Network Psychometric and Item Response Theory (IRT) Approach to Validating the Russian Adult Version of the Structure of Temperament Questionnaire (STQ-77Ru)

A. Gallyamova & D. Grigoryev[*]

HSE University, Moscow, Russia

Abstract. This study examines the psychometric properties of the Russian adult version of the Structure of Temperament Questionnaire (STQ-77Ru) within a diverse community sample. Employing both network psychometrics and Item Response Theory (IRT), we analyzed data from 3,442 Russian participants, aged between 18 and 81 years (M = 38, SD = 11). Network psychometrics were used to explore the complex interdependencies among temperament traits, while IRT provided detailed insights into the difficulty and discrimination of individual test items. Measurement equivalence tests supported high consistency in factor loadings and intercepts across different sexes and educational levels, enhancing the tool's applicability across demographics. This study further evaluated the ability of STQ-77Ru to distinguish between demographic groups, revealing substantial variations in temperament traits between males and females and across different levels of educational attainment. These findings validate the questionnaire's effectiveness in diverse settings and, with some reservations, underscore the potential of STQ-77Ru for delivering precise and reliable psychological assessments of temperament traits within the Russian community.

Keywords: Temperament Traits, Functional Ensemble of Temperament, Structure of Temperament Questionnaire, Network Analysis, Item Response Theory

1. Introduction

Neurochemical processes have a significant role in regulating human behavior, with neurotransmitter imbalances often leading to psychological disorders like depression and anxiety (Trofimova & Sulis, 2018). Similarly, subtle imbalances may influence traits in healthy individuals, shaping their temperament. The Functional Ensemble of Temperament (FET) is a prominent model devoted to this subject. Rooted in the functional constructivism approach, FET views all types of behavior as dynamic, generative processes created anew each time based on existing opportunities and situational demands (Trofimova, 2021b).

FET suggests that standard statistical methods often fall short in fully capturing the complex underlying processes of psychological traits (Trofimova & Araki, 2022). This complexity arises partly because multiple hormonal systems are involved in shaping any single trait. The model highlights how neurochemical systems that differentially regulate various brain structures contribute to different behavioral patterns, including temperament traits (Trofimova, 2016, 2018, 2019, 2021a, 2021b, 2022; Trofimova & Gaykalova, 2020; Trofimova & Robbins, 2016).

Despite the recognized limitations of the psychometric approach, survey assessments of biological-based features can still offer meaningful insights into individual differences (Figueredo et al., 2015). Consequently, Rusalov and Trofimova (2007) introduced the Structure of Temperament Questionnaire (STQ-77) as the most developed tool to assess these variations. The structure of STQ-77 was based on the functional neurochemistry summarized in the neurochemical framework of FET. They proposed that temperament is composed of 12 components that could be divided into four main groups: endurance (ergonicity), behavioral

[*] Corresponding author: e-mail: dgrigoryev@hse.ru
DOI: 10.38098/nsom_2025_05_01_03

orientation, speed of integration of actions (plasticity, tempo), and emotionality.

Each domain captures distinct aspects of temperament. Endurance reflects the ability to sustain activity over time; behavioral orientation involves the preference for specific types of stimuli and reinforcers. Tempo or plasticity describes an individual's speed in performing activities, while emotionality refers to emotional dispositions amplifying the previous three regulatory aspects of behavior. Notably, these three groups—endurance, orientation, and ease of integration—are specific to different types of activities, meaning they can vary within the same person based on the activity. For example, someone who shows high intellectual endurance may not necessarily excel in prolonged physical tasks.

Consequently, as it was shown in the temperament model and experimental studies by Rusalov (1989, 1997, 2018), these traits are activity-specific covering intellectual, social, and physical aspects. In contrast, emotionality is not activity-specific differing from other traits constituting a standalone factor. It includes three subfactors: impulsivity, neuroticism, and dispositional satisfaction, each contributing to how a person reacts emotionally under various circumstances (Rusalov & Trofimova, 2011; Trofimova & Sulis, 2011). Further details on these traits are available in Table 1.

Table 1. Descriptions of the FET Traits

Trait	Symbol	Description
Probabilistic Aspects		
Probabilistic Processing	PRO	Ability to handle uncertainty and make predictions.
Plasticity	PL	Flexibility in adapting to change.
Intellectual Endurance	ERI	Stamina for sustained cognitive effort.
Social-Verbal Aspects		
Empathy	EMP	Ability to understand and share others' feelings.
Social Tempo	TMS	Speed of engaging in social-verbal interactions.
Social-Verbal Endurance	ERS	Stamina to maintain social interactions over time.
Physical-Motor Aspects		
Sensation Seeking	SS	Propensity for new and intense experiences.
Physical Tempo	TMM	Speed of physical-motor actions.
Physical Endurance	ERM	Stamina to sustain physical activity.
Emotional Aspects		
Neuroticism	NEU	Propensity for emotional instability and negative feelings.
Impulsivity	IMP	Propensity for spontaneous and hasty actions.
(dispositional) Satisfaction	SLF	Ability to experience pleasure and contentment.

STQ-77 is a streamlined version of the original STQ, tested across multiple cultural and linguistic groups—including Chinese, English, Polish, Portuguese, and Urdu (Araki & Trofimova, 2021; Rusalov & Trofimova, 2007; Trofimova, 2010b; Trofimova & Araki, 2022, 2024). This compact version has demonstrated adequate reliability, internal consistency, and a consistent factor structure across these groups. There are several notable modifications in STQ-77, the latest, compact version of this test as compared to its predecessor, the 150-items STQ (Rusalov, 2018; Rusalov & Trofimova, 2007, 2011).

Initially, in STQ-150, emotionality was categorized as an activity-specific trait, similar to other traits in the original questionnaire. However, in STQ-77, this approach was revised as numerous psychometric studies showed that STQ Emotionality traits always converged into one factor of emotionality (Rusalov, 1997, 2004; Rusalov & Trofimova, 2007; Trofimova, 2010b, Rusalov, 2018). Additionally, STQ-150 did not have any scales related to behavioral orientation, and all the Plasticity scales in STQ-150 were converged into one scale of plasticity in STQ-77. Plasticity describes the ability to adapt or change one's actions swiftly and efficiently. All psychometric studies of various versions of STQ-77 showed the consistent presence of four factors in its structure: Physical Aspects, Socio-verbal Aspects, Mental Aspects, and Emotionality Aspects of behavioral regulation (Araki & Trofimova, 2021; Rusalov & Trofimova, 2007; Trofimova, 2010b). This is in line with the activity-specific approach in structuring temperament components offered by Rusalov (1989, 2018).

STQ-77 has been evaluated among English and Russian-speaking populations, demonstrating adequate reliability coefficients ranging from .65 to .85 (Rusalov & Trofimova, 2007; Trofimova, 2010a, 2010b; Trofimova & Sulis, 2011). However, these evaluations were primarily conducted using student samples, a significant limitation given the demographic homogeneity. Consequently, the current study aims to validate STQ-77Ru within a community sample, employing a comprehensive, long questionnaire frequently used in psychological research.

