Pak.j.stat.oper.res. Vol. 22 No. 3 2026 pp 361-376 DOI: http://dx.doi.org/10.18187/pjsor.v22i3.5067

Pakistan Journal of Statistics and Operation Research

Item response model proposal under a uniform latent variable assumption: Application to school bullying data

Edilberto Cepeda-Cuervo[1], Vicente Núñez-Antón[2]

*Corresponding author

1. Department of Statistics, National University of Colombia, Colombia, ecepedac@unal.edu.co
2. Department of Quantitative Methods, University of the Basque Country (EHU), Spain, vicente.nunezanton@ehu.eus

Abstract

School bullying victimization is a variable that cannot be directly measured. Taking into account that this variable has a lower bound given by the absence of bullying victimization, we propose the use of Item Response Theory (IRT) logistic models, where the ability or latent parameter ranges from 0 to a positive real number R, adequately defining the IRT parameters and also including an empirical anchor estimation procedure. As academic abilities and school bullying victimization can be explained by associated factors such as habits, sex, socioeconomic level and education level of parents, IRT regression models are proposed to be able to perform joint inferences about individual and school characteristic effects.

Key Words: Logistic models; Ability modeling; Item response theory; Bullying data.

Mathematical Subject Classification: 62F15, 62P15.

1. Introduction

In social sciences, research interest usually includes characteristics, such as cognitive development, quality of life or level of victimization, which cannot be directly measured, see Goodboy and Martin(2015) and Wilson(1989). The information for the specific study of these variables is collected by using instruments, like questionnaires or tests, to observe people's behavior/performance, so that questions related to the variable of interest are asked and, in addition, where there is not a defined and/or interpretable scale available in advance, see Bartholomew(1980, 1984).

In order to be able to analyze and model these unobservable variables from observed variables, latent structure models, also known as Item-Response Models (Muñiz, 1997) have been proposed, see Birnbaum(1968), Lazarsfeld(1855) and Lord(1950). These proposals focus on modelling the relationship between the various reactions, items or behaviors observed with the unobservable variable of interest, also called the latent variable. More specifically, item-response theory (IRT) proposes a relationship between an unobserved variable (usually called ability) and the probability of a person correctly responding to any item of a test, see Lord(1980). The most commonly used and best known IRT models are the one-parameter, two-parameter and three-parameter logistic models. All of them assume that the underlying unobserved variable (i.e., the latent variable) ranges from $-\infty$ to ∞ although, in practical analysis, it is assumed that it takes values between -3 and 3, with the ability scaled so that it has a mean equal to 0 and a standard deviation equal to 1, see Harris(1989). Alternative proposals have assumed that the latent variable ranges from -4 to 4, see Ludlow and Haley(1995). These IRT logistic models have an item difficulty parameter, corresponding to the inflection of the item's characteristic curve; that is, the curve of the probability of correctly answering an specific item as a function of the ability. This parameter ranges, theoretically, from $-\infty$ to ∞ (or, practically, from -3 to 3 or from -4 to 4). In these models, parameter estimates are usually obtained under the assumption that subjects' abilities are random samples from an $N(0,1)$ ability distribution, see Bock and Lieberman(1970). Parameter estimates can also be obtained without the requirement of any arbitrary assumption about the distribution of the ability in the sampled population, see Bock and Aitkin(1981), by using estimations from a discrete distribution at a finite number of points (i.e., a histogram).

