Psychological Test and Assessment Modeling, Volume 56, 2014 (3), 287-303

Analysing latent constructs via a derived metric paired comparison approach: An application to students' emotions in mathematics

Alexandra Grand[1]

Abstract

In this article we suggest an approach for the analysis of sets of items using the method of paired comparisons. We applied the proposed approach to a students' survey of emotions typically experienced while learning mathematics by focusing on the *relative* dominance of these emotions. The emotions of interest were: enjoyment, pride, anger, anxiety, boredom and shame which were each measured by a set of items and for which we want to obtain an ordering on a continuum. In a first step we evaluated the quality of items by using a method of non-parametric Rasch model tests. The item sets of the emotions enjoyment, anxiety and boredom met the properties of a Rasch model. As a result of fitting Rasch models, we obtained person "emotion" parameter estimates. We then derived for each individual metric paired comparison responses from the obtained person parameter estimates and directly modelled these derived relative responses by fitting a beta regression model. This model is similar to generalized linear models (GLMs). The proposed model accounts for bounded metric paired comparison data in (0,1) where subject covariates and object-specific covariates can also be incorporated. We found that there is a tendency, the higher the positive discrepancy between the self concept of maths ability and the averaged perceived maths ability of students the more enjoyment and the less anxiety is typically experienced while learning mathematics.

Keywords: beta regression model, latent constructs, logistic Bradley-Terry model, metric paired comparisons, Rasch model

[1] *Correspondence concerning this article should be addressed to:* Mag. Alexandra Grand, Institute for Statistics and Mathematics, Vienna University of Economics and Business (WU), Welthandelsplatz 1, 1020 Vienna, Austria; e-mail: alexandra.grand@gmx.at

Introduction

In many sciences, for example, psychology, social sciences or economics (especially in the field of marketing), items with a graded response format are commonly used in self-reporting surveys. Usually, the intention is to obtain an absolute measure for a latent construct (e.g. attitudes, abilities, values, emotions, attributes, etc.) on the basis of an item set. A two-point rating scale with the response categories *disagree*, *agree* or a four-point rating scale with the categories *strongly dissatisfied*, *dissatisfied, satisfied*, *strongly satisfied*, may be examples to name only some. Usually two-point up to seven-point rating scales are used in surveys depending on the research question. To come up with a measure or measures representing the latent construct(s) one common technique is to build the sum of the scores of the items in a set. In practice it is often assumed that all items of a set refer to one latent dimension, that respondents do not differ in their interpretation of labelled response categories and that the response categories are equidistant and ordered. Concerning the latter issue, Rasch (1961) presented an unidimensional model for multicategorical data which defines the probability that person i responds to item k in category h ($h = 0, \dots, m$). He mentioned that this model can also be extended to multidimensional cases (see also e.g. Kubinger, 1989; Andersen, & Olsen, 2001). Approaches to give up the (questionable) assumptions commonly used in item response theory (IRT) are to fit a Rasch model (Rasch, 1960), in the case of dichotomous response categories, a Rating Scale model (Andrich, 1978) or a Partial Credit model (Masters, 1982) for polytomous data (see also e.g. Kubinger, 1989; van der Linden, 1997) and then test the assumptions of Rasch model fit via parametric or non-parametric methods. These are unidimensionality, local stochastic independence, monotone increasing item characteristic curves, specific objectivity (see also Scheiblechner, 2009) and sufficient statistics (for details see e.g. Fischer, 1974; Fischer & Molenaar, 1995). From such approaches we obtain reasonable person parameter estimates on an interval scale level. Another technique for the analysis of sets of responses made on a labelled κ -point response scale is to focus on the relative importance, preference etc. of items in a set. Dittrich, Francis, Hatzinger, and Katzenbeisser (2007), for example, proposed a paired comparison method where dichotomous or trichotomous paired comparison responses were derived from Likert-scale responses with the aim to obtain a relative ordering of a set of items. In paired comparison studies individuals are asked to repeatedly decide for one of two presented objects (of a set of J objects), i.e. either for object j or for object k in a given comparison (jk) over $\binom{J}{2} = \left(J \bullet \left(J - 1\right)\right)/2$ paired comparisons. There are only two possible responses (preference for object j, denoted by (jk)j or preference for object k, (jk)k) in each comparison by simply ignoring the degree of preference. Dittrich et al. (2007) derived judgements via pairwise comparison of J items of a given item set where, for example, a response score of 1 of item j only expresses a lower importance than a response score of 3 of item k, for a given person i. Equal response scores indicate no preference for one of the two items compared, i.e. a tie. For modelling paired comparison data the well-known Bradley-Terry (BT) model (Bradley & Terry, 1952) is commonly used. The BT model is defined by the probability of preferring object j to object k in the comparison of object (jk). It can be seen as a special case of the Rasch model, where not persons are compared against items but items are compared against items, and as a special case of the model of Cox (1970), for details see e.g. Kubinger (1989).

In this article we refer to sets of items or statements, each measuring a latent construct, and present a method where we first use Rasch models to obtain person parameter estimates of the latent constructs of interest and second use a BT model to obtain a relative ordering of these latent constructs. We explicitly model the *degree of relative* preference, dominance, importance, ability etc. of a set of latent constructs using the method of metric paired comparisons (see e.g. Grand & Dittrich, 2014). Let J latent constructs be the objects of interest for which we build $\binom{J}{2}$ paired comparisons. By pairwise comparing the person parameter estimates in each comparison for each individual, we obtain *derived* relative responses (which can be thought to be made) on a bounded metric paired comparison scale indicating the degree of dominance of one latent construct (i.e. object) compared to another. As these responses do not originate from real paired comparison tasks, but were afterwards constructed, we termed such responses *derived*. The metric paired comparison approach allows us to model the degree of preference in paired comparisons reported on a metric bounded response scale.