Additionally, this study employs network psychometric and Item Response Theory (IRT) approaches. The use of network analysis allows us to examine the interconnections between different temperament traits and items, providing a more nuanced understanding of their structure within STQ-77Ru. The IRT approach is advantageous because it offers detailed insights into the properties of individual scale items by STQ-77Ru, such as their difficulty and discriminatory ability. Finally, this study examines the effectiveness of STQ-77Ru in distinguishing between different demographic groups, specifically between males and females, and across levels of educational attainment. Together, these elements are crucial for enhancing the precision and reliability of STQ-77Ru, facilitating more tailored and effective psychological assessments.

2. Method

Samples

The dataset encompassed 3,442 responses from Russian participants aged between 18 and 81 years, with a mean age of 38 years (*SD* = 11). The sex distribution consisted of 48% males and 52% females. Regarding educational levels, participants were categorized from *'Incomplete secondary'* to *'Academic degree.'* Interestingly, the data showed a nearly even split in educational achievement among the participants: 45% lacked higher education degrees, while 55% held them. This presents a unique opportunity to compare these groups, given this rare balance in educational distribution.

Procedure

Our study followed the ethical guidelines set by COPE and the APA, aligning with Russian university and national regulations. According to these regulations, no ethics clearance was required for this type of survey research, as it did not involve medical data. Data collection took place over two phases—in January 2023 and May 2023—and was conducted by an independent commercial research company through an online survey. Participants were drawn from the company's proprietary respondent pool and received financial compensation for their time. Prior to beginning the survey, all participants were given detailed instructions explaining the study's purpose, confidentiality measures, implied consent, and contact information, ensuring transparency and informed participation.

Measures

The Structure of Temperament Questionnaire (STQ-77), developed by Rusalov and Trofimova (2007), consists of 77 statements spread across 12 temperament scales, with each scale having 6 items except for a validity scale that contains 5 items. Participants rate these statements on a 4-points Likert scale ranging from 1 = *strongly disagree* to 4 = *strongly agree*. The questionnaire categorizes the scales into four groups: (1) probabilistic aspects (Probabilistic Processing [PRO], Plasticity [PL], Intellectual Endurance [ERI]), (2) social-verbal aspects (Empathy [EMP], Social-Verbal Tempo [TMS], Social Endurance [ERS]), (3) physical-motor aspects (Sensation Seeking [SS], Motor Tempo [TMM], Physical-Motor Endurance [ERM]), and (4) emotional aspects (Neuroticism [NEU], Impulsivity [IMP], [dispositional] Satisfaction [SLF]). Additionally, the validity scale, as proposed by the authors, suggests that scores between 15 and 20 could indicate a positive impression bias, deeming responses invalid.

Data Processing

The analysis began with meticulous data screening, removing records with uniform responses, such as identical answer sequences, to ensure data quality. Network analysis using the *qgraph* package (Epskamp et al., 2012) then investigated the complex interconnections between trait variables and items from STQ-77Ru through a visual and statistical approach. In this analysis, each trait variable and item was represented as a node, with edges depicting the relationships between them based on partial correlations. These relationships were further refined using the graphical LASSO technique, which applies regularization to estimate a sparse inverse covariance matrix, helping to clarify which connections are most critical by eliminating weaker ones. It applies a penalty to the partial correlation coefficients, which effectively shrinks smaller correlations towards zero, leaving only the strongest and most significant connections. This results in a cleaner, more interpretable network diagram where only meaningful links are shown. This analysis also focused on centrality measures (expected influence, strength, betweenness, closeness) to identify key variables that act as bridges or maintain central roles within the network structure.

Expected Influence calculates the sum of edge weights connected to a node, considering both positive and negative connections. Nodes with a high absolute value of expected influence are pivotal; changes in these nodes could significantly impact the network, particularly in terms of trait flow and integration. Strength reflects the sum of the absolute values of all links connected to a node, indicating its overall connectivity and potential influence. Betweenness identifies nodes that act as bridges along the shortest paths between other nodes, which is crucial for understanding how certain traits or items mediate the relationships between other elements within the network. Closeness measures the average length of the shortest path from a node to all other nodes, highlighting those that can quickly affect or be affected by others in the network. A *z*-score of these centrality measures indicates how many *SD*s a node's centrality measure is from the mean centrality score across the network. This standardization helps to identify which nodes are significantly more central or influential than the average, thus providing insight into the structure and dynamics of the network.

Next, IRT was applied using the mirt package (Chalmers, 2012) to evaluate the precision of the STQ-77Ru scales in measuring temperament traits. The Item Characteristic Curves (ICCs) were central to this analysis, illustrating how the probability of item endorsement varies with an individual's latent trait level, providing key insights into item difficulty and discrimination. Item difficulty is indicated by the latent trait level at which an item receives a 50% endorsement probability, and item discrimination is reflected by the steepness of the curve around this midpoint. Steeper curves indicate better discrimination, effectively distinguishing between individuals with similar, but distinct latent trait levels. This methodological approach enhances the scale's reliability and validity by pinpointing which items accurately measure and are sensitive to variations in temperament traits. We then estimated the internal consistency of the scales using Cronbach's α to ensure that each STQ-77Ru scale reliably measured its intended construct across different demographic groups.

We then estimated measurement equivalence of the STQ-77Ru scales across sex and educational attainment using the invariance alignment technique within the *sirt* package (Robitzsch, 2024). This analysis evaluated the consistency of factor loadings and the uniformity of intercepts, supporting that differences in scores truly reflected variations in the underlying constructs rather than differences in item interpretation.

Finally, the differences between groups categorized by sex and educational attainment were estimated using Mahalanobis *D* (Del Giudice, 2022). This statistic measures the multivariate distance between the centroids of each group, functioning as a multivariate effect size indicator. Mahalanobis *D* serves a similar purpose to Cohen's *d*, but is suited for analyses involving multiple temperament traits simultaneously. This approach enables a comprehensive evaluation of the variability and distinctions between the groups based on combined characteristics.

3. Results

Data Preparing

During the data screening process, 11 records were removed due to uniform responses, such as sequences of identical answers (e.g., 1,1,1,1...). This step was vital to eliminate potential biases and ensure the validity of the subsequent data analysis.

Network Analysis

The resulting networks, as illustrated in Figure 1-3 (see also Supplementary Material), is a complex tapestry, revealing the centrality and interconnectedness of each variable within and between the communities, as well as how these might vary across different sociodemographic groups.

Expected Influence and Central Impact. Expected Influence, akin to eigenvector centrality, measures a trait's potential to influence the network, factoring in the centrality of its connections. Trait like TMM stand out, indicating that it may exert considerable sway through the quality and strength of the ties.