The need for testing reliable procedures based on statistical models has increased in the different fields. Item-based statistical models are used to measure the quality of education, development of communicative skills and mathematical ability, among others. In this paper, we propose the use of these models to measure the intensity of school bullying victimization. However, since there is a lower bound of victimization given by absence of bullying and an upper bound assumed by the test, an IRT model with ability or latent parameter ranging from $-\infty$ to $+\infty$ is not consistent with the conditions present in this specific application. As a consequence, a new model specification should be considered to be able to take into account the specific conditions of the test and also to be able to determine the true distribution of the latent variable (i.e., the ability or bullying victimization). This is the main motivation leading us to propose alternative IRT models where the range for the individual ability goes from 0 to R, where R is a positive real number. In the former case, the model is defined by assuming a logarithmic transformation for the abilities and, in the latter case, by assuming a logit, probit or log-log transformation after rescaling the abilities from the interval $(0, R)$ to the interval $(0, 1)$. Therefore, a special modelling approach is considered under a Bayesian perspective, where uniform prior distributions are assumed for the difficulty and subject latent parameters. In this model, parameter estimates are obtained under the assumption that subjects' abilities are random samples from an a uniform ability distribution, $U(0, R)$, which represent a general and not too restrictive informative prior distribution for the abilities than that of the normal standard distribution. Additionally, this paper also introduces an anchoring procedure for the scale introduced by Beaton and Allen(1992), as well as a proposal for empirically obtaining the performance levels, which allows for an easier and a more appealing interpretation for the latent parameter scale.

The proposed IRT models are fitted to the Bogotá's school bullying data, finding that they are able to capture the skewness of the latent parameters (i.e., bullying victimization). Moreover, after the estimation of the parameters involved in the models and the anchoring processes of the items, see Andrade et al.(2000) and Beaton and Johnson(1992), the values of the scale that allows us to have an appealing and understandable interpretation with respect to the construct (i.e., bullying victimization) is appropriately selected. In addition, since bullying victimization can be explained by some of the associated factors, such as habits, sex, socioeconomic level and education level of parents, IRT regression models can be seen as reasonable models to analyze the Bogotá's bullying data. We start by introducing the three-parameter logistic models, where the range for the individual latent parameter (ability) goes from 0 to ∞, and from 0 to R, where R is a positive real number. We then present the ability scale and our modelling proposals. Finally, we present the results from the analysis of the Bogotá bullying data and end with some conclusions and practical recommendations.

In summary, the present work was designed to make three main contributions. Firstly, our goal was to propose alternative more robust IRT models that adapt their range to the needs of the specific application under study. Secondly, we propose a Bayesian alternative estimation method for the proposed IRT models, together with an anchoring empirical estimation procedure for the scale. Thirdly, we illustrate the usefulness of the proposed procedure by applying it to the Bogota's school bullying data, the motivating dataset for our proposals.

2. Three Parameter Logistic Model

We propose alternative item-response models to study the school bullying victimization phenomenon. The proposed models present a well motivated and reasonable alternative to analyze sets of dichotomic items taking into account specific theoretical settings in educational and psychological tests, where the latent and difficulty parameters range in a positive bounded interval of the real numbers. These models provide parameter estimates in a scale with an easy and understandable interpretation for practitioners in the area, who are the ones constructing, applying and interpreting these specific questionnaires.

2.1. IRT Model With Abilities Ranging in the Real Numbers Set

We start by presenting the traditional three-parameter IRT model. Let us assume that u_{ij}, $i = 1, \ldots, I$, and $j = 1, \ldots, n$, are $n \times I$ binary random variables, where i indicates an item and j indicates a subject. That is, $u_{ij} = 1$ if subject j solves item i correctly, and $u_{ij} = 0$, otherwise. In this model, the probability that subject j with ability θ_j solves item i correctly, with difficulty parameter b_i, is given by:

$$P(u_{ij} = 1|\theta_j, \xi_i) = c_i + (1 - c_i)\frac{1}{1 + e^{-Da_i(\theta_j - b_i)}},\ i = 1, \ldots, I, j = 1, \ldots, n, \tag{1}$$

where $\xi_i = (a_i, b_i, c_i)$ are the parameters for item i, with $0 \leq c_i < 1$ and $a_i > 0$, see Birnbaum(1968). In this model, θ_j and b_i can take on values between $-\infty$ and $+\infty$. The latent continuum θ is called ability, and θ_j represents the ability of the j-th subject. In addition, D is a constant which can be arbitrarily set. However, it is customary to set $D = 1.7$. The following considerations provide a basis for interpreting the parameters involved in the IRT model:

1. As $c_i = \lim\limits_{\theta_j \to -\infty} P(u_{ij} = 1|\theta_j, \xi_i)$, c_i can be defined as the probability of random correct responses.

