Intelligence 56 (2016) 72–81

Contents lists available at ScienceDirect

Intelligence

Task difficulty prediction of figural analogies

Diego Blum [a,b,*], Heinz Holling [a], Maria Silvia Galibert [b], Boris Forthmann [a]

[a] Westfälische Wilhelms-Universität Münster. Fliednerstraße 21 -, 48149 Münster, Germany
[b] Universidad de Buenos Aires. Av. Independencia 3065 - C1225AAM CABA, Argentina

ARTICLE INFO

Article history:
Received 27 April 2015
Received in revised form 4 March 2016
Accepted 5 March 2016
Available online xxxx

Keywords:
Figural analogies
Rules
Item difficulty
Rasch model
LLTM

ABSTRACT

The purpose of this psychometric study is to explain performance on cognitive tasks pertaining Analogical Reasoning that were taken into consideration during the construction of a Test of Figural Analogies. For this purpose, a general Linear Logistic Test Model (LLTM) was mainly used for data analysis. A 30-itemed Test of Figural Analogies was administered to a sample of 422 students from Argentina, and eight of these items were administered along with a Matrices Test to 84 participants mostly from Germany. Women represented 77% and 76% of each respective sample. Indicators of validity and reliability show acceptable results. Item difficulties can be predicted by a set of nine Cognitive Operations to a satisfactory extent, as the Pearson correlation between the Rasch model and the LLTM item difficulty parameters $r = .89$, the mean prediction error is slightly different between the two models, and there is an overall effect of the number of combined rules on item difficulty ($F_{(3,23)} = 15.16, p < .001$) with an effect size $\eta^2 = .66$ (large effect). Results suggest that almost all rotation rules are highly influential on item difficulty.

1. Introduction

Figural matrices are often applied to the measurement of General Intelligence components (Freund, Hofer, & Holling, 2008). Among these components, Analogical Reasoning is of particular importance for many psychometricians, since Analogical Reasoning tests have been historically labeled as good measures of *g* (Cattell, 1971; Spearman, 1904; Sternberg, 1977; Wolf Nelson & Gillespie, 1991). Examples of tests that contain figural matrices and/or are said to measure Analogical Reasoning by using figural stimuli are abundant in the psychometric literature (e.g., Brown, Sherbenou, & Johnsen, 2010; Cattell & Cattell, 1960; Freund et al., 2008; Raven, Raven, & Court, 1998; Wechsler, 1997).

The Analogical Reasoning process consists basically of making inferences about unknown entities by using previous knowledge recollected from other similar and well-known entities. The new domain or least-known case is commonly referred to as the *target analog*, whereas the source domain or better-known case is labeled as the *source analog* (Gentner, 1983; Gick & Holyoak, 1980; Holyoak & Koh, 1987). Generally, each domain comprises relations between elements, and analogies are actually comparisons between these relations (Sternberg, 1977). This is also known as the A:B::C:D analogies (A is to B as C is to D). When a problem is based on finding the missing element D of the analogy (i.e., A:B::C:?), then C:D becomes the target analog and A:B becomes the source analog. What needs to be extrapolated from one domain to the other is the compound of structural relations that binds these two entities, and not just superficial data (Gentner, 1983).

The basic problem A:B::C:? can be applied to different types of contents, namely: verbal, pictorial and figural (Wolf Nelson & Gillespie, 1991). Figural analogy tests have been profusely studied in the psychometric field. In fact, many figural matrix tests are said to measure Analogical Reasoning, such as Raven's Progressive Matrices (Raven, Raven, & Court, 1998). It has also become increasingly important to study the relation between these measures and the underlying cognitive components that trigger them. In this line of thinking, a set of Cognitive Operations should be matched with the item properties during item construction, since by doing so, item parameters can be predicted by a cognitive theory independently from their calibration (Embretson, 1999; Irvine, 2002). This also helps to demonstrate the construct validity of items (Baghaei & Kubinger, 2015).

It is possible to model item difficulty as a function of the difficulty of Cognitive Operations by means of the Linear Logistic Test Model (LLTM. Fischer, 1973), which is an extension of the Rasch (1980) model. In an LLTM, item difficulty parameters β are dependent on a linear combination of basic parameters α, which are in turn the difficulty estimates of the Cognitive Operations. In this sense, if two Cognitive Operations

---

[*] Corresponding author.
E-mail addresses: blumworx@gmail.com (D. Blum), holling@uni-muenster.de (H. Holling), galibert@psi.uba.ar (M.S. Galibert), boris.forthmann@uni-muenster.de (B. Forthmann).

affect item 1 and only one of them affects item 2, then the difference in terms of difficulty between these two items is given by the additional basic parameter (α_2) that is present in just one of them (Fischer, 1995).

The ordinary Rasch model for dichotomous data can be expressed in the following way:

$$P_i(X_{vi}=1)=\frac{\exp(\theta_v-\beta_i)}{1+\exp(\theta_v-\beta_i)}=\mathrm{Expit}(\theta-\beta_i)$$

where the probability P_i that the v'th person solves the i'th item correctly is given by the Expit function applied to the difference between the ability of that person (θ_v) and the difficulty of that item (β_i). Besides, the LLTM splits item difficulty into the following linear combination (Kubinger, 2008):

$$\beta_i=\sum_{j=1}^{p}\omega_{ij}\alpha_j$$

where the i'th item difficulty parameter (β_i) depends on the sum of products between the j'th basic parameter (α_j) and its weight on the i'th item (ω_{ij}) (Scheiblechner, 1972). This means that the difficulty parameter of an item depends on the sum of the difficulties of those Cognitive Operations that are present in the item to a certain extent (that is, the sum of those basic parameters with respective weights $\omega_{ij}>0$, in contrast to those with weight $\omega_{ij}=0$). A simple study can be carried out if the basic parameters display only two possible weights according to the presence ($\omega_{ij}=1$) or absence ($\omega_{ij}=0$) of their respective Cognitive Operations in the task.

