Electronic Journal of Applied Statistical Analysis
EJASA, Electron. J. App. Stat. Anal.
http://siba-ese.unisalento.it/index.php/ejasa/index
e-ISSN: 2070-5948
DOI: 10.1285/i20705948v14n2p339

A discretization method for continuous latent traits

By Golia

Published: 20 November 2021

Electronic Journal of Applied Statistical Analysis
Vol. 14, Issue 02, November 2021, 339-358
DOI: 10.1285/i20705948v14n2p339

A discretization method for continuous latent traits

Silvia Golia*

University of Brescia, Department of Economics and Management
C.da S.Chiara, 50 - 25122 Brescia, Italy

Published: 20 November 2021

The paper presents a new discretization method applicable to measures of continuous latent traits estimated using a measurement model belonging to the item response theory (IRT) approach. The reasons of this proposal are twofold: first, the need to discretize a continuous variable due to the use of methodologies primarily designed to handle categorical data (for example Bayesian Networks) or the increase of efficiency and effectiveness of the learning algorithms, second, the discretizers available in literature are not able to reproduce the peculiarities of the target variables of this paper. The idea underlying the proposed method is to use the information from an IRT model in order to forecast the answer of a subject to a characterizing item; the obtained response is the category assigned to the subject in the discretized version of her/his continuous latent trait. The performance of this discretizer is compared to the performance of other common unsupervised discretization methods, with respect to a global single-item measure, that is assumed to represent an observed discretized version of the continuous latent trait.

keywords: Discretization, Continuous latent trait, IRT models, Global single-item.

1 Introduction

When dealing with a survey composed of many sections referred to several latent traits of interest, for each of them generally a set of items, built to measure it, or an overall

*Corresponding author: silvia.golia@unibs.it

ISSN: 2070-5948
http://siba-ese.unisalento.it/index.php/ejasa/index

item is included and the set of items very rarely also includes the overall item. It is of interest to underlie that the overall item, under certain circumstances, can be considered as an approximation of the discrete version of the continuous latent trait of interest (Diamantopoulos et al., 2012). When the overall item is absent, it is possible to obtain a measure of the latent trait making use, for example, of models following the item response theory (Bartolucci et al., 2016); the obtained measure is continuous. As one will see shortly, the use of discretized variables, instead of the original continuous ones, is desirable in many applicative contexts. This consideration motivated the need to discretize continuous latent traits, which is a challenging task and it is of interest when dealing with social phenomena, which involve many descriptors.

Discretization is the process that transforms a quantitative variable into a qualitative one, producing a partition on the range of the values taken by the variable; an association between each interval in the partition and a numerical discrete value is then established. Once the discretization has been performed, the new variable can be treated as an ordinal one. Discretization can be viewed as one of the possible data preprocessing techniques; these techniques can substantially improve the overall quality of the relations extracted from the data and/or the time required for the analysis (Han et al., 2012). Discretization can, or must, be applied before using many statistical models; in fact, there are many models that are primarily designed to handle categorical data (Dougherty et al., 1995; Liu et al., 2002), such as, for example, Bayesian Networks (BNs) or Naive Bayes (NB). Both these models study the relations between the variables of interest and they allow the coexistence of discrete and continuous variables in the dataset under study. Nevertheless, in the case of BNs, hybrid databases force to constraint the parent-child relationships among variables, imposing that a discrete variable may only have discrete parents (Kjærulff and Madsen, 2013), and this can be an unrealistic constraint in many applications. For both BN and NB, it is necessary to estimate probabilities and continuous variables are difficult to handle; to circumvent this problem they are commonly assumed to be normally distributed, but this hypothesis does not always reflect the real nature of these variables. Moreover, even if the models can deal with continuous variables, the learning process is less efficient and effective (García et al., 2013). For example, when building a decision tree, the coexistence of discrete and continuous variables implies that, in the splitting procedure, the continuous variables are more easily chosen, as they assume more values than the other type of variables. So, discretizing continuous variables either before the decision tree induction or during the process of tree building can be a good strategy. Moreover, the results obtained from a decision tree without continuous variables are more compact and can be more closely examined and used (Liu et al., 2002). Two other examples are found in Bartolucci et al. (2015) and Lustgarten et al. (2008). In the analysis of genomic and proteomic biomedical data, Lustgarten et al. (2008) have shown that machine learning classification algorithms, such as Support Vector Machines and Random Forests, benefit from the discretization of continuous variables. Bartolucci et al. (2015) discretized bibliometric indicators to avoid strong parametric assumptions and because the discretized variables offered some robustness to measurement errors.

The continuous variables considered in this paper are continuous latent traits, which are common in socio-economic and psychological contexts. In fact, in the social sciences it is common to find surveys composed of several questionnaires, related to distinct latent traits, plus overall questions and personal characteristics of the respondents. These latent traits can be estimated from the related questionnaires by a measurement model, and individually or jointly used in successive analyses. So, depending on the chosen model and/or efficiency and effectiveness reasons, it could be necessary to discretize them when one wants to use them jointly with other variables that are of categorical nature. As stated previously, these continuous latent variables are generally estimated via a measurement model based on data coming from the administration of ad hoc questionnaires to a sample of the target population, so they are not directly observed. The questions (items) included in such questionnaires typically admit response categories which are on ordinal scale, such as the Likert-type scale, and the measurement model can belong to the item response theory (IRT) approach (Bartolucci et al., 2016). The IRT approach is based on the idea that the response probability to an item is a function of the subject's location on the latent continuum representing the continuous latent trait of interest and of some parameters characterizing the item. The problem addressed in this paper is to identify a good method able to discretize this kind of variable. The resulting discretized variable must mimic the evidence that higher scores correspond to higher levels of the continuous latent trait, so the ordering of the categories matters. Moreover, when the population globally owns high or low levels of the continuous latent trait, the expected distribution of the discretization of the continuous latent trait should be skewed.