2. If $\theta_j = b_i$, we have that, from equation (1),

$$P(u_{ij} = 1|\theta_j, \xi_i) = \frac{1 + c_i}{2}. \tag{2}$$

   Therefore, b_i can be interpreted as the ability necessary for a subject to solve item i with probability equal to $\frac{1+c_i}{2}$. As a consequence of equation 2, for the same probability of random correct responses c_i and given that items with larger difficulty values need larger ability to have a correct answer probability equal to $\frac{1+c_i}{2}$, large values of b_i correspond to items with large difficulty.

3. If we assume that $P(u_{ij} = 1|\theta_j, \xi_i)$ is a function $f(\cdot)$ of θ_j, the first order derivative of f evaluated at b_i is given by:

$$f'(b_i) = \frac{1}{4}(1 - c_i)Da_i.$$

   In this way, given b_i and c_i, a_i can be assumed as a measure of discrimination. That is, if $c_1 = c_2 = c$ and $a_1 < a_2$, an item with parameters c and a_2 discriminates more than an item with parameters c and a_1.

In equation (1), if $c_i = 0$, then there is no random possibility of correct answer and, therefore, the resulting model is called a two-parameter univariate logistic model or Birnbaum's model, see Van der Linden and Hambleton(1997). If we set $a_i = a$ and $c_i = 0$ in equation (1), we obtain the one parameter univariate logistic model, also known as Rasch model, see Rasch(1980). The Rasch model describes how the probability of correct answers depends on the subject's overall ability and the level of difficulty of the questions. Moreover, one of the most important theoretical merits of the Rasch model is the so-called "specific objectivity," see Fisher(1995). This means that the item parameters do not depend on the characteristics of the person answering the test, and that the personal parameters do not depend on the items, specifically chosen from a set of items. Consequently, the three-parameters are independent of the average ability of the sample and its respective dispersion.

In addition, given an item and known parameters, the curve describing the relationship between the subjects' ability parameter and the agreement probability is called the Item Characteristic Curve, see Hambleton et al.(1991). Figure 1 shows this curve for an item with $D = 1.7$ and parameters $a_i = 1.5$, $c_i = 0.25$ and $b_i = 1$. The curve shows that the probability of a correct answer $P(u_{ij} = 1|\theta_j)$ increases when θ_j increases. This means that, for example, in mathematical performance, the probability of a correct response increases with the subject's mathematical ability.

