



NATIONAL RESEARCH UNIVERSITY
HIGHER SCHOOL OF ECONOMICS

*Yulia Tyumeneva, Julia Kuzmina,
Elena Kardanova*

IRT ANALYSIS AND VALIDATION OF THE GRIT SCALE: A RUSSIAN INVESTIGATION

BASIC RESEARCH PROGRAM

WORKING PAPERS

SERIES: PSYCHOLOGY
WP BRP 24/PSY/2014

Yulia Tyumeneva¹, Julia Kuzmina², Elena Kardanova³

IRT ANALYSIS AND VALIDATION OF THE GRIT SCALE: A RUSSIAN INVESTIGATION⁴

Previous work has already established the reliability and validity of the Grit scale, based on the classical test theory (Duckworth, Peterson, Matthews, & Kelly, 2007). This study tests the psychometric properties and the dimensionality of the scale by employing an IRT analysis of data collected from a Russian sample of 3383 15 year-old students (Study 1). The results showed that the Grit Scale is two-dimensional, but unlike the early results, both the Consistency of Interests and Perseverance of Effort subscales are quite separate measures for different traits. The construct validation (Study 2) included a retest of the obtained internal structure of the scale via CFA, and a determination of whether two subscales show reasonable patterns of correlations with external measures. Collectively, the data supports the reliability and the validity of both subscales as separate measures.

JEL Classification: Z.

Keywords: grit scale, dimensionality, IRT, CFA, construct validity.

¹ National Research University Higher School of Economics, the Graduate School of Education (HSE), the International Laboratory for Educational Policy Research: Senior researcher.

² National Research University Higher School of Economics, the Graduate School of Education (HSE), the International Laboratory for Educational Policy Research: Research Fellow.

³ National Research University Higher School of Economics, the Graduate School of Education (HSE), the Center of Education Quality Monitoring, Director.

⁴ Financial support from the Basic Research Program of the National Research University Higher School of Economics is gratefully acknowledged.

Introduction

Grit, persistence, hardiness and tenacity belong to those personality traits which are typically associated with human attainment. Due to modern society's current interest in people's achievements, several questionnaires have been developed to assess people's attainment-oriented personal traits. The Grit scale, developed by A. Duckworth et.al (2007, 2009) and her colleagues (Duckworth, et al., 2007; Duckworth & Quinn, 2009) is one within a family of similar self-report measures, which focus on people's abilities to keep maintain long-term goals and overcome difficulties. Some of these questionnaires refer to entities not exactly covering exactly the concept of grit: Personal Views Survey III-R (Maddi et al., 2006); The tenacity scale by Baum and Locke (2004); A motivational persistence scale (Constantin, et al., 2012); The Passion Scale (Vallerand et al., 2003), the KUT (Klein et al., Unidimensional, Target-free) measure of commitment (Klein, et al., 2014). Other instruments have been developed for specific populations, such as a scale for measuring persistence in children (Lufi & Cohen, 1987), or the Career Advancement Ambition Scale (DesRochers & Dahir, 2000). Some scales are out of date, as in the questionnaire measure of persistence disposition (Mukherjee, 1974), or a scale for measuring persistence (Wang, 1932). Some of the modern grit-relevant scales, which are part of the multidimensional questionnaires, consist of only a few items (e.g. The Motivation and Engagement Scale by Liem and Martin (2012)), which can enlarge the error of measurement significantly.

The Grit Scale, developed by Duckworth et. al (2007), has an advantage over scales from this family by having, as the authors claim, face validity for varied populations, a low probability of ceiling effects in high-achieving groups, and a precise focus on the construct of grit (Duckworth, et al., 2007). The authors assume that grit as a personality trait is composed of two underlying constructs: the ability to overcome difficulties and a stability of interest over time. The latter provides a distinguishing of the long-term interest in goal achievement as a key component of grit from other possible motivations, e.g. fear of censure or avoidance of changes completely.

According to the theory, the Grit Scale consists of two subscales: Consistency of Interests and Perseverance of Effort – each consists of 6 items (in the full version) or 4 items (in the short 8-item version (Duckworth, et al., 2007; Duckworth & Quinn, 2009)). The items are based on

behavior characteristics of high-performing individuals given by professionals of varied domains.

Both versions of the scale have demonstrated very good psychometric characteristics: $\alpha = 0.85$ for the overall scale and 0.84 and 0.78 for the subscales of Consistency of Interests and Perseverance of Effort, respectively (and for the short version 0.73 and 0.83, respectively). Those researchers who used the scale in their own studies reported the alpha 0.77 – 0.80 for the full scale, 0.63 – 0.80 for Consistency of Interests, and 0.60 – 0.79 for Perseverance of Effort (Pozzebonet al., 2013; Maddi, et al., 2012; Eskreis-Winkler, et al., 2014).

However, it is interesting that despite quite a low percentage of the total variance ($\approx 20\%$) shared between the Consistency of Interests and Perseverance of Effort factors reported by Duckworth et al (2007), authors tend to treat their instrument as single, using the overall score for the scale in all future calculations. Unfortunately, we have not found any other data on the correlation between the factors, but we believe that the low correlation between subscales may suggest that the scale measures two overlapping but distinct constructs, rather than one common trait at least for the studied samples. Anyway, based on the available data a question of the dimensionality requires additional verification.

In terms of validity, the Grit score has predicted educational attainment; the number of lifetime career changes; the completion of some training programs (Duckworth, et al., 2007); retention in the military, workplace, school and marriage (Eskreis-Winkler, et al., 2014), students' metacognition (Arslan et al., 2013); teachers' efficiency (Robertson-Kraft, et al., 2014; Duckworth, et al., 2009); the well-being of residents in general surgery (Salles, et al., 2014), and job satisfaction in doctors (Reed, et al., 2012).

Despite the impressive psychometric and predicting characteristics obtained for the Grit scale we've not found any supporting data from the modern test theory – Item Response Theory (IRT). It's well known that the Classical test theory (CTT) exploits several serious assumptions about test scores and the standard error of measurement. First, under CTT the sum of item scores is treated as a measure of trait, without any justification whether it is appropriate for this purpose or not. Second, assessment of item and scale properties heavily relies on the sample being used. It means that any changes in sample characteristics would lead to changing in psychometric features of both item and the total scale characteristics. Third, the standard error of measurement

is treated as equal across all trait levels, while the error (or, in other words, precision of measurement) as well other item and test statistics differ throughout the trait continuum.

IRT methods and procedures overcome these limitations (Embretson & Reise, 2000). Moreover, from practical point of view, a big advantage of IRT modeling is that it provides a way for constructing personality and cognitive ability instruments and examining their measurement characteristics, including investigation of the quality of response categories in Likert-type scales and dimensionality study (Linacre, 2002; Smith, 2002). So we resorted to the IRT approach to overcome the restrictions of CTT and strengthen a methodological foundation of the Grit scale.

Aside from developing a Russian version of the Grit scale, which would be valuable in itself, in this study we also aimed to examine the reliability and dimensionality of the scale through an IRT analyses (Study 1). Once the psychometric quality of the Grit scale was confirmed, we estimated all respondents using the IRT scaling. The students' estimates are on an interval scale, using a scale unit called logit. Using the students' IRT estimates, we confirmed the multidimensional structure through a confirmatory factor analysis. A further construct validation (Study 2) included the determination of whether two subscales show reasonable patterns of correlations with external measures.