Strength Centrality and Connectivity. Strength centrality quantifies the sum of connection weights for a trait, denoting its connectivity within its community. TMM and IMP show strong centrality, highlighting their substantial ties within their respective communities and underscoring their significant roles in the network's cohesiveness and stability

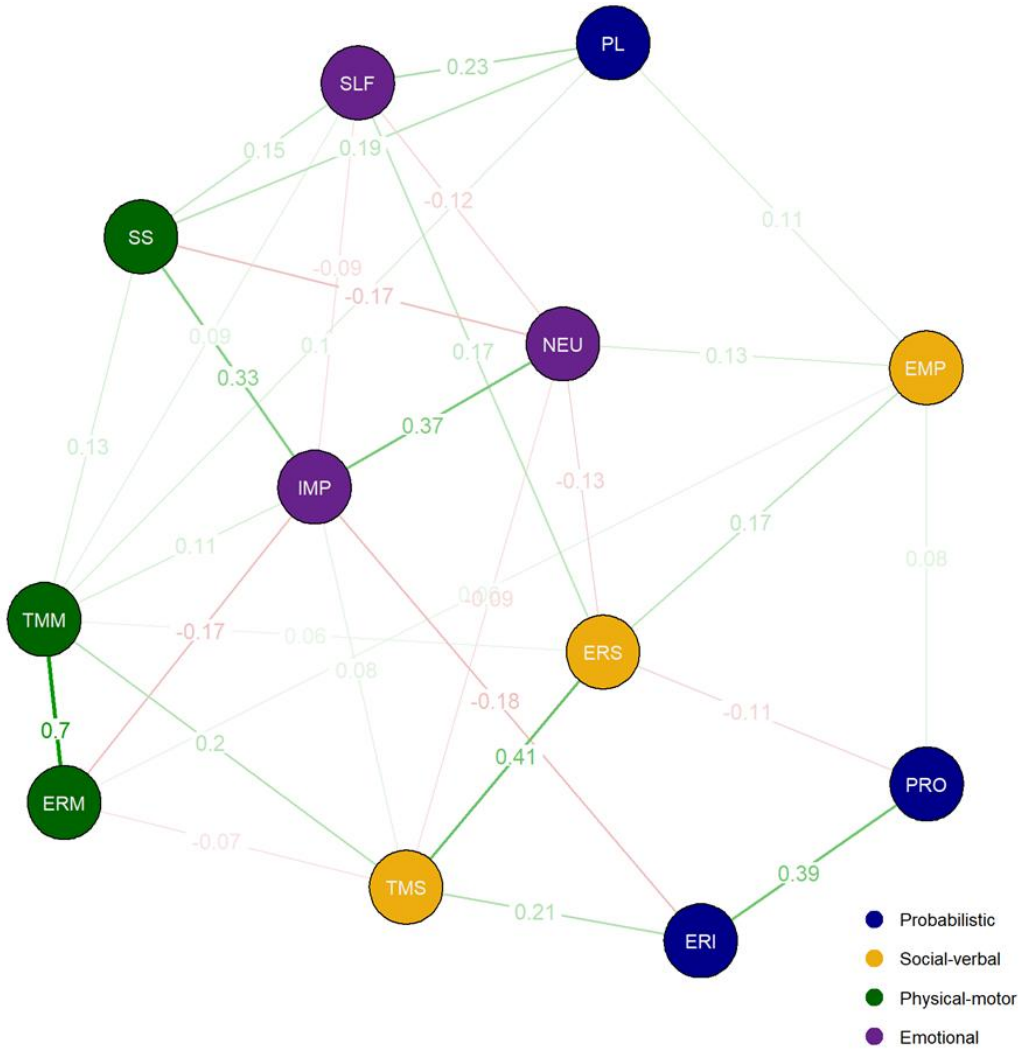

Figure 1. Network Plot for Variables by STQ-77Ru for the Total Sample (N = 3431)

Betweenness Centrality and Community Bridges. Betweenness centrality identifies traits that serve as bridges within the broader network, connecting different communities. Traits with high betweenness, such as IMP, may have a regulatory or gatekeeping role, influencing the flow of interactions between communities rather than solely within them.

Closeness Centrality and Network Integration. Closeness centrality measures the degree to which a trait is near all other traits in the network, indicative of how integrated it is within the overall structure. IMP's close centrality suggests it is well-positioned to rapidly exchange information or influence with other traits, reflecting its central role both within its immediate community and in the network at large.

Our analysis uncovered clear patterns of cohesion within FET framework communities: all traits, except for PL, demonstrated strong intragroup connections. In the probabilistic aspect, PRO and ERI shared close ties linked to cognitive functions and endurance. The social-verbal cluster, containing EMP, TMS, and ERS, highlighted integrated empathetic and communicative processes. SS, TMM, and ERM in the physical-motor aspect were tightly interwoven, suggesting sensory engagement and motor endurance. Lastly, NEU, IMP, and SLF in the emotional aspect were closely associated, indicative of a shared spectrum in emotional dynamics.