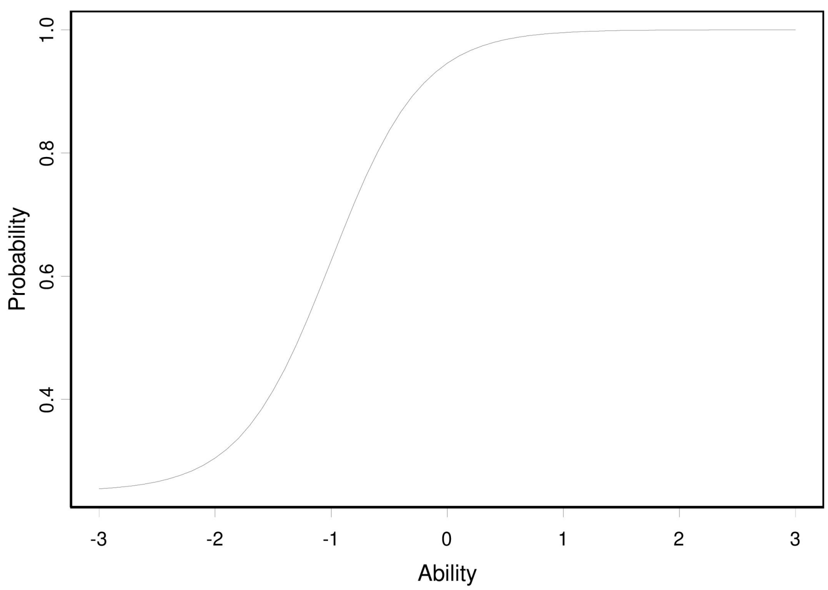

Figure 1: Item Characteristic Curve for $a_i = 1.5$, $b_i = 1$, $c_i = 0.25$ and $D = 1.7$.

3. Alternative IRT Response Model Proposals

In model (1), the latent variable is assumed to have a range in the set of real numbers and, in addition, to follow a standard normal distribution, which represents a strong prior assumption for the latent variable. Moreover, we should take into account two very relevant issues with regard to this matter. First of all, in the evaluation of communicative skills in a foreign language or in the study of student bullying victimization, the latent variable does not follow, in general, a standard normal distribution, providing instead a clear evidence of following, for example, a left or a right skewed distribution. However, in the usual IRT data analysis, a standard normal prior distribution is assumed for the latent variable, for either frequentist or Bayesian estimation methods. In both cases, this represents a strong assumption that may have a very relevant influence on the parameter estimates. Second of all, as will be seen in Section 5 for the specific application motivating our proposals, in the scholar bullying victimisation case, or even in other multiple applications, there is a lower and an upper bound for the latent variable, which are determined by the framework of the test. In summary, if we consider the aforementioned issues, and to be able to overcome the limitations they may produce in the parameters estimation procedure, we propose an IRT model with abilities ranging in a bounded open interval $(0, R)$, where R is positive a real number. In addition, a uniform distribution is assumed for the latent variable providing.

3.1. IRT model with uniform prior distributions

The three-parameter uniform model is defined as a three-parameter normal model, assuming that the probability that subject j with ability θ_j solves item i, with difficulty parameter b_i, is given by:

$$P(u_{ij} = 1|\theta_j, \xi_i) = c_i + (1 - c_i)\frac{1}{1 + e^{-Da_i[\text{logit}(\theta_j/R) - \text{logit}(b_i/R)]}},\ 0 < \theta_j < R,\ 0 < b_j < R, \tag{3}$$

$i = 1, \ldots, I,\ j = 1, \ldots, n$, where $\xi_i = (a_i, b_i, c_i)$ are the parameters of item i, with $0 \leq c_i < 1$ and $a_i > 0$. Here, θ_j and b_i take on values in the open interval $(0, R)$. The latent continuum θ represents the ability, and θ_j is the ability of the j-th subject. As $c_i = \lim\limits_{\theta_j \to 0^+} P(u_{ij} = 1|\theta_j, \xi_i)$, c_i can be defined as the probability of random correct responses. Moreover, if $\theta_j = b_i$, from equation (3) it follows that:

$$P(u_{ij} = 1|\theta_j, \xi_i) = \frac{1 + c_i}{2}.$$

In this way, b_i can be interpreted as the subject's ability necessary to solve item i with probability equal to $\frac{1+c_i}{2}$. Larger values of b_i correspond to a larger item difficulty.

If we assume that $P(u_{ij} = 1|\theta_j, \xi_i)$ is a function $f(\cdot)$ of θ_j, the first-order derivative of f with respect to θ_j, evaluated at b_i, is given by:

$$f'(b_i) = \frac{1}{4}(1 - c_i) D a_i \frac{R}{(R - b_i) b_i}.$$

Therefore, given c_i and b_i, a_i can be treated as measure of discrimination. If $c_1 = c_2$, $b_1 = b_2$ and $a_1 < a_2$, where the subscript represents the item number, then item 2 discriminates more than item 1 between subjects with different ability levels. That is, the discrimination depends on the difficulty parameter. Figure 2 shows the characteristic curve for the uniform item response model.

