Unclassified **EDU/WKP(2016)23**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **27-Nov-2016**

English text only

DIRECTORATE FOR EDUCATION AND SKILLS

EDU/WKP(2016)23
Unclassified

MEASUREMENT PROPERTIES OF NON-COGNITIVE SCALES IN THE POLISH FOLLOW-UP STUDY ON PIAAC (POSTPIAAC)

OECD Education Working Paper No.149

By Marta Palczyńska and Karolina Świst
Instytut Badań Edukacyjnych, Educational Research Institute, Poland

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

Marta Palczyńska, m.palczynska@ibe.edu.pl
Karolina Świst, k.swist@ibe.edu.pl

JT03406292

English text only

OECD EDUCATION WORKING PAPERS SERIES

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

---

www.oecd.org/edu/workingpapers

ABSTRACT

There is a growing literature providing evidence that not only cognitive skills but also non-cognitive skills are important for economic and social outcomes. This paper assesses the measurement properties of the Big Five and Grit scales in a large representative sample of adults in Poland. The data from the Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) include longitudinal information on PIAAC respondents in Poland and additional background information not available in the international study. The results presented in this paper show that not all the criteria concerning the reliability, validity and comparability of these scales have been satisfied, though the personality measures significantly contribute to explaining the variability in policy-relevant outcomes. Most of the questions discriminate well between people possessing a high and a low level of a given trait, though reverse-worded items perform weaker. The Big Five theoretical five-factor structure was not replicated; however, a six-factor model with an additional factor loading reverse-worded items fits the data. In case of Grit, a bi-factor model, which has an equivalent interpretation to the second-order theoretical structure, holds. The scales are not fully measurement invariant. The results confirm earlier findings from the literature that differences in personality traits are clearly associated with differences in life outcomes. For most of the outcomes, the Big Five traits outperform cognitive skills in predictive power. Only educational attainment is more strongly related to cognitive skills, while for wages, the predictive power of personality and cognitive skills is similar. The paper provides recommendations for incorporating measures of personality traits into competence surveys.

RÉSUMÉ

De plus en plus de travaux montrent l'importance du rôle des compétences non seulement cognitives, mais aussi non cognitives, dans les retombées économiques et sociales. Ce document évalue les propriétés de mesure des échelles *Big Five* et Grit dans un large échantillon représentatif d'adultes en Pologne. Les données issues de l'étude polonaise menée à la suite du Programme pour l'évaluation internationale des compétences des adultes (postPIAAC) comprennent des données longitudinales sur les répondants au PIAAC en Pologne et des informations contextuelles supplémentaires non disponibles dans l'étude internationale. Les résultats présentés dans ce document montrent que tous les critères concernant la fiabilité, la validité et la comparabilité de ces échelles n'ont pas été respectés, même si les mesures relatives à la personnalité contribuent de manière significative à l'explication de la variabilité d'un ensemble de retombées pertinentes sur le plan de l'action publique. La plupart des questions font bien la distinction entre les individus présentant un niveau élevé ou faible concernant un trait de personnalité donné, bien que les items à formulation inversée se révèlent moins efficaces. Le cadre conceptuel à 5 facteurs de l'échelle *Big Five* n'a pas été répliqué ; toutefois, un modèle à 6 facteurs, comprenant un facteur supplémentaire intégrant les items à formulation inversée, a été appliqué. Dans le cas de l'échelle Grit, un modèle bi-factoriel, dont l'interprétation est équivalente à la structure conceptuelle de deuxième ordre, est appliqué. Les échelles ne sont pas totalement invariantes en termes de mesures. Les résultats confirment les conclusions de travaux antérieurs montrant une association nette entre les différences de traits de personnalité et les différences de retombées à l'échelle individuelle. Pour la plupart des retombées, les traits de personnalité retenus dans l'échelle *Big Five* s'avèrent des variables prédictives plus probantes que les compétences cognitives. Seul le niveau de formation présente une corrélation plus étroite avec les compétences cognitives, tandis que pour la rémunération, le pouvoir prédictif des traits de personnalité et celui des compétences cognitives sont similaires. Ce document propose un ensemble de recommandations pour l'intégration de mesure des traits de personnalité dans les évaluations des compétences.

TABLE OF CONTENTS

Tables

Figures

MEASUREMENT PROPERTIES OF NON-COGNITIVE SCALES IN THE POLISH FOLLOW-UP STUDY ON PIAAC (POSTPIAAC)

1. Introduction

There is a consensus that cognitive skills have important effects on economic and social outcomes. The evidence is ample, both in national and cross-national data. For over two decades comparable international surveys on cognitive skills have been conducted[1] with the most recent being the Programme for the International Assessment of Adult Competencies (PIAAC) co-ordinated by the OECD. The programme confirmed the importance of cognitive skills for economic and social outcomes in 24 participating countries/economies (da Costa et al., 2014; OECD, 2013).

However, it has also been recognised that cognitive skills as measured by achievement tests are not the only determinants of a successful life. There is a growing literature providing evidence on the importance of non-cognitive skills for life outcomes (for reviews, see Almlund et al., 2011; Borghans et al., 2008).

The striking causal evidence on the predictive power of non-cognitive skills has come from the Perry Preschool Program. This experimental intervention, targeted at disadvantaged three-year-old children, included weekly home visits for enriching children–parent interactions and preschool education with a curriculum aimed at fostering children's cognitive and socio-emotional skills. The programme lasted two years and both treatment and control groups have been followed through to age 40 (Heckman et al., 2010a). The programme did not boost IQ in the long run but produced significant treatment effects for educational and economic outcomes and crime (Heckman et al., 2010b). Heckman, Pinto and Savelyev (2013) show that it is actually persistent changes in non-cognitive skills that account for a large part of the effectiveness of the programme.

In addition, the General Educational Development (GED) programme in the United States provides evidence of the importance of non-cognitive skills for educational and economic outcomes. This programme offers high school dropouts the possibility of assuring themselves that their cognitive skills are equivalent to those of high school graduates. Despite the fact that their cognitive skills are similar, they nevertheless perform worse in the labour market than the high school graduates with no college. Heckman and Kautz (2012) show that the non-cognitive skills of GED recipients are closer to high school dropouts than to graduates, which results in the differences in the labour market performance.

Almlund et al. (2011) summarise why personality should not be ignored in the research of life outcomes. First, personality traits have competing predictive power for important outcomes to measures of cognition. Moreover, very often performance in achievement tests depends not only on cognition but also on personality. The authors also highlight the relevance of personality traits for policy interventions, as these traits are more malleable than cognition; hence such interventions aimed at boosting non-cognitive skills can be a way of addressing social problems.

One of the challenges in personality assessment in large scale studies is the measurement of personality traits. Unlike standardised achievement tests, personality tests are based on self-reported measures, which are less reliable. The only international survey that collects information, both on cognitive and non-cognitive skills in a homogeneous way across countries, is the Skills Towards Employment and Productivity (STEP) Study conducted in developing countries. The preliminary findings show that more conscientious, emotionally stable and grittier (determined) workers find their first job faster. Likewise, non-cognitive skills are associated with higher wages (World Bank, 2014).

However, we still lack a comprehensive study that systematically evaluates the combined impact of cognitive and non-cognitive skills in developed countries. While the first cycle of PIAAC made an assessment across three domains - literacy, numeracy and problem solving in technology-rich environments - no personality measures are available. The second cycle of PIAAC is planned for 2018–2023 and will be an opportunity to build on experience from the previous cycle. One of the possible changes might be the inclusion of new areas of assessment, particularly non-cognitive skills.

The objective of this paper is to assess the measurement properties of the Big Five and Grit scales, namely their reliability and validity in a large representative sample of adults in Poland. The paper is organised as follows. Section 2 provides an overview of the context and existing empirical evidence on the linkages between non-cognitive skills and life outcomes. Section 3 describes the postPIAAC study and the psychological scales used. Section 4 examines the psychometric properties of the Big Five and Grit scales. First, reliability and structural validity is assessed before the analysis of measurement equivalence (representativeness) and discriminant validity. Section 5 presents the evidence on the analytical importance of personality measures: the criterion validity for education, life satisfaction, trust, health and labour market outcomes. Section 6 concludes.

2. The importance of personality traits

This section discusses the relationship between personality and cognitive skills and provides the empirical evidence on the link between personality and the life outcomes analysed in the paper.

There has been a long-lasting debate on the degree of overlap between the constructs of personality and cognitive ability. The literature suggests two channels of influence of personality traits on achievement test scores. The first is via an indirect effect on knowledge: that the acquisition of cognitive skills depends on personality traits (Cunha and Heckman, 2008). The second is that performance in achievement and intelligence tests depends not only on cognitive ability but also on personality traits and motivation (Borghans, Meijers and Ter Weel, 2008; Segal, 2006).

Scientists often refer to intelligence, cognitive ability and achievement test scores interchangeably. Although there is a high correlation of scores between these different tests of cognitive ability, Borghans et al. (2011) stress that performance in achievement tests depends both on IQ and personality.[2] As proposed by Almlund et al. (2011), the term "intelligence tests" should be reserved for tests with emphasis on fluid intelligence, while "achievement tests" should be for tests of mainly crystallised intelligence. In this paper, by "cognitive skills" we refer to the score on the PIAAC test, which is a standardised achievement test.

A meta-analysis by Ackerman and Heggestad (1997) shows a positive association of general intelligence with Openness and Extraversion, and a negative one with Anxiety, a trait associated with Neuroticism. No relationship with Conscientiousness has been reported. In the case of mathematical/numerical ability they find a positive correlation with Extraversion and a negative one with Anxiety and Conscientiousness. With regard to cognitive skills measured in PIAAC, di Francesco et al. (2015) show a positive correlation of cognitive skills with Openness and Conscientiousness[3] in the Italian PIAAC sample. Using the German PIAAC sample, Rammstedt, Danner and Martin (2016) also show a positive correlation with Openness but a negative one with Conscientiousness. They also find a negative relationship between cognitive skills and Neuroticism and Extraversion. All these correlations range between 0.05 and 0.13. In sum, Openness is positively and Neuroticism negatively related to cognitive ability, while the results on Conscientiousness and Extraversion are mixed.

We now turn to evidence on the power of personality to predict important outcomes. We will focus on studies which examine the Big Five or Grit scales, or traits which can be linked to them, as these are the

traits analysed in this paper. Some caution must be taken when comparing the size of the effects as often the covariates controlled and the metrics reported differ between the studies.

Traits connected with motivation and interest in new ideas and learning might cause people to study for longer. From the Big Five dimensions, Openness to Experience is the best personality predictor of educational attainment (Goldberg et al., 1998; O'Connell and Sheikh, 2011; Van Eijck and de Graaf, 2004). The first two studies also report a positive but much weaker association of Conscientiousness with years of education, while O'Connell and Sheikh (2011) report no such association. Three aforementioned studies control for different covariates: Goldberg et al. (1998) include only age, gender and ethnicity into specification; the next study controls additionally for parents education and occupation; while the last one controls for gender and cognitive functioning. Van Eijck and de Graaf (2004) and O'Connell and Sheikh (2011) find that Extraversion and Neuroticism are negatively linked to educational attainment. A meta-analysis by Poropat (2009), which evaluates the relationship between the Big Five traits and academic performance, shows that Conscientiousness is most strongly related to grades. There is relatively less research on the criterion validity of the Grit scale. Eskreis-Winkler et al. (2014) show that grittier individuals are more likely to graduate from high school, and that the relationship holds also when controlling for academic Conscientiousness. Moreover, individuals with a higher level of Grit have smaller chances of dropping out of education or the labour market (becoming NEETs - not in education, employment, or training) at age 18–20 (Mendolia and Walker, 2014).

As with educational attainment, many studies find a relationship between the Big Five dimensions and labour market outcomes. Conscientiousness is positively related to job performance, training proficiency and personnel data, while Openness and Extraversion predict training proficiency (Barrick and Mount, 1991). Various studies have shown that there is a wage penalty for Agreeableness and Neuroticism (Mueller and Plug, 2006; Nyhus and Pons, 2005; O'Connell and Sheikh, 2011). Studies on samples from the United States also note the positive association between Openness and wages (Mueller and Plug, 2006; O'Connell and Sheikh, 2011). The importance of the traits related to Neuroticism / Emotional Stability, such as self-esteem and locus of control for the wage setting, is confirmed by Drago (2011) and Heineck and Anger (2010). While most of the literature examines the effect in the United States and Western Europe, the study by Cunningham, Torrado and Sarzosa (2016) is an exception. Using the Peruvian National Skills and Labor Market Survey, they find that Openness and Emotional Stability are related to wages, and that an aggregate of these two – plasticity – is related to employment. There is also a negative link between traits connected to Agreeableness and wages. Additionally, the study includes one of the Grit subscales: Perseverance of Effort – which is positively correlated with employment when controlling for Conscientiousness. When analysing workers' well-being, a meta-analysis by Judge et al. (2002) indicates that extraverts are more likely to be satisfied with their work, while neurotic individuals are less likely to be satisfied.

Quite a large body of the literature reports a link between personality and health. As noted by Almlund et al. (2011), the mechanism of influence is not well recognised, but there is some evidence suggesting that personality affects health-related behaviours, like smoking, eating habits and exercise. Hampson et al. (2007) find that Extraversion, Agreeableness and Conscientiousness, measured during childhood, predicts self-rated health and healthy behaviours in midlife. Part of this effect is indirect through educational attainment. Using longitudinal data and concentrating on Conscientiousness, Takahashi et al. (2013) find that changes in this trait are correlated with health-related behaviours and improvements in self-rated health. In addition to examining the association with self-rated health, other studies have concentrated on the relationship between personality and longevity/mortality. A meta-analysis by Roberts et al. (2007) indicates that Conscientiousness, Extraversion and Agreeableness have positive effects on longevity, while Neuroticism has a negative effect. All personality traits were stronger predictors of mortality than socio-economic status and were comparable to IQ. Overall, among the Big Five traits, Conscientiousness is the strongest predictor of health outcomes.

Health can be viewed as both an individual and a social outcome. Another important social outcome is interpersonal trust: one of the measures of social capital which boosts economic growth. There is scarce evidence on the relationship between personality and trust. The exception is Dohmen et al. (2008) who examine this link using a German representative sample. According to their study, Openness and Agreeableness are positively linked to trust, while Conscientiousness and Neuroticism are negatively linked.

One of the important components of individual well-being, next to economic situation and social outcomes, is life satisfaction. Numerous studies indicate that personality can explain individual differences in life satisfaction. A meta-analysis indicates that Neuroticism and Extraversion are the strongest predictors of subjective well-being among the Big Five traits (DeNeve and Cooper, 1998). More recently, using the British Cohort Study 1970, Prevoo and ter Weel (2015) show that Extraversion in childhood is the strongest predictor of life satisfaction around 30 years later, while Neuroticism does not have a significant influence.

In sum, non-cognitive skills play an important role in predicting a wide range of economic and social outcomes. Therefore, these skills can be considered as an important complement to the cognitive skills already measured within the PIAAC framework. Inclusion of measures of non-cognitive skills would allow us to gather comparable evidence for a wider variety of countries.

3. Data and personality measures

PostPIAAC research design and sample

We analyse the data from the Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC). The main goals of the study were to gather longitudinal information on PIAAC respondents in Poland and to collect additional background information not available in the international study.

PIAAC is an international survey measuring adults' skills. Twenty-four countries/economies participated in Round 1 and a further nine in Round 2. The first cycle covered the assessment in three domains: literacy, numeracy and problem solving in technology-rich environments, which are defined as the key information-processing skills necessary to participate in present-day society (OECD, 2013a). The aim of the programme is to provide data on the level and distribution of skills among adult populations in the participating countries and to investigate the relationship between proficiency and meaningful life outcomes, such as employment status, wages or social trust. Additionally, it seeks to find factors important for the development, maintenance and decline of basic cognitive skills. In order to serve these needs the assessment is accompanied by an extensive background questionnaire collecting information on educational attainment and training participation, current or previous job, skills use at home and at work, socio-economic background and diverse social outcomes.

The background questionnaire (BQ) of postPIAAC is based on the PIAAC international questionnaire with many additional questions and some modifications. The methodology of collecting labour market outcomes has not been changed. Regarding other outcomes, a question on life satisfaction was added, and the question on political efficacy was removed, leaving the rest of the social indicators unchanged (health and trust). The important extension of the postPIAAC study is the inclusion of measures of non-cognitive skills. The study includes two self-reporting scales: the Big Five Inventory-Short (BFI-S) (Gerlitz and Schupp, 2005; John, Donahue and Kentle, 1991) and the short eight-item Grit scale (Grit-S) (Duckworth and Quinn, 2009). The choice of personality measures to be included was driven by expert advice and the possibilities of comparability to other longitudinal studies with PIAAC samples in other countries (BFI-S is included in the German PIAAC-L and the Canadian Longitudinal and International Study of Adults).

The BQ was administered as a computer-assisted personal interview (CAPI). The study included parts with direct assessment, both on computer (a working memory test and a basic ICT skills test) and on paper (a coding speed test, a Big Five personality test and a self-assessment of skills). The Grit test was part of the BQ.

The target population for PIAAC included all non-institutionalised individuals aged 16–65, residing in Poland during the period of data collection in 2011–2012. The target population of postPIAAC were PIAAC respondents who lived in Poland during the fieldwork conducted between October 2014 and February 2015. PIAAC respondents who had either died or emigrated between the interviewers' visits were classified as ineligible in postPIAAC. The weighting process of the postPIAAC sample was based on PIAAC guidelines (OECD, 2013b).

The final weight in PIAAC was taken as the person base weight in postPIAAC. The next step involved correcting for non-response in order to reduce potential bias arising from differences between respondents and non-respondents. Using a classification tree methodology, adjustment cells were constructed and which were homogeneous with respect to the response rate. The calibration referred to the population estimates produced by PIAAC with respect to age, gender and proficiency score. The weighting process ensured that the average PIAAC results are replicated between the original and postPIAAC sample with regards to standard characteristics such as gender, age or educational attainment. Additionally, replication weights (paired jackknife) were computed in order to facilitate the estimation of variance.

Of the initial 9 366 respondents in PIAAC, 5 224 completed postPIAAC interviews in 2014/2015. Of this group 4 551 and 5 121 completed the Big Five and Grit tests respectively and included information on age, gender and education. The scales' properties were investigated based on these samples. After further selecting individuals with valid answers to the relevant outcome questions, we kept a working sample of 4 454 for the analysis of life outcomes. The working sample for job quality outcomes was further reduced to 3 026 and 2 059 respectively for job satisfaction and wages.

The analysis is based on postPIAAC data combined with the proficiency estimates from PIAAC. The interval between the interviews for an individual respondent is from 2.5 to 3.5 years. The analysis thus assumes that respondents' cognitive skills have not changed significantly between the two waves.

Personality measures

The Big Five is one of the most popular frameworks for describing human personality. It is based on the psycholexical traditions (John and Srivastava, 1999). The model identifies five dimensions of personality (Costa and McRae, 1985; John and Srivastava, 1999; McCrae and Costa Jr, 1999), which are:

- Agreeableness: trust, straightforwardness, altruism, compliance, modesty and tender-mindedness.
- Conscientiousness: competence, order, dutifulness, achievement-striving, self-discipline and deliberation.
- Extraversion: gregariousness, assertiveness, activity, excitement-seeking, positive emotions and warmth.
- Neuroticism: anxiety, anger, hostility, depression and self-consciousness.
- Openness: ideas, fantasy, aesthetics, actions, feelings and values.

There are many self-reporting measures based on the Big Five model: one of the most well-known is the Revised NEO Personality Inventory (NEO-PI-R) which consists of 240 items. Although the NEO-PI-R allows for a detailed measurement of personality traits, it is inconveniently long. Therefore, shortened questionnaires are often used, such as the NEO-FFI (NEO Five-Factor Inventory) which consists of 60

items (Costa and McCrae, 1992) and the Big Five Inventory (BFI) with 44 items (John et al., 1991). Completing these questionnaires takes around about five minutes.

Recently, psychological measures have often been included in large scale social surveys (Wagner, Frick and Schupp, 2007). Questionnaires have to be as short as possible due to the time-constraints and the wide range of topics covered by such surveys (Rammstedt and John, 2007). In the postPIAAC study, we used the BFI-S (Gerlitz and Schupp, 2005), which consists of 15 items. BFI-S had already been used in many household surveys, such as the German Socio-Economic Panel Study and the Household, Income and Labour Dynamics in Australia (HILDA) Survey.

Hahn, Gottschling and Spinath (2012) describe the psychometric properties of the BFI-S. Its items load on the respective scales from the BFI, so structural validity is established. A shortened scale correlates with the same constructs as the original scale (so that convergent validity is ensured). Moreover, both the BFI-S and the BFI are unrelated to the same constructs,[4] so discriminant validity is obtained. The authors also find test–retest correlations for all the personality subscales (for the interval of 18 months). High test-retest correlations were also obtained by Cobb-Clark and Schurer (2012). In the sample of working-age adults (HILDA survey), personality (as measured by the Big Five) is stable over a four-year period.

Most studies concerning Big Five questionnaires are validated in homogeneous student samples (Rammstedt, Goldberg and Borg, 2010). Replicating the Big Five structure is however harder in more heterogeneous groups with a different educational background. The examples of studies which fail to replicate the Big Five structure in a non-student sample are:[5] Körner, Geyer and Brähler, (2002); Lang, Lüdtke and Asendorpf (2001); Mottus, Allik and Pullmann (2007); Rolland, Parker and Stumpf (1998); Tokar et al. (1999) and Toomela (2003). This effect might be attributed to the response style: less educated respondents tend to agree more while completing the questionnaire. Rammstedt et al. (2010) show that, when the answers were corrected for acquiescence, the Big Five model was replicated in all the groups, independently of their educational level.

In order to conduct reliable comparisons between the analysed groups, measurement invariance (MI) needs to be obtained. However, research on MI on BFI-S is rather limited. Lang et al. (2011) investigate it with respect to administration mode (face-to-face assessment, self-administered, computer-assisted telephone interview—CATI). CATI is not invariant in the eldest group, as it probably demanded increased mental effort. Specht, Egloff and Schmukle (2011) check MI for different age groups and for rank-order stability (of four years). The authors obtain strict measurement invariance[6] for both grouping variables.

The other non-cognitive measure used in the postPIAAC study is Grit, which is defined as perseverance and passion for long-term goals. People with a high Grit level sustain their interest and effort in an activity, despite challenges, failures and a lack of positive feedback (Duckworth et al., 2007; Duckworth and Quinn, 2009). The original self-report measure of Grit (Grit-O, Duckworth et al., 2007) consisted of 12 items. Grit is a second-order factor, including the first-order latent constructs of Consistency of Interest and Perseverance of Effort.

Duckworth and Quinn (2009) show that a shortened version of Grit (Grit-S), which consists of eight items, displays adequate internal consistency. A second-order latent factor model is well fitted to data from various subsamples, which supports the structural validity. The shortened scale (Grit-S) was therefore used in the postPIAAC study.

Grit is a similar construct to one of the Big Five factors—Conscientiousness. Conscientiousness is defined as a "socially prescribed impulse control that facilitates task- and goal-related behaviour" (John and Srivastava, 1999: 121). It consists of such traits as being careful, thorough, responsible, organised and

planful. Additionally, it (just like Grit) includes volitional traits such as being hardworking, achievement-oriented and persevering (Barrick and Mount, 1991).

However, Duckworth et al. (2007) argue that Grit and Conscientiousness are conceptually different. Grit emphasises stamina: being able to sustain effort and interest in projects which take a lot of time to complete. Nevertheless, both constructs correlate strongly. Duckworth and Quinn (2009) show that the correlations of the Conscientiousness with Consistency of Interest and Perseverance of Effort subscales are respectively 0.64 and 0.74. According to Ivcevic and Brackett (2014) the correlation of Grit and Conscientiousness is 0.44. Duckworth and Quinn (2009) also analyse the incremental validity of Grit. When controlling for Conscientiousness and the other Big Five traits, they find that grittier individuals obtained more education than those of the same age. Credé, Tynan and Harms (2016) criticize Grit for the lack of construct validity and show that it is highly correlated with Conscientiousness. However, they find that Perseverance of Effort subscale explains the variance in academic performance (basing on GPAs in high school, college, and individual grades) even after controlling for Conscientiousness.

Duckworth and Quinn (2009) find evidence for the predictive validity and test–retest stability of Grit-S. People who are grittier have higher educational attainment and make less career changes, when controlled for other personality traits and age. We can observe this effect in various subsamples. Children with a high Grit level are more likely to win in spelling competitions and grittier adolescents obtain higher GPAs (Grade Point Average). Cadets with higher Grit drop out less frequently during training. Grit-S also has a high consensual validity (i.e. correlations between self-reports, peer reports and family reports are medium to large) and test–retest stability (comparable with NEO Five-Factor Inventory).

In order to ensure comparability, any analysis with latent variables requires that test items have been appropriately translated and culturally adapted (e.g. Hambleton and Patsula, 1999). There are differences in the Polish and English versions of the BFI-S, especially in the case of negatively-worded items (see Table A.1 in Annex A).

Two items are stronger in Polish. The Polish translation of "sometimes a bit rude to others" (Agreeableness) omits "a bit" and "somewhat lazy" (Conscientiousness) is translated into Polish as "lazy". Consequently, answers may be biased due to the social desirability factor. The item "reserved" (Extraversion) is translated as "reserved towards people". Therefore, the Polish item has stronger, possibly negative, connotations than does keeping one's opinions and emotions to oneself. It is possible that people may disagree with the statement due to a social desirability factor or because being introverted does not necessarily imply disliking people or having a social phobia. The item "relaxed, able to deal with stress" (Neuroticism) is translated as "calm in stressful situations", which changes its meaning. The original item suggests that when someone is stressed he or she manages to relax and deal with this feeling; however, the Polish translation implies that stress does not affect the person in any way.

Unlike the BFI-S translation, the Grit-S translation is equivalent to its original version (Table A.1 in Annex A).

4. The psychometric properties of BFI-S and Grit-S scales

This section examines the psychometric properties of the BFI-S and Grit-S scales, both at the item level and at the scale level. First, we perform the analysis within the Item Response Theory (IRT) paradigm. Then, we assess reliability and check for structural validity and measurement invariance of both scales. Finally, we perform the analysis of discriminant validity.

Item-level analysis

The following section considers the performance of particular BFI-S and Grit-S items. We show both basic descriptive statistics and results of IRT analysis. Although IRT is mainly applied in educational research, it is also a valuable diagnostic tool of psychological measures, especially those using categorical ordered Likert scales.

This methodology allows to assess item performance with item-fit coefficients, which show whether the model underestimates or overestimates the latent trait level (Reise, 1990). It is also possible to determine to what extent the observed item response patterns deviate from those predicted by the model.[7]

We analyse the data with a graded response model (GRM, Samejima, 1969), appropriate for categorical ordered responses. A model is estimated for each item category and gives information about the probability of giving an answer of at least a value of x. As a result, we obtain item characteristic curves for m-1 item categories. This model can be written as follows (note that this is the inverse of a two-parameter logistic model (Samejima, 1969):

$$P_x(u_i \leq x \mid \theta, a_i, b_{i,x}) = \frac{-1}{1+e^{-a_i(\theta - b_{i,x})}},$$

where: a_i -a discrimination parameter, $b_{i,x}$ – difficulty parameters for each category level, x – value attributed to each category level, u_i –answer given by respondent.
Therefore, models for each item category have different difficulty parameters $b_{i,x}$, but discrimination parameter (a_i) is shared.

The properties of BFI-S items

Table 1 shows descriptive statistics for each BFI-S item. Respondents tended to agree rather than disagree in the case of all but two items (“reserved” and “relaxed, able to deal with stress”), which are reverse-worded items.