A practical example for using the proposed method would be the measurement of relative satisfaction with J different attributes (the objects) of a holiday-booking web site of a destination. It might be of interest which of the attributes is ranked highest according to satisfaction and, for example, if the relative ordering of attributes differs for females and males or for people with more or less experience in online-booking. The J latent constructs could be each assessed by a set of items where persons can state their degree of agreement on a 4-point rating scale. By applying the suggested method person satisfaction parameter estimates for each attribute were obtained. From these parameter estimates continuous paired comparison responses were derived and a metric paired comparison model fitted. As a result parameter estimates of the attributes indicating the relative amount of satisfaction were obtained. These can be located on a continuum where the distances between the attributes can be interpreted in a reasonable way, which might be a useful guidance for improvement and further development of the holiday-booking web site.

The aim of this article is to present the approach of derived metric paired comparisons where possible effects of subject covariates and/or object-specific covariates (see e.g. Dittrich, Hatzinger, & Katzenbeisser, 1998) on the relative preference ordering can be modelled.

In this study we were interested in emotions (i.e. achievement emotions) that may typically be experienced by individuals in the context of academic learning. Referring to the control-value theory, emotions can be defined as multi-component constructs, sets of interrelated psychological processes with affective, cognitive, motivational and physiological components (cf. Pekrun, 2006; Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011). Achievement emotions can be classified by the *object focus* (activity vs. outcome emotions), the *valence* (positive vs. negative emotions) and the *degree of activation* implied (activating vs. deactivating emotions), see e.g. Pekrun, Frenzel, Goetz, and Perry (2007).

Emotions are assumed to play an important role in the process of learning. Positive activating emotions like enjoyment or pride correlate positively with interest, intrinsic and extrinsic motivation and academic achievement. They are positively related to the effort spent on academic tasks, metacognitive learning strategies, critical thinking, self-regulated learning and they can facilitate creative and flexible problem solving strategies. Learning related enjoyment has a negative correlation with task-irrelevant thinking and may have a positive effect on the flow experience (cf. e.g. Pekrun, Goetz, Titz, & Perry, 2002a, 2002b). However, emotions are linked to the learning process and they can possibly be influenced by shaping the learning environments of students (cf. e.g. Pekrun et al. 2002a; Pekrun et al. 2007).

We suppose that especially learning for mathematics is charged with emotions. Each student of the WU (Vienna university of business and economics) has to pass a test in mathematics as one prerequisite to pursue her/his study. At each semester several maths courses are provided for students to prepare for the test. Students also have the opportunity to use the e-learning platform Learn@WU where all course materials are available and where sequences of the mathematics lectures can be viewed (lecturecasts). They can simulate tests, check their knowledge by solving mathematical tasks, obtain solutions and have the opportunity to discuss problems at a forum.

There may be a great diversity of emotions that can be experienced by students in the context of learning mathematics. Pekrun et al. (2002a), for example, named achievement emotions that are most often reported in qualitative studies. Following these findings, we limited the range and selected the emotions *enjoyment*, *pride*, *anger*, *anxiety*, *boredom* and *shame* for our self-reporting online survey. There are several methods to assess emotions, for example qualitative interviews, observations (e.g. facial expressions) or psychophysiological approaches (e.g. brain imaging techniques). However, in practice self-reported surveys are commonly used for assessing individual experienced emotions.

The aim of this study is to obtain a ranking of specified learning related emotions in mathematics on a continuum of relative dominance. We are further interested if this ordering is different for various groups of students. In the following section we introduce the steps and methods required for transforming graded response data into metric paired comparison data and then continue with the application section.

Transforming graded response data into metric paired comparison data

The idea is to obtain for each individual person parameter estimates of latent constructs of interest and then pairwise compare these on a metric bounded paired comparison scale. This approach can be described in three steps as follows:

1. Selecting Rasch model conform sets of items and obtaining person parameter estimates.

   First we test if the sets of items are each Rasch model conform (i.e. that the data fit the model). In the case of items with dichotomous response categories, i.e. binary observed responses, we usually fit a Rasch model and in the case of items where the responses are made on a categorical κ–point rating scale, $\kappa > 2$, we fit a Rating Scale or Partial Credit model for each item set separately. We then check for the properties of the Rasch model. Commonly used parametric methods are, for example, Andersen LR tests, Wald tests, item-fit indices, person-fit indices, graphical model checks (see e.g. Mair & Hatzinger, 2007; Kubinger, 1989). For a small sample size non-parametric goodness-of-fit tests for the Rasch model (e.g. for invariance, local dependence and/or homogeneity, discrimination) may be an appropriate alternative (see e.g. Rasch, 1960; Rasch, 1961; Ponocny, 2001; Koller & Hatzinger, 2013).

   We suggest estimating the item parameters of the Rasch-family models through a conditional maximum likelihood (CML) method and using unconditional maximum likelihood (ML) for person parameter estimation (where the item parameter estimates are assumed to be known from prior CML estimation). One problem may be that there exist no person parameter estimates for zero and full scorer, but we can approximate values by using spline interpolation, for example.

2. Deriving metric paired comparison responses from the obtained person parameter estimates of the latent constructs of interest (i.e. the objects in the paired comparisons).

   From J objects we obtain $\binom{J}{2}$ paired comparisons. For example, having three objects we build three possible paired comparisons: (12), (13) and (23), where the objects are labelled with 1, 2 and 3 for ease of notation. In each paired comparison responses can be derived on a bounded metric paired comparison scale $\left[-a, a\right]$, with a predefined known value of a. We determine the bounds $-a$, a by the lowest and the highest approximated value over all person parameter estimates, respectively. Metric paired comparison responses are *derived* by comparing the estimated person parameters of the objects of interest.