Latent class analysis (LCA) allows to empirically identify a discrete latent variable from two or more categorical observed variables (McCutcheon, 1987), so it can be considered, at first glance, as an alternative way to reach the goal of this paper. Nevertheless, LCA assumes that the underlying latent variable is discrete, and one of its issues consists in identifying the correct number of categories of the latent trait. The latent class model assumes that the population from which the subjects are sampled, can be partitioned in a fixed number of mutually exclusive and homogeneous subpopulations, represented by the latent classes. So, each latent class contains observations that are similar to each other, but different from those of the other classes. Moreover, within each latent class, the observed variables are supposed to be stochastically independent. The latent classes are assumed to be the levels of the discrete latent variable measured by the categorical observed variables. The set of latent classes can be viewed as a partition of the original population into disjoint subpopulations and nothing forces these classes to be ordered along a continuum, therefore, the obtained discrete variable is a nominal variable. There are two drawbacks which limit the application of the LCA approach to the context of this paper. The first one is connected with the nature of the considered latent traits, which are continuous and not discrete, whereas the second one is connected to the nature of the discretized version of these continuous latent traits. The discretized latent trait must be an ordinal variable, with increasing scores corresponding to increasing levels of the latent trait, as previously explained.

In literature there were proposed many discretization methods, classifiable following different criteria, that is static versus dynamic, univariate versus multivariate, supervised versus unsupervised, global versus local, direct versus incremental and splitting versus merging. A static discretizer operates prior and independently to the learning algorithm, whereas a dynamic discretizer acts together with the learning algorithm. A univariate discretizer works with a single variable at a time, whereas a multivariate one considers more than one attributes simultaneously. Supervised discretizer takes into account the class label of the outcome, whereas the unsupervised one does not. Global discretizers use all the available observations, whereas local ones consider partial information. A direct discretizer produces a number of intervals that must be determined a priori, whereas an incremental one begins with a simple discretization and passes through an improvement process requiring a stopping rule. Splitting and merging methods refer to the procedure used to make new intervals. García et al. (2013) provide a detailed survey of the discretization methods available in literature, classified following the above mentioned criteria. Now, due to the nature of the variables analyzed in this paper, a class label is not available, so the discretizers that can be used must be unsupervised. Moreover, given that the interest is focused on one variable at time and the procedure does not run together with a learning algorithm, useful discretizers must be univariate and static.

The method proposed in this paper originates from the consideration that the available unsupervised, univariate and static methods do not take into account the peculiarities of these measures and the characteristics that the resulting discretized variables must have. The idea underlying it, is to use the information from an IRT model in order to forecast the answer of a subject to a characterizing item; the obtained response is the category assigned to the subject in the discretized version of her/his continuous latent trait. The proposed discretizer fixes a priori the number of modalities of the discretized variable equal to the number of the response categories used in the questionnaire of reference for the latent trait of interest; in general this number is common to all the questions in the questionnaire. This is one of its characteristics, linked to the way used to discretize the continuous latent trait, and it can be seen as a limit. Nevertheless, there are no real reasons to ask that the discretized variable has more categories than those used in the related items, and if one instead needs a smaller number, then the final categories can be conveniently collapsed. So, the proposed method can be classified as an unsupervised, univariate, static and direct discretizing method. It is useful when the questionnaire used to measure the continuous latent trait of interest does not contain an overall question, otherwise this overall question can be used as a discretized version of the continuous latent trait. Nevertheless, it is rare that a questionnaire, devoted to measuring a latent trait, includes the set of items as well as the overall item.

The rest of the paper is organized as follows. A brief overview of the most common unsupervised, univariate, static and direct methods that can be applied to an IRT measure available in literature, and the description of the new method proposed in this paper are presented first. Then theories which support the use of a global single-item measure as an observed discretized version of the continuous latent trait are discussed. The tools implemented to measure the resemblance between the global single-item measure and the related discretized measure are presented and utilized to evaluate the performances of the discretizers described in the paper applied on synthetic and real datasets.

Conclusions end the paper.

2 Discretizer of an IRT measure

The most common unsupervised, univariate, static and direct discretization methods available in literature are the Equal Width Discretizer (EWD) and the Equal Frequency Discretizer (EFD). In both cases the n observations are sorted in ascending order $(x_{(1)}, x_{(2)}, \ldots x_{(n)})$ and the range of the variable x, $[x_{(1)}, x_{(n)}]$, is divided into a user-supplied number k of intervals; the difference between EWD and EFD consists in the way these intervals are created. For the EWD the intervals are of equal width and the level j is assigned to the subject i if and only if $x_1 + \frac{(j-1)(x_{(n)}-x_{(1)})}{k} < x_i \leq x_1 + \frac{j(x_{(n)}-x_{(1)})}{k}$. For the EFD the intervals contain approximately same number (about n/k) of subjects with adjacent values and the level j is assigned to the subject i if and only if $Q_{j/k} < x_i \leq Q_{(j+1)/k}$, where Q_q is the quantile of order q, with $q = 1/k, 2/k, \ldots, (k-1)/k$.

Another standard unsupervised method is the cluster-based discretization (Gan et al., 2007; Dillon and Goldstein, 1984). There are many clustering algorithms available, and the one used in this paper is the k-means clustering method. This is a popular clustering method, classified as a partitional or nonhierarchical clustering method, which groups the observations in such a way that each cluster has a center called the mean. The number of clusters k is assumed to be fixed and defined by the researcher. It can be divided into two steps, the initial and the iteration step (Hartigan and Wong, 1979). In the initial step the algorithm randomly assigns the observations into k clusters, whereas in the iteration step it computes the distance between each observation and each cluster and assigns the observation to the nearest cluster.

The above mentioned methods do not take into account the peculiarities of the variables to be discretized (i.e. they are continuous latent traits) and the characteristics that the resulting discretized variables must have. In fact, the discretized version of the latent variable must mimic the evidence that higher scores correspond to higher levels of the latent trait and, if the population globally owns high or low levels of the latent trait, its expected distribution should be skewed.