Table 1. Descriptive statistics of BFI-S items

Item	Mean	Std. dev	Jknife S.E.	95% Confidence interval	
				LB	UB
Agreeableness					
Sometimes a bit rude to others*	4.97	1.81	0.03	4.91	5.04
Forgiving	5.05	1.74	0.04	4.98	5.12
Considerate and kind to others	5.61	1.32	0.03	5.56	5.67
Conscientiousness					
A thorough worker	6.21	1.13	0.02	6.16	6.25
Somewhat lazy*	5.37	1.75	0.03	5.30	5.43
Effective and efficient in completing tasks	6.02	1.07	0.02	5.97	6.06
Extraversion					
Communicative, talkative	4.43	1.90	0.04	4.35	4.51
Outgoing, sociable	4.72	1.69	0.04	4.65	4.79
Reserved*	2.98	1.54	0.03	2.92	3.04
Neuroticism					
A worrier	4.84	1.85	0.04	4.77	4.92
Nervous	4.38	1.94	0.04	4.29	4.47
Relaxed, able to deal with stress*	3.08	1.70	0.04	3.00	3.15
Openness					
Original, someone who comes up with new ideas	4.89	1.48	0.03	4.82	4.95
Someone who values artistic, aesthetic experiences	5.21	1.64	0.04	5.14	5.29
Imaginative	5.02	1.62	0.04	4.95	5.09

Notes: *Reverse-worded item. Upper bound (UB). Lower bound (LB).

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Fitting Graded Response Model (GRM) allows for an analysis of the discrimination and difficulty parameters (IRT parameters are shown in Annex B, ICC plots in Annex C).[8] All the negatively-worded items show a worse IRT model fit than positively-worded items. What is more, the item "reserved" (Extraversion) has unusual observed response patterns given the latent trait level (see Figure C.9 in Annex C). For the "reserved" item, the probability of disagreeing strongly is the highest both in groups with the highest latent trait level (i.e. definitely extroverted) and the lowest latent trait level (i.e. definitely introverted). Moreover, nearly no one used the category "agree strongly". The middle categories on the ICC plot are flat, as the item does not discriminate between people with high and low level of extraversion. It seems likely that we can attribute this effect to the translation of the item (see section 3 and Annex A).

The other reverse-worded items also fit poorly, probably also due to problems with translation (for details see section 3). Negatively-worded items have the lowest discrimination parameters in each subscale. The ability to differentiate between respondents with high and low latent trait level is poor for all these items.

The properties of Grit-S items

Grit is a second-order construct, so we show the properties of its two subscales (first-order latent constructs) – Consistency of Interest and Perseverance of Effort.

We observe acquiescence (especially in the Perseverance of Effort subscale), as the mean of all the items was higher than 2.5 (see Table 2).[9]

Table 2. Descriptive statistics of Grit-S items

Item	Mean	Std. dev.	Jknife S.E.	95% Confidence interval	
				LB	UB
Consistency of interest					
New ideas and projects sometimes distract me from previous ones*	3.05	1.11	0.02	2.99	3.08
I have been obsessed with a certain idea or project for a short time but later lost interest*	3.45	1.14	0.03	3.40	3.50
I often set a goal but later choose to pursue a different one*	3.50	1.10	0.02	3.46	3.55
I have difficulty maintaining my focus on projects that take more than a few months to complete*	3.41	1.18	0.03	3.35	3.46
Perseverance of effort					
Setbacks don't discourage me	3.49	1.15	0.02	3.45	3.54
I am a hard worker	4.06	0.87	0.02	4.02	4.09
I finish whatever I begin	4.10	0.91	0.02	4.07	4.13
I am diligent	4.03	0.87	0.02	4.00	4.07

Notes: * Reverse-worded items. Upper bound (UB). Lower bound (LB).

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

We observe a poor fit for the item "setbacks don't discourage me" (Figure D1 in Annex D), which is probably caused by the administration of the item in CAPI mode. Grit-S items were read to a respondent by an interviewer. Answering may have demanded increased mental effort due to the use of a double negative in the formulation of the item. Therefore, the response pattern was disturbed. This item has also the lowest discrimination in the Perseverance of Effort subscale. The item "new ideas and projects sometimes distract me from previous ones" (Consistency of Interests) has the lowest discrimination of all Grit-S items; however, the reason for that is not clear.

Why do negatively-worded items seem to function in somehow different way than their positively-worded counterparts? One cause might be cultural adaptation (translation) of these items, described earlier, another is the psychometric properties of reverse-worded items.

The practice of using mixed-worded items is widespread, as it is believed to decrease acquiescence bias (Nunnally, 1978). However, it is assumed implicitly that there are no differences between a positively scored item and its opposite ("I am happy" vs "I am not sad") as well as between two types of negatively scored items ("I am sad" and "I am not happy') (Wong, Rindfleisch and Burroughs, 2003).

Additionally, reverse-worded items decrease scale internal consistency (Schriesheim and Eisenbach, 1995) or load on a separate factor during factor analysis (Herche and Engelland, 1996). The effects have been observed independently of the acquiescence effect (Terborg and Peters, 1974). Such an impact is present only for mixed-worded scales - those consisting only of negatively-worded items are not affected (Dalal and Carter, 2015). Can the lower discrimination of reversely-worded items influence results of the subsequent analyses with non-cognitive measures? Although they perform worse, deleting such items may decrease the validity of the scale as a whole. The only item which presents a truly distorted response pattern in our analysis is "reserved" from the Extraversion subscale – as people with extremely low and extremely high levels of the latent trait tend to answer in a similar manner. Therefore, we expect that Extraversion subscale will perform better without this particular item in the case that simple sum scores by factor are used.

Item-level analysis allowed for an investigation of each item's performance. In case of the BFI-S questionnaire, the reverse-worded items perform worse than the positive ones. All the Grit-S items except "setbacks don't discourage me" perform well and give the expected data pattern. This particular item has distorted response patterns due to the presence of a double negative and the manner of administration of the item which was potentially confusing for respondents.

Negatively-worded items can influence the results of subsequent analyses. In the literature it is often found that they tend to load on specific, additional factors and decrease the reliability of the whole mixed-worded scales. Thus we might expect that the BFI-S, which consists of four mixed-worded subscales and one positively-worded subscale (Openness), would be affected, especially in the analysis of reliability and structural validity. The results for Grit-S should not be influenced, as one of its subscales consists of positively-worded items only and the other of negatively-worded items only.

Scale-level (reliability) analysis

Scale reliability is a necessary (but not sufficient) condition for scale validity (Moss, 1994). We assess reliability with several statistics: RIR (item-rest correlation), Cronbach's alpha when item is removed from the scale, standardised Cronbach's alpha coefficient, hierarchical and total omega coefficients.

Cronbach's alpha coefficient (Cronbach, 1951) remains one of the most popular methods of reliability assessment. However, alpha requires several assumptions, which are not always fulfilled or possible in psychological research:[10] relying on alpha itself might lead to erroneous conclusions – mostly the underestimation of population ("true") level of reliability (Schmitt, 1996; Sijtsma, 2009). What is more, alpha depends on the number of items in a given scale (Cortina, 1993). Therefore, we present omega coefficients (Revelle and Zinbarg, 2009; Zinbarg et al., 2005) as well. Hierarchical omega accounts for the variance explained by general factor, total omega for the variance explained both by the general factor and the specific factors. They can be especially useful for assessing Grit-S reliability, as it has a second-order structure.

The reliability of BFI-S items

The descriptive statistics show acquiescence for each BFI-S subscale (means ranging from 4.04 to 5.86), especially for Conscientiousness, Agreeableness and Openness. The standard deviations of the results are the highest for Neuroticism and the lowest for Conscientiousness (Table 3).

Table 3. Descriptive statistics of BFI-S subscales

			Jknife	[95% Confidence interval]	
Subscale	Mean	Std. dev	S. E.	LB	UB
Agreeableness	5.21	1.08	0.02	5.17	5.25
Conscientiousness	5.86	0.98	0.02	5.83	5.90
Extraversion	4.04	1.13	0.02	4.00	4.09
Neuroticism	4.10	1.30	0.03	4.04	4.16
Openness	5.04	1.16	0.03	4.99	5.09

Notes: Upper bound (UB). Lower bound (LB).

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

The results of reliability analysis[11] are shown in Table 4. They are consistent with the item-level analysis described in the previous subsection.

Table 4. Reliability of BFI-S subscales (N= 4551)

Item	RIR	Alpha. when item deleted	Alpha std.	Omega hierarchical	Omega total
		Agreeableness			
Sometimes a bit rude to others*	0.16	0.45			
Forgiving	0.23	0.32	0.42	0.48	0.52
Considerate and kind to others	0.34	0.16			
		Conscientiousness			
A thorough worker	0.43	0.39			
Somewhat lazy*	0.31	0.66	0.61	0.63	0.77
Effective and efficient in completing tasks	0.44	0.39			
		Extraversion			
Communicative, talkative	0.32	0.05			
Outgoing, sociable	0.29	0.13	0.36	0.47	0.57
Reserved*	0.06	0.54			
		Openness			
Original, someone who comes up with new ideas	0.42	0.46			
Someone who values artistic, aesthetic experiences	0.35	0.56	0.59	0.60	0.68
Imaginative	0.43	0.44			
		Neuroticism			
A worrier	0.33	0.44			
Nervous	0.43	0.27	0.52	0.56	0.62
Relaxed, able to deal with stress*	0.27	0.53			

Note: Item-rest correlation (RIR).

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Given the cut-off value of alpha=0.7 proposed by Nunnally (1978) for use in basic research,[12] none of the subscales can be considered reliable. However, such levels of reliability are common in empirical research using NEO-PI-R facets (McCrae et al., 2010). As the alpha underestimates reliability and is a function of scale length, omega can be a better indicator of scale reliability. In most cases, general and specific factors (omega total) explain between 52-77% of the variance of the observable results. However, unlike the case of alpha, there are no such rules of thumb for omega coefficients.

The Conscientiousness and Openness subscales have the highest reliability level, according to all the coefficients: alpha, omega hierarchical and omega total. The reliability of Extraversion subscale is the lowest. It has also the largest differences between alpha and omega hierarchical indicators – possibly due to lack of tau-equivalence (see more in Sheng and Sheng, 2012).

The reverse-worded items have the lowest item-rest correlations for all the subscales (except Openness, consisting of positively-worded items only). Excluding the negatively-worded items improves the reliability of all the subscales. These items perform also worse during IRT analysis, as they have the lowest discrimination parameters.

The reliability of Grit-S items

Descriptive statistics of Grit-S indicate that respondents tended to agree with the items – as in case of BFI-S. The standard deviation is higher for Consistency of Interest subscale (Table 5).

Table 5. Descriptive statistics of Grit-S subscales

(sub)Scale	Mean	Std. dev	Jknife S. E.	[95% Confidence interval] LB	UB
Consistency of interest	3.35	0.78	0.02	3.32	3.38
Perseverance of effort	3.92	0.68	0.01	3.89	3.95

Notes: Upper bound (UB). Lower bound (LB).

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

According to Duckworth and Quinn (2009), Grit is a second-order construct, which consists of two subscales – Consistency of Interest and Perseverance of Effort. Therefore, omega (both hierarchical and total) is the most appropriate indicator of composite reliability. We present both coefficients (for the general Grit factor and the subscales) in Table 6. The variance explained by the general factor is the same as in the case of the Neuroticism subscale (0.52). The general factor (Grit) and specific factors (Consistency of Interest and Perseverance of Effort) explain about 80% of the variance of the observed results, which is comparable to the best performing BFI-S subscale – Conscientiousness. The reliability (measured by the coefficient alpha) of the two subscales is quite similar – about 0.65-0.67. There are some slight differences in variance of the observed results explained by the two specific factors (see the values of omega hierarchical and total indicators). The Perseverance of Effort subscale explains more variance than does the Consistency of Interest subscale.

Two items have lower item-rest correlations – "new ideas and projects sometimes distract me from previous ones" (Consistency of Interest) and "setbacks don't discourage me" (Perseverance of Effort subscale). These items also have the lowest discrimination parameters (see Annex B). For the item "setbacks don't discourage me", the double negative formulation probably lowered the RIR coefficient. In case of the item "new ideas and projects sometimes distract me from previous ones", the reason is not clear.

Table 6. Reliability of Grit-S scale (N=5121)

Item	RIR	Alpha std.	Omega hierarchical	Omega total	Omega hierarchical	Omega total
Consistency of interest						
New ideas and projects sometimes distract me from previous ones*	0.22					
I have been obsessed with a certain idea or project for a short time but later lost interest*	0.43					
I often set a goal but later choose to pursue a different one*	0.48	0.65	0.64	0.71		
I have difficulty maintaining my focus on projects that take more than a few months to complete*	0.46				0.52	0.79
Perseverance of effort						
Setbacks don't discourage me	0.13					
I am a hard worker	0.40	0.67	0.70	0.79		
I finish whatever I begin	0.54					
I am diligent	0.50					

Note: Item-rest correlation (RIR).

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

The conclusions of the reliability (scale-level) analysis are similar to those of the item-level analysis. The negatively-worded items lowered their subscale's reliability. However, the Extraversion subscale was affected in the most significant way. Conscientiousness and Openness subscales are the most reliable, according to all the coefficients: alpha, omega hierarchical and omega total. Reliability values are similar to the values reported for the facets of the NEO-PI-R scale, which proved its validity in practice. Grit-S reliability is acceptable for its two subscales.

Structural validity

The following section discusses the structural validity of the Grit-S and BFI-S measures – i.e. whether the dimensional structure claimed by the theory was observed in the data. Theoretically, Grit is a second-order construct and the BFI-S measures five latent personality traits which are orthogonal. We conduct confirmatory factor analyses (CFA)[13] to check the goodness of fit of the proposed theoretical solution.

The CFA results show that the theoretical Big Five factor structure does not hold for the postPIAAC data. An orthogonal model with five factors for categorical variables gives poor fit values (Table 7). It is possible to improve the model with modification indices. However, they perform inconsistently across different samples and rely strongly on chance (MacCallum, Roznowski and Necowitz, 1992). Therefore, we fit two alternative solutions – a five-factor model where the correlations between factors are freely estimated and a six-factor solution, where all negatively-worded items are loaded by the separate factor. A graphical presentation of the six-factor model can be found in Annex E. We obtained the theoretical Big Five factor structure with the separate, artificial factor representing respondents' tendency to react differently on the reverse-worded items (irrespective of the item content).

In the previous sections, we observed that all the negatively-worded items perform worse than positively-worded items from the same subscales. Several studies suggest that negatively-worded items tend to load on separate factors during exploratory and confirmatory factor analysis, as well when a multitrait-multimethod is used (e.g. Dalal and Carter, 2015; DiStefano and Motl, 2006). Consequently, multidimensional models, accounting for the separate factor loading negatively-worded items, offer a better fit to the data (Chan, 2009; DiStefano and Motl, 2006; Marsh, 1986).

The "adjusted" solution accounting for an additional factor loading reverse-worded items (six-factor oblique) fits the observed data best (Table 7).

Table 7. Goodness of fit of the three alternative BFI-S models

Model	TLI	CFI	RMSEA
Five-factor orthogonal	0.450	0.542	0.199
Five-factor oblique	0.625	0.725	0.127
Six-factor oblique	0.860	0.905	0.077

Notes: Comparative Fit Index (CFI). Tucker Lewis Index (TLI). Root Mean Square Error of Approximation (RMSEA).

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Rammstedt, Goldberg, and Borg (2010) argue that the Big Five model does not hold among respondents with lower educational levels, unless their responses are corrected for acquiescence. However, the five-factor orthogonal model is not replicated in the postPIAAC sample in any of the educational groups.[14] When we corrected the results for acquiescence, as described in Rammstedt et al. (2010), the model did not converge. Consequently, the problems with replicating the BFI-S theoretical structure cannot be explained by acquiescence in groups with lower educational levels.

CFA shows that the proposed second-order structure of Grit does not fit the data well, as the model is not identified. The more appropriate model (RMSEA=0.057, CFI= 0.993, TLI= 0.984) has a bi-factor structure described by Canivez (2015) or Gignac and Watkins (2013). Each item loads on a general factor (which represents the target Grit construct) and specific factors (in this case Interest and Effort subscales). Specific factors may represent the variance which is not accounted by general factor (Reise, Moore and Haviland, 2010). Hence, a bi-factor model accounts both for single target construct and multidimensionality present in the data. A graphical representation of the model and the factor loadings are shown in Annex E.

Although we interpret them in a similar way, the second-order and bi-factor models are not mathematically equivalent (Chen, West and Sousa, 2006; Gustafsson and Balke, 1993). Reise et al. (2010) argue that both models describe the latent construct in a distinct way. Second-order models assume that the variance of items can be explained by weighted combination of primary traits. In a bi-factor model, the variance of items is explained both by general factor and group-specific factors.

The better fit of the bi-factor model is not a great threat to structural validity. Both models have similar interpretations. Moreover, it is common to observe evidence for the existence of a general dimension and multidimensionality at the same time (Chen et al., 2006; Reise et al., 2010).

The theoretical structure of the BFI-S does not hold for the postPIAAC sample, unless we fit the model with an additional factor accounting for the negatively-worded items. Multidimensional models with a separate dimension accounting for the presence of reverse-worded items provide better fit to the observable data. Although the theoretical structure (second-order factor) was not observed, the better fitting model (bi-factor model) can be treated as equivalent to the theoretical model. Therefore, there are no threats to structural validity of the Grit-S questionnaire.

Measurement invariance

Measurement invariance (equivalence) is necessary in order to make valid and reliable intergroup comparisons. Lack of invariance means that we are measuring different constructs in the groups that are compared. This subsection investigates the measurement invariance of the BFI-S and Grit-S scales.

Mellenbergh (1989) gives a formal definition of measurement invariance. A set of observable indicators Y forms a psychometric tool which measures the latent variable Z. The values of another variable, X, determine group membership. Invariance (with respect to X) is defined as the independency of Y and X, conditional on the latent trait Z. Therefore: $f(Y \mid Z = z) = f(Y \mid \mathrm{Z} = \mathrm{z}, \mathrm{X} = \mathrm{x})$ for each Z and X value. The function indicates that only trait Z is responsible for the differences in observable Y variables.

Both the results of a simple mean differences test or complicated structural equation models can be biased due to non-equivalence (Vandenberg and Lance, 2000). Thus, any type of intergroup comparisons requires measurement invariance analysis.

Testing invariance means answering four questions (Steinmetz et al., 2009):

- Whether we measure the same construct in different groups (i.e. configural invariance is reached). Testing metric, scalar and latent means invariance requires configural invariance.
- Whether the model parameters (factor loadings, measurement errors) are the same in the measured groups (i.e. metric invariance is reached). Metric invariance allows for comparing correlations of observable variables with latent constructs between analysed groups.

- Whether the presence of bias (response style) is the same in the analysed groups (i,e. scalar invariance is obtained). When scalar invariance is present, we can compare manifest scale scores reliably.
- Whether we can interpret differences in observable means as differences in latent means (i.e. whether invariance of latent means is obtained).

Each step of invariance testing requires the introduction of additional model constraints, so they have to be performed in an order (from configural to latent means).[15] According to the majority of researchers, obtaining scalar invariance is necessary to perform valid group comparisons (Meredith, 1993; Mullen, 1995).

We test four levels of invariance: configural, metric, scalar and latent means with multigroup confirmatory factor analysis (MCFA, French and Finch, 2008). The grouping variables are: gender, age,[16] educational level, literacy and numeracy skills.[17]

The data does not support the theoretical models described by their authors (Duckworth and Quinn, 2009; Gerlitz and Schupp, 2005; Costa and McRae, 1985; see previous subsection). So, we test the invariance only for six-factor model (BFI-S) and bi-factor model (Grit-S).

Most fit indices (CFI, TLI, RMSEA) used to determine model fit rely on the chi-square test, which is sensitive to the sample size. The bigger the sample, the more probable it is that the invariance hypothesis will be rejected (basing on these fit indices). Therefore, Cheung and Rensvold (2002) proposed another criterion that is more robust in the case of large samples – delta CFI.[18] Delta CFI value lower than 0.002 indicates that proposed invariance model fits the data (Cheung and Rensvold, 2002). We use only this criterion of model fit – as our sample size is large (4 551 observations for BFI-S, 5 121 for Grit-S). We treat BFI-S and Grit-S questions as continuous variables.[19]

Measurement invariance of BFI-S

The BFI-S scale is metrically invariant for all the grouping variables except for age for which configural invariance was established (Table 8). In case of educational level, we cannot draw valid conclusions.

The solution indicates that the same personality traits are measured in different groups and that the relationship between latent constructs and observable variables is the same (factor loadings are equal). However, the bias (response style) level is different among the analysed groups (as delta CFI > 0.002). What is more, the differences in the observed means cannot be interpreted as differences in the latent variable means (for the last hypothesis, delta CFI > 0.002 in all analysed groups).

Table 8. Invariance testing of six- factor BFI-S model

Variable	Delta CFI (metric vs. configural invariance)	Delta CFI (scalar vs. matric invariance)	Delta CFI (latent means vs. scalar invariance)	Level of invariance
Gender	-0.015	0.037	no convergence	metric
Age	0.004	0.031	0.020	configural
Educational level		no convergence		NA
Literacy level	-0.004	0.012	0.006	metric
Numeracy level	-0.015	0.014	0.007	metric

Note*:* Comparative Fit Index (CFI).

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Measurement invariance of Grit-S

In case of Grit-S only the bi-factor model (described in the previous subsection) was tested for measurement invariance[20] (see Table 9 for the results) as the second-order model did not converge.

Table 9. Invariance testing of bi-factor Grit-S model

Variable	Delta CFI (metric vs configural)	Delta CFI (scalar vs metric invariance)	Delta CFI (latent means vs scalar invariance)	Level of invariance
Gender	0.002	0.002	0.015	scalar
Age	0.003	0.011	0.011	configural
Educational level	0.001	0.009	0.008	metric
Literacy level	0.002	0.012	0.01	metric
Numeracy level	0.004	no convergence		configural

Note: Comparative Fit Index (CFI).

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

We found scalar invariance for gender and metric invariance for education and literacy levels. We observed configural invariance for the remaining variables. In case of gender, the measured constructs have similar meaning, and their relationship with the observable variables manifests itself in the same way. The systematic bias is the same in both groups. However, we cannot interpret the observed differences in mean as latent means differences. For the education and literacy level, we found metric invariance. The Grit construct has the same meaning in the analysed groups. Its relationship with the observable variables is independent of the education and literacy level. Age and numeracy level are configurally invariant, which means that the same constructs are measured in the analysed groups (but nothing more).

The BFI-S and Grit-S are mostly metrically invariant. The same constructs of personality and Grit are measured in the analysed groups and the strength of relationship between latent constructs and observable variables is the same. However, the level of response bias may differ. At the same time, direct intergroup comparisons for different age groups (and numeracy level in case of Grit-S) may be biased, as only configural invariance was reached. We cannot determine the level of invariance for different education level groups for the BFI-S, as the model did not converge.

Can the lack of full (scalar) measurement equivalence influence the results of other analyses using the BFI-S and Grit-S? Other approaches to invariance testing (partial invariance and Bayesian invariance – see Byrne, Shavelson and Muthén, 1989; Steenkamp and Baumgartner, 1998; Asparouhov and Muthén, 2014; Muthén and Asparouhov, 2013) indicate that the criteria used within the MCFA paradigm are too strict. All the solutions other than full scalar invariance are said to be biased (Borsboom, 2006). However, scalar invariance is rarely obtained in practice.

Moreover, lack of full (scalar) measurement invariance does not imply that all subsequent analyses with the BFI-S and Grit-S are biased. For example, Millsap (2007) distinguishes between measurement invariance and prediction invariance. The latter means that the relationship between observable variables Y and dependent variable V does not vary on other group characteristics (X). Therefore, prediction invariance concerns only observable variables. Millsap (2007) proves that it is possible to obtain prediction invariance, even when full measurement invariance is not observed.

As we did not obtain full measurement invariance, care should be taken when making direct comparisons of latent personality and Grit constructs (as response bias can be different in these particular groups). However, lack of full measurement invariance does not necessarily influence the results of analysis on observable variables.

Discriminant validity: The association between non-cognitive skills and competences

To what extend is the personality of an individual linked to his cognitive abilities? As noted earlier the evidence in the literature is mixed and the result often depends on the measure of cognitive ability employed.

This subsection uses the factor scores from the best fitted models (see section 4). The lack of correlation between Openness and cognitive skills is surprising as it is often reported in the literature (Table 10). All the other Big Five traits are negatively associated with both literacy and numeracy with the correlations ranging from -0.06 to -0.13. Grit is not correlated with cognitive skills.

Table 10. Correlations of the personality traits with literacy and numeracy

	Literacy	Numeracy
Conscientiousness	-0.134***	-0.109***
Extraversion	-0.059**	-0.063**
Agreeableness	-0.121***	-0.096***
Openness	-0.019	-0.011
Neuroticism	-0.078***	-0.096***
Grit	-0.018	-0.015

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Correlations are averaged across the ten plausible values, N=4454, the correlation between literacy and numeracy is 0.853.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Next, we examine whether personality traits contribute to explaining the level of cognitive ability after adjusting for education and demographic characteristics. First, we include in the model control variables only (results 1 and 4 in Table 11), second, control variables and the Big Five traits (results 2 and 5) and finally control variables and Grit (results 3 and 6). The models with personality traits explain very little additional variance in cognitive skill (less than 1%). When we look at the relationships between cognitive ability and personality traits after controlling for age, gender and education most of the relationships are insignificant. Only Extraversion and Conscientiousness retain their negative association with cognitive ability (the latter with the numeracy only). The direction of the Extraversion relationship is at odds with the results of Ackerman and Heggestad (1997) while the negative numeracy-Conscientiousness association confirms their findings.

Table 11. Relationship between literacy/numeracy and personality traits

	(1) Literacy (std)	(2) Literacy (std)	(3) Literacy (std)	(4) Numeracy (std)	(5) Numeracy (std)	(6) Numeracy (std)
Conscientiousness (std)		-0.086			-0.098*	
Extraversion (std)		-0.057*			-0.058*	
Agreeableness (std)		0.006			0.041	
Openness (std)		0.062			0.053	
Neuroticism (std)		-0.020			-0.027	
Grit (std)			-0.036			-0.036
Observations	4 454	4 454	4 454	4 454	4 454	4 454
R^2	0.260	0.267	0.262	0.236	0.242	0.237

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Control variables: age, age squared, gender, years of education. Ten plausible values. Literacy, numeracy and non-cognitive skills are standardised.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Overall, the results suggest that the measures of personality traits analysed and cognitive skills in PIAAC are very weakly related. The reason for that could be either poor measurement of personality traits or the fact that they indeed assess distinct constructs. The latter explanation suggests that inclusion of personality measures may have additional power of predicting life outcomes as they are another source of individual differences than cognitive skills. The relationship between personality traits and the important life outcomes will be the subject of investigation in the next section.

5. The relationship between non-cognitive skills and life outcomes

The previous section examined the psychometric properties of the scales. We now turn to those aspects of the scales which determine their analytical utility and their policy relevance. This section examines the relationship between non-cognitive skills and different aspects of individual and social well-being, including education, labour market participation, employment, job satisfaction, wages, health, interpersonal trust and overall life satisfaction. It also compares criterion validity of cognitive skills and non-cognitive skills. Finally, we provide insights into the incremental effect of Grit after controlling for the Big Five dimensions.