   For example: Let $\hat{\theta}_{1,i} = 1.9$ be the person parameter estimate of *object 1* and $\hat{\theta}_{2,i} = .4$ the estimate of *object 2* of individual i. The derived response $y_{12,i}^{d}$, denoted by d, in the paired comparison (12) of individual i is defined as: $y_{12,i}^{d} = \hat{\theta}_{2,i} - \hat{\theta}_{1,i} = -1.5$ (see Figure 1).

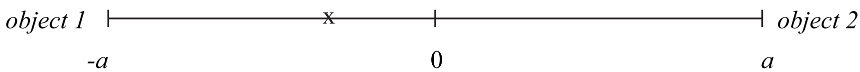

Figure 1:
Derived response in the comparison of *object 1* and *object 2* on a bounded metric paired comparison scale.

Derived responses < 0 indicate preference for the first object (j) and derived responses > 0 preference for the second object (k) in a given paired comparison (jk). A tied response (i.e. no preference) is indicated by $y_{jk,i}^{d} = 0$. We then say that *object 1* is preferred over *object 2* by 1.5 units.

3. Modelling derived metric paired comparison data.

For the analysis of metric paired comparison data we suggest using a beta regression model (Smithson & Verkuilen, 2006) for continuous response variables restricted to the interval $(0,1)$. The model structure is similar to those of generalized linear models (GLMs) and includes two linear predictors: one for the location parameter μ and the other for the precision parameter ϕ (cf. e.g. Smithson & Verkuilen, 2006; Cribari-Neto & Zeileis, 2010; Simas, Barreto-Souza, & Rocha, 2010):

$$g_1(\mu) = \boldsymbol{x}^T \boldsymbol{\lambda} = \eta_1 \quad ,$$
$$g_2(\phi) = \boldsymbol{z}^T \boldsymbol{\gamma} = \eta_2 \quad ,$$

where $g(\bullet)$ is a link function, $\boldsymbol{\lambda}$ and $\boldsymbol{\gamma}$ are vectors of unknown parameters and $\boldsymbol{x}$ and $\boldsymbol{z}$ are covariate vectors.

First we have to transform each derived random variable $Y_{jk,i}^{d}$ in $[-a,a]$ into the interval $(0,1)$. Following the proposed transformation of Smithson and Verkuilen (2006) we take two transformation steps. In a first step the derived response variable $Y_{jk,i}^{d}$, is squeezed so that we obtain a one times (indicated by *) transformed random variable $Y_{jk,i}^{d*}$ in the interval $[0,1]$ and in a second step we obtain a two times transformed variable $Y_{jk,i}^{d**}$ in $(0,1)$ by:

$$Y_{jk,i}^{d} \overset{\frac{Y_{jk,i}^{d}+a}{2a}}{\rightarrow} Y_{jk,i}^{d*} \overset{\frac{Y_{jk,i}^{d*}\bullet(n-1)+0.5}{n}}{\rightarrow} Y_{jk,i}^{d**},$$

where n is the sample size (cf. Smithson & Verkuilen, 2006; Grand & Dittrich, 2014). A tie is represented by $Y_{jk,i}^{d**}=.5$, a value close to 0 indicates the most favourable and a value close to 1 the least favourable response for object j. Let the two times transformed response $y_{jk,i}^{d**}$ be the realization of the random variable $Y_{jk,i}^{d**}$ for which we assume a beta distribution.

Second, we specify the transformed random variable in the interval (0,1) as the response variable and apply a logit-linear Bradley-Terry model (Bradley & Terry, 1952) to the logistic mean structure of a beta regression model. The beta regression model for paired comparisons for judge i for the comparison (jk) is defined by:

$$\operatorname{logit}\left(\mu_{\mathrm{jk,i}}\right)=\ln\left(\frac{\frac{\pi_j}{\pi_j+\pi_k}}{\frac{\pi_k}{\pi_j+\pi_k}}\right)=\ln\left(\frac{\pi_j}{\pi_k}\right)=\lambda_j-\lambda_k=\eta_{1jk,i}, \quad \text{where} \quad \lambda_j=\ln\pi_j. \tag{1}$$

The parameter λ_j characterizes object j. For identifiability we set λ_J to be zero. The relationship between the λ 's and the π 's, the worth parameters, is given by: $\pi_j=\frac{\exp\left(\lambda_j\right)}{\sum_{j=1}^{J}\exp\left(\lambda_j\right)}$. We labelled model (1) the BBTR model, i.e. beta Bradley-Terry regression model (see Grand & Dittrich, 2014) where we assume independence between the derived judgements of the n individuals and between the derived paired comparisons. The design structure of a BBTR model is shown in Grand and Dittrich (2014).

Application to a data set of students emotions' in mathematics

The online-survey consisted of six item sets each with four or five items. The emotions of interest were: enjoyment, pride, anger, anxiety, boredom and shame. In total students were asked to respond to 29 statements (items) on a four-point rating scale ranging from *strongly disagree* to *strongly agree*. The items were chosen from the questionnaire of Pekrun, Goetz, and Perry (2005). We slightly altered the statements where necessary and translated them into German. The subject covariates used in this study are:

- gender (*sex*): categorical subject covariate (*sex*: male, female).
- comparative ability (*cab*): numeric subject covariate derived by comparing the self concept of ability in mathematics (*ab*) and the perceived average of students' ability in maths (*pab*), each measured on a self-reported scale ranging from 0 = low abil-

ity to 100 = high ability. It can be calculated as follows: *cab* = *ab* – *pab*, where a negative sign indicates that student *i* has rated her/his ability in maths below the averaged perceived ability of other students and vice versa.