In equation (3), if $c_i = 0$, there is no random possibility of correct responses, and the resulting model is called the two-parameter univariate logistic model or the C-Birnbaum model. If $a_i = a$ and $c_i = 0$, we have the one-parameter univariate logistic model or the C-Rasch model. The C-Rasch model describes how the probability of correct responses depends on the subject's overall ability and the level of difficulty of the questions.

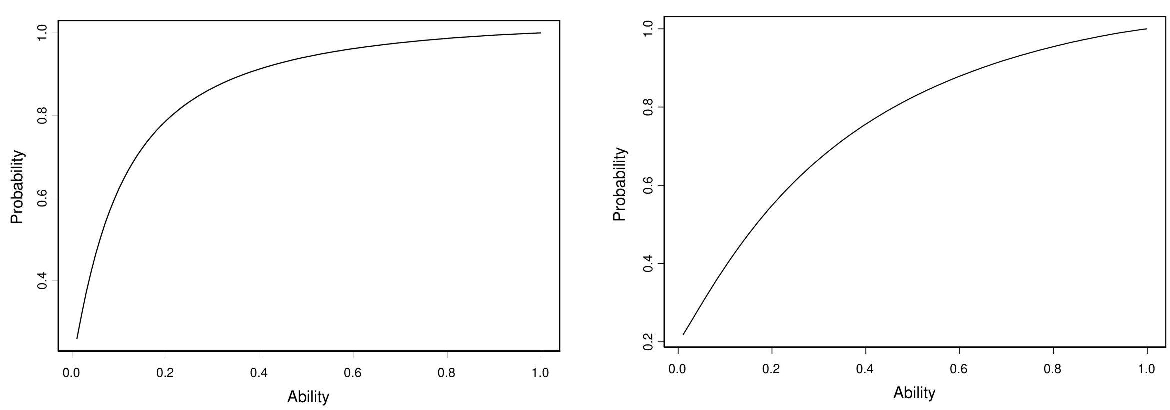

Figure 2: Item Characteristic Curve for the uniform item response model with $D = 1.7$, and $a_i = 1.5$, $b_i = 0.5$ and $c_i = 0.25$ (left panel), and $a_i = 1.5$, $b_i = 1$ and $c_i = 0.25$ (right panel).

In the three-parameter uniform model in equation (3), there are additional possible choices for the transformation functions, apart from the logit specification, $\text{logit}(\theta/R) = \log(\theta/(R - \theta))$. Some examples of this are the probit function $g(\theta) = \Phi^{-1}(\theta/R)$, where $\Phi(.)$ is the cumulative distribution function of a standard normal random variable; the complementary log-log function $g(\theta) = \log[-\log((R - \theta)/R]$, and the log-log function $g(\theta) = -\log[\log(R) - \log(\theta)]$, among others.

Along these lines, and from a practical point of view, a good choice for R is usually $R = 1$. Alternative choices for possible values for R are, for example, $R = 5$ or $R = 10$. In all cases, and given that different values of R correspond to different equivalent scales, results are comparable. Therefore, a specific criterion to be able select the value of R

would depend on the convenience for the communication and interpretation of the specific results of the latent variable values.

4. Ability Scales

We start by mentioning that, given that the ability scale does not have a practical interpretation on its own (i.e., in an educational context), it is necessary to define a practical ability scale that is characterized by sets of items that provide a pedagogical interpretation in the theoretical context of the test. The ability scale is defined by a *set of values* $\theta_0 < \theta_1 < \cdots < \theta_P$ of θ, labelled as "anchor levels", selected by the researcher analyzing the data under study, see Andrade et al.(2000). In our specific applications, the interpretation of the scale, determined by the bullying victimization levels, is possible by taking into account the psychological interpretation of the items that are associated to each level.

