thing effect (Grant & Schwartz, 2011; Pierce & Aguinis, 2013). While exceptionally high intelligence in itself may not necessarily be harmful, it may become problematic for some individuals with additional risk factors, including personality, socialization, culture, or some biological predispositions (e.g., autism spectrum disorder). Extremely intelligent individuals might experience feelings of isolation due to functioning differently at a fundamental level compared to their peers (Falck, 2020, p. 82–85; Flakus et al., 2021). A sense of belongingness, identified as a fundamental psychological need according to Self-Determination Theory (Ryan & Deci, 2017), may be compromised. Research on perceived desirability based on the prospective intelligence of potential partners indicates that ratings of desirability peak at the 90th percentile, with a subsequent decrease from the 90th to the 99th percentile, because of compatibility concerns and concerns about the social skills of the extremely intelligent potential partners (Gignac et al., 2018; Gignac & Callis, 2020; Gignac & Starbuck, 2018).

Similarly, teachers also perceive gifted students as less extroverted, emtionally stable and agreeable (Baudson & Preckel, 2013). It has been proposed that individuals with high intelligence might struggle with issues of unidimensional identity based solely on their high intellectual ability (Coakley, 1992; Czerwiński et al., 2025). This notion finds support in studies indicating that adolescents formally identified as gifted show worse mental health compared to their non-identified peers

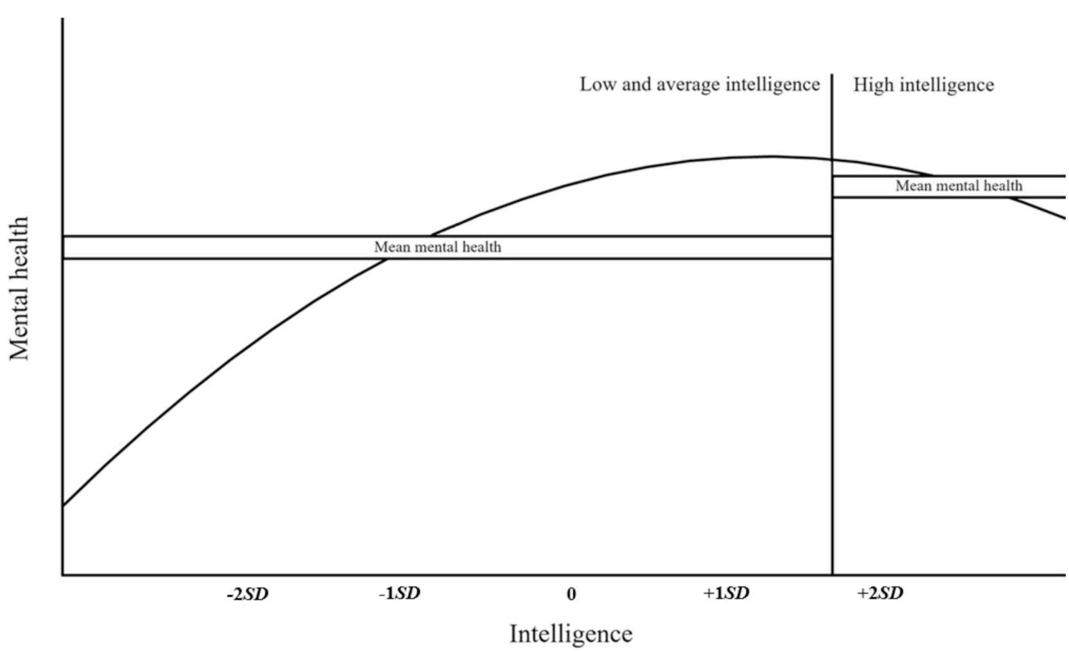

Fig. 1. Visualization of how dichotomizing a continuous variable can fail to detect a nonlinear effect.

(Freeman, 2006; Lavrijsen & Verschueren, 2023).

1.3. Longitudinal effects

One area that could be the root of the problems potentially experienced by individuals with extremely high intelligence is social functioning and interpersonal relationships. Indicators of social functioning, such as loneliness, social support, and social anxiety, have consistently demonstrated associations with changes in mental health over time (Beesdo et al., 2007; Holahan & Moos, 1981; Hu et al., 2020; Lee et al., 2021). Establishing satisfying social relationships can have a preventive effect against the exacerbation of any psychological distress. While individuals with high intelligence do not necessarily score lower on any traditional measure of social functioning, they still might experience some feelings of exclusion and isolation due to being treated differently by their environment or having a harder time finding peers they could relate to (Jackson, 1998; Jackson & Peterson, 2003). A study by Major et al. (2014) revealed a reverse U-shaped relationship between intelligence and sociability. Autism spectrum disorder, which is characterized by experiencing difficulties with social interactions, is also more common at extreme values of intelligence (Billeiter & Froiland, 2023). As such, it is plausible that the relationship between intelligence and the rate of change in mental health over time might also be nonlinear. Higher intelligence has been proven to predict a slower increase in depressive symptoms over time (Aichele et al., 2018); however, no previous study has tested the possible nonlinearity of this relationship.

1.4. The role of gender

Another aspect which could prove to be relevant in the relationship between intelligence and mental health is the role of gender. With gender roles being ingrained in the social world, men and women are likely to have different experiences with how their social environments react to their high intelligence. Notably, women remain underrepresented in the science, technology, engineering, and mathematics (STEM) fields, at high managerial positions, and receive lower average pay (Freeman & Garces-Basca, 2015). The disparity in salary has also been shown to be present in the SMPY cohort, with men earning about 50 % more than women (Lubinski et al., 2014). These are likely the results of systemic factors such as stereotype threat and implicit biases regarding gender (Nosek et al., 2009; Steele, 1997).

Furthermore, highly intelligent individuals, especially highly intelligent women, are less inclined to decide to become parents (Kanazawa, 2014). Given the greater social pressure for women to have children, it is more likely that their decision to remain childless will be met with disapproval from their environment. Though there are no gender differences in the means of intelligence, data suggests that women have a lesser variance in intelligence test scores (Halpern, 2011). In consequence, extremely highly intelligent women may feel more alienated due to being in the minority. All of these challenges may be further exacerbated by a higher prevalence of social anxiety in women (Asher et al., 2017).

For example, in school, gifted girls are more likely to try to hide their abilities to avoid standing out from their peers (Callahan & Hébert, 2020; Reis & Hébert, 2008). Gifted women are also more likely to lose confidence in their own abilities over the course of their educational journey. They tend to attribute their academic successes to luck or hard work, while attributing failures to a perceived lack of ability (Callahan & Hébert, 2020; Reis & Hébert, 2008). In contrast, gifted men attribute their academic success to their ability and failures to a lack of effort. Women, in general, exhibit a higher tendency for rumination (Johnson & Whisman, 2013), have nightmares more frequently (Schredl & Reinhard, 2011), and are more likely to experience emotional exhaustion from work-related burnout (Purvanova & Muros, 2010). They have a higher prevalence of affective disorders, which has been attributed to several factors, from social, such as gender role socialization, to biological, such as increased genetic vulnerability (McLean & Anderson, 2009). The Social Role Theory posits that differences between men and women stem from gender role beliefs they are prescribed through biosocial processes, and to which they are socialized into (Eagly & Wood, 2011). There is also data suggesting that, when investigated linearly, the positive relationship between intelligence and mental health is stronger for women than for men (Hatch et al., 2007). It is likely that there also might be a difference in the functional form of the relationship.

1.5. The present study

The aim of this study was to investigate the nonlinearity of the relationship between intelligence and mental health and the role of gender in that relationship. Additionally, longitudinal data were analyzed to test whether the changes in mental health will also be predicted nonlinearly by intelligence. The study was not pre-registered. Based on previous research and the theoretical frameworks described above, the following hypotheses were tested:

1. Intelligence is nonlinearly related to mental health and changes in mental health in such a way that the relationship is positive until it reaches an inflection point and changes direction at high values of intelligence (a reverse U-shape as presented in Fig. 1).
2. This effect is more pronounced for women than it is for men.

2. Method

2.1. Participants and procedure

Data from the National Longitudinal Survey of Youth 1979 (NLSY79) and the National Longitudinal Survey of Youth 1997 (NLSY97) provided by the United States Bureau of Labor Statistics were used (Bureau of Labor Statistics, U.S. Department of Labor, 2021, 2023). The data is freely available at https://www.nlsinfo.org/investigator/.

NLSY79 consists of a nationally representative sample of approximately 12,500 individuals born between 1957 and 1964 in the United States. The first wave of the study was conducted in 1979. The following waves were conducted annually until 1994 and biennially thereafter. Eight thousand four hundred seventy-four participants had complete data on all variables of interest. A total of 4178 (49.3 %) were men, and 4296 (50.7 %) were women. In terms of ethnicity, 2574 (30.4 %) were Black, 1620 (19.1 %) were Hispanic, and 4280 (50.5 %) were Non-Black / Non-Hispanic. The mean age of the participants in 1979 was $M =$ 17.55 years ($SD = 2.25$).

NLSY97 consists of a nationally representative sample of approximately 9000 individuals born between 1980 and 1984 in the United States. The first wave of the study was conducted in 1997. The following waves were conducted annually until 2011 and biennially thereafter. Six thousand four hundred seventy-two participants had complete data on all variables of interest at the first measurement occasion. A total of 3267 (50.2 %) were men, and 3205 (49.5 %) were women. In terms of ethnicity, 1668 (25.8 %) were Black, 1227 (19.0 %), were Hispanic, 66 (1.0 %) were Mixed (Non-Hispanic), and 3511 (54.2 %) were Non-Black / Non-Hispanic. The mean age of the participants in 1997 was $M =$ 14.93 years ($SD = 1.39$).

Brown et al. (2021) also used the data from the NLSY79 and the NLSY97, but there are key differences from the analyses presented in this paper. The study by Brown et al. (2021) did not use structural equation modeling (SEM), did not account for the multi-factor structure of the Mental Health Inventory-5 (MHI-5; Berwick et al., 1991), did not test moderation by gender, and did not test longitudinal effects.

