Statistical Theory — Final Project

Sharon Kamensky 211576723 and Roni Navon 207858937

Department of Statistics and Data Science

August 2025

Abstract

The aim of this study was to examine whether personality profiles explain variance in overall life happiness. Initially, intuitive profiles were constructed based on theoretical considerations, followed by empirical factors derived through exploratory factor analysis (EFA) with oblimin rotation. Subsequently, background variables were examined in relation to happiness, revealing that only education level was significantly associated. Accordingly, the contribution of the profiles block was compared to that of education, with education included as a control variable. Analyses were based on complete cases ($n = 1006$), using OLS regression with heteroskedasticity-consistent standard errors (HC3) and Benjamini–Hochberg false discovery rate correction. Results showed that the profiles block explained a small but statistically significant additional proportion of variance beyond education ($\Delta R^2 \approx 0.020$, $f^2 \approx 0.021$), with the *materialist_image* profile having the largest individual effect ($\beta \approx 0.092$, $p \approx 0.023$), although this did not remain significant after FDR correction. While effect sizes were small, the findings suggest that personality profiles provide incremental explanatory value for happiness beyond education. The complete project materials, including code and data processing scripts, are available on GitHub (click here).

1 Introduction

Happiness is a multifaceted construct shaped by socio-demographic conditions and personality traits. The present study aimed to examine whether personality profiles can explain variance in *Happiness in life* beyond the effect of education, integrating both theory-driven (intuitive) and empirically derived (EFA) profiles within a unified regression framework.

Several intuitive personality profiles ($n = 10$) were initially constructed based on theoretical considerations. Of these, only two met the internal consistency threshold (Cronbach's $\alpha \geq 0.60$) — *materialist image* and *artistic empath*. After testing their associations with overall life happiness, only *materialist image* showed a statistically significant correlation and was therefore retained for the regression analyses.

To broaden the scope of the analysis, an exploratory factor analysis (EFA) with oblimin rotation ($\delta = 0.6$) was conducted, yielding four factors (Factor1–Factor4) and one additional composite factor, STEM, representing a combined orientation toward science, technology, engineering, and mathematics. All four EFA-derived factors and the STEM composite were included in the regression models regardless of their individual correlations with *Happiness in life*, in order to evaluate their potential joint contribution in a multivariate context.

In the next stage, background variables such as education, age, gender, and number of siblings were examined for their associations with *Happiness in life*. Education emerged as the only statistically significant background variable, and was therefore included as a control variable in the main models. Robust inference was conducted using HC3 standard errors, with significance levels adjusted via the Benjamini–Hochberg false discovery rate (BH–FDR) procedure. Both individual coefficients and the joint contribution of the profiles block were tested, as well as potential *Profile* × *Education* interactions.

2 Methods

2.1 Sample, Data Source, and Variable Construction

Data were drawn from a large-scale questionnaire including dozens of items covering demographics, attitudes, and lifestyle preferences. The dependent variable (*Happiness in life*) was the total score for Happiness in life (1–5 scale; higher values indicate greater happiness). Analyses used a complete-case approach, excluding respondents with missing data on the dependent or predictor variables.

Education. Education was coded as a four-level categorical variable and retained as a covariate in all models, based on preliminary screening of background variables.

Intuitive profiles. Ten theory-driven (intuitive) personality profiles were initially constructed by selecting conceptually related questionnaire items (e.g., *social, spiritual, materialist, family-oriented*) and averaging the responses for each respondent to produce a continuous score per profile. Internal consistency was assessed using Cronbach's α ($\alpha \geq 0.60$ as the threshold). Two profiles met this criterion; of these, only *Materialist image* showed a statistically significant correlation with happiness and was retained for further analysis.