In addition, the specific anchor levels, θ_p, $p = 1, 2, \ldots, P$, depend on the conditional probabilities of a correct answer $P(u = 1|\theta = \theta_p)$ and $P(u = 1|\theta = \theta_{p-1})$, see Beaton and Allen(1992) and Andrade et al.(2000). More specifically, an item u is an anchor item of level θ_p if and only if the following conditions hold:

1. $P(u = 1|\theta = \theta_p) \geq 0.65$,
2. $P(u = 1|\theta = \theta_{p-1}) < 0.5$, and
3. $P(u = 1|\theta = \theta_p) - P(u = 1|\theta = \theta_{p-1}) \geq 0.30$,

where the anchor values θ_p and θ_{p-1} are related to conditions 1 and 2, and the third condition is related to the difference between the two previous probabilities.

We propose an empirical procedure for the selection of the ability levels for a given set of items. This method, originally introduced by Cervantes et al.(2008), claims that, for a given item, it is possible to search for a pair of points, (θ_ℓ, θ_h), in the ability scale that meets the necessary criteria for the anchoring in such a way that any ordered pair of points (θ_1, θ_2), for which $\theta_1 \leq \theta_\ell$ and $\theta_2 \geq \theta_h$, would allow anchoring the item at level θ_2 related to level θ_1, and where, in addition, we have that $\theta_1 > \theta_\ell$ and $\theta_2 < \theta_h$ would not allow anchoring the item. Moreover, the pair (θ_ℓ, θ_h) provides the shortest interval in the ability scale for which it is possible to anchor the item. For the one and two-parameter models, the pair is given by the values of (θ_ℓ, θ_h) such that $P(u = 1|\theta_\ell) = 0.35$ and $P(u = 1|\theta_h) = 0.65$, respectively. For the three-parameter models, this interval is given by the values of θ such that:

1. $P(u = 1|\theta_\ell) = \frac{c+1}{2} - 0.15$ and $P(u = 1|\theta_h) = \frac{c+1}{2} + 0.15$, if $c < 0.3$
2. $P(u = 1|\theta_\ell) = 0.35$ and $P(u = 1|\theta_h) = 0.65$, if $0.3 \leq c < 0.35$
3. $P(u = 1|\theta_\ell) = 0.5 - \epsilon$ and $P(u = 1|\theta_h) = 0.8 - \epsilon$, if $0.35 \leq c < 0.5$ and $\epsilon > 0$

If $c \geq 0.5$, no pair (θ_ℓ, θ_h) meets all the three criteria included above. Once the intervals for each item have been determined, the procedure is to find P and θ_p, $p = 1, 2, \ldots, P$ such that it is possible to include the items in P levels. More specifically, we describe the steps for the specific procedure as follows:

1. Sort the items in ascending order, according to the θ_ℓ's
2. Take the smallest θ_ℓ at the lowest level θ_0
3. If $p = 1$, take the candidate θ^* for θ_1 to be the θ_h corresponding to pair (θ_0, θ_h). If $p = 2, \ldots, P$, take the candidate θ^* for θ_p to be the θ_h corresponding to pair (θ_p, θ_h).
4. Compare the candidate with the interval found for the next item:
   (a) If $\theta_\ell < \theta^* < \theta_h$, let $\theta^* = \theta_h$ and repeat the comparison with the interval for the next item

   (b) If $\theta^* \geq \theta_h$, keep the candidate and make the comparison with the following interval
   (c) If $\theta^* \leq \theta_\ell$, then θ_p determines the p-th level of performance

5. Repeat steps 3 and 4 to determine the levels until having compared all of the items

This procedure is proposed as a tool to be able to describe the ability scale so that the researcher can find points that clearly define the different performance levels, see Beaton and Allen(1992). A representation of the intervals for the different items, ordered according to their lower limit, allows for the possibility of identifying levels of achievement, determined by "jumps" between the anchor intervals. If the proposed procedure does not allow the finding of the different ability levels, eliminating some of the items of the test would make it possible to define the specific levels of performance. The interpretation of the items anchored in levels seeks to provide a generalization of the abilities for the population under study.

