



Test Partnership  
**TPAQ Series**  
Psychometric Properties

2018

# 1 Construct Validity

The IPIP-NEO-120 (Johnson, 2014) is a validated measure of the Big-5 model of personality, specifically the OCEAN model (Costa & McCrae, 1992). The questionnaire is comprised of 120 items, with multiple sub-scales per superordinate construct. The sub-scales are combined to provide a measure of each of the following constructs: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. These scales were used to validate the corresponding TPAQ scales and their sub-scales: Openness to Experience, Conscientiousness, Extraversion, Agreeableness, and Emotional Stability.

To validate the additional measures included in the TPAQ, appropriate scales were accessed via the International Personality Item Pool (IPIP, n.d.). Relevant sub-scales were selected from a combination of available IPIP Inventories and single scale constructs, to validate the TPAQ Resilience, TPAQ Integrity, TPAQ Emotional Intelligence and TPAQ Industriousness measures.

## Openness to Experience

To validate Test Partnership's TPAQ Openness to Experience scales, the following sub-scales were taken from the IPIP-NEO-120, Openness to Experience (Johnson, 2014): Imagination (O1), Artistic Interests (O2), Adventurousness (O3), Intellect (O4) and Liberalism (O5).

**Table 1.1: Correlations between TPAQ Scales and the IPIP-NEO-120 Scales**

TPAQ Scale	O1	O2	O3	O4	O5	Openness to Experience
Imaginative	.72*					
Artistic		.61				
Spontaneity			.52			
Inquisitive				.58*		
Free-Thinking					.36*	
Openness TPAQ-45						.74
Openness TPAQ-27						.70

*Note: All correlations are significant at the 0.001 level. Correlations have not been corrected for attenuation, or restriction of range. \* = scales included in the TPAQ-27, whereas all are included in the TPAQ-45.*

Openness to Experience is a composite of the sub-scales shown and provides a measure of general Openness to Experience. As the TPAQ is available in two forms, a composite measure has been provided for both the TPAQ-45 and the TPAQ-27.

All correlation coefficients (excluding Free-Thinking, which can be interpreted as “moderate”) can be interpreted as ranging from large to very large (Cohen, 1988), strongly supporting the construct validity of the TPAQ Openness to Experience scales.

### Conscientiousness

To validate Test Partnership’s TPAQ Conscientiousness scales, the following sub-scales were taken from the IPIP-NEO-120, Conscientiousness (Johnson, 2014): Self-Efficacy (C1), Orderliness (C2), Dutifulness (C3), Self-Discipline (C4) and Cautiousness (C5).

Conscientiousness is a composite of the sub-scales shown and provides a measure of general Conscientiousness. As the TPAQ is available in two forms, a composite measure has been provided for both the TPAQ-45 and the TPAQ-27.

**Table 1.2: Correlations between TPAQ Scales and the IPIP-NEO-120 Scales**

TPAQ Scale	C1	C2	C3	C4	C5	Conscientiousness
Self-Confidence	.61					
Methodical		.51*				
Diligence			.59			
Discipline				.71*		
Vigilance					.60*	
Conscientiousness TPAQ-45						.80
Conscientiousness TPAQ-27						.76

*Note: All correlations are significant at the 0.001 level. Correlations have not been corrected for attenuation, or restriction of range. \* = scales included in the TPAQ-27, whereas all are included in the TPAQ-45.*

All correlation coefficients can be interpreted as ranging from large to very large (Cohen, 1988), strongly supporting the construct validity of the TPAQ Conscientiousness scales.

## Extraversion

To validate Test Partnership's TPAQ Extraversion scales, the following sub-scales were taken from the IPIP-NEO-120, Extraversion (Johnson, 2014): Friendliness (E1), Gregariousness (E2), Assertiveness (E3), Activity Level (E4), Excitement Seeking (E5) and Cheerfulness (E6).

Extraversion is a composite of the sub-scales shown and provides a measure of general Extraversion. As the TPAQ is available in two forms, a composite measure has been provided for both the TPAQ-45 and the TPAQ-27.