The methodological approach used to analyse the relationship between non-cognitive skills and life outcomes involves using standard multivariate methods such as ordered logit, binary logit and linear regression, depending on the nature of the outcome analysed. The specifications account for demographic and socio-economic characteristics. The limitation of this approach is that it estimates the association between personality traits and outcomes rather than isolating causal relationships. One way of dealing with reverse causality is to use the early assessment of traits to explain later life outcomes using prospective longitudinal data. Almlund et al. (2011) note that this approach does not necessarily solve the problem as the trait in question may evolve over time and, therefore, the outcome may be affected by its current value. However, there is some evidence that personality traits are stable over time (Cobb-Clark and Schurer, 2012). While the caveats concerning causal inferences based on cross-sectional data need to be kept in mind, the accumulated empirical evidence on this topic from longitudinal studies (see Roberts et al., 2007 for a review) strongly suggests that personality traits have an influence on a range of life outcomes.

This section uses the factor scores from the best fitted models, six-factor oblique for the BFI-S and bi-factor for Grit-S (see section 4). The additional sixth reverse factor is treated as an uninterpretable adjustment for the reverse items and is not included in the analysis. There are seven specifications for each outcome. The first one includes only control variables. The next three specifications separately consider the associations between the outcome with numeracy,[21] with Big Five and with Grit. Specifications 5 and 6 include numeracy and Big Five or Grit respectively, while the last column shows the incremental validity of Grit when controls are included for numeracy and the Big Five. The first part of this section focuses on the specifications 1-6 while the results on the incremental impact of Grit (specification 7) are covered in a separate subsection.

In the case of continuous outcomes, the variance explained is compared between the models. The non-linear probability models are compared based on analogues to ordinary least squares (OLS) R^2: McFadden's R^2 and McKelvey and Zavoina's R^2 as well as on information criteria.[22] Thanks to information criteria, we can identify the model which is more likely to generate the observed data. The complex survey design and sampling weights have been accounted for in the estimations of the parameters.

Most of the outcomes are analysed for all respondents with valid responses (n=4 454). However, this is not true for the economic outcomes. Employment is analysed for people active on the labour market and job quality outcomes (job satisfaction and wages) for employed persons with appropriate data. Categories of ordinal variables categories with a frequency below 5% were merged with the neighbouring category.

Table 12. Summary statistics of the outcomes

Variable	Obs	Mean	Std. dev.	Min	Max
Years of education	4 454	13.19	2.86	6	21
LM participation	4 454	0.72	0.45	0	1
Employment	3 356	0.93	0.25	0	1
Job satisfaction	3 026	1.98	0.65	1	3
Hourly wages (PLN)	2 059	17.18	10.30	4.9	83.3
Health	4 454	3.15	0.89	1	5
Trust	4 454	2.85	1.37	1	6
Life satisfaction	4 454	3.63	1.03	1	5

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Criterion validity

Educational attainment

Personality traits and cognitive skills are measured at the same point in time for the entire sample. This means that older individuals finished their formal education many years previously while the young are often still in education. Therefore, we can only examine whether personality helps to explain individual variation in educational attainment measured by completed years of schooling which is a censored measure for the younger cohorts. The variance explained by cognitive skills rivals that explained by measured personality traits (Table 13). Including personality traits in regressions with cognitive skills explains only a small amount of additional variance. Openness is associated with higher levels of education while Extraversion, Agreeableness and Neuroticism are associated with lower levels of education (specification 5). The absence of an effect for Conscientiousness is not in accordance with empirical literature. However, Grit is positively correlated with years of education.

Controls are included for socio-economic status (SES) which is believed to be one of the main determinants of educational attainment (Van Eijck and de Graaf, 2004). Without controlling for SES, the standardised effects of Big Five traits are around 1.5 times higher (see Table F.1 in Annex F).

Table 13. Educational attainment and cognitive and non-cognitive skills (linear regression)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Numeracy (std)		0.329***			0.322***	0.329***	0.319***
Conscientiousness (std)			-0.043		-0.015		-0.071
Extraversion (std)			-0.086**		-0.062**		-0.063**
Agreeableness (std)			-0.080		-0.083*		-0.060
Openness (std)			0.172***		0.149***		0.150***
Neuroticism (std)			-0.051**		-0.038*		-0.024
Grit (std)				0.082***		0.080***	0.096***
Observations	4 355	4 355	4 355	4 355	4 355	4 355	4 355
R^2	0.296	0.388	0.310	0.302	0.397	0.394	0.405

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Control variables: age, age squared, gender, father's education and mother's education. First PV (plausible value) for numeracy. Years of education, numeracy and non-cognitive skills are standardised.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Labour market outcomes

Several studies suggest that personality traits predict labour market outcomes. The decision to participate or not in the labour market appears to be related to individual personality traits. Controlling for education, age and gender, conscientious individuals are more likely to be active on the labour market while agreeable and neurotic individuals are less likely (Table 14). The relationship between cognitive skills as measured by PIAAC numeracy scores and labour force participation is positive but not statistically significant.

Among individuals active in the labour market, cognitive skills and personality traits are unrelated to employment after controlling for basic socio-demographic factors (Table 15). This lack of a clear association with employment can be partially explained by the differences in job search behaviours between individuals with internal and external locus of control – a trait linked to Neuroticism. People with internal locus of control tend to search for work more intensively. At the same time, they have higher reservation wages (Caliendo et al., 2015). Therefore, the effect on the duration of unemployment is ambiguous.

The analysis uses the International Labour Organization (ILO) definitions of activity in the labour market and employment. Slightly different results are obtained if we turn to the self-declared main activity of individuals (see Tables F.2 and F.3 in Annex F).

Table 14. Labour force participation and cognitive and non-cognitive skills (logit)

Odds ratios showing the likelihood of being active in the labour market

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Numeracy (std)		1.136			1.129	1.139	1.129
Conscientiousness (std)			1.557**		1.564**		1.560**
Extraversion (std)			0.949		0.957		0.957
Agreeableness (std)			0.717*		0.717*		0.718*
Openness (std)			1.129		1.123		1.123
Neuroticism (std)			0.745***		0.747***		0.748***
Grit (std)				1.112		1.115*	1.005
Observations	4 454	4 454	4 454	4 454	4 454	4 454	4 454
McFadden's R^2	0.253	0.254	0.267	0.254	0.268	0.256	0.268
McKelvey and Zavoina's R^2	0.385	0.389	0.405	0.386	0.408	0.390	0.408
BIC	18237506.8	18201025.9	17897266.8	18206185.1	17864854.9	18168266.9	17864815.7

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Control variables: age, age square, gender, years of education. First PV (plausible value) for numeracy. Numeracy and non-cognitive skills are standardised.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table 15. Employability and cognitive and non-cognitive skills (logit)

Odds ratios showing the likelihood of being employed

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Numeracy (std)		1.134			1.136	1.134	1.136
Conscientiousness (std)			1.442		1.441		1.457
Extraversion (std)			1.021		1.031		1.031
Agreeableness (std)			0.690		0.695		0.691
Openness (std)			0.973		0.965		0.965
Neuroticism (std)			1.081		1.083		1.080
Grit (std)				0.989		0.990	0.982
Observations	3 356	3 356	3 356	3 356	3 356	3 356	3 356
McFadden's R^2	0.0647	0.0661	0.0684	0.0647	0.0698	0.0661	0.0699
McKelvey and Zavoina's R^2	0.136	0.138	0.142	0.136	0.145	0.139	0.146
BIC	6729088.6	6719116.4	6702100.8	6728988.9	6692035.8	6719035.7	6691793.0

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Control variables: age, age squared, gender, years of education. First PV (plausible value) for numeracy. Numeracy and non-cognitive skills are standardised.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table 16. Wages and cognitive and non-cognitive skills (linear regression)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Numeracy (std)		0.057**			0.053**	0.056**	0.052**
Conscientiousness (std)			0.114**		0.115**		0.133***
Extraversion (std)			0.006		0.007		0.007
Agreeableness (std)			-0.135***		-0.131***		-0.137***
Openness (std)			0.002		0.000		-0.001
Neuroticism (std)			-0.039**		-0.036*		-0.041**
Grit (std)				-0.022		-0.018	-0.029*
Observations	2 059	2 059	2 059	2 059	2 059	2 059	2 059
R^2	0.371	0.380	0.384	0.373	0.391	0.381	0.394

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Control variables: age, age squared, gender, years of education, 1-digit ISCO. ISCO=0 excluded, top and bottom 1% of wage distribution excluded. First PV (plausible value) for numeracy. Logarithm of wages. Numeracy and non-cognitive skills are standardised.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table 17. Job satisfaction and cognitive and non-cognitive skills (ordered logit)

Odds ratios showing the likelihood of reporting higher level of job satisfaction comparing to lower level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Numeracy (std)		1.092			1.110	1.112	1.122
Conscientiousness (std)			1.564**		1.566**		1.414*
Extraversion (std)			1.279*		1.282*		1.288*
Agreeableness (std)			0.773		0.776		0.805
Openness (std)			1.015		1.012		1.019
Neuroticism (std)			0.784***		0.784***		0.806**
Grit (std)				1.302***		1.309***	1.202***
Observations	3 026	3 026	3 026	3 026	3 026	3 026	3 026
McFadden's R^2	0.0126	0.0133	0.0317	0.0212	0.0327	0.0222	0.0363
McKelvey and Zavoina's R^2	0.0280	0.0295	0.0699	0.0470	0.0721	0.0492	0.0802
BIC	26 469091.8	26450447.6	25956496.4	26238279.2	25930227.0	26211312.1	25833061.6

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ Control variables: age, age squared, gender, years of education, 1-digit ISCO. ISCO=0 excluded. First PV (plausible value) for numeracy. Numeracy and non-cognitive skills are standardised.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Once a person is employed, the important question is the quality of their job. The most-often used indicators of job quality are wages and job satisfaction. The Big Five traits explain as much of the variation in wages as cognitive skills (Table 16). Adding personality traits to the model with cognitive skills improves it slightly (1% of variance). Conscientious individuals are more likely to earn more. On the contrary, Agreeableness and Neuroticism are associated with lower wages. These results are in line with empirical studies on Big Five from other countries. The effect size of personality traits on wages are higher than the effect of numeracy – e.g. one standard deviation increase in Conscientiousness is associated with 11.5% increase in wages while it is 5.3% for numeracy. Cognitive skills might affect wages indirectly via choice of occupation. When we compare the influence of cognitive and non-cognitive skills on wages without controlling for occupation the effect of Conscientiousness on wages is still higher than the effect of numeracy but the difference decreases (10.8% and 7.2% increase in wages respectively).Grit is unrelated to wages in our sample.

The next dimension of job quality examined is job satisfaction which is by definition subjective. Results in Table 17 show that personality traits do better at predicting job satisfaction than do cognitive skills. Conscientiousness and Extraversion are linked to higher job satisfaction while Neuroticism is linked to lower job satisfaction. Grit is associated with higher levels of job satisfaction.

Social outcomes

In addition to examining the relationships of personality measures to educational and labour market outcomes, it is also useful to examine their relationship to social outcomes. Following OECD (2007, 2013a), interpersonal trust and health are considered as social outcomes.

Measures of fit provides strong support for an association of health with personality traits, especially with Big Five (see specifications 2-4 in Table 18). In line with the earlier research on PIAAC (da Costa et al., 2014; OECD, 2013a), numeracy proficiency is positively linked to health. Lower levels of Conscientiousness and Extraversion are associated with lower levels of health while lower levels of Neuroticism are associated with higher level of health. A standard deviation increase in Grit, results in an increase of the odds of reporting better health by a factor of 1.34, holding all other variables constant.

Table 18. Health and cognitive and non-cognitive skills (ordered logit)

Odds ratios showing the likelihood of reporting higher level of health comparing to lower level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Numeracy (std)		1.173**			1.186***	1.192***	1.193***
Conscientiousness (std)			1.380*		1.401**		1.219
Extraversion (std)			1.177**		1.186**		1.184**
Agreeableness (std)			0.866		0.859		0.908
Openness (std)			0.919		0.917		0.921
Neuroticism (std)			0.689***		0.690***		0.712***
Grit (std)				1.340***		1.350***	1.267***
Observations	4 454	4 454	4 454	4 454	4 454	4 454	4 454
McFadden's R^2	0.144	0.146	0.159	0.153	0.161	0.155	0.166
McKelvey and Zavoina's R^2	0.340	0.343	0.367	0.357	0.372	0.361	0.381
BIC	45460921.1	45355654.6	44704685.6	45004451.1	44586321.7	44878268.6	44332605.4

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Control variables: age, age squared, gender, years of education, employment status. First PV (plausible value) for numeracy. Numeracy and non-cognitive skills are standardised.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table 19. Interpersonal trust and cognitive and non-cognitive skills (ordered logit)

Odds ratios showing the likelihood of reporting higher level of trust comparing to lower level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Numeracy (std)		1.217***			1.205***	1.208***	1.201***
Conscientiousness (std)			0.863		0.876		0.978
Extraversion (std)			1.114		1.126		1.124
Agreeableness (std)			1.078		1.071		1.025
Openness (std)			0.897		0.891		0.890
Neuroticism (std)			0.829***		0.832***		0.809***
Grit (std)				0.826***		0.830***	0.823***
Observations	4 454	4 454	4 454	4 454	4 454	4 454	4 454
McFadden's R^2	0.0118	0.0145	0.0164	0.0151	0.0189	0.0177	0.0219
McKelvey and Zavoina's R^2	0.0353	0.0439	0.0494	0.0459	0.0572	0.0537	0.0664
BIC	62691017.0	62514779.5	62394334.4	62480288.2	62236068.6	62316837.8	62050332.6

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Control variables: age, age squared, gender, years of education, employment status. First PV (plausible value) for numeracy. Numeracy and non-cognitive skills are standardised.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

The model with the Big Five fits the observed data about trust best (specifications 2-4 Table 19). As with health, numeracy is positively correlated with trust. Neuroticism is the only Big Five trait associated with trust. Neurotic people are more likely to report lower levels of trust. Similarly, Grit is negatively associated with trust.

Life satisfaction

In contrast to health and trust, life satisfaction is not related to cognitive skills. However, the associations with personality traits are very strong. For a standard deviation increase in Conscientiousness, the odds of reporting higher life satisfaction increase by 71%, holding all other variables constant. Also Extraversion has a positive relationship with life satisfaction while Neuroticism a negative one. Grit is strongly positively related to life satisfaction.

Table 20. Life satisfaction and cognitive and non-cognitive skills (ordered logit)

Odds ratios showing the likelihood of reporting higher level of life satisfaction comparing to lower level

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Numeracy (std)		0.956			0.981	0.983	0.996
Conscientiousness (std)			1.712***		1.710***		1.412**
Extraversion (std)			1.269***		1.268**		1.276**
Agreeableness (std)			0.824		0.824		0.878
Openness (std)			0.992		0.992		0.998
Neuroticism (std)			0.766***		0.766***		0.802***
Grit (std)				1.635***		1.634***	1.457***
Observations	4 454	4 454	4 454	4 454	4 454	4 454	4 454
McFadden's R^2	0.0214	0.0215	0.0473	0.0446	0.0473	0.0446	0.0591
McKelvey and Zavoina's R^2	0.0601	0.0605	0.128	0.124	0.128	0.124	0.159
BIC	55020911.7	55011553.3	53562592.2	53718224.5	53560947.2	53716829.1	52899931.9

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Control variables: age, age squared, gender, years of education, employment status. First PV (plausible value) for numeracy. Numeracy and non-cognitive skills are standardised.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

The incremental validity of Grit

The construct of Grit is often related conceptually to the Big Five factor of Conscientiousness and the empirical correlations between the two are high (see section 3). An important question, therefore, is whether Grit provides some extra information when we control for the Big Five traits. The correlation between Grit and Conscientiousness in the postPIAAC sample is 0.37 indicating that 14% of variation in scores across these scales is shared.

In non-nested models (specifications 3 and 4 in the previous subsection), the direction of the relationships between Grit and Conscientiousness and the outcomes of interest is the same when both are significant. But what is the incremental validity of Grit after controlling for the Big Five? For the continuous outcomes the comparison of nested specifications 5 and 7 (Table 13 and Table 16) from the previous subsection indicates whether the relationship of Big Five traits with outcomes changes after adjusting for Grit. In the non-linear probability models, this comparison cannot be easily interpreted in this way. As the logit models are non-linear, the change in parameters in the nested models after controlling for additional variable is partially due to confounding and partially due to rescaling of coefficients (Karlson, Holm and Breen, 2012). In order to deal with the problem, we will use the method proposed by Karlson, Holm and Breen (2012) which permits the decomposition of these effects and interpretation of the effect differences in parameters in line with those in linear models.

Looking at the full specifications adjusted for the Big Five factors, Grit is positively associated with educational attainment, job satisfaction, health and life satisfaction while negatively associated with trust. However, it is unrelated to the other labour market outcomes: labour force participation, employment and wages.

Does Grit confound the effect of the Big Five on outcomes? First, looking at the continuous outcomes, when Grit is added to the model predicting educational attainment, the relationship with Agreeableness and Neuroticism becomes insignificant (Table 13). In contrast, controlling for Grit in the wage equation does not change the significance of the associations with the Big Five traits and Grit does not improve the model (Table 16). Surprisingly, adding Grit strengthens the positive effect of Conscientiousness on wages even though it has a negative effect itself.

In order to deal with the problem, we will use the KHB method proposed by Karlson, Holm and Breen (2012) which permits the decomposition of these effects and interpretation of the effect differences in parameters in line with those in linear models. For the non-linear models, the KHB method confirms the results of the simple comparison of changes in parameters after controlling for Grit: Conscientiousness is mediated more strongly by Grit than by any other Big Five trait (Table F.4 in Annex F). Concerning the other factors from the Big Five model, their effect on the given outcome due to Grit is not statistically significant. The percentage reduction which measures the percentage change in the coefficient of each Big Five factor attributable to confounding net of rescaling is lower than 5% for all significant trait-outcome relationships except for Conscientiousness and Neuroticism. It means that adding Grit to the model does not change the relationship between the outcomes and the other Big Five traits: Openness, Extraversion and Agreeableness.

Describing the effects in more detail, 25% of the total effect of Conscientiousness on job satisfaction net of rescaling is due to Grit, yet its effect is still significant. In case of health 41% of the total effect of Conscientiousness is due to Grit and its effect becomes insignificant after controlling for Grit. In contrast, the effect of Conscientiousness on life satisfaction is still significant after controlling for Grit which accounts for 40% of the effect. In case of Neuroticism the mediation effect is weaker. 11% and 19% of the total effect of this trait on job satisfaction and life satisfaction respectively is due to Grit and the relationships of Neuroticism to these outcomes remain significant after account is taken of Grit.

Do personality measures moderate the effect of cognitive skills on relevant outcomes?

Cognitive skills are positively correlated with some of the analysed outcomes, i.e. educational attainment, wages, health and trust (see previous subsections). One would expect that these effects might vary depending on the personality of an individual. We include interactions between the cognitive and non-cognitive skills in order to capture how non-cognitive skills potentially moderate the effect of cognitive skills on outcomes.

Non-cognitive skills moderate the relationships between cognitive skills and two outcomes: labour force participation and wages. These effects are consistent – for both literacy and numeracy. The more neurotic a worker is the lower the returns to cognitive skills. It would be expected that individuals more vulnerable to stress, less self-confident and worrying might not be able to fully put their skills to the productive use. The relationship between cognitive skills and labour force participation is moderated by Conscientiousness. The more conscientious the person, the stronger is the relationship between cognitive skills and labour force participation. The potential explanation could be that conscientious individuals have stronger need of using their higher skills on the labour market.

The evidence presented in this section suggests that non-cognitive skills are significantly related to meaningful life outcomes. Table 21 summarises the main results. Conscientiousness is positively linked to most of the outcomes while Neuroticism shows a clear negative association with all of the outcomes except for employment. Extraversion is positively related to self-rated outcomes such as job or life satisfaction and health while Openness is associated only with educational attainment. The results also confirm a negative relationship of Agreeableness and wages. Grit is also significantly related to most of these outcomes, but adding it to the specification does not change the relationships with Big Five traits for most of them. Most of the results also hold when simple sum scores by factor are used in the criterion validity analysis (see Table F.5 in Annex F).

Table 21. Summary of skill / trait– outcome relationships (factor scores)

	Education		LF participation		Employability		Wages	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Numeracy (std)	+++	+++					++	++
Conscientiousness (std)			++	++			++	+++
Extraversion (std)	--							
Agreeableness (std)	-		-	-			---	---
Openness (std)	+++	+++						
Neuroticism (std)	-		---	---			-	--
Grit (std)	x	+++	x		x		x	-

	Job satisfaction		Health		Trust		Life satisfaction	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Numeracy (std)			+++	+++	+++	+++		
Conscientiousness (std)	++	+	++				+++	++
Extraversion (std)	+	+	++	++			++	++
Agreeableness (std)								
Openness (std)								
Neuroticism (std)	---	--	---	---	---	---	---	---
Grit (std)	x	+++	x	+++	x	---	x	+++

Notes: +/- $p < 0.05$, ++/-- $p < 0.01$, +++/--- $p < 0.001$ (1) Numeracy and Big Five; (2) Numeracy, Big Five and Grit.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

6. Conclusions

It is widely acknowledged that non-cognitive skills play an important role in determining meaningful life outcomes. However, a comprehensive investigation of all the measurement properties of the Big Five and Grit-S scales in a heterogeneous sample of adults has not been undertaken to date. The purpose of this analysis was to evaluate the properties of these non-cognitive scales measured by self-reporting in the Polish follow-up study on PIAAC and to provide information about the value of the inclusion of personality measures in a large scale assessment of skills. The results presented in this paper show that not all the criteria concerning the reliability, validity and comparability of these scales have been satisfied, though the personality measures significantly contribute to explaining the variability in policy-relevant outcomes.

Item-level analysis indicated that most of the questions discriminate well between people possessing a high and a low level of a given trait, though reverse-worded items perform weaker. This might be due to translation and adaptation problems that should be addressed in future studies. The reliability of scales is moderate but satisfactory given their length. Removing the problematic items improves the reliability of the scales.

The BFI-S theoretical five-factor structure was not replicated in the Polish adult population sample. However, there are indications in the literature that reverse-worded items form a separate factor. A six-factor model with an additional factor loading reverse-worded items fits the postPIAAC data. In case of Grit-S, the second-order theoretical structure does not hold in our sample, though a bi-factor model emerged that has an equivalent interpretation. The best fitting models were used in further multivariate analyses presented in the paper.

The scales are not fully measurement invariant. Items have the same meanings across the investigated groups and the correlations with latent traits can be compared between most of the groups, though care should be taken in directly comparing manifest scale scores. The results regarding discriminant validity suggest that the measures of personality traits analysed and the cognitive skills in the Survey of Adult Skills (PIAAC) are very weakly related and confirm that they assess distinct constructs.

Overall, the results confirm earlier findings from the literature that differences in personality traits are important in explaining differences in life outcomes. For most of the outcomes, the Big Five traits outperform cognitive skills in predictive power. Only educational attainment is more strongly related to cognitive skills, while for wages, the predictive power of personality and cognitive skills is similar. However, it has to be noted that due to time span between cognitive tests and measurement of the outcomes the effects might have been diluted. On the other hand, the results are striking if we take into account that the personality traits are measured with much lower precision than cognitive skills.

Including personality traits in the explanatory models with cognitive skills improves the model fit only slightly. In most of the cases the effect does not exceed 1% of the additional variance explained. The stronger effect is observed for subjective, self-declared outcomes such as health or job satisfaction.

The results presented describe relationships between non-cognitive skills and life outcomes, though drawing any conclusions about causality is not possible with the available data. We are aware that the lack of a causal model is a serious limitation for designing policy interventions and for policy analysis. Nevertheless, the results stress the importance of the topic and might motivate further work on appropriate methods.

The effects of particular traits are largely in agreement with predictions. They confirm similar effects observed in previous studies conducted in the United States and Western Europe. Conscientiousness is positively related to most of the outcomes analysed while Neuroticism has a negative relationship. Extraverted individuals are more likely to finish formal education earlier. They are also more satisfied with their life and job and feel healthier. When it comes to investigating the effects of Agreeableness it is associated with lower levels of education and negative labour market outcomes. Openness is strongly and positively related to educational attainment. It is worth noting that even weak effects of personality traits on meaningful outcomes might be important due to their cumulative effects across a person's life (Roberts et al., 2007).

As expected, the Grit-S scale has less predictive power than the Big Five traits. It is incrementally valid for most of the outcomes in that it explains the variance above and beyond the variance explained by the Big Five traits. When both scales are included in a regression, the gains are marginal. For most of the outcomes, the direction and strength of the relationship of Grit to life outcomes is similar to that of Conscientiousness. Taking into account the length of the scales (eight and three items for Grit and Conscientiousness respectively), the use of the Big Five model which covers different aspects of personality seems to be preferable to the use of Grit.

In sum, given the potential benefits and relatively small burden on respondents in terms of required time it seems advisable to incorporate measures of personality traits into competence surveys. The analysis of psychometric properties showed the importance of careful national adaptations with a special focus on negatively-worded items.

Notes

[1] These have included studies on school-aged children: PISA (Programme for International Student Assessment), TIMSS (Trends in International Mathematics and Science Study), PIRLS (Progress in International Reading Literacy Study); and on the adult population: IALS (International Adult Literacy Survey), ALL (Adult Literacy and Lifeskills Survey), PIAAC.

[2] Salkever (2015) shows that the impact of personality reported by Borghans et al. (2011) is weaker under rigorous re-analysis.

[3] Only these two factors were measured in the Italian PIAAC follow-up.

[4] The authors check for convergent and discriminant validity with respect to NEO-PI-R and Coping Inventory for Stressful Situations (CISS).

[5]A summary is provided by Rammstedt, Goldberg and Borg (2010).

[6] Factor loadings, intercepts and error variances are constrained to be equal between groups.

[7] We can analyse discrepancy of observable item response patterns and item characteristic curves predicted by the model (Annex C and D).

[8] The analysis is performed in Stata 13.1 with svy module. IRT analysis is performed with uirt.ado package (Kondratek, 2016).

[9] Descriptive statistics are obtained in Stata 13.1 with svy module, IRT analysis is performed with uirt.ado package (Kondratek, 2016).

[10] These are: tau-equivalence (measuring construct with the same units, the same precision and possibly different error variance) or essential tau-equivalence (same units, different precision, different amount of error), uncorrelated errors and normality (Sheng and Sheng, 2012).

[11] Analyses were performed in R (R Core Team, 2013) with *psych* package (Revelle, 2015).

[12] This value should be the cut-off for basic research. For applied settings reliability should be higher than 0.9 and 0.95 should be the desirable standard (Nunnally, 1978; Lance, Butts and Michels, 2006).

[13] For categorical variables.

[14] For ISCED 1-2 level (CFI=0.506, TLI=0.424, RMSEA=0.113), for ISCED 3-4 level (CFI=0.533, TLI=0.455, RMSEA=0.104), for ISCED 5-6 level (CFI=0.683, TLI=0.631, RMSEA=0.089).

[15] It is also possible to perform the procedure backwards – from fully constrained model to the model with no constraints. Nevertheless, analysis has to be performed stepwise.

[16] We decided to recode "age" variable in order to perform multigroup CFA properly (the model requires categorical grouping variables). Although measurement invariance analysis can be performed for continuous variables in MIMIC model, the method can detect only uniform bias only (see Barendse, Oort and Garst, 2010; Kim, Yoon and Lee, 2012; Woods, 2009).

[17] Some of the variables had to be recoded, as the sample size of some categories was too small to perform measurement invariance analysis. Age was recoded basing on quartiles; education level into 2 groups (ISCED 1-4 and ISCED 5-6) and literacy and numeracy also into 2 groups (levels 0-1 – group 1, 2-5 – group 2).

