

# MSP Research Note

## TLQ Reliability, Validity and Norms

### Introduction

This research note describes the reliability and validity of the TLQ. Evidence for the reliability and validity of is presented against some of the key the criteria in the EFPA Review Model for the Description and Evaluation of Psychological Tests (Bartram, 2002). The EFPA Review Model was produced to support and encourage the process of harmonising the reviewing of tests. It provides a standard set of criteria to assess the quality of tests. These cover the common areas of test review such as norms, reliability and validity.

### Reliability

#### Internal consistency reliabilities

Table 1 shows the TLQ scale reliabilities. The internal consistency for the questionnaire as a whole is 0.94 and the median internal consistency for the TLQ scales is 0.70. The reliability of the questionnaire when considered as a single measure is excellent, and the reliabilities of the scales are adequate to good as defined by the EFPA review model (Bartram, 2002).

The TLQ scale sten score SEms range from 1.23 to 1.73. This means that there is a 68% likelihood that the person's true score on the scales will be 1-1.75 Stens either

side of the observed score. The Sten score SEM band around the TLQ total score is smaller (about 1) because the reliability is significantly higher (0.94).

**Table 1.** Internal consistency reliabilities for the TLQ (n = 5,000)

Scale	Alpha	Mean	SD	Raw score SEM	Sten score SEM
Transformational leadership	0.94	248.16	27.83	6.82	0.67
Empathy	0.66	31.69	3.57	2.08	1.50
Motivation	0.78	32.00	4.39	2.06	1.23
Performance	0.60	28.86	4.22	2.67	1.63
Communication	0.69	29.25	4.51	2.51	1.64
Empowerment	0.60	28.13	3.97	2.51	1.73
Commitment	0.77	33.98	4.24	2.03	1.28
Feedback	0.77	32.52	4.69	2.25	1.13
Achievement	0.70	31.74	4.35	2.38	1.53

## Construct Validity

### Scale intercorrelations

Table 2 shows the intercorrelations of the CSQ scales. There are moderate to high correlations between the scales ranging from 0.53 to 0.74.

**Table 2.** Inter-correlations of TLQ scales (n=5,000)

Scale	Empathy	Motivation	Performance	Communication	Empowerment	Commitment	Feedback	Achievement
Empathy	1.00	0.59	0.54	0.56	0.60	0.66	0.66	0.55
Motivation		1.00	0.65	0.69	0.57	0.65	0.74	0.74
Performance			1.00	0.71	0.58	0.59	0.67	0.56
Communication				1.00	0.56	0.60	0.68	0.62
Empowerment					1.00	0.53	0.57	0.55
Commitment						1.00	0.67	0.68
Feedback							1.00	0.65

In order to determine how well a questionnaire differentiates between the different dimensions it is designed to measure, it is necessary to correct the correlations for unreliability. A correlation needs to be divided by the square root of the product of the two variables' reliability to determine what the correlation between the two variables would be if the variables' reliabilities were perfect. If two scales share less than 50% reliable variance, then we can be reasonably certain that they are independent.

Using 50% as a benchmark, it is apparent that there is substantial overlap between the TLQ scales as all the scale pairs share 50% or more common variance with each other (Table 3).

**Table 3.** Percentage of common reliable variance for TLQ scales (N = 5,000)

Scale	Motivation	Performance	Communication	Empowerment	Commitment	Feedback	Achievement
Empathy	0.68	0.74	0.69	0.91	0.86	0.86	0.65
Motivation		0.90	0.88	0.69	0.70	0.91	1.00
Performance			0.87	0.93	0.75	0.97	0.75
Communication				0.76	0.68	0.87	0.80
Empowerment					0.61	0.70	0.72
Commitment						0.76	0.86
Feedback							0.78

### Standard error of difference

The standard error of difference (SEd) helps us determine the size of the gap that you need to see between a person's scores on any two scales before you can conclude that the difference is real. The SEd depends on the reliability of the scales – the higher the reliability the smaller the SEd is. If there are two full SEs between the scores on two scales, then there is a 95% likelihood that there is a real difference.

All the SEs in Table 4 are around 2 so a difference of around 4 or more in TLQ raw scores constitutes a real difference. This is equivalent to a difference of 3 Stens.

**Table 4.** Standard error of difference of TLQ scales (n = 5,000)

Scale	Motivation	Performance	Communication	Empowerment	Commitment	Feedback	Achievement
Empathy	1.94	2.22	2.22	2.29	1.97	1.88	2.14
Motivation		2.04	2.05	2.12	1.78	1.67	1.96
Performance			2.31	2.38	2.07	1.98	2.24
Communication				2.38	2.08	1.99	2.24
Empowerment					2.15	2.07	2.31
Commitment						1.71	1.99
Feedback							1.90

### Factor analysis

Principal factors extraction with varimax rotation was performed on the TLQ scales on a sample of 5,000 respondents (see Table 5). The Kaiser-Meyer-Olkin measure of sampling adequacy was 0.82, well above 0.6 required for a good factor analysis. The variables were on the whole well-defined by the factor solution. Community values were moderate to fairly high. One factor was extracted accounting for 67% of the variance indicating that transformational leadership style is a single construct.

