Journal of the International Neuropsychological Society (2011), **17**, 562–563.

doi:10.1017/S1355617711000361

LETTER TO THE EDITOR

Neuropsychologists Must Keep Their Eyes on the Reliability of Difference Measures

Bruno Kopp

Cognitive Neurology, University of Technology Carolo-Wilhelmina Braunschweig, Braunschweig, Germany, Department of Neurology, Braunschweig Hospital, Braunschweig, Germany

(Received February 14, 2011; Final Revision February 18, 2011; Accepted February 22, 2011)

Sánchez-Cubillo et al. (2009) conducted a study in which 41 healthy older subjects performed a battery of neuropsychological tests, including the Trail Making Test (TMT), the Digit Symbol subtest (WAIS-III), the Digits Forward and Backward subtests (WAIS-III), a Finger Tapping Test, a Stroop Test, and a task-switching paradigm akin to the Wisconsin Card Sorting Test (cf. Strauss, Sherman, and Spreen, 2006). The results of correlation and regression analyses suggested that TMT_A requires mainly visuo-perceptual abilities, TMT_B primarily reflects working memory and task-switching abilities, while the TMT_{B-A} difference score provides a relatively pure indicator of task-switching abilities. The use of the TMT_{B-A} difference score should help clinicians to interpret abnormal performance in terms of a failure of this specific cognitive mechanism.

Unfortunately, there is the danger that the reliability of difference scores will be unacceptably low because the reliability of a difference score is simply a function of the average reliability of its two components and of the correlation between them (Crawford, Sutherland, and Garthwaite, 2008). Under the circumstance of a common standard deviation of the two components used to form the difference, the formula for the reliability of a difference score, $r(B-A)$, is:

$$r(B - A) = \frac{\frac{r(AA) + r(BB)}{2} - r(BA)}{1 - r(BA)} \quad (1),$$

where $r(AA)$ and $r(BB)$ are the reliabilities of the two components, and $r(BA)$ is the correlation between them (Crawford et al., 2008). Thus, if difference scores compare measures of two related constructs, the correlation between the components will be substantial, and it may eventually approach the reliabilities of the components in its magnitude. Given this situation, the variance of the difference score will predominantly be measurement error variance, simply because the numerator of Equation 1 will approach zero.

The available data point in this direction. Reynolds' (2002) estimates of internal consistencies and correlations that were obtained in the normative sample of the Comprehensive Trail Making Test (CTMT) are presented in Table 1. Applying Reynolds' (2002) data to Eq. 1, a reliability of the difference score TMT_{B-A} of $r(B-A) = 0.32$ is predicted, which is unacceptably low for any clinical purpose. Crawford et al. (2008) obtained even lower estimates for the reliabilities of TMT difference scores from the Delis-Kaplan Executive Functioning System (D-KEFS; Delis, Kaplan, and Kramer, 2001). Specifically, the reliability estimates of the difference score D-KEFS-$TMT_{Number–Letter–Switching}$ (TMT_B analogue) minus D-KEFS-$TMT_{Number–Sequencing}$ (TMT_A analogue) equaled in three different age groups .10, −.06, and −.08, respectively (Crawford et al., 2008).

Adequate reliability is fundamental whenever the cognitive status of an individual is assessed. The advocated TMT_{B-A}

Table 1. Estimates of CTMT internal consistencies and correlations in the normative sample ($N = 1.664$; Reynolds, 2002)

	Internal consistencies	Correlations (with Trail 2)
Trail 1 (equals TMT_A)	.74	.70
Trail 2 (equals TMT_A)	.77	—
Trail 3 (equals TMT_A)	.72	.70
Trail 4 [a]	.70	.61
Trail 5 (equals TMT_B)	.70	.59[b]

Note. TMT_A is defined here as a Number-Sequencing task. TMT_B is defined as a Number-Letter-Switching task. Trail 1 essentially mimics the TMT_A, whereas Trails 2 and 3 are similar to the TMT_A but introduce distractor items. The inclusion of distractor items is deemed to be of no relevance at this point. The internal consistency estimates of the three TMT_A analogues thus average to .74.

[a]Trail 4 of the CTMT does not fit into the TMT_A/TMT_B dichotomy.

[b]In the sample of Sánchez-Cubillo et al. (2009), the correlation between the TMT_A and the TMT_B scores amounted to .73.

Correspondence and reprint requests to: Bruno Kopp, Cognitive Neurology, University of Technology Carolo-Wilhelmina Braunschweig, and Department of Neurology, Braunschweig Hospital, Salzdahlumer Str. 90, 38126 Braunschweig, Germany. E-mail: b.kopp@klinikum-braunschweig.de

difference score should be considered un-interpretable in those contexts, due to its expectably unacceptable reliability, and, despite its superior construct validity. We deliberately designed the Brunswick Trail Making Test (BTMT) to maximize internal consistency (above the level of .90), by adjusting test length (Kopp, Rösser, and Wessel, 2008). To conclude, neuropsychologists should be very reluctant to use difference scores which compare measures of two related constructs, due to the potential trade-off between gains in validity and losses in reliability. Needless to emphasize that this argument applies to all domains of neuropsychological assessment.

REFERENCES

Crawford, J.R., Sutherland, D., & Garthwaite, P.H. (2008). On the reliability and standard errors of measurement of contrast measures from the D-KEFS. *Journal of the International Neuropsychological Society*, *14*, 1069–1073.

Delis, D.C., Kaplan, E., & Kramer, J.H. (2001). *Examiner's manual for the Delis-Kaplan executive function system*. San Antonio, TX: The Psychological Corp.

Kopp, B., Rösser, N., & Wessel, K. (2008). Psychometric characteristics and practice effects of the Brunswick Trail Making Test. *Perceptual and Motor Skills*, *107*, 707–733.

Reynolds, C.R. (2002). *Comprehensive Trail-Making Test (CTMT): examiner's manual*. Austin, TX: PRO-ED.

Sánchez-Cubillo, I., Periáñez, J.A., Adrover-Roig, D., Rodríguez-Sánchez, J.M., Ríos-Lago, M., Tirapu, J., & Barceló, F. (2009). Construct validity of the Trail Making Test: Role of task-switching, working memory, inhibition/interference control, and visuomotor abilities. *Journal of the International Neuropsychological Society*, *15*, 438–450.

Strauss, E., Sherman, E., & Spreen, O. (2006). *A compendium of neuropsychological tests*. Oxford: Oxford University Press.

https://doi.org/10.1017/S1355617711000361 Published online by Cambridge University Press