Validation of a unidimensional and probabilistic measurement scale for proenvironmental behaviour by travellers

Please cited as: **Gaborieau, J.B., Pronello, C. (2019): Validation of a unidimensional and probabilistic measurement scale for pro-environmental behaviour by travellers. Transportation. ISSN 0049-4488 DOI 10.1007/s11116-019-10068-w. 39 pp.**

Validation of a unidimensional and probabilistic measurement scale for pro-environmental behaviour by travellers

Gaborieau, J.B., Pronello, C.

Abstract

In the current debate, ecological themes have become a key element that can influence public policy, as recent events involving green activist groups have shown. Public policies targeted to education, along with focused advertising, can strongly influence people's beliefs and their emotional reactions. Understanding individual behavioural responses is therefore of the utmost importance for policy makers wishing to encourage more sustainable mobility. They could be greatly assisted by an effective measure of ecological behaviour giving them a better understanding of the determinants of travel behaviour, enabling them to analyse the impact of adopted policies. Ideally, such a measure should be simple to use, and it should be usable across different cultural and geographical contexts so as to allow comparisons between different countries.

This paper seeks to determine whether the General Ecological Behaviour (GEB) questionnaire – as a dichotomous multi-items Rasch scale for ecological behaviour measurement – is valid for use in a different cultural context. We refer to the relevant literature, and we describe our approach in detail so that it may easily be adopted by interested practitioners. The research was done in the metropolitan area of Torino (Italy), where a multimodal real-time smartphone application to assist travellers and encourage them towards more sustainable mobility was being developed and trialled. Within this framework, an investigation was done into the pro-environmental behaviour of the participants in the app trial. Our aim was to determine whether a general pro-environmental attitude can legitimately be assessed using Item Response Theory and, notably, the Rasch model. Results suggest that, using an Item Response Theory model, GEB is a questionnaire that is able to effectively measure pro-environmental behaviour by travellers. There are no discrepancies between pro-social behaviour (a trait that is known to correlate with environmentally friendly attitudes and that the GEB questionnaire seeks to measure) and actual environmentally friendly behaviour; one-dimensionality, item reliability, and the absence of simple differential item functioning are all good indicators of a model that functions well. GEB has shown its potential in providing an understanding of people's attitudes towards environmental issues and of how this information might be used to better tailor public policies in a number of sectors, in particular transport.

Keywords

travel behaviour, General Ecological Behaviour, sustainable mobility, pro-environmental behaviour, attitude measurement, Rasch model.

1. Introduction

Despite increased awareness of environmental issues, the environment continues to be harmed by human behaviour. Changes to individual behaviour in an otherwise unchanged world are not enough to remove the threats caused by a damaged environment. Sustainability requires a collective effort to change current travel behaviour. Changes in travel behaviour towards more environmentally friendly choices are a necessary but insufficient condition for sustainability, and they need to be measured to quantify their effect.

Policy makers wishing to encourage sustainability need measuring instruments (as well as theories and methods) to quantify the level of commitment to pro-environmental behaviour, and this is especially true in the transport sector. This level of commitment is a reflection of the extent to which individuals consent to making a certain effort, and of the socio-cultural, industrial and political context. Social psychologists have shown that individual behaviour arises from deeply held beliefs, from social dynamics and from emotional responses. In particular, there are numerous theories in which the direct precursor to behaviour is *behavioural intention*, and this in its turn is preceded by certain *attitudes towards the object of behaviour* (Ajzen, 1991; Garling et al., 2003; Bamberg and Moser, 2007).

Psychosocial factors and travel behaviour

Hunecke et al. (2007) used a hierarchical regression analysis to assess the effects of psychological variables when social demographics and infrastructures are controlled, and they concluded that the inclusion of attitudinal variables in a model of individuals' choices of transport modes accounted for an additional 14% of explained variance with respect to a model based on sociodemographic and infrastructural factors alone. This finding revealed our poor understanding of the effect of psychological factors on travel behaviour, and we are probably still missing key factors that could improve our understanding (Pronello and Gaborieau, 2018).

Hunecke et al. (2010) claimed that "behaviour specific attitudes and beliefs are better predictors of behaviours than values or general environmental concerns". Pronello and Camusso (2011) found that an educational policy "could trigger the positive behavioural intention to increase the use of bike and bus (40%) and the commitment to act in a different way, consistent with their general beliefs, strengthening, as a consequence, their behavioural control". However, the finding by Harland et al. (1999) that environmental concern and knowledge may influence moral norms does

not always hold true. This can be attributed to a diversity of cultural backgrounds and habits in countries where environmental sensitivity is still not broadly shared (Pronello and Camusso, 2011). It is confirmed by the literature that points to differences both between northern and southern European countries (European Commission, 2005; Korfiatis et al., 2004; Wright and Klÿn, 1998) and between different regions within the same country (Steg and Gifford, 2005) as regards attitudes to compliance with the rules and concern for the environment. Recently, Haustein and Nielsen (2016) clustered the population of the European Union into eight mobility styles that differ in travel-related choices, socioeconomic factors, IT-affinity and life satisfaction. EU countries were grouped into six regional clusters according to these eight mobility styles. This gives an indication that diverse cultural backgrounds within Europe are likely to require different policy approaches when implementing Sustainable Urban Mobility Plans (SUMPs).

Another element to consider is the context (such as the availability of public transport) that influences pro-environmental behaviour when there are trade-offs to be made between protecting the environment and personal convenience, as shown by Pronello and Camusso (2011). The authors confirmed the argument made by Ajzen (1988) that general attitudes are poor predictors of behaviour, given that specific choices are governed by constraints (e.g. a lack of public transport) that play a big role in disrupting the relationship between attitudes and behaviour.

The discussion above shows that the situation is complex and that a number of variables have a role to play in inducing pro-environmental behaviour. Measuring commitment to pro-environmental behaviour remains challenging, and there is uncertainty about how to handle this kind of metric in studies of travel behaviour. In order to understand either behaviour or behavioural intention relating to mobility and the use of different modes of transport, the studies mentioned above mainly used attitudinal and intention questions (using Likert scales), or sought to determine how much people were prepared to pay in order to pollute less (Pronello and Camusso, 2011). However, Stern (2000) stressed the importance of distinguishing between measures of intent and measures assessing actual impact.

Pro-environmental behaviour and the attitude-behaviour gap

Pro-environmental behaviour, like all types of behaviour, implies complex combinations of skills and competencies that interact with affective and cognitive process. This complexity is reflected in the scientific literature, where differences between actual (i.e. observable) behaviours and cognitive behaviours (the way in which thoughts and feelings influence decision making), attitudes, and intentions are often blurred (Pronello and Gaborieau, 2018; Heimlich and Ardoin, 2008). Some authors argue that inconsistency between attitudes and actual behaviour – the attitude-behaviour gap – is inherently due to biased measures that use reflective and introspective measurement tools (Otto

et al, 2018). Pronello and Camusso (2011) showed how general attitudes are unable to provide a satisfactory explanation of travel behaviour, and their finding has more recently been confirmed by Kroesen and Chorus (2018), who showed that specific attitudes are strongly correlated with specific behaviour, but that behaviour has a greater causal influence on attitudes than attitudes have on behaviour. These authors conclude that attitude-behaviour conceptualization, if not unfounded, remains weak. Furthermore, in the absence of a common measurement tool, it has been accepted that pro-environmental behaviour needs to be measured at a domain-specific level to increase the consistency between attitudes and behaviour (Ajzen, 1991; Axelrod and Lehman, 1993). Sometimes a single-item measure is used (Van Liere and Dunlap, 1978; Vining and Ebreo, 1992), making measurements inconsistent across studies. Otto et al. (2016) show that although single-item measures can be useful in some instances, using them to investigate pro-environmental behaviour can yield contradictory results.

Measuring pro-environmental behaviour

An effective measure of pro-environmental behaviour, allowing cross-study comparisons, should be consistent across domains and cover a wide set of different relevant behaviours. Some authors have tried to develop a general measure of pro-environmental behaviour (Maloney and Ward, 1973; Leonard-Barton, 1981; Ramsey, 1993) aggregating behaviours across domains in a single indicator (Maloney and Ward, 1973), using multidimensional scales (Ramsey, 1993) or factor analysis (Green-Demers et al., 1997), but none of these scales is able to provide an accurate, consistent and insightful measure of pro-environmental behaviour (Kaiser, 1998).

Most papers focusing on pro-environmental behaviour use self-reporting survey techniques. Although some authors report a low correlation between self-reported and actual behaviour (Corall-Verdugo, 1997), it would appear that using dichotomous items about practice (I do/I do not) or ownership (I own/I do not own) can give an accurate estimate of true behavioural patterns (Gamba and Oskamp, 1994; Hirst and Goeltz, 1985; Fuji et al, 1985; Kaiser et al., 2010).

The General Ecological Behaviour (GEB) scale (Kaiser, 1998), which built upon previous works by Fejer and Stroschein (Fejer, 1989; Fejer and Stroschein, 1991, cited by Kaiser, 1998), was developed using the Rasch Model for Scale Measurement (Rasch, 1960) with the objective of providing a unidimensional and probabilistic measure of pro-environmental behaviour. Two assumptions underlie the GEB, namely that 1) pro-environmental behaviours differ in terms of the difficulty of performing them, and 2) pro-environmental behaviours are constrained by external factors such as the sociocultural context, political agendas and the built environment. We argue that the use of such a scale can help to overcome the difficulty described above, where both psychological factors (people's commitment to pro-environmental behaviour) and environmental factors

(constraints affecting subgroups of people) need to be taken into account in assessing influences on travel behaviour. Recently, a study suggested that an adapted version of the GEB – a self-reported behavioural measure – was substantially more reliable than classical introspective attitude measures and might pave the way to the development of new psychological theories (Otto et al, 2018).

Arnold et al. (2018) mention the studies by Kaiser et al. (2007) that showed an 85%-95% overlap between GEB scores and scores from conventional measures of behavioural intention to engage in pro-environmental activities. These measures differ in their dimensionality. GEB is a one-dimensional measure based on an assumption that behaviour reflects an individual's generic intention to protect the environment. In contrast, multidimensional measures allow for different subjective goals or intentions (Stern, 2000).

This paper is part of wider research as part of the European *Opticities* project (www.opticites.com), which was set up to investigate whether new technologies, and in particular a multimodal real time navigator (ATIS - Advanced Traveller Information Systems), could drive more intermodal, sustainable travel behaviour. The first objective of this research was to determine whether Item Response Theory and, in particular, the Rasch model – in a different cultural context from that analysed by Kaiser (1998) – could provide a legitimate measure of general attitudes to the environment and thus become a common method allowing comparisons between different countries.

The second objective was to investigate the factors driving decision making, comparing models of psychosocial correlations using Structural Equation Modelling. The third objective was to produce a psychosocial-based segmentation of potential ATIS users based on various psychological constructs that play a role in determining individual mobility patterns.

Our focus in the present paper is the first objective outlined above. We were seeking to validate the General Ecological Behaviour (GEB) questionnaire as a dichotomous multi-item Rasch scale for measuring pro-environmental behaviour. The transport context used for our research is Italy, and our paper includes a step-by-step methodological guide to Rasch models for the benefit of those unfamiliar with IRT. To our knowledge, this is the first transport-focused research paper to apply such a comprehensive methodology, and we hope that it will be useful for practitioners who wish to adopt attitude measures using IRT. In-progress publications suggest its usefulness as a substitute for specific behavioural-intention measures within theory-based structural equation models.

The paper is organized as follows: the following section will present the methodology used to design the questionnaire, fit the Rasch model and test the various assumptions. Section 3 will present the results obtained applying the Rasch model for scale measurement. Section 4 will discuss the potential use of the GEB as an attitude measure in psychology-based theories of decision making and

for cross-cultural comparisons of behaviour. We refer to the existing literature and present our conclusion.

2.Methodology

The research was conducted in the metropolitan area of Torino (Italy) as part of the European *Opticities* project (http://www.opticities.com/) that oversaw the development of a smartphone multimodal real-time application to assist travellers and promote a more sustainable mobility. The ultimate goal was to understand the effect of the app on travel behaviour, and to this end the app was trialled in the Torino metropolitan area on a sample of the population. Within this framework, our research involved looking at the pro-environmental behaviour of our sample in order to determine whether a general pro-environment attitude may legitimately be assessed using Item Response Theory and, in particular, the Rasch model. A three-step methodology comprised: 1) survey design; 2) sample selection and survey administration; 3) model estimation and testing for general ecological behaviour measure.

2.1 Survey design

A three-step survey was contained within the Opticities project: 1) the *ex-ante* survey, carried out before the app trial, collected data on users' mobility patterns and attitudes as well as their requirements in relation to the "*TUeTO*" app. This survey was based on a web-questionnaire and six focus groups and was implemented in October-December 2014; 2) the *in-itinere* survey, carried out during the app trial, focused on the use of the latest version of the app itself, with the aim of monitoring problems and bugs and seeing how participants reacted to using the app when travelling. To this end, monthly web-questionnaires were administered to the participants between February and June 2016; 3) the *ex-post* survey, carried out in July-September 2016 after the app trial, aimed at assessing changes to potential travel behaviour as well as changes of perception, expectation, and preferences driven by the use of the app. These changes were assessed through a web-questionnaire and six focus groups, analogous to the *ex-ante* survey.

Our research involved including an additional web-survey, carried out just before the *in-itinere* survey and designed to help us analyse general attitudes towards the environment and the ecological behaviour of the selected sample using the General Ecological Behaviour (GEB) web-questionnaire.

The GEB questionnaire is a version adapted to the Italian context based on Kaiser and Wilson (2000). It consists of 40 dichotomous (yes/no) items (Table 1) grouped into seven different categories. Seven items represent pro-social behaviours (CS1-CS7) while the other 33 items represent pro-

environmental behaviours, separated into six ecological domains: garbage handling (R1-R6), water and power saving (AE1-AE7), consumerism (CE1-CE6), garbage inhibition (RR1-RR5), environmental activism and volunteering (V1-V4) and transport (T1-T5). It will be remarked that the questionnaire seeks to measure pro-social behaviour and relevant pro-environmental behaviour on a single scale. The idea of combining two different types of behaviour in this way is supported by findings that pro-environmental values are highly correlated with social-value orientation (Vugt et al., 1995; Gärling et al, 2003; Bamberg and Moser, 2007).

In relation to original questionnaire, one additional item was included (R6 – "I sort plastic waste for recycling") and seven items were adapted to the Italian context. Two of the seven items were modified: 1) RR5 "Usually, I buy water in returnable bottles", substituting water for milk in the original questionnaire. In Italy water is generally consumed from bottles and a different behaviour has to be engaged if plastic or glass bottles are used; 2) T3 "When possible, I do not use a car for distances less than 30km", substituting 30 km for 20 miles.

Table 1: Structure of the GEB questionnaire

2.2 Sample selection and survey administration

The subjects used in this research were participants in the *Opticities* project. A stratified sampling plan of convenience was used to select 150 users of different types of transport with reference to the following criteria:

- gender;
- age: classes related to people having different technological skills;
- profession/educational level/income;
- presence of children under 14 in the household;
- mode used by travellers: motorised, public transport (PT), soft modes, intermodal (motorised + PT);
- residential location: city centre, suburbs, extra-urban locations, considering also the geographical position (north, east, south, and west). It is important to observe the origins and destinations to better choose the people's profiles, also in terms of their residential location.