2.1 The Rating Scale Model

In this paper the continuous latent trait is measured thanks to the Rating Scale Model (RSM) (Andrich, 1978). The RSM belongs to the family of the Rasch models and it is able, as all the other models in the family, to turn raw scores into linear and reproducible measures so, if the data fit the model, the obtained measures are objective and expressed in logits (Wright and Master, 1982). Following the RSM, given an item i with $m+1$ response categories $(c = 0, 1, \cdots, m)$, the probability of the subject s with level of latent trait θ_s (denoted also as the ability of the subject s) to respond in category c is given by:

$$P(X_{si} = c) = p_{sic} = \frac{exp\left\{c(\theta_s - \delta_i) - \sum_{j=0}^{c} \tau_j\right\}}{\sum_{l=0}^{m} exp\left\{l(\theta_s - \delta_i) - \sum_{j=0}^{l} \tau_j)\right\}} \tag{1}$$

where δ_i represents the difficulty of item i and the τ_j are called thresholds ($\tau_0 \equiv 0$ and $\sum_{j=1}^{m} \tau_j = 0$). The m thresholds are equal for all the items, implying that all the items in the test have the same set of labels for the response alternatives, a common situation in many application fields. All the parameters are expressed in the same scale (logit) and this allows comparisons. The abilities θ_s and the difficulties δ_i can be represented on the same continuum, so it is possible to evaluate the difficulty of the items relative to each other and also relative to the abilities distribution. In fact, looking at (1), when the subject ability is bigger than the difficulty of a given item, it is more probable that this subject chooses the response to the item between the highest response categories, on the contrary the same subject will choose her/his response between the lowest response categories if the difficulty of the item is higher than the her/his ability.

The choice of using the RSM as measurement model, relies on its belonging to the class of Rasch-type models, its compact parametrization and the fact that it possesses almost all the statistical properties desirable for a model for polytomous responses (Bartolucci et al., 2016).

The estimate of the parameters involved in (1) can be obtained making use of the joint maximum likelihood estimation method. In order to make the model identifiable, it is necessary to impose some constraints on its parameters. Three possible constraints could be used: to set the difficulty of the first item equal to zero ($\delta_1 = 0$), to set the average difficulty of the items equal to zero ($\sum \delta_i = 0$), or to set the average ability of the subjects equal to zero ($\sum \theta_i = 0$). These three alternatives are equivalent and in this work the chosen identifiability constraint is the second one, that is the average difficulty of the items is set equal to 0.0 logits, so each δ_i can be interpreted as the difficulty of item i with respect to the average difficulty (Wright and Master, 1982; Bartolucci et al., 2016). This choice generates a more general setting, in which no reference item, between the available ones, must be singled out. Moreover, this choice allows to give a meaning to the comparison between the estimated mean level of the latent trait and the average difficulty of the items (fixed to be zero) in terms of a sort of quantification of the overall level of latent trait owned by the sample. If the estimated mean level of the latent trait is significantly higher (lower) than zero, this suggests that the questionnaire is not well calibrated for the sample of respondents, because it is overall easy (difficult) to score high (low) to the items (Bond and Fox, 2007). In the original context in which the model was born, that is intelligence tests, a bad calibrated test should be revised and some items substituted. Nevertheless, if the questionnaire can not be revised, as in the case in which the latent trait to be measured is described by the facets reproduced by the items composing the questionnaire (for example satisfaction, fairness, burnout, etc.), a significant distance from zero of the mean level of measured latent trait can be interpreted as an indication of how the latent trait is owned by the subjects in the sample. One considers, for example, job satisfaction; if the estimated mean satisfaction is significantly higher (lower) than zero, one can conclude that the workers are (not) overall satisfied with their job. Then, if the sample of subjects has an overall high (low) level of latent trait, for example if the sample is overall satisfied (dissatisfied), the frequency distribution of a discretized version of the latent trait should be concentrated in the higher (lower) response categories so, it should be skewed. This characteristic should be captured by any used discretizer; the distribution of the discretized measure should mimic this type of shape in the distribution.

2.2 The proposed method

The proposed method originates from the considerations regarding the peculiarities of the expected distribution of the discretized version of the continuous latent trait explained above. The idea is to forecast the answers of the subjects to a characterizing item, able to discriminate the subjects according to their level of latent trait, and to consider the obtained response categories as the discretized version of the continuous latent trait of interest. Equation (1) and the maximum likelihood estimates of the θ_s and τ_j parameters allow to estimate the response probabilities of the subjects to a characterizing item. From each of these response probabilities records it is possible to identify the most probable response category, which becomes the discretized version of the continuous latent trait. Now, the question is how to choose this characterizing item. The estimates of the parameters in (1) are obtained under the constraint that the sum of the difficulty parameters is zero, therefore, an item with zero difficulty ($\delta_{i^*} = 0$) corresponds to an item with average difficulty, so, it can be considered a good candidate for the characterizing item. In fact, an item too difficult (easy) for the sample of subjects, that is an item with a positive (negative) δ_{i^*} far from zero, would generate, for almost all the subjects, responses concentrated on the first (last) response categories, and this would not allow to adequately differentiate the respondents in terms of their latent trait. An item with a mean difficulty ($\delta_{i^*} = 0$) should not suffer from this drawback. So, the characterizing item proposed in this study is the item i^* with difficulty equal to zero: $\delta_{i^*} = 0$.

The proposed discretization method is described as follows. Given the estimated thresholds $\{\hat{\tau}_j\}_{j=1}^m$, the estimated measure of the latent trait $\{\hat{\theta}_s\}_{s=1}^n$, where n indicates the sample size, and $\delta_{i^*} = 0$, equation (1) allows to calculate, for each subject s, her/his response probability record $\{p_{si^*c}\}_{c=0}^m$. The discretization of $\hat{\theta}_s$ for the subject s, say d_s, is represented by the most probable response category, that is

$$d_s = \arg\max_c p_{si^*c}. \tag{2}$$

Following the taxonomy recalled in the introduction, the proposed discretization method belongs to the class of static, unidimensional, unsupervised, global, direct discretizers.

3 Adequacy of the global single-item measure

In order to verify the goodness of the proposed discretization technique, the most natural way should be to compare the new discretized variable with the observed discrete version of the continuous latent trait of interest. The challenge is to find such a variable. Intuitively, a global, or overall, single-item measure, that summarizes the essence of the construct under study, could be seen as an approximation of the searched discrete version of the latent trait measured by the set of items that are proposed for this purpose. As an example, a global single-item for the job satisfaction could be "How satisfied are you with your job as a whole?", and it is reasonable to admit that the respondent can implicitly make a synthesis of her/his job satisfaction when she/he answers to this kind of question.