[18] The difference in CFI value between base model (no constraints) and model with measurement invariance constraints.

[19] Testing measurement invariance for Grit-S and BFI-S questions as categorical variables resulted in problems with model convergence.

[20] Configural model is used as the base model. RMSEA values for all the grouping variables are between 0.055-0.06, CFI values between 0.88-0.89.

[21] The results with literacy are not qualitatively different (results available upon request).

[22] As concluded by Long and Freese (2001: 148) *" (…) for ordinal outcomes, McKelvey and Zavonia's R^2 most closely approximates the R^2 obtained by estimating the linear regression model on the underlying latent variable"*.

REFERENCES

Ackerman, P. L. and E. D. Heggestad (1997), "Intelligence, personality, and interests: Evidence for overlapping traits", *Psychological Bulletin*, Vol. 121/2, p. 219.

Almlund, M., A. L. Duckworth, J. Heckman, and T. Kautz (2011), "Personality psychology and economics", in *Handbook of Economics of Education*, Hanushek, E. A., S. J Machin and L. Woessmann (eds), Vol. 4, Elsevier BV.

Barendse, M., F. Oort and G. Garst (2010), "Using restricted factor analysis with latent moderated structures to detect uniform and nonuniform measurement bias; a simulation study", *AStA Advances in Statistical Analysis*, Vol. 94/2, pp. 117–127.

Barrick, M. R. and M. K. Mount (1991), "The big five personality dimensions and job performance: A meta-analysis", *Personnel Psychology*, Vol. 44/1, pp. 1–26.

Borghans, L., A. L. Duckworth, J.J. Heckman, and B. Ter Weel (2008), "The economics and psychology of personality traits", *Journal of Human Resources*, Vol. 43/4, pp. 972–1059.

Borghans, L., B. H. Golsteyn, J. Heckman and J. E. Humphries (2011), "Identification problems in personality psychology", *Personality and Individual Differences*, Vol. 51/3, pp. 315–320.

Borghans, L., H. Meijers and B. Ter Weel (2008), "The role of noncognitive skills in explaining cognitive test scores", *Economic Inquiry*, Vol. 46/1, pp. 2–12.

Borsboom, D. (2006), "When does measurement invariance matter?", *Medical Care*, Vol. 44/11, S176-S181.

Caliendo, M., D. A. Cobb-Clark and A. Uhlendorff (2015), "Locus of control and job search strategies", *Review of Economics and Statistics*, Vol. 97/1, pp. 88–103.

Canivez, G. L. (2016), "Bifactor modeling in construct validation of multifactored tests: Implications for understanding multidimensional constructs and test interpretation" in K. Schweizer and C. DiStefano (Eds.), *Principles and Methods of Test Construction: Standards and Recent Advancements*, Hogrefe Publishers, Göttingen, Germany, www.ux1.eiu.edu/~glcanivez/Adobe%20pdf/Publications-Papers/Canivez%20%282016%29%20Bifactor%20Modeling.pdf.

Chan, D. (2009), "So why ask me? Are self-report data really that bad", *Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in the Organizational and Social Sciences*, pp. 309–336.

Chen, F. F., S.G. West and K. H. Sousa (2006), "A comparison of bifactor and second-order models of quality of life", *Multivariate Behavioral Research*, Vol. 41/2, pp. 189–225.

Cheung, G. W and R. B. Rensvold (2002), "Evaluating goodness-of-fit indexes for testing measurement invariance", *Structural Equation Modeling*, Vol. 9/2, pp. 233–255.

Cobb-Clark, D. A. and S. Schurer (2012), "The stability of big-five personality traits", *Economics Letters*, Vol. 115/1, pp. 11–15.

Cortina, J. M. (1993), "What is coefficient alpha? An examination of theory and applications", *Journal of Applied Psychology*, Vol. 78/1, p. 98.

Costa, P. T. and R. R. McCrae (1992), "Four ways five factors are basic", *Personality and Individual Differences*, Vol. 13/6, pp. 653–665.

Costa, P. T and R. R. McRae (1985), *NEO Personality Inventory–Form R*.

Credé, M., M. C. Tynan and P. D. Harms (2016), "Much ado about grit: A meta-analytic synthesis of the grit literature", *Journal of Personality and Social Psychology*.

Cronbach, L. J. (1951), "Coefficient alpha and the internal structure of tests", *Psychometrika*, Vol. 16/3, pp. 297–334.

Cunha, F. J. and J. J Heckman (2008), "Formulating, identifying and estimating the technology of cognitive and noncognitive skill formation", *Journal of Human Resources*, Vol. 43/4, pp. 738–782.

Cunningham, W., M. P. Torrado and M. Sarzosa (2016), "Cognitive and non-cognitive skills for the Peruvian labor market: Addressing measurement error through latent skills estimations", *World Bank Policy Research Working Paper*, 7550.

da Costa, P. D., M. Rodrigues, E. Vera-Toscano and A. Weber (2014), *Education, Adult Skills and Social Outcomes: Empirical Evidence from the Survey on Adult Skills (PIAAC 2013)*, Publications Office of the European Union, Luxembourg.

Dalal, D. K. and N. T. Carter (2015), "Negatively worded items negatively impact survey research", in *More Statistical and Methodological Myths and Urban Legends*, pp. 112–132.

DeNeve, K. M. and H. Cooper (1998), "The happy personality: A meta-analysis of 137 personality traits and subjective well-being", *Psychological Bulletin*, Vol. 124/2, p. 197.

Di Francesco, G., F. Roma, O. Giancola, S. Mineo, M. Bastianelli and F. Fraccaroli (2015), "The relationship between competences and non-cognitive dimensions. The experience of the "PIAAC Italia" survey", *Osservatorio Isfol*, Vol. 3, pp. 123–144.

DiStefano, C. and R. W. Motl (2006), "Further investigating method effects associated with negatively worded items on self-report surveys", *Structural Equation Modeling*, Vol. 13/3, pp. 440–464.

Dohmen, T., A. Falk, D. Huffman and U. Sunde (2008), "Representative trust and reciprocity: Prevalence and determinants", *Economic Inquiry*, Vol. 46/1, pp. 84–90.

Drago, F. (2011), "Self-esteem and earnings", *Journal of Economic Psychology*, Vol. 32/3, pp. 480–488.

Duckworth, A. L., C. Peterson, M. D. Matthews and D. R. Kelly (2007), "Grit: Perseverance and passion for long-term goals", *Journal of Personality and Social Psychology*, Vol. 92/6, pp. 1087–1101, http://doi.org/10.1037/0022-3514.92.6.1087.

Duckworth, A. L. and P. D. Quinn (2009), "Development and validation of the Short Grit Scale (GRIT-S)", *Journal of Personality Assessment*, Vol. 91/2, pp. 166–174.

Eskreis-Winkler, L., E. P. Shulman, S. A. Beal and A. L. Duckworth (2014), "The grit effect: Predicting retention in the military, the workplace, school and marriage", *Frontiers in Psychology*, Vol. 5.

French, B. F. and W. H. Finch (2008), "Multigroup confirmatory factor analysis: Locating the invariant referent sets", *Structural Equation Modeling*, Vol. 15/1, pp. 96–113.

Gerlitz, J.-Y., and J. Schupp (2005), "Zur Erhebung der Big-Five-basierten persoenlichkeitsmerkmale im SOEP", *DIW Research Notes*, Vol. 4.

Gignac, G. E. and M. W. Watkins (2013), "Bifactor modeling and the estimation of model-based reliability in the WAIS-IV", *Multivariate Behavioral Research*, Vol. 48/5, pp. 639–662, http://doi.org/10.1080/00273171.2013.804398.

Goldberg, L. R., D. Sweeney, P. F. Merenda and J. E. Hughes (1998), "Demographic variables and personality: The effects of gender, age, education, and ethnic/racial status on self-descriptions of personality attributes", *Personality and Individual Differences*, Vol. 24/3, pp. 393–403.

Gustafsson, J.-E., and G. Balke (1993), "General and specific abilities as predictors of school achievement", *Multivariate Behavioral Research*, Vol. 28/4, pp. 407–434.

Hahn, E., J. Gottschling and F. M. Spinath (2012), "Short measurements of personality – Validity and reliability of the GSOEP Big Five Inventory (BFI-S)", *Journal of Research in Personality*, Vol. 46/3, pp. 355–359, http://doi.org/10.1016/j.jrp.2012.03.008.

Hambleton, R. K., and L. Patsula (1999), "Increasing the validity of adapted tests: Myths to be avoided and guidelines for improving test adaptation practices", *Association of Test Publishers*, Vol. 1/1, pp. 1-13.

Hampson, S. E., L. R. Goldberg, T. M. Vogt and J. P. Dubanoski (2007), "Mechanisms by which childhood personality traits influence adult health status: Educational attainment and healthy behaviors", *Health Psychology*, Vol. 26/1, p.121.

Heckman, J. J. and T. Kautz (2012), "Hard evidence on soft skills", *Labour Economics*, Vol. 19/4, pp. 451-464.

Heckman, J. J., R. Pinto and P. A. Savelyev (2013), "Understanding the mechanisms through which an influential early childhood program boosted adult outcomes", *The American Economic Review*, Vol. 103/6, pp. 1–35.

Heckman, J., S. H. Moon, R. Pinto, P. A. Savelyev and A. Yavitz (2010a), "The rate of return to the HighScope Perry Preschool Program", *Journal of Public Economics*, Vol. 94/1, pp. 114–128.

Heckman, J., S. H. Moon, R. Pinto, P. Savelyev and A. Yavitz (2010b), "Analyzing social experiments as implemented: A reexamination of the evidence from the HighScope Perry Preschool Program", *Quantitative Economic*s, Vol. 1/1, pp. 1–46.

Heineck, G. and S. Anger, S. (2010), "The returns to cognitive abilities and personality traits in Germany", *Labour Economics*, Vol. 17/3, pp. 535–546.

Herche, J. and B. Engelland (1996), "Reversed-polarity items and scale unidimensionality", *Journal of the Academy of Marketing Science*, Vol. 24/4, pp. 366–374.

Ivcevic, Z. and M. Brackett (2014), "Predicting school success: comparing conscientiousness, grit, and emotion regulation ability", *Journal of Research in Personality*, Vol. 52, pp. 29–36.

John, O. P., E. M. Donahue and R. L. Kentle (1991), "*The big five inventory—versions 4a and 54"*, University of California, Berkeley, Institute of Personality and Social Research, Berkeley, CA.

John, O. P. and S. Srivastava (1999), "The Big Five trait taxonomy: History, measurement, and theoretical perspectives", *Handbook of Personality: Theory and Research*, Vol. 2, pp. 102–138.

Judge, T. A., D. Heller and M. K. Mount (2002), "Five-factor model of personality and job satisfaction: A meta-analysis", *Journal of Applied Psychology*, Vol. 87/3, p. 530.

Karlson, K. B., A. Holm and R. Breen (2012), "Comparing regression coefficients between same-sample nested models using logit and probit a new method", *Sociological Methodology*, Vol. 42/1, pp. 286-313.

Kim, E. S., M. Yoon and T. Lee (2012), "Testing measurement invariance using MIMIC likelihood ratio test with a critical value adjustment", *Educational and Psychological Measurement*, Vol. 72/3, pp.469–492.

Kondratek, B. (2016), UIRT: Stata module to fit unidimensional Item Response Theory models, Statistical Software Components S458247, Boston College Department of Economics.

Körner, A., M. Geyer and E. Brähler (2002), "Das NEO-Fünf-Faktoren Inventar", *Diagnostica*, Vol. 48/1, pp. 19–27.

Lance, C. E., M. M. Butts and L. C. Michels (2006), "The sources of four commonly reported cutoff criteria what did they really say?", *Organizational Research Methods*, Vol. 9/2, pp. 202–220.

Lang, F. R., D. John, O. Lüdtke, J. Schupp and G. G. Wagner (2011), "Short assessment of the Big Five: Robust across survey methods except telephone interviewing", *Behavior Research Methods*, Vol. 43/2, pp. 548–567.

Lang, F. R., O. Lüdtke and J. B Asendorpf (2001), "Testgüte und psychometrische Äquivalenz der deutschen Version des Big Five Inventory (BFI) bei jungen, mittelalten und alten Erwachsenen", *Diagnostica*, Vol. 47/3, pp. 111–121.

Long, J. S., and J. Freese (2001), *Regression Models for Categorical Dependent Variables Using Stata*, Stata press.

MacCallum, R. C., M. Roznowski and L. B. Necowitz (1992), "Model modifications in covariance structure analysis: The problem of capitalization on chance", *Psychological Bulletin*, Vol. 111/3, p. 490.

Marsh, H. W. (1986), "Negative item bias in ratings scales for preadolescent children: A cognitive-developmental phenomenon", *Developmental Psychology*, Vol. 22/1, p. 37.

McCrae, R. R. and Jr, P. T. Costa (1999), "A five-factor theory of personality", *Handbook of Personality: Theory and Research*, Vol. 2, pp. 139–153.

McCrae, R. R., J. E. Kurtz, S. Yamagata and A. Terracciano (2010), "Internal consistency, retest reliability, and their implications for personality scale validity", *Personality and Social Psychology Review*.

Mellenbergh, G. J. (1989), "Item bias and item response theory", *International Journal of Educational Research*, Vol. 13/2, pp. 127–143.

Mendolia, S. and I. Walker (2014), "Do NEETs need grit?", *IZA Discussion Paper*, No. 8740.

Meredith, W. (1993), "Measurement invariance, factor analysis and factorial invariance", *Psychometrika*, Vol. 58/4, pp. 525–543.

Millsap, R. E. (2007), "Invariance in measurement and prediction revisited", *Psychometrika*, Vol. 72/4, pp. 461–473.

Moss, P. A. (1994), "Can there be validity without reliability?", *Educational Researcher*, Vol. 23/2, pp. 5-12.

Mottus, R., J. Allik and H. Pullmann (2007), "Does personality vary across ability levels? A study using self and other ratings", *Journal of Research in Personality*, Vol. 41/1, pp. 155–170.

Mueller, G. and E. Plug (2006), "Estimating the effect of personality on male and female earnings", *Industrial & Labor Relations Review*, Vol. 60/ 1, pp. 3 22, http://doi.org/10.1177/001979390606000101.

Mullen, M. R. (1995), "Diagnosing measurement equivalence in cross-national research", *Journal of International Business Studies*, pp. 573–596.

Nunnally, J. (1978), *C.(1978), Psychometric Theory*, McGraw-Hill, New York.

Nyhus, E. K. and E. Pons (2005), "The effects of personality on earnings", *Journal of Economic Psychology*, Vol. 26/3, pp. 363–384, http://doi.org/10.1016/j.joep.2004.07.001.

O'Connell, M. and H. Sheikh (2011), "Big Five personality dimensions and social attainment: Evidence from beyond the campus", *Personality and Individual Differences*, Vol. 50/6, pp. 828–833.

OECD (2013a), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264204256-en.

OECD (2013b), *Technical Report of the Survey of Adult Skills (PIAAC)*, OECD. www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf

OECD (2007), *Understanding the Social Outcomes of Learning*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264034181-en.

Poropat, A. E. (2009), "A meta-analysis of the five-factor model of personality and academic performance", *Psychological Bulletin*, Vol. 135/2, p. 322.

Prevoo, T. and B. ter Weel (2015), "The importance of early conscientiousness for socio-economic outcomes: Evidence from the British Cohort Study", *Oxford Economic Papers*, Vol. 67/4, pp. 918-948.

R Core Team. (2013), *R Core Team. 2013. R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, www.R-project.org.

Rammstedt, B., D. Danner and S. Martin (2016), "The association between personality and cognitive ability: Going beyond simple effects", *Journal of Research in Personality*, Vol. 62, pp. 39–44.

Rammstedt, B., L. R. Goldberg and I. Borg (2010), "The measurement equivalence of Big-Five factor markers for persons with different levels of education", *Journal of Research in Personality*, Vol. 44/1, pp. 53–61, http://doi.org/10.1016/j.jrp.2009.10.005.

Rammstedt, B. and O. P. John (2007), "Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German", *Journal of Research in Personality*, Vol. 41/1, pp. 203–212, http://doi.org/10.1016/j.jrp.2006.02.001.

Reise, S. P. (1990), "A comparison of item-and person-fit methods of assessing model-data fit in IRT", *Applied Psychological Measurement*, Vol. 14/2, pp. 127–137.

Reise, S. P., T. M. Moore and M. G. Haviland (2010), "Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores", *Journal of Personality Assessment*, Vol. 92/6, pp. 544–559.

Revelle, W. (2015), *Package "psych."*, ftp://mozilla.c3sl.ufpr.br/CRAN/web/packages/psych/psych.pdf.

Revelle, W. and R. E. Zinbarg (2009), "Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma", *Psychometrika*, Vol. 74/1, pp. 145–154.

Roberts, B. W., N. R. Kuncel, R. Shiner, A. Caspi, and L. R. Goldberg (2007), "The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes", *Perspectives on Psychological Science*, Vol. 2/4, pp. 313–345.

Rolland, J. P., W. D. Parker and H. Stumpf (1998), "A psychometric examination of the French translations of NEO-PI-R and NEO-FFI", *Journal of Personality Assessment*, Vol. 71/2, pp. 269-291.

Salkever, D. (2015), "Interpreting the NLSY79 empirical data on "IQ" and "achievement": A comment on Borghans et al.'s "Identification problems in personality psychology", *Personality and Individual Differences*, Vol. 85, pp. 66–68.

Samejima, F. (1969), "Estimation of latent ability using a response pattern of graded scores", *Psychometrika Monograph Supplement*.

Schmitt, N. (1996), "Uses and abuses of coefficient alpha", *Psychological Assessment*, Vol. 8/4, p. 350.

Schriesheim, C. A. and R. J. Eisenbach (1995), "An exploratory and confirmatory factor-analytic investigation of item wording effects on the obtained factor structures of survey questionnaire measures", *Journal of Management*, Vol. 21/6, pp. 1177–1193.

Segal, C. (2006), *Motivation, Test Scores and Economic Success*, https://repositori.upf.edu/bitstream/handle/10230/897/1124.pdf?sequence=1

Sheng, Y. and Z. Sheng (2012), "Is coefficient alpha robust to non-normal data?", *Frontiers in Psychology*, Vol. 3, http://doi.org/10.3389/fpsyg.2012.00034.

Sijtsma, K. (2009), "On the use, the misuse, and the very limited usefulness of Cronbach's alpha", *Psychometrika*, Vol. 74/1, pp. 107–120.

Specht, J., B. Egloff and S. C. Schmukle (2011), "Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five", *Journal of Personality and Social Psychology*, Vol. 101/4, p. 862.

Steinmetz, H., P. Schmidt, A. Tina-Booh, S. Wieczorek and S. H. Schwartz (2009), "Testing measurement invariance using multigroup CFA: Differences between educational groups in human values measurement", *Quality & Quantity*, Vol. 43/4, pp. 599–616, http://doi.org/10.1007/s11135-007-9143-x.

Takahashi, Y., G. W. Edmonds, J. J. Jackson and B. W. Roberts (2013), "Longitudinal correlated changes in conscientiousness, preventative health-related behaviors, and self-perceived physical health", *Journal of Personality*, Vol. 81/4, pp. 417–427.

Terborg, J. R. and L. H. Peters (1974), "Some observations on wording of item-stems for attitude questionnaires", *Psychological Reports*, Vol. 35/1, pp. 463–466.

Tokar, D. M., A. R. Fischer, A. F. Snell and N. Harik-Williams (1999), "Efficient assessment of the five-factor model of personality: Stuctural validity analyses of the NEO five-factor inventory (form s)", *Measurement and Evaluation in Counseling and Development*, Vol. 32/1, p.14.

Toomela, A. (2003), "Relationships between personality structure, structure of word meaning, and cognitive ability: A study of cultural mechanisms of personality", *Journal of Personality and Social Psychology*, Vol. 85/4, p. 723.

Van Eijck, K. and P. M. de Graaf (2004), "The Big Five at school: The impact of personality on educational attainment", *The Netherlands Journal of Social Sciences*, Vol. 40/1, pp. 24–40.

Vandenberg, R. J. and C. E. Lance (2000), "A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research", *Organizational Research Methods*, Vol. 3/1, p.4–70.

Wagner, G. G., J. R. Frick and J. Schupp (2007), *The German Socio-Economic Panel study (SOEP)-Evolution, Scope and Enhancements*, German Socio-Economic Panel Study (SOEP), DIW Berlin, www.diw.de/documents/publikationen/73/diw_01.c.60184.de/diw_sp0001.pdf.

Wong, N., A. Rindfleisch and J. E. Burroughs (2003), "Do reverse-worded items confound measures in cross-cultural consumer research? The case of the material values scale", *Journal of Consumer Research*, Vol. 30/1, pp.72–91.

Woods, C. M. (2009), "Evaluation of MIMIC-model methods for DIF testing with comparison to two-group analysis", *Multivariate Behavioral Research*, Vol. 44/1, pp. 1–27.

World Bank (2014), S*TEP Skills Measurement: Snapshot 2014*, World Bank.

Zinbarg, R. E., W. Revelle, I. Yovel and W. Li (2005), "Cronbach's α, Revelle's β, and Mcdonald's ωH: Their relations with each other and two alternative conceptualizations of reliability", *Psychometrika*, Vol. 70/1, pp.123–133, http://doi.org/10.1007/s11336-003-0974-7.

ANNEX A. THE ORIGINAL VERSIONS IN ENGLISH AND THE POLISH ADAPTATION OF BFI-S AND GRIT-S QUESTIONNAIRES

Table A.1 The original version (Gerlitz and Schupp, 2005) and the Polish adaptation of BFI-S (Strus, Cieciuch, Rowiński)

English	Polish
Agreeableness	
Sometimes a bit rude to others*	czasami jest niegrzeczny dla innych*
Forgiving	łatwo przebacza
Considerate and kind to others	prawie dla każdego jest uprzejmy i życzliwy
Conscientiousness	
A thorough worker	dokładnie wykonuje swoją pracę
Somewhat lazy*	jest leniwy*
Effective and efficient in completing tasks	skutecznie wykonuje swoje zadania
Extraversion	
Communicative, talkative	dużo mówi
Outgoing, sociable	jest towarzyski i wylewny
Reserved*	zachowuje rezerwę w stosunku do ludzi*
Neuroticism	
A worrier	dużo się martwi
Nervous	łatwo się denerwuje
Relaxed, able to deal with stress*	zachowuje spokój w napiętych sytuacjach*
Openness	
Original, someone who comes up with new ideas	jest oryginalny i pomysłowy
Someone who values artistic, aesthetic experiences	ceni artystyczne i estetyczne doświadczenia
Imaginative	ma bogatą wyobraźnię

Note: *Reverse-worded item.

Table A.2 The original version (Duckworth and Quinn, 2009) and the Polish adaptation of Grit-S

English	Polish
Consistency of interest	
New ideas and projects sometimes distract me from previous ones*	Czasami nowe pomysły i plany odrywają mnie od poprzednich.
I often set a goal but later choose to pursue a different one*	Często wybieram jakiś cel, ale potem dążę do czegoś innego.
I have difficulty maintaining my focus on projects that take more than a few months to complete*	Trudno mi skupić się na zadaniach, które wymagają wielu miesięcy pracy.
I have been obsessed with a certain idea or project for a short time but later lost interest*	Mam obsesję na punkcie jakiegoś pomysłu czy planu przez krótki czas, ale potem przestaję się tym interesować
Perseverance of effort	
Setbacks don't discourage me	Przeszkody mnie nie zniechęcają.
I finish whatever I begin	Zawsze kończę to, co zacząłem / zaczęłam.
I am diligent	Jestem pilny/a.
I am a hard worker	Jestem bardzo pracowity/a.

Note: *Reverse-worded item.

ANNEX B. IRT PARAMETERS OF BFI-S AND GRIT-S ITEMS

Table B.1 IRT parameters of the Agreeableness subscale

Item	Coef.	SE	z	95% CI	
				LB	UB
Sometimes a bit rude to others*					
a	0.64	0.05	13.59	0.55	0.74
b1	-4.82	0.28	-17.38	-5.37	-4.28
b2	-3.59	0.18	-19.49	-3.96	-3.23
b3	-2.26	0.09	-24.49	-2.44	-2.08
b4	-0.82	0.05	-18.06	-0.91	-0.73
b5	0.24	0.11	2.27	0.03	0.45
b6	2.10	0.14	15.00	1.83	2.37
Forgiving					
a	0.89	0.06	13.98	0.76	1.01
b1	-3.73	0.19	-19.64	-4.10	-3.35
b2	-2.64	0.13	-20.87	-2.89	-2.39
b3	-1.73	0.06	-26.59	-1.85	-1.60
b4	-0.59	0.03	-18.45	-0.65	-0.53
b5	0.42	0.08	5.54	0.27	0.56
b6	1.66	0.03	54.37	1.60	1.72
Considerate and kind to others					
a	2.70	0.42	6.48	1.88	3.51
b1	-2.80	0.17	-16.69	-3.12	-2.47
b2	-2.31	0.12	-20.06	-2.54	-2.08
b3	-1.79	0.07	-26.11	-1.92	-1.65
b4	-1.01	0.03	-39.22	-1.06	-0.96
b5	-0.26	0.01	-29.48	-0.28	-0.24
b6	0.75	.	.	.	.

Notes: *- Reverse-worded item, a – discrimination parameter, b – difficulty parameters for item response categories. Coef. – coefficient; SE - standard error; 95% CI - 95% confidence interval; z - z-test statistic; LB - lower bound; UB - upper bound.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table B.2 IRT parameters of the Conscientiousness subscale

Item	Coef.	SE	z	95% CI	
				LB	UB
A thorough worker					
a	2.26	0.13	17.98	2.01	2.50
b1	-3.02	0.13	-23.51	-3.27	-2.77
b2	-2.92	0.09	-31.45	-3.10	-2.74
b3	-2.40	0.06	-38.84	-2.53	-2.28
b4	-1.75	0.04	-46.05	-1.83	-1.68
b5	-0.99	0.02	-52.57	-1.02	-0.95
b6	0.03	0.10	0.30	-0.17	0.23
Somewhat lazy*					
a	0.99	0.04	22.66	0.91	1.08
b1	-3.75	0.12	-31.00	-3.98	-3.51
b2	-2.65	0.08	-31.91	-2.82	-2.49
b3	-1.79	0.05	-34.09	-1.89	-1.69
b4	-0.87	0.04	-23.65	-0.94	-0.79
b5	-0.17	0.04	-4.02	-0.25	-0.09
b6	0.95	0.12	7.67	0.71	1.19
Effective and efficient in completing tasks					
a	2.40	0.13	17.90	2.13	2.66
b1	-3.33	0.12	-28.04	-3.57	-3.10
b2	-2.84	0.09	-32.46	-3.02	-2.67
b3	-2.45	0.06	-43.41	-2.56	-2.34
b4	-1.70	0.03	-54.32	-1.76	-1.64
b5	-0.74	0.02	-40.58	-0.77	-0.70
b6	0.50	0.13	3.74	0.24	0.77

Notes: *- Reverse-worded item, a – discrimination parameter, b – difficulty parameters for item response categories. Coef. – coefficient; SE - standard error; 95% CI - 95% confidence interval; z - z-test statistic; LB - lower bound; UB - upper bound.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table B.3 IRT parameters of the Extraversion subscale

Item	Coef.	SE	z	95% CI	
				LB	UB
Communicative, talkative					
a	1.76	0.35	4.97	1.07	2.46
b1	-2.13	0.19	-11.10	-2.51	-1.75
b2	-1.45	0.11	-13.41	-1.66	-1.24
b3	-0.77	.	.	.	.
b4	0.04	0.03	1.40	-0.02	0.11
b5	0.68	0.10	6.84	0.48	0.87
b6	1.27	.	.	.	.
Outgoing, sociable					
a	1.41	0.24	5.84	0.94	1.89
b1	-2.81	0.27	-10.50	-3.34	-2.29
b2	-2.04	0.17	-12.24	-2.36	-1.71
b3	-1.26	0.05	-24.61	-1.36	-1.16
b4	-0.22	.	.	.	.
b5	0.64	0.11	5.67	0.42	0.86
b6	1.56	.	.	.	.
Reserved*					
a	0.10	0.04	2.51	0.02	0.18
b1	-15.65	2.92	-5.36	-21.37	-9.92
b2	-4.92	.	.	.	.
b3	4.88	5.35	0.91	-5.61	15.37
b4	17.35	9.32	1.86	-0.92	35.63
b5	25.95	13.68	1.90	-0.86	52.77
b6	37.81	.	.	.	.