In view of the likelihood that some participants would withdraw (this did in fact occur), more than 150 were contacted. Thus, 159 subjects participated in the *ex-ante* survey to determine user requirements for a real-time multi-modal navigator. After this *ex-ante* survey a few participants withdrew, and the 142 remaining took part in the app trial and *in-itinere* survey. The final step in the three-step survey was the evaluation of effects on subjects' travel behaviour (the *ex-post* survey).

The GEB questionnaire was made available to the 159 subjects (those who had taken part in the *ex-ante* survey) in early February 2016, that is to say shortly before the start of the *in-itinere* survey. The participants received an e-mail containing the link to the LimeSurvey platform allowing them to fill in the questionnaire, although they were also given the choice of submitting their replies on paper. Responses were collected as soon as they were submitted in early February 2016 (over a week). 131 of the 159 subjects agreed to respond (81.8%).

2.3 Database construction

Despite having a good level of motivation to complete the questionnaire to the best of their ability, not all the respondents answered all the questions, and there were 273 missing values (5.2%) out of 5240 item-responses (40 items x 131 respondents). Such data may be considered as *structurally missing data* or *Missing Not At-Random* (MNAR), where answers are missing for very specific reasons that are related to self-knowledge about investigated behaviours, as shown below. Where respondents did not answer we hypothesized that there were specific reasons for this relating to self-knowledge about specific behaviours featuring in the questionnaire, and we were able to cross-check with our available information (e.g. driving licenses) to confirm some of our hypotheses.

Although the non-imputation of missing data can give good results in some circumstances, such as when sample sizes are large and missing values are few (~5%) (Soysal et al., 2016), we chose to recode missing values, in view of the type of analysis and the model used to measure ecological behaviour. Some substitutions could not be avoided where non-parametric tests do not allow missing values. We explored several methods and, because of the typology of the missing answers (MNAR), we concluded that manual imputation was the best approach, given the following structural reasons behind the non-responses.

The most problematic items were:

- 25-CE5 ("I use phosphate-free laundry detergent"), with 35 missing values (26.7%), which were filled with "No", on the assumption that anyone who does not know whether (s)he uses a phosphate-free laundry detergent will not buy it deliberately;
- 4-CS4 ("If I were an employer, I would not hesitate to hire a person previously convicted of crime"), with 31 missing values (23.6%), which were filled with "No" for all of them, on the assumption that the doubt or unwillingness to answer is revealing of the hesitation itself;
- 31-RR5 ("Usually, I buy water in returnable bottles"), with 30 missing values (22.9%), which were filled with "Yes" for all of them, on the assumption that these people do not buy bottled water at all, and that this is a pro-environment choice analogous to buying returnable bottles;
- 17-AE4 ("I turn off the heating at night"), with 15 missing values (11.5%), which were filled with "No" for all them, on the assumption that these respondents live in apartments connected

to a central heating system lacking the possibility of individual control (often the case in Torino);

- 36-T1 ("Usually, I do not drive my automobile in the city"), 37-T2 ("I usually drive on freeways at speeds under 100 km/h") and 38-T3 ("When possible, I do not use a car for distance lower than 30km"), each of them with 11 missing values (8.3%). After consideration of respondent's driving licence, car ownership and usage frequency from the *Opticities* questionnaire, these missing values were filled with "Yes", on the assumption that the respondents do not drive in general.

After the above changes in the database (manual imputation of missing data), the percentage of missing values fell to 2.4% of the item-responses. These were then filled with "No", on the assumption that not answering certain items reveals either that the behaviour is, in general, not engaged or engaged by chance without a strong intention to behave in that way.

Finally, in the case of questions formulated such that "No" was the pro-environment response, values were reversed ("Yes" in place of "No" and "No" in place of "Yes") in order that throughout the questionnaire "Yes" corresponded to more pro-environmental behaviour.

2.4 The Rasch model as a measure of general ecological behaviour

The estimation of general environmentally friendly behaviour, based on the data collected by the questionnaire, was done using the Rasch Model for scale measurement (Rasch, 1960).

The Rasch Model is a special case of Item Response Theory (IRT, also known as Latent Trait Theory), which in psychometrics is the alternative paradigm to Classical Test Theory (CTT). Whereas in CTT all items are considered equivalent and treated in aggregation, IRT treats items differently, according to their relative difficulty, and focuses on the interaction between an item's difficulty and the ability of individuals, termed θ_n. IRT is therefore a theory based on the idea that the probability of a respondent's answer to an item can be described as a function of that individual's location on the latent trait and of one or more parameters characterizing the item. The item-response function corresponds to an Item Characteristic Curve (ICC). IRT has a number of advantages over CTT: there is less inconsistency when applying items to different samples (Revelle, 2011); IRT produces fewer measurement errors than CTT (Magno, 2009); and individuals and items are calibrated on a common scale, which facilitates the interpretation of the measured variables (Embretson, 1996). It is thus possible to compare individuals in terms of probability of response, which is much more informative than saying that someone is one standard deviation above the mean score.

The Rasch model (Rasch, 1968) is the simplest case of IRT and it assumes only one parameter per item – the difficulty β_i – thus it is sometimes referred to in the literature as the *one-parameter*

logistic IRT model. Additional parameters used in *two-* or *three-parameter IRT* include *discrimination* (slope of the ICC) and *pseudo-guessing* parameters (that force a lower asymptotic limit, so that the probability never reaches zero).

Formally, considering a dichotomous random variable where $x = 1$ denotes a correct answer and $x = 0$ an incorrect one, the probability of individual n answering correctly for item i is given by equation (1):

$$P(x_{ni} = 1) = \frac{e^{x_{ni}(\theta_n - \beta_i)}}{1 + e^{(\theta_n - \beta_i)}}. \quad (1)$$

where θ_n is the ability of person n and β_i is the difficulty of item i.

See Fisher (1997) for details of the assumptions behind the Rasch model (one-dimensionality; monotonic functions; local stochastic independence; sufficiency of a simple sum statistic; dichotomy of the items).

The Rasch model, although now used in a wide variety of scientific fields (Andrich, 2004), uses a specific vocabulary, derived from educational science. We should therefore point out that in our application of this method on the GEB questionnaire, there are no *correct* or *incorrect* answers, but rather the *performance* or *non-performance* of given behaviours. Similarly, *difficulty* corresponds to the effort required of an individual to engage in a given behaviour, and *ability* corresponds to the particular location of an individual on a general unidimensional latent trait giving rise to specific behaviour. This measure will respond to the criteria of a Campbellian attitude measurement (Campbell, 1963), as it is derived only by measuring specific attitude-relevant practices (see also Kaiser and Byrka, 2015).

Parameter estimation. Statistical methods for estimating Rasch model parameters may be seen as combinatorial computations, across all items and all individuals, of the logistic equation (1). Various estimation methods exist, and for our needs we chose WINSTEPS[1] and the eRm package for R (Mair and Hatzinger, 2007a; Mair et al., 2018). WINSTEPS uses two consecutive estimation methods: the Normal Approximation Estimation Algorithm (PROX; Linacre, 1994a), recognised for its efficiency, followed by a Joint Unconditional Maximum Likelihood Estimation (JMLE or UCON; Wright and Douglas, 1977). As for eRm, its core Rasch Model estimation method is implemented with a Conditional Maximum Likelihood function (CML; Mair and Hatzinger, 2007b). A detailed mathematical description of these estimation method is given in Gaborieau (2016).

[1] http://www.winsteps.com

Rasch Model Fit. The aim is to determine whether items within the General Ecological Behaviour questionnaire are valid for assessing a Rasch measure of a one-dimensional latent trait. To this end, we follow the general guidelines proposed by Linacre (2005). After estimating the parameters for both items and individuals, we observe and analyse the *point-biserial correlation* and the *fit statistics*.

Point-biserial correlation: a positive answer to more-difficult items should correlate positively with measures relating to the individual. The point-biserial correlation is an adaptation of Pearson's correlation when one of the variables is dichotomous (Jaspen, 1946) and is given by equation (2):

$$r_{pbi} = \frac{\sum_{n=1}^{N}(X_{ni} - \bar{X}_i)(\theta_n - \bar{\theta})}{\sqrt{\sum_{n=1}^{N}(X_{ni} - \bar{X}_i)^2 \sum_{n=1}^{N}(\theta_n - \bar{\theta})^2}}, \quad (2)$$

where $X_{ni}$is the observation of individual n on item i, $\underline{X_i}$ is the mean of the X_{ni} on item i, θ_n is the trait measure for individual n and $\underline{\theta}$ is the mean of θ_n. As X_{ni}=$E_{ni} \pm W_{ni}$, the expected observation and its variance, we can compute the expected point-biserial correlation (Olsson et al., 1982) with equation (3):

$$E(r_{pbi}) \approx \frac{\sum_{n=1}^{N}(E_{ni} - \bar{X}_i)(\theta_n - \bar{\theta})}{\sqrt{\sum_{n=1}^{N}((E_{ni} - \bar{X}_i)^2 + W_{ni}) \sum_{n=1}^{N}(\theta_n - \bar{\theta})^2}} \quad (3)$$

Fit statistics: two kinds of mean squared fit statistics are calculated, namely OUTFIT (standing for *Outlier-sensitive fit statistics*) mean square and INFIT (*Inlier-pattern-sensitive fit statistics*) mean square. These give an indication of how well the model fits the observed data. Both OUTFIT and INFIT are based on classical χ^2 fit statistics, as reported by Wright and Panchapakesan (1969), which makes a transformation into Z-statistics possible. Equations (4) report the formulas for both OUTFIT (U_i) and INFIT (V_i) for each item:

$$U_i = \frac{\sum_{i=1}^{N} {Z_{ni}}^2}{N}, V_i = \frac{\sum_{i=1}^{N} {Z_{ni}}^2 {W_{ni}}^2}{\sum_{i=1}^{N} {W_{ni}}^2}, \quad (4)$$

where Z_{ni} is the standardised residual between the model and the observation, and ${W_{ni}}^2$ is the variance of X_{ni}. OUTFIT is a traditional sum of squared standardized residuals, sometimes reported as *non-weighted mean square error*; it is sensitive to unexpected responses at some distance from the item parameter (an individual with a low measure on the latent trait engaging in a difficult behaviour, or a person with a high measure failing to engage in an easy behaviour), whereas INFIT is considered as the *information-weighted mean square error*. INFIT statistics are weighted according to the quantity of statistical information provided by an item, that is to say its variance ${W_{ni}}^2$ (an easy

behaviour that most people will engage in will have a low variance and low information). INFIT statistics are sensitive to unexpected responses close to the item parameter (Smith et al., 2008). INFIT and OUTFIT mean square statistics have an expected value of 1.0 and a range that goes from 0.0 to positive infinity (Bond and Fox, 2013). Values greater than 1.0 indicate more variation in the observed data than predicted by the model, and this is referred as *underfit*, where response patterns are unpredictable. In contrast, values lower than 1.0 correspond to a variation in the observed data that is lower than predicted by the model, and this is referred to as *overfit*, where response patterns are too predictable, close to what will be expected with a Guttman pattern[2]. Although the range of acceptable values for INFIT and OUTFIT statistics are still open to debate (Smith et al., 1998; Karabatsos, 2000; Smith and Suh, 2003), the reference values commonly used are those proposed by Wright et al. (1994), where acceptable mean square values are between 0.5 and 1.5 (Table 2).

Table 2: Interpreting INFIT and OUTFIT statistics

The corresponding standardised Z-score – corresponding to the probability of the mean square following a unit-normal deviation when the data fit the Rasch model – is expressed using Wilson-Hilferty cube root transformation (Wislon and Hilferty, 1931) (Equation (5)):

$$z(U_i) = \left(U_i^{\frac{1}{3}} - 1\right)\left(\frac{3}{\sigma_i}\right) + \left(\frac{\sigma_i}{3}\right), z'(V_i) = \left(V_i^{\frac{1}{3}} - 1\right)\left(\frac{3}{\sigma'_i}\right) + \left(\frac{\sigma'_i}{3}\right), \quad (5)$$

where σ_i and σ'_i stand respectively for the standard deviation of $U_i$and V_i and are not explicitly given here (refer to Wang and Chen, 2005). Z-score is interpreted as a classical t-statistic, where a value of 1.96 corresponds to a two-sided significance of 5%.

Observed and expected correlations, together with INFIT and OUTFIT statistics, allow us to focus our validation process on specific items, but they cannot be used to blindly accept or reject a given item. As explained by Linacre (2006), when dealing with real-world observations misfits are to be expected, and validation of the Rasch Model should be done to make sense of data, given that there will not be a perfect fit between observations and the model.

Rasch Model Testing. Different categories of tests, parametric and non-parametric were conducted to validate, on the one hand, the assumptions of the Rasch Model – including the one-dimensionality of the measure and subgroup homogeneity – and, on the other hand, the reliability of

[2]A Guttman scale (Guttman, 1950) is a deterministic version of the Rasch scale. In a Guttman scale, if the items are ranked by difficulty, the following will be true: 1) if a given answer is correct, then all easier answers are also correct and 2) if a given answer is incorrect, then all more difficult answers are also incorrect. Thus, knowing the last correct answer provides all information needed to know the answers to other items and the ability of the respondent in relation to the trait measured by the scale.

the measure. The tests were chosen according to whether they were implemented in R, and so as to validate all of the general assumptions underlying the Rasch model in order to be certain of its robustness.

The tests for one-dimensionality included a Principal Components Analysis on the residuals produced by the Rasch Model, a Martin-Löf Test, and a non-parametric T_{md} test. The test for local stochastic independence was performed with the non-parametric T_{11} test. Subgroup homogeneity was tested using two splitting criteria (mean score and gender) and assessed by a graphical model check, an Andersen Likelihood-Ratio test, and a non-parametric T_{10} test. The KR-20, individual separation, and item separation reliability statistics were then calculated. Item Characteristic Curves will be presented below.

Testing one-dimensionality: one-dimensionality is one of the foundations of the Rasch model and the assumption to be checked first and foremost. In the ideal case of a perfect Rasch scale, the Rasch dimension – i.e. the latent measure the Rasch model is estimating – is the only dimension in the data and all other unexplained variance should only be random noise. Two different tests were conducted:

- according to Linacre (2005), one-dimensionality may be assessed by performing a Principal Component Analysis on the matrix of inter-item correlations of the standardized residuals produced by the model. The PCA evaluation produces components that are, in this case, known as "contrasts", as a reminder that they are derived from the residuals and not from the raw data matrix;
- Martin-Löf (1970) proposed a test of one-dimensionality: for D disjoint sets of items, the hypothesis that the items measure the same one-dimensional latent construct can be tested using the likelihood ratio test, based on equation (6) (Martin-Löf, 1970, cited by Christensen et al., 2002):

$$LR = 2\left(\sum_{r_1=0}^{k_1}\cdots\sum_{r_D=0}^{k_D} n_{r_1\cdots r_D}\ln\left(\frac{n_{r_1\cdots r_D}}{N}\right) - \sum_{r=0}^{k} n_r \ln\left(\frac{n_r}{N}\right) - \ln\Lambda\left(\hat{\beta}|R\right) + \sum_{d=1}^{D}\ln\Lambda\left(\hat{\beta}_d|R_d\right)\right) \qquad (6)$$

$R_{1\cdots d}$ being the raw score from subset $D_{1\cdots D}$ composed of $k_{1\cdots D}$ items and $n_{r1\cdots D}$ the number of person with raw score $R_{1\cdots d}$.