**Table 1.3: Correlations between TPAQ Scales and the IPIP-NEO-120 Scales**

TPAQ Scale	E1	E2	E3	E4	E5	E6	Extraversion
Outgoing	.71						
Sociable		.72*					
Assertive			.64*				
Energetic				.55*			
Adventurous					.52		
Cheerful						.63	
Extraversion TPAQ-45							.84
Extraversion TPAQ-27							.78

*Note: All correlations are significant at the 0.001 level. Correlations have not been corrected for attenuation, or restriction of range. \* = scales included in the TPAQ-27, whereas all are included in the TPAQ-45.*

All correlation coefficients can be interpreted as ranging from large to very large (Cohen, 1988), strongly supporting the construct validity of the TPAQ Extraversion scales.

## Agreeableness

To validate Test Partnership's TPAQ Agreeableness scales, the following sub-scales were taken from the IPIP-NEO-120, Agreeableness (Johnson, 2014): Trust (A1), Morality (A2), Altruism (A3), Cooperation (A4) and Sympathy (A5).

Agreeableness is a composite of the sub-scales shown and provides a measure of general Agreeableness. As the TPAQ is available in two forms, a composite measure has been provided for both the TPAQ-45 and the TPAQ-27.

**Table 1.4: Correlations between TPAQ Scales and the IPIP-NEO-120 Scale**

TPAQ Scale	A1	A2	A3	A4	A5	Agreeableness
Trusting	.71					
Principled		.49				
Altruistic			.58*			
Cooperative				.48*		
Sympathetic					.56*	
Agreeableness TPAQ-45						.77
Agreeableness TPAQ-27						.70

*Note: All correlations are significant at the 0.001 level. Correlations have not been corrected for attenuation, or restriction of range. \* = scales included in the TPAQ-27, whereas all are included in the TPAQ-45.*

All correlation coefficients can be interpreted as ranging from moderate to very large (Cohen, 1988), strongly supporting the construct validity of the TPAQ Agreeableness scales.

### Emotional Stability

To validate Test Partnership’s TPAQ Emotional Stability scales, the following sub-scales were taken from the IPIP-NEO-120, Neuroticism (Johnson, 2014): Anxiety (N1), Anger (N2), Depression (N3), Self-Consciousness (N4), Immoderation (N5) and Vulnerability (N6).

Neuroticism is a composite of the sub-scales shown and provides a measure of general Neuroticism. As the TPAQ is available in two forms, a composite measure has been provided for both the TPAQ-45 and the TPAQ-27.

All correlation coefficients, as shown in Table 1.5, can be interpreted as ranging from large to very large (Cohen, 1988), strongly supporting the construct validity of the TPAQ Emotional Stability scales, via divergent validity.

**Table 1.5: Correlations between TPAQ Scales and the IPIP-NEO-120 Scales**

TPAQ Scale	N1	N2	N3	N4	N5	N6	Neuroticism
Relaxed	-.68						
Calm		-.73*					
Positivity			-.75				
Assurance				-.62*			
Restraint					-.63		
Composure						-.71*	
Emotional Stability TPAQ-45							-.87
Emotional Stability TPAQ-27							-.80

*Note: All correlations are significant at the 0.001 level. Correlations have not been corrected for attenuation, or restriction of range. \* = scales included in the TPAQ-27, whereas all are included in the TPAQ-45.*

## Resilience

To validate Test Partnership's TPAQ Resilience scales, the following measures of the Core Self Evaluation (CSE) Model were accessed via the International Personality Item Pool (IPIP, n.d.): Locus of Control (R1), Emotional Stability (R2), Self-Efficacy (R3) and Self-Esteem (R4).

Resilience is a composite of the CSE sub-scales shown and provides a measure of general Resilience. As the TPAQ is available in two forms, a composite measure has been provided for both the TPAQ-45 and the TPAQ-27.

All correlation coefficients (excluding Confidence, which can be interpreted as "moderate") can be interpreted as ranging from large to very large (Cohen, 1988), strongly supporting the construct validity of the TPAQ Resilience scales. The full range of results is presented in Table 1.6.