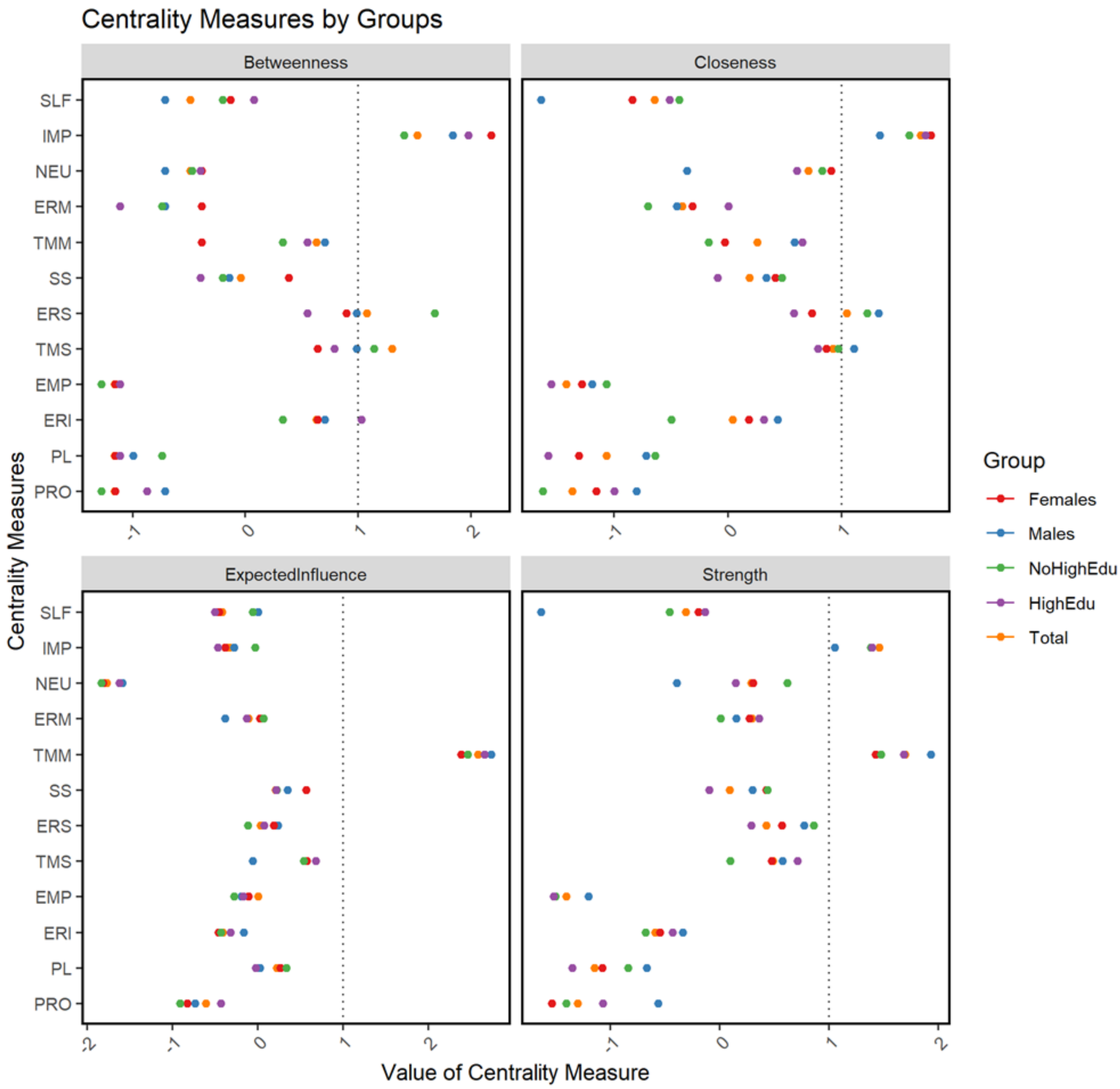

Figure 2. Centrality Measures (Expected Influence, Strength, Betweenness, Closeness) of Network Plots for STQ-77Ru by the Groups

Item Level Analysis. Our network analysis at the item level within STQ-77Ru suggested variances that may indicate both the strengths and potential weaknesses in the scale's item quality. While most items demonstrated strong intragroup connections, indicative of coherent and reliable scales, items related to IMP and SLF showed a sparser connectivity. This could reflect either a nuanced expression of these traits or, alternatively, issues with item formulation that warrant further psychometric evaluation. The notable scatter of PL items across the network raises additional concerns about their convergent validity, suggesting these items may not capture a singular construct as effectively as those from other domains. Despite these concerns, the dense interconnections among the remaining items signal robust scale quality, capturing well-defined constructs consistent with the theoretical underpinnings of STQ-77Ru.

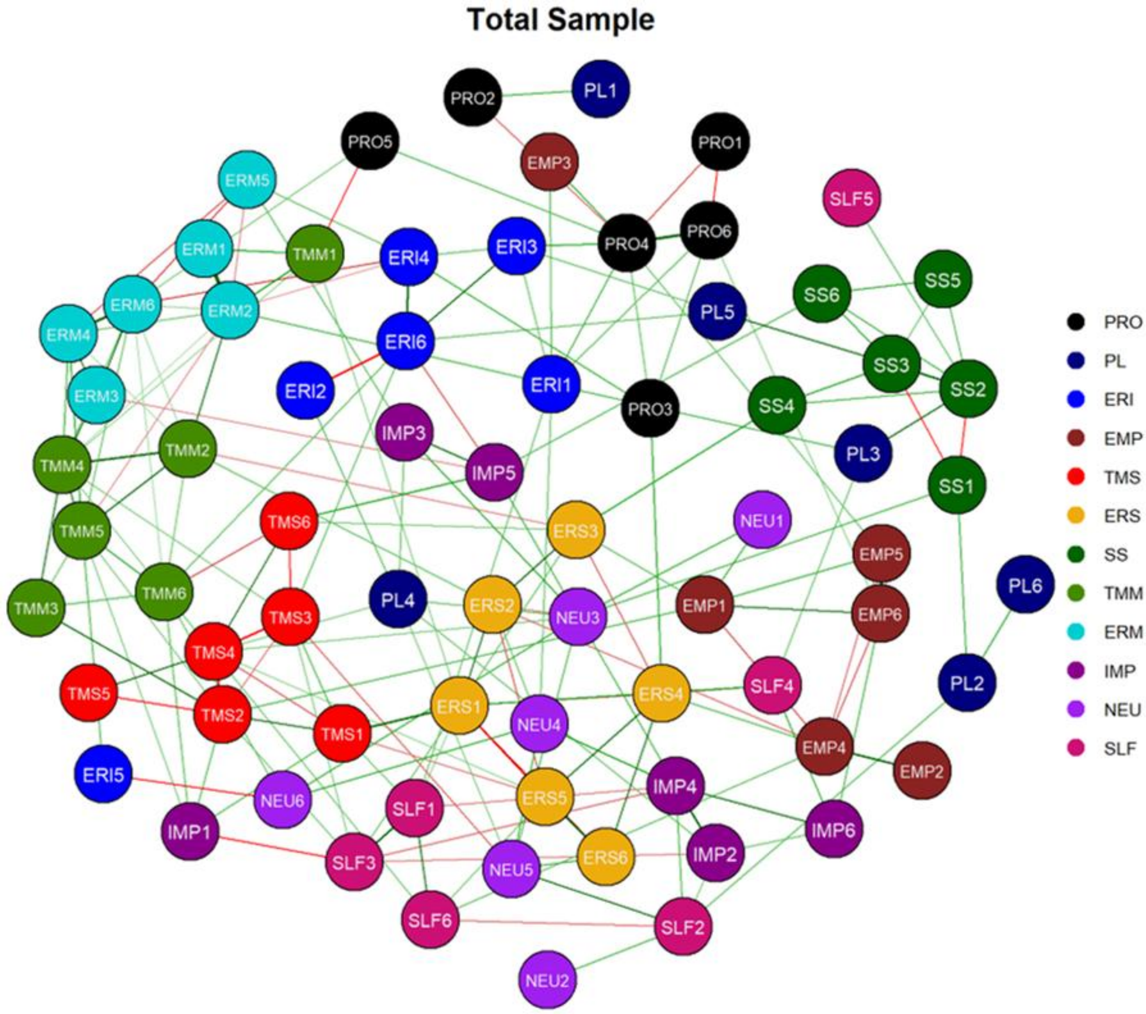

Figure 3. Network Plot of Items by STQ-77Ru for the Total Sample (*N* = 3431)

IRT Analysis

The IRT analysis highlighted substantial variation in item efficiencies within the STQ-77Ru scales (see Supplementary Material and Figure 4). Notable were items with steeply inclined ICCs, such as ERI6 and TMM4, which demonstrated superior discrimination. These items were adept at discerning even marginal differences in the respective temperament traits. In stark contrast were items like PL2, characterized by their flatter ICCs, signifying a diminished discriminative capability and thus marking them as candidates for re-evaluation or refinement.

Diving into the probabilistic dimensions, the scales pertaining to PRO, PL, and ERI disclosed a tapestry of performance across discrimination and difficulty indices. The PL scale was a focal point of critique, with items like PL2 revealing subpar discriminative qualities. Conversely, PL5's steep curve was a beacon of precise measurement at certain trait levels, indicating its robustness in capturing the trait it was intended to measure.