2.2. Instruments

2.2.1. The Armed Forces Qualification Test

The Armed Forces Qualification Test (AFQT; Bayroff & Anderson, 1963) was used in both NLSY79 and NLSY97 as a measure of intelligence. AFQT consists of four subscales from the Armed Services Vocational Aptitude Battery (ASVAB) which measure cognitive abilities: arithmetic reasoning, mathematics knowledge, word knowledge, and paragraph comprehension. The AFQT scores are highly correlated with the Wechsler Adult Intelligence Scale (WAIS; McGrevy et al., 1974) and are commonly used as a proxy measure of intelligence in psychological research (e.g., Deary et al., 2007). The test was administered in the first waves of both studies. The original standardization procedure is described at: https://www.nlsinfo.org/content/cohorts/nlsy79/topical-guide/education/aptitude-achievement-intelligence-scores. The scores were coded as centiles so an inverse normal distribution transformation was used to obtain standardized z-scores for the purposes of this study. The scale showed high test-retest reliability (0.84–0.93) in previous studies (Defense Manpower Data Center, 2008).

2.2.2. Center for Epidemiological Studies – Depression

A 7-item version of the Center for Epidemiological Studies – Depression (CES—D; Radloff, 1977) was used in NLSY79. The instrument consists of 7 items asking how often the participant has felt a certain way during the past week (e.g., "I felt sad") with a 4-point Likert-type response format. Data from the 1992 measurement wave was included in the study. The participants of NLSY79 were measured with CES-D on several different occasions; however, the time intervals between each measurement varied significantly for each individual. Due to the difficulty of dealing with individually varying time intervals in a satisfying manner, only the data from when the test was first administered in 1992 were included in the analyses. The Cronbach's alpha reliability coefficient was of 0.77 in the sample used.

2.2.3. Mental Health Inventory-5

Mental Health Inventory-5 (MHI-5; Berwick et al., 1991) was used in NLSY97. The scale has a two-factor structure (Meybodi et al., 2011; Rivera-Riquelme et al., 2019). Three items measure psychological distress (e.g., "How much of the time, during the last month, have you felt downhearted and blue?"), and two items measure well-being (e.g., item "How much of the time, during the last month, have you been a happy person?"). A 4-point Likert-type response format was used. The test was administered in the years 2000, 2002, 2004, 2006, 2008, 2010, 2015, and 2017. Although data from 2019 were also available, they were not included in the analyses, as there was only data from around 20 % of participants compared to 2017. The items were summed to create total scores for each of the factors. For psychological distress, the Cronbach's alpha reliability coefficients in each subsequent wave were of 0.69, 0.70, 0.70, 0.69, 0.70, 0.72, 0.72, and 0.71, respectively. For well-being, the split-half Spearman-Brown's reliability coefficients in each subsequent wave were of 0.70, 0.70, 0.72, 0.73, 0.74, 0.75, 0.75, and 0.76, respectively.

2.3. Statistical analyses

Analyses were conducted in R 4.3.0 (R Core Team, 2023) with the *lavaan* and *sirt* packages, and visualized with the *ggplot2* package. All tests were two-tailed, and the significance level was set to $\alpha = 0.05$. The code for the analyses is available through Open Science Framework at the following link: https://osf.io/bq3pz/?view_only=ce75b1edf3fc4bcf81d727702901de39

To begin with, CES-D 7 and MHI-5 were tested for measurement invariance. CES-D 7 showed metric invariance between genders, and MHI-5 showed full scalar invariance across measurement occasions and metric invariance between genders in the present sample (see Supplementary material).

2.3.1. Polynomial and piecewise regression

For the NLSY79 data, SEM was applied to model the relationship between intelligence and depression. Intelligence and a quadratic term of intelligence were used as independent observed variables predicting a latent variable comprised of CES-D items. To calculate the ΔR^2 and the ΔR^2_{mo}, linear model with intelligence and depressiveness was also calculated. If the quadratic term was significant in a given model, then another model that replaced polynomial regression with piecewise regression (also referred to as segmented regression or interrupted regression) was tested. Piecewise regression is constructed of linear segments that are divided by breakpoints and provides regression coefficients for each of those segments (McZgee & Carleton, 1970). The estimated inflection point from the quadratic regression determined the placements for these breakpoints. Several researchers proposed this procedure to address the concerns that a significant quadratic coefficient might not be sufficient evidence for a U-shaped (or reverse U-shaped) relationship as the relationship might still be non-significant after the inflection point (Iribarren et al., 1996; Nelson & Simonsohn, 2014; Nickel et al., 2019). The usage of piecewise regression is the reason for intelligence being treated as an observed variable rather than a latent one, as to the authors' knowledge, applying piecewise regression with a latent predictor is not possible. The estimated inflection points were also transformed to a standardized IQ scale ($M = 100$, $SD = 15$) for more straightforward interpretation. Fig. 2 presents a path diagram of the cross-sectional model tested in the NLSY79 sample using polynomial regression.

For the NLSY97 data, latent growth curve modeling (LGCM) was applied to model changes in psychological distress and well-being from 2000 to 2017. In these models, the intercept represents a variable's initial level, and the slope represents the rate of change over time. Intelligence and a quadratic term of intelligence were used as independent variables predicting the intercepts and slopes of distress and well-being. The change over time was set to be linear. Again, models where the relationships between the predictor and the dependent variables were linear were also tested to calculate the ΔR^2 and the ΔR^2_{mo}. The models for distress and well-being were analyzed independently. Missing data in subsequent measurements were handled using full information maximum likelihood (FIML) estimation. Fig. 3 presents a path diagram of the longitudinal model tested in the NLSY97 sample with the use of polynomial regression.

The Robust Maximum Likelihood (MLR) estimator was used. Simulation studies show that the MLR estimator we used performs well with highly skewed data when the sample size is large (e.g., Bandalos, 2014). The following measures were used to evaluate model fit: the Root Mean Squared Error of Approximation (RMSEA), the Comparative Fit Index (CFI), and the Standardized Root Mean Square Residual (SRMR). Cut-off scores for those indexes for an acceptable fit that were utilized were CFI ≥ 0.90, RMSEA ≤ 0.08, and SRMR ≤ 0.08 (Hu & Bentler, 1999; Schreiber et al., 2006).

Multigroup analyses were performed to test the moderating role of gender. For each regression parameter, a separate nested model was tested with a given regression parameter constrained to be equal across men and women. The Satorra-Bentler scaled chi-square difference statistic (Satorra & Bentler, 2010) was used to test the difference in fit between the nested models and the baseline models independently for each regression parameter. A significant difference suggests that the constrained parameter is different across the groups.

2.3.2. Local structural equation models

As the application of polynomial and piecewise regression for testing nonlinear relationships has come under scrutiny in recent years (Breit et al., 2025; Simonsohn, 2018), local structural equation modeling (LSEM; Hildebrandt et al., 2009; Hildebrandt et al., 2016) was also applied to check the robustness of the results. LSEM is a method for modeling parameters of SEM models as a continuous, non-parametric

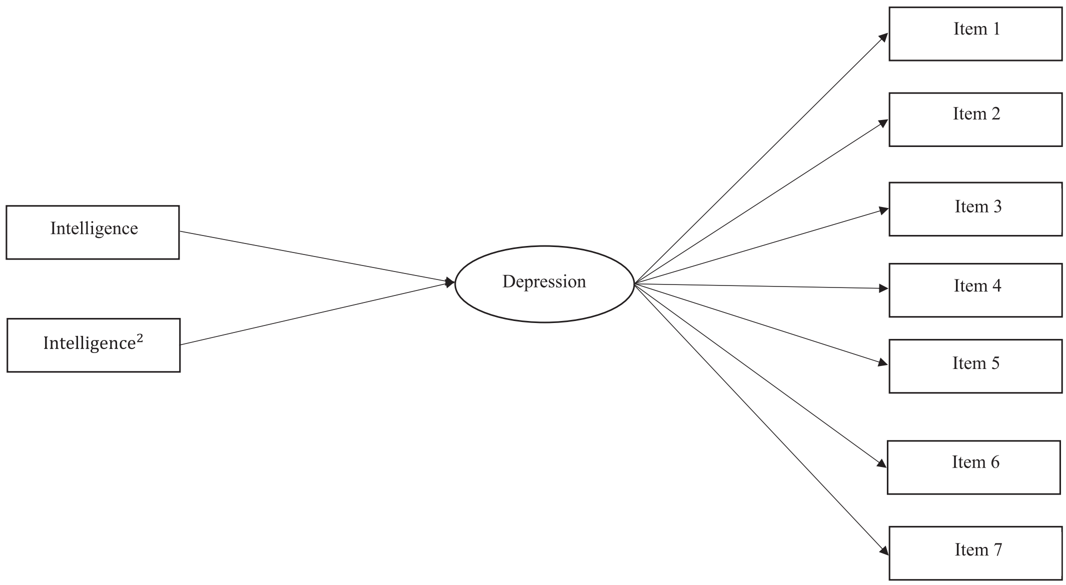

Fig. 2. Path diagram for the cross-sectional model (NLSY79 sample).

