

Measurement equivalence of written C-Tests and Multiple-Choice C-Tests

Hella Klemmert
 Research and Development Section of the Psychological Service
 German Federal Employment Agency, Nürnberg

Hella.Klemmert@arbeitsagentur.de

Background

The Psychological Service of the German Federal Employment Agency (Bundesagentur für Arbeit, BA), uses a variety of psychological tests, among them a written C-Test for adults who speak German as a second language. The test results are used to support decisions about the client's aptitude for particular vocational training.

This paper-pencil C-Test was transformed into a Multiple Choice (MC) C-Test. For routine use in a testing institution, MC C-Tests provide economical advantages. Evaluation and scoring of test results can be fully automated. By modifying the C-Test it has been possible to integrate it into the computer based testing system (DELTA) of the BA.

Research questions:

1. Does the MC C-Test measure the same construct as the written C-Test?
2. Is the difference between the two measures related to personality traits?
3. Is the overall magnitude of differences acceptable for the intended test use?

Methodology

In an experimental study, n = 1061 customers of the Psychological Service of the German Federal Employment Agency completed an established German C-Test and a computerized MC C-Test in randomized order. The results presented here are based on the group that completed at first the MC C-Test and later in the morning the written C-Test (Form B). Each C-Test consists of 6 texts with a total of 120 blanks.

Example screenshot:

Instruction of the computer based MC C-Test



Sample (n = 548):

- 49% male, 51% female
- 79 countries of origin, predominantly (60%) Kazakhstan, Russia, Turkey, Ukraine and Poland
- Age 29.1 ± 7.8 years
- Living in Germany since 8.2 ± 6.3 years
- German language course attended: 73%
- Mean course length 650 h

Results

Homogeneity within the scope of classical test theory (CTT)

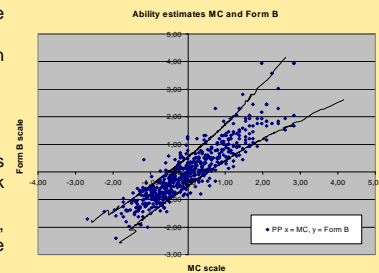
- high correlation between raw scores of the MC C-Test and Form B (r = 0.90)
- factor structure of all 12 texts is clearly onedimensional in exploratory factor analysis (75 % explained variance)
- high part-whole correlations of all texts (> 0.73)
- Cronbach alpha > 0.90
- confirmatory factor analysis indicates that a two factor solution with separate correlated factors for the MC-Version and the written Form B fits statistically significant better than a one factor solution (Chi-Square of difference = 302.80 - 135.55 = 167.25, df = 1, p < 0.001)

Homogeneity within the scope of item response theory

- Estimation of language proficiency according to the partial credit model of Masters (1982)
- difference between the proficiency estimates within chance level (alpha = 5%), if

$$|t| = \frac{|\hat{\theta}_i - \hat{\theta}_j|}{\sqrt{sem_i^2 + sem_j^2}} < 2$$

- the ability graph to the right shows differences within chance level as points between the black control lines
- differences are not completely caused by chance, because about 15% of the points lie beyond the control lines, but departures are not large



Relation between personality traits and differences in proficiency estimates

- no bias with respect to sex, age or language proficiency
- small advantage with the MC-Version for individuals with higher nonverbal reasoning abilities
- possibly, reasoning abilities are helpful in constructing correct solutions on a micro-level, without full text comprehension.

Correlation between personality traits and ability differences (Form B - MC)

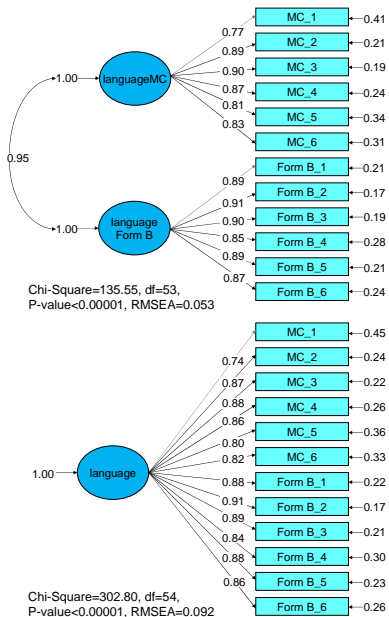
Personality trait:	r	p	N
Language proficiency	.005	> .10	548
Reasoning (non verbal tests)	-.129**	< .01	525
Sex	-.001	> .10	548
Age	.072(*)	< .10	548

Diagnostic judgment

- Form B test results are interpreted by utilizing cutoff values. The cutoff values discriminate between „insufficient“, „sufficient“ and „good“ language proficiency, for vocational trainings with „low“, „medium“ and „high“ linguistic demands respectively.
- With the method of equipercntile equating, raw scores of the new MC C-Test were transformed in comparable Form B raw scores. Thereby the cutoffs can still be used to interpret the (transformed) MC test results.
- Agreement between classification due to Form B and MC ist as high as between two parallel tests with reliability = 0.90.

Disagreement between MC-Version (row) and Form B (column)			Expected Disagreement between two parallel tests with reliability = 0.90			
	0	1	2	0	1	2
Vocational training with low demands on german language proficiency:						
0	4.9%			0	6.8%	
1	7.3%	4.7%		1	6.8%	4.3%
2		4.2%		2		4.3%
Vocational training with medium demands on german language proficiency:						
0	7.3%			0	6.0%	
1	6.0%	2.9%		1	6.0%	2.4%
2		1.3%		2		2.4%
Vocational training with high demands on german language proficiency:						
0	4.0%			0	4.3%	
1	3.3%	1.5%		1	4.3%	1.1%
2		0.9%		2		1.1%

0 = „insufficient“
 1 = „sufficient“
 2 = „good“



Conclusions

MC C-Tests are a viable alternative, if for technical or economical reasons the usual written C-Test format seems not suitable. Analyses according to different test models show that written C-Test and MC C-Test measure very similar but not identical constructs.

Differences are small and not related to sex, age or German language proficiency. To a small degree, higher nonverbal reasoning abilities are related to better results in the MC C-Test. Further research is needed to explore the involved cognitive process.

In the intended diagnostic context, disagreement between the two test formats is not larger than expected between two parallel tests with reliability = 0.90. This size of disagreement can be considered as unavoidable.