Empirical profiles (EFA). To complement the intuitive approach, an exploratory factor analysis (EFA) with oblimin rotation ($\delta = 0.6$) was conducted on the item pool. Four factors (Factor1–Factor4) met the reliability criterion ($\alpha \geq 0.60$) and were supplemented with an additional composite STEM profile, capturing interest in science, technology, engineering, and mathematics. Although only Factor1 correlated significantly with happiness in bivariate tests, all profiles were retained in the multivariate models. This decision was made because predictors that are non-significant individually may still explain variance jointly when modeled together, and excluding them prematurely could underestimate the block's potential explanatory power.

2.2 Handling Missing Data

Analyses were conducted on cases with complete data for all model variables, using listwise deletion. Reverse-coded items were recoded so that higher values consistently indicated stronger endorsement of the construct. Education was treated as a four-level categorical variable (*primary school*, *secondary school*, *college/bachelor degree*, *master's degree*) in the main regression models. For interaction analysis, a binary version was created, contrasting participants with a bachelor's degree or higher (coded 1) versus lower educational attainment (coded 0).

2.3 Scale Construction and Reliability

The *Materialist image* profile and the four EFA-derived STEM factors met the internal reliability threshold (Cronbach's $\alpha \geq 0.60$). Variance inflation factors (VIF) were examined to assess multicollinearity between predictors. Interactions with education were evaluated in a complementary model.

2.4 Preliminary Correlation Tests

Bivariate associations with *Happiness in life* were examined using Spearman's ρ and Kendall's τ. Ordinary least squares (OLS) regression with HC3 robust standard errors was used to estimate β coefficients and 95% confidence intervals. Within-family significance was adjusted using Benjamini–Hochberg false discovery rate (BH–FDR) corrections (applied separately to profile effects and interaction effects). Two-sided tests used $\alpha = 0.05$.

2.5 Background Variable Screening

Background variables (gender, age, residence type, housing type, number of siblings, education) were screened prior to regression. Binary variables were tested using two-sample t-tests, multi-category variables (education) using one-way ANOVA, and continuous variables using Pearson correlations. BH–FDR adjustment was applied within the background-variable family. Only education showed a statistically significant association with happiness and was retained as a covariate.

2.6 Primary Model

The main regression model was an OLS with HC3 robust standard errors, including the personality profiles block (*Materialist image*, Factor1, Factor2, Factor4, STEM) and education as a control variable. All continuous predictors were mean-centered prior to inclusion. A complementary model tested *Profile* × *Education* interactions.

2.7 Multiplicity Control

Significance was controlled separately for main effects and interactions using BH–FDR within each family. Main tests used $\alpha = 0.05$, while exploratory tests used $\alpha = 0.10$.

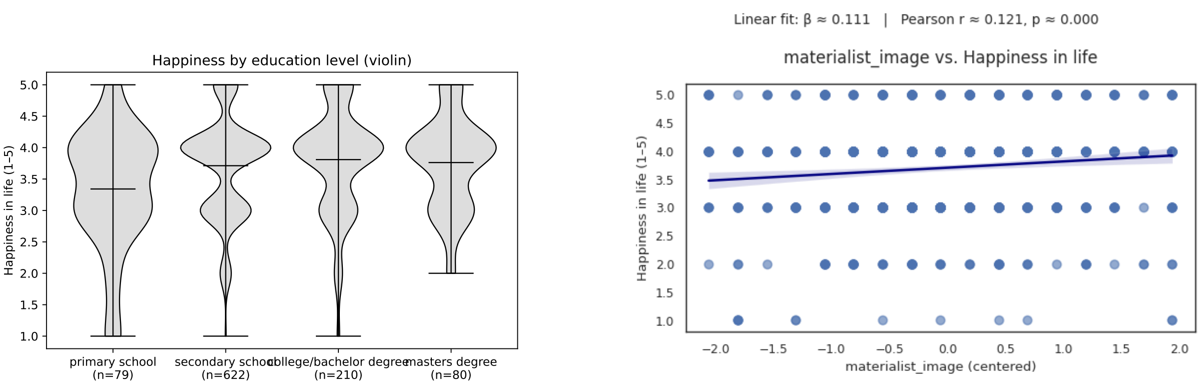
2.8 Model Diagnostics

Assumption checks confirmed that the regression models met standard statistical requirements. Specifically, the Breusch–Pagan test indicated no evidence of heteroskedasticity, the Anderson–Darling test supported the normality of residuals, and Cook's distance values suggested no influential outliers. Variance inflation factors (VIF ≤ 1.58) ruled out problematic multicollinearity among predictors. These diagnostics confirm the robustness and validity of the analytical framework.

