The Test of Everyday Attention for Occupational Assessment

Reliability, Validity and Fairness Summary

TalentLens

ALWAYS LEARNING

PEARSON

TEA-Occ Reliability, Validity and Fairness Summary

Reliability

The TEA-Occ standardisation for Rail Safety Standards Board (RSSB)

The TEA-Occ was completed by 134 candidates attending train driver assessment centres. Cronbach's Alpha was calculated to provide a measure of internal consistency for Lift Counting with Distraction, and a value of 0.84 was obtained. The TEA-Occ subtest Lift Counting with Distraction can therefore be considered to possess good internal consistency reliability. Due to the format of the Telephone Search and Telephone Search While Counting subtests, it is not possible to calculate internal consistency reliability estimates for these tests

Test of Everyday Attention (TEA) original manual

In the original TEA manual test-retest reliability was assessed using alternate forms of the test. Three versions of the test were developed, labelled Version A, Version B and Version C. The TEA-Occ was developed from Version A.

Table 1 shows test-retest reliability coefficients, taken from the original standardisation of the TEA. Table 1 shows the reliability coefficients for the one-week test-retest on versions A and B for 118 participants from the standardisation sample, and for 74 participants from a sample of stroke patients included in the original development and standardisation of the TEA. Results from the TEA subtests that were subsequently developed for the TEA-Occ are reported. Test-retest reliability figures are also given for a subsample of the standardisation sample who were given Version C of the test a further week after receiving Version B; these correlations are between versions B and C.

The reliability of the original TEA is good for almost all subtests for both the control participants and stroke patients (see Table 1). The one exception is the dual-task decrement. The decreased reliability of this task in comparison with the others may be due to the large learning effects from one version to the other. It may be that participants differ in their ability to automate tasks, meaning that not all benefit equally from the experience of first taking the test which results in the lower test-retest reliability observed here.

Table 1: Test-retest reliability from the original TEA standardisation

	Pearson correlations		
	Controls (n=118) A with B	Controls (n=39) B with C	Stroke patients (n=74) A with B
Lift Counting with Distraction	0.71	0.68	0.83
Telephone Search – raw score	0.86	0.90	0.78
Telephone Search While Counting – dual task decrement	0.59	0.61	0.41

It has been assumed for the TEA-Occ standardisation that, due to the high similarity between the subtest administration, scoring and normative data of the TEA and the TEA-Occ, the test-retest properties of the TEA would apply to the TEA-Occ subtests.

Standard errors of measurement

Based on the test-retest reliability evidence available for the TEA, Table 2 shows the standard errors of measurement (SEMs) for each of the three subtests. Adding and subtracting the SEM from an observed score gives a range that is typically referred to as the 'confidence band' or 'confidence interval'. These figures are based on the test-retest between versions A and B of the TEA. The SEMs allow the accuracy of measurement, as evidenced through the test's reliability, to be taken into account when interpreting scores. SEMs are given for both raw and T scores, at 68% and 95% levels of confidence.

Table 2: Standard errors of measurement (SEM) for the TEA-Occ

			Raw score SEM		T score SEM	
	Raw score mean	Raw score SD	68%	95%	68%	95%
Lift Counting with Distraction	8.89	2.01	1.08	2.16	5.39	10.77
Telephone Search – raw score	2.80	1.13	0.42	0.85	3.74	7.48
Telephone Search While Counting – dual task decrement	1.53	2.56	1.64	3.28	6.40	12.81

Validity

Face validity

The subtests in the TEA-Occ make use of everyday materials and contexts, meaning they are realistic and should be subject areas that respondents can readily relate to. The original authors cite this as one of the 'strengths' of the TEA. However, to ensure face validity it is important that test administrators clearly explain the purpose of the test, so allowing respondents to make the connection between the constructs being assessed by the test and the competencies required of a train driver.

Content validity

Development of the TEA was grounded in research on attention and established tasks which had been empirically shown to be sensitive to individual differences in components of attention (e.g. Wilkins *et al.*, 1987). The grounding of the subtests in applied problems that place demand on the attentional system for their successful completion, supports the content validity of the TEA-Occ.

Construct validity

This form of validity was used during the development of the TEA when examining the differential performance between groups (e.g. stroke patients and a control group).

Evidence of construct validity for the TEA-Occ comes primarily from the pilot studies where it has been used as part of a larger assessment of train drivers. Project T948 (RSSB, 2013) reported the correlation between the Lift Counting with Distraction and Telephone Search While Counting (dual task decrement) subtests to be -0.392 (n=177). This provides evidence of internal construct validity, showing that the association of the scores obtained on these two aspects of the TEA-Occ is modest and supporting the argument that these subtests are assessing distinct aspects of the attentional system. Squaring the correlation coefficient gives the degree of shared variance or 'overlap' between the measures. This value is just over 15 percent, meaning that performance on each of these subtests is relatively independent of performance on the other.