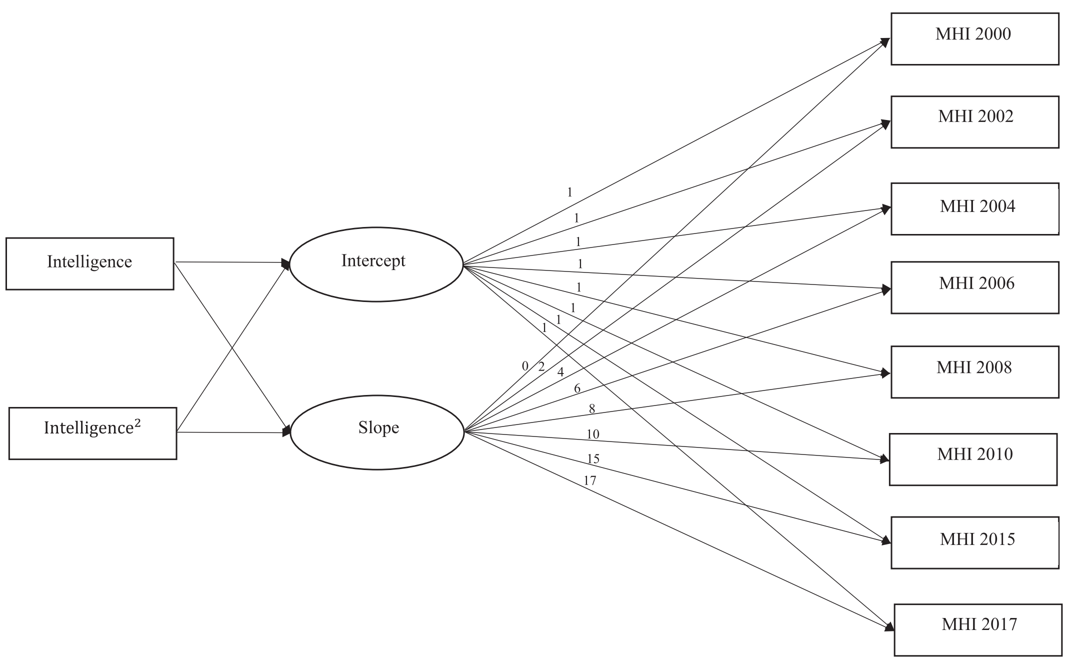

Fig. 3. Path diagram for the longitudinal model with loading constraints on latent variables (NLSY97 sample).

function of a moderator. At each focal point of the moderator, SEM parameters are estimated using a locally weighted subset of the data using a Gaussian kernel function. LSEM can be used to test a nonlinear relationship by exploring how a latent mean changes across moderator values (the moderator in LSEM terms would be the independent variable in polynomial regression terms). Moreover, it can also be used to test for measurement invariance across a continuous variable. Similarly to multigroup factor analysis, by exploring whether relevant model parameters change across the moderator, configural (equality of the measurement model), metric (equality of factor loadings), and scalar invariance (equality of item intercepts) can be established (Hildebrandt et al., 2016; Putnick & Bornstein, 2016).

In the present study, for the NLSY79 sample, the measurement model of CES-D 7 was tested with intelligence as a moderator. For the NLSY97 sample, before testing the LGCM model with intelligence as a moderator, the measurement model for MHI-5 with intelligence as a moderator was tested to check whether the mean values of measured variables are comparable across intelligence values. As MHI-5 showed scalar invariance across time, only the measurement from the earliest wave used (year 2000) was tested this way.

The focal points for intelligence were set every 0.5 points from −2.5 to 2.5. The moderator values range had to be restricted at extreme values for the models to converge. The bandwidth parameter was set to $h = 2$ based on recommendations by Hildebrandt et al. (2016). The MLR estimator was used. To test whether the changes in model parameters across moderator values were statistically significant, the permutation test was performed, with the number of permutations set to 1000 (Good, 2005).

Within the LSEM framework, the joint estimation approach is also available (Robitzsch, 2023). This method has the advantage of

Table 1
Model fit indices for each cross-sectional model using polynomial regression with test results for nested model differences.

Model	$\chi 2$	*df*	CFI	RMSEA with 90 % CI	SRMR	Satorra-Bentler $\Delta\chi^2$	
						$\Delta\chi^2$	*p*
Men only	386.89	26	0.923	0.058 [0.053; 0.062]	0.040	–	–
Women only	488.46	26	0.924	0.064 [0.060; 0.069]	0.041	–	–
Baseline	873.51	52	0.924	0.061 [0.058; 0.064]	0.037	–	–
Linear coefficient constrained to be equal	878.14	53	0.923	0.061 [0.057; 0.064]	0.037	2.85	0.091
Quadratic coefficient constrained to be equal	877.59	53	0.923	0.061 [0.057; 0.064]	0.037	1.47	0.235

computing global model fit estimates and supports traditional measurement invariance testing, where parameters are progressively constrained to equality and changes in model fit are examined (Putnick & Bornstein, 2016). However, at this point in time, the joint estimation approach does not allow for the usage of the MLR estimator. For this reason, the analyses with the joint estimation approach were treated as auxiliary analyses and are only presented in the supplementary material.

3. Results

3.1. The cross-sectional model using polynomial and piecewise regression

Table 1 presents model fit statistics for the models tested in the NLSY79 sample. Table 2 presents standardized coefficients from the baseline models for men and women, the explained variance, as well as estimated inflection points for the nonlinear relationships. The results of the model tested without grouping the sample by gender are presented in the Supplementary material.

The models fit the data well. The relationship between intelligence and depression was U-shaped for both men and women. The regression coefficients did not differ across genders. The results of the models with piecewise regression confirm that for both men and women, the sign of the relationship changes after the breakpoint (from positive to negative).

3.2. The longitudinal model using polynomial and piecewise regression

Table 3 presents model fit statistics for the models tested in the NLSY97 sample. Table 4 presents standardized coefficients from the baseline models for men and women, the explained variance, as well as estimated inflection points for the nonlinear relationships. The visualization of the estimated changes in mental health from the LGCM models is presented in Fig. 4. The zero-order correlation coefficients between all variables included in the analyses, as well as the results of the model tested without grouping the sample by gender, are presented in the Supplementary material.

For the distress subscale of MHI-5, the models had acceptable values of RMSEA and SRMR, but the CFI was slightly below the cut-off point. However, testing a more complex model with the change over time being set to be quadratic resulted in non-convergence of the iterative algorithm. Moreover, evidence suggests that the more complex the tested model is, the CFI values tend to perform worse as model fit indicators (Niemand & Mai, 2018). For this reason and taking into account the fact that the CFI values were only slightly off, the model was considered to be acceptable. For the well-being subscale, the models fit the data well.

The slope values for distress show that, on average, distress decreased by 0.05 points per time point for men and by 0.07 points for women. The relationship between intelligence and the intercept of distress was U-shaped for both men and women. The linear regression coefficient was higher for women, and the quadratic coefficients did not differ between men and women. Despite that, the results of the models with piecewise regression show that for women, the sign of the relationship changes after the breakpoint, while for men, the relationship between intelligence and distress becomes non-significant after the breakpoint. The relationship between intelligence and the slope of distress was linearly positive for both men and women. This means that the higher the intelligence values, the more distress decreased over time. The regression coefficients did not differ between genders.

The slope values for well-being show that, on average, well-being increased by 0.01 points per time point for men and 0.02 points for women. The relationship between intelligence and the intercept of well-being was not significant. The relationship between intelligence and the slope of well-being was linearly negative for men; however, the regression coefficient did not differ across genders. This means that the higher the intelligence values, the less well-being increased over time.

Table 2
Results of the cross-sectional models using polynomial and piecewise regression (standardized regression coefficients with 95 % confidence intervals, explained variance and estimated inflection points).

	NLSY79 - Depression			
	Men		Women	
	β	*p*	*β*	*p*
Intelligence -> mental health	−0.16** [−0.19; −0.13]	<0.001	−0.18** [−0.21; −0.14]	<0.001
Intelligence^2 -> mental health	0.04** [0.02; 0.06]	0.001	0.05** [0.03; 0.08]	<0.001
Intelligence below the breakpoint -> mental health	−0.20** [−0.23; −0.16]	<0.001	−0.22** [−0.25; −0.18]	<0.001
Intelligence above the breakpoint -> mental health	0.03** [0.01; 0.06]	0.005	0.05** [0.03; 0.08]	<0.001
Total R^2 for mental health	0.038		0.043	
ΔR^2 (linear model versus quadratic model) for mental health	0.003		0.004	
ΔR^2_{mo} for mental health	0.079		0.093	
Inflection point / breakpoint[a] (in standard deviation / IQ score)	2.10 / 132		1.70 / 126	

[a] The estimated inflection points from the model with polynomial regression served as the breakpoints for the models with piecewise regression.
** $p < .01$.

Table 3
Model fit indices for each longitudinal model using polynomial regression with test results for nested model differences.

Model	$\chi2$	df	CFI	RMSEA with 90 % CI	SRMR	Satorra-Bentler $\Delta\chi^2$	
						$\Delta\chi^2$	p
Distress							
Men only	476.81	43	0.886	0.053 [0.050; 0.057]	0.053	–	–
Women only	524.12	43	0.899	0.057 [0.053; 0.060]	0.057	–	–
Baseline	999.69	86	0.893	0.055 [0.052; 0.058]	0.055	–	–
Linear coefficient constrained to be equal (intercept)	1013.88	87	0.892	0.055 [0.052; 0.058]	0.056	15.28**	<0.000
Quadratic coefficient constrained to be equal (intercept)	1002.61	87	0.893	0.055 [0.052; 0.057]	0.055	0.70	0.401
Linear coefficient constrained to be equal (slope)	1002.94	87	0.893	0.055 [0.052; 0.057]	0.055	0.57	0.449
Quadratic coefficient constrained to be equal (slope)	1001.24	87	0.893	0.055 [0.052; 0.057]	0.055	0.15	0.703
Well-Being							
Men only	351.96	43	0.924	0.045 [0.041; 0.049]	0.043	–	–
Women only	326.89	43	0.932	0.044 [0.039; 0.048]	0.039	–	–
Baseline	679.04	86	0.928	0.044 [0.041; 0.047]	0.041	–	–
Linear coefficient constrained to be equal (intercept)	680.48	87	0.928	0.044 [0.041; 0.047]	0.041	0.60	0.605
Quadratic coefficient constrained to be equal (intercept)	683.44	87	0.928	0.044 [0.041; 0.047]	0.042	4.08*	0.043
Linear coefficient constrained to be equal (slope)	683.16	87	0.928	0.044 [0.041; 0.047]	0.041	3.40	0.065
Quadratic coefficient constrained to be equal (slope)	683.95	87	0.928	0.044 [0.041; 0.047]	0.041	4.84*	0.028

* $p < .05$.
** $p < .01$.

Table 4
Results of the longitudinal models using polynomial and piecewise regresion (standardized regression coefficients with 95 % confidence intervals, explained variance and estimated inflection points).