Testing subgroup homogeneity and differential item functioning: a good Rasch model should produce similar item difficulty parameters independently of the population sample. To this purpose, Andersen (1973) proposed a Likelihood-Ratio test that consists in arbitrarily splitting the sample into two (or more) disjoint groups *G*. We would expect the estimates of the parameters β_{Gi} to be the same. In this

regard, Rasch himself proposed a graphical model check (Rasch, 1960), that can be obtained by plotting β_{1i} against β_{2i}, where the items should not deviate too much from the diagonal. The test is consequently able to detect *differential item functioning*, which happens when individuals with the same level of an underlying latent trait differ in their response to an item depending on other characteristics. Andersen's LR test is similar to Martin-Löf's, but is based on creating subgroups of individuals rather than subgroups of items. We tested the model using different splitting criteria. First, we divided the sample according to individuals' raw scores in the questionnaire (i.e. the sum of positive answers). One group consisted of respondents having a score of less than or equal to the median score (n = 62), and the other of respondents having a score of more than the median score (n = 69). Second, we divided the sample according to gender, one group consisting of males (n = 76), and the other of females (n = 55).

Non-parametric quasi-exact tests: Ponocny (2001) proposed a family of non-parametric tests using a Monte Carlo algorithm for goodness of fit. Based on the assumptions of sufficient statistics, all matrices with identical margins should have the same parameter estimates. Let A_0 be the observed matrix of size $(n\ items\ \times p\ individuals)$. We can, theoretically, generate all possible matrices with margins as in A_0, denoted $A_s\ \in\ \Omega_{np}$, with $(s = 1, \cdots, S)$. In practice, the generation of all possible matching matrices is computationally very demanding; this is why Ponocny (2001) proposed to simulate a sample of possible matrices with a Monte-Carlo algorithm, which was improved as a Markov Chain Monte-Carlo (MCMC) algorithm by Verhelst (2008). Because these tests are based on a reduced sample of all possible matrices, they are called *quasi-exact tests* and are more reliable than parametric tests for a small sample (Ponocny, 2001). A given test-statistic T is computed both for the observed matrix $A_0(T_0)$ and all generated matrices $A_s(T_s)$. By counting how often T_s has similar or more extreme values than T_0, we can define the *re-sampling p-value* under the null hypothesis "The data conforms to the model", as well as the relative frequency given by equation (7):

$$\mathrm{p} = \frac{1}{\mathrm{S}} \sum_{\mathrm{s}=1}^{\mathrm{S}} \mathrm{t_s}\,, \text{where } \mathrm{t_s} = \begin{cases} 1 & , \quad \text{if } \mathrm{T_s} \geq \mathrm{T_0} \\ 0 & , \quad \text{elsewhere.} \end{cases} \qquad (7)$$

The different tests we conducted on our data matrix are the following:

- T_{10}, global test for subgroup invariance. This test is the non-parametric equivalent of Andersen's LR test described above. The idea is that, within the Rasch model, the quotient $\frac{n_{ij}}{n_{ji}}$ should be approximated by $e^{(\beta_j - \beta_i)}$, where $n_{ij}$is the number of respondents who have a positive answer to item i but not to item j. This holds true for any sub-sample G of respondents. Therefore, we may use the equation (8) in equation (7).

$$T_{10} = \sum_{ij} \left| n_{ij}^{(g_1)} n_{ji}^{(g_2)} - n_{ij}^{(g_2)} n_{ji}^{(g_1)} \right|, over\ all\ pairs\ (i,j), \quad (8)$$

We conducted this test with the same splitting criterion used for Andersen's LR test, i.e., based on median raw score and gender.

- T_{11}, test for local stochastic independence. Good Rasch items should correlate to each other only through the latent dimension they measure, which is a consequence of the one-dimensionality assumption. In other words, the answer to a given item should not be determined by an answer to another item; statistically speaking, correlations of residuals should be zero. Therefore, a test for the violation of local stochastic independence may be expressed as in equation (9):

$$T_{11} = \sum_{ij} \left| r_{ij} - \widetilde{r_{ij}} \right|, over\ all\ pairs\ (i,j), \quad (9)$$

where r_{ij} are the observed inter-item correlation and $\widetilde{r_{ij}}$ its expected value, estimated as a mean r_{ij} for the simulated matrices. The model test is computed by using equation (7) on T_{11} (equation (9) and defined as the relative frequency of T_s having the same or a larger value than in T_0.)

- T_{md}, test for multidimensionality. Developed by Koller and Hatzinger (2013) on the principles formulated by Ponocny (2001) and based on Martin-Löf's test described above, this test is formulated as in equation (10):

$$T_{md} = Cor\left(r_n^{(d_1)}, r_n^{(d_2)}\right), \quad (10)$$

where $r_n^{(d_i)}$ is the raw score of person n on subscale d_i. If the Rasch model holds, the two sub-scaled raw scores should be positively associated. The model test is given by equation (7) and it is defined as the relative frequency of T_s having the same or a smaller correlation value than in T_0.

Reliability: Reliability is expressed as the quotient of true variance over observed variance and shows the level of reproducibility of the measures (Peter, 1979). The method used for estimating the true variance will produce different reliability indexes. We report in our results the following reliability coefficient:

- the KR-20 (Kuder and Richardson, 1937), which is a special case of Cronbach's α for dichotomies, based on raw score variance;

- the individual separation reliability r_θ, virtually equivalent to the KR-20, but based on variance of individuals' abilities;
- the item separation reliability r_β, based on variance of item difficulties.

The model estimates and tests were computed using either WINSTEPS 3.80.1 or the eRm package v0.15-6 for R release 3.2.3. Differences in parameter estimation between WINSTEPS and R are due to different estimation methods: WINSTEPS estimates item and individual parameters simultaneously using a normal approximation method followed by a Joint Maximum Likelihood, while R uses Conditional Maximum Likelihood for item parameters and Joint Maximum Likelihood for individual parameters. This does not interfere with our analysis, since parameter estimates for the two methods are linearly related (Figure 1). We used WINSTEPS to compute fit statistics and to test for one-dimensionality through a PCA, along with subgroup homogeneity and differential item functioning. We used eRm for the Martin-Löf test of one-dimensionality (1970) and for all the quasi-exact tests described above.

Figure 1: Winsteps vs eRm estimates of item and person parameters

3.Results

The GEB questionnaire was completed by 131 participants. The mean age of respondents was 41.4 years (median = 41.0 years, range from 20 to 75 years), 42% were women (N = 55). Table 3 shows the household size, the number of children and the public transport (PT) subscription. More than half of the respondents belonged to households of more than 4 people, while 69 respondents did not have any children. The mean age of respondents' children was 10 years, the youngest being less than one and the oldest 24. 51% of the sample (N=67) did not have a public transport subscription.

Table 3: Descriptive statistics of the sample

Figure 2 shows the attributes of the most frequent trips. The longest trips are "*chains*" (involving a series of different transport modes), while walking is limited to trips shorter than 3 km. Bicycles are used for distances less than 7 km. The most frequent purpose of a trip is travelling to work (79%) or to study (9%). Car is the most used mode – almost 50% of the respondents use a car at least four times a week – followed by public transport (50% of the respondents use public transport at least three times a week).

Figure 2: Attributes of the most frequent trips (mode, distance, purpose and frequency)

Mean net monthly household income was just below 3000 euros (median = 2750 euros). According to the Italian National Statistical Institute[3], average household income in metropolitan areas is around 2720 euros per month, and median household income around 2150 euros per month. A t-test of equal means returned a significance value of p=0.376, indicating that the mean we measured is not different from that of the general population. However, the Mann-Whitney U-test on median values returned a highly significant level: our sample is wealthier overall than the metropolitan Italian population.

Concerning our sample (N=131), its size can be considered suitable for applying the Rasch model, as suggested by Lord (1983) and confirmed by Linacre (1994b) who proposed that, in our case, item calibration may be considered stable within a 0.5 logit deviation with 95% confidence.

Table 4 presents the estimates of item parameters ("MEASURE") from WINSTEPS, together with corresponding observed and expected point-biserial correlations, and INFIT and OUTFIT statistics. Additional information includes the raw score for items ("SCORE") as well as the percentage of observed and expected positive answers for each item ("EXACT MATCH"). In Table 4, items are ordered by increasing observed point-biserial correlation.

It can be seen from Table 4 that some items are problematic, as we now explain:

- Item 27-RR1 ("I re-use plastic grocery bags") shows a high mean square OUTFIT value of 2.02 and a negative correlation with individual measures (-0.07). This would appear to be one of the easiest behaviours (MEASURE = -4.06) and was answered "Yes" by all but one respondent (TOTAL SCORE = 130, TOTAL COUNT = 131), which caused the observed misfit. A possible explanation is the semantic ambiguity of the item; perhaps this person does not use plastic bags to carry groceries home and, therefore, answered "No" to this specific task of re-using them. Considering the acceptable Z-standardised (1.1), and the fact that, although negative, the observed correlation is close to the expected correlation, we decided to retain this item in the final model;
- Item 5-CS5 ("If a friend or a relative was in hospital for a week or two for minor surgery I would visit him or her") shows a value of mean square OUTFIT of 1.65, that is "unproductive but not degrading for the measurement" (Table 2). The observed correlation with individual measures is negative (CORR. = -0.03) but close to the expected correlation (EXP. = 0.07). Similarly to item 27-RR1, the behaviour is considered easy by the model (MEASURE = -

[3] http://dati.istat.it

3.35) and was answered positively by all but two respondents. We also decided to retain this item in the final model;

- Item 14-AE1 ("Before taking a shower, I let the water run so it gets to the temperature I want") may show good mean square statistics (INFIT MNSQ = 1.11, ZSTD = 0.5; OUTFIT MNSQ = 1.48, ZSTD = 1.4) but a negative correlation with individual measures (CORR. = -0.02), quite far from the expected correlation (EXP. = 0.20). This means that negative answers to this items do not correlate with individual measures on the latent trait as it should, and is symptomatic of un-modelled noise (Linacre, 2008). In other words, it may indicate that individuals with a more pro-environmental profile tended to answer "Yes", and those less environmentally aware to answer "No". We decided to exclude this item from the final model;
- Item 25-CE5 ("I use phosphate-free laundry detergent") shows acceptable mean square values (INFIT = 0.85; OUTFIT = 0.80) but very high negative Z-standardised scores (INFIT = -2.4; OUTFIT = -2.6). We concluded that this item probably does not fit the Rasch model (p = 0.016 for INFIT; p = 0.09 for OUTFIT) and we excluded it from the final model.

Table 4: Estimates of Item parameters, INFIT, OUTFIT and bi-serial correlation statistics

With 2 of the 40 items excluded for the reasons given above, the parameters were estimated with this new set of 131 individuals x 38 items where fit-statistics were satisfactory, as shown in Figure 3.

Figure 3: Item map of INFIT statistics for final item selection

Fit statistics can be used in relation to individuals as well as to items. Out of 131 respondents, 8 had high OUTFIT values (from 1.6 to 9.82), but other indicators (INFIT, observed and expected scores and correlations) were good. We thus decided to retain all individuals for the final computation.

3.1 Testing the Rasch model

This section presents the results of the tests for one-dimensionality, local stochastic independence, and differential item functioning or subgroup homogeneity, that is to say the tests for validating the general assumptions behind the Rasch model. It also discusses reliability and presents Item-Characteristic Curves.

Check for one-dimensionality: Table 5 presents the results of the Rasch-residuals-based PCA and Figure 4 shows the corresponding scree plot. *Variance explained by measures* shows how much of the variance within the data is explained by the model. A small variance explained may indicate an

inappropriate test. But if the sample of individuals has a tight range of ability and items have a narrow range of difficulty, the variance explained will be small even for the very best test. The right way to interpret Table 5 is to compare the columns *empirical* (the results from the model fit) and *modelled* (the expected results if the data fitted the Rasch model perfectly). We can check whether there is a problem by breaking down the remaining unexplained variance using a PCA. The term *contrast* is used to talk about *components* or *factors* because of the specific interpretation we can give: contrast might not be a sub-dimension but simply the result of random noise in the data.

Table 5: Results of the PCA performed on residuals

Figure 4: Scree plot of the PCA variance component

Comparing the values contained in the *empirical* and *modelled* columns in Table 5, we do not see any noticeable difference, confirming the good fit of the model. The first contrast has an Eigenvalue of 2.8, explaining 5.0% of total variance. Although a 2.8 Eigenvalue is high enough to warrant investigating this possible second dimension produced by the data, 5.0% of total variance explained is low enough for us to ignore it (Linacre, 2005). By plotting the loading of items on the 1st contrast of the residuals-based PCA (Figure 5), we clearly see that this possible dimension is produced by transport-related items. In fact, item T1 ("Usually, I do not drive my automobile in the city"), T2 ("I usually drive on freeways at speeds under 100 km/h") and T5 ("I walk, ride or take public transport to go to work/university") are quite far away from the general cluster created by the other items. In Figure 5, only the five items with the highest loading have been deciphered.

Figure 5: Item loadings on the first contrast

The non-parametric T_{md} test conducted on 1000 sampled matrices was highly significant (p-value=0.001). The result of this specific test indicates that transport-related items show either low discrimination and/or multidimensionality (Koller and Hatzinger, 2013). However, the Martin-Löf Test for one-dimensionality grouped all items other than those related to transport in one subset, while transport-related items were grouped in a second subset. This test gave a p-value equal to 0.997, comforting the idea of a one-dimensional trait measure by the GEB questionnaire. We thus conclude that transport-related items do not produce a second dimension of measurement although, of course, they do not discriminate very well. Item discrimination shows how much engaging in a specific behaviour (in our case, transport-related items) will lead to high score over the whole questionnaire.

Test for local stochastic independence. The test T_{11} produced a significant result (p-value < $10e^{-4}$) and leads us to reject the hypothesis of local independence. This result means that some of our items in the GEB questionnaire are related to each other. Taking into account the results of dimension exploration (Figure 5), it is reasonable to conclude that transport-related items are subject to other conditions, such as car ownership. In the same way, we could expect that the items concerning the collection and recycling of different types of garbage (glass, paper and plastic) are somewhat correlated, under the influence of another factor, like living in a zone where differential garbage collection is provided by the local authorities.

Test for differential item functioning or subgroup homogeneity. Figure 6 shows the graphical representation of test results for the two alternative splitting criteria used, namely *median raw score* and *gender*. Item parameter estimates for the two subgroups are plotted against each other; red ellipsoids represent the 95% confidence interval of the beta estimates for both dimensions, i.e. both subgroups. The diagonal plain line represents the line along which all points would lie if there were no differences between subgroups. As long as red ellipsoids cross the diagonal plain line, we can conclude that items are homogenous across subgroups, i.e., they have equal difficulty. Neither of the likelihood-ratio tests could lead to rejection of the null hypothesis of subgroup homogeneity (p-value = 0.151 for median raw score splitting and p-value = 0.098 for gender splitting), from which we conclude that items are equally discriminatory for subgroups, which is a positive indication for the quality of the Rasch measure. Ponocny's T_{10} test was performed using the same splitting criteria. The same conclusion can be drawn (p-value = 0.096 for median raw score splitting and p-value = 0.076 for gender splitting). However, examining the right-hand side graph within Figure 6, we remark that item 37-T2 ("I usually drive on freeways at speeds under 100 km/h") is slightly more difficult for men and that item 39-T4 ("If possible, I do not insist on my right of way and make the traffic stop before entering crossroads") is slightly more difficult for women, although this is not significant at the overall level of the questionnaire.