– learning resources (*lr*): originally students could respond in one of four response categories (e-learning (Learn@WU), paper-based learning, peer-learning and lecture learning) according to predominantly used learning resources. Afterwards we merged them into two response categories: e-learning resources and all other resources except of e-learning (i.e. *lr1*: e-learning, *lr2*: no e-learning) because there were no observations for lecture learning and only few students chose the category peer-learning.

Studies of Goetz, Pekrun, Zirngibl, Jullien, Kleine, vom Hofe, and Blum (2004) and Frenzel, Pekrun, and Goetz (2007), for example, showed that gender, individual achievement and the averaged class achievement have an influence on emotions experienced in mathematics. We assume that a positive discrepancy between the self concept of ability and the perceived ability of other students (the social reference frame) may have a positive effect on beliefs of personal competence or control. For example, if students perceive that they are better than their colleagues at university, i.e. show a high comparative ability, this may have a positive effect on students' control-related beliefs regarding achievement in mathematics. An activation of these beliefs will lead to viewing challenging tasks as being manageable and to positive emotions, like enjoyment (Pekrun et al., 2007). In contrast, a low level of control-related beliefs may result in negative emotions, like anxiety or anger (cf. Goetz, Cronjaeger, Frenzel, Lüdtke, & Hall, 2010).

The study was conducted at the end of the winter semester 2012/2013. All students of the WU who are possibly in the stage of learning mathematics received an e-mail, a short time before the maths test took place. They were asked to participate in an online-survey of emotions typically experienced while learning mathematics. By eliminating one response vector with a missing value we obtained a sample size of n = 111 (male = 44, female = 67).

Method

The proposed method including the main steps of transforming dichotomous responses into metric paired comparison data applied to the data set of students' emotions, is illustrated in Figure 2 and will be described in this section.

The sample size of n = 111 was relative small. We dichotomized all items where responses can be made in four response categories to two response categories: *disagree* and *agree* (see also Discussion) and fitted for each of the six item sets a Rasch model with the R-package *eRm* (Mair, Hatzinger, & Maier, 2012). The Rasch model defines the probability that person *i agrees* on item *k*:

$$P\left(X_{ik} = x_{ik} | \theta_i, \beta_k\right) = \frac{\exp\left[x_{ik}\left(\theta_i - \beta_k\right)\right]}{1 + \left(\theta_i - \beta_k\right)} \quad ,$$

where x_{ik} is the observed response (coded with 1 for *agreement* and 0 for *disagreement*) of individual *i*, $i = 1,\ldots,n$. The parameter θ_i is the person parameter or in this study the emotion parameter of person *i* and β_k, $k = 1,\ldots,K$, is the item parameter of item *k* which indicates the difficulty to *agree* on that item.

The item parameters, the β_k 's, were estimated independent from the person "emotion" parameters by conditional maximum likelihood (CML). The θ_i 's, the person "emotion" parameters were estimated using a maximum likelihood method. Note that the emotion parameters for full- and 0-scorers cannot be estimated but instead were spline interpolated to receive a value (for interpolation a minimum of five items is required).

An appropriate method to check for Rasch model conformity in a small sample is the method of non-parametric goodness-of-fit tests. With this method quasi-exact Rasch model tests can be conducted where, simply speaking, an arbitrary chosen number of random samples of data matrices, with the same item and person margins as obtained from the observed data matrix, are simulated and compared with the observed matrix. Only the elements of the data matrices are changed by keeping the margins fixed (see e.g. Verhelst, 2008; Koller, Alexandrowicz, & Hatzinger, 2012; Koller & Hatzinger, 2013). We found that the item sets of the emotions enjoyment, anxiety and boredom are

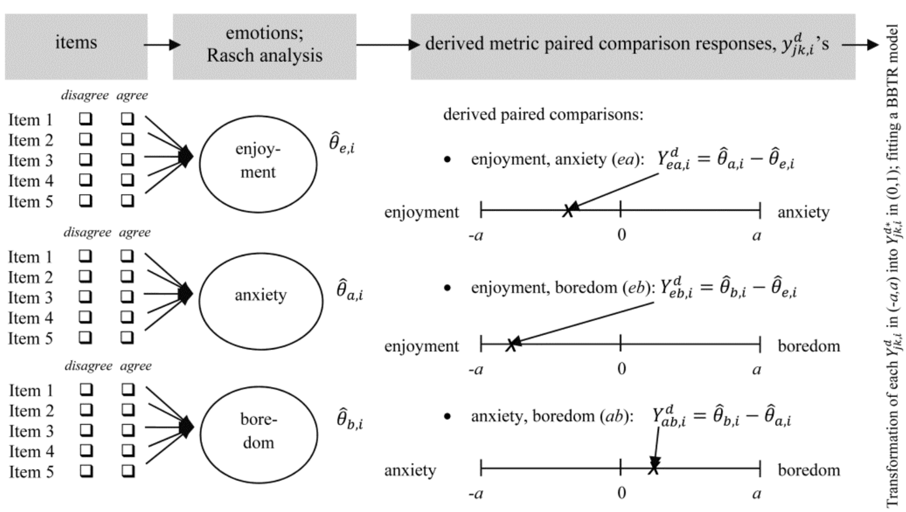

Figure 2:
Overview: transformation into derived metric paired comparison data

Rasch model conform (see Appendix) and built three pairs of emotions, i.e. enjoyment, anxiety (*ea*); enjoyment, boredom (*eb*) and anxiety, boredom (*ab*). In a next step we derived for each individual *i* responses in each paired comparison (*jk*) by comparing the estimated emotion parameter $\hat{\theta}_{j,i}$, characterizing emotion *j,* and $\hat{\theta}_{k,i}$ characterizing, emotion *k*: $y^{d}_{jk,i} = \hat{\theta}_{k,i} - \hat{\theta}_{j,i}$ (see also Figure 2). A negative sign indicated preference for the first emotion *j* and a positive sign preference for emotion *k* in a given paired comparison (*jk*). The derived responses could then be located on a bounded metric paired comparison scale. The bounds of this response scale were specified by the theoretical values –8.4 and +8.4, i.e. the highest possible minimum value $\theta_e = -4.5$ and maximum value $\theta_e = 3.9$, respectively, over all emotion parameters.