Notes: *- Reverse-worded item, a – discrimination parameter, b – difficulty parameters for item response categories. Coef. – coefficient; SE - standard error; 95% CI - 95% confidence interval; z - z-test statistic; LB - lower bound; UB - upper bound.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table B.4 IRT parameters of the Neuroticism subscale

Item	Coef.	SE	z	95% CI	
				LB	UB
A worrier					
a	1.05	0.07	15.12	0.91	1.18
b1	-3.11	0.13	-24.20	-3.36	-2.86
b2	-1.99	0.08	-24.63	-2.15	-1.83
b3	-1.21	0.04	-28.17	-1.30	-1.13
b4	-0.34	0.03	-10.76	-0.40	-0.28
b5	0.46	0.06	7.49	0.34	0.59
b6	1.47	0.02	95.91	1.44	1.50
Nervous					
a	3.00	0.48	6.26	2.06	3.94
b1	-1.62	0.06	-26.39	-1.74	-1.50
b2	-1.00	0.04	-27.98	-1.07	-0.93
b3	-0.54	0.02	-28.17	-0.57	-0.50
b4	0.00	0.02	-0.03	-0.05	0.04
b5	0.53	0.04	12.19	0.44	0.61
b6	1.08	.	.	.	.
Relaxed, able to deal with stress					
a	0.70	0.04	16.65	0.62	0.79
b1	-2.55	0.08	-33.73	-2.70	-2.40
b2	-0.80	0.05	-17.18	-0.89	-0.71
b3	0.59	0.11	5.58	0.38	0.80
b4	2.14	0.17	12.37	1.80	2.47
b5	3.30	0.25	13.35	2.81	3.78
b6	4.58	.	.	.	.

Notes: *- Reverse-worded item, a – discrimination parameter, b – difficulty parameters for item response categories. Coef. – coefficient; SE - standard error; 95% CI - 95% confidence interval; z - z-test statistic; LB - lower bound; UB - upper bound.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table B.5 IRT parameters of the Openness subscale

Item	Coef.	SE	z	95% CI	
				LB	UB
Original, someone who comes up with new ideas					
a	1.66	0.08	21.01	1.51	1.82
b1	-2.91	0.09	-32.41	-3.09	-2.73
b2	-2.44	0.06	-39.88	-2.56	-2.32
b3	-1.63	0.03	-51.76	-1.69	-1.57
b4	-0.43	0.03	-16.13	-0.48	-0.38
b5	0.59	0.05	12.84	0.50	0.68
b6	1.55	0.06	27.75	1.44	1.66
Someone who values artistic, aesthetic experiences					
a	1.16	0.05	21.91	1.06	1.26
b1	-3.47	0.11	-30.24	-3.70	-3.25
b2	-2.61	0.08	-32.86	-2.77	-2.46
b3	-1.78	0.05	-38.64	-1.87	-1.69
b4	-0.73	0.03	-23.88	-0.79	-0.67
b5	0.15	0.04	3.50	0.07	0.24
b6	1.16	0.10	11.85	0.96	1.35
Imaginative					
a	1.79	0.09	20.69	1.62	1.96
b1	-2.60	0.08	-33.09	-2.75	-2.44
b2	-2.17	0.06	-37.74	-2.28	-2.06
b3	-1.58	0.04	-45.00	-1.65	-1.52
b4	-0.70	0.02	-29.46	-0.75	-0.65
b5	0.16	0.03	4.60	0.09	0.22
b6	1.16	0.05	22.49	1.06	1.26

Notes: *- Reverse-worded item, a – discrimination parameter, b – difficulty parameters for item response categories. Coef. – coefficient; SE - standard error; 95% CI - 95% confidence interval; z - z-test statistic; LB - lower bound; UB - upper bound.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table B.6 IRT parameters of the Consistency of Interest subscale (Grit-S)

Item	Coef.	SE	z	95% CI	
				LB	UB
New ideas and projects sometimes distract me from previous ones*					
a	0.74	0.04	20.43	0.67	0.81
b1	-3.67	0.06	-60.01	-3.79	-3.55
b2	-0.63	-0.05	12.32	-0.73	-0.53
b3	0.93	0.15	6.32	0.64	1.21
b4	3.53	0.09	37.71	3.35	3.71
I have been obsessed with a certain idea or project for a short time but later lost interest*					
a	1.84	0.07	27.79	1.71	1.97
b1	-2.23	0.03	-65.81	-2.29	-2.16
b2	-0.91	0.02	-39.51	-0.95	-0.86
b3	-0.15	0.03	-4.69	-0.22	-0.09
b4	1.32	0.04	30.03	1.23	1.41
I often set a goal but later choose to pursue a different one*					
a	2.09	0.08	26.67	1.94	2.25
b1	-2.25	0.03	-69.04	-2.32	-2.19
b2	-0.94	0.02	-43.20	-0.98	-0.90
b3	-0.20	0.03	-6.69	-0.26	-0.14
b4	1.26	0.04	28.11	1.17	1.35
I have difficulty maintaining my focus on projects that take more than a few months to complete*					
a	1.51	0.05	29.31	1.41	1.61
b1	-2.41	0.04	-63.28	-2.48	-2.33
b2	-0.95	0.03	-36.21	-1.00	-0.90
b3	-0.27	0.04	-7.28	-0.34	-0.20
b4	1.43	0.05	26.97	1.32	1.53

Notes: *- Reverse-worded item, a – discrimination parameter, b – difficulty parameters for item response categories. Coef. – coefficient; SE - standard error; 95% CI - 95% confidence interval; z - z-test statistic; LB - lower bound; UB - upper bound.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table B.7 IRT parameters of the Perseverance of Effort subscale (Grit-S).

Item	Coef.	SE	z	95% CI	
				LB	UB
Setbacks don't discourage me					
a	0.49	0.03	15.23	0.43	0.55
b1	-6.06	0.21	-28.71	-6.47	-5.65
b2	-2.74	0.10	-28.53	-2.92	-2.55
b3	-0.98	0.16	-6.04	-1.30	-0.66
b4	3.13	0.30	10.51	2.54	3.71
I am a hard worker					
a	1.94	0.06	33.66	1.83	2.05
b1	-3.19	0.06	-57.71	-3.30	-3.08
b2	-2.13	0.03	-68.74	-2.20	-2.07
b3	-1.01	0.02	-45.74	-1.06	-0.97
b4	0.66	0.10	6.52	0.46	0.86
I finish whatever I begin					
a	2.13	0.07	31.71	1.99	2.26
b1	-2.95	0.05	-61.31	-3.04	-2.85
b2	-1.84	0.03	-60.37	-1.90	-1.78
b3	-1.04	0.02	-51.92	-1.08	-1.00
b4	0.43	0.09	4.90	0.26	0.61
I am diligent					
a	3.83	0.20	18.79	3.43	4.23
b1	-2.62	0.04	-70.42	-2.69	-2.54
b2	-1.70	0.02	-77.37	-1.74	-1.66
b3	-0.83	0.02	-43.86	-0.87	-0.79
b4	0.55	0.08	6.89	0.39	0.71

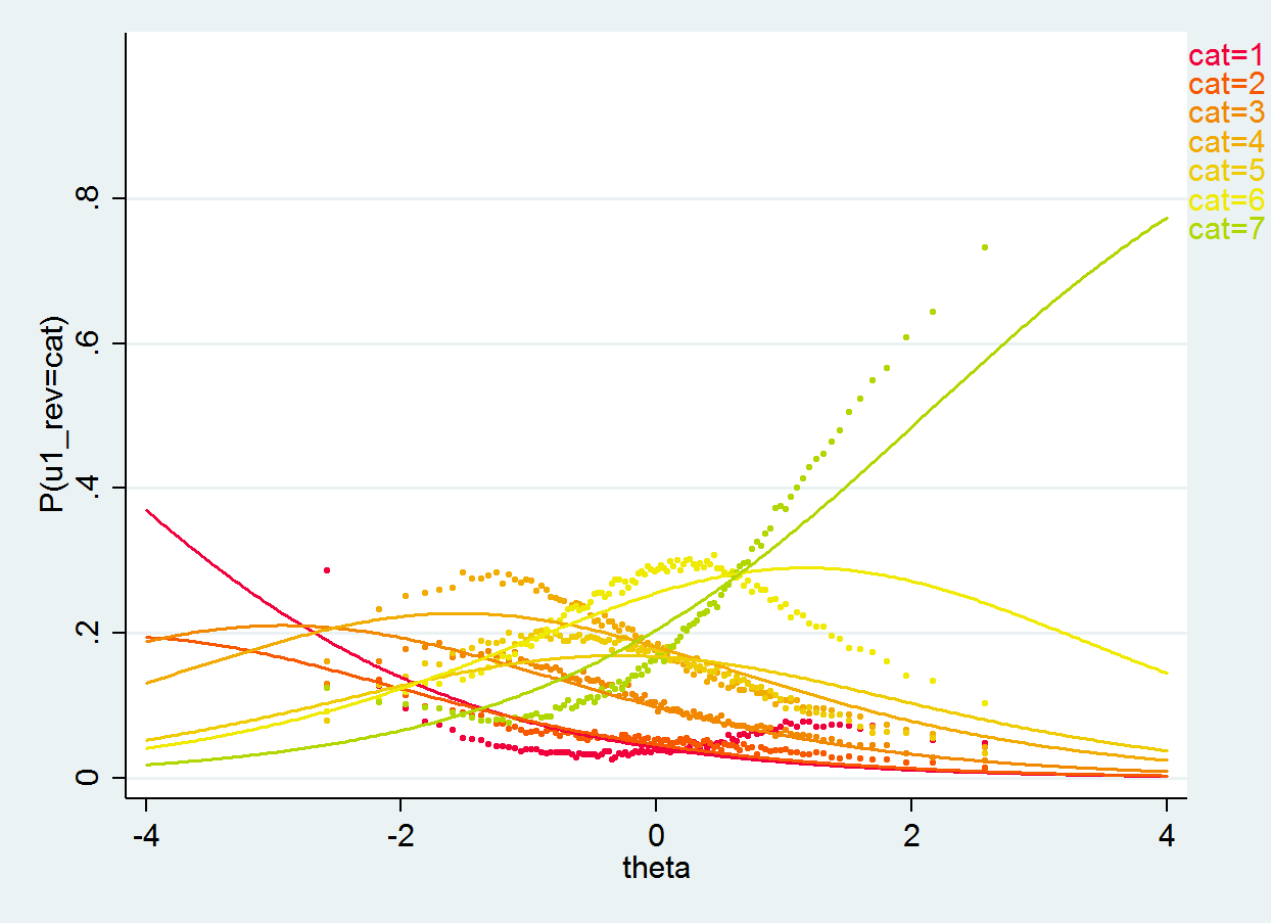
Notes: *- reverse-worded item, a – discrimination parameter, b – difficulty parameters for item response categories. Coef. – coefficient; SE - standard error; 95% CI - 95% confidence interval; z - z-test statistic; LB - lower bound; UB - upper bound.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

ANNEX C. ITEM CHARACTERISTIC CURVES OF BFI-S ITEMS

Agreeableness

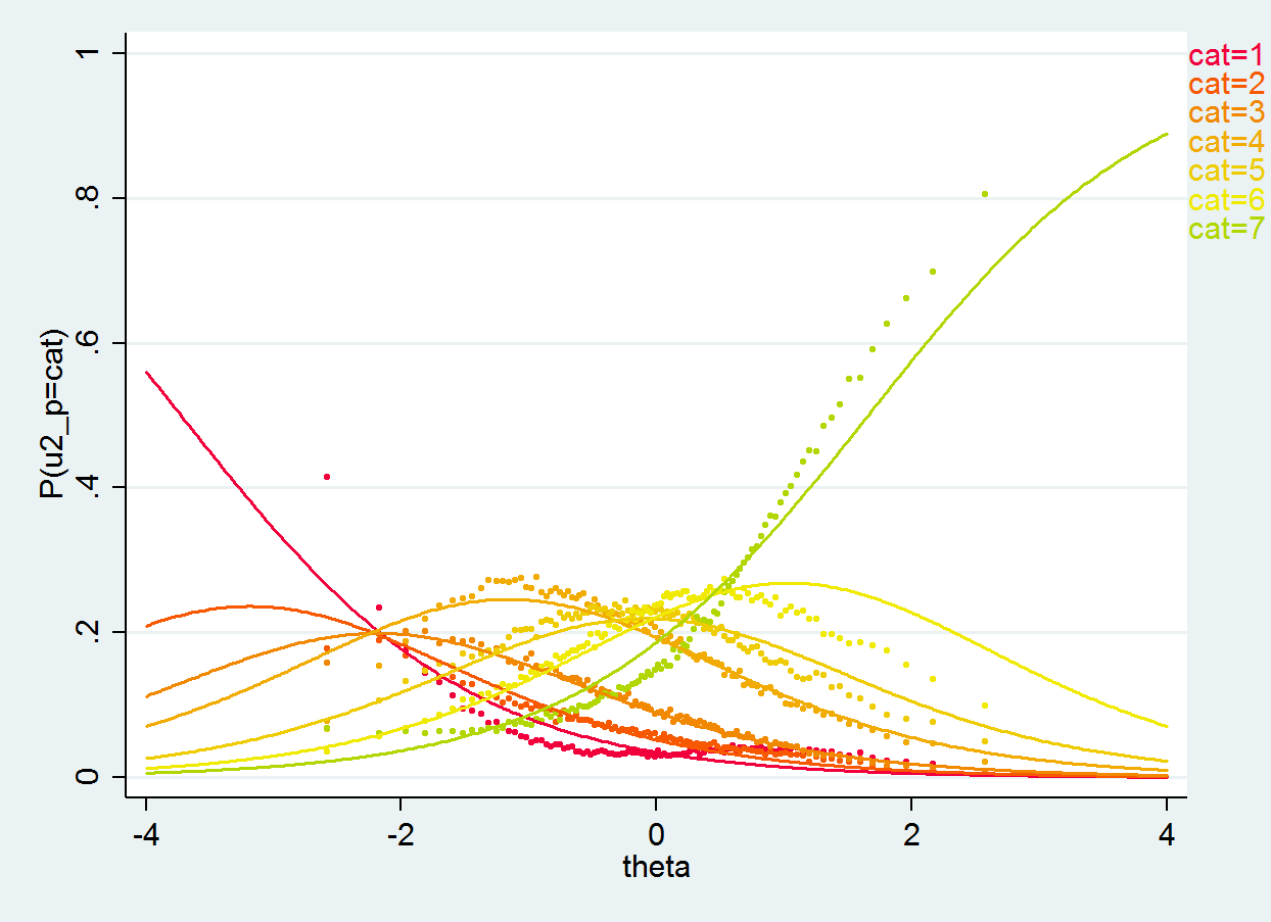
Figure C.1 ICC of "sometimes a bit rude to others" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

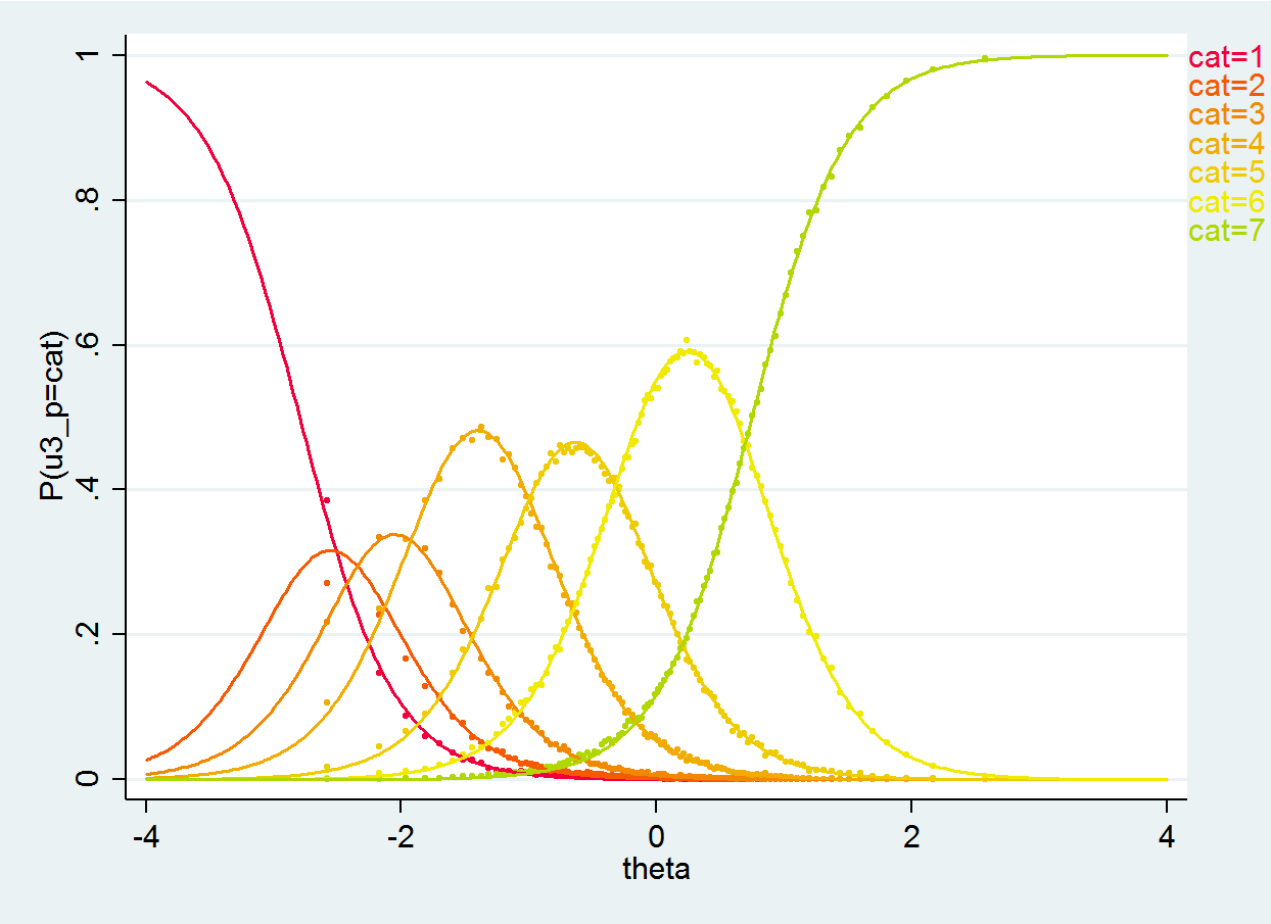
Figure C.2 ICC of "forgiving" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Figure C.3 ICC of "considerate and kind to others" item

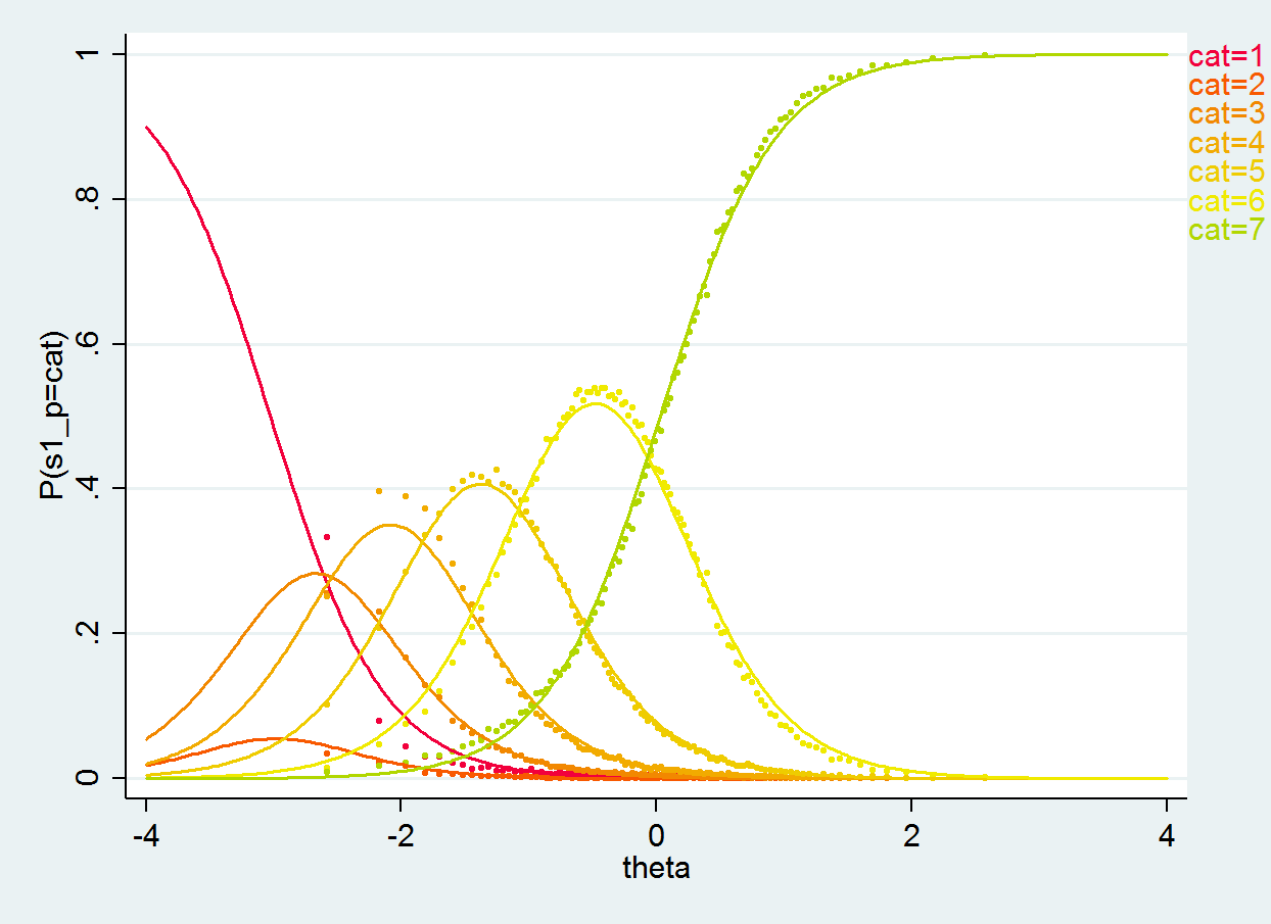


Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Conscientiousness

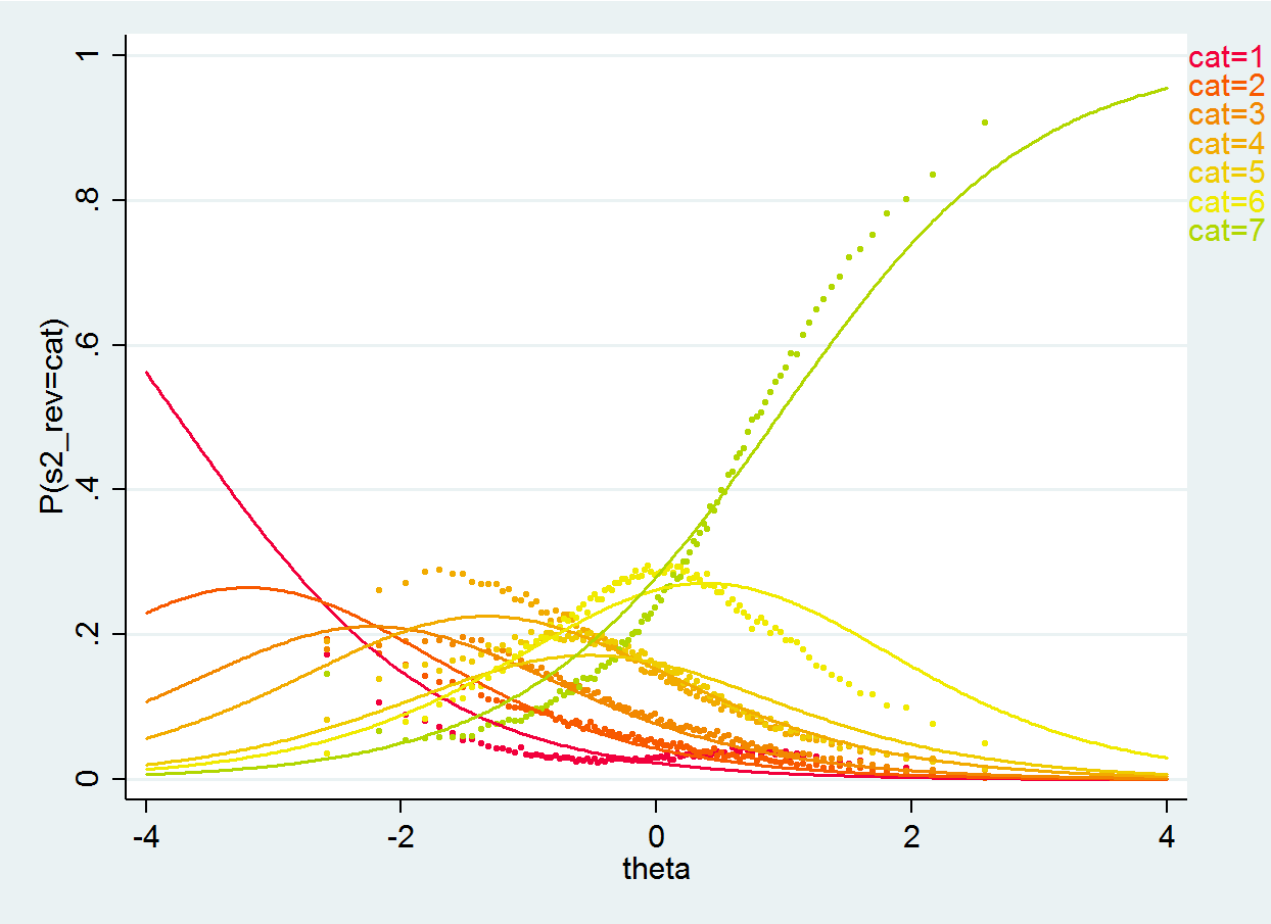
Figure C.4 ICC of "a thorough worker" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