Figure 6: Test results for the two splitting criteria

Reliability. An item separation reliability of 0.96 shows a very good estimate of item hierarchy (Linacre, 2005) or, in other words, it shows that items estimated as more (resp., less) difficult actually are more (resp., less) difficult. The KR-20 value was equal to 0.58 and the value of person separation

reliability was equal to 0.57. This result indicates that items are not very powerful predictors of differences between respondents; this issue will be further discussed in the following paragraphs.

Figures 7 and 8 represent, for each item category, the joint plot of Item Characteristic Curves (ICC). ICCs represent the plot of the logistic equation (1) with estimated β_i. They represent the probability of engaging in a certain behaviour as a function of the position of an individual on the latent trait. In our case this corresponds to the probability of engaging in specific behaviours as a function of a measure of general attitudes towards the environment. ICCs are useful indicators of the most appropriate position for a given item (or behaviour) on the latent trait continuum. In the case of the Rasch model, all the curves have the same shape but vary in terms of position on the latent trait. A 50% probability will correspond to item parameter estimates β_i. Focusing on garbage handling and transport items, we observe that ICC curves overlap for:

- R4 ("I sort paper waste for recycling"), R5 ("I sort glass waste for recycling") and R6 ("I sort plastic waste for recycling");
- T3 ("When possible, I do not use a car for distances less than 30km"), T4 ("If possible, I do not insist on my right of way and make the traffic stop before entering crossroads") and T5 ("I walk, ride or take public transport to go to work/university").

Figure 7: ICC plots for pro-social, garbage handling, power saving and consumerism items

Figure 8: ICC plots for garbage inhibition, activism and volunteering, and transport items

The results show that these items produce the same information. Concerning the other categories, we can observe that items are quite well distributed on the latent dimension.

The Individual-Item Map (Figure 9) shows the individual parameter distribution (upper part of the graphic) and the item parameter value (lower part of the graphic), with items sorted in ascending order of difficulty. When items align, it means that they share the same level of difficulty and, thus, that all except one are superfluous for measurement. It has been confirmed that items related to recycling share the same estimate of difficulty, together with CS5 ("If a friend or a relative had to stay in the hospital for a week or two for minor surgery I would visit him or her"). However, items related to transport that seemed to coincide in Figure 7 gives slightly different parameters values.

Figure 9: Individual-item map of the Rasch Model

Within the Individual-Item Map, when an item is aligned with an individual, this individual is predicted to have a 50% probability of engaging the behaviour. Such an item is said to be *targeted* on the individual. Equivalently, when an item is 1.1 logits more difficult (or easier) than the measure of attitude towards the environment for an individual, this individual has a 25% (or 75%) probability of engaging the behaviour. With these properties in mind, we can draw a few observations from Figure 9:

- first, we can see that at least the eight easiest items are too easy, not targeting anyone, and so they do not contribute anything useful to the GEB measurement;
- second, the existence of gaps between two successive parameters related to item difficulty on the horizontal scale makes it difficult to *fine-tune* estimates for individuals, especially around values of 0.7, 1.2, 1.6, 2.5 and 3 logits; this explains the relatively poor value of the individual separation reliability. This issue will be further discussed in the conclusion.

Finally, Figure 10 gives the individual parameter histogram together with its kernel density plot. This enables us to visualize the distribution of the measured latent trait inside our sample, which is assumed to follow a Gaussian distribution. If the distribution were not Gaussian then estimates would be biased (Seong, 1990; Stone, 1992). In our case the distribution of θ fits a normal distribution of mean $\mu_\theta = 1.14$ and standard deviation $\sigma_\theta = 0.66$ (Jarque-Bera test p-value=0.27, skewness = 0.35, Pearson's kurtosis = 3.00).

Figure 10: Histogram and Kernel density plot of the Rasch Measure

3.1 Concurrent criterion validity

In order to understand if the estimates of the attitude towards the environment measured via the GEB questionnaire are correlated with current transport behaviour, we performed a one-way ANOVA considering three different sub-samples discriminated by the transport mode used by respondents for their most frequent trip. The results show:

- a group of respondents who used 4- or 2-wheeled Private Motorized Vehicles (PMV) as a driver or passenger;
- a group of respondents who used Public Transport (PT: regional train, bus, tram, or metro);
- a group of respondents who used Soft Modes (SM: walking or riding their own or a shared bike).

Of the 131 original respondents, those who used a sequence of different transport modes (a "*trip chain*") were excluded, leaving a sample of 108 respondents. Descriptive statistics for each subgroup and results from the ANOVA are presented in Figure 11 and tables 6 and 7.

Figure 11: Box plots of the measure estimates on the latent trait for each subgroup

Table 6: Descriptive statistics of the measure estimates for each subgroup

Table 7: Results from multiple comparisons Tukey Post-hoc test

Levene's test of homogeneity of variance came out as statistically insignificant (p=0.203), meaning that we were able to reject the null hypothesis, to conclude that subgroups have the same variance, and to carry out the one-way ANOVA, which showed a statistically significant difference between subgroups (df=2, F=6,905, p=.002). A Tukey post-hoc test showed that the GEB score was statistically significantly lower for the group using PMT for their most frequent trip (0.43 ± .81 logit) compared to PT users (0.95 ± .86 logit, p=0.008) or SM users (1.09 ± .56 logit, p=.008). However, there was no statistically significant difference between PT and SM users (p=.799).

4. Discussion and Conclusions

The results reflect the fact that some items were excluded from the analysis and that item parameters were well defined. However, even though the one-dimensionality of the measurement proved to be good enough, the results obtained from our sample showed a violation of local stochastic dependence. This may be due to questions in the GEB questionnaire that are correlated by an independent structural factor (such as car ownership, or provision of differential garbage disposal facilities in an individual's neighbourhood). The influence of independent structural factors on general attitude is not a problem, this aspect being part of Campbell's paradigm of attitude: some behaviours may be more difficult in certain contexts than in others. However, retaining items that are related to each other through independent structural factors is a violation of assumptions made by the formal definition of the Rasch Model. Results concerning transport-related items may indeed be problematic for the Rasch model, especially for the low discrimination they offer. This issue may be partly solved using 2-PL IRT (by producing estimates on the slope of the ICCs, where items with low discrimination are also useful for the measurement). Since transport-related behaviours are an important part of pro-environmental behaviours, we decided to keep them within the group of items retained for the analysis. However, rewording transport-related item or removing some of them in future applications may lead to a better quality of the measuring instrument. Furthermore, we saw that estimates of an individual's pro-environmental attitude cannot be fine-tuned: at least eight items are of no use in producing estimates, and thus can be disregarded, because the difficulty of engaging in these behaviours is too low, and some items of intermediate to high difficulty are missing. Filling

the gaps between difficulties (as evidenced in Figure 9) could lead to better estimates of pro-environmental behaviour on the part of an individual. To this end, an improved GEB could include other forms of pro-environmental behaviour, such as, for example, diet-related behaviours (e.g. limitation of meat and/or dairy consumption), use of technology (e.g. number of smartphones, tablets and computer bought per year or their presence and typical usage in households), holiday-related travel behaviours (limitation of airplane trips for personal reasons) and offsetting emissions (e.g. through the UN Carbon Offset Platform).

All this considered, while it remains a neglected method for measuring attitude within the framework of Item Response Theory, the General Ecological Behaviour questionnaire may be considered a useful tool for assessing the pro-environmental behaviour of travellers: there are no discrepancies between pro-social behaviour and observed pro-environmental behaviour, in line with the current literature in which the two behaviours are strongly correlated. We suggest that a continuity exists between social-value orientation and pro-environmental attitudes; one-dimensionality, item reliability, and the absence of simple differential item functioning are all good indicators of a model that functions well.

GEB thus has interesting potential that has been investigated by other researchers in Switzerland and in Sweden. It is useful to compare results to ascertain whether different cultural contexts can play a role in determining ecological behaviour. One of the advantages of using a Rasch Scale is that, even though some items may differ between studies, results are still comparable in terms of the calibration of a specific item's difficulty, making it possible to directly examine the difference in item calibration across different samples and contexts, including various cultural–linguistic settings and translations. A powerful key and characteristic of a definitive Rasch model (once assumptions have been tested for, items have been calibrated on representative populations, and the questionnaire has been consolidated and validated externally and internally) is that the results are independent of the sample as well as the instrument (Bond and Fox, 2013; Rasch, 1968). Table 8 reports, for a subsample of behaviours, a comparison of difficulties between three populations: Italian (N=131), Swiss (N=445) and Swedish (N=247).

The Swedish sample consisted of 247 adults, the questionnaire was communicated by e-mail in Swedish and participants could fill it out at their own convenience on a voluntary basis (response rate of 51%). The participants' (48.6% male) median age was 42.0 years (M:43.8, range: 17 to 75).

The Swiss sample consisted of 445 members of two associations. The first association exists to promote a transportation system that has as little negative impact on humans and nature as possible, while the second represents the interests of automobile drivers and advocates such things as maintaining roads properly, allowing higher speed limits on freeways, and fighting gasoline tax

increases. The questionnaire in German was communicated via e-mail and participants could fill it out at their own convenience on a voluntary basis (response rate of 82.0%). The participants' (62.5% male) median age was 45.5 years (M: 46.6, range: 20 to 82). Members of this second association were less well represented in the sample (25.8%) than members of the first (74.2%), and so the Swiss sample would appear to be biased toward more ecologically concerned participants.

Given the differences in sample composition and the way that the questionnaire was administered in each case, any differences that these studies might suggest between Italy, Sweden and Switzerland need to be treated with caution. Given the lack of statistical significance of the comparison, its interest lies solely in the potential of the method to measure behavioural attitudes based on the Campbell's definition of "behavioural disposition" (Campbell, 1963) and to interpret this measure within different socio-technical and cultural contexts. However, the time that elapsed between the different studies is a drawback for direct comparison inasmuch as, in the last 15 years, environmental awareness has greatly increased together with effective knowledge about behaviours that are perceived and/or understood to environmentally friendly.

Values for the Swiss and Swedish samples are taken from Kaiser and Biel (2000); the mean GEB score for the Swiss participants was 1.42 (SD=0.94, N=443) and for the Swedish participants 0.61 (SD=0.88, N=246). Thus, the Italian participants were globally more committed to pro-environmental behaviour than the Swedish, but less so than the Swiss participants.

Table 8: Comparison of item difficulties between three samples from different countries.

From Table 8 it can be seen that some behaviours are absolutely easier to perform in Italy than in Switzerland and Sweden:

- 10 - I bring unused medicine back to the pharmacy.
- 11 - I sort paper waste for recycling.
- 16 - In winter, I keep the heating on so that I do not have to wear a sweater.
- 18 - I wait until I have a full load before doing my laundry.
- 29 - If I am offered a plastic bag in a store, I will always take it.
- 34 - In the past, I have pointed out to someone his or her un-ecological behaviour.
- 38 - When possible, I do not use a car for distances less than 30km.

Others are, on the contrary, absolutely more difficult to perform in Italy than in Switzerland or Sweden:

- 8 - I put dead batteries in the garbage.
- 19 - In winter, I leave the windows wide open for long periods of time to let in fresh air.

- 20 - I wash dirty clothes without pre-washing.
- 21 - I use fabric softener with my laundry.
- 24 - I use specific cleaners for different rooms rather than an all-purpose cleaner.
- 25 - I use phosphate-free laundry detergent
- 30 – For shopping, I prefer paper bags to plastic ones.
- 31 - Usually, I buy water in returnable bottles.
- 33 - I am a member of an environmental organization.
- 35 - I sometimes contribute financially to environmental organizations.
- 37 - I usually drive on freeways at speeds lower than 100km/h.

The above comparison shows the potential of GEB in understanding people's attitudes towards environmental issues and how such information can be used in better tailoring public policies in sectors such as environment, energy and transport. Clearly, further research may be needed to remove, replace and add some items that could produce better estimates for individuals; indeed, since the GEB questionnaire was first developed (Kaiser, 1998), some alternative items have been used (Kaiser and Wilson, 2000; Kaiser et al., 2007). We consider that this measure of attitude is far more convincing than the traditional measure; its mathematical model is well defined and satisfies the requirements that we expect from a measurement tool, including:

- specific objectivity: the fact that the measure is sample-free for the agents and test-free for the items;
- additive measurement: a property desirable for any extensive measure that translates an empirical system into a numerical system (see Borsboom and Scholten, 2008 and Karabatsos, 2001 for further discussion);
- being hypocrisy-insensitive: here, *hypocrisy* means declaring a certain general attitude without engaging in corresponding behaviours. We measure self-reported relevant behaviour instead of introspective specific attitudes towards behaviours.

In addition, the GEB's one-dimensional scale is a parsimonious alternative (Arnold et al., 2018), with scores stable over time (Kaiser et al., 2014) that also addresses the criterion of validity. Arnold et. al (2018), according to Steg and Sievers (2000), observe that pro-environmental behaviour should not be limited to a person's intent, but should refer to that person's actual environmental impact. Even though the use of the GEB was motivated by an intention-oriented approach, its validity as an impact-oriented measure of commitment has been partially demonstrated: Kaiser et al. (2003), using Life Cycle Analysis (LCA) on behaviours investigated by the GEB questionnaire, showed that 46 out of 52 items tested were validated as being effectively less damaging to the environment than their alternatives. Arnold et al. (2018) found a negative correlation between self-reported pro-

environmental behaviour (assessed via the GEB) and electricity consumption. Further investigation is needed to correlate the GEB score with an individual's overall consumption of energy and materials, and in particular, with travel behaviour and mobility patterns. Specifically, it could be interesting to consider holiday trips and long-distance travel to understand if GEB correlates with the avoidance of superfluous trips or the self-limitation of travel modes with a heavy carbon footprint (in particular, flying) without deteriorating the uni-dimensionality of the measure. However, flying is a typical example of cognitive dissonance where people find exogenous reasons to justify their travel.

In the literature, a wide debate has been going on for years about the measure of attitudes as well as about the attitude-behaviour gap. Pronello and Gaborieau (2018) observed that a deeper understanding of travel behaviour requires a redefinition of the concepts (attitude and habits) and that "attitude towards a given object is only person-dependent and reflects itself through a set of behaviours transitively ordered according to the level of difficulty (cost) to perform them: in practice, attitudes are measured by means of what people do, not of what they say". This is the Campbell measure of attitudes mentioned above that we have used and applied to the transport sector. We have used the GEB score as a measure of the general attitude towards the environment within existing behavioural theories (Ajzen's Theory of Planned Behaviour, 1991; Schwartz' Norm Activation Theory, 1970; Triandis' Theory of Interpersonal Behaviour, 1977) to explain modal choice and travel behaviour.

A wider use of the GEB questionnaire by practitioners could make it easier to identify good practices and to devise effective public policies and marketing campaigns. Campbell's paradigm broadens policy makers' range of action by enabling them to push people towards a more sustainable lifestyle, either by reducing the difficulties linked to specific behaviours or by increasing people's motivation to engage in them. Moreover, the specific construction of a Rasch scale for measurements allows the development of adaptive surveys that can be used to make questionnaires shorter.

The present paper has certain limitations. First, although the sampling plan was designed to include travel behaviours as diverse as possible, results are not representative of the general population living in the Piedmont region of Italy. In particular, the calibration of parameters may be biased by a wealth factor, that can help or hinder people with regard to specific pro-environmental behaviours. Second, as mentioned above, more research is needed to test the association between the Rasch measure of environmental attitude as measured by the GEB questionnaire and the actual impact of travel behaviours, in terms of resources consumed as well as pollutants and CO_2 emissions. Finally, in order to take into account direct social influences, it would be interesting to research the trade-off mechanism that occurs within households composed of individuals with differing pro-environmental sensibilities.