The association between the TEA-Occ and a range of other psychometric assessments used as part of train driver selection is also reported as part of Project T948 (RSSB, 2013). These associations are summarised in Table 3. Detailed descriptions of the tests included in this analysis are given in the report for Project T948.

Table 3: Associations between the TEA-Occ and other psychometric tests used as part of the train driver selection process

	Group Bourdon[1]			DTG[2]		
TEA-Occ subtest	Production total	Omissions total	Faults total	Part 3 good	Part 3 wrong	Self paced wrong
Lift counting with distraction	0.98 (N=80)	-0.141 (N=80)	-0.086 (N=80)	0.072 (N=91)	-0.142 (N=21)	-0.145 (N=69)
Telephone Search While Counting	-0.28 (N=80)	0.047 (N=80)	0.040 (N=80)	-0.330 (N=91)	0.264 (N=21)	0.404 (N=69)

	TAVTMB[3]		TRP[4]	
TEA-Occ subtest	Overview	Overview	Part 1	Part 2
Lift counting with distraction	0.256 (N=142)	0.256 (N=142)	0.094 (N=93)	0.152 (N=92)
Telephone Search While Counting	-0.287 (N=142)	-0.287 (N=142)	-0.143 (N=93)	-0.337 (N=92)

[1]The Group Bourdon is a paper-based psychometric test distributed by Southeastern. It is designed to measure attention. [2]The Determinations Gerat (DGT) test is designed to assess the operation of hand and foot controls for train driver selection. [3]The Tachistoscopic Traffic test (TAVTMB) is part of the computerised Vienna Test System battery. It assesses visual perception and perceptive speed in traffic situations. [4]The Trainability for Rules & Procedures Test (TRP) is a paper-based psychometric test developed by OPC Assessment Ltd. It is designed to measure trainability, memory and reasoning.

The original TEA manual (Robertson *et al.*, 1994) explored the association between the subtests included in TEA-Occ and the National Adult Reading Test (NART), a measure of verbal intelligence, in the standardisation sample. The results of this analysis are shown in Table 4, and indicate that verbal intelligence has very little effect on performance on any of the subtests.

Table 4: Associations between TEA-Occ subtests and the NART

TEA subtest	Partial correlation with NART (age partialled out)
Lift Counting with Distraction	0.20
Telephone Search (time per target)	-0.20
Telephone Search While Counting (dual task decrement)	-0.03

The discriminant validity of the TEA was examined using two groups during its development: stroke patients and people with closed head injuries. In both cases the results from these groups were compared with controls. The results from these studies are summarised in Tables 5 and 6. T-tests (t) were used to establish the statistical significance of any differences and the probability (p) of these results are reported.

Table 5: Performance on the TEA for stroke patients versus controls in two age groups

	Age 50-64				Age 65-80			
	Means (SDs)				Means (SDs)			
	Control (n=26)	Stroke (n=39)	t	p	Control (n=65)	Stroke (n=41)	t	p
Lift Counting with Distraction	8.18 (2.80)	5.65 (3.20)	3.45	<0.001	7.75 (2.90)	4.69 (3.50)	-4.68	<0.001
Telephone Search – time per target	3.39 (0.70)	5.43 (2.60)	4.60	<0.001	4.22 (1.50)	7.38 (5.9)	3.37	<0.001
Telephone Search While Counting (dual task decrement)	2.03 (3.40)	3.77 (9.50)	ns	ns	2.28 (2.80)	10.50 (14.40)	3.57	<0.001

Table 6: Performance on the TEA for people with closed head injuries and controls

	Control (n=15)	Head injured (n=15)	t	p
Lift Counting with Distraction	7.87	7.47	-0.46	ns
Telephone Search – time per target	2.99	5.50	5.05	0.0001
Telephone Search While Counting (dual task decrement)	1.83	4.84	1.94	0.038

Table 5 shows that there were clear and statistically significant differences between stroke patients and controls in almost all cases, with stroke patients showing reduced attentional functioning. A similar pattern of results was seen for the patients with head injuries, despite the limited numbers available for this study (n=15 per group).

Criterion validity

T948 (RSSB, 2013) reports concurrent criterion validity evidence for the TEA-Occ against the selection criterion of 'attention'. As the report describes "The selection criterion for attention is split into two sub-criteria: selective attention and divided attention. Selective attention is defined as the ability to differentiate between different sources of information and attend selectively to them. Divided attention is defined as the ability to switch attention between different sources of information" (RSSB, 2013, p 170-171). The results of the criterion validity study conducted as part of project T948 are shown in Table 7.