	Distress				Well-Being			
	Men		Women		Men		Women	
	β	p	β	p	β	p	β	p
Mental health <-> rate of change in mental health	−0.56** [−0.62; −0.50]	<0.001	−0.50** [−0.56; −0.44]	<0.001	−0.37** [−0.44; −0.30]	<0.001	−0.43** [−0.49; −0.36]	<0.001
Intelligence -> mental health	−0.10** [−0.14; −0.06]	<0.001	−0.21** [−0.25; −17]	<0.001	−0.04 [−0.08; 0.00]	0.073	−0.02 [−0.07; 0.02]	0.315
Intelligence^2 -> mental health	0.05** [0.02; 0.08]	0.001	0.07** [0.04; 0.10]	<0.001	−0.03 [−0.06; 0.00]	0.079	0.02 [−0.02; 0.06]	0.257
Intelligence -> rate of change in mental health	0.12** [0.07; 0.18]	<0.001	0.16** [0.10.; 0.22]	<0.001	−0.14** [−0.20; −0.08]	<0.001	−0.06 [−0.13; 0.00]	0.056
Intelligence^2 -> rate of change in mental health	−0.03 [−0.07; 0.01]	0.135	−0.02 [−0.07; 0.03]	0.431	0.05 [0.00; 0.09]	0.064	−0.03 [−0.09; 0.02]	0.204
Intelligence below the breakpoint -> mental health	−0.14** [−0.20; −0.09]	<0.001	−0.24** [−0.28; −0.19]	<0.001	–	–	–	–
Intelligence above the breakpoint -> mental health	0.04 [0.00; 0.08]	0.063	0.05** [0.10; 0.22]	0.002	–	–	–	–
Intelligence below the breakpoint -> rate of change in mental health	–	–	–	–	–	–	–	–
Intelligence above the breakpoint -> rate of change in mental health	–	–	–	–	–	–	–	–
Intercept value	4.88		5.37		5.53		5.18	
Slope values	−0.05		−0.07		0.01		0.02	
Total R^2 for mental health	0.019		0.054		0.003		0.001	
ΔR^2 (linear model versus quadratic model) for mental health	0.005		0.008		0.002		0.000	
ΔR^2_{mo} for mental health	0.263		0.148		0.666		0.000	
Total R^2 for the rate of change in mental health	0.020		0.026		0.029		0.005	
ΔR^2 (linear model versus quadratic model) for the rate of change in mental health	0.002		0.003		0.004		0.000	
ΔR^2_{mo} for the rate of change in mental health	0.100		0.115		0.138		0.000	
Inflection point/breakpoint[a] for the relationship with the mental health intercept (in standard deviation / IQ score)	1.00 / 115		1.53 / 123		–		–	
Inflection point/breakpoint[a] for the relationship with the rate of change (in standard deviation / IQ score)	–		–		–		–	

[a] The estimated inflection points from the model with polynomial regression served as the breakpoints for the models with piecewise regression.
** $p < .01$.

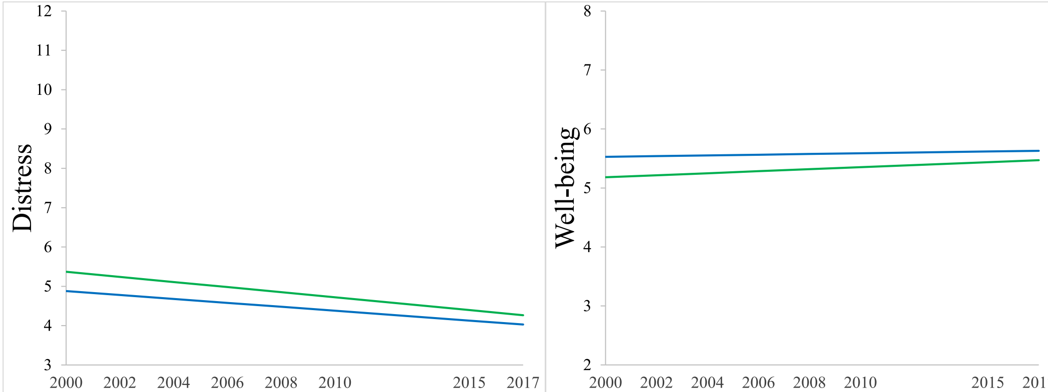

Fig. 4. Visualization of the estimated changes in mental health from the longitudinal models (blue lines = men, green lines = women).
Note. The values on the y-axis represent raw scores (sums of the items) on the MHI-5 inventory. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

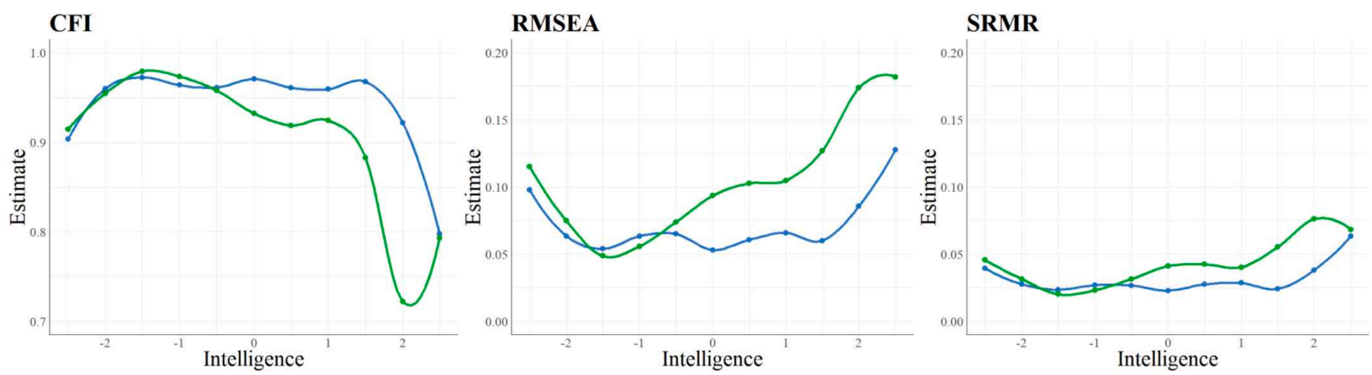

Fig. 5. Model fit indices of CES-D 7 across intelligence levels in the NLSY79 sample (blue lines = men, green lines = women). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Results of permutation tests for the moderating effect of intelligence on unstandardized factor loadings and item intercepts in CES-D 7 measurement model using the NLSY79 data.

Item	Loadings						Intercepts					
	Men			Women			Men			Women		
	M	*SD*	*p(SD)*	*M*	*SD*	*p(SD)*	*M*	*SD*	*p(SD)*	*M*	*SD*	*p(SD)*
I did not feel like eating; my appetite was poor	0.312	0.091**	<0.001	0.388	0.055**	0.009	0.288	0.133**	<0.001	0.490	0.133**	<0.001
I had trouble keeping my mind on what I was doing	0.457	0.068**	<0.001	0.526	0.037	0.184	0.596	0.053**	<0.001	0.706	0.053**	<0.001
I felt depressed	0.571	0.101**	<0.001	0.678	0.096**	<0.001	0.396	0.149**	<0.001	0.555	0.149**	<0.001
I felt that everything I did was an effort	0.372	0.021	0.855	0.456	0.059**	0.004	1.027	0.277**	<0.001	0.945	0.277**	<0.001
My sleep was restless	0.457	0.072**	<0.001	0.520	0.080**	<0.001	0.657	0.060**	<0.001	0.819	0.060**	<0.001
I felt sad	0.486	0.055**	0.005	0.596	0.062**	0.001	0.347	0.085**	<0.001	0.513	0.085**	<0.001
I could not get “going”	0.409	0.020	0.882	0.454	0.045*	0.035	0.422	0.049**	<0.001	0.571	0.049**	<0.001

M = weighted average of a given parameter; *SD* = test statistic of permutation test; *p(SD)* = *p* value of the permutation test.
* $p < .05$.
** $p < .01$.

The squared regression coefficients for both the intercept of well-being and the slope of well-being differed between men and women despite the coefficients being non-significant. However, they did differ in signs.

3.3. LSEM

In the NLSY79 sample, the weighted sample sizes ranged from n_{eff} = 85.40 to n_{eff} = 1906.31 for men, and from n_{eff} = 59.92 to n_{eff} = 1928.97 for women (see supplementary material for exact n_{eff} at each focal point). For men the average CFI value was *M* = 0.962 (*SD* = 0.017, Min = 0.798, Max = 0.972), the average RMSEA was *M* = 0.062 (*SD* = 0.009, Min = 0.053, Max = 0.128), and the average SRMR was *M* = 0.027 (*SD* = 0.004, Min = 0.023, Max = 0.063). For women the average CFI value was *M* = 0.941 (*SD* = 0.037, Min = 0.722, Max = 0.979), the average RMSEA was *M* = 0.084 (*SD* = 0.025, Min = 0.049, Max = 0.182), and the average SRMR was *M* = 0.035 (*SD* = 0.009, Min = 0.020, Max = 0.076). For men, there was a sharp decrease in model fit at high values of intelligence, while for women there was a steadier decline with a sharper decrease at high values of intelligence (see Fig. 5).