To better bridge the gap between personal intent and personal environmental impact and solve the attitude-behaviour gap, we have selected the items that should be included in the travel surveys to capture a reliable and stable attitude towards pro-environmental behaviour. Our research, as mentioned in the introductory section, has continued to focus on a) evaluating GEB validity as a psychometric measure of attitude, used in a model to forecast modal choices, and b) using the GEB measure, together with other psycho-social variables, to cluster our sample in groups that respond differently to the deployment of Multimodal Advanced Traveller Information Systems. Finally, we have designed an augmented and polytomous version of the questionnaire to collect data on a large population of the Piedmont region (in Italy). The data are under analysis and they will be used to fit an IRT model on a big sample and provide a further test of the proposed method.

The ultimate goal is to design tailored policies to make travel more sustainable. Public policies targeted to education – to affect people's beliefs – or advertising – to affect people's emotional response – can potentially have direct impacts on individual behavioural responses. A simple, effective measure of pro-environmental behaviour would allow public authorities to quantify the effectiveness of adopted policy.

Acknowledgments

This research has been carried out within the project "OPTICITIES. Optimise Citizen Mobility and Freight Management in Urban Environments", funded by the European Commission in the VII FP, Challenge 2. Safe and seamless mobility. SST.2013.3-1. Managing integrated multimodal urban transport network FP7-SST-2013-RTD-1.

Conflict of interest

On behalf of all authors, the corresponding author states that there is no conflict of interest.

Authors contribution

J-B. Gaborieau and C. Pronello conceived the idea presented in this paper. The experiment was planned and designed by both authors. The Literature Search was made by J-B-Gaborieau while the methodology was written in part by C. Pronello and in part by J-B. Gaborieau. The analysis was

carried out by J-B. Gaborieau, who wrote a first draft while the final manuscript was written by C. Pronello. C. Pronello supervised the entire work providing feedbacks continuously, helped shaping the Manuscript and revised it.

References

Ajzen, I. (1988). Attitudes, personality, and behavior. Chicago: Dorsey Press.

Ajzen, I. (1991). The theory of planned behavior. Organizational behavior and human decision processes, 50(2), 179-211.

Andersen, E. B. (1973). A goodness of fit test for the Rasch model. Psychometrika, 38(1), 123-140.

Andrich, D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms? Medical care, I7-I16.

Arnold, O., Kibbe, A., Hartig, T., & Kaiser, F. G. (2018). Capturing the environmental impact of individual lifestyles: evidence of the criterion validity of the general ecological behavior scale. Environment and Behavior, 50(3), 350-372.

Axelrod, L. J., & Lehman, D. R. (1993). Responding to environmental concerns: What factors guide individual action? Journal of environmental psychology, 13(2), 149-159.

Bamberg, S., & Möser, G. (2007). Twenty years after Hines, Hungerford, and Tomera: A new meta-analysis of psycho-social determinants of pro-environmental behaviour. Journal of environmental psychology, 27(1), 14-25.Bond, T. G., & Fox, C. M. (2013). Applying the Rasch model: Fundamental measurement in the human sciences. Psychology Press.

Bond, T. G., & Fox, C. M. (2013). Applying the Rasch model: Fundamental measurement in the human sciences. Psychology Press.

Borsboom, D. & Scholten, A. (2008). The Rasch Model and Conjoint Measurement Theory from the Perspective of Psychometrics. Theory & Psychology 18. 111-117. 10.1177/0959354307086925.

Campbell, D. T. (1963). Social attitudes and other acquired behavioral dispositions. In Koch, S. (1963). Psychology: A study of a science. Volume 6. Investigations of man as socius: Their place in psychology and the social sciences. (pp. 94-172). New York, NY, US: McGraw-Hill, xii, 791 pp.

Christensen, K. B., Bjorner, J. B., Kreiner, S., & Petersen, J. H. (2002). Testing unidimensionality in polytomous Rasch models. Psychometrika, 67(4), 563-574.

Corral-Verdugo, V. (1997). Dual 'realities'of conservation behavior: self-reports vs observations of re-use and recycling behavior. Journal of environmental psychology, 17(2), 135-145.

Embretson, S. E. (1996). Item response theory models and inferential bias in multiple group comparisons. Applied Psychological Measurement, 20, 201-212.

European Commission. (2005). The attitudes of European citizens towards environment. Special Eurobarometer 217/ Wave 62.1 TNS Opinion & Social.

Fejer, S. (1989). Aspekte zur Änderbarkeit von Verbraucherver-halten durch Social-Marketing-Eine empirische Analyse eines konkreten Beispiels [Some aspects of consumerism modification by social-marketing-An empirical analysis of a real life example]. Unpublished master's thesis, University of Duisburg, Duisburg, Germany.

Fejer, S., & Stroschein, F. R. (1991). „Die Ableitung einer Guttman-Skala für sozial-und ökologiebewußtes Verhalten ", planung und analyse. Heft, 1, 5-12.

Fisher, W. P. (1997). Physical Disability Construct Convergence Across Instruments: Towards a Universal Mentric. Journal of Outcome Measurement, 1(2), 87-113.

Fuj, E. T., Hennessy, M., & Mak, J. (1985). An evaluation of the validity and reliability of survey response data on household electricity conservation. Evaluation Review, 9(1), 93-104.

Gaborieau, J.B. (2016). Evaluation of the potential modal shift induced by the use of a real time multimodal navigator: psycho-social study of travel behaviour and attitude (Doctoral dissertation, Politecnico di Torino). Retrieved from https://iris.polito.it/

Gamba, R. J., & Oskamp, S. (1994). Factors influencing community residents' participation in commingled curbside recycling programs. Environment and behavior, 26(5), 587-612.

Gärling, T., Fujii, S., Gärling, A., & Jakobsson, C. (2003). Moderating effects of social value orientation on determinants of proenvironmental behavior intention. Journal of environmental psychology, 23(1), 1-9.

Green-Demers, I., Pelletier, L. G., & Menard, S. (1997). The impact of behavioural difficulty on the saliency of the association between self-determined motivation and environmental behaviours. Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 29(3), 157.

Guttman, L. A. (1950). The basis for scalogram analysis. In S. Stouffer (Ed.), Measurement and prediction (vol. IV, pp. 60-90).

Haustein, S., & Nielsen, T. A. S. (2016). European mobility cultures: A survey-based cluster analysis across 28 European countries. Journal of Transport Geography, 54, 173-180.

Harland, P., Staats, H., & Wilke, H. A. (1999). Explaining proenvironmental intention and behavior by personal norms and the Theory of Planned Behavior 1. Journal of applied social psychology, 29(12), 2505-2528.

Heimlich, J. E., & Ardoin, N. M. (2008). Understanding behavior to understand behavior change: A literature review. Environmental education research, 14(3), 215-237.

Hirst, E., & Goeltz, R. (1985). Evaluation of residential energy conservation programs in Minnesota. Evaluation Review, 9(3), 329-347.

Hunecke, M.; Haustein, S.; Grischkat, S.; Böhler, S. (2007) Psychological, sociodemographic, and infrastructural factors as determinants of ecological impact caused by mobility behavior. J. Environ. Psychol. 2007, 27, 277–292.

Hunecke, M., Haustein, S., Böhler, S., & Grischkat, S. (2010). Attitude-based target groups to reduce the ecological impact of daily mobility behavior. Environment and behavior, 42(1), 3-43.

Jaspen, N. (1946). Serial correlation. Psychometrika, 11(1), 23-30.

Kaiser, F. G. (1998). A general measure of ecological behavior 1. Journal of applied social psychology, 28(5), 395-422.

Kaiser, F. G., Brügger, A., Hartig, T., Bogner, F. X., & Gutscher, H. (2014). Appreciation of nature and appreciation of environmental protection: How stable are these attitudes and which comes first? European Review of Applied Psychology/Revue Européenne de Psychologie Appliquée, 64, 269-277.

Kaiser, F. G., & Biel, A. (2000). Assessing general ecological behavior: A cross-cultural comparison between Switzerland and Sweden. European Journal of Psychological Assessment, 16(1), 44.

Kaiser, F. G., & Byrka, K. (2015). The Campbell paradigm as a conceptual alternative to the expectation of hypocrisy in contemporary attitude research. The Journal of social psychology, 155(1), 12-29.

Kaiser, F. G., Byrka, K., & Hartig, T. (2010). Reviving Campbell's paradigm for attitude research. Personality and Social Psychology Review, 14(4), 351-367.

Kaiser, F. G., Doka, G., Hofstetter, P., & Ranney, M. A. (2003). Ecological behavior and its environmental consequences: A life cycle assessment of a self-report measure. *Journal of environmental psychology*, *23*(1), 11-20.

Kaiser, F. G., Oerke, B., & Bogner, F. X. (2007). Behavior-based environmental attitude: Development of an instrument for adolescents. Journal of Environmental Psychology, 27(3), 242-251.

Kaiser, F. G., & Wilson, M. (2000). Assessing People's General Ecological Behavior: A Cross-Cultural Measure 1. Journal of Applied Social Psychology, 30(5), 952-978.

Karabatsos, G. (2000). A critique of Rasch residual fit statistics. Journal of applied measurement, 1(2), 152-176.
Karabatsos, G. (2001). The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. Journal of applied measurement, 2(4), 389-423.
Koller, I., & Hatzinger, R. (2013). Nonparametric tests for the Rasch model: explanation, development, and application of quasi-exact tests for small samples. Interstat, 11, 1-16.
Korfiatis, K.J., Hovardas, T., Pantis, J.D., 2004. Determinants of Environmental Behavior in Societies in Transition: Evidence from five European Countries. Population and Environment 25(6), 563–584.
Kroesen, M., & Chorus, C. (2018). The role of general and specific attitudes in predicting travel behavior-A fatal dilemma? Travel Behaviour and Society, 10, 33-41.
Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. Psychometrika, 2(3), 151-160.
Leonard-Barton, D. (1981). Voluntary simplicity lifestyles and energy conservation. Journal of Consumer Research, 8(3), 243-252.
Linacre, J. M. (1994a). PROX with missing data, or known item or person measures. Rasch Meas Trans, 8(3), 378.
Linacre, J. (1994b). Sample Size and Item Calibration Stability. Rasch Measurement Transactions. 7. 328.
Linacre, J. M. (2005). WINSTEPS Rasch measurement [Computer software]. Chicago: winsteps.com.
Linacre, J. M. (2006). Misfit diagnosis: Infit outfit mean-square standardized. Help for Winsteps Rasch Measurement Software
Linacre, J.M. (2008). The Expected Value of a Point-Biserial (or Similar) Correlation. Rasch Measurement Transactions, 22:1 p. 1154
Lord, F. M. (1983). Small N justifies Rasch model. In New horizons in testing (pp. 51-61). Academic Press.
Magno, C. (2009). Demonstrating the difference between classical test theory and item response theory using derived test data.
Mair, P., & Hatzinger, R. (2007a). Extended Rasch modeling: The eRm package for the application of IRT models in R. Journal of Statistical Software, 20(9), 1-20.
Mair, P., & Hatzinger, R. (2007b). CML based estimation of extended Rasch model with the eRm packahe in R. Psychology Science, 49(1), 26-43.
Mair, P., Hatzinger, R., & Maier M.J. (2018). eRm: extended Rasch Modeling 0.16-2. http://erm.r-forge.r-project.org/
Maloney, M. P., & Ward, M. P. (1973). Ecology: Let's hear from the people: An objective scale for the measurement of ecological attitudes and knowledge. American psychologist, 28(7), 583.
Martin-Löf, P. (1970). Statistiska modeller, anteckningar från seminarier läsåret 1969–70, utarbetade av R. Sundberg, Institutionen för matematisk statistik, Stockholms universitet.
Olsson, U., Drasgow, F., & Dorans, N. J. (1982). The polyserial correlation coefficient. Psychometrika, 47(3), 337-347.
Otto, S., Kröhne, U., & Richter, D. (2018). The dominance of introspective measures and what this implies: The example of environmental attitude. PLOS ONE, 13(2), e0192907. https://doi.org/10.1371/journal.pone.0192907
Otto, S., Neaman, A., Richards, B., & Marió, A. (2016). Explaining the ambiguous relations between income, environmental knowledge, and environmentally significant behavior. Society & Natural Resources, 29, 628-632.
Peter, J. P. (1979). Reliability: A review of psychometric basics and recent marketing practices. Journal of marketing research, 6-17.
Ponocny, I. (2001). Nonparametric goodness-of-fit tests for the Rasch model. Psychometrika, 66(3), 437-459.

Pronello, C., & Camusso, C. (2011). Travellers' profiles definition using statistical multivariate analysis of attitudinal variables. journal of Transport Geography, 19(6), 1294-1308.
Pronello, C., & Gaborieau, J. B. (2018). Engaging in Pro-Environment Travel Behaviour Research from a Psycho-Social Perspective: A Review of Behavioural Variables and Theories. Sustainability (2071-1050), 10(7).
Ramsey, J. M. (1993). The effects of issue investigation and action training on eighth-grade students' environmental behavior. The Journal of Environmental Education, 24(3), 31-36.
Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danish Institute for Educational Research.
Rasch, G. (1968). A mathematical theory of objectivity and its consequences for model construction. In Report from European Meeting on Statistics, Econometrics and Management Sciences, Amsterdam.
Revelle, W. (2011). An introduction to psychometric theory with applications in R, (http://personality-project.org/r/book/)
Schwartz, S. H. (1970). Moral decision making and behavior. *Altruism and helping behavior*, 127-141.
Seong, T.-J. (1990). Sensitivity of Marginal Maximum Likelihood Estimation of Item and Ability Parameters to the Characteristics of the Prior Ability Distributions. *Applied Psychological Measurement, 14*(3), 299–311.
Smith, A. B., Rush, R., Fallowfield, L. J., Velikova, G., & Sharpe, M. (2008). Rasch fit statistics and sample size considerations for polytomous data. BMC Medical Research Methodology, 8(1), 33.
Smith, R. M., & Suh, K. K. (2003). Rasch fit statistics as a test of the invariance of item parameter estimates. Journal of applied measurement, 4(2), 153-163.
Smith, R. M., Schumacker, R. E., & Bush, M. J. (1998). Using item mean squares to evaluate fit to the rasch model. Journal of Outcome Measurement, 2(1), 66-78.
Soysal, S., Akın A. Ç. & İnal, H. (2016). Impact Of Missing Data On Rasch Model Estimations. Turkish Online Journal of Educational Technology. 2016.
Steg, L., & Gifford, R. (2005). Sustainable transportation and quality of life. Journal of transport geography, 13(1), 59-69.
Steg, L., & Sievers, I. (2000). Cultural theory and individual perceptions of environmental risks. Environment and behavior, 32(2), 250-269.
Stern, P. C. (2000). Toward a coherent theory of environmentally significant behavior. Journal of Social Issues, 56, 407-424.
Stone, C. A. (1992). Recovery of marginal maximum likelihood estimates in the two-parameter logistic response model: An evaluation of MULTILOG. Applied Psychological Measurement, 16(1), 1-16.
Triandis, H. C. (1977). *Interpersonal behavior*. Monterey, CA: Brooks/Cole Publishing Company.
Van Liere, K. D., & Dunlap, R. E. (1978). Moral Norms and Environmental Behavior: An Application of Schwartz's Norm-Activation Model to Yard Burning 1. Journal of Applied Social Psychology, 8(2), 174-188.
Van Vugt, M., Meertens, R. M., & Van Lange, P. A. (1995). Car Versus Public Transportation? The Role of Social Value Orientations in a Real-Life Social Dilemma. Journal of Applied Social Psychology, 25(3), 258-278.
Verhelst, N. D. (2008). An efficient MCMC algorithm to sample binary matrices with fixed marginals. Psychometrika, 73(4), 705-728.
Vining, J., & Ebreo, A. (1992). Predicting recycling behavior from global and specific environmental attitudes and changes in recycling opportunities 1. Journal of applied social psychology, 22(20), 1580-1607.