**Table 1.6: Correlations between TPAQ Scales and the CSE Scales**

TPAQ Scale	R1	R2	R3	R4	Resilience
Self-Directed	.51*				
Collected		.62			
Confidence			.45*		
Self-Esteem				.67*	
Resilience TPAQ-45					.80
Resilience TPAQ-27					.76

Note: All correlations are significant at the 0.001 level. Correlations have not been corrected for attenuation, or restriction of range. \* = scales included in the TPAQ-27, whereas all are included in the TPAQ-45.

### Integrity

To validate Test Partnership's TPAQ Integrity scales, the following sub-scales were taken from Lee and Ashton's (2004) Hexaco Personality Inventory (HEXACO-PI), Honesty-Humility Facets: Sincerity (H1), Fairness (H2) Greed Avoidance (H3) and Modesty (H4).

Integrity is a composite of the sub-scales shown and provides a measure of general Integrity. As the TPAQ is available in two forms, a composite measure has been provided for both the TPAQ-45 and the TPAQ-27.

**Table 1.7: Correlations between TPAQ Scales and HEXACO-PI Scales**

TPAQ Scale	H1	H2	H3	H4	Honesty-Humility
Honesty	.47*				
Fairness		.51*			
Greed Aversion			.55		
Modesty				.50*	
Integrity TPAQ-45					.73
Integrity TPAQ-27					.69

Note: All correlations are significant at the 0.001 level. Correlations have not been corrected for attenuation, or restriction of range. \* = scales included in the TPAQ-27, whereas all are included in the TPAQ-45.

All correlation coefficients (excluding Honesty, which can be interpreted as “moderate”) can be interpreted as ranging from large to very large (Cohen, 1988), strongly supporting the construct validity of the TPAQ Integrity scales.

### Emotional Intelligence

To validate Test Partnership’s TPAQ Emotional Intelligence scales, the following sub-scales were taken from Barchard’s (2001) Components of Emotional Intelligence (CEI): Positive Expressivity (EI1), Negative Expressivity (EI2), Attending to Emotions (EI3), Emotion-Based Decision-Making (EI4) and Empathic Concern (EI5).

Emotional Intelligence (EI) is a composite of the sub-scales shown and provides a measure of general Emotional Intelligence. As the TPAQ is available in two forms, a composite measure has been provided for both the TPAQ-45 and the TPAQ-27.

**Table 1.8: Correlations between TPAQ Scales and CEI Scales**

TPAQ Scale	EI1	EI2	EI3	EI4	EI5	EI
Positive Expression	.55					
Negative Expression		.61				
Emotional Awareness			.66*			
Intuitive				.58*		
Empathetic					.51*	
EI TPAQ-45						.75
EI TPAQ-27						.71

*Note: All correlations are significant at the 0.001 level. Correlations have not been corrected for attenuation, or restriction of range. \* = scales included in the TPAQ-27, whereas all are included in the TPAQ-45.*

All correlation coefficients can be interpreted as ranging from moderate to very large (Cohen, 1988), strongly supporting the construct validity of the TPAQ Emotional Intelligence scales.

### Industriousness

To validate Test Partnership’s TPAQ Industriousness scales, appropriate scales were accessed via the International Personality Item Pool (IPIP, n.d.). The following sub-scales



were taken from a combination of available IPIP Inventories and single scale constructs (VIA, Peterson & Seligman, 2004; HPI, Hogan & Hogan, 1992; Grit-S, Duckworth & Quinn, 2009; TCI, Cloninger, et al., 1994; IPIP-NEO-120, Johnson, 2014): Perseverance (I1), Competitive (I2), Grit (I3), Initiative (I4) and Achievement-Striving (I5).

Industriousness is a composite of the sub-scales shown and provides a measure of general Industriousness. As the TPAQ is available in two forms, a composite measure has been provided for both the TPAQ-45 and the TPAQ-27.

All correlation coefficients, as shown in Table 1.9, can be interpreted as ranging from moderate to very large (Cohen, 1988), strongly supporting the construct validity of the TPAQ Industriousness scales.