Study I. Developing, IRT and confirmatory factorial analysis of the Russian version of the Grit scale

Developing instrument. The procedure of the Grit scale translation into Russian was in line with the guide of Van de Vijver and Hambleton (1996) and included a primary translation into Russian, a review of this translation by psychologists experienced in the field and a small pilot group of respondents, to reach of a reconciled version of translation, and back translation to English language and comparison of the original and the backtranslated version. At the stage of reviewing the primary Russian version the participants were asked in detail about a questionnaire, clarity of items, the convenience of the response scale, and their impression of the questions.

As a result of this process two items were reformulated to be more consistent with cultural features of perception of diligence as a personal trait. Although a diligent person is held

in respect and diligence is an encouraged trait which is encouraged in Russian culture, calling yourself a diligent person or hard worker is perceived as bragging. Openly calling somebody a diligent worker is to be understood as antiphrasis, in an ironic or derisive sense. This was why two items with a direct reference to these traits (“*I am a hard worker*” and “*I am diligent*”) were received with resistance by the participants. The items were reformulated to reduce their straightforwardness and categoricalness: “*I am a hard worker*” → “*Without irony I am a hard worker*”, “*I am diligent*” → “*At work I am diligent*”.

We also found that the item “*Setbacks don’t discourage me*” confused some of our respondents because of its double negative wordings in conjunction with the negative answer options (“*Do not agree at all*” and “*Not fully agree*”). To avoid the confusion the item has been reformulated to be in positive wording. We also introduced “as a rule” to emphasize regularity of the situation. Eventually the item was worded: “*As a rule, setbacks do discourage me*”.

Except these three items, in other respects the backtranslation and the original form were found to be remarkably similar (See the items of the original version of the Grit Scale in Appendix A).

IRT modeling of the Grit scale

Method

Participants

A sample of Russian students ($n=3383$, 49.3% were girls and 50.7% were boys) from the 9th grades of regular secondary schools took part in a longitudinal national survey of school and university graduates (Kuzmina et al., 2012), where the Grit scale was implemented. The mean age of the students was 15.13 ($SD = 0.47$).

Instrument

The developed Russian version of the Grit scale included 12 items that referred to 2 subscales: 6 items to Consistency of Interests and 6 items to Perseverance of Effort. Each item is rated on a 5-point Likert scale ranging from 1 (*do not agree at all*) to 7 (*completely agree*). Table 1 shows the 12 items of the Grit scale.

Table 1. The twelve items of the Grit scale.

Consistency of Interests

I often set a goal but later choose to pursue a different one*

New ideas and new projects sometimes distract me from previous ones*

I become interested in new pursuits every few months*

My interests change from year to year*

I have been obsessed with a certain idea or project for a short time but later lost interest*

I have difficulty maintaining my focus on projects that take more than a few months to complete*

Perseverance of Effort

Without irony I am a hard worker

I have achieved a goal that took years of work

I have overcome setbacks to conquer an important challenge

As a rule, setbacks do discourage me

I finish whatever I begin

At work I am diligent

* Item was reverse scored.

IRT model

The Grit scale employs a 5-point rating scale with 1 point for Strongly Disagree category to 5 points for Strongly Agree category. To construct test and scales we use Rating Scale Model that is a Rasch-type model designed especially for Likert-type scales. The RSM is an extension of the Rasch dichotomous model for polytomous items. According to Linacre (2002),

$$\log(P_{nik}/P_{ni(k-1)})=B_n - D_i - F_k ,$$

where P_{nik} is the probability that person n on encountering item i would response in category k , $P_{ni(k-1)}$ is the probability that the response would be in category $k-1$, B_n is the ability of person n , D_i is the difficulty of item i and F_k is the k^{th} step calibration, i.e. a rating scale threshold defined as the location corresponding to the equal probability of observing adjacent categories $k-1$ and k .

A detailed description of the model can be found in Wright and Masters (1982). The reasons for choosing the RSM are both psychometrical and practical. Firstly, the RSM is the Rasch model and so has optimal metric properties, and secondly, from practical point of view, it is useful to empirically determine the quality of response categories in Likert-type scales (Bond & Fox, 2001).

Winsteps software that is a Rasch-model computer program was used for parameters estimation, analysis of the quality of response categories and scales construction (Linacre, 2003).

Results of IRT analysis

Dimensionality study. At the first stage of the research we examined the dimensionality of the Grit scale by conducting a principal component analysis (PCA) of the standardized residuals (Linacre, 1998; Ludlow, 1985; Smith, 2002). Theoretically, if the unidimensionality assumption is hold, then correlations between item-level residuals should be near zero. If there is no second dimension remaining in the residual variation, then the principal component analysis should generate eigenvalues all near one and the percentage of variance across the components should be uniform. The analysis of the eigenvalues of the Grit scale residual correlation matrix for 12 components revealed that there is one component with eigenvalue of 3.3 while eigenvalues for other components ranged from 1.3 to 0.6 (the exception was the eigenvalue for the last component that was 0.012). In addition, the percentage of variance for the first component was 27% while for other components the variance was roughly evenly split across components.

Thus, the scale is two-dimensional and consists of two subscales. Table 2 shows standardized residuals loadings for all items.

Table 2. Standardized residuals loadings for the Grit test items

Item	Loading
I often set a goal but later choose to pursue a different one*	-0.54
Without irony I am a hard worker	0.50
New ideas and new projects sometimes distract me from previous ones*	-0.28
I have achieved a goal that took years of work	0.60
I become interested in new pursuits every few months*	-0.07
I have overcome setbacks to conquer an important challenge	0.57
My interests change from year to year*	-0.43
As a rule, setbacks do discourage me*	-0.52
I have been obsessed with a certain idea or project for a short time but later lost interest*	-0.64
I have difficulty maintaining my focus on projects that take more than a few months to complete*	-0.44
I finish whatever I begin	0.67
At work I am diligent	0.70

* Item was reverse scored.

In accordance with previous research on the Grit scale (Duckworth, Peterson, Matthews, & Kelly, 2007; Duckworth & Quinn, 2009), it consists of two subscales; Consistency of Interests and Perseverance of Effort. Our findings confirm the two-factor structure, but the item “*As a rule, setbacks do discourage me*” moved from the Perseverance of Effort to the Consistency of Interests subscale. Subsequently, the Perseverance of Effort subscale consisted of 5 items, while the other 7 items formed the Consistency of Interests subscale. The item “*I become interested in new pursuits every few months*” has a negative, but very small loading. It will be analyzed more carefully through a confirmatory factor analysis and reported in Study 2.

Therefore, the analysis showed that the Grit scale is multidimensional and consists of two subscales. Further analysis should be conducted on each subscale separately.

Analysis of the Consistency of Interest subscale

The Consistency of Interest subscale consists of seven items with negative loadings, as shown in Table 2. First, we conducted the dimensionality study for the subscale using the same approach as for whole Grit scale. The subscale is unidimensional and can be productive for measurement if it has good psychometric properties. We then conducted a series of analyses aimed at investigating the quality of the items and the subscale as a whole.