The scales probing social-verbal aspects (EMP, TMS, ERS) predominantly exhibited a commendable level of discrimination, illustrating the scales' adeptness in mapping the landscape of social-verbal attributes. However, there was a need for refinement in some areas, such as with the EMP scale where EMP3 displayed reduced discriminative efficiency at median trait levels. On the other hand, items like EMP5 and EMP6 were discriminatively efficient across a wider trait spectrum, enhancing their utility in the scale.

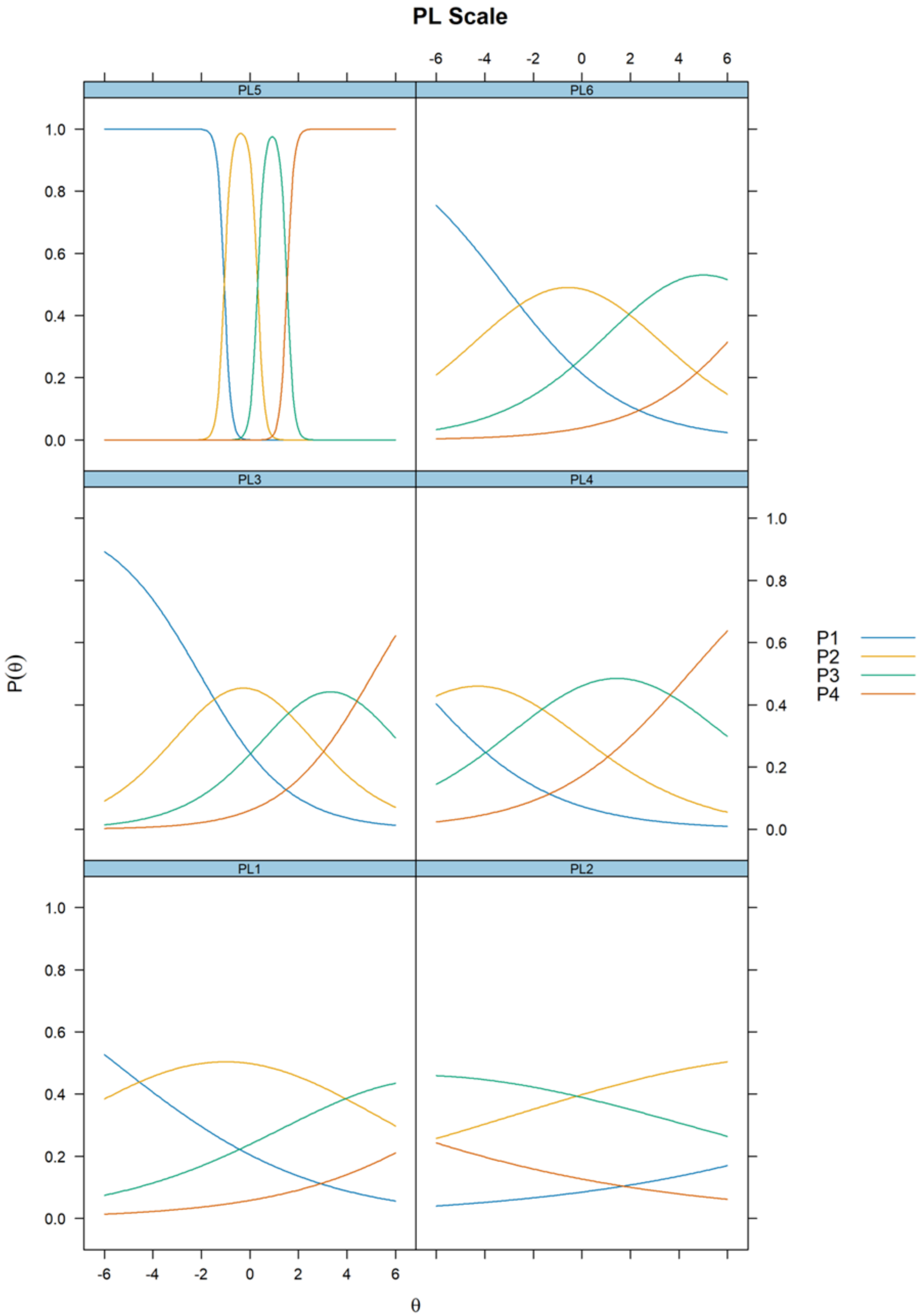

Figure 4. Item Characteristic Curves by the PL scale from STQ-77Ru (N = 3431)

In the realm of physical-motor traits, there was notable variability in how SS, TMM, and ERM items responded across different trait levels. The SS scale, with items such as SS2, showed precise discrimination, especially at moderate trait levels. In contrast, the TMM scale revealed potential improvement areas, highlighted by TMM1's limited discrimination ability, which may not accurately reflect the nuances of motor tempo.

Emotionally-oriented scales such as NEU, IMP, and SLF showcased a broad array of trait level representations. NEU3, for instance, had a steep curve indicating high discrimination, though NEU2 had a somewhat less acute sensitivity to trait variations, suggesting a need for adjustment to optimize difficulty and discrimination. Attention was drawn to certain items that did not capture the full intended range of their respective traits, calling into question their difficulty calibration. Additionally, observed redundancies, such as between TMM3 and TMM6, suggest a potential overlap in item content, challenging their unique contributions to their scales.

Analysis of Internal Consistency of Scales

Finally, the internal consistency of the STQ-77Ru scales, estimated using Cronbach's α across various demographic groups including sex and educational attainment, showed a range from low to excellent reliability (see Table 2 and 3). Notably, scales like ERM demonstrated high reliability across all groups (α = .86), whereas PL consistently reported low reliability (α = .24). Slight variations were observed between sexes and educational attainment, with higher education typically associated with slightly higher reliability in scales such as ERI and PL.

Measurement Equivalence

The invariance alignment technique was used to evaluate the equivalence of the measurement model across groups defined by sex and educational attainment, focusing on factor loadings (metric invariance) and intercepts (scalar invariance). Results showed exceptional measurement equivalence; for sex, R^2 values of .997 for loadings and .999 for intercepts demonstrated consistent representation of constructs between male and female groups. Similarly, for educational attainment, R^2 values of .999 and 1.000 for loadings and intercepts, respectively, indicated nearly identical factor structures across those with and without higher education. These findings support the model's robustness and suitability for comparative studies and pooled analyses, ensuring that differences in construct measurements are not due to sex-based or educational biases, thereby enhancing the reliability and validity of any substantive conclusions drawn from the model.

Analysis of Group Means Comparisons

Our analysis of group means comparisons are available in Table 2 and 3. Significant differences were found in the FET traits between males and females, except for ERI. Males scored higher on traits such as PRO, PL, ERS, and all physical-motor aspects (SS, TMM, ERM). They also showed higher levels of SLF. On the other hand, females outscored males in EMP, TMS, NEU, and IMP, indicating a female-based divergence in these social-verbal and emotional temperament traits. When comparing individuals based on their level of educational attainment, notable differences emerged in several temperament traits. Those without higher education scored higher in SS, IMP, and SLF, suggesting that these traits were more pronounced in this group. In contrast, individuals with higher education demonstrated higher scores in PRO, ERI, and TMS, indicating these traits were more pronounced among those with advanced educational backgrounds.