Table 5 presents the results of permutation tests for the moderating

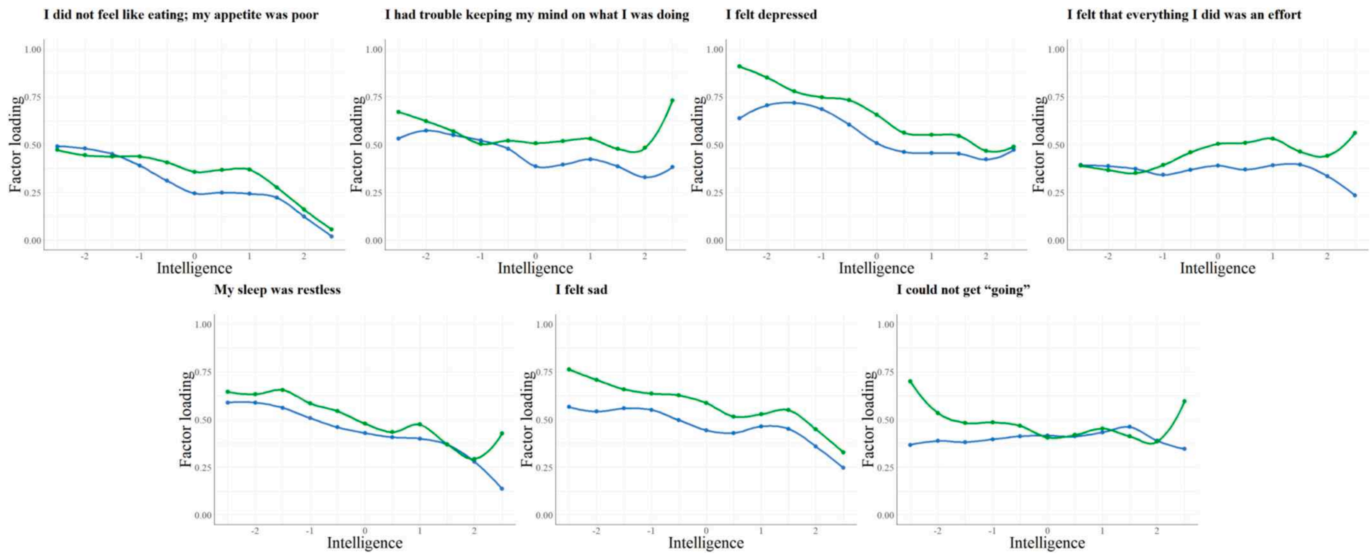

Fig. 6. Unstandardized factor loadings of CES-D 7 across intelligence levels in the NLSY79 sample (blue lines = men, green lines = women). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

effect of intelligence on the unstandardized factor loadings and intercepts of the CES—D. For men, there were significant changes in the loadings of five items, and for women, in the loadings of six items. The general trend was that factor loadings decreased as intelligence values increased. (see Fig. 6). Therefore, there was no metric invariance across intelligence values. The results of the joint estimation approach also supported the lack of metric invariance across intelligence for both men and women (see supplementary material). This means that values of depression are not comparable at different levels of intelligence. However, for information and comparison (with different analytical approaches and other studies) purposes, the change of the latent mean of depression across intelligence is still presented in the supplementary material.

For the NLSY97 sample, the presented results refer to the MHI-5 measurement taken in 2000. The weighted sample sizes ranged from $n_{eff} = 77.85$ to $n_{eff} = 1250.76$ for men, and from $n_{eff} = 56.70$ to $n_{eff} = 1242.43$ for women (see supplementary material for exact n_{eff} at each focal point). For men the average CFI value was $M = 0.958$ ($SD = 0.011$, Min = 0.942, Max = 0.978), the average RMSEA was $M = 0.112$ ($SD = 0.018$, Min = 0.074, Max = 0.132), and the average SRMR was $M = 0.033$ ($SD = 0.006$, Min = 0.020, Max = 0.041). For women the average CFI value was $M = 0.974$ ($SD = 0.010$, Min = 0.894, Max = 0.984), the average RMSEA was $M = 0.093$ ($SD = 0.015$, Min = 0.067, Max = 0.150), and the average SRMR was $M = 0.026$ ($SD = 0.004$, Min = 0.019, Max = 0.044). Overall, the model fit the data at all levels of intelligence for both men and women (see Fig. 7). Although RMSEA values were above the commonly recommended cut-off point, simulation studies show that for less complex models with few degrees of freedom, it often falsely indicates a poor fitting model (Kenny et al., 2015).

Table 6 shows the results of permutation tests for the moderating effect of intelligence on unstandardized factor loadings and intercepts of MHI-5 using the measurement from the year 2000. For both men and women, there were significant changes in the loadings of three items. The general trend was that factor loadings decreased as intelligence values increased. (see Fig. 8). Therefore, there was no metric invariance across intelligence values. The results of the joint estimation approach also supported the lack of metric invariance across intelligence for both men and women (see supplementary material). This means that values of distress and well-being are not comparable at different levels of intelligence. However, for information and comparison (with different analytical approaches and other studies) purposes, the change of latent means of distress and well-being across intelligence is still presented in the supplementary material.

With no measurement invariance of MHI-5 across intelligence, the longitudinal LSEM models were not performed, as any results from those could not be distinguished from statistical artifacts. The results of LSEM models tested without grouping the sample by gender are available in the supplementary material.

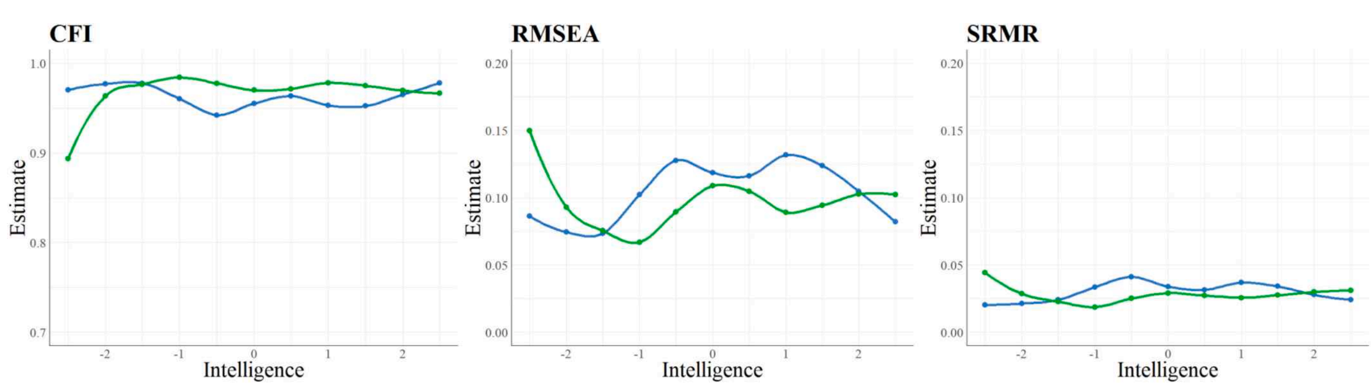

Fig. 7. Model fit indices of MHI-5 across intelligence levels in the NLSY97 sample (blue lines = men, green lines = women). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 6
Results of permutation tests for the moderating effect of intelligence on unstandardized factor loadings and item intercepts in MHI-5 measurement model using the NLSY97 data from the year 2000.

	Item	Loadings						Intercepts					
		Men			Women			Men			Women		
		M	*SD*	*p(SD)*	*M*	*SD*	*p(SD)*	*M*	*SD*	*p(SD)*	*M*	*SD*	*p(SD)*
Distress subscale	How much of the time during the last month have you been a very nervous person?	0.347	0.032	0.261	0.346	0.033	0.101	1.669	0.031**	<0.001	1.900	0.133**	<0.001
	How much of the time during the last month have you felt downhearted and blue?	0.538	0.062**	<0.001	0.553	0.071**	<0.001	1.733	0.025**	<0.001	2.002	0.053**	<0.001
	How much of the time during the last month have you felt so down in the dumps that nothing could cheer you up?	0.418	0.047*	0.010	0.428	0.047**	<0.001	1.396	0.104**	<0.001	1.489	0.149**	<0.001
Well-being subscale	How much of the time during the last month have you felt calm and peaceful?	0.468	0.039	0.058	0.450	0.030	0.120	2.253	0.038**	<0.001	2.492	0.277**	<0.001
	How much of the time during the last month have you been a happy person?	0.561	0.050**	0.009	0.555	0.062**	0.002	2.121	0.029**	<0.001	2.264	0.049**	<0.001

M = weighted average of a given parameter; *SD* = test statistic of permutation test; *p(SD)* = *p* value of the permutation test.
* $p < .05$.
** $p < .01$.

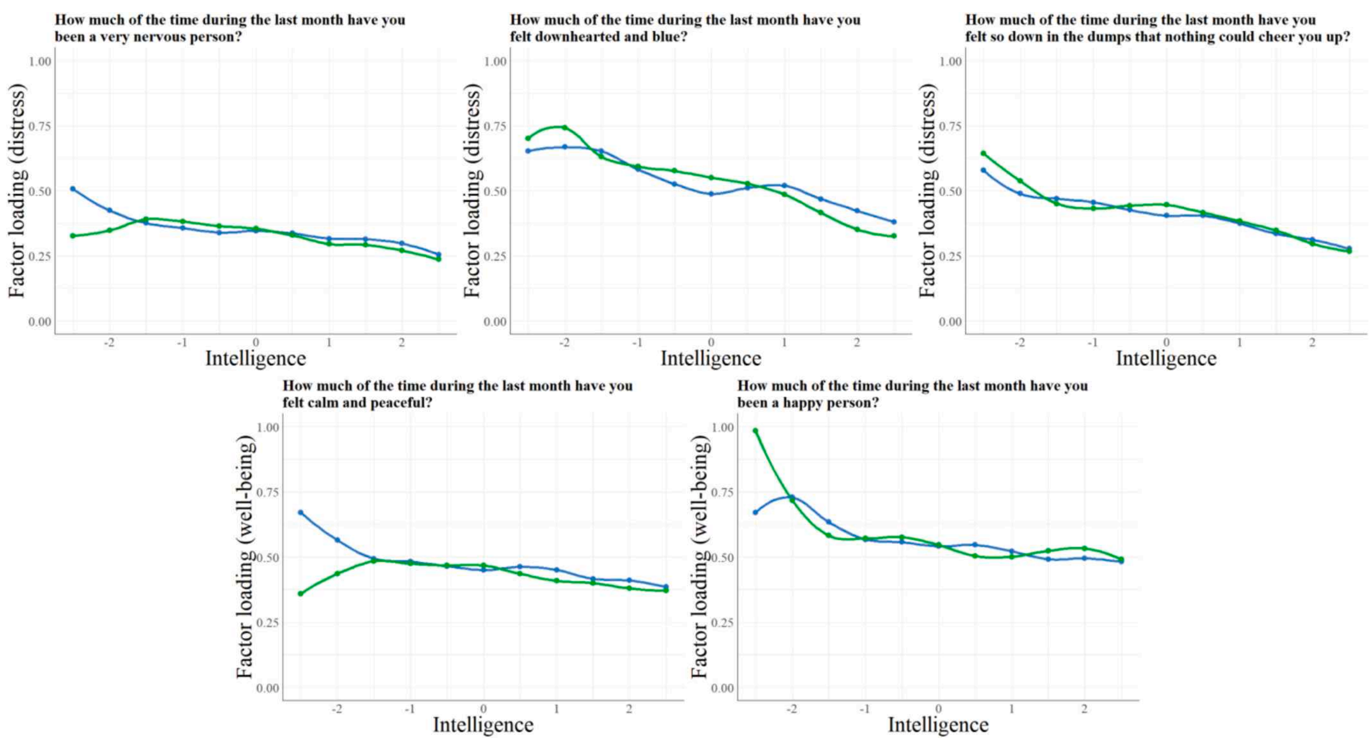

Fig. 8. Unstandardized factor loadings of MHI-5 across intelligence levels in the NLSY97 sample (blue lines = men, green lines = women). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4. Discussion