We began the analysis of the scale from a model fit analysis. The Rasch goodness-of-fit analyses rely principally upon standardized residuals – the difference between the observed response and the response expected under the model (Wright & Masters, 1982). The residuals are squared and summarized in the form of un-weighted and weighted mean squares (in terms of Winsteps output: OUTFIT MNSQ and INFIT MNSQ, respectively). Generally, items with values between 0.7 and 1.3 are considered an acceptable indication of good fit between the data and the model RSM. Our analysis showed that the values of both INFIT and OUTFIT MNSQ item statistics for all items are in an acceptable range, with a mean of 1.00, where SD = 0.11 for the INFIT MNSQ statistics, and a mean 1.02 and SD=0.12 for the OUTFIT MNSQ statistics. This result indicates that all items in our subscale (including the item “*I become interested in new pursuits every few months*” mentioned above) fit the model, in accordance with the selected criteria. (Appendix B provides summary statistics on each item of the Consistency of Interests subscale).

A similar approach was used to analyze person-fit. The mean and standard deviation for the fit statistics were 1.02 and 0.74 (INFIT), respectively, and 1.02 and 0.75 (OUTFIT),

respectively. Taking into account the small number of items in the scale, even one or two unexpected responses could generate a large value of person-fit statistics. Therefore, we decided to assign the value of 2.0 as a critical value for MNSQ person statistics, rather than using 1.3 for the MNSQ item statistics. Values greater than 2.0 signal the distortion of the measurement system (noise is greater than useful information), while values in the range of 1.5 – 2.0 mean are unproductive for measurement, but not degrading.

In total, six hundred thirty six subjects (17% of the sample) were removed based on the 2.0 criteria for the MNSQ person statistics. When we analyzed the response profiles of these students, we confirmed that large values of fit statistics were caused by one or two unexpected responses, which could be explained by carelessness or by chance. After excluding these people, the mean and standard deviation for the fit statistics became 1.00 and 0.54 (both INFIT and OUTFIT), respectively. Further analysis of the subscale properties was therefore conducted with a sample of 3191 individuals.

The next step was to investigate the quality of the response categories. In the RSM context, a set of criteria had been posed for verifying the useful functioning and diagnosing malfunctioning of a rating scale: (a) sufficient frequency and regular distribution of responses (the optimal is a uniform distribution); (b) the average measures by which a category should increase monotonically with the rating scale categories; (c) categories should fit the model well ; and (d) the Rasch step calibrations (or threshold estimates) must advance monotonically with the categories (if this is not so, then some of the categories are inefficient).

Table 3 provides a summary of the rating scale’s category structure and indicates that the five categories’ response system is adequate.

Table 3. Summary of category structure of the rating scale.

Category number	Category label	Observed count	Observed %	Average measure	Infit mnsq	Outfit mnsq	Threshold estimate
1	SA	1585	7	-1.15	0.98	0.98	NONE
2	A	5015	23	-0.44	1.01	1.01	-1.95
3	N	6562	30	0.10	0.99	1.01	-0.44
4	D	6272	29	0.65	1.00	1.00	0.41
5	SD	2491	11	1.43	1.00	1.00	1.98

The same information can also be seen in Figure 1, which shows the category probability curves for the scale. We can observe that thresholds are ordered and each category has the

probability of being selected by the respondents. This implies that, as one moves up the ability continuum, each category in turn becomes the most probable response.

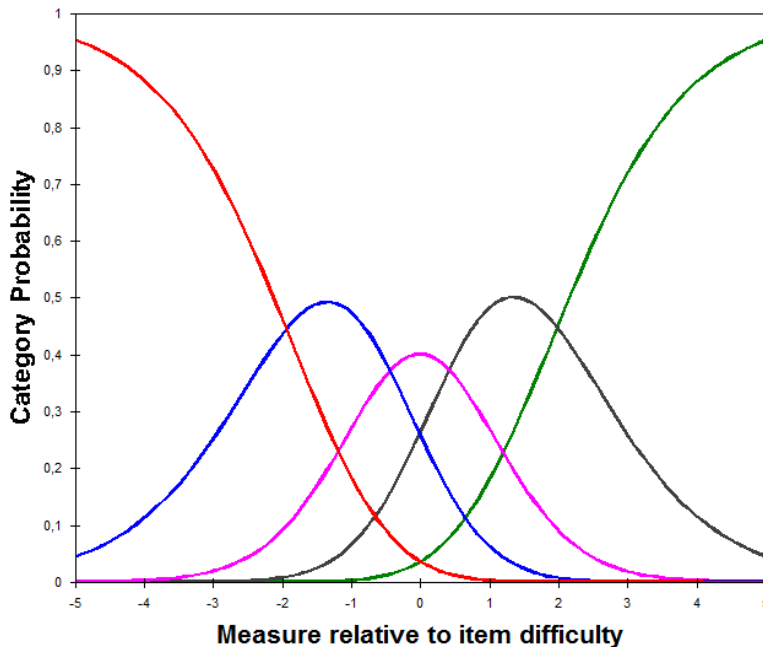


Figure 1. Category probability curves for the rating scale

In summary, the evidence suggests effective functioning of the categories of the Consistency of Interest subscale.

The next section of analysis was devoted to the properties of the whole scale. The Rasch analysis provides a reliability index and alternative statistics, a separation index, that characterize the ability of the scale to reliably distinguish between individual participants (Wright & Masters, 1982; Stone, 2004). The person reliability is 0.71, meaning that the proportion of observed student variance considered true is 71%. (This index is close in value and interpretation to the classical reliability: $\alpha = 0.76$ for the subscale).

The last IRT analysis was performed to investigate whether the items' functioning varies across male and female respondents, that is, whether a differential item functioning (DIF) exists across gender groups. Therefore all items were analyzed on DIF.

The DIF-analysis of the Consistency of Interest subscale items was conducted with two statistics available in Winsteps. These are the Student's t-test and Mantel-Haenzel statistics (Smith. R. M., 2004; Dorans N.J., 1989). The results indicated that the items do not demonstrate significant DIF in regard to representatives of different gender (see Appendix B for details).

In conclusion, the 7-item Consistency of Interest subscale shows good psychometric properties and can be used for a person's measurement.

Analysis of the Perseverance of Effort subscale

The Perseverance of Effort subscale consists of 5 items with positive loadings in Table 1. All steps of this subscale analysis were similar to the analysis of the Consistency of Interest subscale. Only a condensed version of the full results is presented here (see Appendix C for details).

Seven hundred sixty-one subjects (20% of the sample) were removed, based on the same criteria as for the first subscale. Therefore, an analysis of the subscale properties was conducted with a sample of 3066 persons. The subscale is unidimensional, all items fit the model, and all categories function effectively. The person reliability is 0.78, and classical reliability is $\alpha = 0.84$ for the subscale.

The 5-item Perseverance of Effort subscale also shows good psychometric properties and can be used to measure this person's characteristic.

Discussion of Study 1

This study has showed that the Grit scale consists of two subscales; the Consistency of Interests and the Perseverance of Effort, as predicted based on Duckworth's report (Duckworth, et al., 2007). The two subscales both have good psychometric quality, meaning that both subscales are uni-dimensional and have high reliability, and that all categories function effectively, since the items do not demonstrate DIF. The subscales can therefore be used to measure a respondents' correspondent traits.