The purpose of exploring the validity of a single-item measure, compared to a multi-item measure, as a measure of latent constructs has been pursued in several research fields, including job satisfaction (Wanous et al., 1997), quality of life (deBoer et al., 2004), well-being and life satisfaction (Fors and Kulin, 2016; Lucas and Donnellan, 2012), burnout (Waddimba et al., 2016), citizen satisfaction (Van Ryzin, 2004), social identification (Postmes et al., 2013) and management research (Fuchs and Diamantopoulos, 2009). Following the Bagozzi and Heatherton construct representation (Bagozzi and Heatherton, 1994), a single-item measure is an individual indicator defined under the so-called *total disaggregation model*, that is an indicator that can not be further decomposed to lower-level constituents; in the context of job satisfaction, an example is given by the item "How satisfied are you with your job as a whole?". Several studies (see, for example, Fuchs and Diamantopoulos (2009) and the references therein) have shown that the single-item measure can have acceptable psychometric properties such as reliability and validity. Moreover, in order to identify conditions under which the use of a single item is worth consideration, the nature of the construct and the research objectives mainly have to be taken into account. A latent construct can be of two types, concrete or abstract. A concrete construct is characterized by a virtually unanimous agreement among respondents as to what characteristic is being measured, so it is perceived as homogeneous, whereas an abstract construct means somewhat different things to different respondents and it is sensed as heterogeneous Rossiter (2002). Sackett and Larson (1990) and Rossiter (2002) argued that, when a construct can be judged to be concrete, the use of a single-item measure could be considered reasonable. Connected to the concept of concreteness there is the concept of complexity of the construct. In general, it is inappropriate to use a single-item measure when dealing with a complex construct, nevertheless, with highly complex constructs, a global single-item question may be appropriate (Sloan et al., 2002). The reason lies in the fact that with a highly complex construct, it is possible that a multi-item measure is not able to capture all the dimensions of the construct, resulting in an incomplete evaluation of it (Nagy, 2002). Regarding the research objectives, if the interest is focused on understanding the general nature of a construct, a single-item global measure may be appropriate (Lee et al., 2000). In synthesis, depending on the nature of the construct operationalized, the use of a global single-item measure is often adequate for the purpose (Diamantopoulos et al., 2012).

4 Measures of resemblance between discrete measures

Once the validity of the global single-item measure as an observed discretized version of the continuous latent trait is established, it is necessary to identify tools able to quantify the resemblance between this discrete measure and the discretized measures produced by the application of all the methods described in the paper.

For the subject s, let o_s be her/his response to the global single item, let d_s be the discretization of her/his estimated measurement $\hat{\theta}_s$, obtained from the application of a discretizer, and let pm_s be the indicator of her/his perfect match, so that $pm_s = 1$ if $o_s = d_s$ and zero otherwise. The first indicator of resemblance considered in this work is the percentage of Perfect Match (%PM) between the os and the ds given by

$$\%PM = \frac{\sum_{s=1}^{n} pm_s}{n} * 100.$$

This indicator does not take into account the fact that the categories are ordered, so the Mean Absolute Difference (MAD) between the os and the ds and the Similarity Index proposed by Gower (1971) are computed. The MAD between the os and the ds is defined as

$$MAD = \frac{\sum_{s=1}^{n} |o_s - d_s|}{n},$$

and it gives an idea of the mean distance between the two variables. The Similarity Index is a normalized indicator with values near 1 indicating high degree of similarity.

5 Empirical evidences

This section reports some evidence that the proposed method performs better than the three standard ones considered. Both synthetic and real datasets were analyzed. In order to produce synthetic data which do not favor, by construction, the proposed method, the mean and standard deviation of the ability and the difficulty of the overall item were chosen from what obtained from real contexts.

5.1 Synthetic datasets

This subsection contains the results of a simulation study aimed to investigate three scenarios that can show up. The first scenario (Condition I) concerns the situation in which the population owns an overall high level of the latent trait, which translates, following the reasoning reported in subsection 2.1, in an average ability significantly higher than zero and a much lower value of the difficulty parameter of the corresponding overall item. The second scenario (Condition II) describes an opposite situation of the first scenario, where the population owns an overall low level of the latent trait which translates in an average ability significantly lower than zero and a much higher value of the difficulty of the corresponding overall item. The last scenario (Condition III) describes an intermediate condition between the previous ones, which translates in both the mean ability and the difficulty of the overall item around zero.
In what follows, the simulation design is described. One sample of 1000 subjects was drawn from a normal distribution with mean and standard deviation choosen according to the three scenarios previously described (Table 1). These abilities were fixed throughout each condition and were referred to as the true abilities θ. A set of 15 difficulty parameters $\{\delta_i\}_{i=1}^{15}$ was drawn from a continuous uniform distribution on the interval

from -1.9 to 1.9 and transformed so that the sum of the parameters was equal to zero. The values used in this study were [-1.7684, -1.4726, -0.9496, -0.8373, -0.5323, -0.3092, -0.2662, -0.1237, 0, 0.3092, 0.7783, 0.9029, 1.0733, 1.3682, 1.8274]. Moreover, the difficulty parameter of the overall item was added to this set; Table 1 reports its value for the three scenarios. A six-level response scale for each item was considered and the

Table 1: Choices for the mean and standard deviation of the abilities and the difficulty parameter of the overall item

	Average Ability	St. Dev. Ability	Difficulty
Condition I	0.89	1.12	-0.59
Condition II	-0.68	1.88	0.03
Condition III	0.10	1.21	-0.17

corresponding threshold parameters τ_j were set equal to [-1, -0.5, 0, 0.5, 1]. The data simulation was performed using the software R 4.1.1 (R Core Team, 2021). Figure 1 reports, as an example, the distributions of five simulated overall items, which show the different shapes of the distributions related to the three different scenarios. As expected, the first two conditions imply a distribution with a pronounced asymmetry.