**Table 5.** Factor loadings and communalities for principal components extraction on TLQ scales (n = 5,000)

Scale	F1	Communality
Empathy	0.78	0.62
Motivation	0.86	0.75
Performance	0.81	0.66
Communication	0.83	0.69
Empowerment	0.75	0.56
Commitment	0.82	0.68
Feedback	0.86	0.75
Achievement	0.82	0.67

## Criterion-related validity

Table 6 shows the correlations between TLQ scale scores and job appraisal ratings. This is based on a sample of 4,487 respondents who completed the TLQ on the internet. Respondents were asked to report how their manager assessed their performance at their last performance appraisal using a 4-point scale (excellent, good, satisfactory, poor) and to assess their own performance. Combined ratings shown in the table are the sum of the manager's assessment and the test taker's self-assessment.

There are statistically significant correlations at 0.2-0.4 between job performance and all eight factors. These correlations are consistent with those reported in the literature for personality assessment instruments. For example, Robertson (1997) notes that even with meta-analytic corrections, the upper limits for the validity of personality variables against overall work performance variables are in the range of 0.25 to 0.4.

Regression analysis was used to help understand the contribution of the different leadership scales to job appraisal ratings. A standard multiple regression was performed between managerially and self-assessed job performance combined as the dependent variable and the TLQ scales as the independent variables.

Table 6 displays the correlations between the variables, the unstandardised regression coefficients (B), the semi-partial correlations ( $sr^2$ ) and R,  $R^2$  and adjusted  $R^2$ . For the combined ratings, R for regression was significantly different from zero,  $F(8, 4,478) = 104.17$ ,  $p < 0.001$ . Only two of the independent variables (the Commitment and Achievement scales) contributed significantly to the prediction of job performance ratings ( $sr^2 = 0.02$ ). The eight scales in combination contributed another 0.14 shared variability. Altogether, 16% of the variability in job performance ratings was accounted for by the TLQ scores.

**Table 6.** Regression of TLQ scale sten scores on job performance ratings (n = 4,487)

Variables	Combined Assessment (DV)	Empathy	Motivation	Performance	Communication	Empowerment	Commitment	Feedback	Achievement	B	Beta	sr <sup>2</sup> (unique)
Empathy	0.24									-0.022	-0.06	0.00
Motivation	0.32	0.57								0.007	0.02	0.00
Performance	0.27	0.53	0.65							0.002	0.01	0.00
Communication	0.29	0.55	0.69	0.71						0.005	0.02	0.00
Empowerment	0.23	0.59	0.57	0.58	0.56					-0.001	0.00	0.00
Commitment	0.35	0.65	0.65	0.58	0.60	0.53				0.048	0.16**	0.01
Feedback	0.32	0.65	0.73	0.66	0.67	0.56	0.66			0.028	0.11	0.00
Achievement	0.36	0.53	0.72	0.55	0.61	0.54	0.67	0.63		0.055	0.19**	0.01
Intercept =										2.65		
Mean	6.64	31.79	32.11	28.95	29.36	28.25	34.10	32.64	31.87	R <sup>2</sup> = 0.16		
Standard deviation	1.23	3.48	4.29	4.21	4.48	3.95	4.14	4.61	4.26	Adjusted R <sup>2</sup> = 0.16		
										R = 0.40		

\*\* P<0.01.

<sup>a</sup>Unique variability = 0.02, shared variability = 0.14.

## Norms

Norms for the TLQ were collected from people taking tests on the MSP website. A very large sample of 5,000 respondents between the ages of 18 and 65 with equal numbers of men and women was created from the database of completed tests. The age, racial and country characteristics of the sample are shown in Tables 7-9. Norms are contained in Tables 10-11.

The key characteristics of the international norms' sample are as follows.

- The mean age of respondents was 37.9 with a standard deviation of 11.0. The majority of respondents were between the ages of 21 and 50 with the largest number in the 31-40 age band. There were small numbers in the sample under 20 and over 60.
- Approximately two thirds of respondents described themselves as White, 10% said they were Asian, 9% said they were Black, and 4% of a mixed background.
- Just over one half of respondents were from the United States. The next largest country groups were the United Kingdom (15%), Australia & New Zealand (7%), Canada (6%), and India (3%).