As we could not derive a response with the value –8.4 or 8.4 (because we did not compare the emotion enjoyment by itself) we transformed the random variable $Y^{d}_{jk,i}$ which takes on values in (–8.4, 8.4) into the interval $(0,1)$ as follows: $Y^{d*}_{jk,i} = \frac{Y^{d}_{jk,i} + 8.4}{2 \bullet 8.4}$. We assumed that the one time transformed random variable $Y^{d*}_{jk,i}$ is beta distributed and fitted a beta Bradley-Terry regression (BBTR) model by using R (R Development Core Team, 2013) and the packages *prefmod* (Hatzinger, 2012) for construction of a corresponding design structure and *betareg* (Zeileis, Cribari-Neto, Grün, Kosmidis, Simas, & Rocha, 2013) for estimating the parameters of interest.

For model selection of nested BBTR models we used a likelihood ratio test. In general, for all analysis in this article we committed a type-I-risk of $\alpha = .05$.

Results

We started by fitting a basic BBTR model (see Table 1, 0-model) without subject covariates to get an initial overview. The object parameter estimates of this model (listed in Table 1) can be interpreted in terms of the higher a negative value the more dominant an emotion and the higher a positive value the less dominant an emotion. To simplify interpretation we reversed (denoted by *r*) these object parameter estimates, i.e. we multiplied them by -1, e.g. $\hat{\lambda}_{enjoyment} * -1 = \hat{\lambda}^{r}_{enjoyment} = .331$. Higher values now indicate a more dominant emotion and vice versa. These reversed object parameter estimates are shown in Figure 3. We can see in Figure 3 that enjoyment was ranked first, followed by anxiety and with distance followed by boredom (enjoyment > anxiety > boredom). However, we were interested if the ordering of the three learning related emotions may be different for various groups of students, i.e. if the incorporation of subject covariates (see also e.g. Dittrich et al., 1998; Francis, Dittrich, Hatzinger, & Penn, 2002; Grand & Dittrich, 2014) would significantly improve the model fit. We started our model selection process with a full BBTR model, i.e. a three-way interaction model consisting of the categorical subject covariates gender (*sex*) and learning resources (*lr*) and the numerical subject covariate comparative ability (*cab*).

Table 1:
Estimates of nested BBTR models

estimates	**two-way interaction model (s.e.)**	**0-model (s.e.)**
enjoyment	–.204 (.201)	–.331 (.060)
anxiety	–.147 (.116)	–.222 (.060)
boredom	0 (NA)	0 (NA)
enjoyment:*lr*2	–.453 (.356)	-
anxiety:*lr*2	–.201 (.206)	-
boredom:*lr*2	0 (NA)	-
enjoyment:*cab*	.000 (.009)	-
anxiety:*cab*	.015*(.005)	-
boredom:*cab*	0 (NA)	-
enjoyment:*lr2*:*cab*	–.017* (.006)	-
anxiety:*lr2*:*cab*	.002 (.006)	-
boredom:*lr2*:*cab*	0 (NA)	-
*lr*1 (e-learning)	–.013 (.143)	-
*lr*2 (no e-learning)	.046 (.206)	-
cab	–.001 (.006)	-
precision submodel:		
phi	7.021 (.512)	5.582 (.400)
log-likelihood	121.6	83.84
number of estimated parameters	12	3

Based on a likelihood ratio test and backward elimination we selected a model with a two-way interaction between learning resources (*lr*) and comparative ability (*cab*) with a constant precision parameter (see Table 1, two-way interaction model). We again reversed the estimated parameters of the selected model. On basis of the reversed object and interaction parameter estimates we calculated the estimates of the worths (the $\hat{\pi}_j s$) by ensuring that the sum of the worths is equal to one for a given level of the subject covariate learning resources (*lr*) and a given value of the covariate comparative ability (*cab*).

Example: For illustrative purposes suppose that we are only interested in the worth of the emotion anxiety. The worth of anxiety for the group of students who uses e-learning (*lr1*) and who rates their ability in maths above the averaged perceived ability of students, i.e. *cab* = 20, can be calculated by:

$$\frac{\exp\left[\hat{\lambda}^r_{anxiety} + \hat{\beta}^r_{anxiety:cab}\, 20\right]}{\sum_{j=1}^{J} \exp\left[\hat{\lambda}^r_j + \hat{\beta}^r_{j:cab}\, 20\right]} = \frac{\exp\left[.147 - .291\right]}{3.085} = .281 \ ,$$

see also Figure 4, left plot. The parameter $\lambda_{anxiety}$ is the parameter of anxiety for the group of students who mainly uses e-learning (*lr1*), i.e. the reference group, the first level of the covariate learning resources (*lr*), and $\beta_{anxiety:cab}$ is the interaction parameter between anxiety and the covariate comparative ability (*cab*) of the e-learning group (*lr1*).