3 Results

3.1 Descriptive exploration: education and profile effects

Figure 1 presents two complementary descriptive views. Panel **(a)** displays *Happiness in life* by education level (violin plot with means), highlighting a modest upward trend. Panel **(b)** shows the bivariate association between *materialist image* and *Happiness in life*, which exhibits a small but positive slope.



(a) *Happiness in life* by education level (violin; means shown).

(b) *Materialist image* vs *Happiness in life* (OLS fit).

Figure 1: Panel **(a)** shows *Happiness in life* by education level. Panel **(b)** shows the bivariate association between *materialist image* and *Happiness in life*.

3.2 Preliminary model: education only

The education-only model (`Education_collapsed`) yielded a statistically significant overall effect, $F(3, 997) = 14.31$, $p < 0.001$, with a small proportion of explained variance ($R^2 = 0.016$, $R^2_{\text{adj}} = 0.013$).

3.3 Extended model: education + personality profiles

In addition to education, five centered personality profile scores were entered: Factor1_c, Factor2_c, Factor4_c, STEM_c, and $\text{materialist\_image}_\text{c}$. The STEM score represents a composite orientation toward science, technology, engineering, and mathematics, derived from the EFA factors.

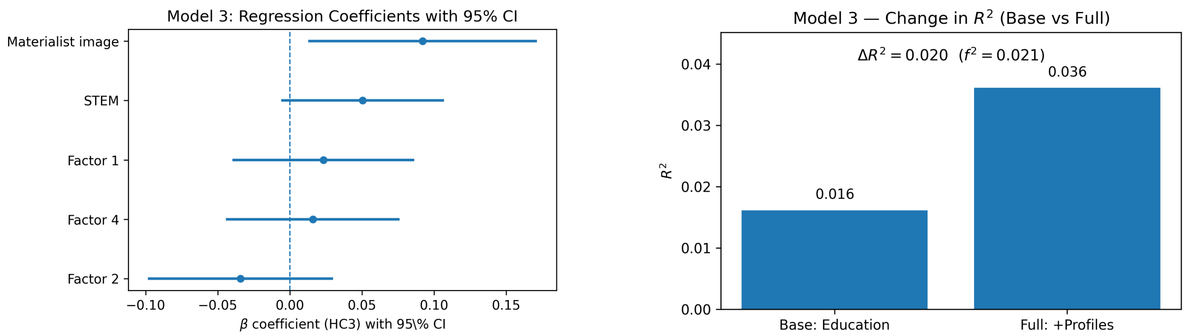
Adding these profiles produced a statistically significant block contribution, $F(5, 997) = 3.46$, $p = 0.004$, with model fit increasing to $R^2 = 0.036$ ($R^2_{\text{adj}} = 0.028$). This corresponds to $\Delta R^2 = 0.020$ ($\Delta R^2_{\text{adj}} = 0.015$) and a small effect size ($f^2 \approx 0.021$).

3.4 Model 3: coefficient estimates

Table 1 reports HC3 estimates; Figure 2a visualizes coefficients with 95% CIs. Only *materialist image* was individually significant before FDR correction; after Benjamini–Hochberg (BH) adjustment, no profiles remained significant ($q \approx 0.115$).

Table 1: Regression coefficients for Model 3 (Education + Personality Profiles; HC3 SEs). BH = Benjamini–Hochberg adjusted p.

Predictor	β	p	BH-adjusted p	95% CI
Materialist image	0.092	0.023	0.115	[0.013,0.171]
STEM	0.050	0.079	0.199	[-0.006,0.107]
Factor 2	-0.034	0.296	0.493	[-0.098,0.030]
Factor 1	0.023	0.469	0.586	[-0.040,0.086]
Factor 4	0.016	0.605	0.605	[-0.044,0.076]



(a) Standardized coefficients (β) with 95% CIs.