An estimation of respondents' measures was conducted separately for each subscale, which resulted in two different estimates for each person; the measure of consistency of interest and the measure of perseverance of efforts. We revealed that the correlation between respondents' measures on two subscales was 0.24. This implies that a low value of correlation does not allow us to construct a common measure of one trait with this scale. The two subscale provide rather measures of different traits than one. The first distinction from Duckworth's results is therefore that a common measure of grit cannot be constructed.

The second distinction we found was the migration of the item «*As a rule, setbacks do discourage me*» from the Perseverance of Effort to the Consistency of Interests subscale. As

previously mentioned, this item underwent two changes in comparison with its original form: the positive wording and the addition of the time adverb “as a rule”, and both changes might result in the item’s migration.

First, after positive rewording, the item became reversed towards the Perseverance of Effort subscale and therefore was recoded. The recoding itself might cause the migration to the other subscale “Consistency of interest” where all items were reversed towards their subscale and therefore recoded. This tendency of pooling of reversed items in a separate factor has been extensively discussed for the last decades (Bagozzi, 1993; Marsh, 1996; Herche and Engelland 1996).

The inclusion of the time adverb “*as a rule*” to the wording of the item emphasized the trait’s regularity and consistency. As a result, the migration of the item could be related to the effect of the new wording (adverbs of regularity) rather than to the content of the items (perseverance – consistency). Indeed, the time adverbs (regularity) such as “*usually*”, “*sometimes*”, “*often*”, “*from time to time*” can be seen in every item of the Consistency of Interest subscale, but they are absent from the Perseverance of Effort subscale. Accordingly, we consider the wording effect as the main reason for the migration of the item. Nonetheless, even with this structure, both subscales showed their effectiveness as measures of corresponding traits.

Study 2. Construct validation of the Grit scale

The aim of this study was to examine: 1) the factor structure of the Grit scale via confirmatory factor analysis (CFA), and 2) the associations between grit and several other constructs which were theoretically and empirically linked with grit, such as attainment, retention, and honor in different training activities. Based on our results obtained through IRT analyses we could expect that the scale provides two separate measures for grit and, thus, all criterion-related evidences shown earlier for the Grit scale validity are assumed to be questionable.

Method

Participants

The sample included 2322 students from Study 1 (50.5% were girls and 49.5% were boys). We excluded the remaining 1055 cases because they did not fit the person statistics shown earlier. The mean age of participants was 15.6 years.

Procedures and instruments

The students of this study were participants in both the TIMSS-2011 and PISA-2012 studies, and were also included in the sample from the National Survey of school and university graduates, as mentioned previously. The instruments to assess the grit-related constructs came from these surveys and were used as a part of them.

To assess educational achievement, we used the TIMSS (8 grade)⁵ and PISA⁶ results (plausible values). Both TIMSS and PISA metrics for the overall mathematics, science or reading scales are based on a mean of 500 points and a standard deviation of 100 points (Mullis et al., 2009; OECD, 2014b). Both TIMSS and PISA use five plausible values for each respondent, indicating possible “true” values of the student’s score on the underlying conceptual dimension (reading, mathematics, science). To produce estimates using plausible values correctly, we use the recommended statistical procedures for calculating coefficients and standard errors (Davier, Gonzalez, Mislevy, 2009).

We measured other school outcomes that are used in several studies, confirming validity measures for grit-related constructs (such as perseverance and tenacity). When students were completing questionnaires of the national survey they indicated whether they had repeated any grades, taken advanced courses in any school subjects, or participated in school subject competitions or subject Olympiads. They also indicated the period of time they spent doing out-of-school educational activities (music, art and sport) (“If you practice music, art or sport, how long have you been doing this?”).

In Russia, a nine-year education is compulsory and after the ninth grade, students have an opportunity to choose whether to stay in high school for two more years and then start university after graduating, or to apply to a vocational school (a technical college) and to receive a vocational education. Alternatively, students may finish their formal education and get a job. We questioned the students in a telephone follow-up study conducted 7 months after the PISA survey, and asked whether they had stayed in high school, moved to vocational education or finished their schooling after the 9th grade. Although we have no strong evidence that shifting to

⁵ Trends in International Mathematics and Science Study measures mathematics and science achievement at the fourth and eighth grades.

⁶ The Programme for International Student Assessment is OECD project assessing competency in three domains: reading, mathematics and science in 15-years-old students.

vocational education itself indicates a lack of grit, there are a great number of studies which claim that mainly low-achieving students follow a vocational trajectory (Altinok, 2011; Kuzcera, 2008). As we demonstrated earlier, student achievements can be predicted by grit-relevant measures, and so it is reasonable to presume that the choice between general and vocational education after 9th grade can be indirectly predicted by grit.

The PISA survey also provided an Index of “Perseverance” which we used to examine a convergent validity of the Grit scale. PISA measures students’ perseverance through their responses to questions about the extent to which they feel they resemble someone who gives up easily when confronted with a problem; who puts off difficult problems; who remains interested in the tasks that he or she starts; who continues to work on a task until everything is perfect, and who does more than is expected of them when confronted with a problem. Student responses ranged on a 5-point Likert scale from “this person is very much like me” to “not at all like me”. The Index of Perseverance provided by PISA is standardized to have a mean of 0 and a standard deviation of 1 across all OECD countries (OECD, 2014a). The observed internal reliability of the PISA perseverance scale (Cronbach’ Alpha) was 0.49 for the Russian sample.

The Russian version of the Grit scale developed in Study 1 was used to assess the consistency of interests and perseverance of effort. The students’ estimates of Consistency of Interests and Perseverance of Effort in logits were used in this study.

Analyses

We tested two models: 1) Model 1 was an initial theoretical model with two factors, each with six items; 2) Model 2 was the two-factor model obtained in Study1, where the Consistency of Interests has seven items and the Perseverance of Effort has five. These models were tested via a confirmatory factor analysis, using Mplus 6.12 (Muthen & Muthen, 2011).

To examine the construct validity we, first, used a correlation analysis of performance in TIMSS and PISA on the one hand and Consistency of Interests and Perseverance of Effort scores on the other hand. Second, we used ANOVA to compare the means for both the Consistency of Interests and Perseverance of Effort subscales in the groups which were contrasted by educational outcomes; grade repetition, general school retention vs. vocational school, attending advanced courses, participating in subject competition/Olympiads, duration of practicing music (or art) and sport.

Third, we assessed the convergent validity as a part of construct validity via correlations between the Consistency of Interests and Perseverance of Effort scores with the PISA Index of Perseverance.

Results

We provided a summary for descriptive statistics in Appendix D.

Confirmatory Factor Analysis

The results of the confirmatory factor analysis showed that Model 2, obtained by an IRT analysis in Study 1, has better fit indices than Model 1 (Table 4). These results support the two-factor model of the Grit scale, with 5 items for the Perseverance of Effort subscale and 7 items for the Consistency of Interests, which differed from the initial model supposed by A. Duckworth.

Table 4. Fit indices for the estimated models.