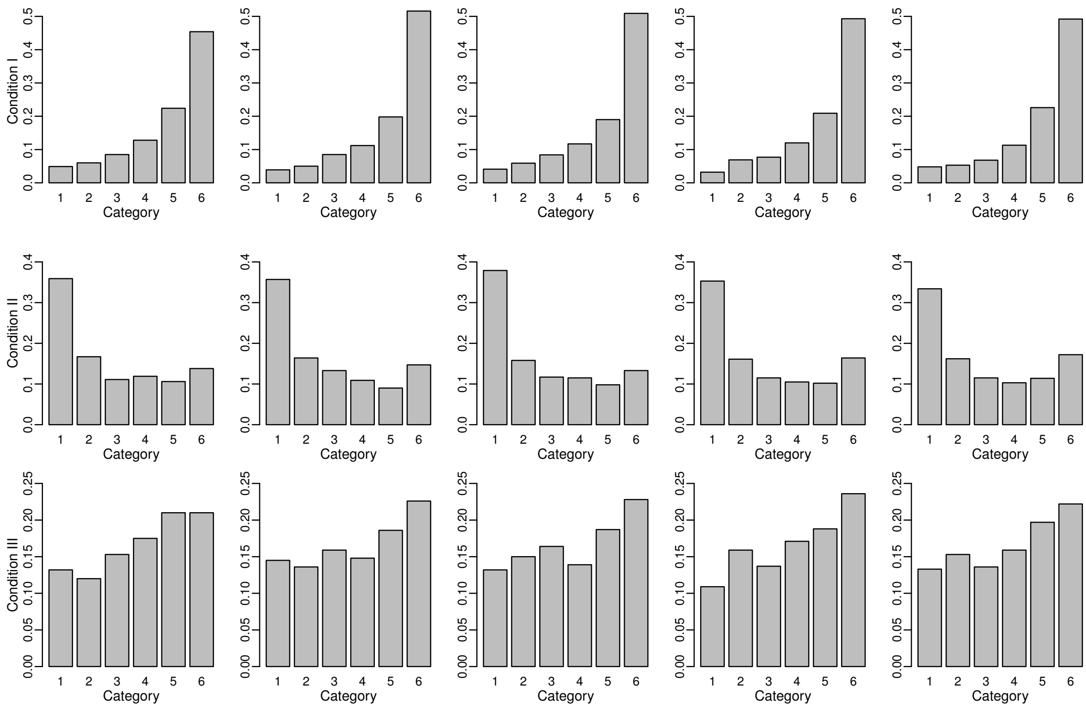

Figure 1: Frequency distribution of the overall item in 5 different outcomes of the simulation under the three conditions

The responses to the first 15 items were used to estimate all the parameters involved in the RSM making use of the joint maximum likelihood estimation method (Wright and Master, 1982) implemented in Winsteps 3.75 (Linacre, 2011) with the sum of item difficulty parameters set equal to 0.0 logits. The responses to the sixteenth item corresponded to the responses to the overall item. The estimated abilities were discretized according to the four methods considered in this study and the measures of resemblance described in Section 4 were computed. This scheme was repeated 100 times and the mean and the standard deviation of the measures of resemblance between the global overall item and the discretized measures are reported in Table 2.

Table 2: Average Measures of resemblance between the global single-item and discretized measures. Standard deviation in brackets

Condition I	%PM	MAD	Similarity Index
IRT	51.5 (0.016)	0.733 (0.028)	0.853 (0.006)
EWD	18.5 (0.025)	1.259 (0.125)	0.748 (0.025)
EFD	24.5 (0.013)	1.469 (0.044)	0.706 (0.009)
k-means	24.4 (0.047)	1.207 (0.199)	0.759 (0.040)
Condition II	%PM	MAD	Similarity Index
IRT	53.3 (0.015)	0.660 (0.027)	0.868 (0.005)
EWD	30.0 (0.019)	0.953 (0.030)	0.809 (0.006)
EFD	42.1 (0.015)	0.818 (0.035)	0.836 (0.007)
k-means	32.8 (0.029)	0.922 (0.101)	0.816 (0.020)
Condition III	%PM	MAD	Similarity Index
IRT	39.7 (0.016)	0.886 (0.032)	0.823 (0.006)
EWD	24.6 (0.026)	1.073 (0.073)	0.785 (0.015)
EFD	38.0 (0.014)	0.885 (0.032)	0.823 (0.006)
k-means	34.3 (0.025)	0.889 (0.066)	0.822 (0.013)

In the first two conditions the proposed method outperforms the competitors, whereas in Condition III the method performs slightly better than the EFD, which has similar performance. The explanation to this behaviour can be found in the inspection of the frequency distribution of the overall item under this condition shown in the last row of Figure 1; it is possible to note that almost all the categories have similar frequencies so it is reasonable to expect that the EFD has a certain power. Looking at the %PM in Table 2, Condition III seems to be the condition in which the proposed method performs worse, even if it is still preferable to its competitors.

5.2 Real datasets

In this subsection the performances of the proposed method are tested on real data, so, in order to be able to do that, surveys containing, for a given latent trait, a set of items as well as the overall item, were searched. Eight different psychological and social constructs, i.e. burnout, avoidant attachment, distributive fairness, perceived well-being, worker, student, customer and life satisfaction, were taken into account and analyzed. Data regarding worker satisfaction and distributive fairness come from two different surveys, the first one, the Survey on the Italian Social Cooperatives (ICSI), is a national one, whereas the second one, the Survey on the Quality of Work (QdL), was a local one. ICSI was carried out in 2007 (Carpita and Golia, 2012) and it involved paid workers employed in Italian social cooperatives of type A and B. QdL was held in 2013 and was attended by workers of an Italian Municipality. The global single item of the worker satisfaction was "How satisfied are you with your job?", whereas the global single item of the distributive fairness was "Do you consider the remuneration you receive to be adequate?". Burnout data come from a survey held in 2009 concerning social workers working in Veneto, a region in Northern Italy (Bressan et al., 2011). The global single item was "How often do you experience the following feeling: I feel emotionally drained from my work". Data concerning the avoidant attachment come from a survey carried out between March and June 2016, at three nursing homes located in Lombardia (Northern Italy). The respondents were auxiliary nurses. The global single item was "I feel nervous when people start to get too close". Data regarding life satisfaction come from the Opinions and Lifestyle Survey (Office for National Statistics, 2016); the respondents were components of households aged 16 and over living in Great Britain and they were asked to rate their feelings towards different aspects of their lives. The data were collected between April and May, 2015 and the global single item was "Overall, how satisfied are you with your life nowadays?". Student satisfaction data come from the responses given to a course evaluation questionnaire administered in an Italian University in 2015. The global single item was "Overall, are you satisfied with the course?". Data concerning the customer satisfaction come from the annual Mayor and Chief Executive Officer Survey held in 2015 (Queensland Government, 2015). Respondents of the survey were mayors and chief executive officers of Queensland's 77 local government offices. The global single item was "Overall, how satisfied are you with the advice and services provided to your council by the department over the past 12 months?". Data regarding the perceived well-being come from the European Quality of Life Survey, 2011-2012, restricted to the Italian citizens, carried out between September 2011 and February 2012 (European Foundation for the Improvement of Living and Working Conditions, 2014). Following the note at page 18 of Eurofound (2013), the global single item considered here was "Taking all things together, how happy would you say you are?".

