Learning and Individual Differences 29 (2014) 8–17

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

On gender differences in mental rotation processing speed

Rudolf Debelak [a,*], Georg Gittler [b], Martin Arendasy [c]

[a] Department of Psychology, SCHUHFRIED GmbH, Mödling, Austria
[b] Department of Applied Psychology, Health Development, Enhancement and Intervention, University of Vienna, Austria
[c] Psychological Methods and Computer-based Modeling Group, University of Graz, Austria

ARTICLE INFO

Article history:
Received 23 February 2013
Received in revised form 8 September 2013
Accepted 11 October 2013

Keywords:
Mental rotation ability
Mental rotation speed
IRT model
Gender differences

ABSTRACT

There is a wide consensus in the literature that gender differences can be observed in tasks measuring mental rotation ability. A possible explanation of this finding is the presence of gender differences in the processing speed of mental rotation tasks. In two studies, we investigated the dimensionality and the presence of gender differences in mental rotation processing speed in two mental rotation tasks. By applying a joint modeling approach for responses and response times, we found that, in both tasks, mental rotation ability and mental rotation processing speed can be regarded as unidimensional constructs. We replicated previous findings that gender differences in mental rotation ability can be observed in both tasks, although we could not find gender differences in mental rotation processing speed. Our results thus indicate that the observed gender differences in mental rotation ability cannot be explained by gender differences in mental rotation processing speed.

1. Introduction

Spatial abilities constitute an important component in current models of human intelligence (cf. Carroll, 1993; Johnson & Bouchard, 2005; McGrew, 2005). Studies on gender differences in human intelligence indicate that many spatial ability measures exhibit considerable gender differences in favor of male subjects. Meta-analytic studies show that gender difference is particularly pronounced in case of three-dimensional mental rotation tasks (e.g., Linn & Petersen, 1985; Voyer, Voyer, & Bryden, 1995). Gender differences have been generally found to be smaller in other spatial ability tasks (e.g. spatial orientation tasks; cf. Coluccia & Louse, 2004). Although there is evidence that the observed gender difference in favor of male subjects is stable across cohorts (cf. Masters & Sanders, 1993; Voyer et al., 1995), age (cf. Linn & Petersen, 1985), and culture (Silverman, Choi, & Peters, 2007), there is evidence that the magnitude of the gender difference varies with item design characteristics (cf. Arendasy & Sommer, 2010; Voyer & Doyle, 2010) and general design characteristics such as time limit. Further variables which influence the size of the gender differences include the response format (Glück & Fabrizii, 2010; Titze, Heil, & Jansen, 2010) and the experience with similar mental rotation tasks (e.g. Quaiser-Pohl, Geiser, & Lehmann, 2006). Numerous explanations have been proposed to account for the observed male superiority in three-dimensional mental rotation performance (for an overview: Halpern, 2011). While some studies emphasize the role of biological causes for the observed gender differences (e.g. Jaušovec & Jaušovec, 2012; Peters, 2005), several studies investigated differences in the strategies female and male respondents use when working on three-dimensional mental rotation tasks (e.g. Arendasy, Sommer, & Gittler, 2010; Geiser, Lehmann, Corth, & Eid, 2008). The influence of these reported differences in test taking strategy on the psychometric characteristics of mental rotation tests is still a topic of current research. At least for some mental rotation tasks, unidimensionality and measurement invariance across both gender groups has been demonstrated (Arendasy & Sommer, 2010; Bors & Vigneau, 2011; Gittler & Arendasy, 2003). Some models attributed gender differences in mental rotation tasks to gender differences in speed-accuracy tradeoff. This explanation is based on findings, which indicate that gender differences in favor of male subjects decrease in effect size once time limits had been removed from the test (cf. Goldstein, Haldane, & Mitchell, 1990). Although this finding has not usually been consistently replicated (e.g., Masters, 1998), a recent meta-analysis conducted by Voyer (2011) indicate that gender differences in paper–pencil mental rotation tasks indeed decrease in size when the psychometric measures were administered without time limits. This finding could be due to at least two different reasons: (1) the removal of time limits may allow female respondents that are not well trained in this population to utilize effective mental rotation strategies (cf. Arendasy, Sommer, Hergovich, & Feldhammer, 2011; Arendasy et al., 2010), or (2) the observed reduction of the gender difference in the untimed administration condition could be due to a ceiling effect in the male population (Voyer, 2011), meaning that male respondents are not able to further improve an already excellent test performance when time limits are removed. Because none of these previous studies assessed mental rotation processing speed, it is hard to differentiate between these two explanations. With the model by Goldstein et al. (1990) as basis, one would expect that gender differences in mental rotation processing speed are either more pronounced or of the same magnitude than gender differences in mental rotation accuracy. By contrast, if the

* Corresponding author at: Department of Psychology, SCHUHFRIED GmbH, Hyrtlstraße 45, 2340 Mödling, Austria.

1041-6080/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.lindif.2013.10.003

reduced effect size of the gender difference in three-dimensional mental rotation performance is mainly attributable to a ceiling effect in the male population in case of untimed mental rotation tasks, one would expect either no or small effect sizes of the gender differences in mental rotation processing speed. Under this model, observed gender differences in mental rotation accuracy should be large in magnitude compared with the processing speed measure.

Some authors have already reported gender differences in mental processing speed based on results found in elementary cognitive tasks (for a review, see Roivainen, 2011). However, processing speed is generally considered to be a complex, multi-dimensional construct (e.g. Danthiir, Wilhelm, & Roberts, 2012; Danthiir, Wilhelm, Schulze, & Roberts, 2005; Roberts & Stankov, 1999), so it remains unclear whether these findings can be generalized to the processing speed in chronometric or psychometric mental rotation tasks.