(b) Change in R^2: base vs. full model.

Figure 2: Panel **(a)** shows profile coefficients with 95% CIs; panel **(b)** shows the gain in explained variance when adding the profiles block ($\Delta R^2 \approx 0.020$, $f^2 \approx 0.021$), which represents a small effect size according to Cohen's (1988) guidelines.

Nevertheless, all profiles were retained in the model to allow a full evaluation of the profiles block's joint effect. This approach acknowledges that while individual predictors may lack statistical significance in isolation, their combined contribution can still yield a meaningful improvement in model fit, as reflected in the significant ΔR^2.

3.5 Multicollinearity check

Variance inflation factors indicated no concerning collinearity (all VIF ≤ 1.58). The strongest pairwise correlation was $|r| \approx 0.55$ (Factor1_c with $\text{materialist\_image}_c$). A compact visualization of the correlation matrix is provided in Figure 3.

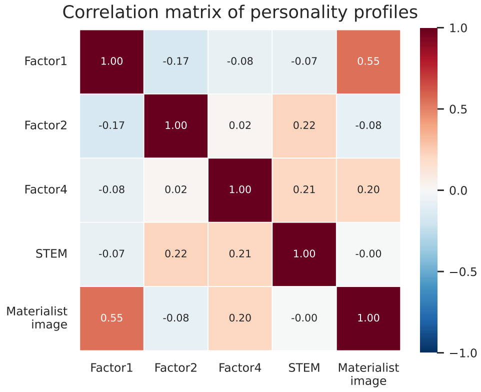


Figure 3: Correlation matrix among personality profiles (Pearson). Magnitudes are modest, consistent with low VIF values.

3.6 Summary of main findings

- Education is significantly associated with *Happiness in life* (small R^2).
- Adding profiles significantly improves fit ($\Delta R^2 \approx 0.020$; $f^2 \approx 0.021$, small effect).
- Only *materialist image* was individually significant before FDR; no profiles remained significant after BH correction ($q \approx 0.115$).

4 Discussion

The findings of this study suggest that personality profiles can provide a modest but meaningful contribution to explaining variance in overall happiness, even after accounting for a key background variable such as education. In particular, the *materialist image* profile showed a positive and statistically significant association with happiness (prior to FDR correction), while the remaining profiles did not exhibit individually significant effects. These results may indicate that the links between certain personality traits and subjective well-being are not always linear or consistent, and may depend on the social or economic contexts in which these traits are expressed. This is also why we chose to retain and examine the entire profiles block in the model, despite the lack of significance for most profiles in individual analyses, in order to assess their potential joint contribution in a multivariate context.

Several limitations should be acknowledged. First, the cross-sectional design precludes causal inference: it remains unclear whether personality traits influence happiness, whether happiness shapes these traits, or whether both are influenced by a third factor. Second, all measures were based on self-report, which may introduce biases related to perception, recall, or social desirability. Third, unmeasured confounding variables—such as health status, major life events, or other personality characteristics—may partially account for the observed associations. Future research could benefit from expanding the measurement framework to include multi-informant reports or behavioral measures, as well as longitudinal designs that allow examination of temporal dynamics and causal directions. Additionally, incorporating physiological indicators or large-scale behavioral data (e.g., from digital platforms) could deepen our understanding of how personality traits relate to different levels of happiness across diverse social and cultural contexts.

References

[DeNeve and Cooper, 1998] DeNeve, K. M. and Cooper, H. (1998). The happy personality: A meta-analysis of 137 personality traits and subjective well-being. *Psychological Bulletin*, 124(2):197–229.

[Sabo, 2018] Sabo, M. (2018). Young people survey. `https://www.kaggle.com/datasets/miroslavsabo/young-people-survey`. Accessed: 2025-08-14.