

VCA CPG Assessment Exercises

RELIABILITY ANALYSIS

ELEANOR RUTHERFORD, C.PSYCHOL.

Table of Contents

Introduction	2
Types of Reliability	2
Test-retest reliability	2
Inter-rater reliability.....	2
Internal reliability	2
Level of Reliability.....	3
Results	3
Conclusion and Recommendation	4
References.....	5

Introduction

Data has been collected on the performance 66 candidates for crew manager and 49 candidates at station manager across several fire services in England. In line with Standards for The Design and Delivery of Assessment (British Psychological Society, Division of Occupational Psychology) reliability analysis has been conducted on the following exercises in order to determine their level of accuracy:-

- Crew Manager Case Study-sample size=48
- Crew Manager Role Play-sample size=66
- Station Manager Case Study-sample size=49
- Station Manager Role Play-sample size=42

Types of Reliability

There are 3 main types of reliability used within psychometric analysis. These are outlined below for reference, but it should be noted that not all of these are appropriate within the quality assurance process for assessment centre exercises as standards for these are separate from psychometric instruments such as intelligence tests and personality questionnaires.

Test-retest reliability

This is used within psychometric tests which measure stable characteristics which are not expected to change over time, for example personality. Test-retest reliability is not appropriate for characteristics which may be open to change due to practice effects or change due to personal development by the individual. Furthermore, under normal circumstances assessment centre exercises are not generally used twice on the same cohort or where some but not all, of the cohort may have previously completed the exercise. This ensures fairness for all candidate but prevents the collection of data for test-retest reliability. For all these reasons assessment centre exercises are not evaluated for test-retest reliability.

Inter-rater reliability

This is used where two or more raters are used to assess a trait, behaviour or skill. Since the current exercises are assessed using one assessor this analysis is not possible.

Internal reliability

Internal reliability is used to assess the internal consistency of a scale. It is widely used within psychometrics and research. It allows the consistency of behavioural indicators to be

measured and ensures that levels of random error within a measurement are low. It should be noted that reliability is a property of the tests scores rather than the measurement itself (Pedhazur and Schmelkin, 1991). There are multiple measures of internal reliability including split half, Kuder Richardson and Cronbach's Alpha. Cronbach's alphas is one of the most popular measurements used as it is robust and easily applied to this type of data.

Level of Reliability

Cronbach's alpha can produce a coefficient anywhere between -1 and 1 (Streiner, 2003), though it would be extremely unusual to see results below zero. Bartram (1995) provides the following guide for reliability results:-

correlation	Verbal label
.85	Excellent
.8	Good
.7	Reasonable
.6	Adequate
Under .6	Inadequate

Results

Exercise	Result Cronbach's Alpha Coefficient	Level (as per Bartram, 1995)
Crew Manager Case Study	.886	Excellent
Crew Manager Role Play	.866	Excellent
Station Manager Case Study	.933	Excellent
Station Manager Role Play	.932	Excellent

The results demonstrate that each of the exercises currently demonstrates good internal reliability and can be considered accurate. Item analysis was also carried out on all the behavioural indicators within the exercise by checking mean scores for all indicators to see if any behavioural indicator scores were significantly above or below the general trend.

The following was found:-

Station manager case study-behavioural indicator service delivery 2 was a slightly lower mean than other indicators (0.9 versus a general range from 1.5 to 1.8).

All other behavioural indicators on all other exercises were as expected.

A further check was then carried out using the inter-item correlations to ensure that no indicator had consistent correlations below the .3 level.

Station manager case study SD2-correlations. 6 out of 11 correlations were below .3 level.

All other behavioural indicators on all other exercises were as expected.

Lastly a check was carried out on the item-total statistics to ensure that no behavioural indicator was weakening the overall score for Cronbach's alpha.

Station Manager case study SD2-this was the only behavioural indicator where Cronbach's alpha would be improved very slightly by removing the indicator (without this alpha would be .942)

All other behavioural indicators on all other exercises were as expected.

Conclusion and Recommendation

The exercises are currently demonstrating an excellent level of internal reliability indicating accuracy as per standards required by the British Psychological Society (BPS), Division of Occupational Psychology (DOP).

Checks were also carried out on each behavioural indicator within the exercises and these were found to be satisfactory with no requirement to remove or change any of the indicators at the current time for both the crew manager exercises and the station manager role play. These checks essentially ensure that all behavioural indicators within the assessors marking sheets are contributing positively to the evaluation of the candidate. The results for the station manager case study were as expected for 11 of the competencies with only one competency being borderline. This suggests that this behavioural indicator is producing less positive evidence, however the differences are slight and there are numerous reasons why this could be happening. This issue has been discussed with the exercise designer and changes have been applied to this behavioural indicator to see if this will bring the indicator in line with the others in the exercise. The exercises will be subject to ongoing monitoring to see if the changes made are appropriate.

In line with BPS DOP standards for long term evaluation, data should continue to be collected and reviewed as a minimum on an annual basis. Where possible performance data should also be requested on candidates to allow long term validation work to be carried out to supplement the reliability studies.

References

Bartram, D. (1995) *Review of Personality Assessment Instruments (level B)*. British Psychological Society

British Psychological Society, Division of Occupational Psychology (2015) *The Design and Delivery of Assessment Centres: A Standard produced by the British Psychological Society's Division of Occupational Psychology*. British Psychological Society

Pedhazur, E.K., & Schmelkin, L.P. (1991) *Measurement Design and Analysis: An Integrated Approach* Lawrence Erlbaum Associates Ltd.

Streiner, D.L., (2003) Starting at the Beginning: An Introduction to Coefficient Alpha and Internal Consistency. *Journal of Personality Assessment*. 80 (1) 99-103