Indices	Model 1	Model 2
χ^2	667.6	463.5
df (degree of freedom)	53	53
RMSEA (Root Mean Square Error of Approximation)	.07	.06
RMSEA (90% C.I.)	.06/.07	.05/.06
CFI (Comparative Fit Index)	.87	.92
SRMR (Standardized Root Mean Square Residual)	.84	.05

Correlations with external measures

The following validation consisted of several parts. Firstly, both Consistency of Interests and Perseverance of Effort subscales were correlated with the TIMSS and PISA performance. We found small positive and significant correlations (r s ranged 0.05 – 0.11, $p < 0.05$ at least) between the subscales and TIMSS and PISA achievements (Table 5).

Table 5. Correlations between Consistency of Interests and Perseverance of Effort subscales and achievements in TIMSS and PISA.

Achievement	Consistency of Interest	Perseverance of efforts
TIMSS Math	0.10**	.09**
TIMSS Science	0.10**	.09**
PISA Math	0.11**	.07**
PISA Science	0.10**	.05*
PISA Reading	0.09**	.05*

Note. * $p < 0.05$; ** $p < 0.01$

Second, the Consistency of Interests and PI scores were compared in the group of students stratified by the measurements of educational outcome: grade repetition, taking advanced courses, participating in school subject competition/Olympiad, and studying in general/vocational school after 9th grade. The results of the t-test showed that the PI subscale differentiated between the groups in all measures (Table 7), while the Consistency of Interests did not show statistically significant differences between students who repeated any grades ($M = 0.11$, $SD = 0.99$) and those who did not ($M = 0.18$, $SD = 0.88$), $t(2245) = 0.28$, $p \leq 0.05$, and between those who took advanced courses in school ($M = 0.2$, $SD = 0.88$) and those who did not ($M = 0.16$, $SD = 0.92$), $t(2060) = -0.85$, $p \leq 0.05$. (Table 7).

Table 6. Results of the t-test for the Consistency of Interests subscale.

Grouping Variable		Consistency of Interest					
		M	SD	N	95% CI for Mean Difference	t	Df
Grade repetition	Yes	0.11	0.99	27	-0.29, 0.38	0.28	2245
	No	0.18	0.88	2220			
Educational track after 9 th grade	General	0.24	0.89	1410	-0.26, -0.09	-4.19**	2037
	Vocational	0.06	0.87	629			
Take Advanced Courses	Yes	0.20	0.88	1439	-0.12, 0.05	-0.85	2060
	No	.16	0.92	623			
Participation in Olympiads	Yes	.24	0.92	948	-.17, -0.01	-2.42*	2072
	No	.15	0.87	1126			

Note. * $p < .05$, ** $p < .01$.

Table 7. Results of t-test for Perseverance of Efforts subscale.

Grouping Variable		Perseverance of Efforts					
		M	SD	N	95% CI for Mean Difference	t	df
Grade repetition	Yes	0.39	1.91	27	0.02, 1.21	2.04*	2245
	No	1.00	1.56	2220			
Educational track after 9 th grade	General	1.13	1.51	1410	-0.48, -0.19	-4.56**	2037
	Vocational	0.79	1.58	629			
Taking Advanced Courses	Yes	1.14	1.54	1439	-0.51, -0.24	-5.16**	2060
	No	0.76	1.53	623			
Participating in Olympiads	Yes	1.21	1.55	948	-0.47, -0.21	-5.01**	2072
	No	0.87	1.52	1126			

Note. * $p < 0.05$, ** $p < 0.01$.

A one-way analysis of variance was conducted, to compare the Consistency of Interests and Perseverance of Effort means in the groups of students with different durations of music (or art) and sports out-of-school practice. The analysis was significant for the Perseverance of Effort scores, with regard to each type of out-of-school practice (for music $F(2,2009) = 5.25, p < .01$; for sport $F(2,2021) = 12.433, p < .01$), while the ANOVA revealed no significant differences in the Consistency of Interests scores between groups with regard to music (or art) practice ($F(2,2009) = 1.017, p = 0.362$). The means and standard deviations are presented in Tables 8 and 9.

Table 8. Means (M), standard deviations (SD), and ANOVA results for Consistency of Interests and Perseverance of Effort by duration of music (or art) practice

Subscales	Duration of music (or art) practice			Between group effects	
	No activity (n=1754)	Less than one year (n=56)	More than one year (n=202)	F	p
	M (SD)	M (SD)	M (SD)		
CI	0.18 (0.89)	0.09 (0.99)	0.26 (0.85)	1.017	0.362
PE	0.98 (1.54)	1.32 (1.61)	1.31 (1.55)	5.25	0.005

Note. CI – Consistency of Interest, PE – Perseverance of Efforts

Table 9. Means (M), standard deviations (SD), and ANOVA results for Consistency of Interests and Perseverance of Effort by duration of sport practice.

Subscales	Duration of sport practice			Between group effects	
	No activity (n=1118)	Less than one year (n=318)	More than one year (n=588)		
	M (SD)	M (SD)	M (SD)	F	p
CI	0.14 (0.91)	0.13 (0.85)	0.33 (0.87)	10.367	.000
PE	0.90 (1.55)	0.94 (1.50)	1.29 (1.57)	12.433	.000

Note. CI – Consistency of Interest, PE – Perseverance of Efforts

A post-hoc Bonferroni test showed that students who did not practice music (or art) at all had significantly higher Perseverance of Effort scores than those who practiced music (or art) more than one year (mean difference = -0.33, $p = 0.012$, 95% CI for mean difference -0.61, -0.05). Students who played sport less than one year had less consistency of interests and less perseverance of effort than those who played sport for longer than a year (for Consistency of Interests: mean difference = -0.20, $p = 0.003$, 95% CI for mean difference -0.35, -0.06; for Perseverance of Effort: mean difference = -0.35, $p = 0.004$, 95% CI for mean difference -0.61, -0.09).

Finally, the correlations with the PISA Index of Perseverance were equal to 0.24 ($p < .01$) for Consistency of Interests and 0.34 ($p < .01$) for Perseverance of Effort, showing the medium effect size.

Discussion of Study 2

The aim of Study 2 was to confirm the two-dimensional model of the Russian version of the Grit scale produced by IRT analyses when Consistency of Interests consists of 7 and Perseverance of Effort of 5 items. The CFA supported the two-factor scale structure. A validity study was also undertaken, where we treated the Consistency of Interests and Perseverance of Effort subscales separately, relying on our results which had been obtained in the previous analyses. We found that the Perseverance of Effort subscale discriminated the groups well, and contrasted all the educational outcomes which were measured in a hypothesized direction. Students that stayed on the general education track after the 9th grade; who took advanced courses; participated in school subject competitions and Olympiads; did not repeat any grades,

and practiced music and sports over a year, had significantly higher Perseverance of Effort scores than students who did not. The Consistency of Interests subscale failed to discriminate between those who repeated a grade, took an advanced course and those who practiced music, from those who did not. The remaining criteria were distinguishable by the Consistency of Interests subscale.

One possible explanation for the Consistency of Interests's failure to discriminate between these outcomes is that this kind of behavior does not require any particular consistency of interests. Indeed, the formal schooling, including music or art schooling⁷, occurs regardless of whether one is interested in it consistently and steadily or not. On the other hand, the personal efforts and perseverance should still be applied in order to move along an educational trajectory without grade retention or dropping out.

The Consistency of Interests subscale's poor discrimination between the advanced course takers should not be surprising either. Consistency of interests implies the continuance of interests over time. Questioning whether or not a student has ever taken an advanced course does not refer to its continuance. A person could take several advanced courses and interrupt their studies immediately, which would demonstrate an inconsistency of pursuits. Therefore, attending advanced courses is likely to be a criterion of perseverance of effort rather than their stability over time.