The main aim of this study was to investigate whether the relationship between intelligence and mental health is nonlinear. Although the results from polynomial and piecewise regression suggested that the relationship of intelligence with depression and distress is nonlinear, LSEM models provided evidence of the lack of measurement invariance in men and women for both CES-D and MHI-5. The higher the intelligence level of the individual answering the items, the more the items on both instruments seem to lose validity (lower factor loadings) as mental health indicators. Therefore, any result showing nonlinearity of the relationship between intelligence and mental health, or any results showing a lack thereof, cannot be considered reliable evidence when using these measures. Consequently, the study's hypothesis could not be addressed. It is possible that the inconsistencies in the results of previous studies on this subject are a manifestation of this effect and that other measures of mental health also lack measurement invariance across intelligence. In line with this argument, Czerwiński et al. (2025) had already suggested that the presence of nonlinearity in the relationship between intelligence and mental health might be tied to the scale used to measure mental health.

The presented results corroborate with studies reporting problems with clinical diagnosis of mental disorders in individuals with high intelligence (Bishop & Rinn, 2019; Foley-Nicpon et al., 2011). Amend and Beljan (2009) argued that some behaviors that are treated as symptoms of psychopathology, such as hyperfixation, are simply regular characteristics of gifted children. Studies also show that giftedness can mask disorders such as attention deficit–hyperactivity disorder (ADHD; Mullet & Rinn, 2015). Therefore, common indicators of psychopathology might not be appropriate for individuals with high intellectual ability. The deteriorating model fit of the CES-D 7 could even indicate that the structure of depression differs at higher levels of intelligence, but more research is needed to eliminate the possibility that this effect is exclusive to this specific instrument. Moreover, the sharp decrease in model fit indices might be simply the result of low weighted sample size at these focal points, and caution is advised when interpreting this specific result.

It is also possible that the revealed lack of measurement invariance comes from the interpretation of the wording of different items or even the instructions of questionnaires. The process of responding to a psychometric measure could be considered a cognitive task in its nature. Moreover, the autism spectrum disorder has a complex relationship with intelligence (Burger-Veltmeijer et al., 2011; Crespi, 2016) and is often characterized by some level of impairment of the semantic function of language (Boucher, 2003). However, the available evidence shows mixed results on measurement invariance of psychometric measures between individuals with autism spectrum disorder and those without it, with evidence for different scales showing support from strict invariance to complete lack of invariance (Breiner et al., 2025; Pelton et al., 2020; White et al., 2015).

The observed pattern of factor loadings decreasing with intelligence levels can also be somewhat reminiscent of the Spearman's law of diminishing returns (Spearman, 1927). Also known as the ability differentiation hypothesis, the effect of factor loadings of the general factor of intelligence decreasing with higher values of intelligence is well documented by empirical research (Blum & Holling, 2017; Breit et al., 2022). On the other hand, a study on the issue of the dimensionality of the Rosenberg Self-Esteem Scale showed that with higher cognitive abilities, the proportion of variance explained by the general factor of self-esteem increased and the scale became unidimensional (Gnambs & Schroeders, 2017). This was attributed to a better ability to understand negatively worded items. Overall, in conjunction with the results of the present study, these effects may suggest that individuals with varying levels of intelligence employ different strategies when answering self-reported psychometric questionnaires. Future studies should test the measurement invariance across intelligence of not only other mental health measures, but also scales measuring other psychological constructs.

4.1. Limitations and strengths of the study

The present study utilized two highly representative samples from large-scale cohort studies. The instruments used to measure mental health showed metric measurement invariance across genders and scalar across time. To the author's knowledge, this is the first study to attempt to test whether the nonlinear relationship between intelligence and mental health differs for men and women, as well as to incorporate longitudinal effects in an investigation of the nonlinearity of the relationship between intelligence and mental health. More importantly, to the author's knowledge, it is also the first study to test for measurement invariance of mental health measures across intelligence scores.

However, this study has some limitations. For one, the lack of measurement invariance of CES-D 7 and MHI-5 across intelligence meant the study's original aim could not be accomplished. As such, future studies should aim to develop a mental health scale that is invariant across intelligence and reexamine the nonlinearity of the relationship between intelligence and mental health, the rate of change in mental health, and the moderating role of gender. Alternatively physiological or behavioral indicators of mental health could be used. The intelligence test used was also taken under low-stakes conditions, meaning that the performance motivation of participants could affect their results (Kyllonen & Kell, 2018; Segal, 2012), although the size of this effect has been argued to be negligible (Bates & Gignac, 2022). The samples were also limited only to the United States. Gender roles are heavily influenced by culture and differ from country to country, which could have influenced the results. Furthermore, the experience of having high intelligence may vary across different cultures, influenced by disparities in education systems, values, or societal perceptions of highly intelligent individuals. As such, future research should test the nonlinearity of the relationship between intelligence and mental health and the role of gender in it across different cultures, especially in regions that are not Western, educated, industrialized, rich, and democratic (WEIRD; Henrich et al., 2010). The sample sizes utilized in this study would be considered large by psychological research standards. However, at the extreme values of intelligence, the weighted sample sizes in LSEM analyses were quite small, which could affect model estimates. Particularly, the sharp decrease of model fit indices of CES-D 7 at extreme values of intelligence could be caused by the smaller weighted sample sizes at those focal points. On the other hand, the general trends seen in factor loading decreases were relatively stable across intelligence levels, suggesting these effects are substantive and not a statistical artifact. Further studies on the topic are recommended, preferably with even larger samples.

5. Conclusions

In conclusion, the study's original aim of testing the nonlinearity of the relationship between intelligence and mental health could not be reliably accomplished. The tools used to measure mental health proved to lack measurement invariance across intelligence for both men and women. The factor loadings decreased with intelligence, meaning that the scores on those scales are not comparable between individuals at different levels of cognitive abilities. If this effect is not exclusive to CES-D 7 and MHI-5 used in the study and is present in other commonly used psychometric instruments this can have a substantial impact on psychological research. Lack of measurement invariance of psychometric instruments across intelligence can compromise the results of studies, not only testing for nonlinear relationships of intelligence, but also linear ones. Therefore it is crucial to test measurement invariance across intelligence of both measures of mental health and of other psychological constructs.

CRediT authorship contribution statement

Stanisław K. Czerwiński: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Roman Konarski:** Writing – review & editing, Supervision, Methodology, Formal analysis, Conceptualization. **Paweł A. Atroszko:** Writing – review & editing, Supervision, Project administration, Methodology, Conceptualization.

Ethical considerations

The study was carried out in accordance with the Declaration of Helsinki. All used data were anonymous and freely available online. No sensitive data were used.

Funding statement

This research was funded by a research grant from National Science Centre, Poland (2021/41/N/HS6/00821).

Declaration of competing interest

All authors declare that they have no conflict of interest regarding this manuscript.

Acknowledgements

Language editing was done with the help of ChatGPT, a language model developed by OpenAI in San Francisco, CA, USA.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.intell.2025.101963.

Data availability statement

The data used in the study is managed by the United States Bureau of Labor Statistics and is freely available at https://www.nlsinfo.org/investigator/.

References

Aguinis, H., Beaty, J. C., Boik, R. J., & Pierce, C. A. (2005). Effect size and power in assessing moderating effects of categorical variables using multiple regression: A 30-year review. *Journal of Applied Psychology, 90*(1), 94–107. https://doi.org/10.1037/0021-9010.90.1.94

Aichele, S., Ghisletta, P., Corley, J., Pattie, A., Taylor, A. M., Starr, J. M., & Deary, I. J. (2018). Fluid intelligence predicts change in depressive symptoms in later life: The Lothian Birth Cohort 1936. *Psychological Science, 29*(12), 1984–1995. https://doi.org/10.1177/0956797618804501

Altman, D. G., & Royston, P. (2006). The cost of dichotomising continuous variables. *BMJ, 332*(7549). https://doi.org/10.1136/bmj.332.7549.1080

Amend, E. R., & Beljan, P. (2009). The antecedents of misdiagnosis: When normal behaviors of gifted children are misinterpreted as pathological. *Gifted Education International, 25*(2), 131–143. https://doi.org/10.1177/026142940902500204

Asher, M., Asnaani, A., & Aderka, I. M. (2017). Gender differences in Social Anxiety Disorder: A review. *Clinical Psychology Review, 56*, 1–12. https://doi.org/10.1016/j.cpr.2017.05.004

Bandalos, D. L. (2014). Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(1), 102–116. https://doi.org/10.1080/10705511.2014.859510

Bates, T. C., & Gignac, G. E. (2022). Effort impacts IQ test scores in a minor way: A multi-study investigation with healthy adult volunteers. *Intelligence, 92*, Article 101652. https://doi.org/10.1016/j.intell.2022.101652

Baudson, T. G., & Preckel, F. (2013). Teachers' implicit personality theories about the gifted: An experimental approach. *School Psychology Quarterly, 28*(1), 37–46. https://doi.org/10.1037/spq0000011

Bayroff, A. G., & Anderson, A. A. (1963). *Development of armed forces qualification test 7 and 8*. U.S. Army Personnel Research Office.