Table 2. Group Means Comparisons Between Males and Females

	Total (N = 3431)		Males (n = 1647)		Females (n = 1784)		t	d
	M (SD)	α	M (SD)	α	M (SD)	α		
Probabilistic aspects								
PRO	16.45 (3.33)	.61	17.14 (3.12)	.59	15.80 (3.39)	.60	12.08***	0.41
PL	13.99 (2.26)	.24	14.11 (2.14)	.21	13.87 (2.37)	.27	3.21***	0.11
ERI	17.37 (2.83)	.56	17.33 (2.76)	.59	17.41 (2.89)	.55	-0.89	-0.03
Social-verbal aspects								
EMP	14.45 (2.92)	.58	14.33 (2.79)	.58	14.55 (3.02)	.59	-2.22*	-0.08
TMS	16.67 (3.47)	.73	16.48 (3.41)	.74	16.84 (3.53)	.72	-2.99***	-0.10
ERS	14.91 (3.85)	.78	15.05 (3.68)	.78	14.78 (4.00)	.78	2.03*	0.07
Physical-motor aspects								
SS	13.12 (3.57)	.72	13.48 (3.46)	.72	12.79 (3.65)	.73	5.63***	0.19
TMM	15.05 (3.92)	.82	15.28 (3.83)	.83	14.83 (3.99)	.82	3.40***	0.12
ERM	15.51 (4.27)	.86	15.82 (4.24)	.87	15.23 (4.28)	.85	4.01***	0.14
Emotional aspects								
NEU	15.51 (2.97)	.57	15.06 (2.83)	.55	15.93 (3.03)	.57	-8.70***	-0.30
IMP	14.47 (3.25)	.68	14.09 (3.11)	.68	14.83 (3.34)	.67	-6.71***	-0.23
SLF	13.81 (2.87)	.50	14.30 (2.66)	.45	13.36 (2.99)	.52	9.76***	0.33

Note. ***p < .001, **p < .01, *p < .05; the Mahalanobis D summarizing the overall distance in temperament traits between males and females was 0.72.

Table 3. Group Means Comparisons by Educational Attainment

	Total (N = 3431)		No Higher Education (n = 1542)		Higher Education (n = 1889)		t	d
	M (SD)	α	M (SD)	α	M (SD)	α		
Probabilistic aspects								
PRO	16.45 (3.33)	.61	16.22 (3.33)	.60	16.63 (3.32)	.62	-3.59***	-0.12
PL	13.99 (2.26)	.24	14.04 (2.31)	.23	13.94 (2.23)	.26	1.35	0.05
ERI	17.37 (2.83)	.56	16.78 (2.82)	.52	17.85 (2.75)	.57	-11.19***	-0.39
Social-verbal aspects								
EMP	14.45 (2.92)	.58	14.55 (2.95)	.58	14.36 (2.89)	.59	1.84	0.06
TMS	16.67 (3.47)	.73	16.44 (3.52)	.73	16.85 (3.43)	.73	-3.40***	-0.12
ERS	14.91 (3.85)	.78	14.96 (3.85)	.77	14.87 (3.86)	.79	0.62	0.02
Physical-motor aspects								
SS	13.12 (3.57)	.72	13.47 (3.60)	.71	12.83 (3.52)	.73	5.21***	0.18
TMM	15.05 (3.92)	.82	15.13 (4.00)	.83	14.99 (3.86)	.82	1.04	0.04
ERM	15.51 (4.27)	.86	15.47 (4.32)	.86	15.55 (4.23)	.86	-0.52	-0.02
Emotional aspects								
NEU	15.51 (2.97)	.57	15.60 (3.07)	.58	15.44 (2.88)	.55	1.52	0.05
IMP	14.47 (3.25)	.68	14.73 (3.28)	.67	14.26 (3.21)	.68	4.20***	0.14
SLF	13.81 (2.87)	.50	13.96 (2.96)	.51	13.68 (2.80)	.49	2.84***	0.10

Note. ***p < .001, **p < .01, *p < .05; the Mahalanobis D summarizing the overall distance in temperament traits between the no higher education group and the higher education group was 0.46.

We used the Mahalanobis D to investigate sex and educational attainment differences across the FET traits. The analysis revealed noticeable differences between males and females (D = 0.72) and smaller differences between educational groups (D = 0.46), indicating distinct patterns of variability across these demographics. For sex differences, the probability that a randomly selected male scores higher than a female on the combined variables was approximately 69.4%, with a moderate probability of correct classification (PCC = 0.64). This suggests meaningful yet noticeable distinctions between males and females. The overlap in group distributions (OVL = 0.72) and heterogeneity measures (H^2 = 0.67) further indicated that certain variables significantly drive these differences. In contrast, differences between educational attainment were less pronounced, with higher overlap coefficients (OVL = 0.82) and a lower PCC (0.59), suggesting subtler distinctions between the groups. The significant heterogeneity coefficient (H^2 = 0.84) implies that a few variables predominantly contribute to these differences.

4. Discussion

The present study provided a psychometric evaluation of the STQ-77Ru, combining both network psychometric and IRT approaches to validating the structure and measurement precision of the questionnaire. Our results offer several insights into the intricate network of temperament traits as conceptualized within the FET framework (e.g., Trofimova & Robbins, 2016) and their manifestations within the large community Russian sample.

Network Analysis: Centrality and Connectivity

In general, intragroup closeness of the temperament traits measured by STQ-77Ru suggests that traits within the same aspect of the FET framework are more likely to interact with each other, sharing functional domains and potentially neurobiological pathways, which may have implications for the understanding of behavior and temperament as organized, systemic constructs. Overall, network analysis is valuable for assessing observed traits rooted in evolutionary and biological adaptations, as previously demonstrated in models such as the evolved human motives model (Aunger et al., 2025). This analysis also revealed significant findings regarding the centrality of temperament traits measured by STQ-77Ru. Traits like TMM displayed considerable expected influence, signifying a central role in the trait network. IMP's high betweenness centrality emphasized its function as a bridge, potentially moderating interactions between traits, while its closeness centrality underscores its potential rapid impact within the network. TMM refers to learned integration, and IMP refers to premature integration; both mobility-related traits in FET are associated with dopaminergic systems (Trofimova & Araki, 2022). The prominent roles of TMM and IMP within the temperament network suggest that variations in these traits could significantly alter the overall network dynamics, potentially leading to shifts in behavioral patterns across a broad spectrum.

IRT Analysis: Item Efficiency and Scale Precision

The IRT analysis was instrumental in evaluating item-level performance, with certain items demonstrating superior discrimination. This underlines the capacity of STQ-77Ru to differentiate between individuals with varying levels of the latent traits being measured. Nonetheless, the variability observed in item discrimination and difficulty across the scales points to potential areas for refinement. In particular, items related to PL warrant a closer examination due to their suboptimal discriminative quality and potential issues with convergent validity. The suboptimal performance of the PL items within the STQ-77Ru might be attributed to the complexity of their wording. In our context of a lengthy questionnaire with the community sample, which encompasses not only 77 items on various temperament traits but also additional measures, respondents might struggle with understanding the PL items. This challenge can lead to issues in item discrimination and difficulty, affecting the overall precision and validity of the scale. Enhanced clarity and simplification of item wording could potentially improve respondent comprehension and, subsequently, the discriminative capacity of these items.