The worth of anxiety for students, with *cab* = 20, who do not mainly use e-learning resources (*lr2*) can be obtained by:

$$\frac{\exp\left[\hat{\lambda}^r_{anxiety}+\hat{\lambda}^r_{anxiety:lr2}+\left(\hat{\beta}^r_{anxiety:cab}+\hat{\beta}^r_{anxiety:lr2:cab}\right)20\right]}{\sum_{j=1}^{J}\exp\left[\hat{\lambda}^r_{j}+\hat{\lambda}^r_{j:lr2}+\left(\hat{\beta}^r_{j:cab}+\hat{\beta}^r_{j:lr2:cab}\right)20\right]}=\frac{\exp\left[.348-.334\right]}{4.695}=.216 \ ,$$

where $\lambda_{anxiety:lr2}$ is the interaction parameter between anxiety and the second level of the covariate learning resources *lr* (i.e. *lr2*, the no e-learning group). This parameter indicates the change of $\lambda_{anxiety}$ for group *lr2*. The parameter $\beta_{anxiety:lr2:cab}$ is the interaction parameter between anxiety, the second level of the covariate *lr* and the covariate *cab*. The worth parameters of the selected model for the group of students who mainly uses e-learning as resource for learning mathematics and the group who mainly uses other resources (paper-based learning, peer-learning), denoted by *no e-learning*, are shown in

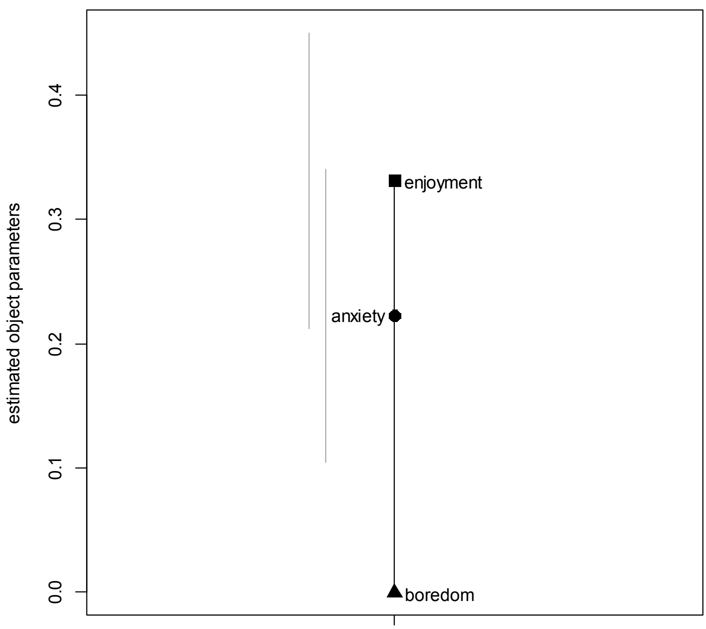

Figure 3:
Plot with a 95 % - confidence line for the estimated object parameters. Note that the object boredom has no confidence line as it is defined as the reference object (i.e. set to be zero) in this model fit

Figure 4. The x-axis of the worth plot (see Figure 4) shows the covariate comparative ability (*cab*) ranging from -70 to $+50$. A value of *cab* below zero means that the maths ability of a student is below the averaged perceived ability, zero indicates equivalence and a value above zero can be interpreted that a student rated her/his ability above the maths ability of other students. The y-axis shows the estimated worth parameters, where the higher the value the more dominant an emotion while learning mathematics. We can see in Figure 4 that for both groups the ordering of the emotions varies according to the value of the covariate comparative ability. For students who report equivalence (*cab* = 0), i.e. whose self concept of ability is conform with the perceived averaged ability of other students, enjoyment is the first ranked emotion followed by anxiety and the least experienced emotion boredom. For students with *cab* = 0 who mainly use e-learning resources the worths of the three emotions are relatively close together, which means that enjoyment, anxiety and boredom are similarly dominant while learning mathematics. Whereas for students with *cab* = 0 who do not mainly use e-learning resources the worths differ, showing gaps between the emotions. In general, there is a tendency that the higher the comparative ability the less anxiety and the more enjoyment is experienced while learning mathematics, whereas this tendency seems to be much stronger for the no e-learning group (see Figure 4). For students with low comparative ability who

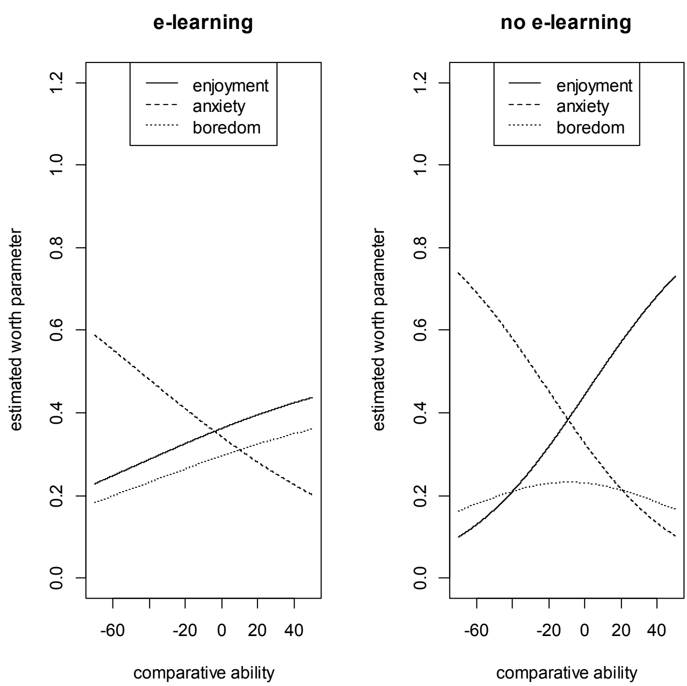

Figure 4:
Plot of worth parameters according to the covariate comparative ability for the group of students who mainly use e-learning and for those who do not

mainly use other resources than e-learning boredom is ranked on the second place whereas for students who use e-learning resources boredom is the least experienced emotion. A more detailed look at the object parameter estimates for students who rate their maths ability much lower than the averaged perceived ability of students (i.e. *cab* = -60) and who predominantly use e-learning shows that the distance between the estimated parameters of the emotions anxiety and enjoyment (i.e. $\Delta_{anxiety,enjoyment} = \left|\left(\hat{\lambda}^r_{anxiety} + \hat{\beta}^r_{anxiety:cab} * -60\right) - \left(\hat{\lambda}^r_{enjoyment} + \hat{\beta}^r_{enjoyment:cab} * -60\right)\right| = .802$) is less than for students who do not use e-learning (i.e. $\Delta_{anxiety,enjoyment} = 1.682$). For students who rated their comparative ability relatively high (e.g. *cab* = 40) the object parameter estimates of the first ranked emotion enjoyment and the second ranked emotion boredom differ far less ($\Delta_{enjoyment,boredom} = .194$) for students who mainly use e-learning resources than for students who do not ($\Delta_{enjoyment,boredom} = 1.316$).