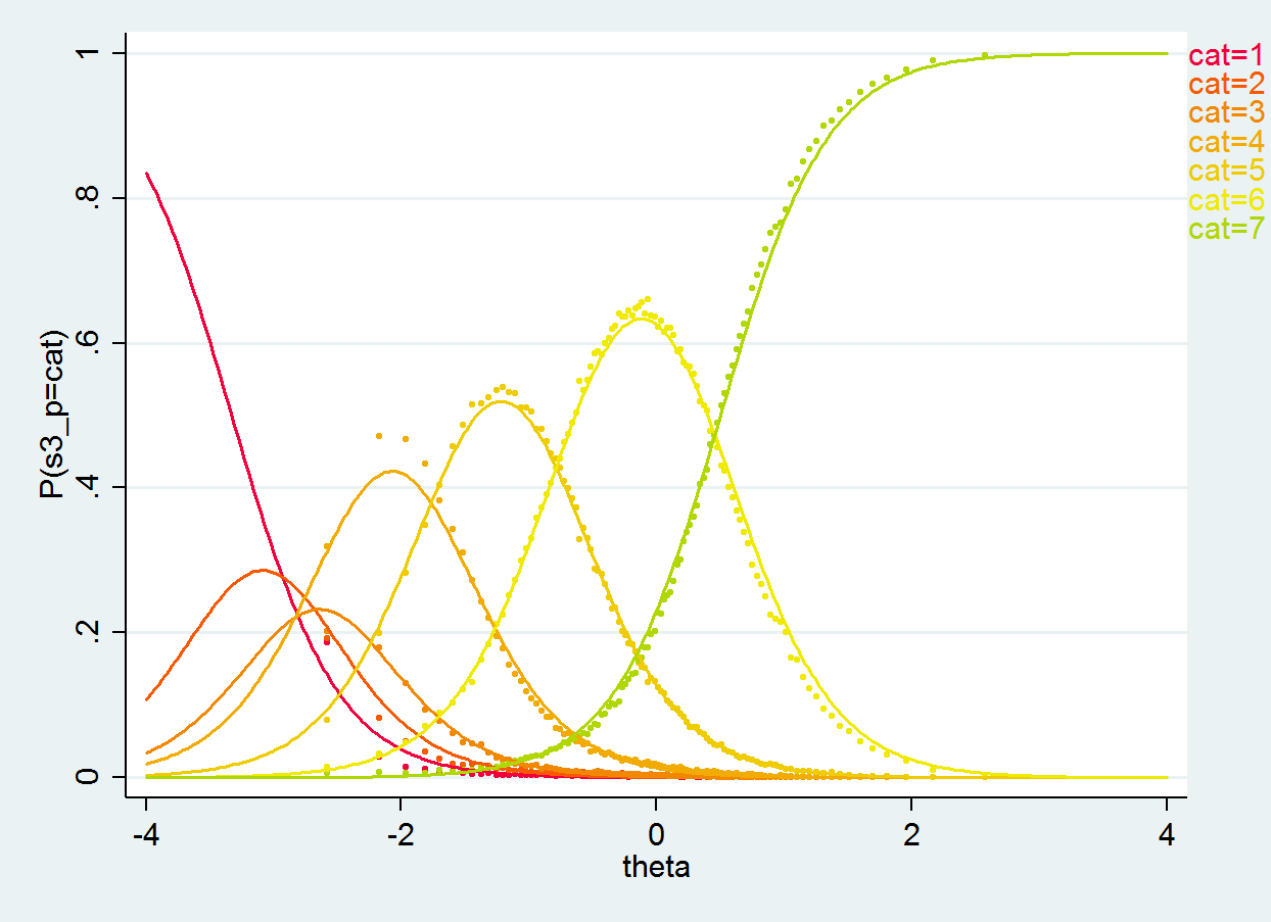
Figure C.5 ICC of "somewhat lazy" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Figure C.6 ICC of “effective and efficient in completing tasks" item

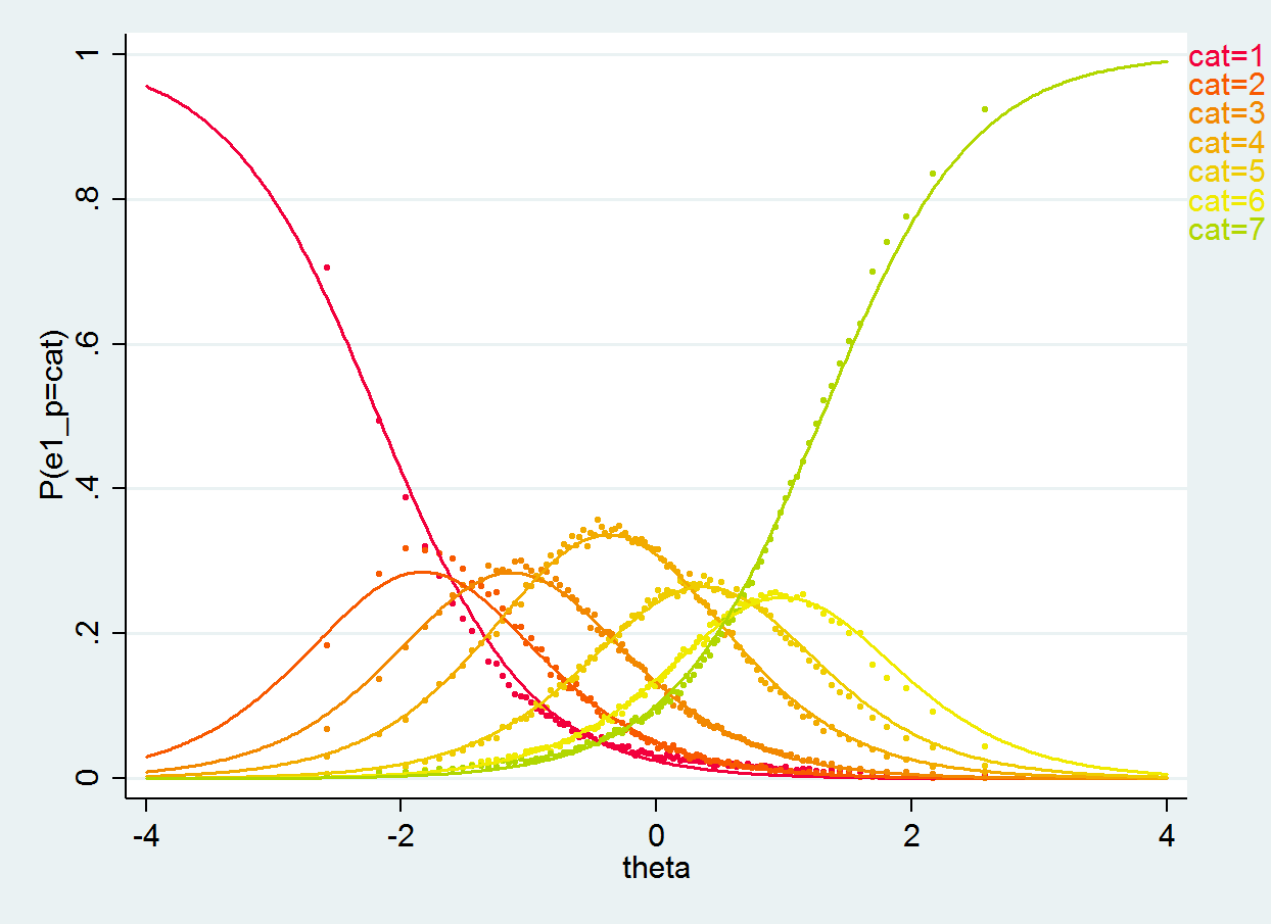


Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Extraversion

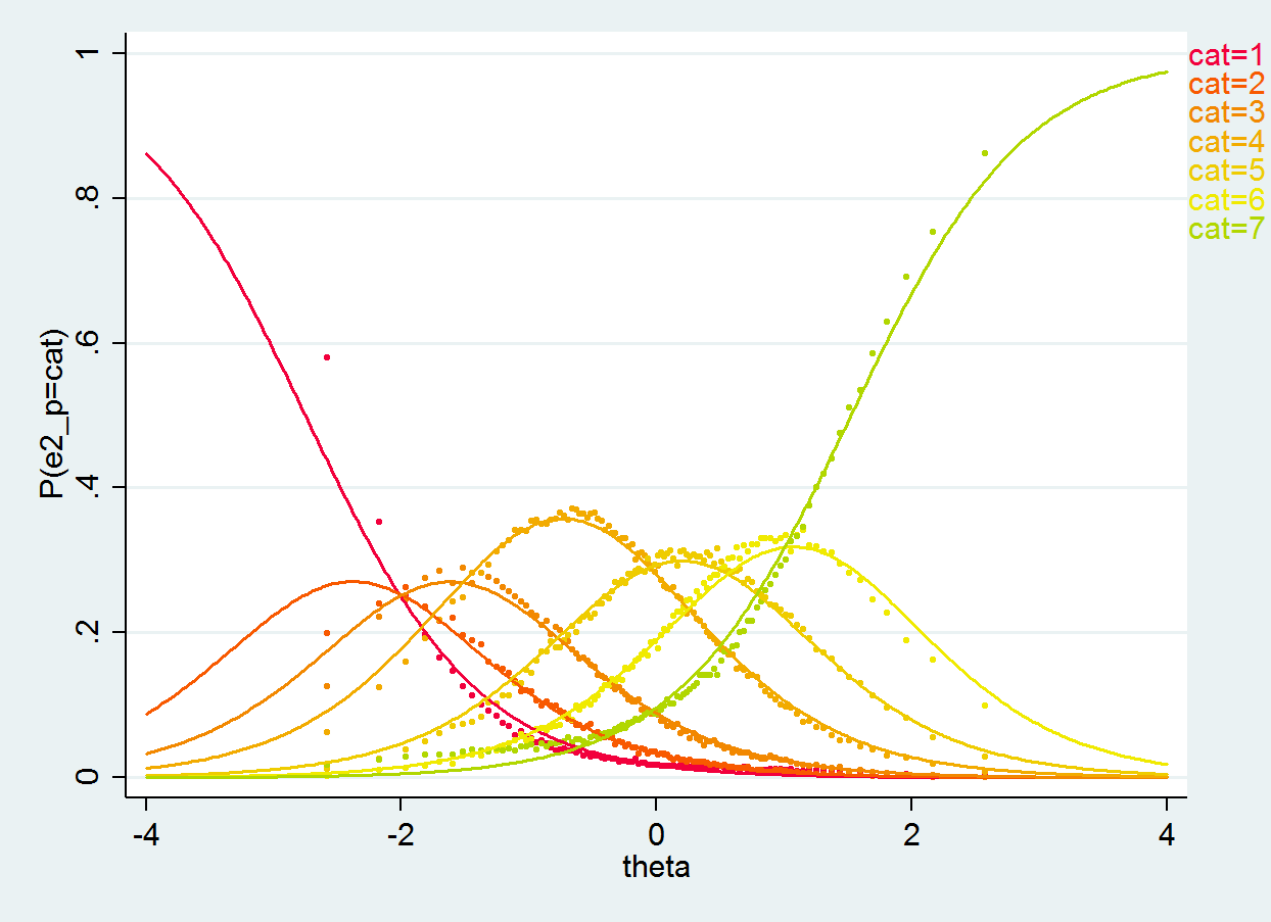
Figure C.7 ICC of "communicative, talkative" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

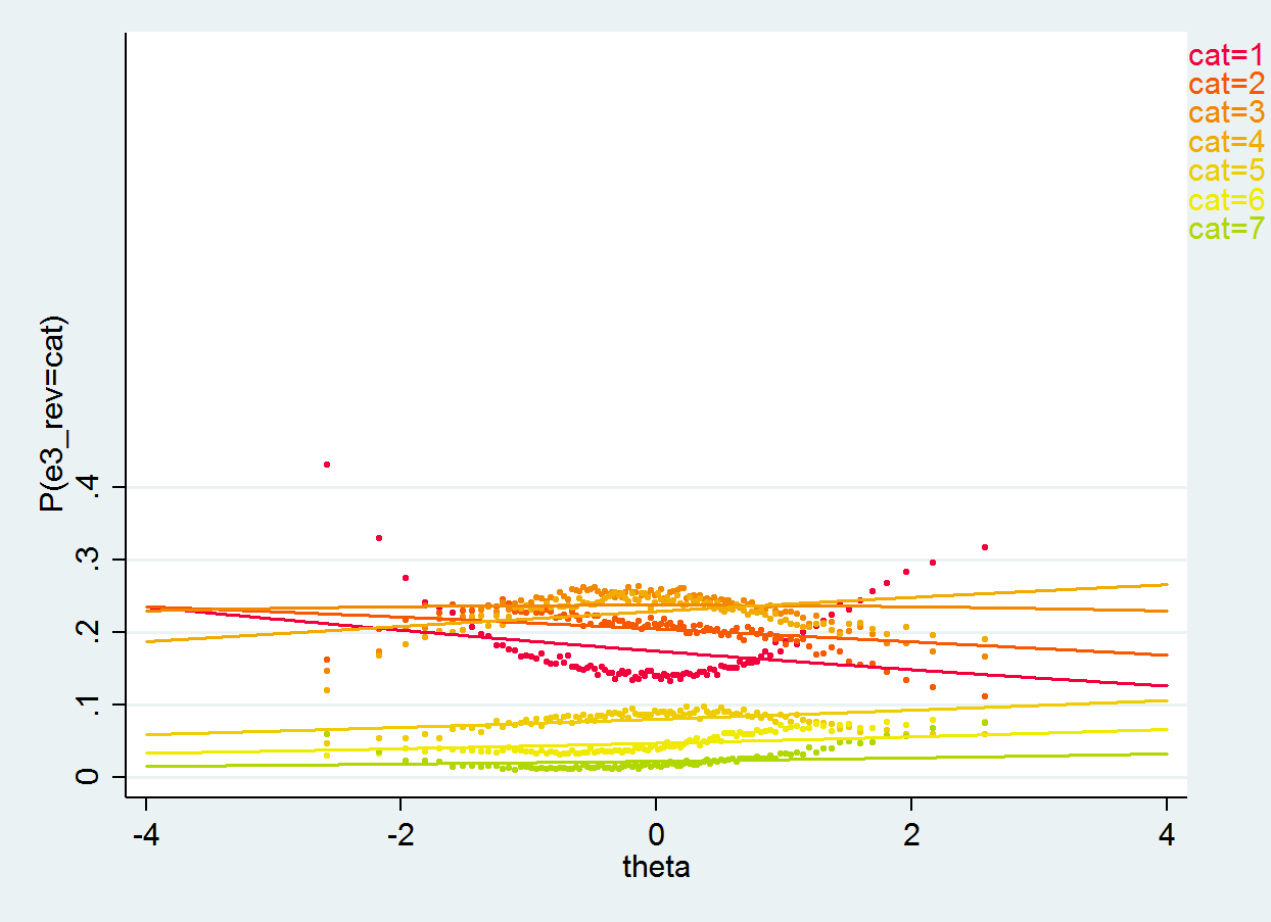
Figure C.8 ICC of "outgoing, sociable" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Figure C.9 ICC of "reserved" item

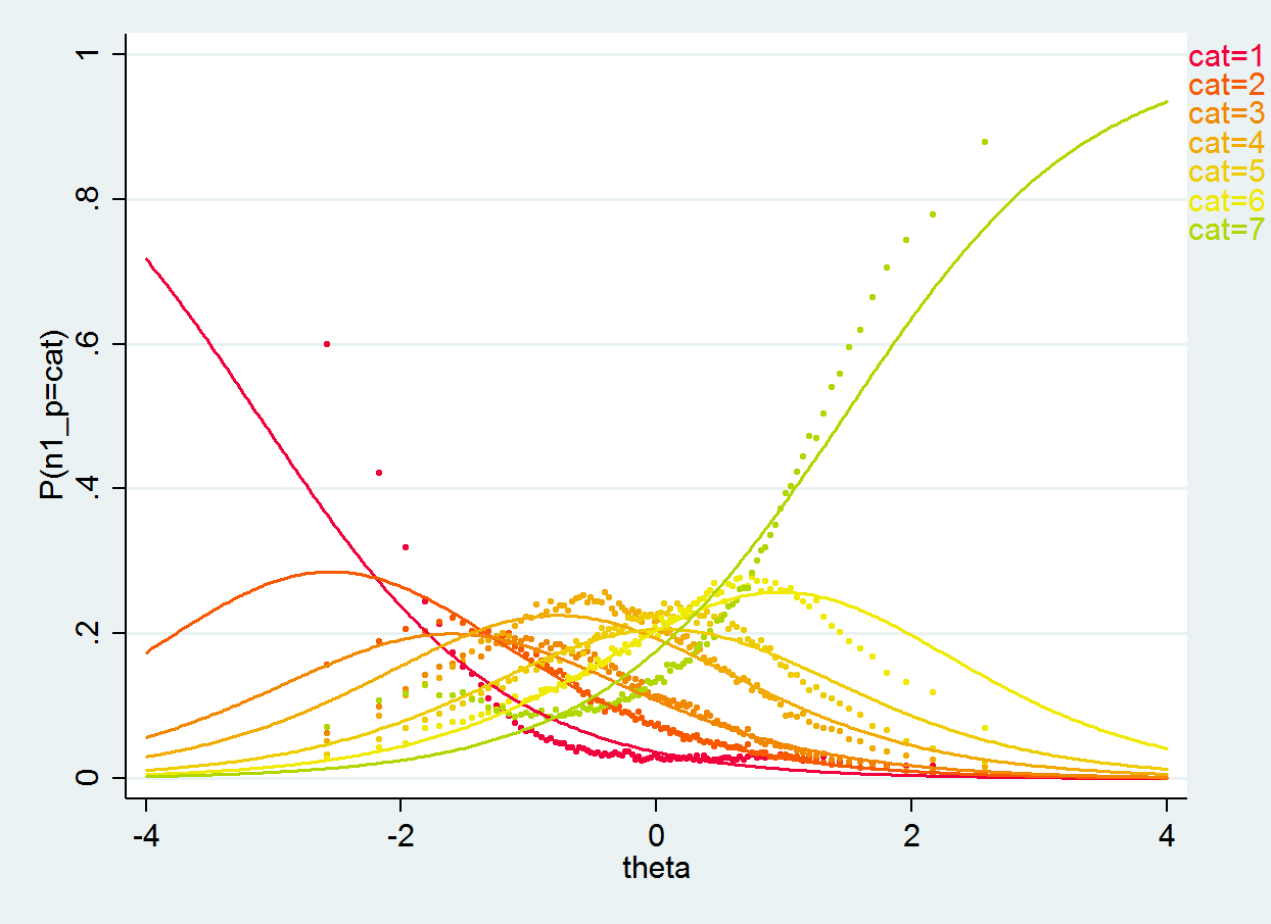


Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Neuroticism

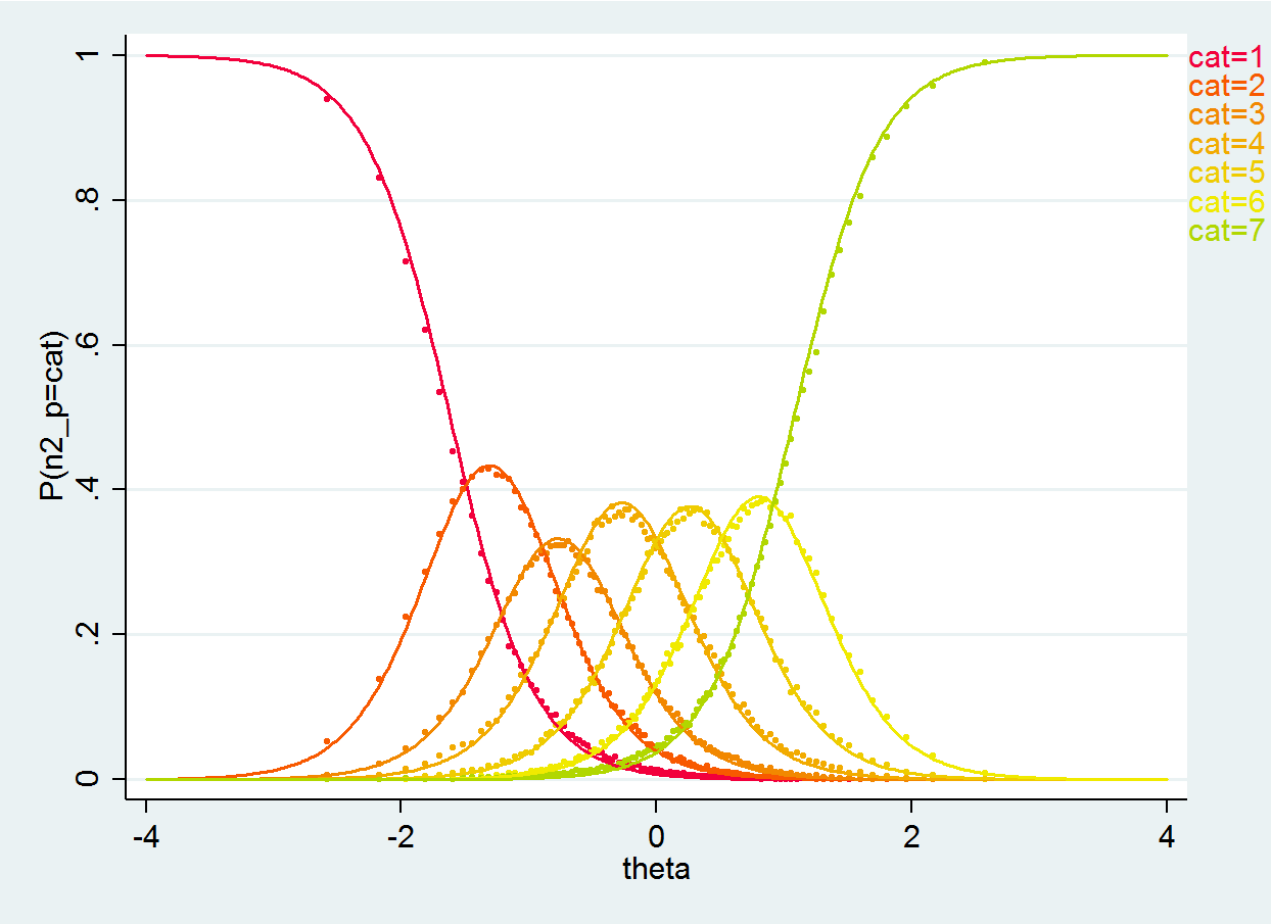
Figure C.10 ICC of "a worrier" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

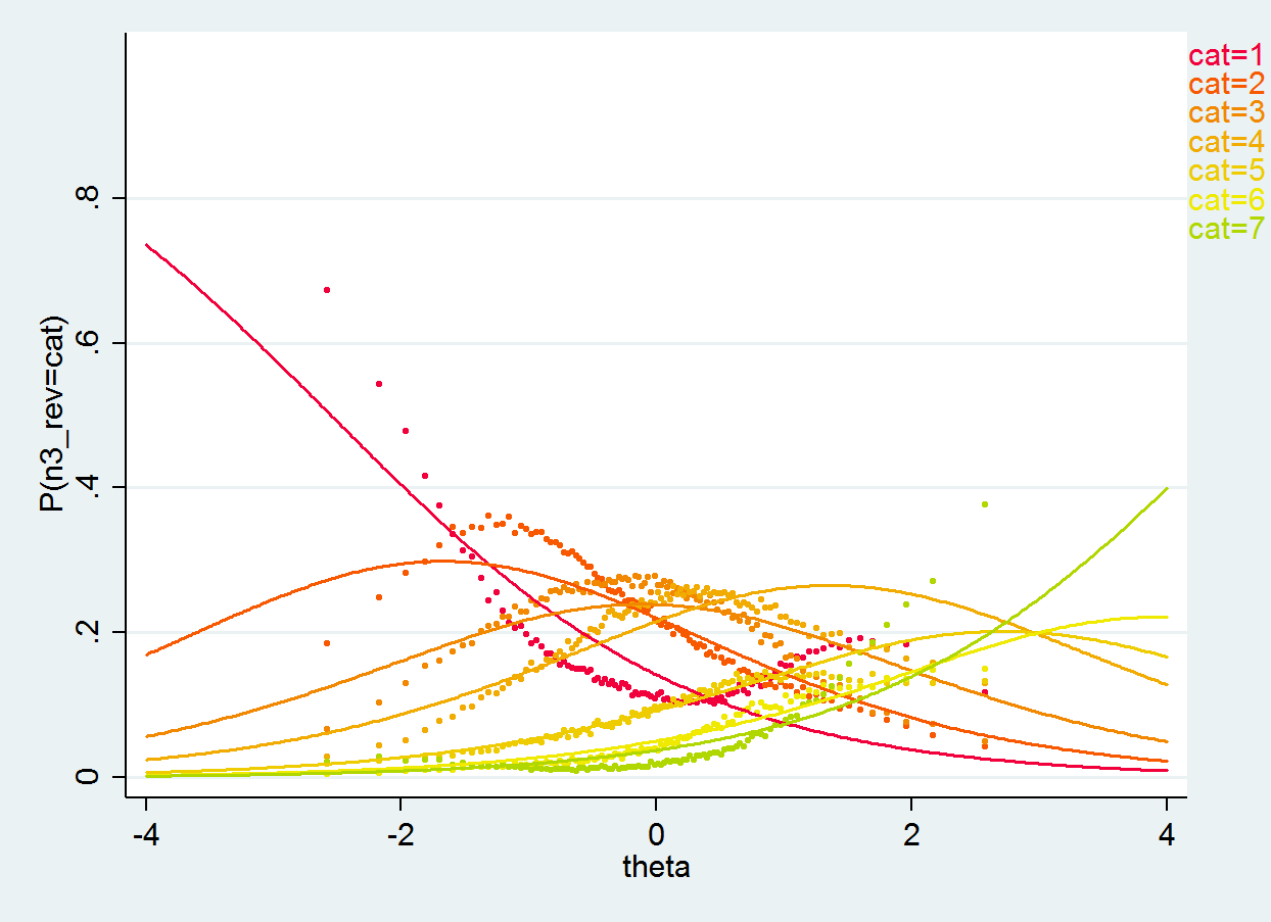
Figure C.11 ICC of "nervous" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Figure C.12 ICC of "relaxed, able to deal with stress" item