Wang, W. C., & Chen, C. T. (2005). Item parameter recovery, standard error estimates, and fit statistics of the WINSTEPS program for the family of Rasch models. Educational and Psychological Measurement, 65(3), 376-404.
Wilson, E. B., & Hilferty, M. M. (1931). The distribution of chi-square. Proceedings of the National Academy of Sciences, 17(12), 684-688.
Wright, B. D., & Douglas, G. A. (1977). Best procedures for sample-free item analysis. Applied Psychological Measurement, 1(2), 281-295.
Wright, B. D., Linacre, J. M., Gustafson, J. E., & Martin-Lof, P. (1994). Reasonable mean-square fit values. Rasch measurement transactions, 8(3), 370.
Wright, B., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. Educational and Psychological measurement, 29(1), 23-48.
Wright, M., & Klÿn, B., (1998). Environmental Attitude – Behaviour Correlations in 21 Countries. Journal of Empirical Generalisations in Marketing Science 3, 42–60.

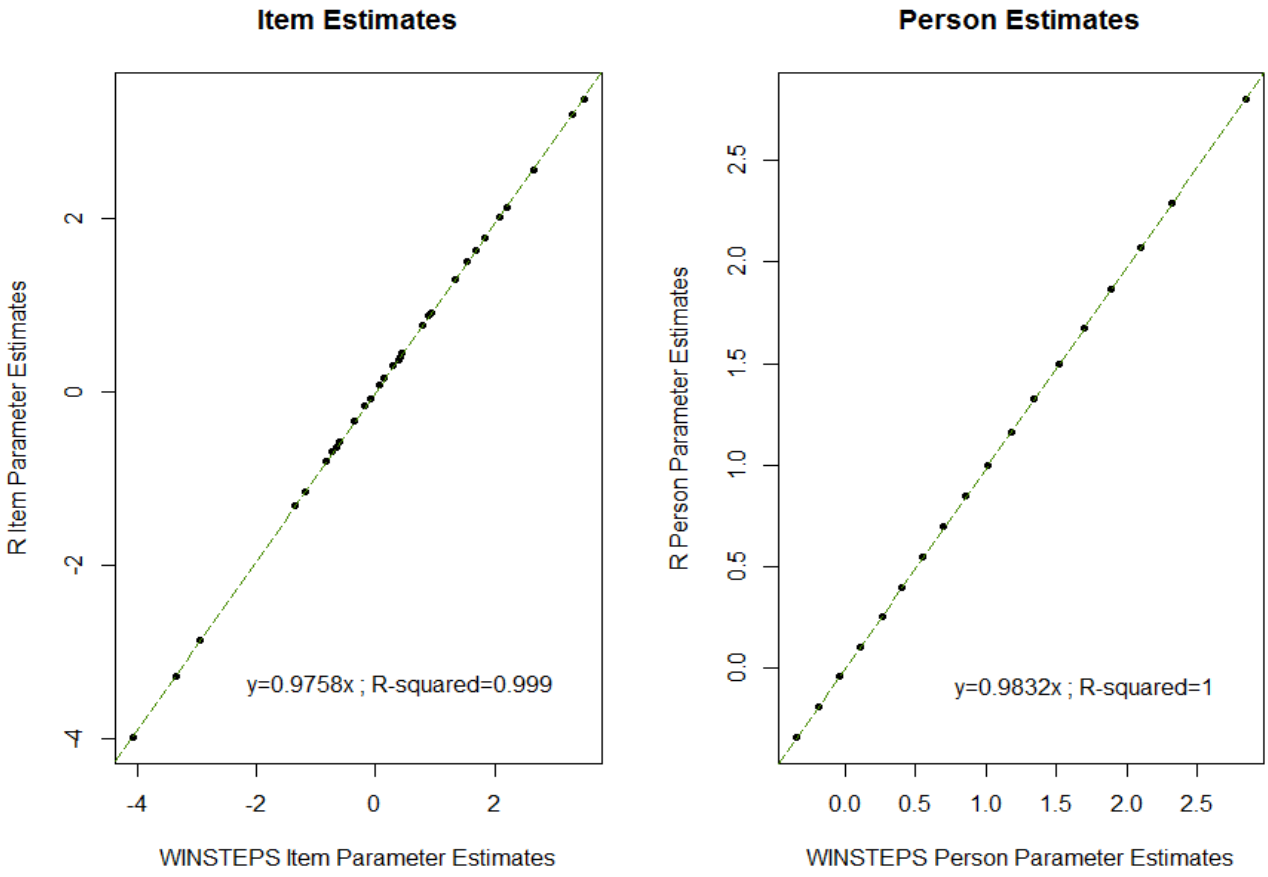

Figure 1: Winsteps vs eRm estimates of item and person parameters

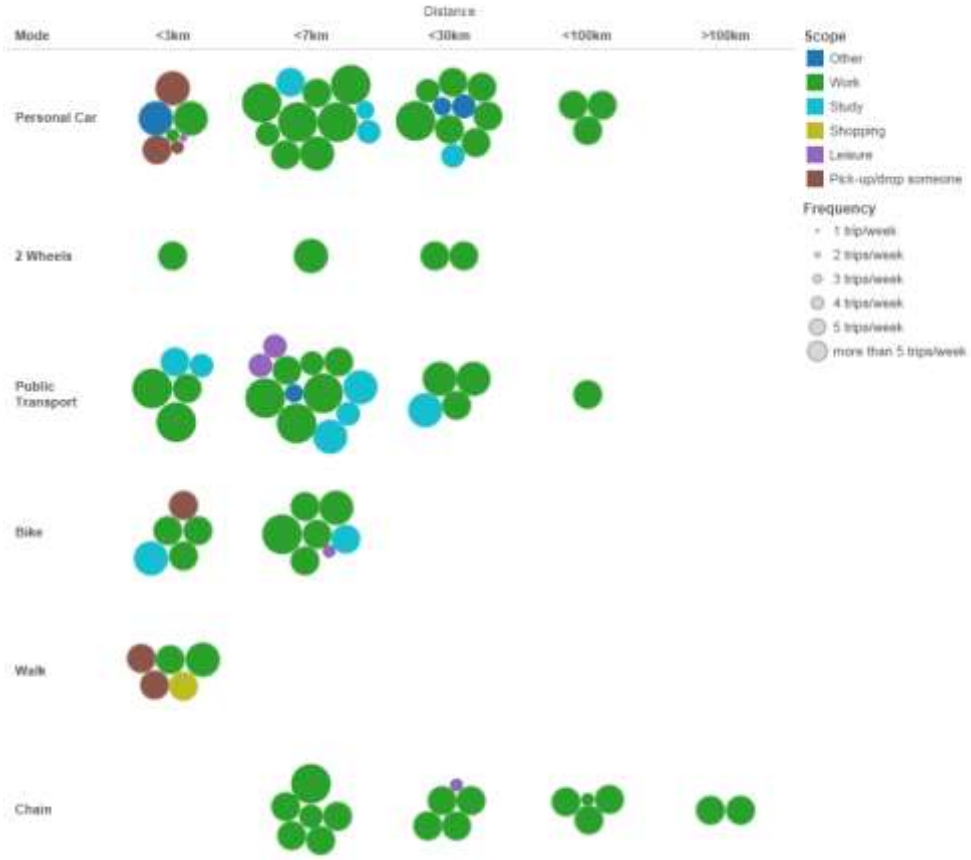

Figure 2: Attributes of the most frequent trips (mode, distance, purpose and frequency)