**Table 1.9: Correlations between TPAQ Integrity Scales and Validity Scales**

TPAQ Scale	I1	I2	I3	I4	I5	Industriousness
Persistence	.69*					
Drive		.53				
Perseverance			.70			
Initiative				.70*		
Goal Focus					.57*	
Industriousness TPAQ-45						.84
Industriousness TPAQ-27						.80

*Note: All correlations are significant at the 0.001 level. Correlations have not been corrected for attenuation, or restriction of range. \* = scales included in the TPAQ-27, whereas all are included in the TPAQ-45.*

### Summary

All correlation coefficients for the TPAQ scales can be interpreted as ranging from moderate to very large, as correlations of 0.1 are considered small, 0.3 are considered moderate and 0.5 and above are considered to be large (Cohen, 1988); supporting the construct validity of the TPAQ scales

## 2 Group Differences

The following protected groups were investigated for evidence of average group differences and unfair discrimination:

- *Gender*: Participants that reported to be male were compared against participants reporting to be female.
- *Ethnicity*: Participants that reported being white were compared against those who reported to be BME participants.
- *Age*: Participants that reported being under the age of 40 were compared against the average scores of participants reporting to be over the age of 40.

TPAQ Scale	Group Differences in terms of Cohen's <i>d</i>		
	Gender	Age	Ethnicity
Imaginative	-0.10	0.38	0.07
Artistic	-0.17	0.02	-0.25
Spontaneity	-0.02	-0.02	-0.17
Inquisitive	0.16	-0.06	0.06
Free-Thinking	-0.01	0.18	0.15
Self-Confidence	0.23	-0.27	-0.24
Methodical	-0.06	-0.12	-0.21
Diligence	0.03	-0.28	-0.14
Discipline	0.05	-0.28	-0.23
Vigilance	0.11	-0.21	-0.13
Outgoing	0.03	-0.21	-0.10
Sociable	0.13	-0.13	-0.22
Assertive	0.17	-0.23	-0.13
Energetic	-0.05	-0.10	-0.08
Adventurous	0.16	0.25	-0.23

TPAQ Scale	Group Differences in terms of Cohen's <i>d</i>		
	Gender	Age	Ethnicity
Cheerful	-0.07	-0.18	-0.13
Trusting	0.02	-0.41	-0.04
Principled	-0.16	-0.26	-0.08
Altruistic	-0.21	-0.27	-0.07
Cooperative	-0.11	-0.27	-0.01
Sympathetic	-0.34	-0.19	-0.05
Relaxed	0.28	-0.31	-0.13
Calm	0.20	-0.24	-0.08
Positivity	0.17	-0.29	-0.20
Assurance	0.22	-0.28	-0.16
Restraint	0.11	-0.24	-0.10
Composure	0.31	-0.41	0.01
Self-Directed	0.14	-0.20	0.02
Collected	0.32	-0.34	-0.17
Confidence	0.20	-0.33	-0.19
Self-Esteem	0.13	-0.35	-0.17
Honesty	-0.04	-0.30	-0.07
Fairness	-0.09	-0.28	-0.17
Greed Aversion	-0.13	-0.40	0.30
Modesty	-0.14	-0.23	0.20
Positive Expression	-0.31	-0.10	0.01
Negative Expression	-0.11	-0.13	0.07
Emotional Awareness	-0.24	-0.06	0.10
Intuitive	-0.39	-0.02	0.33

TPAQ Scale	Group Differences in terms of Cohen's <i>d</i>		
	Gender	Age	Ethnicity
Empathetic	-0.27	-0.23	0.03
Persistence	0.14	-0.21	-0.27
Drive	0.20	-0.14	-0.24
Perseverance	0.09	-0.15	-0.31
Initiative	0.06	-0.36	-0.14
Goal Focus	0.06	-0.10	-0.24

*Table 2.1: Average Score Effect Sizes Across Protected Groups*

### Summary

All effect sizes for each of the TPAQ scales can be interpreted as being small or non-existent (Cohen, 1988). Group differences across Age, Gender and Ethnicity generally report small to negligible group differences between focal and reference groups. None of the effect sizes were greater than 0.5, and therefore do not suggest evidence of practically significant differences between the protected groups included in analyses.

### 3 Reliability

The Rasch model provides two separate forms of reliability: person reliability and item reliability.