1.1. Formulation of the problem

In this article, we want to evaluate these two conflicting hypotheses using an item response theory model that enables the simultaneous estimation of accuracy and processing speed parameters (Klein Entink, Fox & van der Linden, 2009). Another advantage of this psychometric approach is the possibility of simultaneously evaluating the dimensionality of accuracy and processing speed measures of mental rotation performance, which have been debated in the literature for some time because of the possibility of solving mental rotation tasks using different solution strategies (cf. Arendasy et al., 2010; Geiser et al., 2008). We investigated this problem in two separate studies, which used different computerized mental rotation tasks. There were two reasons for selecting these tasks: First, past studies have provided evidence that both tasks show favorable psychometric characteristics, like unidimensionality (e.g. Arendasy & Sommer, 2010; Gittler & Fischer, 2011). Second, no ceiling effects have been reported for both tasks under untimed conditions, which allowed us to test the hypothesis of Voyer (2011) that ceiling effects in male respondents cause the reduction of gender differences in three-dimensional mental rotation tasks.

2. Method

2.1. A multivariate multilevel approach for modeling speed and ability

In the literature, multiple approaches for modeling speed have been described (for an overview of early approaches, see van der Linden & Hambleton, 1997). This study chose an approach that has been originally proposed by Klein Entink, Fox, et al. (2009). This approach estimates parameters on three separate levels. We will provide a rough overview of each model in the following sections. More detailed introductions can be found in the literature (e.g. Fox, 2010; Klein Entink, Fox, et al., 2009).

On the first level of this approach, two separate models for responses and response times are defined. In our study, the model for responses is the one-parameter normal ogive model, which defines a person parameter θ_i, which marks the ability of person i to answer items correctly. The model further defines one item parameter b_k for each item k, which defines the respective item's difficulty. This model is closely related to the Rasch model (Rasch, 1960; cf. Embretson & Reise, 2000). In the one-parameter normal ogive model, the probability that person i answers item k correctly is given by the following:

$$P = (+|\theta_i, b_k) = \Phi(\theta_i - b_k). \quad (1)$$

In this formula, $\Phi()$ denotes the cumulative distribution function of the standard normal distribution, i.e. the probability that a standard normally distributed variable is smaller than the value given by $\theta_i - b_k$.

The response times are described by the two-parameter log-normal model. In this model, two item parameters are defined for each item that describes the respective item's time intensity and time discrimination. An item's time intensity denotes the average amount of time which is generally needed to answer this item, while its time discrimination denotes to which extent the observed average response time of an item changes between fast and slow responders. For each person i, a speed parameter is defined.

In the two-parameter log-normal model, the log response time T_{ik} of a person i working on item k is given by the following:

$$T_{ik} = -\phi_k\zeta_i + \lambda_k + \varepsilon_{ik}. \quad (2)$$

In formula (2), ζ_i denotes the respondent's speed, λ_k denotes an item's time intensity, and ϕ_k is an item's time discrimination. It follows from the negative sign of the $\phi_k\zeta_i$ term that a higher speed parameter leads to smaller response times if an item's time intensity remains constant. ε_{ik} is a residual term, which is assumed to be normally distributed with an item-specific variance.

On the second level, the approach of Klein Entink, Fox, et al. (2009) defines additional models to investigate the relationship between the person and item parameters of the first level models. In these models, the variances and covariances of the item and person parameters are estimated.

The third level allows investigating the influence of person covariates on the observed person parameters. In our study, the original model for the responses and response times was further expanded to contain gender as a distinct person covariate G_i (which took the value 0 for the male population and 1 for the female population) and used to measure the influence of gender on speed and ability by a linear regression model:

$$\theta_i = \gamma_{00} + G_i \times \gamma_{01} + e_{oi} \quad (3)$$

$$\zeta_i = \gamma_{10} + G_i \times \gamma_{11} + e_{1i}. \quad (4)$$

In this model, e is a residual term, which is assumed to be normally distributed. The γ terms are regression coefficients which are to be estimated.

2.2. Model selection and estimation

Under the presented Bayesian framework, several criteria have been proposed for model selection, one of them being the Deviance Information Criterion (DIC) (Spiegelhalter, Best, Carlin, & van der Linde, 2002; see also Fox, 2010; Gelman, Carlin, Stern, & Rubin, 2004). Model selection based on this criterion tends to prefer less complex models which show a good fit to the data. It has been already used in a number of studies for model selection (e.g., Goldhammer & Klein Entink, 2011). A detailed discussion of the DIC has been provided by Fox (2010), Gelman et al. (2004), and others.

In our study, we estimated all model parameters using a Gibbs sampling approach, which has been implemented in the software package cirt (Klein Entink, 2011) for the statistical software R (R Development Core Team, 2011). This approach is based on the principal idea of simulating the multivariate posterior distribution of all model parameters. The distribution of values drawn from the Gibbs sampler converges to the posterior distribution; therefore, convergence has to be tested. Values, which were drawn before convergence was reached, are denoted as burn-in phase and usually not used for further analysis. Based on the drawn values, the mean of the posterior distribution (i.e. the expected a posteriori value, EAP) and the highest posterior density (HPD) intervals can be calculated. HPD intervals are the smallest intervals that contain a given percentage (e.g., 95%) of the values of the posterior distribution and can be used to test the statistical significance of model parameters.