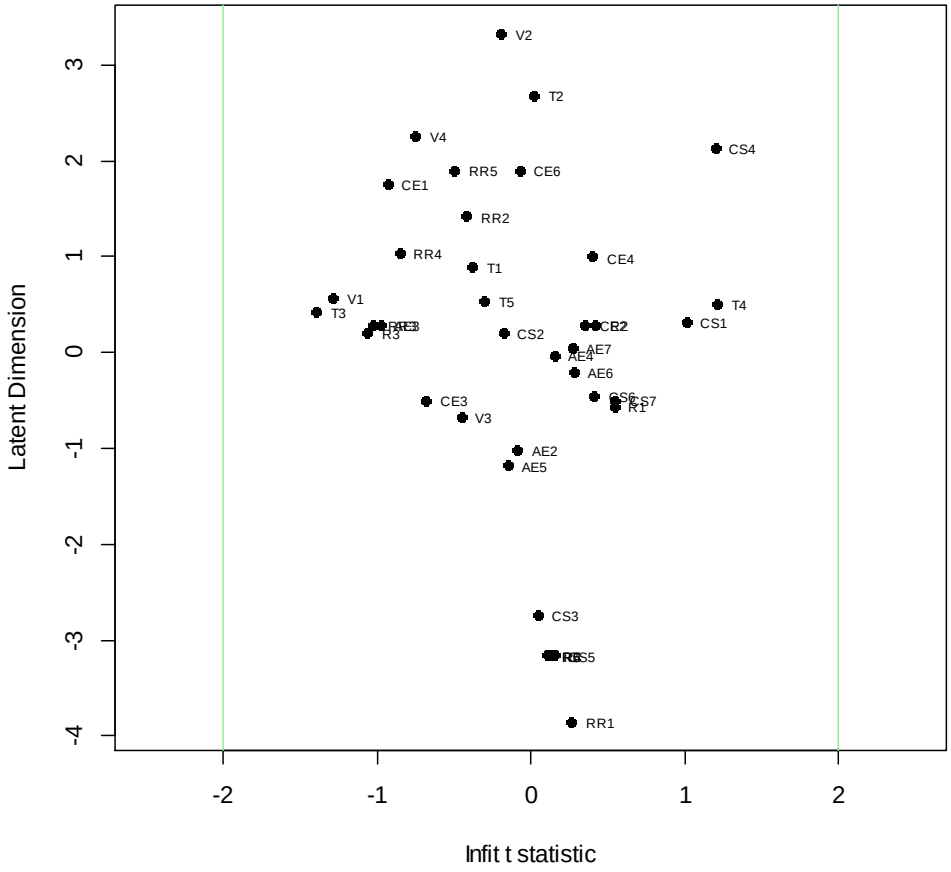

Figure 3: Item map of INFIT statistics for final item selection

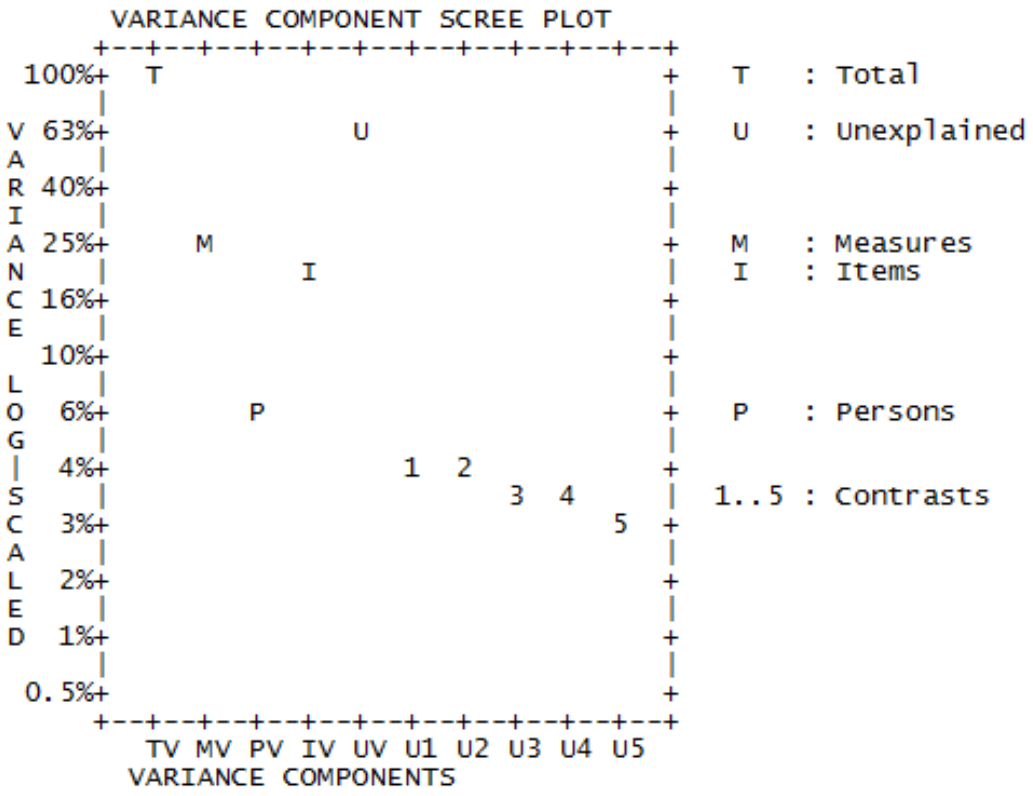

Figure 4: Scree plot of the PCA variance component

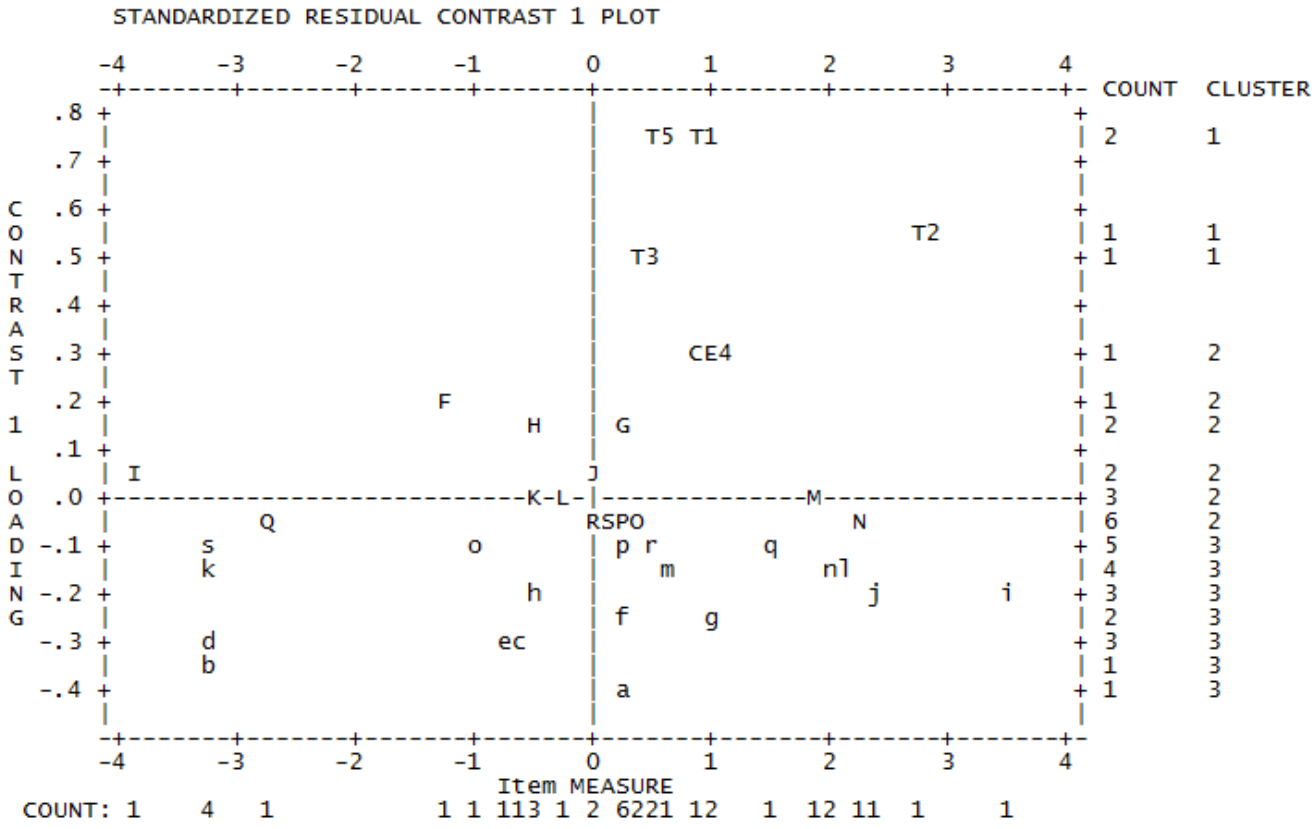

Figure 5: Item loadings on the first constrast

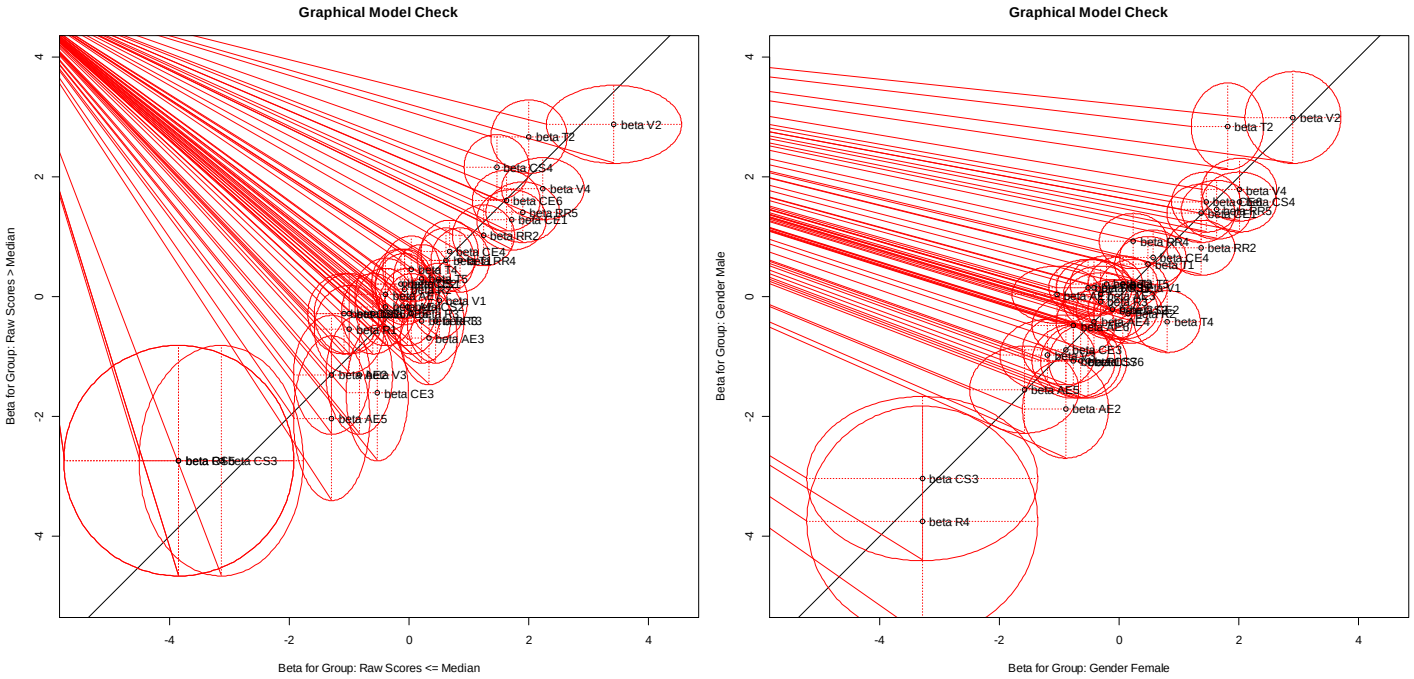

Figure 6: Test results for the two splitting criteria.

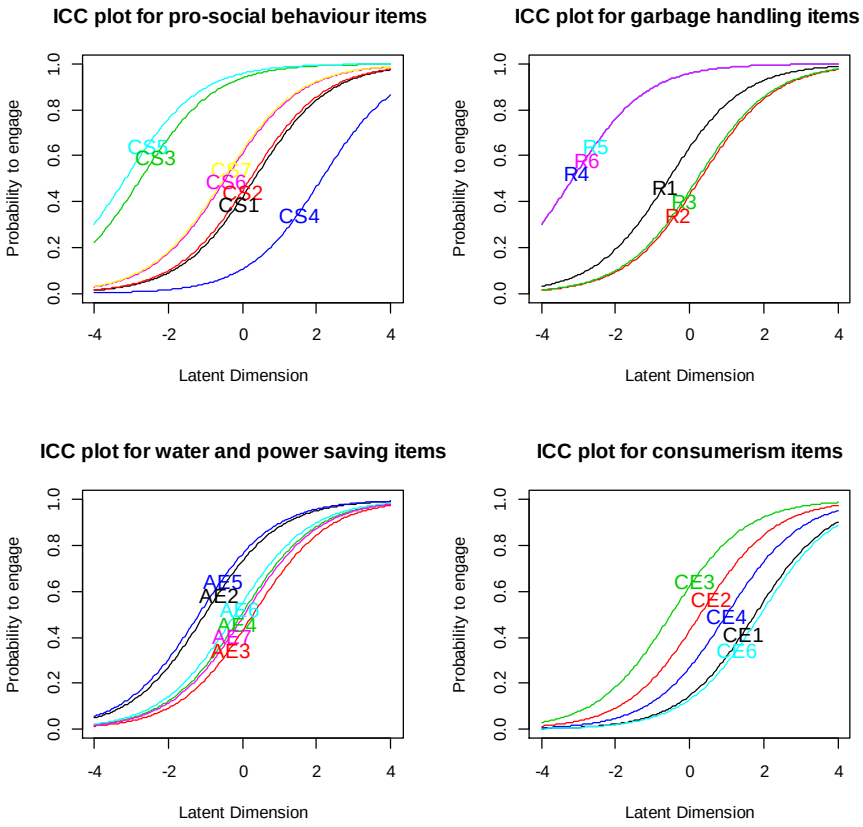

Figure 7: ICC plots for pro-social, garbage handling, power saving and consumerism items

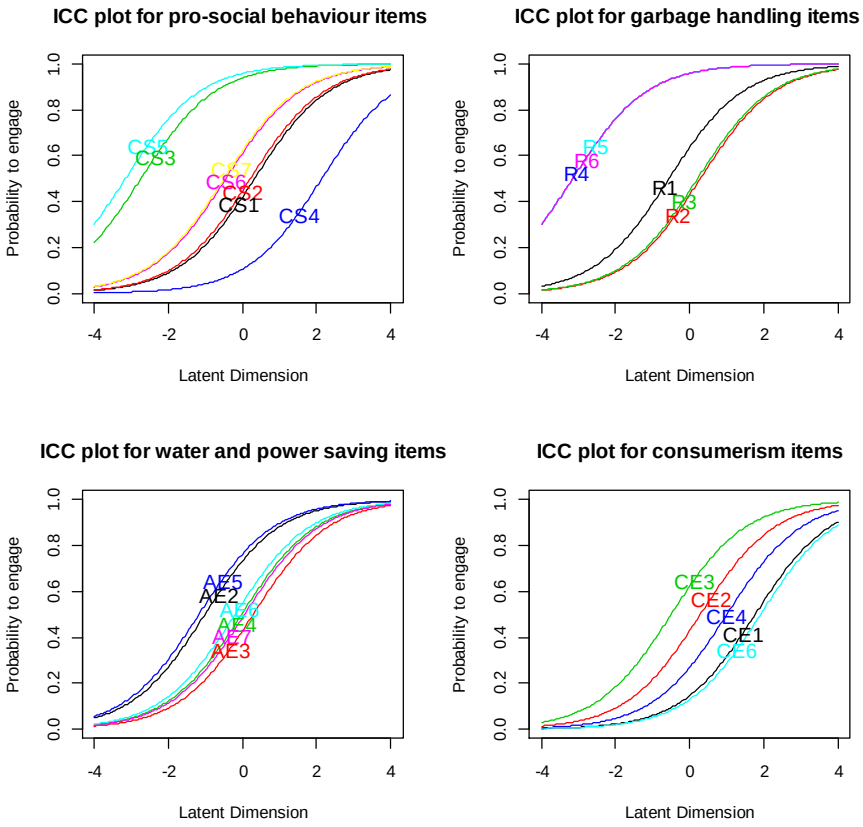

Figure 7: ICC plots for pro-social, garbage handling, power saving and consumerism items

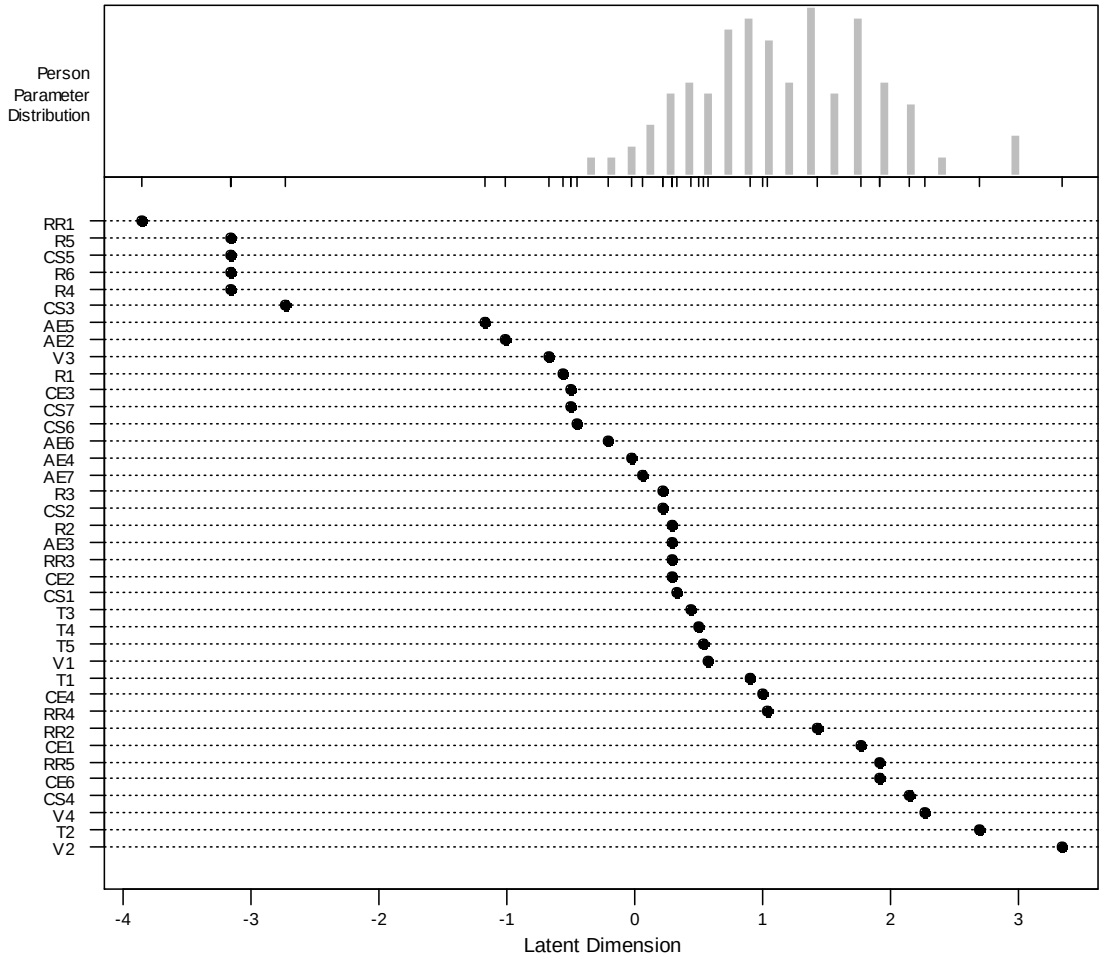

Figure 9: Individual-item map of the Rasch Model

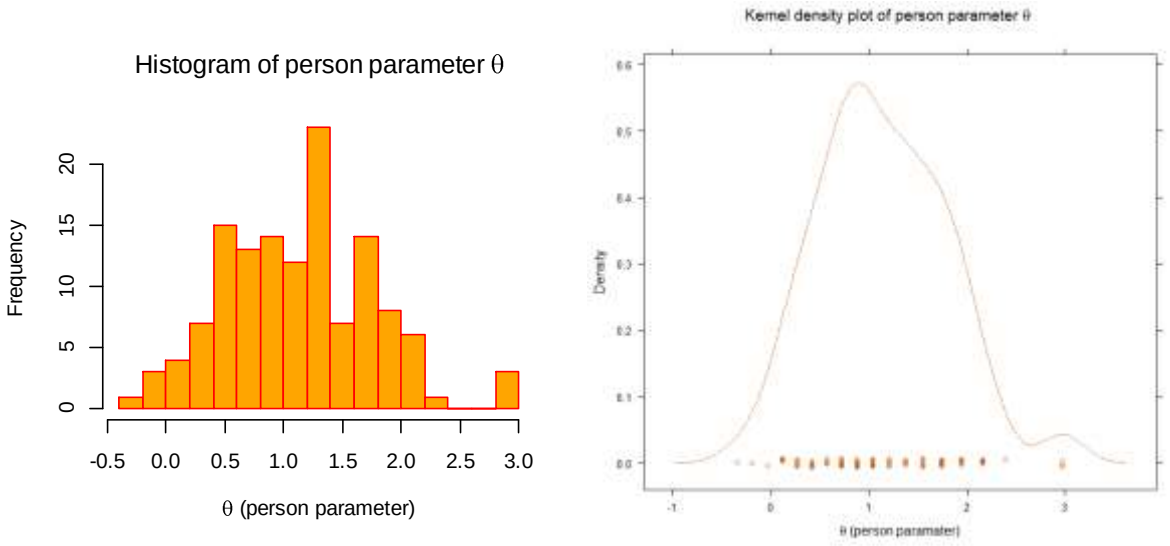

Figure 10: Histogram and Kernel density plot of the Rasch Measure

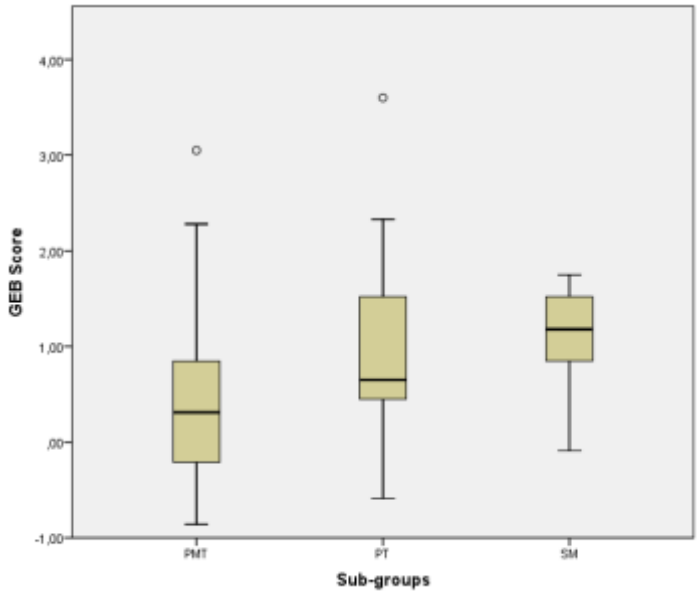

Figure 11: Box plots of the measure estimates on the latent trait for each subgroup