In summary, the proposed procedure to be able to empirically obtain the different performance levels (here, bullying victimization levels) allows us to have a closer approximation of the attributes present in the measurement instrument that effectively compose it. This procedure also allows to achieve the main purpose we have for these specific proposals, which is the anchorage of the item scale in the sense of: 1) finding, for a performance level, a set of items that belongs to each of these levels; 2) facilitating the formulation of a theory about the relationship among the different performance levels of the attribute evaluated in the individuals and the items that compose the test; and 3) facilitating the communication and interpretation of the results of the evaluation to people with no expertise in the specific area.

4.1. Modeling of abilities

In most applications and areas, response data are often obtained in combination with other input variables. For example, response data are obtained from respondents together with school information, and the objective is to be able to perform a joint inferential process about individual and school effects for a given outcome variable. In a Bayesian framework, different sources of information can be handled efficiently, accounting for their level of uncertainty. We will justify that the flexibility of a Bayesian modelling approach, together with the powerful computational methods available for it, provides an attractive set of tools for analyzing response data.

A subject's ability can be explained by associated available variables such as, for example, habits, practice time, socioeconomic level and parents' educational level, with the use of the model $\theta_j = \mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta} + \epsilon$, where $\mathbf{x}_j^{\mathrm{T}} = (x_{1j}, \ldots, x_{rj})$ is an explanatory vector of variables, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_r)^T$ is a vector of unknown regression parameters that needs to be estimated, and $\epsilon \sim N(0, \sigma^2)$. Therefore, model (1) can be then written as:

$$p(u_{ij} = 1|\boldsymbol{\beta}, \xi_i, \epsilon, \mathbf{x}_j) = c_i + (1 - c_i)\frac{1}{1 + e^{-Da_i\left(\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta} + \epsilon - b_i\right)}}, \tag{4}$$

for $i = 1, \ldots, I$ and $j = 1, \ldots, n$. In the three-parameter uniform model in equation (3), if we set logit$(\theta_j/R) = \mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta} + \epsilon$, we can obtain a regression model similar to that in equation (4).

Interpreting the regression parameters is something where additional insights should be provided. The effect of a unit change in the k-th covariate; that is, a change in the covariate value from from x_{kj} to $x_{kj} + 1$, can be illustrated for the specific case in model (4). If we assume that $\tilde{\mathbf{x}}_j^{\mathrm{T}} = (x_{1j}, \ldots, x_{kj} + 1, \ldots, x_{rj})$, this model can be written as:

$$p(u_{ij} = 1|\boldsymbol{\beta}, \xi_i, \epsilon, \tilde{\mathbf{x}}_j) = c_i + (1 - c_i)\frac{1}{1 + e^{-Da_i\beta_k}e^{-Da_i\left(\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta} + \epsilon - b_i\right)}}. \tag{5}$$

Therefore, if $\beta_k > 0$, larger values of x_{kj} are associated with a larger probability of correct responses, given that $e^{-Da_i\beta_k} < 1$ and that $p(u_{ij} = 1|\boldsymbol{\beta}, \xi_i, \epsilon, \tilde{\mathbf{x}}_j) > p(u_{ij} = 1|\boldsymbol{\beta}, \xi_i, \epsilon, \mathbf{x}_j)$. If $\beta_k < 0$, larger values of x_{kj} are associated with a smaller probability of correct responses, given that $e^{-Da_i\beta_k} > 1$ and that $p(u_{ij} = 1|\boldsymbol{\beta}, \xi_i, \epsilon, \tilde{\mathbf{x}}_j) < p(u_{ij} = 1|\boldsymbol{\beta}, \xi_i, \epsilon, \mathbf{x}_j)$. In this specific parameter interpretation, we have to remember that it is assumed the all of the remaining covariates are held fixed.