A further question, which ought to be discussed in this section, is the low correlations between the TIMSS and PISA performance which we obtained for both the Consistency of Interests and Perseverance of Effort, despite prior work indicating moderate associations of the subscales and academic attainment operationalized by GPA and SAT (Duckworth, et al., 2007). One interpretation might suppose that PISA and TIMSS measures are more multi-determined, integrating a wider range of sources for variance, than SAT and GPA. To support this interpretation, we can first refer to a number of studies that argue PISA measures are substantially more complex outcomes of student experiences rather than only academic outcome (e.g., Fuchs, & Wößmann, 2008; Feuer, 2013). In addition to this, OECD found the weak association between students' perseverance and mathematics performance observed in PISA

⁷ In Russian Federation the music or art schooling has a three-level system of apprenticeship with standard curriculum within a given period approved with proper documentation (diploma of primary, secondary or tertiary musical (art) education).

when the perseverance are measured via PISA Index of Perseverance. They reported that about 5.7% of variation in student performance over OECD countries is explained by the perseverance, and even less so for Russia, i.e. 1.1% (OECD, 2014). Thus, these correlations might be affected by the measurement features of the PISA and TIMSS assessments.

General Discussion

Based on IRT modeling, the present studies provides data which examining dimensionality, reliability and validity of the Grit Scale in a Russian sample. Taken as a whole, the results demonstrated first that the scale provide two separate measures for two different constructs; “consistency of interests” and “perseverance of effort” rather than the common trait “grit”. Both subscales have good psychometric properties in terms of uni-dimensionality, reliability and functioning of answer categories. Second, both subscales displayed their validity, discriminating between groups of respondents with different educational outcomes. It is worth noting that subscales proved to be sensitive enough to predict only the relevant outcomes. The unique patterns of associations, with external measures formed by each subscale, supports the main finding that the subscales function effectively as separate measures of separate traits.

References

- Altinok, N. (2012). General versus vocational education: Some new evidence from PISA 2009. Final draft of the background paper to the education for all global monitoring report 2012.
- Arslan, S., Akin, A., & Çltemel, N. (2013). The predictive role of grit on Metacognition in Turkish university students. *Studia Psychologica*, 55 (4), 311-320.
- Bagozzi, R. P. (1993). An examination of the psychometric properties of measures of negative affect in the PANAS-X scales. *Journal of Personality and Social Psychology*, 65, 836–851. doi:10.1037/0022-3514.65.4.836.
- Baum, J. R., & Locke, E. A. (2004). The relationship of entrepreneurial traits, skill, and motivation to subsequent venture growth. *Journal of Applied Psychology*, 89, 587–598. doi:10.1037/0021-9010.89.4.587.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch model*. Mahwah, NJ, USA: Lawrence Erlbaum.
- Constantin, T., Holman, A., & Hojbotă, M. (2012). Development and validation of a motivational persistence scale. *Psihologija*, 45(2), 99–120. doi:10.2298/PSI1202099C.
- Desrochers, S., & Dahir, V. (2000). Ambition as a motivational basis of organizational and professional commitment: Preliminary analysis of a proposed career advancement ambition scale. *Perceptual and Motor Skills*, 91(2), 563–570. doi:10.2466/pms.2000.91.2.563.
- Dorans, N. J. (1989). Two New Approaches to Assessing Differential Item Functioning: Standardization and the Mantel--Haenszel Method. *Applied Measurement in Education*, 2(3), 217-233. doi:10.1207/s15324818ame0203_3.
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92(6), 1087–1101.

- Duckworth, A. L., & Quinn, P. D. (2009). Development and validation of the short grit scale (grit-S). *Journal of Personality Assessment, 91*(2), 166–174.
doi:10.1080/00223890802634290.
- Duckworth, A. L., Quinn, P. D., & Seligman, M. E. P. (2009). Positive predictors of teacher effectiveness. *Journal of Positive Psychology, 4*, 540–547.
doi:10.1080/17439760903157232.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates, Publishers.
- Eskreis-Winkler, L., Shulman, E. P., Beal, S. A., & Duckworth, A. L. (2014). The grit effect: Predicting retention in the military, the workplace, school and marriage. *Frontiers in Psychology, 5*, 36. doi:10.3389/fpsyg.2014.00036.
- Feuer, M. (2013). *Validity issues in international large-scale assessments: Truth and consequences*. [PDF document] Retrieved from
http://144.81.87.152/c/18486/pdf/19633_ets_tc_validityconference_Feuer%202012_03_28.pdf
- Fuchs, T., & Wößmann, L. (2008). *What accounts for international differences in student performance? A re-examination using PISA data*. Heidelberg: Physica-Verlag HD.
- Herche, J., & Engelland, B. (1996). Reversed-polarity items and scale Unidimensionality. *Journal of the Academy of Marketing Science, 24*, No. 4, 366-374.
doi:10.1177/0092070396244007.
- Klein, H. J., Cooper, J. T., Molloy, J. C., & Swanson, J. A. (2014). The assessment of commitment: Advantages of a unidimensional, target-free approach. *Journal of Applied Psychology, 99*, 222–238. doi:10.1037/a0034751.
- Kuczera, M. (2008). *VET in PISA: Results from PISA 2003 and 2006*. Paris: Organisation for Economic Co-operation and Development. Retrieved from
<http://www.oecd.org/education/innovation-education/41538731.pdf>.

- Kuzmina, Y., Popov, D., & Tyumeneva, Y. A. (2012). Pathways to life. The monitoring of school and university graduates' educational and work trajectories. Working papers by University of Leeds. Series "Education-Line". Retrieved from <http://www.leeds.ac.uk/educol/documents/215839.pdf>.
- Liem, G. A. D., & Martin, A. J. (2012). The motivation and engagement scale: Theoretical framework, psychometric properties, and applied yields. *Australian Psychologist*, *47*, 1, 3-13. doi:10.1111/j.1742-9544.2011.00049.x.
- Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, *2*, 266-283.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, *3*, 85-106.
- Ludlow, L. H. (1985). A strategy for the graphical representation of Rasch model residuals. *Educational and Psychological Measurement*, *45*(4), 851-859. doi:10.1177/0013164485454015.
- Lufi, D., & Cohen, A. (1987). A scale for measuring persistence in children. *Journal of Personality Assessment*, *51*, 178-185. doi:10.1207/s15327752jpa5102_2.
- Maddi, S. R., Harvey, R. H., Khoshaba, D. M., Lu, J. L., Persico, M., & Brow, M. (2006). The personality construct of hardiness, III: Relationships with repression, innovativeness, authoritarianism, and performance. *Journal of Personality*, *74*, 575-598. doi:10.1111/j.1467-6494.2006.00385.x.
- Maddi, S. R., Matthews, M. D., Kelly, D. R., Villarreal, B., & White, M. (2012). The role of hardiness and grit in predicting performance and retention of USMA cadets. *Military Psychology*, *24*(1), 19-28. doi:10.1080/08995605.2012.639672.
- Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifacts? *Journal of Personality and Social Psychology*, *70*, 810-819. doi:10.1037/0022-3514.70.4.810.