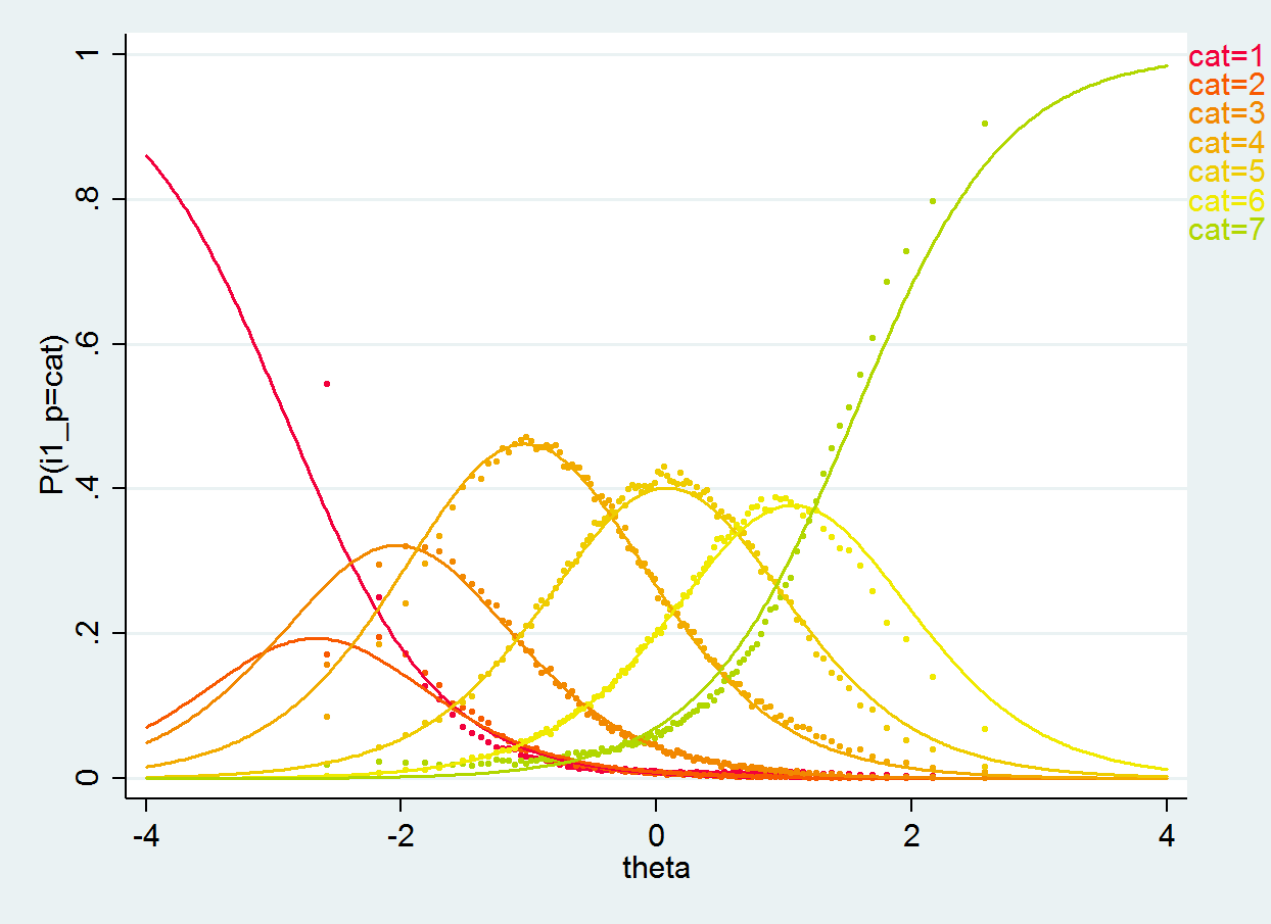


Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Openness

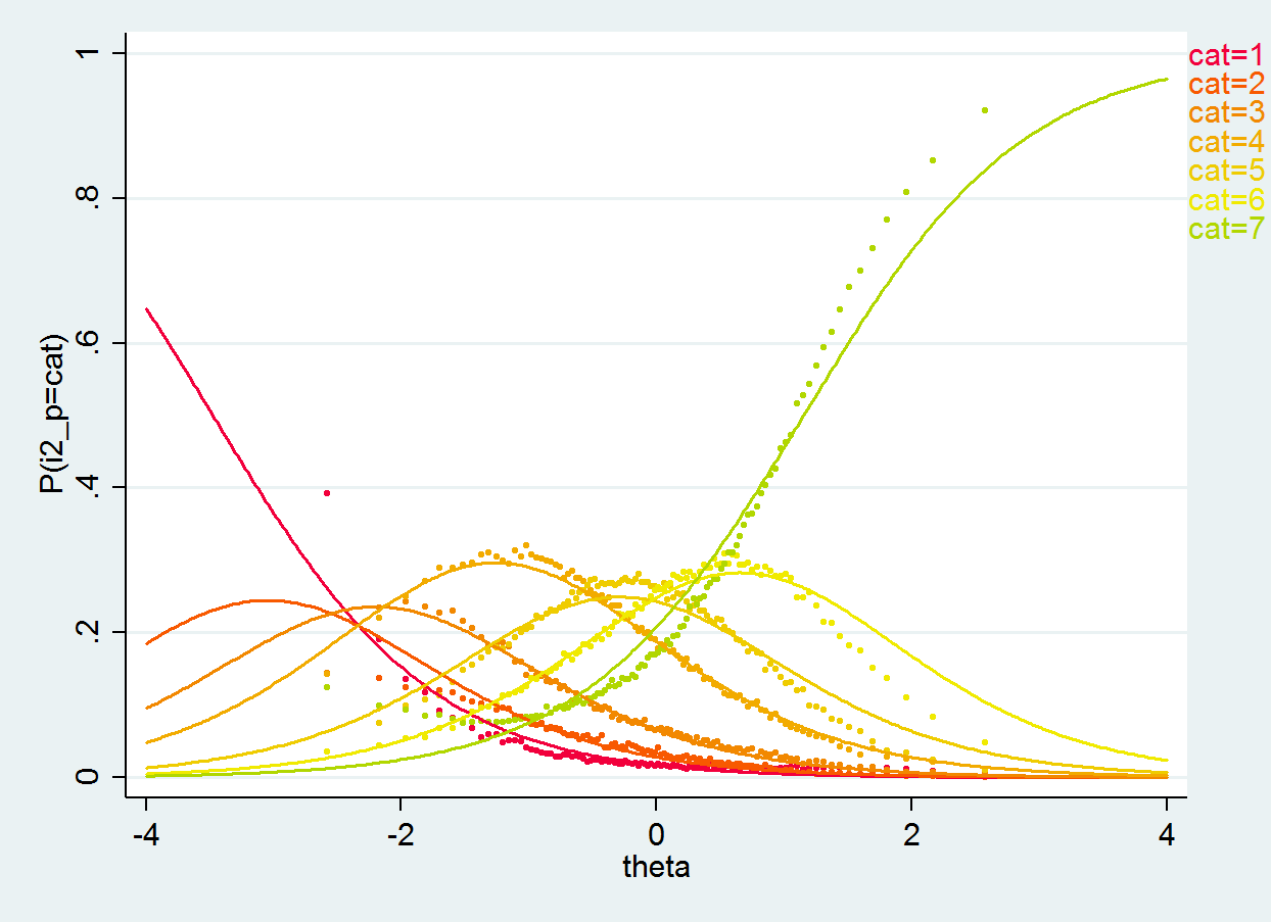
Figure C.13 ICC of "original, someone who comes up with new ideas" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

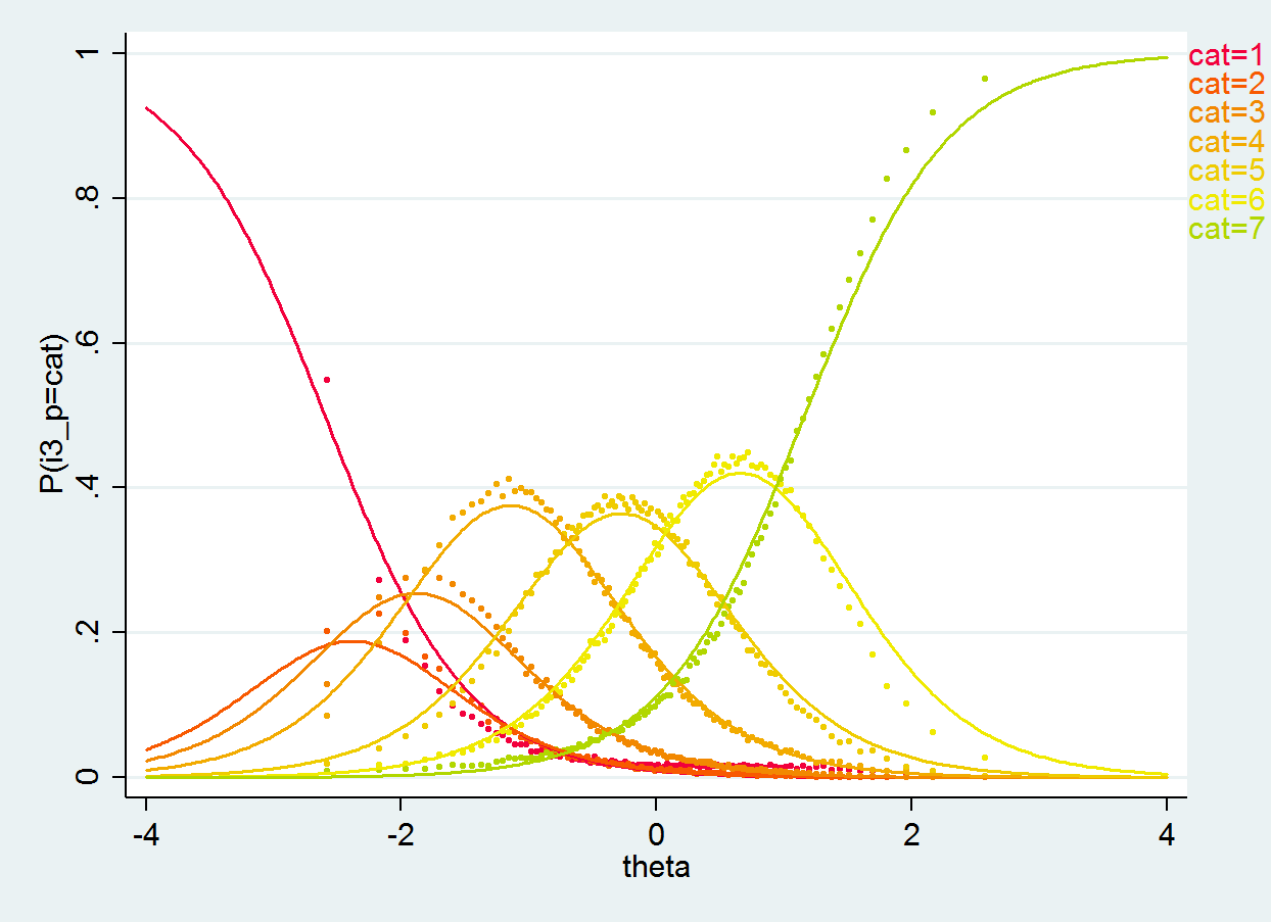
Figure C.14 ICC of "someone who values artistic, aesthetic experiences" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Figure C.15 ICC of "imaginative" item



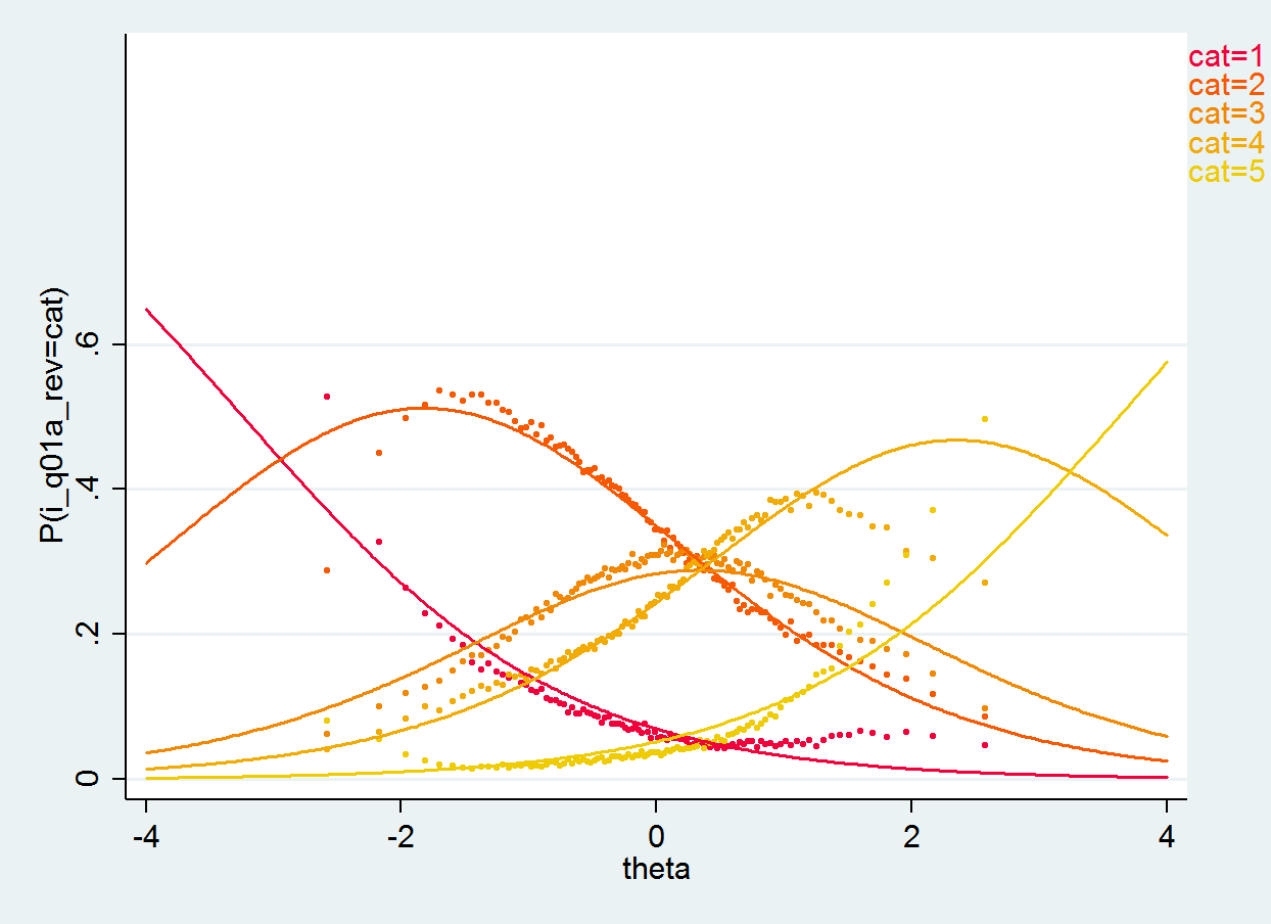
Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "disagree strongly", Cat7 - "agree strongly".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

ANNEX D. ITEM CHARACTERISTIC CURVES OF GRIT-S ITEMS

Consistency of Interest subscale

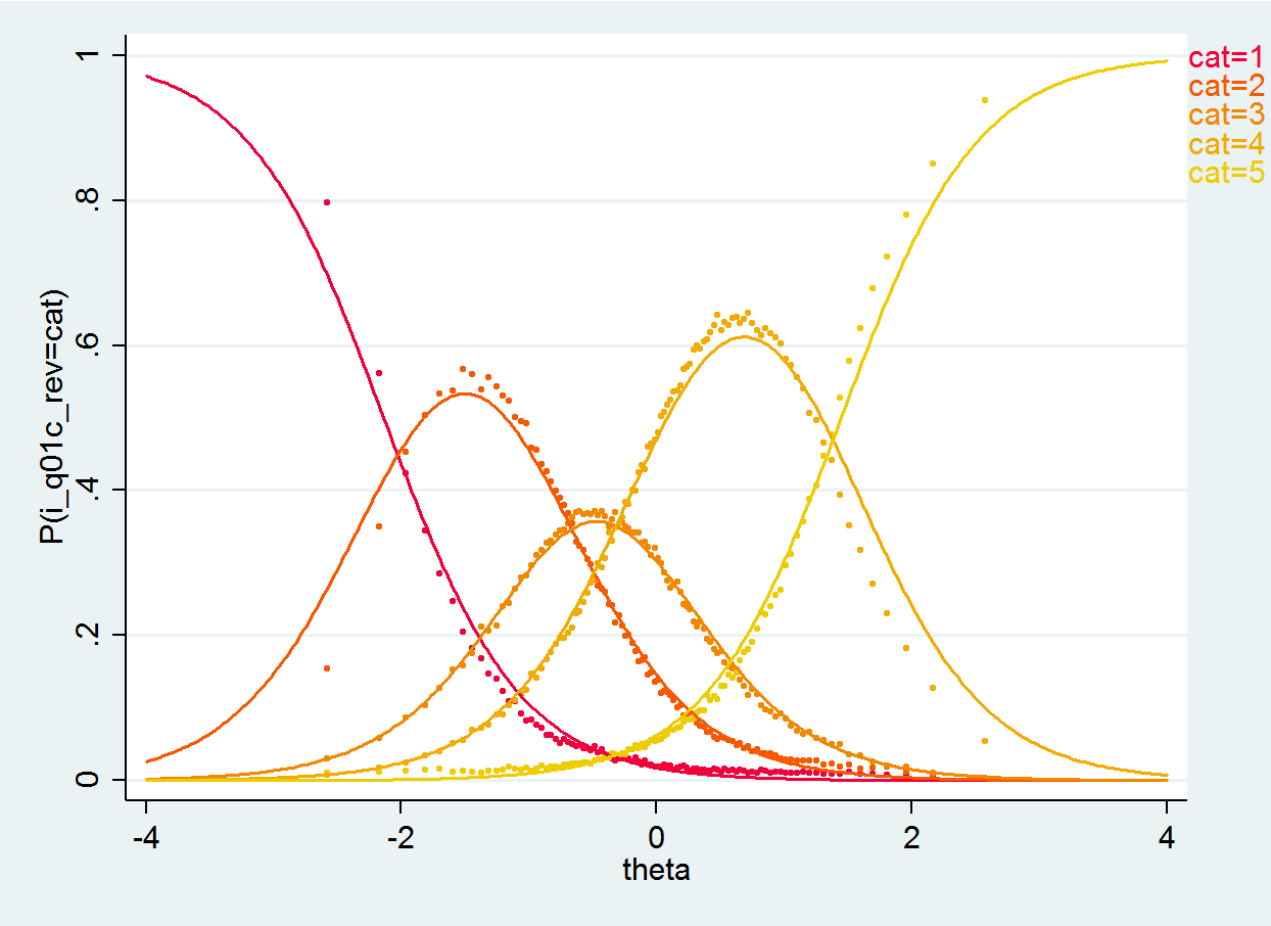
Figure D.1 ICC of "new ideas and projects sometimes distract me from previous ones" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "not like me at all". Cat5 - "very much like me".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

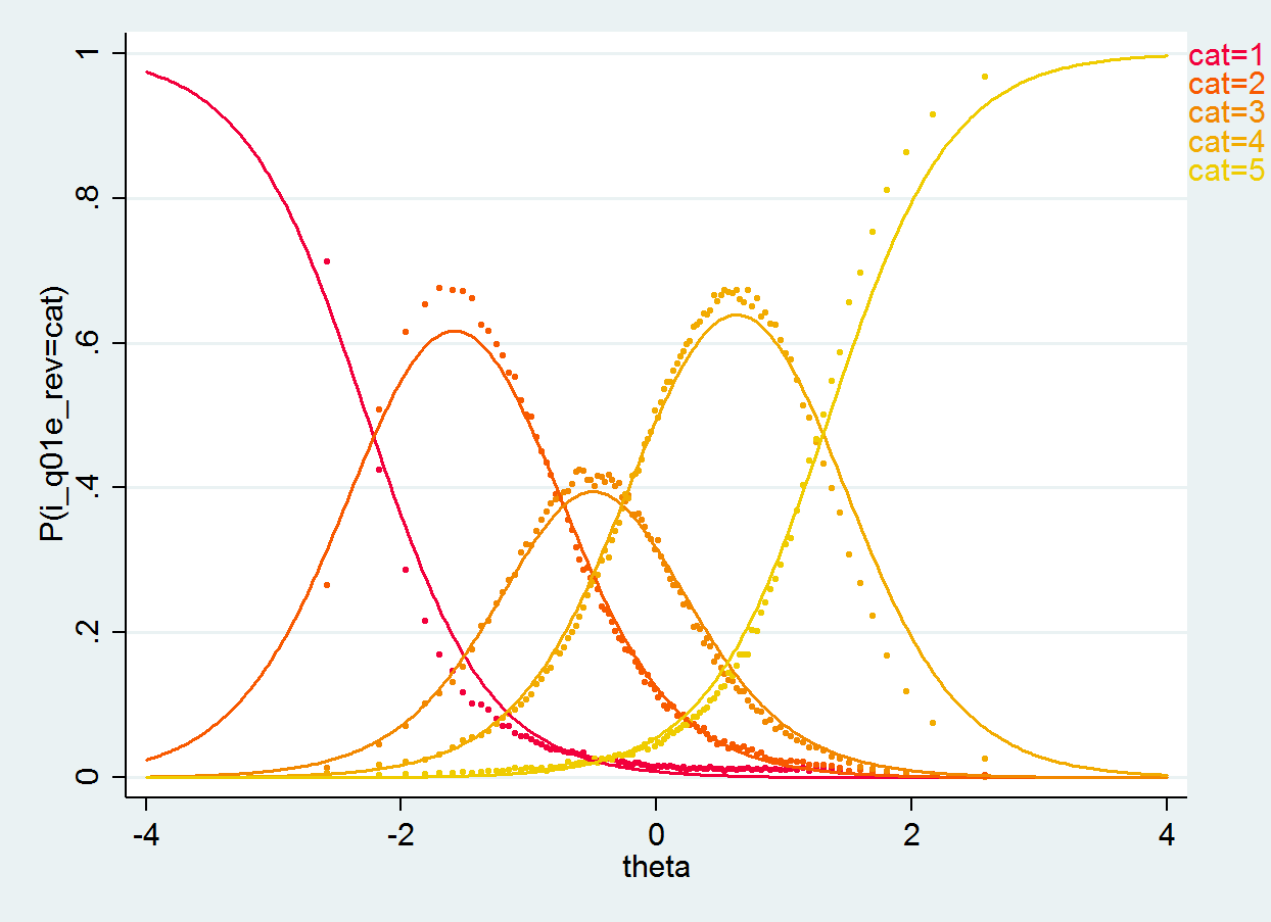
Figure D.2 ICC of "I have been obsessed with a certain idea or project for a short time but later lost interest" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "not like me at all". Cat5 - "very much like me".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

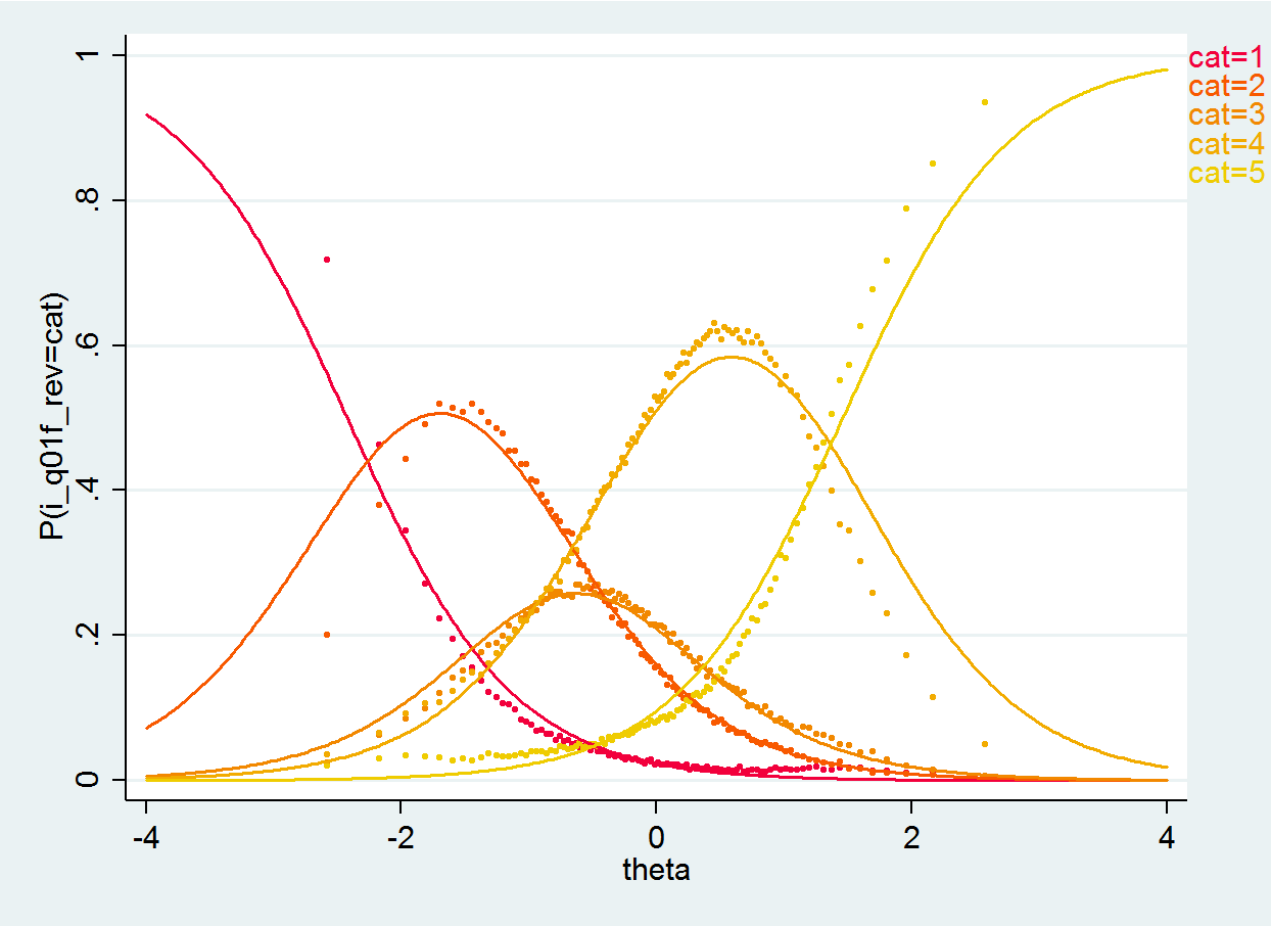
Figure D.3 ICC of "I often set a goal but later choose to pursue a different one" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "not like me at all". Cat5 - "very much like me".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Figure D.4 ICC of "I have difficulty maintaining my focus on projects that take more than a few months to complete" item

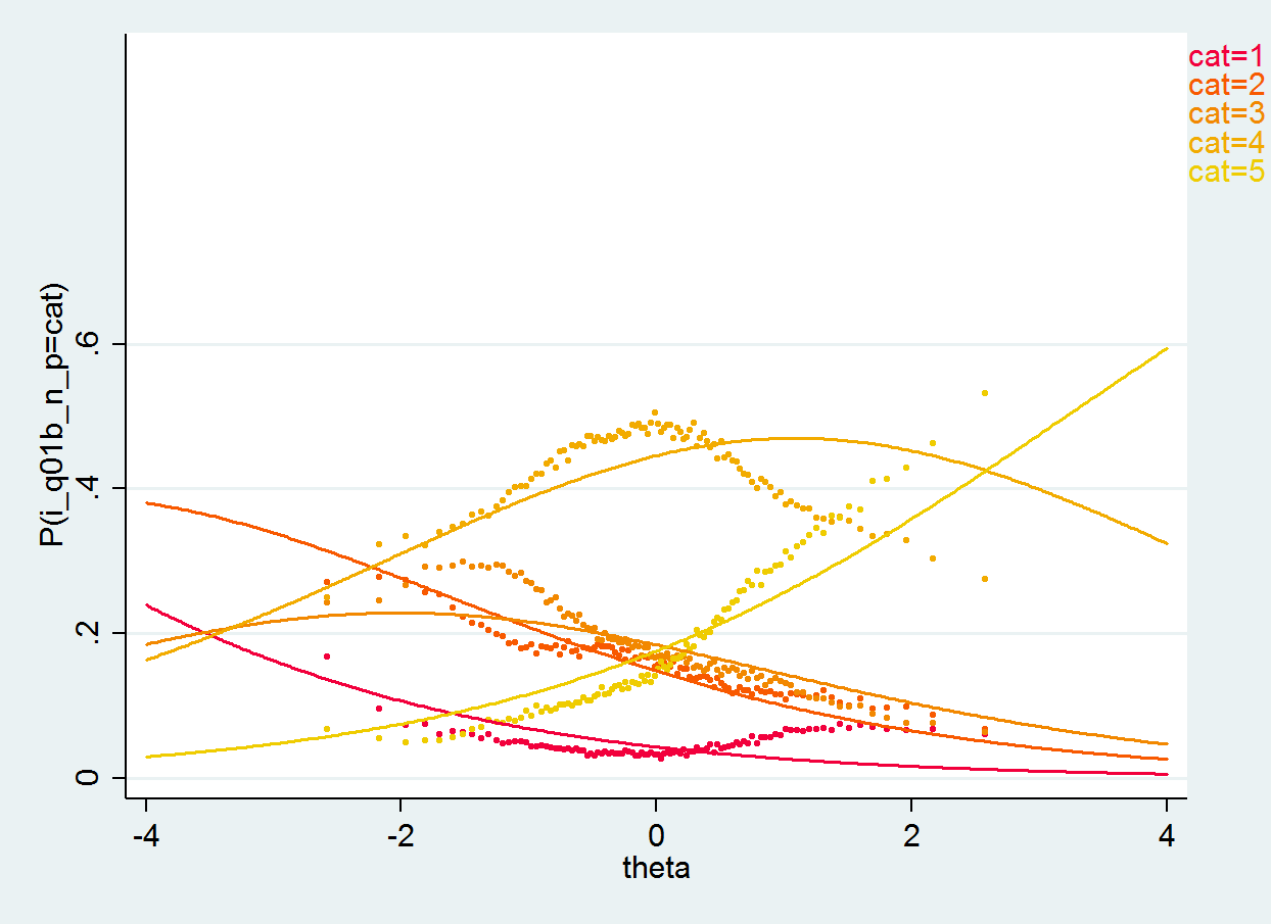


Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "not like me at all". Cat5 - "very much like me".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Perseverance of Effort subscale