Table 7: Criterion validity of the TEA-Occ against train driver selection criterion of attention

	Criterion: Selective attention	
	Average validity	Largest observed correlation
Telephone Search	No significant correlations	No significant correlations
Lift Counting with Distraction	0.20 (performance) 0.37 (training)	0.69 (performance)
	Criterion: Divided attention	
	Average validity	Largest observed correlation
Telephone Search While Counting - time per target score	0.20 (performance)	0.25 (performance)
Telephone Search While Counting – dual task decrement	0.22 (performance)	0.32 (performance)

As part of establishing the validity of the new set of assessments proposed for train driver selection, a range of performance data was collected on participants. Ratings were made by driver managers on a number of indicators of operational driving performance, using a scale ranging from 1 (very poor) to 5 (excellent). The key criteria in the validation of the TEA-Occ were 'Train driving procedure-based work' and 'Preparation and disposal of trains', selected on the basis of being the two most relevant for the TEA-Occ measures and the ones that showed associations in the predicted direction. The association of the TEA-Occ with these criteria is shown in Table 8, indicating the suitability of the TEA-Occ for assessment of these criteria.

Table 8: Criterion validity of the TEA-Occ against 'Train driving procedure-based work' and 'Preparation and disposal of trains'

TEA-Occ criterion validity	Operational performance data	
	Train driving procedure-based work (N=65)	Preparation and disposal of trains (N=82)
Lift Counting with Distraction – total number of correctly counted strings	0.37**	0.34**
Telephone Search While Counting – correctly counted strings of tones	0.25*	0.20*
Telephone Search While Counting - Dual task decrement	-	-0.20*

*p<0.5, **p<0.001

Fairness

Information relevant to the fairness of the TEA-Occ is available from the original TEA manual, and the RSSB T628 and T948 projects. T948 also examined the effect of pass rates according to group membership, if the recommended pass scores were applied. This data was evaluated against the 'four-fifths rule'. The four-fifths rule compares the proportion of the majority group that is successful at any stage of a selection procedure to the proportion of the minority group that is successful. If the proportion of the minority group that is successful is less than four-fifths of the majority group, then adverse impact is said to be occurring. The four-fifths rule is not used in UK law, but is rather a guide to evaluating fairness.

Age: The original TEA manual (Robertson *et al.*, 1994) reported that older people tended to do better on the Telephone Search While Counting subtest. Decrement in performance was seen in people aged over 50, and particularly in those aged over 65. In the project T628 trials of the TEA-Occ (RSSB, 2010) it was found "there are no obvious differences on any test in the 21 - 50 age range but over 50s do less well on all the sub-tests" (p 35). However, project T948 concluded that the evidence for age effects on Lift Counting with Distraction was inconclusive (RSSB, 2013).

When the pass scores were applied to the T948 project sample, 96 percent of those aged 50 and under were seen to pass Lift Counting with Distraction compared to only 76 percent of those aged 51 or older. The pass rates for Telephone Search While Counting (dual task decrement) were 92 percent for those aged 50 and under and 81 percent for those aged 51 or older. RSSB (2013) reports that if these figures remain stable, it would mean that Lift Counting with Distraction would fail to meet the four-fifths rule test for potential adverse impact, though Telephone Search While Counting would comply with this guide.

Gender: None of the research on the TEA or the TEA-Occ has found evidence of gender difference on any subtest. When the pass scores were applied to the T948 project sample, 93 percent of males and 100 percent of females passed Lift Counting with Distraction, with the pass rates for Telephone Search While Counting (dual task decrement) being 91 percent and 100 percent. RSSB (2013) reports that these figures mean both subtests comply with the four-fifths rule. However, it should be noted that the number of females in this sample was very small (n=5), which means these results should be treated with caution.

Ethnic group: Although project T628 (RSSB, 2010) reported no evidence of ethnic group effects on subtest performance, the application of the recommended pass scores was seen to result in differences in project T948 (RSSB, 2013). The pass rates for Lift Counting with Distraction were 94 percent for 'Whites' and 67 percent for 'Others' ('Non-whites'). For Telephone Search While Counting (dual task decrement), the pass rates were 92 percent ('Whites') and 67 percent ('Others'). These figures showed that both subtests failed to meet the four-fifths rule, suggesting that adverse impact may be occurring. However, interpretation of these results is hampered by the small number of people from 'non-white' backgrounds (n=6) and treating all 'non-whites' as a homogenous group for the purposes of analysis.

Overall, current evidence suggests that the TEA-Occ subtests do not have adverse effects according to group membership. However, as the project T948 report notes "Every effort was made to collect information from females, older candidates and ethnic groups. However, people from these demographic groups are so poorly represented in the train driver population that it was only possible to obtain a very small sample" (RSSB, 2013, p213). Due to the small sample sizes, especially for females and non-whites, analysis of fairness needs to continue when larger sample sizes are available following implementation of the revised psychometric assessment process.

References

Robertson, I. H., Ward, T., Ridgeway, V. and Nimmo-Smith, I. (1994). *The Test of Everday Attention.* London: Pearson.

RSSB (2010). *T628 Driver Selection: Development Phase – Updated selection criteria and validation study.* London: RSSB.

RSSB (2013). *T948 Driver selection: Implementation phase – Final report and technical annexes*. London: RSSB.

Contact Us

Customer Services

For all registration and order enquiries please contact our Customer Services team. Details can be found online at www.talentlens.com