Beesdo, K., Bittner, A., Pine, D. S., Stein, M. B., Höfler, M., Lieb, R., & Wittchen, H.-U. (2007). Incidence of social anxiety disorder and the consistent risk for secondary depression in the first three decades of life. *Archives of General Psychiatry, 64*(8), 903. https://doi.org/10.1001/archpsyc.64.8.903

Berwick, D. M., Murphy, J. M., Goldman, P. A., Ware, J. E., Barsky, A. J., & Weinstein, M. C. (1991). Performance of a five-item mental health screening test. *Medical Care, 29*(2), 169–176. https://doi.org/10.1097/00005650-199102000-00008

Billeiter, K. B., & Froiland, J. M. (2023). Diversity of intelligence is the norm within the autism spectrum: Full scale intelligence scores among children with ASD. *Child Psychiatry & Human Development, 54*, 1094–1101. https://doi.org/10.1007/s10578-021-01300-9

Bishop, J. C., & Rinn, A. N. (2019). The potential of misdiagnosis of high IQ youth by practicing mental health professionals: A mixed methods study. *High Ability Studies, 31*(2), 213–243. https://doi.org/10.1080/13598139.2019.1661223

Blum, D., & Holling, H. (2017). Spearman's law of diminishing returns. A meta-analysis. *Intelligence, 65*, 60–66. https://doi.org/10.1016/j.intell.2017.07.004

Boucher, J. (2003). Language development in autism. In *, Vol. 1254. International congress series* (pp. 247–253). Elsevier.

Breiner, C. E., McQuaid, G. A., Wallace, G. L., & Zickgraf, H. F. (2025). Screening for avoidant/restrictive food intake disorder symptoms among autistic adults: Measurement invariance with a comparison general sample. *Autism Research.* https://doi.org/10.1002/aur.70039

Breit, M., Brunner, M., Molenaar, D., & Preckel, F. (2022). Differentiation hypotheses of intelligence: A systematic review of the empirical evidence and an agenda for future research. *Psychological Bulletin, 148*(7–8), 518–554. https://doi.org/10.1037/bul0000379

Breit, M., Preuß, J., Scherrer, V., & Preckel, F. (2025). Why the use of segmented regression analysis to explore change in relations between variables is problematic: A simulation study. *Psychological Methods, 30*(3), 622–635. https://doi.org/10.1037/met0000576

Brown, M. I., Wai, J., & Chabris, C. F. (2021). Can you ever be too smart for your own good? Comparing linear and nonlinear effects of cognitive ability on life outcomes. *Perspectives on Psychological Science, 16*(6), 1337–1359. https://doi.org/10.1177/1745691620964122

Bureau of Labor Statistics, U.S. Department of Labor. (2021). *National Longitudinal Survey of youth 1997 cohort, 1979–2019 (rounds 1–19)* [Data set]. Ohio State University

Bureau of Labor Statistics, U.S. Department of Labor. (2023). *National Longitudinal Survey of youth 1979 cohort, 1979–2020 (rounds 1–29)* [Data set]. Ohio State University

Burger-Veltmeijer, A. E. J., Minnaert, A. E. M. G., Houten-Van, V., & den Bosch, E. J. (2011). The co-occurrence of intellectual giftedness and autism spectrum disorders. *Educational Research Review, 6*(1), 67–88. https://doi.org/10.1016/j.edurev.2010.10.001

Callahan, C. M., & Hébert, T. P. (2020). Gender issues. In J. Plucker, & C. Callahan (Eds.), *Critical issues and practices in gifted education* (pp. 213–228). Routledge.

Coakley, J. (1992). Burnout among adolescent athletes: A personal failure or social problem? *Sociology of Sport Journal, 9*(3), 271–285. https://doi.org/10.1123/ssj.9.3.271

Crespi, B. J. (2016). Autism as a disorder of high intelligence. *Frontiers in Neuroscience, 10*. https://doi.org/10.3389/fnins.2016.00300

Czerwiński, S. K., Atroszko, P. A., & Konarski, R. (2025). Mental health of intellectually gifted individuals: Investigating the nonlinearity of the relationship between intelligence and general mental health. *Health Psychology Report, 13*(1), 39–54. https://doi.org/10.5114/hpr/187337

Deary, I. J., Irwing, P., Der, G., & Bates, T. C. (2007). Brother–sister differences in the G factor in intelligence: Analysis of full, opposite-sex siblings from the NLSY1979. *Intelligence, 35*(5), 451–456. https://doi.org/10.1016/j.intell.2006.09.003

Defense Manpower Data Center. (2008). *ASVAB technical bulletin no. 3 CAT-ASVAB forms* (pp. 5–9).

Eagly, A. H., & Wood, W. (2011). Social role theory. In P. A. M. Van Lange, A. W. Kruglanski, & E. T. Higgins (Eds.)*, Vol. 2. Handbook of theories in social psychology* (pp. 458–476). Sage Publications Ltd.

Falck, S. (2020). *Extreme intelligence: Development, predicaments, implications*. Routledge, Taylor & Francis Group.

Flakus, M., Danieluk, B., Baran, L., Kwiatkowska, K., Rogoza, R., & Schermer, J. A. (2021). Are intelligent peers liked more? Assessing peer-reported liking through the network analysis. *Personality and Individual Differences, 177*, Article 110844. https://doi.org/10.1016/j.paid.2021.110844

Foley-Nicpon, M., Allmon, A., Sieck, B., & Stinson, R. D. (2011). Empirical investigation of twice-exceptionality: Where have we been and where are we going? *Gifted Child Quarterly, 55*(1), 3–17. https://doi.org/10.1177/0016986210382575

Freeman, J. (2006). Giftedness in the long term. *Journal for the Education of the Gifted, 29*(4), 384–403. https://doi.org/10.4219/jeg-2006-246

Freeman, J., & Garces-Basca, R. M. (2015). Gender differences in gifted children. In M. Neihart, S. I. Pfeiffer, & T. L. Cross (Eds.), *The social and emotional development of gifted children: What do we know?* Prufrock Press Inc.

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

Gale, C. R., Batty, G. D., Mcintosh, A. M., Porteous, D. J., Deary, I. J., & Rasmussen, F. (2012). Is bipolar disorder more common in highly intelligent people? A cohort study of a million men. *Molecular Psychiatry, 18*(2), 190–194. https://doi.org/10.1038/mp.2012.26

Gale, C. R., Batty, G. D., Tynelius, P., Deary, I. J., & Rasmussen, F. (2010). Intelligence in early adulthood and subsequent hospitalization for mental disorders. *Epidemiology, 21*(1), 70–77. https://doi.org/10.1097/ede.0b013e3181c17da8

Gignac, G. E., & Callis, Z. M. V. (2020). The costs of being exceptionally intelligent: Compatibility and interpersonal skill concerns. *Intelligence, 81*, Article 101465. https://doi.org/10.1016/j.intell.2020.101465

Gignac, G. E., Darbyshire, J., & Ooi, M. (2018). Some people are attracted sexually to intelligence: A psychometric evaluation of sapiosexuality. *Intelligence, 66*, 98–111. https://doi.org/10.1016/j.intell.2017.11.009

Gignac, G. E., & Starbuck, C. L. (2018). Exceptional intelligence and easygoingness may hurt your prospects: Threshold effects for rated mate characteristics. *British Journal of Psychology, 110*(1), 151–172. https://doi.org/10.1111/bjop.12342

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality and Individual Differences, 102*, 74–78. https://doi.org/10.1016/j.paid.2016.06.069

Gnambs, T., & Schroeders, U. (2017). Cognitive abilities explain wording effects in the Rosenberg self-esteem scale. *Assessment, 27*(2), 404–418. https://doi.org/10.1177/1073191117746503

Good, P. (2005). *Permutation, parametric, and bootstrap tests of hypotheses*. Springer.

Grant, A. M., & Schwartz, B. (2011). Too much of a Good thing: The challenge and opportunity of the inverted U. *Perspectives on Psychological Science, 6*(1), 61–76. https://doi.org/10.1177/1745691610393523

Halpern, D. F. (2011). *Sex differences in cognitive abilities*. Psychology Press.

Hatch, S. L., Jones, P. B., Kuh, D., Hardy, R., Wadsworth, M. E. J., & Richards, M. (2007). Childhood cognitive ability and adult mental health in the British 1946 birth cohort. *Social Science & Medicine, 64*(11), 2285–2296. https://doi.org/10.1016/j.socscimed.2007.02.027

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences, 33*(2–3), 61–83. https://doi.org/10.1017/s0140525x0999152x

Hildebrandt, A., Lüdtke, O., Robitzsch, A., Sommer, C., & Wilhelm, O. (2016). Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behavioral Research, 51*(2–3), 257–258. https://doi.org/10.1080/00273171.2016.1142856

Hildebrandt, A., Wilhelm, O., & Robitzsch, A. (2009). Complementary and competing factor analytic approaches for the investigation of measurement invariance. *Review of Psychology, 16*(2), 87–102.

