ELSEVIER

Acta Psychologica 117 (2004) 79–94

acta psychologica

www.elsevier.com/locate/actpsy

Gender differences on the mental rotations test: a factor analysis ☆

Daniel Voyer *, Kristin A. Saunders

Department of Psychology, University of New Brunswick, Bag Service #45444, Fredericton, Canada E3B 6E4

Received 1 December 2003; received in revised form 29 March 2004; accepted 12 May 2004

Abstract

The purpose of the present study was to examine possible gender differences in strategy when completing the mental rotations test. Two experiments examined gender differences and the factor structure on outcomes that can be obtained on this test. Experiment 1 involved large groups testing and Experiment 2 used small groups. Factor analytic results in both experiments generally supported the notion that items with one wrong and one blank response or one correct and one blank reflect reluctance to guess, whereas one correct and one wrong or two wrong answers reflect propensity to guess. Even though the factor structure was the same in males and females, the data provided mitigated support for the hypothesis that males have a higher propensity to guess and females show a greater reluctance to guess. Findings are discussed in terms of their implications for the interpretation of gender differences on the MRT.

PsycINFO classification: 2340; 2220
Keywords: Human sex differences; Factor analysis; Cognitive ability

1. Introduction

A number of meta-analyses conducted within the last 20 years clearly support the existence of gender differences in favor of males in spatial abilities (Hedges & Nowell, 1995; Linn & Petersen, 1985; Voyer, Voyer, & Bryden, 1995). In both the Linn and

☆ The work presented here was supported by a grant from the Natural Sciences and Engineering Research Council of Canada to D. Voyer.
* Corresponding author. Tel.: +1-506-453-4974; fax: +1-506-447-3063.
E-mail address: voyer@unb.ca (D. Voyer).

0001-6918/$ - see front matter
doi:10.1016/j.actpsy.2004.05.003

Petersen (1985) and Voyer et al. (1995) samples of studies, the mental rotations test (MRT: Vandenberg & Kuse, 1978), a paper and pencil test of three-dimensional mental rotation, produced the largest magnitude of gender differences, especially when a stringent approach to scoring was used (Voyer et al., 1995). This test has also showed stable gender differences across time, in contradiction with the claims that the magnitude of these differences has declined in recent years (Feingold, 1988).

The MRT was designed by Vandenberg and Kuse (1978) as a paper-and-pencil version of the Shepard and Metzler (1971) three-dimensional mental rotation task. It consists of multiple-choice items composed of a target figure, two correct alternatives, and two distractors. Correct alternatives are always identical to the target, but are shown in various rotations. Participants are required to identify the two correct alternatives. One might believe that a task that requires two responses for the same target figure is somewhat crude compared to the classic Shepard and Metzler (1971) task in which participants make a same/different judgment for pairs of rotated stimuli. Specifically, the Shepard and Metzler task allows an estimation of response time and accuracy as a function of angle of rotation, whereas the MRT only provides a measure of overall accuracy unrelated to the angle of rotation of the stimulus pairs. However, in what appears to be the first study examining the relation between performance in a classic mental rotation (MR) task and performance on the MRT, work currently in progress in our laboratory suggests that these two tasks are closely related. So far, 77 female and 63 male undergraduate students have been tested on a classic computerized MR task using the Shepard and Metzler drawings as well as on the MRT and on the primary mental abilities spatial relations subtest (PMA-SR: Thurstone, 1958), a paper-and-pencil test of two-dimensional MR. As expected, significant gender differences emerge on all measures (all p's <0.01). However, the critical finding is that performance on the MRT is significantly correlated with overall performance on the computerized MR task ($r = 0.65$, $p < 0.01$) and with the PMA-SR ($r = 0.62$). These values compare favorably with those reported by Vandenberg and Kuse (1978) when they examined the correlation between the MRT and other paper-and-pencil spatial measures. This suggests that the MRT measures similar processes to those assessed by the classic MR task. However, Voyer et al. (1995) reported that the MRT produces much larger gender differences than the classic MR task (which they called "generic mental rotation"). This fact alone suggests that the classic task and the MRT differ on critical components; at least as far as gender differences are concerned.

One such difference actually refers to the need to produce two responses per item on the MRT. Specifically, Vandenberg and Kuse (1978) suggested that counting only items where both choices are correct in computing the test score provided a correction for guessing. Thus, other possible combinations of responses are assumed to reflect guessing to some extent. This notion was taken further by Voyer (1997) who examined more closely the meaning of possible outcomes on that test. The fact that each item on the MRT requires two answers creates a number of possible outcomes. Participants can produce two correct (CC), one correct and one wrong (CW), two wrong (WW), one correct and one blank (CB), one wrong and one blank (WB), or two blank (BB) answers on any given item. From this perspective, Voyer