Table 1: Structure of the GEB questionnaire

N°	Item description	Item Code
	Pro-social behaviour	
1(=).	Sometimes I give money to panhandlers.	CS1
2(=).	From time to time I give money to charity.	CS2
3(=).	If an elderly or disabled person enters a crowded PT vehicle, I offer him/her my seat.	CS3
4(=).	If I were an employer, I would not hesitate hiring a person previously convicted of crime.	CS4
5(=).	If a friend or a relative had to stay in the hospital for a week or two for minor surgery I would visit him or her.	CS5
6(=).	Sometimes I ride public transport without paying a fare.	CS6 (-)
7(=).	I would feel uncomfortable if people from another ethnicity were my neighbours.	CS7 (-)
	Ecological garbage handling	
8(=).	I put dead batteries in the garbage.	R1 (-)
9(=).	I make use of rechargeable batteries.	R2
10(=).	I bring unused medicine back to the pharmacy.	R3
11(*).	I sort paper waste for recycling.	R4
12(*).	I sort glass waste for recycling.	R5
13(+).	I sort plastic waste for recycling.	R6
	Water and power saving	
14(*).	Before taking a shower, I let the water run so it gets to the temperature I want.	AE1 (-)
15(=).	I prefer to shower rather than take a bath.	AE2
16(=).	In winter, I keep the heat on so that I do not have to wear a sweater.	AE3 (-)
17(*).	I turn off the heat at night.	AE4
18(=).	I wait until I have a full load before doing my laundry.	AE5
19(=).	In winter, I leave the windows wide open for long periods of time to let in fresh air.	AE6 (-)
20(=).	I wash dirty clothes without pre-washing.	AE7
	Ecologically aware consumerism	
21(=).	I use fabric softener with my laundry.	CE1 (-)
22(=).	If there are insects at home, I kill them with a chemical insecticide.	CE2 (-)
23(=).	I use a chemical air freshener in my bathroom.	CE3 (-)
24(=).	I use specific cleaners for different rooms rather than an all-purpose cleaner.	CE4 (-)
25(=).	I use phosphate-free laundry detergent.	CE5

26(*).	I always look to buy vegetables from biological agriculture.	CE6
	Garbage inhibition	
27(=).	I re-use plastic grocery bags.	RR1
28(=).	I sometimes buy beverages in cans.	RR2 (-)
29(=).	If I am offered a plastic bag in a store, I will always take it.	RR3 (-)
30(=).	For shopping, I prefer paper bags to plastic ones.	RR4
31(*).	Usually, I buy water with returnable bottles.	RR5
	Environmental activism	
32(=).	I often talk with friends about problems related to the environment.	V1
33(=).	I am a member of an environmental organization.	V2
34(=).	In the past, I have pointed out to someone his or her un-ecological behaviour.	V3
35(=).	I sometimes contribute financially to environmental organizations.	V4
	Transport	
36(=).	Usually, I do not drive my automobile in the city.	T1
37(=).	I usually drive on freeways at speeds lower than 100km/h.	T2
38(*).	When possible, I do not use a car for distances less than 30km.	T3
39(=).	If possible, I do not insist on my right of way and make the traffic stop before entering crossroads.	T4
40(=).	I walk, ride or take public transport to go to work/university	T5

(-) items positively formulated as environmentally damaging, recoded

(=) unmodified items from Kaiser and Wilson (2000)

(*) adapted items from Kaiser and Wilson (2000)

(+) new items

Table 2: Interpreting INFIT and OUTFIT statistics

Interpretation of parameter-level mean-square fit statistics:	
>2.0	Distorts or degrades the measurement system.
1.5 - 2.0	Unproductive for construction of measurement, but not degrading.
0.5 - 1.5	Productive for measurement.
<0.5	Less productive for measurement, but not degrading. May produce misleadingly good reliabilities and separations.

Table 3: Descriptive statistics of the sample

Gender	N
Male	76
Female	55
Household Size	**N**
1	27
2	34
3	23
>4	47
Children in household	**N**
0	69
1	21
2	37
3	4
PT subscription	
none	67
weekly	5
monthly	18
yearly	39
lifetime	2

Table 4: Estimates of Item parameters, INFIT, OUTFIT and bi-serial correlation statistics

ENTRY			MODEL	INFIT		OUTFIT		POINT-BIS. CORR.		EXACT MATCH (%)		ITEM
N°	SCORE	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	OBS.	EXP.	OBS.	EXP.	NAME
27	**130**	**-4.06**	**1.01**	**1.02**	**0.35**	**2.02**	**1.08**	**-0.07**	**0.05**	**99.2**	**99.2**	**RR1**
5	**129**	**-3.35**	**0.72**	**1.03**	**0.27**	**1.65**	**0.93**	**-0.03**	**0.07**	**98.5**	**98.5**	**CS5**
14	**12**	**3.49**	**0.31**	**1.11**	**0.52**	**1.48**	**1.44**	**-0.02**	**0.20**	**90.8**	**90.8**	**AE1**
8	109	-0.72	0.24	1.10	0.71	1.38	1.72	0.03	0.22	83.2	83.2	R1
7	108	-0.66	0.24	1.09	0.68	1.26	1.29	0.05	0.23	82.4	82.5	CS7
11	129	-3.35	0.72	1.00	0.24	1.06	0.35	0.07	0.07	98.5	98.5	R4
4	36	2.08	0.2	1.13	1.39	1.15	1.22	0.09	0.28	70.2	73.5	CS4
6	107	-0.61	0.23	1.08	0.60	1.17	0.92	0.09	0.23	81.7	81.7	CS6
3	128	-2.94	0.59	0.99	0.17	0.87	0.04	0.13	0.09	97.7	97.7	CS3
12	129	-3.35	0.72	1.00	0.22	0.67	-0.18	0.13	0.07	98.5	98.5	R5
1	89	0.19	0.2	1.09	1.19	1.11	1.08	0.14	0.28	64.9	69.2	CS1
13	129	-3.35	0.72	0.99	0.22	0.60	-0.28	0.14	0.07	98.5	98.5	R6
19	102	-0.36	0.22	1.06	0.52	1.07	0.51	0.16	0.25	77.9	77.9	AE6
39	84	0.38	0.19	1.08	1.12	1.14	1.47	0.16	0.29	63.4	66.8	T4
37	24	2.65	0.23	1.04	0.33	1.15	0.82	0.17	0.25	80.9	81.9	T2
15	116	-1.19	0.28	1.01	0.10	1.03	0.18	0.18	0.19	88.6	88.6	AE2
18	118	-1.36	0.3	0.99	0.05	0.90	-0.25	0.21	0.18	90.1	90.1	AE5
22	90	0.15	0.2	1.03	0.46	1.08	0.73	0.21	0.28	67.2	69.8	CE2
9	90	0.15	0.2	1.04	0.52	1.04	0.36	0.22	0.28	68.7	69.8	R2
20	96	-0.09	0.21	1.04	0.40	1.00	0.01	0.23	0.26	73.3	73.6	AE7
33	14	3.31	0.29	1.01	0.11	0.86	-0.43	0.23	0.21	89.3	89.3	V2
17	98	-0.18	0.21	1.03	0.31	0.98	-0.12	0.23	0.26	73.3	75.0	AE4
2	92	0.07	0.2	1.00	0.00	0.99	-0.03	0.28	0.27	71.8	71.0	CS2
24	69	0.89	0.18	1.02	0.31	1.01	0.20	0.28	0.30	61.1	62.8	CE4
26	42	1.84	0.2	1.01	0.16	0.99	-0.04	0.28	0.29	68.7	69.7	CE6
40	83	0.41	0.19	0.98	-0.25	1.04	0.43	0.30	0.29	73.3	66.3	T5
34	111	-0.84	0.25	0.95	-0.27	0.86	-0.57	0.31	0.21	84.7	84.8	V3
36	72	0.79	0.18	0.99	-0.13	1.00	0.01	0.31	0.30	57.3	63.1	T1
28	56	1.33	0.19	0.99	-0.12	0.98	-0.25	0.32	0.30	64.9	64.0	RR2
31	42	1.84	0.2	0.97	-0.37	0.96	-0.40	0.34	0.29	67.2	69.7	RR5
30	68	0.93	0.18	0.97	-0.51	0.95	-0.72	0.35	0.30	61.8	62.7	RR4
23	108	-0.66	0.24	0.93	-0.47	0.84	-0.81	0.35	0.23	82.4	82.5	CE3

35	33	2.21	0.21	0.94	-0.54	0.90	-0.75	0.37	0.28	76.3	75.6	V4
21	46	1.69	0.19	0.95	-0.70	0.93	-0.76	0.37	0.30	69.5	67.4	CE1
16	90	0.15	0.2	0.94	-0.69	0.89	-1.05	0.38	0.28	68.7	69.8	AE3
10	92	0.07	0.2	0.92	-0.91	0.88	-1.08	0.39	0.27	73.3	71.0	R3
29	90	0.15	0.2	0.93	-0.91	0.87	-1.26	0.40	0.28	71.8	69.8	RR3
32	82	0.45	0.19	0.93	-1.15	0.90	-1.22	0.40	0.29	71.0	65.9	V1
38	86	0.3	0.19	0.91	-1.24	0.87	-1.39	0.42	0.28	71.0	67.7	T3
25	**50**	**1.54**	**0.19**	**0.85**	**-2.44**	**0.80**	**-2.65**	**0.53**	**0.30**	**72.5**	**65.8**	**CE5**
MEAN	84.5	0.00	0.29	1.00	0.00	1.03	0.00	-	-	77.6	77.6	-
S.D.	33.3	1.82	0.2	0.06	0.7	0.25	0.9	-	-	11.9	11.7	-

Items in **bold** are problematic

Table 5: Results of the PCA performed on residuals

INPUT: 131 Person 40 Item REPORTED: 131 Person 38 Item				
Table of STANDARDIZED RESIDUAL variance (in Eigenvalue units)				
	--Empirical--			Modelled
Total raw variance in observations	55.6	100.0%		100.0%
Raw variance explained by measures	17.6	31.6%		31.6%
Raw variance explained by persons	4.3	7.8%		7.7%
Raw Variance explained by items	13.3	23.9%		23.8%
Raw unexplained variance (total)	38.0	68.4%	100.0%	68.4%
Unexplained variance in 1st contrast	2.8	5.0%	7.3%	
Unexplained variance in 2nd contrast	2.3	4.1%	6.0%	
Unexplained variance in 3rd contrast	2.1	3.8%	5.5%	
Unexplained variance in 4th contrast	1.9	3.3%	4.9%	
Unexplained variance in 5th contrast	1.7	3.1%	4.6%	

Table 6: Descriptive statistics of the measure estimates for each subgroup

	N	Mean	Std. Deviation	Std. Error	95 % CI for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
PMT	48	0.4346	0.81452	0.11757	0.1981	0.6711	-0.86	3.05
PT	41	0.9522	0.85554	0.13361	0.6822	1.2222	-0.59	3.60
SM	19	1.0926	0.55504	0.12733	0.8251	1.3602	-0.09	1.75
Total	108	0.7469	0.83546	0.08039	0.5875	0.9062	-0.86	3.60

Table 7: Results from multiple comparisons Tukey Post-hoc test

Multiple Comparisons

	(I) Mode used for Most Frequent trip	(J) Mode used for Most Frequent trip	Mean Difference (I-J)	Std. Error	Sig.	95% CI	
						Lower bound	Upper bound
Tukey HSD	PMT	PT	-0.51761*	0.16861	0.008	-0.9185	-0.1168
		SM	-0.65805*	0.21490	0.008	-1.1689	-0.1471
	PT	PMT	0.51761*	0.16861	0.008	0.1168	0.9185
		SM	-0.14044	0.22004	0.799	-0.6636	0.3827
	SM	PMT	0.65805*	0.21490	0.008	0.1471	10.1689
		PT	0.14044	0.22004	0.799	-0.3827	0.6636

* the mean difference is significant at the 0.05 level.

Table 8: Comparison of item difficulties between three samples from different countries.

N°	Item description	Item Code	Behaviour difficulty		
			Italy	Switzerland	Sweden
	Sample size		131	445	247
	Ecological garbage handling				
8.	I put dead batteries in the garbage. (-)	R1	-0.72	-2.86	-2.1
10.	I bring unused medicine back to the pharmacy.	R3	0.07	0.62	0.27
11.	I sort paper waste for recycling.	R4	-3.35	-2.09	-1.74
12.	I sort glass waste for recycling.	R5	-3.35	-3.55	-2.44
	Water and power saving				
15.	I prefer to shower rather than to take a bath.	AE2	-1.19	-0.79	-1.38
16.	In winter, I keep the heat on so that I do not have to wear a sweater. (-)	AE3	0.15	0.29	0.64
18.	I wait until I have a full load before doing my laundry.	AE5	-1.36	0.37	-1.27
19.	In winter, I leave the windows wide open for long periods of time to let in fresh air. (-)	AE6	-0.36	-0.51	-1.41
20.	I wash dirty clothes without pre-washing.	AE7	-0.09	-0.95	-0.25
	Ecologically aware consumerism				
21.	I use fabric softener with my laundry. (-)	CE1	1.69	0.51	0.99
22.	If there are insects at home, I kill them with a chemical insecticide. (-)	CE2	0.15	0.19	-0.7
23.	I use a chemical air freshener in my bathroom. (-)	CE3	-0.66	-0.61	-1.38
24.	I use specific cleaners for different rooms rather than an all-purpose cleaner. (-)	CE4	0.89	0.13	0.48
25.	I use phosphate-free laundry detergent.	CE5	1.54	-0.59	-0.4
	Garbage inhibition				
28.	I sometimes buy beverage in cans. (-)	RR2	1.33	0.2	1.35
29.	If I am offered a plastic bag in a store, I will always take it. (-)	RR3	0.15	0.93	1.11
30.	For shopping, I prefer paper bag to plastic ones.	RR4	0.93	-0.08	0.75
31.	Usually, I buy water with returnable bottles.	RR5	1.84	2.96	2.56
	Environmental activism				
32.	I often talk with friends about problems related to the environment.	V1	0.45	0.24	1.38
33.	I am a member of an environmental organization.	V2	3.31	1.36	2.78

34.	In the past, I have pointed out to someone his or her un-ecological behaviour.	V3	-0.84	-0.31	0.81
35.	I sometimes contribute financially to environmental organizations.	V4	2.21	-0.35	1.87
	Transport				
36.	Usually, I do not drive my automobile in the city.	T1	0.79	-0.32	0.79
37.	I usually drive on freeways at speeds lower than 100km/h.	T2	2.65	2.19	1.25
38.	When possible, I do not use a car for distances less than 30km.	T3	0.3	0.67	0.92

(-) items positively formulated as environmentally damaging – Intended as "I refrain from"