- Mukherjee, B. N. (1974). A questionnaire measure of persistence disposition. *Indian Journal of Psychology, 49*, 263–278.
- Mullis, I. V. S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 assessment frameworks*. Chestnut Hill, MA, TIMSS & PIRLS International Study Center, Boston College.
- Muthén, L. K., & Muthén, B. (1998-2012). Mplus User's Guide. *Seventh Edition*. Los Angeles, CA: Muthén & Muthén.
- Organisation for Economic Co-operation and Development (OECDa). (2014). *PISA 2012 results : Ready to learn: Students' engagement, drive and self-beliefs – Volume III, PISA*. Paris, OECD Publishing.
- Organisation for Economic Co-operation and Development (OECDb). (2014). *Pisa 2012 results: What students know and Can Do (volume I, revised edition, February 2014): Student performance in mathematics, reading and science, PISA*. Paris, OECD Publishing.
- Pozzebon, J., Damian, R. I., Hill, P. L., Lin, Y., Lapham, S., & Roberts, B. W. (2013). Establishing the validity and reliability of the project Talent Personality Inventory. *Frontiers in Psychology, 4*, 968. doi:10.3389/fpsyg.2013.00968.
- Reed, A. J., Schmitz, D., Baker, E., Nukui, A., & Epperly, T. (2012). Association of “grit” and satisfaction in rural and nonrural doctors. *Journal of the American Board of Family Medicine, 25*, 832–839. doi:10.3122/jabfm.2012.06.110044.
- Robertson-Kraft, C., & Duckworth, A. (2014). True grit: Trait-level perseverance and passion for long-term goals predicts effectiveness and retention among novice teachers. *Teachers College Record, 116* (3). Retrieved from <http://www.tcrecord.org/library>
- Salles, A., Cohen, G. L., & Mueller, C. M. (2014). The relationship between grit and resident well-being. *American Journal of Surgery, 207*, 251–254. doi:10.1016/j.amjsurg.2013.09.006.

- Smith, E. V. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, 3, 2, 205-231.
- Smith, R. M. (2004). Detecting item bias with the Rasch model. *Journal of Applied Measurement*, 5(4), 430-449.
- Stone, M. H. (2003). Substantive scale construction. In E. V. Smith, & R. M. Smith (Eds.), *Introduction to Rasch measurement* (pp.282–297). Maple Grove, MN: JAM Press.
- Vallerand, R. J., Blanchard, C., Mageau, G. A., Koestner, R., Ratelle, C., Leonard, M., Gagne, M., & Marsolais, J. (2003). Les passions de l'ame: On obsessive and harmonious passion. *Journal of Personality and Social Psychology*, 85, 756–767. doi:10.1037/0022-3514.85.4.756.
- Van de Vijver, F., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, 1, 89.
- von Davier, M., Gonzalez, E., & Mislevy, R. J. (2009). What are plausible values and why are they useful? In *IERI monograph series. Issues and methodologies in large-scale assessments, Vol. 2*, Hamburg/Princeton, NJ: IEA-ETS Research Institute, 9-36.
- Wang, C. K. A. (1932). A scale for measuring persistence. *Journal of Social Psychology*, 3, 79–90. doi:10.1080/00224545.1932.9919133.
- Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis*. Chicago, IL: MESA Press.

Appendix A

The twelve items of the original version of the Grit Scale (Duckworth, et al., 2007).

«Consistency of Interests» Subscale

1. I often set a goal but later choose to pursue a different one.
2. New ideas and new projects sometimes distract me from previous ones.
3. I become interested in new pursuits every few months.
4. My interests change from year to year.
5. I have been obsessed with a certain idea or project for a short time but later lost interest.
6. I have difficulty maintaining my focus on projects that take more than a few months to complete.

«Perseverance of Effort» Subscale

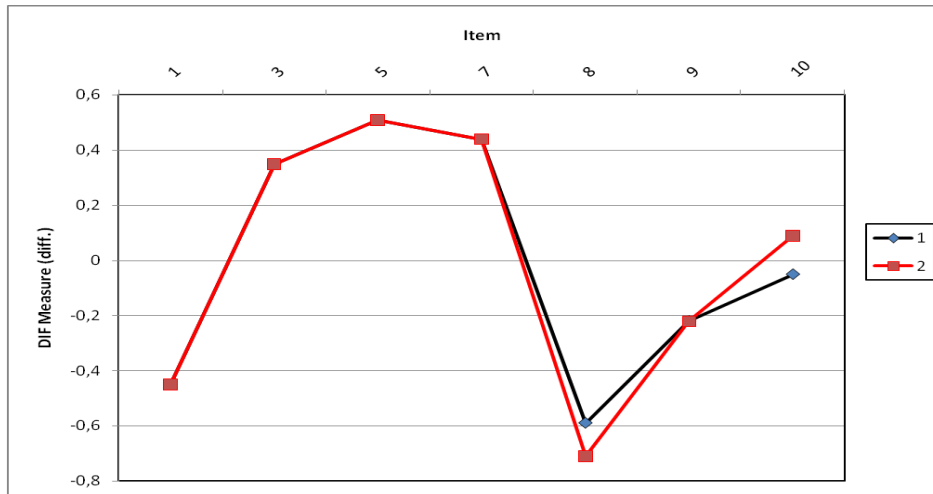
1. I have achieved a goal that took years of work.
2. I have overcome setbacks to conquer an important challenge.
3. I finish whatever I begin.
4. Setbacks don't discourage me.
5. I am a hard worker.
6. I am diligent.

Appendix B

Consistency of Interest subscale: Item Statistics

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	OUTFIT MNSQ	PT-MEAS. CORR.	Item
1	9635	3776	.33	.02	.88	.90	.57	I often set a goal but later choose*
3	11553	3770	-.22	.02	.96	.99	.54	New ideas and new projects sometimes distract me *
5	12122	3765	-.38	.02	1.11	1.15	.50	I become interested in new pursuits every few months*
7	11966	3773	-.33	.02	1.08	1.07	.61	My interests change from year to year*
8	9172	3764	.45	.02	1.09	1.12	.51	Setbacks do discourage me*
9	10189	3765	.16	.02	.81	.82	.63	I have been obsessed with a certain idea but later lost interest
10	10681	3732	-.01	.02	1.09	1.11	.52	I have difficulty maintaining my focus *

Comparison of item difficulties for male and female groups of respondents: none item demonstrates “practically significant” DIF (>0.5 logits):



1. I often set a goal but later choose
3. New ideas and new projects sometimes distract me
5. I become interested in new pursuits every few months
7. My interests change from year to year
8. Setbacks do discourage me
9. I have been obsessed with a certain idea but later lost interest
10. I have difficulty maintaining my focus