The study of difficulty of rule-based figural items using an LLTM approach can be found in recent research (Bertling, 2016; Freund et al., 2008; Graßhoff, Holling, & Schwabe, 2010; Zeuch, 2010; Zeuch, Holling, & Kuhn, 2011). Similar research performed before the year 2000 is not abundant (Bertling, 2016); nevertheless, interesting results can be quoted. For instance, it was demonstrated that figural items with spatial displacement transformations increase item difficulty, whereas those with distortions do exactly the opposite (Novick & Tversky, 1987; Whitely & Schneider, 1981). Mulholland, Pellegrino, and Glaser (1980) studied the causes of item difficulty in geometric analogy problems, and concluded that the number of item elements, as well as the number of transformations, had a significant effect on error rates. Tanzer, Gittler, and Ellis (1995) calibrated a spatial ability test with the LLTM and, when comparing American and Austrian test takers' responses, no cultural differences were found regarding item complexity estimated by this linear model.

Embretson (1999, 2002) regressed item difficulties of abstract items on cognitive model variables and found that the number of rules, the abstract correspondence and the structural overlay of shapes display positive and significant regression coefficients. Freund et al. (2008) tested figural matrix items that hold between two and four rules, found Rasch model fit and explained item difficulties with the LLTM. Their research revealed that rules of complete addition, addition of one element, addition of two elements, progression of position and progression of form display significant basic parameters. After perfecting the work of Beckmann (2008); Bertling (2016) performed LLTM analyses on a Figural Analogy Test and found that item difficulties can be explained either by a model which only holds spatial displacement rules as Cognitive Operations, or by a model which also includes additional rules. In parallel, the works of Graßhoff et al. (2010); Zeuch (2010), and Zeuch et al. (2011) show similar lines of research. For a methodological example on how the LLTM is modeled upon item responses in abstract reasoning tests, such as the Viennese Matrices (Formann & Piswanger, 1979), see Kubinger (2008).

Most of the studies just mentioned were made on tests containing two kinds of variations among items, namely, rule-based variations and visual complexity variations. This means that not only do items differ according to the number and type of rules applied, but they are also different with respect to the number and type of elementary shapes they comprise. Furthermore, items with larger number of rules frequently hold larger number of shapes as well. On the other hand, an ideal measurement of *g*, Analogical Reasoning or related variables should only depend on the manipulation of Cognitive Operations (i.e., the rules) and not on the alteration of visual complexity. Therefore, the presence of these item-to-item shape variations may confound the results by adding an explanation of performance that is not due to the intended measures. In fact, Mulholland et al. (1980) found significant main and interaction effects on error rates when manipulating the number of constituent elements and the number of transformations of geometric analogies, thus proving that these superficial elements may also contribute to performance. The increase of spatial relation complexity would raise the amount of information processed by the working memory, thus leading to much greater probabilities of answering incorrectly. In the present research, visual complexity variation among items is minimized, which will be described below.

Moreover, to the best of our knowledge, no research on Figural Analogical Reasoning has mentioned the use of a systematic strategy or pseudocode to build item options, which could be very useful to not bias the testee's decision making when answering a multiple-choice item, as well as to favor the correct manipulation of item difficulty. For this matter, a Solutions Combination Design, also described in a recent paper by the first author of the present manuscript (Blum, Lozzia, Abal, & Attorresi, 2015), has been utilized in this research.

Despite the LLTM potential contributions to difficulty prediction and cognitive theory, research with the LLTM has become highly infrequent (Kubinger, 2008). In the following pages, two psychometric studies of Analogical Reasoning are presented together, which include the administration of a new Test of Figural Analogies and difficulty prediction through the LLTM, among other analyses. Precedents of this test as well as proposals for theoretical and operational definitions yielding to the construction of figural analogy items can be found in the manuscripts of Blum, Abal, Lozzia, Picón Janeiro, and Attorresi (2011), Blum, Galibert, Abal, Lozzia, and Attorresi (2011), and Blum, Lozzia et al. (2015), even though they do not show an LLTM study. It is also interesting to mention that working with similar visual complexities, building item options systematically and knowing potential predictors of item difficulty based on Cognitive Operations may benefit the creation of computer algorithms to automatically generate figural items (Arendasy, 2005; Freund et al., 2008; Gierl & Lai, 2012; Lai, Alves, & Gierl, 2009).

2. Method

2.1. Participants

Two samples were studied, the first one containing 422 students. These students were undertaking the first year of Psychology at the Universidad de Buenos Aires, Argentina, in April 2014. A paper-based Test of Figural Analogies of 30 items (Figs. 1 and 2) was administered to these students during the first class of Statistics. From this sample, 97 people (23%) were men and 325 (77%) women. The mean age level was 22.75 (*SD* = 6.56), with standardized skew and kurtosis statistics (that is, these statistics divided by their correspondent standard errors) of 28.17 and 56.04 respectively.

The second study consisted of 84 participants, 70 of whom were students from the Westfälische Wilhelms-Universität Münster in Germany, another 9 were Argentineans either living in Germany or in the city of Buenos Aires, and the rest were from other countries. The study took place between September and November 2015 and it was done on-line through the platform Concerto v3.9.14 (Kosinski, Lis, Mahalingam, Sun, & Rust, 2012). Eight items of the Test of Figural Analogies were administered along with eight additional items of a Matrices Test automatically generated by means of Hofer's (2004) MatrixDeveloper, the latter being first presented by Freund et al.