Internal Consistency: Reliability across Demographics

The internal consistency of the scales ranged from low to excellent, with education likely emerging as a factor associated with reliability. The consistently low alpha for the PL scale across all demographics also raises concerns regarding the internal coherence of the scale's items, which may not be accurately capturing the intended construct. However, the PL items, which currently demonstrate weak internal consistency, may benefit from being selected to capture more distinct and nuanced aspects of Plasticity without relying heavily on shared variance. Enhancing the scale in this way could improve its predictive accuracy and relevance. This approach aligns with findings from Altgassen et al. (2024), which emphasize the benefits of focusing on unique variances, promising more precise and personalized assessment tools that move beyond traditional high communality constraints.

Measurement Equivalence

Our initial steps ensured the measurement equivalence of data across demographic variables. The high R^2 values obtained through invariance alignment techniques established robust measurement equivalence across sex and educational attainment. This indicates a strong consistency in the factor structure of STQ-77Ru, suggesting that the scales function similarly across different groups, thus validating the use of STQ-77Ru for cross-group comparisons.

Group Differences Predictions: Sex and Educational Attainment

The analysis of group differences in temperament, as assessed by STQ-77Ru, found significant variations that align with the predictions of the FET framework (see Trofimova, 2015). Both males and females demonstrated distinct temperament profiles, in accordance with FET: traits expected to be more pronounced in males (e.g., physical-motor aspects) were supported to be higher among male participants. Conversely, females exhibited higher levels of traits typically associated with their sex (e.g., EMP, TMS, NEU). Educational disparities also reflected the FET's theoretical projections, with the presence or absence of higher educational achievements coinciding with differences in temperament traits such as ERI, SS, and IMP. These congruencies between these empirical findings and the FET's hypotheses lends further support to accurately capture and predict sex- and education-related variations in temperament as measured by STQ-77Ru.

5. Conclusions

This study provides valuable insights into the psychometric properties of STQ-77Ru with a large Russian community sample, reinforcing its utility while also indicating areas for further refinement. In general, our findings have several implications for the FET framework and the application of STQ-77Ru in research and practice. The robustness of the STQ-77Ru's structure supports its theoretical underpinnings and offers a valid instrument for assessing the FET temperament traits in the Russian context. However, the identified weaknesses necessitate a careful examination of certain items and scales, particularly in terms of item formulation and scale coherence.

Conflict of Interest: The authors declare they have no conflicts of interest.

Funding: The publication was prepared within the framework of the Academic Fund Program at HSE University (grant № 24-00-013 "Adaptive Foundations of Culture: Toward Understanding Cultural Orientations through the Lens of Life History Trade-Offs").

Ethics Statement: This study was conducted in compliance with the ethical standards of COPE and APA. The procedure was in line with Russian regulations; as per university and national Russian regulations, no ethics clearance was required for this type of survey research (if it did not include medical data).

CRediT author statement: Albina Gallyamova: Project administration, Conceptualization, Validation, Investigation, Writing – original draft, Writing – review & editing. Dmitry Grigoryev: Supervision, Conceptualization, Methodology, Investigation, Data Curation, Software, Formal Analysis, Visualization, Writing – review & editing.

The authors have read and approved the final version and are responsible for all aspects of the manuscript.

Acknowledgments: The authors sincerely thanks Irina Trofimova (McMaster University) for her valuable comments and insights on this paper.

Highlights:

- Validated STQ-77Ru with network analysis and IRT on 3,442 adults (ages 18-81).
- Network analysis showed key traits like Impulsivity and Physical Tempo as central.
- Confirmed measurement invariance across sex and educational levels.
- STQ-77Ru effectively differentiates temperament traits across demographics.
- Findings support STQ-77Ru's validity for assessing temperament in Russian adults.

References

1. Altgassen, E., Olaru, G., & Wilhelm, O. (2024). What if there were no personality factors? Comparing the predictability of behavioral act frequencies from a big-five and a maximal-dimensional item set. *European Journal of Personality, 38*(2), 291–305. https://doi.org/10.1177/08902070231163283
2. Arakia, M. E., & Trofimova, I. N. (2021). Validation of the Portuguese version of the Structure of Temperament Questionnaire (STQ-77PT) based on a Brazilian sample. *Natural Systems of Mind, 1*(1), 35–47. https://doi.org/10.38098/nsom_2021_01_03_04
3. Aunger, R., Gallyamova, A., & Grigoryev, D. (2025). Network psychometric-based identification and structural analysis of a set of evolved human motives. *Personality and Individual Differences, 233*, 112921. https://doi.org/10.1016/j.paid.2024.112921
4. Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory package for the R environment. *Journal of Statistical Software, 48*(6), 1–29.

   https://doi.org/10.18637/jss.v048.i06
5. Del Giudice, M. (2022). Measuring sex differences and similarities. In D. P. VanderLaan & W. I. Wong (Eds.), *Gender and sexuality development: Contemporary theory and research*. Springer. https://doi.org/10.1007/978-3-030-84273-4_1
6. Epskamp, S., Cramer, A. O., Waldorp, L. J., Schmittmann, V. D., & Borsboom, D. (2012). qgraph: Network visualizations of relationships in psychometric data. *Journal of Statistical Software, 48*(4), 1–18.

   https://doi.org/10.18637/jss.v048.i04
7. Figueredo, A. J., De Baca, T. C., Black, C. J., García, R. A., Fernandes, H. B. F., Wolf, P. S. A., & Anthony, M. (2015). Methodologically sound: Evaluating the psychometric approach to the assessment of human life history. *Evolutionary Psychology, 13*(2), 299–338.

   https://doi.org/10.1177/147470491501300202
8. Robitzsch, A. (2024). *sirt: Supplementary Item Response Theory models* (Version 4.1-15) [Computer software]. Comprehensive R Archive Network. https://CRAN.R-project.org/package=sirt
9. Rusalov, V. M. (1989). Object-related and communicative aspects of human temperament: A new questionnaire of the structure of temperament. *Personality and Individual Differences, 10*(8), 817–827.

   https://doi.org/10.1016/0191-8869(89)90017-2
10. Rusalov, V. M. (2018). Functional systems theory and the activity-specific approach in psychological taxonomies. *Philosophical Transactions of the Royal Society B: Biological Sciences, 373*(1744), 20170166.

    https://doi.org/10.1098/rstb.2017.0166
11. Rusalov, V. M., & Trofimova, I. N. (2007). *The structure of temperament and its measurement: The theory and the manual of the structure of temperament questionnaire (STQ).* Psychological Services Press.
12. Rusalov, V. M., & Trofimova, I. N. (2011). On representation of types of psychological arousal in various models of temperament. *Psychological Journal, 32*(3), 74–84.
13. Trofimova, I. N. (2010a). An investigation into differences between the structure of temperament and the structure of personality. *The American Journal of Psychology, 123*(4), 467–480.