#### Person Reliability

Person reliability is analogous to traditional notions of test reliability, indicating a level of precision when measuring the personality traits of the persons being measured. Rasch person reliability can be interpreted in the same frame of reference as Cronbach's Alpha, with a minimum of .70 as an acceptable level of test reliability. All estimated person reliabilities are presented in Table 3.1 below

TPAQ Scale	Sample Size	SD	S.E.	Reliability
Imaginative	10,512	1.35	0.30	.95
Artistic	10,522	1.25	0.30	.94
Spontaneity	10,355	1.28	0.31	.94
Inquisitive	10,257	1.36	0.30	.95
Free-Thinking	10,485	1.07	0.31	.92
Self-Confidence	10,379	1.51	0.30	.96
Methodical	10,587	1.15	0.29	.94
Diligence	10,494	1.22	0.30	.94
Discipline	10,658	1.28	0.30	.95
Vigilance	10,474	1.30	0.30	.95
Outgoing	10,556	1.33	0.29	.95
Sociable	10,536	1.27	0.30	.94
Assertive	10,676	1.21	0.29	.94
Energetic	10,567	1.10	0.29	.93
Adventurous	10,519	1.35	0.32	.94
Cheerful	10,499	1.25	0.29	.95
Trusting	10,681	1.28	0.33	.93
Principled	9,939	1.32	0.27	.96
Altruistic	10,460	1.34	0.28	.96
Cooperative	9,962	1.22	0.30	.94
Sympathetic	10,259	1.38	0.27	.96
Relaxed	10,657	1.23	0.30	.94

TPAQ Scale	Sample Size	SD	S.E.	Reliability
Calm	10,220	1.68	0.31	.97
Positivity	9,836	1.92	0.32	.97
Assurance	10,567	1.29	0.30	.95
Restraint	10,558	1.24	0.30	.94
Composure	10,367	1.59	0.31	.96
Self-Directed	10,289	1.20	0.29	.94
Collected	10,621	1.25	0.29	.95
Confidence	10,413	1.60	0.31	.96
Self-Esteem	10,112	1.57	0.30	.96
Honesty	10,229	1.25	0.28	.95
Fairness	10,210	1.22	0.27	.95
Greed Aversion	10,508	1.29	0.32	.94
Modesty	10,421	1.23	0.32	.93
Positive Expression	10,424	1.39	0.29	.96
Negative Expression	10,741	1.04	0.30	.92
Emotional Awareness	10,409	1.46	0.29	.96
Intuitive	10,722	1.13	0.31	.92
Empathetic	10,398	1.36	0.28	.96
Persistence	10,329	1.60	0.32	.96
Drive	10,479	1.26	0.30	.94
Perseverance	10,235	1.66	0.32	.96
Initiative	10,527	1.49	0.32	.95
Goal Focus	9,950	1.56	0.29	.97

Table 3.1: TPAQ Scale Person Reliability Measures

## Summary

All estimated reliabilities exceed the minimum required standard of .70, with the lowest estimated reliability at .92. This level of reliability ensures a high level of accuracy and precision when assessing candidates, regardless of the scales used or the population being measured. Although these coefficients are already large, they are conservative values, as the method used to calculate the reliability underestimates reliability compared to Cronbach's alpha.

## Item Reliability

Rasch item reliability is based on the initial calibration samples used to generate item difficulty parameters. As the difficulty calibrations generated during item trialling are the difficulty parameters for the live assessment, the item reliabilities from initial calibration can be directly reported, rather than estimated. Item difficulty reliabilities for each TPAQ scale are shown in Table 3.2 below:

TPAQ Scale	Sample Size	Difficulty SD	S.E.	Item Reliability
Imaginative	10,512	0.28	0.02	.99
Artistic	10,522	0.48	0.02	1.00
Spontaneity	10,355	0.51	0.02	1.00
Inquisitive	10,257	0.26	0.02	.99
Free-Thinking	10,485	0.52	0.02	1.00
Self-Confidence	10,379	0.44	0.02	1.00
Methodical	10,587	0.29	0.02	.99
Diligence	10,494	0.46	0.02	1.00
Discipline	10,658	0.51	0.02	1.00
Vigilance	10,474	0.48	0.02	1.00
Outgoing	10,556	0.46	0.02	1.00
Sociable	10,536	0.56	0.02	1.00
Assertive	10,676	0.33	0.02	1.00
Energetic	10,567	0.29	0.02	1.00
Adventurous	10,519	0.46	0.02	1.00
Cheerful	10,499	0.29	0.02	.99
Trusting	10,681	0.45	0.02	1.00
Principled	9,939	0.29	0.02	.99
Altruistic	10,460	0.34	0.02	1.00
Cooperative	9,962	0.31	0.02	1.00
Sympathetic	10,259	0.33	0.02	1.00
Relaxed	10,657	0.46	0.02	1.00
Calm	10,220	0.32	0.02	.99
Positivity	9,836	0.33	0.02	.99
Assurance	10,567	0.36	0.02	1.00
Restraint	10,558	0.34	0.02	1.00

TPAQ Scale	Sample Size	Difficulty SD	S.E.	Item Reliability
Composure	10,367	0.37	0.02	1.00
Self-Directed	10,289	0.36	0.02	1.00
Collected	10,621	0.39	0.02	1.00
Confidence	10,413	0.40	0.02	1.00
Self-Esteem	10,112	0.44	0.02	1.00
Honesty	10,229	0.29	0.02	1.00
Fairness	10,210	0.35	0.02	1.00
Greed Aversion	10,508	0.48	0.02	1.00
Modesty	10,421	0.47	0.02	1.00
Positive Expression	10,424	0.25	0.02	.99
Negative Expression	10,741	0.23	0.02	.99
Emotional Awareness	10,409	0.20	0.02	.99
Intuitive	10,722	0.34	0.02	1.00
Empathetic	10,398	0.32	0.02	1.00
Persistence	10,329	0.36	0.02	1.00
Drive	10,479	0.42	0.02	1.00
Perseverance	10,235	0.37	0.02	1.00
Initiative	10,527	0.42	0.02	1.00
Goal Focus	9,950	0.21	0.02	.99

Table 3.2: TPAQ Scale Item Reliability Measures

## Summary

A general recommendation for Rasch item reliability is a minimum of .90. The observed item reliabilities range from .99-1.00, exceeding the minimum standard for Rasch item reliability. This high level of reliability ensures that item difficulty calibrations are accurate and precise, permitting their use in high stakes selection and assessment. It also ensures that item calibrations are highly stable, allowing these initial items to be used in common item linking, as part of ongoing trials for item bank expansion.





## References

- Barchard, K. A. (2001). *Emotional and social intelligence: Examining its place in the nomological network*. Unpublished Doctoral Dissertation: Department of Psychology; University of British Columbia; Vancouver, BC; Canada.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. 2nd.
- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (Neo-PI-R) and NEO Five-Factor Inventory (NEO-FFI): Professional manual*. Odessa, FL: Psychological Assessment Resources.
- Cloninger, C. R., Przybeck, T. R., Svrakic, D. M., & Wetzel, R. D. (1994). *The Temperament and Character Inventory (TCI): A guide to its development and use*. St. Louis, MO: Center for Psychobiology of Personality, Washington University.
- 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.
- Hogan, R., & Hogan, J. (1992). *Hogan Personality Inventory Manual (2nd ed.)*. Tulsa, OK: Hogan Assessment Systems.
- International Personality Item Pool. (n.d.). International Personality Item Pool: A Scientific Collaboratory for the Development of Advanced Measures of Personality Traits and Other Individual Differences (<http://ipip.ori.org/>). Internet Web Site.
- Johnson J. A. (2014). Measuring thirty facets of the Five Factor Model with a 120-item public domain inventory: Development of the IPIP-NEO-120. *Journal of Research in Personality*, 51, 78-89.
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO Personality Inventory. *Multivariate Behavioural Research*, 39, 329-358.
- Peterson, C., & Seligman, M. E. P. (2004). *Character strengths and virtues: A handbook and classification*. New York: Oxford University Press/Washington, DC: American Psychological Association.