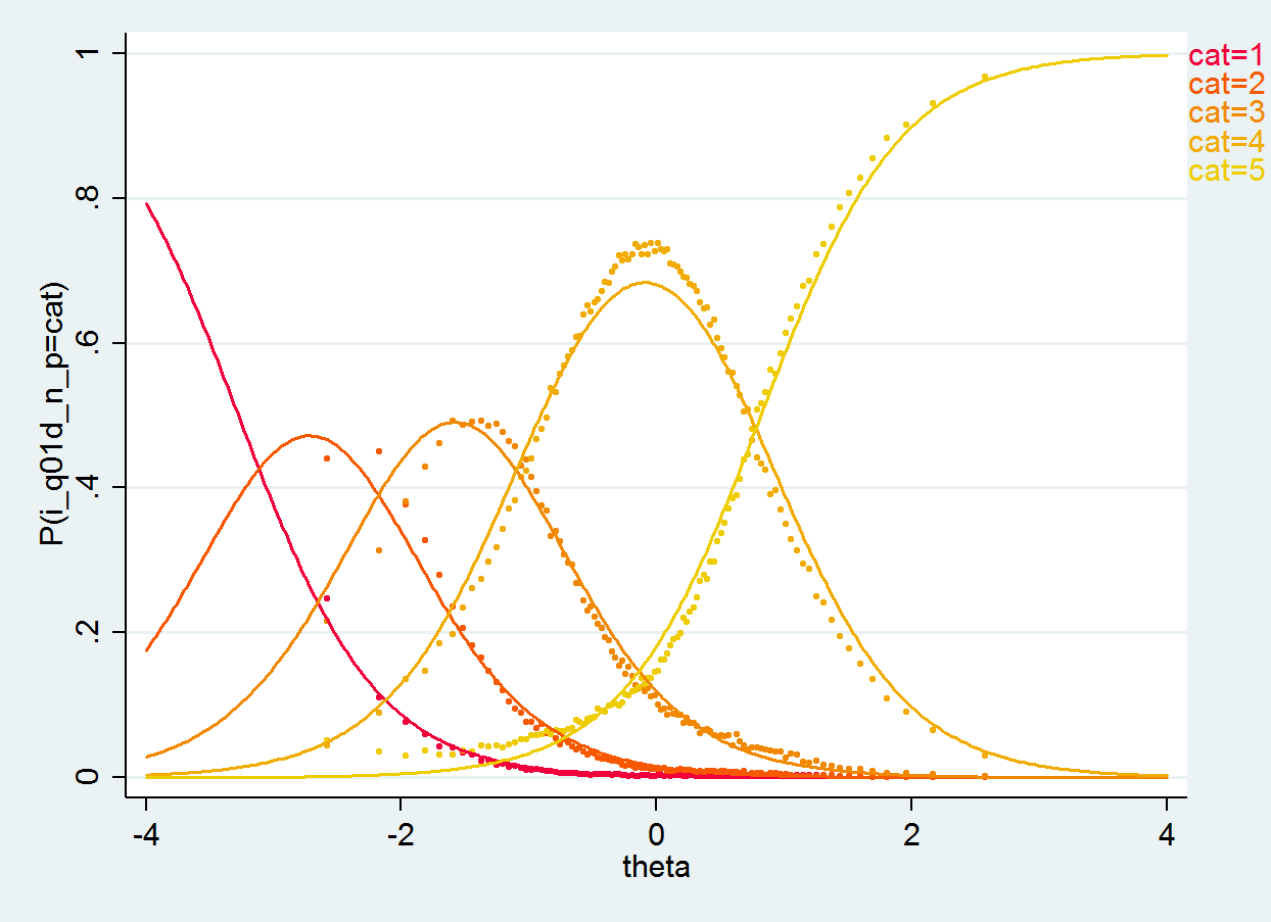
Figure D.5 ICC of "setbacks don't discourage me" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "not like me at all". Cat5 - "very much like me".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

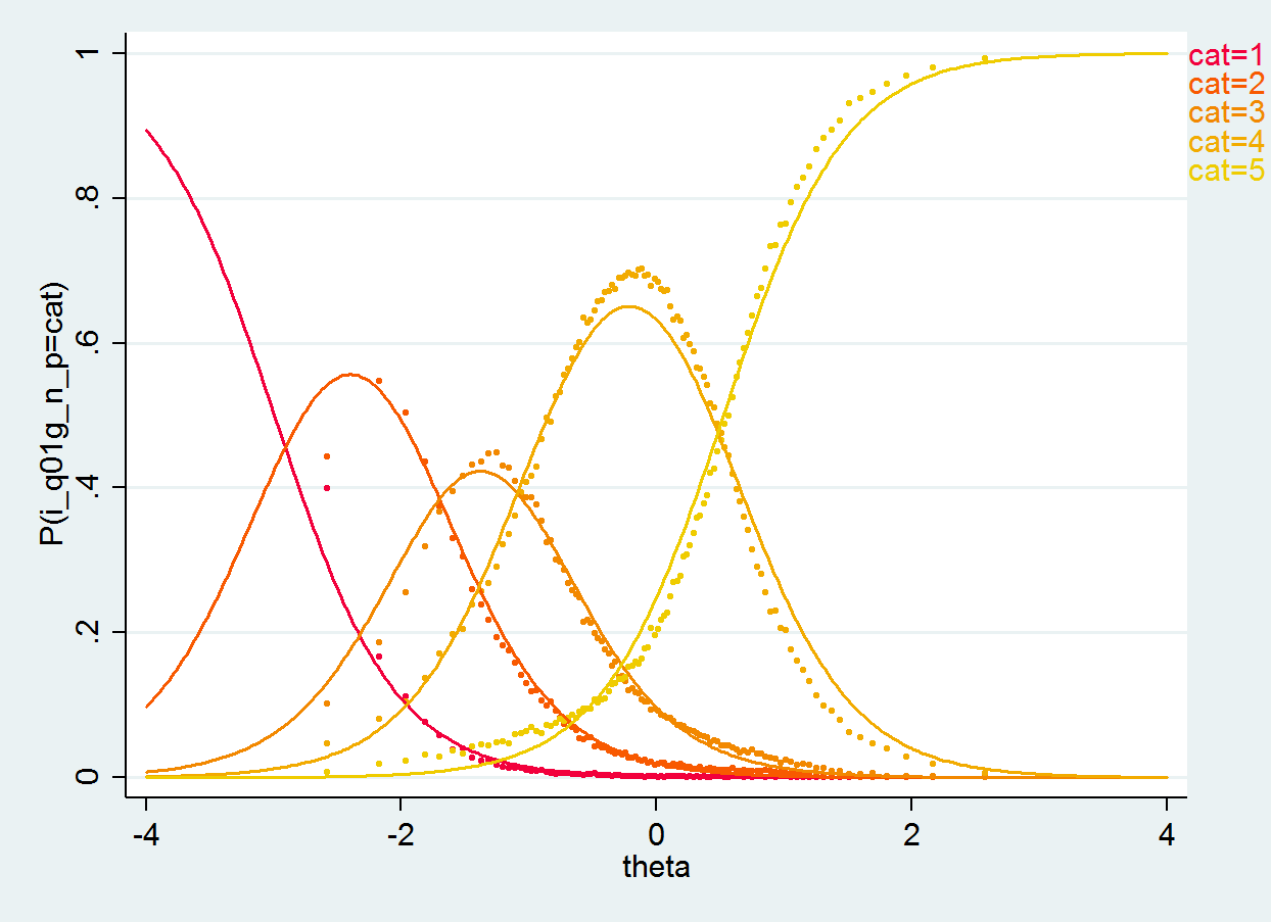
Figure D.6 ICC of "I am a hard worker" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "not like me at all". Cat5 - "very much like me".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

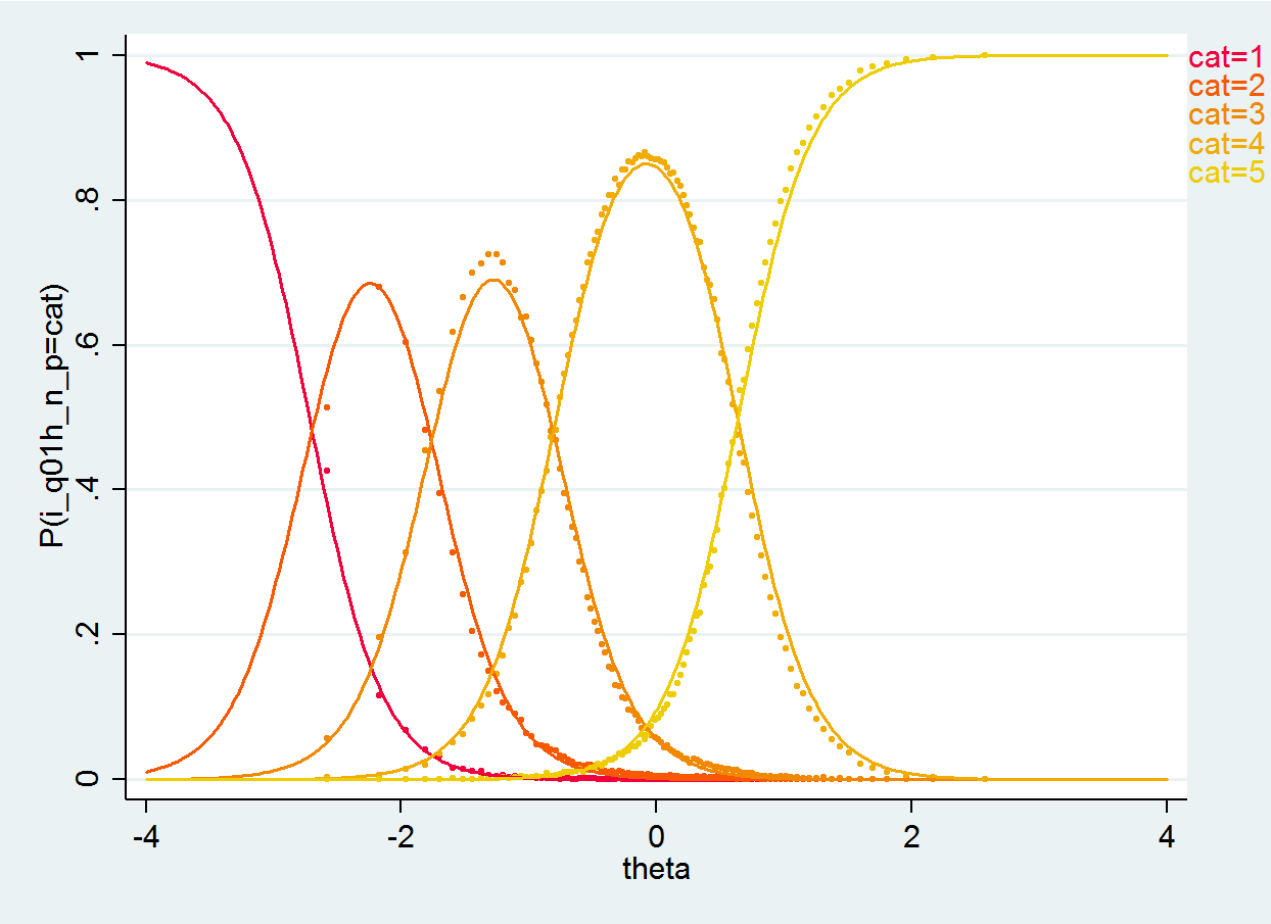
Figure D.7 ICC of "I finish whatever I begin" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "not like me at all". Cat5 - "very much like me".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Figure D.8 ICC of "I am diligent" item



Notes: Axis X - estimated latent trait level. Axis Y - probability of giving an answer of at least a value of x. Cat1 - "not like me at all". Cat5 - "very much like me".

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

ANNEX E. MODIFIED BFI-S AND GRIT-S MODELS AND THEIR PARAMETERS

In order to present graphs[1] and tables in a clear and consistent way, we use items' abbreviation in Annex E. Abbreviations are shown in Table E.1 and Table E.2.

Table E.1 BFI-S items' abbreviations

Item	Abbreviation
Agreeableness	
Sometimes a bit rude to others	a1
Forgiving	a2
Considerate and kind to others	a3
Conscientiousness	
A thorough worker	c1
Somewhat lazy	c2
Effective and efficient in completing tasks	c3
Extraversion	
Communicative, talkative	e1
Outgoing, sociable	e2
Reserved	e3
Neuroticism	
A worrier	n1
Nervous	n2
Relaxed, able to deal with stress	n3
Openness	
Original, someone who comes up with new ideas	o1
Someone who values artistic, aesthetic experiences	o2
Imaginative	o3

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

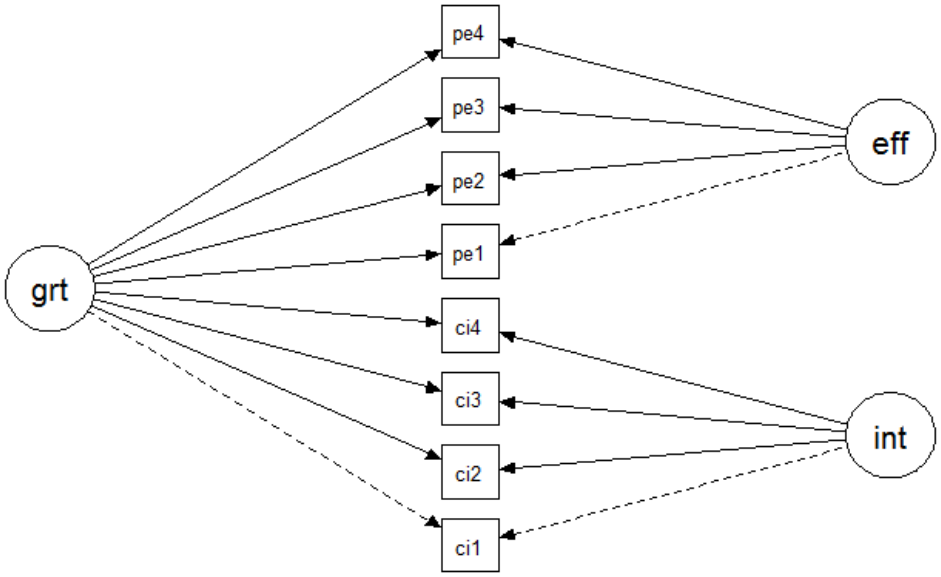
Table E.2 Grit-S items' abbreviations

Item	Abbreviation
Consistency of interest	
New ideas and projects sometimes distract me from previous ones*	ci1
I have been obsessed with a certain idea or project for a short time but later lost interest	ci2
I often set a goal but later choose to pursue a different one	ci3
I have difficulty maintaining my focus on projects that take more than a few months to complete	ci4
Perseverance of effort	
Setbacks don't discourage me	pe1
I am a hard worker	Pe2
I finish whatever I begin	Pe3
I am diligent	Pe4

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

[1] Obtained in R environment with *semPlot* package (Epskamp, 2014).

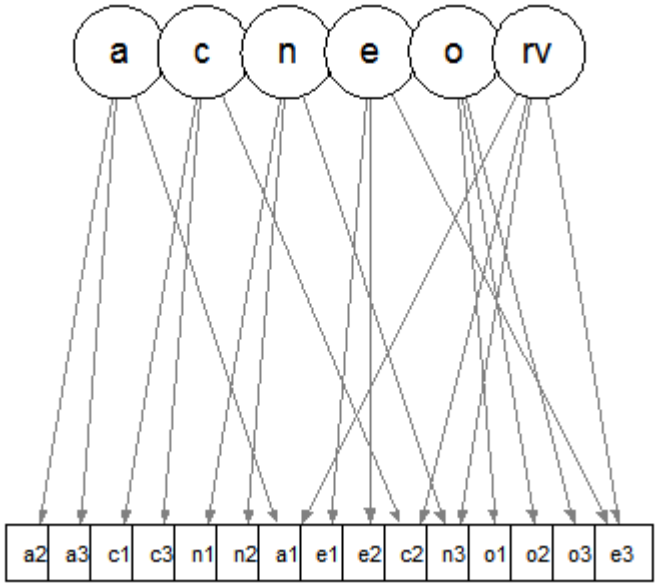
Figure E.1 A path diagram of the Grit-S bi-factor solution



Notes: grt – Grit, eff – Perseverance of Effort, int – Consistency of Interest.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Figure E.2 A path diagram of the modified BFI-S solution



Notes: a – Agreeableness, c - Conscientiousness, n – Neuroticism, e – Extraversion, o – Openness, rv – factor loading reversely-worded items.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table E.3 Factor loadings of the modified BFI-S solution

Latent variable	Item	Std. estimate	SE	Estimate/SE
a				
	a1	1.268	0.089	14.174
	a2	0.395	0.013	29.529
	a3	0.652	0.012	55.480
c				
	c1	0.696	0.011	62.078
	c2	0.952	0.043	22.268
	c3	0.802	0.010	82.327
n				
	n1	0.503	0.014	34.975
	n2	0.791	0.017	46.969
	n3	0.606	0.022	27.255
e				
	e1	0.551	0.015	37.080
	e2	0.744	0.017	44.160
	e3	0.288	0.023	12.487
o				
	o1	0.693	0.012	58.340
	o2	0.569	0.013	45.191
	o3	0.597	0.013	47.713
rv				
	e3	0.531	0.022	23.823
	u1	1.133	0.089	12.717
	s2	0.599	0.045	13.434
	n3	0.596	0.021	29.042

Notes: a – Agreeableness, c- Conscientiousness, n – Neuroticism, e – Extraversion, o – Openness, rv – "reversed" factor.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table E.4 Latent correlations in the modified BFI-S solution

Correlations	Std. estimate	SE	Estimate/SE
e with n	0.197	0.019	10.163
e with o	0.587	0.017	34.961
e with c	0.315	0.019	16.835
e with a	0.328	0.021	15.821
n with o	0.087	0.019	4.613
n with c	0.138	0.019	7.191
n with a	0.090	0.023	3.998
o with c	0.493	0.016	30.409
o with a	0.513	0.018	28.996
c with a	0.818	0.011	72.439
rv with e	-0.459	0.021	-22.258
rv with n	-0.390	0.023	-17.080
rv with o	-0.590	0.016	-35.840
rv with a	-0.738	0.019	-38.621
rv with c	-0.766	0.022	-35.075

Notes: a – Agreeableness, c- Conscientiousness, n – Neuroticism, e – Extraversion, o – Openness, rv – "reversed" factor.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table E.5 Parameters of the Grit-S bi-factor model

	Std. estimate	SE	Estimate/SE
Consistency of interest			
New ideas and projects sometimes distract me from previous ones*	0.252	0.036	6.953.
I have been obsessed with a certain idea or project for a short time but later lost interest*	0.561	0.044	12.863
I often set a goal but later choose to pursue a different one*	0.337	0.049	6.859
I have difficulty maintaining my focus on projects that take more than a few months to complete*	0.115	0.059	1.934
Perseverance of effort			
Setbacks don't discourage me	0.196	0.017	11.866
I am a hard worker	0.641	0.014	45.996
I finish whatever I begin	0.535	0.021	25.557
I am diligent	0.768	0.016	46.633
Grit			
New ideas and projects sometimes distract me from previous ones	0.273	0.025	10.981
Setbacks don't discourage me	0.519	0.032	16.098
I have been obsessed with a certain idea or project for a short time but later lost interest	0.619	0.022	27.899
I am a hard worker	0.684	0.027	25.511
I often set a goal but later choose to pursue a different one	0.170	0.017	10.309
I have difficulty maintaining my focus on projects that take more than a few months to complete	0.328	0.020	16.644
I finish whatever I begin	0.559	0.024	23.179
I am diligent	0.458	0.022	20.920

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

ANNEX F. CRITERION VALIDITY ANALYSIS – ADDITIONAL RESULTS

Table F.1 Educational attainment and cognitive and non-cognitive skills (linear regression) without controlling for SES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Numeracy (std)		0.447***			0.428***	0.447***	0.426***
Conscientiousness (std)			-0.122*		-0.066		-0.117*
Extraversion (std)			-0.138***		-0.088**		-0.088**
Agreeableness (std)			-0.089		-0.085		-0.065
Openness (std)			0.310***		0.237***		0.237***
Neuroticism (std)			-0.073***		-0.048*		-0.036
Grit (std)				0.068**		0.069**	0.087***
Observations	4 454	4 454	4 454	4 454	4 454	4 454	4 454
R^2	0.086	0.277	0.128	0.090	0.299	0.282	0.305

Notes: * p < 0.05, ** p < 0.01, *** p < 0.001. Control variables: age, age squared, gender. First PV (plausible value) for numeracy. Years of education, numeracy and non-cognitive skills are standardised.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table F.2 Self-declared labour force participation and cognitive and non-cognitive skills (logit)

Odds ratios showing the likelihood of being active on the labour market

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Numeracy (std)		1.019			1.015	1.022	1.016
Conscientiousness (std)			1.396*		1.397*		1.366
Extraversion (std)			1.009		1.010		1.009
Agreeableness (std)			0.782		0.782		0.788
Openness (std)			1.063		1.062		1.063
Neuroticism (std)			0.794***		0.795***		0.800***
Grit (std)				1.120*		1.120*	1.041
Observations	4 452	4 452	4 452	4 452	4 452	4 452	4 452
McFadden's R^2	0.263	0.263	0.271	0.265	0.271	0.265	0.271
McKelvey and Zavoina's R^2	0.398	0.399	0.410	0.400	0.410	0.400	0.410
BIC	18513299.6	18512499.4	18315512.5	18477083.2	18315007.6	18476001.7	18311100.1

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Control variables: age, age square, gender, years of education. First PV (plausible value) for numeracy. Numeracy and non-cognitive skills are standardised.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table F.3 Self-declared employability and cognitive and non-cognitive skills (logit)

Odds ratios showing the likelihood of being employed

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Numeracy (std)		1.280**			1.269**	1.280**	1.269**
Conscientiousness (std)			1.553*		1.542*		1.504*
Extraversion (std)			0.933		0.950		0.950
Agreeableness (std)			0.680		0.688		0.695
Openness (std)			0.957		0.943		0.944
Neuroticism (std)			0.950		0.959		0.965
Grit (std)				1.063		1.064	1.044
Observations	3 224	3 224	3 224	3 224	3 224	3 224	3 224
McFadden's R^2	0.0614	0.0676	0.0667	0.0619	0.0724	0.0681	0.0726
McKelvey and Zavoina's R^2	0.127	0.136	0.136	0.127	0.143	0.135	0.142
BIC	9567001.0	9503830.7	9513342.6	9561730.9	9455047.8	9498411.0	9452723.1

Notes: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Control variables: age, age square, gender, years of education. First PV (plausible value) for numeracy. Numeracy and non-cognitive skills are standardised.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table F.4 Incrementality analysis – KHB results

	(1) Labour force participation	(2) Employment	(3) Job satisfaction
Conscientiousness (std)			
Reduced	1.565**	1.440	1.587**
Full	1.560**	1.457	1.414*
Diff	1.003	0.988	1.122
Extraversion (std)			
Reduced	0.957	1.031	1.287**
Full	0.957	1.031	1.288**
Diff	1.000	1.000	0.999
Agreeableness (std)			
Reduced	0.717*	0.695	0.771
Full	0.718*	0.691	0.805
Diff	0.999	1.005	0.959
Openness (std)			
Reduced	1.123	0.965	1.007
Full	1.123	0.965	1.019
Diff	1.000	1.001	0.988
Neuroticism (std)			
Reduced	0.747***	1.083	0.784***
Full	0.748***	1.080	0.806***
Diff	0.999	1.003	0.973
Observations	4 454	3 356	3 026
Conf.-Ratio Con.	1.006	0.969	1.331
Conf.-Perc.	0.604	-3.232	24.88
Conf.-Ratio Ext.	0.999	1.004	0.996
Conf.-Perc.	-0.105	0.436	-0.366
Conf.-Ratio Agr.	1.003	0.987	1.194
Conf.-Perc.	0.307	-1.333	16.25
Conf.-Ratio Opn.	0.999	0.981	0.368
Conf.-Perc.	-0.116	-1.902	-171.5
Conf.-Ratio Neu.	1.002	1.033	1.128
Conf.-Perc.	0.220	3.202	11.37

Table F.4 Incrementality analysis – KHB results (*continued*)

	(4) Health	(5) Trust	(6) Life satisfaction
Conscientiousness (std)			
Reduced	1.403**	0.871	1.765***
Full	1.219	0.978	1.412**
Diff	1.151**	0.891**	1.250***
Extraversion (std)			
Reduced	1.186**	1.122	1.281***
Full	1.184**	1.124	1.276***
Diff	1.002	0.998	1.004
Agreeableness (std)			
Reduced	0.862	1.070	0.807*
Full	0.908	1.025	0.878
Diff	0.949	1.044	0.920
Openness (std)			
Reduced	0.915	0.895	0.986
Full	0.921	0.890	0.998
Diff	0.993	1.006	0.989
Neuroticism (std)			
Reduced	0.689***	0.831***	0.761***
Full	0.712***	0.809***	0.802***
Diff	0.967	1.028	0.949
Observations	4 454	4 454	4 454
Conf.-Ratio Con.	1.708	6.142	1.647
Conf.-Perc.	41.45	83.72	39.29
Conf.-Ratio Ext.	1.014	0.983	1.016
Conf.-Perc.	1.417	-1.732	1.557
Conf.-Ratio Agr.	1.548	2.748	1.642
Conf.-Perc.	35.41	63.61	39.09
Conf.-Ratio Opn.	1.087	0.950	5.568
Conf.-Perc.	7.965	-5.292	82.04
Conf.-Ratio Neu.	1.098	0.872	1.239
Conf.-Perc.	8.890	-14.71	19.27

Notes: Exponentiated coefficients; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).

Table F.5 Summary of skill /trait – outcome relationships with traits measured by sum scores by factor

	Education		LF participation		Employability		Wages	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Numeracy (std)	+++	+++					++	++
Conscientiousness (std)			++	++			+	+
Extraversion (std)								
Agreeableness (std)	-	-					--	--
Openness (std)	+++	+++						
Neuroticism (std)	-		---	---			--	--
Grit (std)	x	+++	x		x		x	

Table F.5 Summary of skill /trait – outcome relationships with traits measured by sum scores by factor (*continued*)

	Job satisfaction		Health		Trust		Life satisfaction	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Numeracy (std)			+++	+++	+++	+++		
Conscientiousness (std)	++	+	+		--		+++	+++
Extraversion (std)	++	++	+	+			+++	+++
Agreeableness (std)					++	++		
Openness (std)					-	-	++	+
Neuroticism (std)	---	---	---	---	---	---	---	---
Grit (std)	x	++	x	+++	x	--	x	+++

Notes: +/- $p < 0.05$, ++/-- $p < 0.01$, +++/--- $p < 0.001$ (1) Numeracy and Big Five; (2) Numeracy, Big Five and Grit.

Source: Polish Follow-up Study on the Programme for International Assessment of Adult Competencies (postPIAAC) (2015).