    https://doi.org/10.5406/amerjpsyc.123.4.0467
14. Trofimova, I. N. (2010b). Exploration of the activity-specific model of temperament in four languages. *International Journal of Psychology and Psychological Therapy, 10*(1), 77–94.
15. Trofimova, I. N. (2015). Do psychological sex differences reflect evolutionary bisexual partitioning? *The American Journal of Psychology, 128*(4), 485–514.

    https://doi.org/10.5406/amerjpsyc.128.4.0485
16. Trofimova, I. N. (2016). The interlocking between functional aspects of activities and a neurochemical model of adult temperament. In M. C. Arnold (Ed.), *Temperaments: Individual differences, social and environmental influences and impact on quality of life* (pp. 77–147). Nova Science Publishers.
17. Trofimova, I. N. (2019). An overlap between mental abilities and temperament traits. In D. McFarland (Ed.), *General and specific mental abilities* (pp. 77–114). Cambridge Scholars Publishing.
18. Trofimova, I. N. (2021a). Contingent tunes of neurochemical ensembles in the norm and

pathology: Can we see the patterns? *Neuropsychobiology, 80*(2), 101–133. https://doi.org/10.1159/000513688

19. Trofimova, I. N. (2021b). Functional constructivism approach to multilevel nature of biobehavioural diversity. *Frontiers in Psychiatry, 12*, 641286.

    https://doi.org/10.3389/fpsyt.2021.641286

20. Trofimova, I. N. (2022). Transient nature of stable behavioural patterns, and how we can respect it. *Current Opinion in Behavioral Sciences, 44*, 101109.

    https://doi.org/10.1016/j.cobeha.2022.101109

21. Trofimova, I. N., & Araki, M. E. (2022). Psychometrics vs neurochemistry: A controversy around mobility-like scales of temperament. *Personality and Individual Differences, 187*, 111446.

    https://doi.org/10.1016/j.paid.2021.111446

22. Trofimova, I. N., & Araki, M. E. (2024). The importance of activity-specific differentiation between orientation-related temperament traits. *Current Psychology, 43*(9), 7913–7923.

    https://doi.org/10.1007/s12144-023-04996-1

23. Trofimova, I. N., & Gaykalova, A. A. (2021). Emotionality vs. other biobehavioural traits: A look at neurochemical biomarkers for their differentiation. *Frontiers in Psychology, 12*, 781631. https://doi.org/10.3389/fpsyg.2021.781631

24. Trofimova, I. N., & Robbins, T. W. (2016). Temperament and arousal systems: A new synthesis of differential psychology and functional neurochemistry. *Neuroscience & Biobehavioral Reviews, 64*, 382–402.

    https://doi.org/10.1016/j.neubiorev.2016.03.008

25. Trofimova, I. N., & Sulis, W. (2011). Is temperament activity-specific? Validation of the Structure of Temperament Questionnaire-Compact (STQ-77). *International Journal of Psychology and Psychological Therapy, 11*(3), 389–400.

26. Trofimova, I. N., & Sulis, W. (2018). There is more to mental illness than negative affect: Comprehensive temperament profiles in depression and generalized anxiety. *BMC Psychiatry, 18*(1), 125.

    https://doi.org/10.1186/s12888-018-1695-x

Подход на основе сетевой психометрики и теории ответов на задания (IRT) к валидизации русскоязычной версии Опросника Структуры Темперамента для взрослых (STQ-77Ru)

Галлямова А.А., Григорьев Д.С.

НИУ ВШЭ, Москва, Россия

Абстракт: В этом исследовании изучаются психометрические свойства русскоязычной версии Опросника Структуры Темперамента для взрослых (STQ-77Ru) на разнообразной российской выборке. Используя как сетевую психометрику, так и теорию ответов на задания (IRT), мы проанализировали данные 3442 россиян в возрасте от 18 до 81 года (M = 38, SD = 11). Сетевая психометрика использовалась для комплексного изучения взаимозависимостей между чертами темперамента, в то время как IRT предоставила подробные сведения о сложности и дискриминативности отдельных тестовых заданий (пунктов опросника). Тесты эквивалентности измерений подтвердили инвариантность шкал в факторных нагрузках и константах по полу и уровням образования, что обосновывает применимость STQ-77Ru в разных демографических группах. В этом исследовании дополнительно оценивалась способность STQ-77Ru различать демографические группы, выявляя существенные отличия в чертах темперамента между мужчинами и женщинами и на разных уровнях образования. Полученные результаты подтверждают эффективность методики и, с некоторыми оговорками, подчеркивают потенциал STQ-77Ru для предоставления точной и надежной психометрической оценки черт темперамента в российской выборке.

Ключевые слова: черты темперамента, функциональный ансамбль темперамента, структура опросника темперамента, сетевой анализ, теория ответа на задания

Citation: Gallyamova, A. & Grigoryev, D. (2025). Network Psychometric and Item Response Theory (IRT) Approach to Validating the Russian Adult Version of the Structure of Temperament Questionnaire (STQ-77Ru). *Natural Systems of Mind, 5(1),* 27–41. doi: 10.38098/nsom_2025_05_01_03

Information about the authors

Albina Gallyamova, Junior Research Fellow, Center for Sociocultural Research, HSE University; 20 Myasnitskaya Str., Moscow, 101000, Russia; ORCID: 0000-0002-8775-7289, ResearcherID: GLV6876-2022, ScopusID: 58182813400, e-library SPIN-code: 6639-2529; e-mail: aagallyamova@hse.ru

Dmitry Grigoryev, PhD (Psychology), Research Fellow, Center for Sociocultural Research, HSE University; 20 Myasnitskaya Str., 101000 Moscow, Russia. ORCID: 0000-0003-4511-7942, ResearcherID: K-3338-2015, ScopusID: 57191706675, e-library SPIN-code: 1807-9739, e-mail: dgrigoryev@hse.ru

Информация об авторах

Галлямова Альбина Аликовна, младший научный сотрудник, Центр социокультурных исследований, Национальный исследовательский университет «Высшая школа экономики», Мясницкая ул., 20, 101000 Москва, Россия. ORCID: 0000-0002-8775-7289, ResearcherID: GLV6876-2022, ScopusID: 58182813400, e-library SPIN-код: 6639-2529; e-mail: aagallyamova@hse.ru

Григорьев Дмитрий Сергеевич, научный сотрудник, Центр социокультурных исследований, Национальный исследовательский университет «Высшая школа экономики»; Россия, 101000, Москва Мясницкая ул., 20; ORCID: 0000-0003-4511-7942, ResearcherID: K-3338-2015, ScopusID: 57191706675, e-library SPIN-код: 1807-9739; e-mail: dgrigoryev@hse.ru

УДК и определить соответствие специальности ВАК:

УДК 159.9.072.59

5.3.1.27. Темперамент и характер. Диагностика темперамента. Структура и типология характера. Акцентуация характера, их диагностика