A Rasch analysis was performed on each dataset, ended with the identification of the proper number of response categories and number of items. Then, all the parameters involved in the RSM were estimated making use of the joint maximum likelihood estimation method (Wright and Master, 1982) implemented in Winsteps 3.75 (Linacre, 2011) with the sum of item difficulty parameters set equal to 0.0 logits. Table 3 reports, for each latent trait, the number of subjects, items and categories (k) used in the Rasch analysis and the mean (standard deviation in parenthesis) of the estimated measures of the latent traits.

Table 3: Number of subjects, items and categories (k) used and the mean (standard deviation) of obtained Rasch measures.

Latent Trait	Num. Subj.	Num. Items	k	Mean (Std. Dev.)
Worker Satisfaction (ICSI)	3980	11	5	0.89 (1.42)
Worker Satisfaction (QdL)	398	7	7	0.10 (1.21)
Distributive Fairness (ICSI)	3666	7	6	-0.68 (1.88)
Distributive Fairness (QdL)	392	4	6	-0.51 (3.42)
Burnout	770	7	5	-1.55 (1.90)
Avoidant Attachment	107	7	5	-1.20 (1.37)
Student Satisfaction	256	10	4	3.69 (1.87)
Customer Satisfaction	107	5	5	1.95 (2.12)
Life Satisfaction	2042	7	5	1.25 (1.53)
Well-being	2238	6	7	1.17 (1.42)

The mean level of the estimated latent traits is, in almost all the cases, far from zero, suggesting that the shape of almost all the distributions of the global single-item measure should be skewed; the shape of frequency distribution of the global single-item measures, shown in Figure 2 (first column), seem to support this intuition. When the mean level of latent trait is significantly higher (lower) than zero, the respondents to an overall question should concentrate their answers on the higher (lower) response categories, with the consequence that the corresponding distribution of the responses should be negatively (positively) skewed.

For each latent trait considered, Figure 2 showes, in addition to the frequency distribution of the global single-item measure, the frequency distribution of its discretization versions obtained applying the proposed method (IRT), EWD, EFD and k-means method. The value k was set equal to the number of response categories used in the estimation of the measures and reported in Table 3.

From a graphical inspection of Figure 2 it is evident as the proposed discretization method, reported in the second column of the figure, is able to produce a discretization that gives a distribution which is more similar to the one of the global single-item measure than the other three competitors.

Tables 4 and 5 report the values of the three measures of resemblance introduced in Section 4. With reference to the results shown with the k-means method, the cluster-based categorization has been performed a high number of times and the most frequent result has been reported.

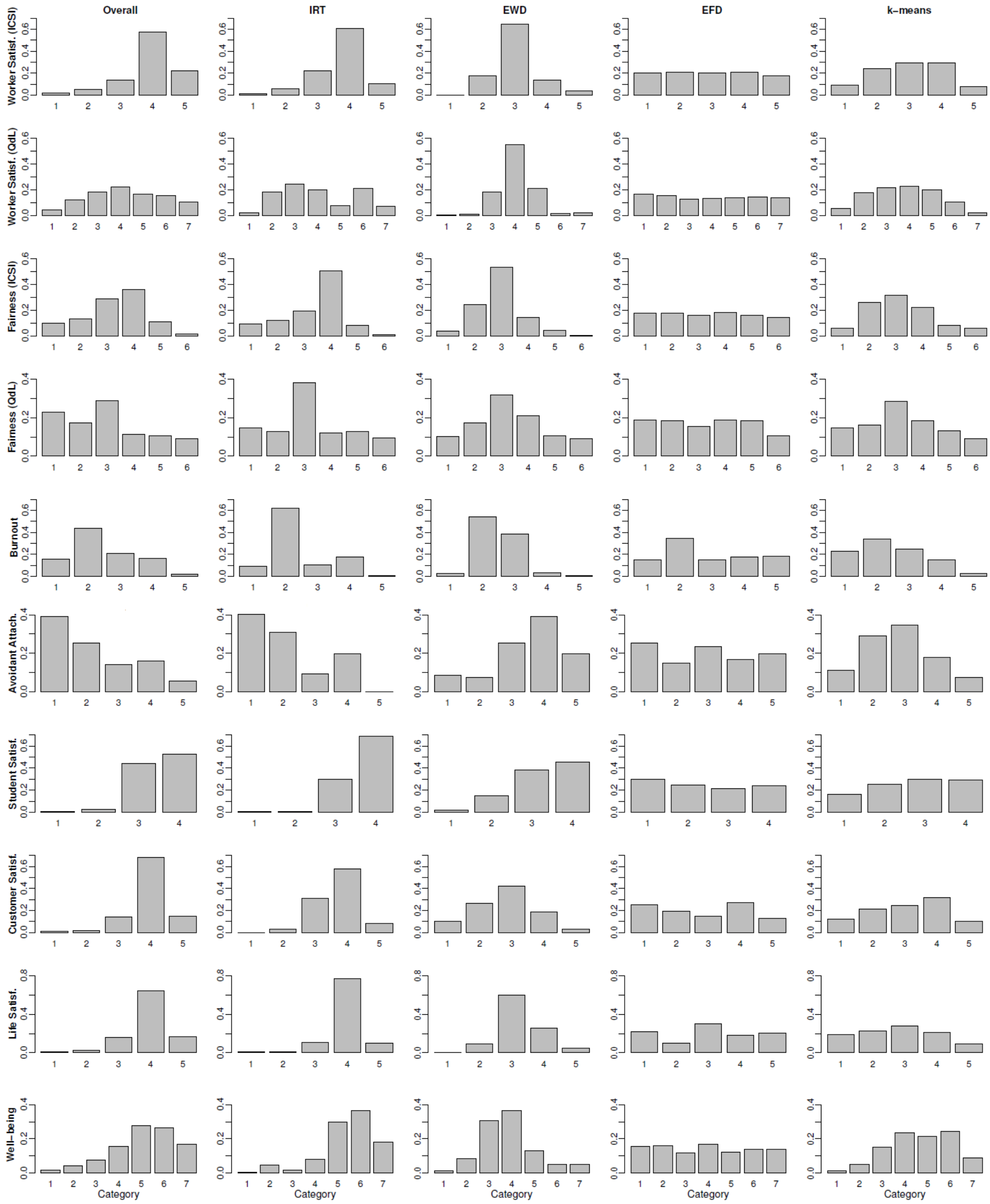

Figure 2: Frequency distribution of the global single-item measure (Overall) and its discretization versions obtained applying the proposed method (IRT), EWD, EFD and k-means method