Appendix C

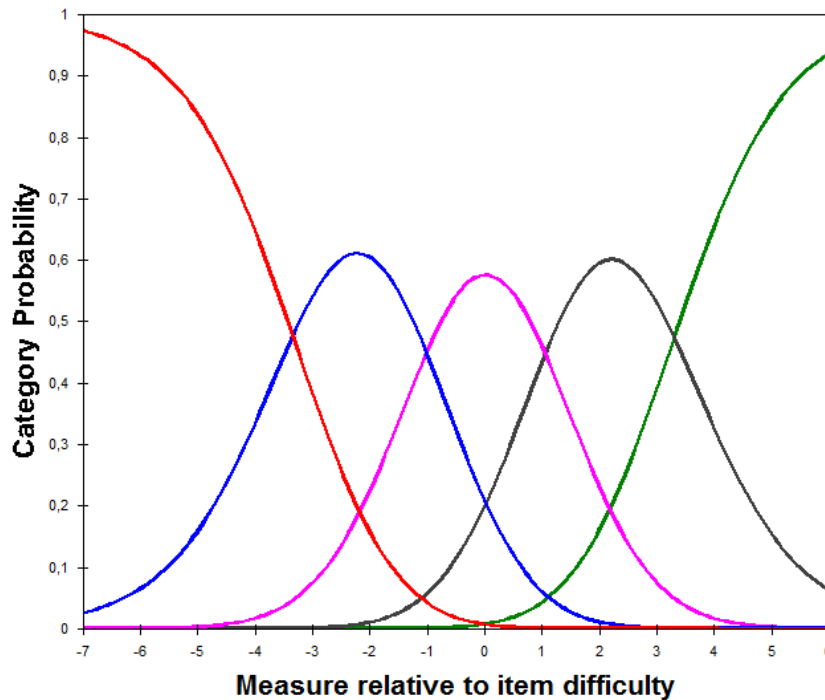
The Perseverance of Effort subscale: Item Statistics

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.	INFIT MNSQ	OUTFIT MNSQ	PT-MEAS. CORR.	Item
2	10902	3775	.76	.02	1.04	1.07	.61	Without irony I am a hard worker
4	13348	3773	-.18	.02	1.20	1.18	.63	I have achieved a goal that took years of work
6	12909	3761	-.02	.02	1.10	1.10	.63	I have overcome setbacks to conquer an
11	13093	3740	-.13	.02	.89	.89	.68	I finish whatever I begin
12	13803	3739	-.42	.02	.77	.75	.71	At work I am diligent

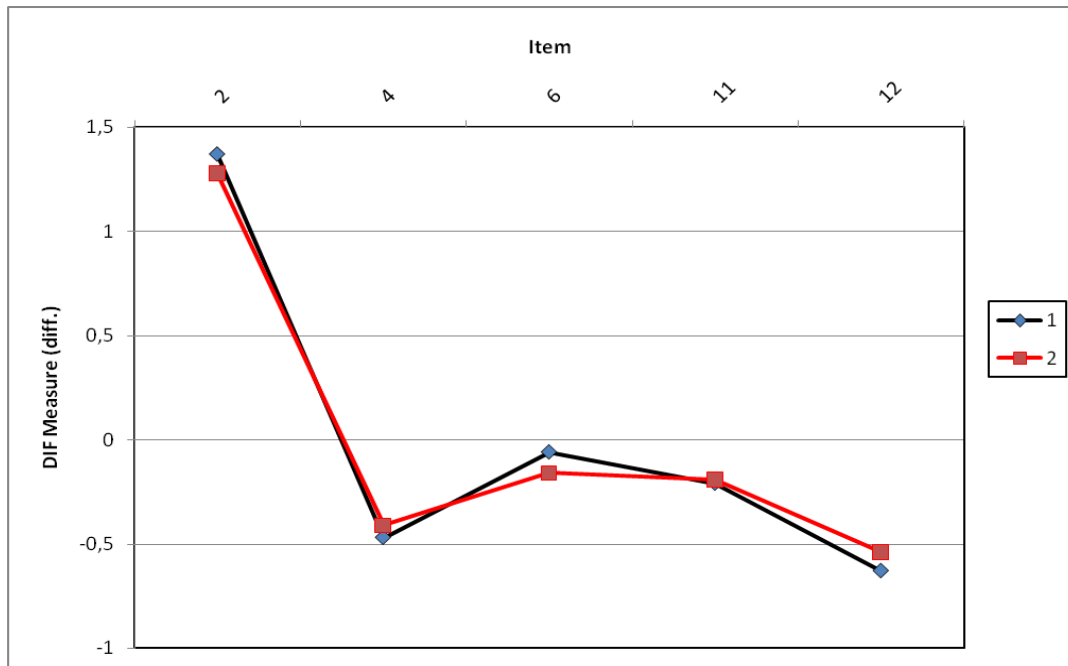
The Perseverance of Effort subscale: Summary of category structure of the rating scale.

Category number	Category label	Observed count	Observed %	Average measure	Infit mnsq	Outfit mnsq	Threshold estimate
1	SA	464	3	-2.48	1.01	1.01	NONE
2	A	2210	15	-0.88	1.07	1.07	-3.35
3	N	4859	32	0.39	0.98	0.98	-1.02
4	D	5067	34	1.75	1.00	0.99	1.06
5	SD	2411	16	3.20	0.93	0.94	3.30

The Perseverance of Effort subscale: Category probability curves for the rating scale



Comparison of item difficulties for male and female groups of respondents: none item demonstrates “practically significant” DIF (>0.5 logits):



2. Without irony I am a hard worker

4. I have achieved a goal that took years of work

6. I have overcome setbacks to conquer an important challenge

11. I finish whatever I begin

12. At work I am diligent

Appendix D

Descriptive statistics.

Table 1. Mean, standard errors and standard deviation for PISA and TIMSS results for the total sample and for the restricted sample (after an exclusion of respondents with poor goodness-of-fit characteristics).

Variables	Total sample		Restricted sample	
	Mean (s.e)	Standard deviation	Mean (s.e.)	Standard deviation
TIMSS Math	538.9 (3.6)	81.1	542.9 (4.1)	77.8
TIMSS Science	542.5 (3.3)	76.6	544.8 (3.3)	74.6
PISA Math	492.2 (0.3)	84.9	498.4 (0.4)	83
PISA Science	488.9 (0.8)	82.6	494.3 (0.6)	80.3
PISA Read	473.6 (0.6)	87.4	480.1 (0.5)	85.3
PISA Index of Perseverance of Efforts	0.33 (0.02)	1.05	0.31 (0.02)	0.96

Table 2. Distribution of variables for the total sample and for the restricted sample (after an exclusion of respondents with poor goodness-of-fit characteristics).

Variables		Total sample		Restricted sample	
		N	%	N	%
Grade repetition	Yes	77	2%	27	1%
	No	4318	98%	2220	99%
Educational track after 9 th grade	General	2697	65%	1410	68%
	Vocational	1354	33%	629	30%
Take advanced course	Yes	2809	68%	1439	70%
	No	1300	32%	623	30%
Participation in Olympiads	Yes	1797	57%	948	46%
	No	2341	43%	1126	54%
Duration of music (or art) activity	No activity	3487	87%	1754	87%
	Less than one year	105	3%	56	3%
	More than one year	406	10%	202	10%
Duration of sport practice	No activity	2215	55%	1118	55%
	Less than one year	606	15%	318	16%
	More than one year	1200	30%	588	29%

For correspondence: Yulia Tyumeneva, National Research University Higher School of Economics, International Laboratory for Educational Policy Analysis, senior researcher. E-mail: jutu@yandex.ru, mobile phone: 89104075140

Any opinions or claims contained in this Working Paper do not necessarily reflect the views of HSE.

© Tyumeneva, Kuzmina, Kardanova, 2014