Intelligence 38 (2010) 574–581

Contents lists available at ScienceDirect

Intelligence

Evaluating the contribution of different item features to the effect size of the gender difference in three-dimensional mental rotation using automatic item generation ☆

Martin E. Arendasy *, Markus Sommer

Psychological Methods and Computer-Based Modeling Group, Dept. of Psychology, University of Graz, Universitätsplatz 2, A-8010 Graz, Austria

ARTICLE INFO

Article history:
Received 21 May 2009
Received in revised form 8 May 2010
Accepted 6 June 2010
Available online 21 August 2010

Keywords:
Mental rotation
Spatial ability
Gender differences
Automated item generation

ABSTRACT

In complex three-dimensional mental rotation tasks males have been reported to score up to one standard deviation higher than females. However, this effect size estimate could be compromised by the presence of gender bias at the item level, which calls the validity of purely quantitative performance comparisons into question. We hypothesized that the effect of gender bias at the level of distinct item design features could lead to either an over- or underestimation of reported effect sizes of the gender difference in three-dimensional mental rotation. Using automatic item generation we conducted a series of psychometric experiments in which we independently manipulated one out of four different item design features that have exhibited a gender bias in the previous studies (study 1). This was done in a between-subjects design. The results indicated that gender bias caused by item design features linked to the perceptual stadium of mental rotation led to an overestimation of the effect size of the gender difference while item design features associated with the encoding and transformational stadium resulted in an underestimation of the effect size of the gender difference. In study 2 we tested the hypothesis that the gender difference still remains while controlling for the item design features causing gender bias. The results suggest that a significant portion of the gender difference may be attributable to perceptual and encoding processes involved in mental rotation.

1. Introduction

Mental rotation tasks have shown to exhibit a large gender difference in meta-analytic studies (cf. Linn & Petersen, 1985; Voyer et al., 1995; Masters & Sanders, 1993). The reported male superiority in mental rotation tasks turned out to be stable across cohorts (cf. Masters & Sanders, 1993; Voyer et al., 1995), age (cf. Geiser et al., 2008; Linn & Petersen, 1985), language and cultures (cf. Silverman et al., 2007) and depended on task characteristics. For instance, the effect size was larger for complex three-dimensional mental rotation tasks (d = .70 to d = .94) than for two-dimensional mental rotation tasks (d = .31 to d = .44) and chronometric mental rotation tasks (d = 0.37). The most pronounced gender difference in favor of males has been reported for the Vandenberg and Kruse test (1978), which require respondents to construct an accurate mental representation of the object and maintain this representation in the course of a mental transformation process. In this particular test males have been reported to score up to one standard deviation higher than females.

However, the estimates of the effect size could be compromised by the presence of gender bias at the item level (differential item functioning: Holland & Wainer, 1993). If a gender bias exists for a particular mental

☆ We wish to thank our reviewers Wendy Johnson, Richard Lynn and Paul Irwing for their help in improving this paper.

* Corresponding author.
E-mail address: martin.arendasy@uni-graz.at (M.E. Arendasy).

0160-2896/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2010.06.004

rotation test, the test may not measure the same latent trait for females and males. This calls the validity of a purely quantitative comparison of the performance of females and males into question. In order to ensure that quantitative comparisons are valid, one needs to be able to measure the performance of females and males on the same scale with a common origin and unit of measurement. This psychometric precondition is referred to as *full score equivalence* (van de Vijver & Poortinga, 2005). Evaluation of the full score equivalence of three-dimensional mental rotation tests for females and males provided mixed results. While some studies indicate that full score equivalence can be assumed (cf. Arendasy, 2000, 2005; Arendasy & Sommer, in press), others failed to investigate whether their test exhibits a gender bias (cf. Bejar, 1990; Rettig et al., 1989) or found evidence that females are systematically disadvantaged by certain item sets (cf. Geiser et al., 2006; Hosenfeld et al., 1997). In the latter case the findings indicated that the gender bias can be attributed to certain item design features that affected the cognitive processes of females and males in a different manner. Moreover, Karmer and Smith (2001) argued that full score equivalence may not suffice on its own. One also needs to demonstrate that the performance of females and males is equally affected by item features that are assumed to exert an influence on the cognitive processes used to work the test items. This additional precondition has been referred to as *construct representation equivalence* (Embretson, 1998; Karmer & Smith, 2001).

1.1. Using automatic item generation to ensure full score- and construct representation equivalence

One means to ensure full score equivalence and construct representation equivalence for females and males is to use automatic item generation (AIG: Irvine & Kyllonen, 2002; Arendasy & Sommer, in press). Within this framework the construction of an item generator starts with a definition of the latent trait and the cognitive processes that characterize it. This process is referred to as the construction of a *cognitive model* (Embretson, 1998, 2005; Gorin, 2006). The psychometrician then chooses an item format to measure the latent trait. Afterwards, the cognitive model is condensed into a more specific *cognitive item model* by outlining the item features that are assumed to trigger the cognitive processes used to solve the test items (e.g. Arendasy & Sommer, in press; Arendasy et al., 2008; Embretson, 1998, 2005; Gorin, 2006). These item features are commonly referred to as *radicals* (Irvine, 2002). The main goal in defining and systematically varying radicals is to *maximize* the *construct-related variance* in the item parameters.

Arendasy & Sommer (2005, 2007, in press) recently argued that the specification of radicals is not sufficient to ensure full score equivalence and construct representation equivalence. One also needs to define a set of item features which must be omitted in the item construction process to *minimize interfering variance* arising from non-construct-related cognitive processes. These item features have been referred to as functional constraints (Greeno et al., 1993).

1.2. Procedural framework for the automatic generation of three-dimensional mental rotation tasks

In this section we outline the different stages in the automatic construction of gender-fair three-dimensional mental rotation tasks.

Definition of the latent trait:
Lohman (1994) defined three-dimensional mental rotation as the ability to construct an accurate representation of the object that has to be transformed and maintain this representation throughout the entire transformation process.

Choice of the item format:
Based on prior studies we decided to use automatically generated endless loops as a general item format (cf. Arendasy, 2000; Gittler & Arendasy, 2003). In general, endless loop items depict two objects referred to as reference object and comparison object (cf. Fig. 1).
Each task consisted of a reference object (left figure) and a comparison object (right figure). The respondents had to indicate the direction in which the reference object needs to be rotated to obtain the comparison object. The answer alternatives were: rotate to the right, rotate to the left, rotate up and rotate down.

Specification of the cognitive model:
In order to outline the cognitive processes involved in solving three-dimensional mental rotation tasks we utilized a modified version of the model proposed by Just and Carpenter (1985). The cognitive model proposes five processing stadium, which mutually interact with each other: (1) perception stadium, (2) search stadium, (3) encoding stadium, (4) mental transformation stadium and (5) conformational stadium.
The first phase is referred to as *perception stadium*. It is concerned with the extraction of the three-dimensional

Fig. 1. Example of an automatically generated endless loop item with the reference figure on the left and the comparison figure on the right.