Table 4: Measures of resemblance between the global single-item and discretized measures

Latent Trait / Discretization Method	%PM	MAD	Similarity Index
Worker Satisfaction (ICSI)			
IRT	56.4	0.511	0.872
EWD	20.4	1.005	0.749
EFD	28.5	1.199	0.700
k-means	30.8	1.012	0.747
Worker Satisfaction (QdL)			
IRT	44.5	0.754	0.874
EWD	32.4	0.940	0.843
EFD	35.2	0.972	0.838
k-means	36.9	0.844	0.859
Distributive Fairness (ICSI)			
IRT	60.5	0.452	0.910
EWD	44.1	0.622	0.876
EFD	37.1	0.783	0.843
k-means	52.8	0.540	0.892
Distributive Fairness (QdL)			
IRT	59.2	0.536	0.893
EWD	57.1	0.556	0.889
EFD	50.5	0.679	0.864
k-means	56.9	0.556	0.889
Burnout			
IRT	60.6	0.440	0.890
EWD	53.1	0.495	0.876
EFD	44.9	0.697	0.826
k-means	57.8	0.462	0.884

With reference to the proposed discretizer (IRT), in few cases the percentage of perfect match is less than 50%, indicating that at least half of the sample is correctly represented by the corresponding identified response category. In very few cases, the same is observed when the three other discretizers are used. Moreover, the MAD and the Similarity

Table 5: Measures of resemblance between the global single-item and discretized measures

Latent Trait / Discretization Method	%PM	MAD	Similarity Index
Avoidant Attachment			
IRT	56.1	0.579	0.807
EWD	16.8	1.383	0.654
EFD	29.0	1.009	0.748
k-means	35.5	0.879	0.780
Student Satisfaction			
IRT	76.6	0.234	0.922
EWD	62.9	0.379	0.877
EFD	27.0	1.148	0.617
k-means	33.6	0.863	0.712
Customer Satisfaction			
IRT	58.9	0.467	0.844
EWD	13.1	1.262	0.685
EFD	29.9	1.318	0.671
k-means	31.8	1.065	0.734
Life Satisfaction			
IRT	71.4	0.309	0.923
EWD	31.9	0.748	0.813
EFD	29.2	1.119	0.72
k-means	24.2	1.219	0.695
Well-being			
IRT	44.1	0.785	0.869
EWD	17.9	1.408	0.765
EFD	22.6	1.511	0.748
k-means	35.3	0.894	0.851

Index of the proposed discretizer are, respectively, the lowest and the highest for all the latent traits. So, even when the IRT discretizer does not do a great job in reproducing the output of the corresponding global single-item measure, it outperforms its three competitors.

6 Conclusions

The paper proposes a new discretization method applicable to the estimates of continuous latent traits obtained using a measurement model belonging to the IRT class of models. The necessity of a suitable discretizer for this kind of variables originated from the consideration that these measures have peculiarities that can hardly be captured by the standard unsupervised, univariate and static discretizers.

The idea underlying the proposed method is to use the information from an IRT model in order to forecast the answer of a subject to a characterizing item; the obtained response is the category assigned to the subject in the discretization of her/his latent trait. The measurement model utilized in the paper is the Rating Scale Model and the difficulty of the characterizing item was set equal to zero. In order to verify the goodness of this proposal, the new discretized variable was compared with a global single-item measure, under the hypothesis that this item is a possible observed discrete version of the latent variable, making use of suitable indicators of resemblance between discrete variables.

The performances of the proposed discretizer plus three standard ones were evaluated making use of synthetic and real datasets, the last ones referring to eight distinct latent traits. The graphical inspection of the distributions as well as the analysis of the similarities between discretized and observed measures through the three indicators recalled in Section 4 demonstrate that, at least in the considered cases and under the three simulation conditions, the proposed method outperforms the three standard ones considered here, so it represents an improvement respect what is available in literature. These results suggest that the proposed method has chances to give a good discretization of the underlying continuous latent variable measured by a model belonging to the family of Rasch models.

The idea underlying the proposed method can be applied to measures obtained with each model belonging to the IRT class of models, so further research can explore the validity of the method applied to measures estimated with different measurement models. Moreover, another line of future research is to extend the comparison of the proposed method to approaches that take into account proper target functions, as shown in Boari and Rusconi (2015) and Carpita and Manisera (2012).

Acknowledgement

The author wishes to thank Marco Sandri and the anonymous reviewers for their useful advices.

References

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

Bagozzi, R. P., and Heatherton, T. F. (1994). A general approach to representing multifaceted personality constructs: application to state self-esteem. *Structural Equation Modeling*, 1(1):35–6.

Bartolucci, F., Dardanoni, V., and Peracchi, F. (2015). Ranking scientific journals via latent class models for polytomous item response data. *Journal of Royal Statistical Society A*, 178(4):1025–1049.

Bartolucci, F., Bacci, S., and Gnaldi, M. (2016). Statistical analysis of questionnaires. A unified approach based on R and Stata. CRC Press.

Boari, G., and Nai Ruscone, M. (2015) A procedure simulating Likert scale item responses. *Electronic Journal of Applied Statistical Analysis*, 8(3):288–297.

Bond, T. G., and Fox, C. M. (2007). Applying the Rasch Model. Fundamental measurement in the human sciences, 2nd edn. Lawrence Erlbaum Associates, Mahwah, NJ .