Discussion

The presented approach is a method for analysing latent constructs. It is suitable in situations where someone is interested in explicitly modelling relative judgements and obtaining a relative preference ordering of a set of latent constructs.

This method is appropriate for data sets with no missing values which meet the requirements of a Rasch model. Metric paired comparisons can be analysed by fitting a beta regression model in a similar manner to generalized linear models, provided that a corresponding design structure has been built up. Subject covariates and object-specific covariates can be incorporated into the suggested model and model selection can easily be done through likelihood ratio tests of nested models.

The sample size of $n = 111$ was relatively small so that we decided to evaluate the quality of the items by non-parametric goodness-of-fit tests of Rasch models. We collapsed the four-point into two-point items and fitted Rasch models. Of course, the items with a dichotomous response scale still have to be checked in original by collecting a new sample to give reasonable statements about the Rasch model properties. Therefore the findings in this article should be interpreted carefully. Moreover, the excluded emotion item sets in this study should be inspected in more detail to find possible reasons for non Rasch conformity.

We found that gender (*sex*) has no significant effect on the ordering of the learning related emotions and selected a two-way interaction model with the covariates learning resources (*lr*) and comparative ability (*cab*). For students with high comparative ability who do not mainly use e-learning resources, enjoyment is, with great distance, the first ranked emotion followed by boredom. Students who mainly use other learning resources, e.g. paper-based resources, may obtain further information, explanations and more insights in the theory behind, have the opportunity to easily take notes, draw pictures etc. and actively practice learning. Students might enjoy a higher level of activity which can be set compared to a restricted interactive form of learning with compressed information provided at the e-learning platform. However, possible reasons regarding differences in students' emotions according to learning resources have to be further explored and analysed in following studies.

References

Andersen, E., & Olsen, L. (2001). The life of Georg Rasch as a mathematican and as a statistican. In A. Boomsma, M. van Duijn & T. Snijders (Eds.), *Essays on item response theroy* (pp. 3-24), New York: Springer.

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*, 561-573.

Bradley, R., & Terry, M. (1952). Rank analysis of incomplete block designs: I. The method of paired comparisons. *Biometrika, 39*, 324-345.

Cox, D.R. (1970). *The analysis of binary data*. London: Methuen.

Cribari-Neto, F., & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software, 34*, 1-24.

Dittrich, R., Hatzinger, R., & Katzenbeisser, W. (1998). Modelling the effect of subject-specific covariates in paired comparison studies with an application to university rankings. *Applied Statistics*, *47*, 511-525.

Dittrich, R., Francis, B., Hatzinger, R., & Katzenbeisser, W. (2007). A paired comparison approach for the analysis of sets of Likert-scale responses. *Statistical Modelling*, 7, 3-28.

Fischer, G. (1974). *Einführung in die Theorie psychologischer Tests*. [Introduction into theory of psychological tests]. Bern, Stuttgart, Wien: Hans Huber.

Fischer, G., & Molenaar, I. (1995). *Rasch models: Foundations, recent developments, and applications*. New York: Springer.

Francis, B., Dittrich, R., Hatzinger, R., & Penn, R. (2002). Analysing partial ranks by using smoothed paired comparison methods: an investigation of value orientation in Europe. *Applied Statistics*, 51, 319-336.

Frenzel, A., Pekrun, R., & Goetz, T. (2007). Girls and mathematics – A "hopeless" issue? A control-value approach to gender differences in emotions towards mathematics. *European Journal of Psychology of Education, 22*, 497-514.

Goetz, R., Pekrun, R., Zirngibl, A., Jullien, S., Kleine, M., vom Hofe, R., & Blum, W. (2004). Leistung und emotionales Erleben im Fach Mathematik. [Achievement and emotional experience in mathematics]. *Zeitschrift für Pädagogische Psychologie, 18*, 201-212.

Goetz, T., Cronjaeger, H., Frenzel, A., Lüdtke, O., & Hall, N. (2010). Academic self-concept and emotion relations: Domain specificity and age effects. *Contemporary Educational Psychology, 35*, 44-58.

Grand, A., & Dittrich, R. (2014). Modelling assumed metric paired comparison data – Application to learning related emotions. *Austrian Journal of Statistics*. Accepted.

Hatzinger, R. (2012). *prefmod: Utilities to Fit Paired Comparison Models for Preferences.* R package version 0.8-31, URL http://cran.R-project.org/web/packages/prefmod/index.html.

Koller, I., & Hatzinger, R. (2013). Nonparametric tests for the Rasch model: explanation, development, and application of quasi-exact tests for small samples. *Interstat* (November), 1-16.

Koller, I., Alexandrowicz, R., & Hatzinger, R. (2012). *Das Rasch Modell in der Praxis: Eine Einführung mit eRm* [The Rasch model in practical applications: An introduction with eRm]. Wien: Facultas WUV, UTB.

Kubinger, K.D. (1989). Aktueller Stand und kritische Würdigung der Probabilistischen Testtheorie [Critical evaluation of latent trait theory]. In K.D. Kubinger (Ed.), *Moderne Testtheorie – Ein Abriss samt neuesten Beiträgen* [Modern psychometrics – A brief survey with recent contributions] (pp. 19-83), Weinheim: Psychologie Verlags Union.