Holahan, C. J., & Moos, R. H. (1981). Social support and psychological distress: A longitudinal analysis. *Journal of Abnormal Psychology, 90*(4), 365–370. https://doi.org/10.1037//0021-843x.90.4.365

Horwood, J., Salvi, G., Thomas, K., Duffy, L., Gunnell, D., Hollis, C., Lewis, G., Menezes, P., Thompson, A., Wolke, D., Zammit, S., & Harrison, G. (2008). IQ and non-clinical psychotic symptoms in 12-year-olds: Results from the ALSPAC birth cohort. *British Journal of Psychiatry, 193*(3), 185–191. https://doi.org/10.1192/bjp.bp.108.051904

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hu, S., Cai, D., Zhang, X. C., & Margraf, J. (2020). Relationship between social support and positive mental health: A three-wave longitudinal study on college students. *Current Psychology, 41*(10), 6712–6721. https://doi.org/10.1007/s12144-020-01175-4

Iribarren, C., Sharp, D., Burchfiel, C. M., Sun, P., & Dwyer, J. H. (1996). Association of serum total cholesterol with coronary disease and all-cause mortality: Multivariate correction for bias due to measurement error. *American Journal of Epidemiology, 143*(5), 463–471. https://doi.org/10.1093/oxfordjournals.aje.a008766

Jackson, P. S. (1998). Bright star — Black sky a phenomenological study of depression as a window into the psyche of the gifted adolescent. *Roeper Review, 20*(3), 215–221. https://doi.org/10.1080/02783199809553894

Jackson, P. S., & Peterson, J. (2003). Depressive disorder in highly gifted adolescents. *Journal of Secondary Gifted Education, 14*(3), 175–186. https://doi.org/10.4219/jsge-2003-429

Johnson, D. P., & Whisman, M. A. (2013). Gender differences in rumination: A meta-analysis. *Personality and Individual Differences, 55*(4), 367–374. https://doi.org/10.1016/j.paid.2013.03.019

Kanazawa, S. (2014). Intelligence and childlessness. *Social Science Research, 48*, 157–170. https://doi.org/10.1016/j.ssresearch.2014.06.003

Karpinski, R. I., Kolb, A. M. K., Tetreault, N. A., & Borowski, T. B. (2018). High intelligence: A risk factor for psychological and physiological overexcitabilities. *Intelligence, 66*, 8–23. https://doi.org/10.1016/j.intell.2017.09.001

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research, 44*(3), 486–507. https://doi.org/10.1177/0049124114543236

Keyes, K. M., Platt, J., Kaufman, A. S., & McLaughlin, K. A. (2017). Association of Fluid Intelligence and psychiatric disorders in a population-representative sample of US adolescents. *JAMA Psychiatry, 74*(2), 179. https://doi.org/10.1001/jamapsychiatry.2016.3723

Kyllonen, P., & Kell, H. (2018). Ability tests measure personality, personality tests measure ability: Disentangling construct and method in evaluating the relationship between personality and ability. *Journal of Intelligence, 6*(3), 32. https://doi.org/10.3390/jintelligence6030032

Lavrijsen, J., & Verschueren, K. (2023). High cognitive ability and mental health: Findings from a large community sample of adolescents. *Journal of Intelligence, 11*(2), 38. https://doi.org/10.3390/jintelligence11020038

Lee, S. L., Pearce, E., Ajnakina, O., Johnson, S., Lewis, G., Mann, F., … Lewis, G. (2021). The association between loneliness and depressive symptoms among adults aged 50 years and older: A 12-year population-based Cohort Study. *The Lancet Psychiatry, 8*(1), 48–57. https://doi.org/10.1016/s2215-0366(20)30383-7

Liu, H., & Yuan, K.-H. (2021). New measures of effect size in moderation analysis. *Psychological Methods, 26*(6), 680–700. https://doi.org/10.1037/met0000371

Lubinski, D., & Benbow, C. P. (2006). Study of mathematically precocious youth after 35 years: Uncovering antecedents for the development of math-science expertise. *Perspectives on Psychological Science, 1*(4), 316–345. https://doi.org/10.1111/j.1745-6916.2006.00019.x

Lubinski, D., Benbow, C. P., & Kell, H. J. (2014). Life paths and accomplishments of mathematically precocious males and females four decades later. *Psychological Science, 25*(12), 2217–2232. https://doi.org/10.1177/0956797614551371

Lubinski, D., Benbow, C. P., McCabe, K. O., & Bernstein, B. O. (2023). Composing meaningful lives: Exceptional women and men at age 50. *Gifted Child Quarterly, 67*(4), 278–305. https://doi.org/10.1177/00169862231175831

Major, J. T., Johnson, W., & Deary, I. J. (2014). Linear and nonlinear associations between general intelligence and personality in project talent. *Journal of Personality and Social Psychology, 106*(4), 638–654. https://doi.org/10.1037/a0035815

Martin, L. T., Burns, R. M., & Schonlau, M. (2010). Mental disorders among gifted and nongifted youth: A selected review of the epidemiologic literature. *Gifted Child Quarterly, 54*(1), 31–41. https://doi.org/10.1177/0016986209352684

McGrevy, D. F., Knouse, S. B., & Thompson, R. A. (1974). *Relationships among an individual intelligence test and two air force screening and selection tests (Tech. Rep.)*. Brooks Air Force Base, Tx: Air Force Human Resources Lab.

McLean, C. P., & Anderson, E. R. (2009). Brave men and timid women? A review of the gender differences in fear and anxiety. *Clinical Psychology Review, 29*(6), 496–505. https://doi.org/10.1016/j.cpr.2009.05.003

McZgee, V. E., & Carleton, W. T. (1970). Piecewise regression. *Journal of the American Statistical Association, 65*(331), 1109–1124.

Meybodi, F. A., Saeedi, Z., Behjati, Z., Noorbala, F., Dastbaravardec, A., & Enjedany, E. (2011). Reliability and validity of a Farsi version of 18- item mental health inventory. *Procedia - Social and Behavioral Sciences, 30*, 1425–1429. https://doi.org/10.1016/j.sbspro.2011.10.277

Mullet, D. R., & Rinn, A. N. (2015). Giftedness and ADHD: Identification, misdiagnosis, and dual diagnosis. *Roeper Review, 37*(4), 195–207. https://doi.org/10.1080/02783193.2015.1077910

Naggara, O., Raymond, J., Guilbert, F., Roy, D., Weill, A., & Altman, D. G. (2011). Analysis by categorizing or dichotomizing continuous variables is inadvisable: An example from the natural history of unruptured aneurysms. *American Journal of Neuroradiology, 32*(3), 437–440. https://doi.org/10.3174/ajnr.a2425

Nelson, L., & Simonsohn, U. (2014). Thirty-somethings are shrinking and other U-shaped challenges [Blog post]. Retrieved from http://datacolada.org/2014/09/17/27-thirty-somethings-are-shrinkingand-other-u-shaped-challenges/.

Nickel, L. B., Roberts, B. W., & Chernyshenko, O. S. (2019). No evidence of a curvilinear relation between conscientiousness and relationship, work, and health outcomes. *Journal of Personality and Social Psychology, 116*(2), 296–312. https://doi.org/10.1037/pspp0000176

Niemand, T., & Mai, R. (2018). Flexible cutoff values for fit indices in the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 46*(6), 1148–1172. https://doi.org/10.1007/s11747-018-0602-9

Nosek, B. A., Smyth, F. L., Sriram, N., Lindner, N. M., Devos, T., Ayala, A., & Greenwald, A. G. (2009). National differences in gender–science stereotypes predict national sex differences in science and math achievement. *Proceedings of the National Academy of Sciences, 106*(26), 10593–10597. https://doi.org/10.1073/pnas.0809921106

Pelton, M. K., Crawford, H., Robertson, A. E., Rodgers, J., Baron-Cohen, S., & Cassidy, S. (2020). A measurement invariance analysis of the Interpersonal Needs Questionnaire and acquired capability for suicide scale in autistic and non-autistic adults. *Autism in Adulthood, 2*(3), 193–203. https://doi.org/10.1089/aut.2019.0055

Pierce, J. R., & Aguinis, H. (2013). The too-much-of-a-good-thing effect in management. *Journal of Management, 39*(2), 313–338. https://doi.org/10.1177/0149206311410060

Purvanova, R. K., & Muros, J. P. (2010). Gender differences in burnout: A meta-analysis. *Journal of Vocational Behavior, 77*(2), 168–185. https://doi.org/10.1016/j.jvb.2010.04.006

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team. (2023). *R: A language and environment for statistical computing (version 4.3.0)*. [Computer software]. R Foundation for Statistical Computing.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*(3), 385–401. https://doi.org/10.1177/014662167700100306

Reis, S. M., & Hébert, T. P. (2008). Gender and Giftedness. In S. I. Pfeiffer (Ed.), *Handbook of giftedness in children*. Boston, MA: Springer. https://doi.org/10.1007/978-0-387-74401-8_14.

Rivera-Riquelme, M., Piqueras, J. A., & Cuijpers, P. (2019). The revised Mental Health Inventory-5 (MHI-5) as an ultra-brief screening measure of bidimensional mental health in children and adolescents. *Psychiatry Research, 274*, 247–253. https://doi.org/10.1016/j.psychres.2019.02.045

Robitzsch, A. (2023). Estimating local structural equation models. *Journal of Intelligence, 11*(9), 175. https://doi.org/10.3390/jintelligence11090175

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

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika, 75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Schredl, M., & Reinhard, I. (2011). Gender differences in nightmare frequency: A meta-analysis. *Sleep Medicine Reviews, 15*(2), 115–121. https://doi.org/10.1016/j.smrv.2010.06.002

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323–338. https://doi.org/10.3200/joer.99.6.323-338

Segal, C. (2012). Working when no one is watching: Motivation, test scores, and economic success. *Management Science, 58*(8), 1438–1457. https://doi.org/10.1287/mnsc.1110.1509

Simonsohn, U. (2018). Two lines: A valid alternative to the invalid testing of U-shaped relationships with quadratic regressions. *Advances in Methods and Practices in Psychological Science, 1*(4), 538–555. https://doi.org/10.1177/2515245918805755

Spearman, C. (1927). *The abilities of man*. MacMillan.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613–629. https://doi.org/10.1037//0003-066x.52.6.613

White, S. W., Lerner, M. D., McLeod, B. D., Wood, J. J., Ginsburg, G. S., Kerns, C., … Compton, S. (2015). Anxiety in youth with and without autism spectrum disorder: Examination of factorial equivalence. *Behavior Therapy, 46*(1), 40–53. https://doi.org/10.1016/j.beth.2014.05.005

Williams, C. M., Peyre, H., Labouret, G., Fassaya, J., Guzmán García, A., Gauvrit, N., & Ramus, F. (2022). High intelligence is not associated with a greater propensity for mental health disorders. *European Psychiatry, 66*(1). https://doi.org/10.1192/j.eurpsy.2022.2343

Zeidner, M. (2021). "Don't worry—Be happy": The sad state of happiness research in gifted students. *High Ability Studies, 32*(2), 125–142. https://doi.org/10.1080/13598139.2020.1733392