Bressan, F., Pedrazza, M., and Neve, E. (1992). Il precorso formativo dell'assistente sociale. Autovalutazione e benessere professionale, Franco Angeli, Milano.

Carpita, M., and Golia, S. (2012). Measuring the quality of work: the case of the Italian social cooperatives. *Quality and Quantity*, 46:1659–1685.

Carpita, M., and Manisera, M. (2012) Constructing indicators of unobservable variables from parallel measurements. *Electronic Journal of Applied Statistical Analysis*, 5(3):320–326.

de Boer, A. G. E. M., van Lanschot, J. J. B., Stalmeier, P.F. M., van Sandick, J. W., Hulscher, J. B. F., de Haes, J. C. J .M., and Sprangers, M. A. G. (2004). Is a single-item visual analogue scale as valid, reliable, and responsive as multi-item scales in measuring quality of life? Quality of Life Research, 13:311–320.

Diamantopoulos, A., Sarstedt, M., Fuchs, C., Wilczynski, P., and Kaiser, S. (2012). Guidelines for choosing between multi-item and single-item scales for construct measurement: a predictive validity perspective. *Journal of the Academy of Marketing Science*, 40:434–449.

Dillon, W., and Goldstein, M. (1984). Multivatiate analysis. Wiley, New York.

Dougherty, J., Kohavi, R., and Sahami, M. (1995). Supervised and unsupervised discretization of continuous features. Proc. Twelfth International Conference on Machine Learning (ICML), Los Altos, CA, Morgan Kaufmann, 194–202.

Eurofound (2013). Third European Quality of Life Survey - Quality of life in Europe: Subjective well-being, Publications Office of the European Union, Luxembourg.

European Foundation for the Improvement of Living and Working Conditions (2014). European Quality of Life Survey, 2011-2012 [computer file], 2nd edition, Colchester, Essex: UK Data Archive [distributor], SN: 7316, http://dx.doi.org/10.5255/UKDA-SN-7316-2, 2014.

Fors, F., and Kulin, J. (2016) Bringing affect back in: measuring and comparing subjective well-being across countries. *Social Indicator Research*, 127:323–339.

Fuchs, C., and Diamantopoulos, A. (2009). Using single-item measures for construct measurement in management research. *Die Betriebswirtschaft/Business Administra-*

tion Review, 69(2):195–210.

Gan, G., Ma, C., and Wu, J. (2007). Data Clustering: Theory, Algorithms, and Applications, ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA.

García, S., Luengo, J., Sáez, J. A., López, V., and Herrera, F. (2013). A survey of discretization techniques: taxonomy and empirical analysis in supervised learning. *IEEE Transaction on Knowledge and Data Engineering*, 25(4):734–750.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27:857–874.

Han, J., Kamber, M., and Pei, J. (2012).Data Mining - Concepts and Techniques, 3rd Edition. Morgan Kaufmann Publ., Waltham, MA, USA.

Hartigan, J. A., and Wong, M. A. (1979). Algorithm AS 136 A K-Means Clustering Algorithm. *Applied Statistics*, 28(1):100–108.

Kjærulff, U. B., and Madsen, A. L. (2013). Bayesian networks and influence diagrams: a guide to construction and analysis. Springer, New York.

Lee, H., Delene, L. M., Bunda, M. A., and Kim, C. (2000). Methods of measuring health-care service quality. *Journal of Business Research*, 48:233–246.

Linacre, J. M. (2011). Winsteps (Version 3.72.3) [Computer Software], Beaverton, Oregon: Winsteps.com.

Liu, H., Hussain, F., Tan, C. L., and Dash, M. (2002). Discretization: An Enabling Technique. *Data Mining and Knowledge Discovery*, 6:393–423.

Lucas, R. E., and Donnellan, M. B. (2012). Estimating the reliability of single-item life satisfaction measures: results from four national panel studies. *Social Indicator Research*, 105:323-331.

Lustgarten, J. L., Gopalakrishnan, V., Grover, H., and Visweswaran, S. (2008). Improving Classification Performance with Discretization on Biomedical Datasets. AMIA 2008 Symposium Proceedings, 445–449.

McCutcheon, A. L. (1987). Latent Class Analysis (Quantitative Applications in the Social Sciences, 64). Sage Publications.

Nagy, M. S. (2002). Using a single-item approach to measure facet job satisfaction. *Journal of Occupational and Organizational Psychology*, 75:77–86.

Office for National Statistics, University of Manchester. Cathie Marsh Institute for Social Research. UK Data Service (2016). SN: 7913, Opinions and Lifestyle Survey, Well-Being Module, April-May 2015: Unrestricted Access Teaching Dataset, 2nd edition. http://doi.org/10.5255/UKDA-SN-7913-2.

Postmes, T., Haslam, S. A., and Jans , L. (2013). A single-item measure of social identification: Reliability, validity, and utility. *British Journal of Social Psychology*, 52:597–617.

Queensland Government (2015). Customer Satisfaction Survey 2015, http://data.qld.gov.au/dataset/customer-satisfaction-survey-2015.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19:305–335.

Sackett, P. R., and Larson, J. R. Jr. (1990). Research strategies and tactics in industrial and organizational psychology. M. D. Dunnette and L. M. Hough, Eds., Handbook of industrial and organization psychology (2nd ed.) Vol. 1, Consulting Psychologists Press, Palo Alto, CA, 419–489.

Sloan, J. A., Aaronson, N., Capelleri, J. C., Fairclough, D. L., and Varricchio, C. (2002). Assessing the clinical significance of single items relative to summated scores. *Mayo Clinic Proceedings*, 77:497–487.

Van Ryzin, G. (2004). The measurement of overall citizen satisfaction. *Public Performance and Management Review*, 27(3):9–28.

Waddimba, A. C., Scribani, M., Nieves, M. A., Krupa, N., May, J. J., and Jenkins, P. (2016). Validation of Single-Item Screening measurement for provided burnout in rural health care network. *Evaluation and the Health Professions*, 39:215–225.

Wanous, J. P., Reicher, A. E., and Hudy, M. J. (1997). Overall Job Satisfaction: How Good Are Single-Item Measures. *Journal of Applied Psychology*, 82:247–252.

Wright, B. D., and Master, G. N. (1982). Rating scale analysis, MESA Press, Chicago.