Mair, P., & Hatzinger, R. (2007). Extended Rasch modeling: The eRm package for the application of IRT models in R. *Journal of Statistical Software, 20*, 1-24.

Mair, P., Hatzinger, R., & Maier, M. (2014). eRm: Extended Rasch Modeling. R package version 0.15-4, URL http://cran.R-project.org/web/packages/eRm/index.html.

Masters, G. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149-174.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review,18*, 315-341.

Pekrun, R., Frenzel, A., Goetz, T., & Perry, R. (2007). The control-value theory of achievement emotions: an integrative approach to emotions in education. In P. Schutz, & R. Pekrun (eds.), *Emotion in education* (pp. 13-36). Amsterdam: Academic Press.

Pekrun, R., Goetz, T., & Perry, R. (2005). *Achievement Emotions Questionnaire (AEQ) - User's manual*.

Pekrun, R., Goetz, T., Frenzel, A., Barchfeld, P., & Perry, R. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology, 36*, 36-48.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. (2002a). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist, 37*, 91-106.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. (2002b). Positive emotions in education. In E. Frydenberg (ed.), *Beyond coping: Meeting goals, visions, and challenges* (pp. 149-173). Oxford: Oxford University Press.

Ponocny, I. (2001). Nonparametric goodness-of-fit tests for the Rasch model. *Psychometrika*, 66, 437-460.

R Development Core Team. (2013). *R: A language and environment for statistical computing* [Computer software manual]. Vienna, Austria, URL http://www.R-project.org.

Rasch, G. (1960). Probabilistc models for some intelligence and attainment tests. *Danish Institute for Educational Research, Copenhagen*.

Rasch, G. (1961). On general laws and the meaning of measurement in psychology. In *Proceedings of the fourth Berkeley symposium on mathematical statistics and probabilitiy* (pp. 321-334). Berkeley: University of California Press.

Scheiblechner, H. (2009). Rasch and pseudo-Rasch models: suitableness for practical test applications. *Psychology Science Quarterly*, 51, 181-194.

Simas, A., Barreto-Souza, W., & Rocha, A. (2010). Improved estimators for a general class of beta regression models. *Computational Statistics and Data Analysis, 54*, 348-366.

Smithson, M., & Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychological Methods, 11*, 54-71.

van der Linden, W. (1997). *Handbook of modern item response theory*. New York: Springer.

Verhelst, N. (2008). An efficient MCMC algorithm to sample binary matrices with fixed marginals. *Psychometrika, 73*, 705-728.

Zeileis, A., Cribari-Neto, F., Grün, B., Kosmidis, I., Simas, A., & Rocha, A. (2013). *betareg*. R package version 3.0-2, URL http://CRAN.R-project.org/package=betareg.

Appendix

Table 2: Outcomes of non-parametric Rasch model tests

non-parametric model tests	**enjoyment items**	**anxiety items**	**boredom items**
subgroup invariance (global) internal split criterion:	$\alpha^* = .05/4 = .0125$	$\alpha^* = .05/4 = .0125$	$\alpha^* = .05/4 = .0125$
• mean (≤ mean, > mean)	.493	.173	.113
external split criteria:			
• sex (male, female)	.040	.227	.367
• comparative ability (< 0, ≥ 0)	.087	.020	.693
• mainly e-learning (yes, no)	.067	.733	.033
subgroup invariance on item level:	$\alpha^* = .05/4 = .0125$	$\alpha^* = .05/4 = .0125$	$\alpha^* = .05/4 = .0125$
• emotion group ≤ mean			
item 1	1.000 (.633)	.920 (.327)	.853
item 2	.840	.273	.707
item 3	.813	.027	.020
item 4	.293	.987 (.060)	1.000 (.533)
item 5	.420	.787	.987 (.153)
• female group			
item 1	.960 (.140)	.980 (.093)	.973 (.140)
item 2	.793	.147	.553
item 3	.720	.940 (.147)	.907 (.400)
item 4	.020	.327	.393
item 5	.680	.213	.167
• group with comparative ability ≥ 0			
item 1	.967 (.053)	.867	.960 (.180)
item 2	.180	.053	.613
item 3	.520	.173	.747
item 4	.993 (.093)	.987 (.073)	.380
item 5	.073	.693	.327
• no e-learning group			
item 1	.053	.820	.507
item 2	.987 (.020)	.707	1.000 (.020)
item 3	.320	.453	.087
item 4	.360	.467	.933 (.287)
item 5	.947 (.213)	.660	.207
item homogeneity split criterion median	$\alpha = .05$ .393	$\alpha = .05$ .340	$\alpha = .05$ .813
local stochastic independence and/or homogeneity (global)	$\alpha = .05$ .093	$\alpha = .05$ .060	$\alpha = .05$ .213
local stochastic independence (item level)	$\alpha^* = .05/\binom{J}{2} = .005$ no item-pair p < .005	$\alpha^* = .05/\binom{J}{2} = .005$ no item-pair p < .005	$\alpha^* = .05/\binom{J}{2} = .005$ no item-pair p < .005
• discrimination	$\alpha^* = .05/5 = .01$	$\alpha^* = .05/5 = .01$	$\alpha^* = .05/5 = .01$
item 1	.227	.827	.333
item 2	.593	.300	.760
item 3	.987	.013	.013
item 4	.187	.880	.973
item 5	.687	.980	.920

Bonferroni-corrections are indicated by α^*. The p-values were based on n = 150 simulated data matrices. Values in parentheses are the p-values obtained from testing if the items for a given group are too difficult as expected under the Rasch model. To be able to replicate these outcomes we fixed for each goodness-of-fit test a different starting point.