**Table 7.** Age and gender characteristics of TLQ norms (n = 5,000)

Gender	18-20	21-30	31-40	41-50	51-60	60-65	Total
Women	102	629	742	652	332	43	2500
	4.1%	25.2%	29.7%	26.1%	13.3%	1.7%	100.0%
Men	113	635	710	686	332	24	2500
	4.5%	25.4%	28.4%	27.4%	13.3%	1.0%	100.0%
Total	215	1264	1452	1338	664	67	5000
	4.3%	25.3%	29.0%	26.8%	13.3%	1.3%	100.0%

**Table 8.** Racial characteristics of TLQ norms (n = 5,000)

Race	Female	Male	Total
White	1625	1608	3233
	65.0%	64.3%	64.7%
Asian	205	312	517
	8.2%	12.5%	10.3%
Black	262	176	438
	10.5%	7.0%	8.8%
Mixed	100	101	201
	4.0%	4.0%	4.0%
Spanish/Hispanic/Latino	83	63	146
	3.3%	2.5%	2.9%
Chinese	39	48	87
	1.6%	1.9%	1.7%
Other	186	192	378
	7.4%	7.7%	7.6%
Total	2500	2500	5000
	100.0%	100.0%	100.0%

**Table 9.** Country composition of TLQ international sample (n = 5,000)

Country	Female	Male	Total
United States	1509	1211	2720
	60.4%	48.4%	54.4%
United Kingdom	348	387	735
	13.9%	15.5%	14.7%
Australia & New Zealand	160	210	370
	6.4%	8.4%	7.4%
Canada	131	148	279
	5.2%	5.9%	5.6%
India	46	96	142
	1.8%	3.8%	2.8%
Other	306	448	754
	12.2%	17.9%	15.1%
All countries	2500	2500	5000
	100.0%	100.0%	100.0%

**Table 10.** TLQ general population norms (n = 5,000)

Scale	Sten										Scale	Mean	SD
	1	2	3	4	5	6	7	8	9	10			
Empathy	0-23	24	25-27	28-29	30-31	32-33	34	35	36-37	38-40	Empathy	31.69	3.57
Motivation	0-21	22-23	24-27	28-29	30-31	32-33	34-35	36-37	38-39	40	Motivation	32.00	4.39
Performance	0-20	21-22	23	24-26	27-28	29-30	31-32	33-34	35-36	37-40	Performance	28.86	4.22
Communication	0-20	21-22	23-24	25-26	27-28	29-31	32-33	34-35	36-37	38-40	Communication	29.25	4.51
Empowerment	0-19	20-21	22-23	24-25	26-27	28-29	30-31	32-33	34-35	36-40	Empowerment	28.13	3.97
Commitment	0-23	24-25	26-29	30-32	33-34	35-36	37	38	39	40	Commitment	33.98	4.24
Feedback	0-21	22-23	24-27	28-30	31-32	33-34	35-36	37-38	39	40	Feedback	32.52	4.69
Achievement	0-22	23	24-26	27-29	30-31	32-33	34-35	36-37	38	39-40	Achievement	31.74	4.35
Transformational leadership	0-188	189-198	199-219	220-236	237-250	251-263	264-275	276-286	287-295	296+	Transformational leadership	248.16	27.83

**Table 11.** Percentiles for TLQ scales (n= 5,000)

Raw Score	Empathy	Motivation	Performance	Communication	Empowerment	Commitment	Feedback	Achievement
8								
9								
10								
11								
12								
13								
14								
15								
16								
17								
18		1	1	1	1			
19		1	1	1	1		1	1
20		1	2	2	3	1	1	1
21	1	2	4	4	5	1	2	1
22	1	2	6	6	8	1	3	2
23	2	4	10	9	12	2	4	3
24	5	7	16	15	19	4	7	7
25	7	9	23	21	26	6	9	10
26	10	12	29	28	34	7	12	13
27	12	15	37	36	43	9	15	17
28	17	19	47	45	53	11	19	22
29	23	25	56	54	62	14	25	28
30	31	32	65	61	71	18	30	35
31	42	39	73	69	80	23	37	44
32	54	51	80	76	86	30	45	53
33	67	62	86	82	91	38	54	63
34	79	71	90	87	95	47	63	71
35	88	78	94	91	97	58	71	79
36	95	86	97	94	99	69	79	87
37	98	91	98	96		79	85	92
38	99	95	99	98		88	91	96
39		97		99		95	95	99
40		99				99	99	

## References

Bartram, D. (2002). *EFPA Review Model for the Description and Evaluation of Psychological Tests: Notes for Reviewers*. [www.efpa.be](http://www.efpa.be): European Federation of Psychologists' Associations.

Robertson, I. T., *Personality and Work Behaviour (1997)*. Keynote Address to 2nd Australian Industrial and Organizational Psychology Conference.

© 2004-2010, myskillsprofile  
[www.myskillsprofile.com](http://www.myskillsprofile.com)