Parameter estimates obtained from the application of these models are helpful to determine, for example, the current status or current characteristics of the system, so that the researcher will be able to determine strategies to improve educational quality. This is one of the main reasons to use these models to study school bullying, so that results will be used to determine if bullying victimization is more present in the early or late stages of the secondary school period, or if there are differences in the bullying victimization between boys and girls. These results will be extremely helpful to establish and prioritize governmental prevention policies.

5. School Bullying Application

5.1. School bullying definition

Based on the Cisneros Autotest, see Oñate and Piñuel(2005), Cepeda-Cuervo et al.(2008) proposed a theoretical framework and a 22-item questionnaire to be able to inquire about the harassment situation of students at school. Each of the items included in this questionnaire corresponds to a "harassment strategy". Moreover, each item consists of a statement and two response options (often and sometimes-never), from which the student chooses one according to the frequency with which, in each case, the harassment situation described in each of the statements occurs. The authors conducted a survey consisting of 22 items related to situations and behaviors that characterize school bullying in 709 public schools in Ciudad Bolívar, a district of Bogotá city. The survey is included in Appendix 1.

5.2. School bullying in Bogotá, Colombia

The population of interest includes primary and middle school students of 709 public schools in Ciudad Bolívar, Colombia, which serves more than 28000 students. The data obtained from a sample of 80 classes, 19 of sixth grade, 12 of seventh grade, 12 of eighth grade, 10 of ninth grade, 10 of tenth grade and 17 of eleventh grade. Therefore, the sample consists of 3226 students aged between 10 and 20 years old, mainly drawn from low level socioeconomic households.

Each item is dichotomic, assigning a value of 0 if the student "sometimes or never" was a victim of the situation described by the item and a value of 1 if the student was a "frequent victim" of that situation. Therefore, given that the probability of a random answer is 0, the probability of victimization is assumed to follow one of the logistic functions given by:

1. **Model 1a:**

$$\begin{aligned} &\text{logit}\left[P(u_{ij}=1|\theta_j,b_i)\right]=\theta_j-b_i, \\ &\theta_j\sim N(0,1)\ \ j=1,\ldots,3226;\ i=1,\ldots,22, \end{aligned} \tag{6}$$

   where θ_j is a measure of student's victimization and b_i is a measure of the "item victimization impact".

2. **Model 2a:**

$$\begin{aligned} &\text{logit}\left[P(u_{ij}=1|\theta_j,b_i)\right]=\text{logit}(\theta_j/5)-\text{logit}(b_i/5), \\ &\theta_j\sim U(0,5)\ \ j=1,\ldots,3226;\ i=1,\ldots,22, \end{aligned} \tag{7}$$

   where θ_j and b_i can be interpreted as in Model 1a. In order to be able to define a completely specified Bayesian model, a uniform prior distribution, $U(0,R)$, is assumed for the difficulty parameters b_i's. More specifically, if $R=1$, a $U(0,1)$ uniform distribution is assumed as a prior distribution for all of the b_i's.

Appendix 2 includes the Winbugs program for fitting the IRT uniform Rasch models. In order to be able to illustrate the estimation process and allow for the reproducibility of the model fits presented here, keeping the confidentiality of the school bullying dataset as much as possible, we have introduced a subset of the original dataset for 30 students. We do believe that the provided Winbugs program for this small sample can help users and practitioners to better illustrate the good performance of the aforementioned proposed model, as well as its possible application to different datasets within this context. For the usual IRT models, when assuming a standard normal distribution, the winbugs corresponding program has already been provided and it is well known in the literature, see Curtis(2010). Initial values for the b_i parameters can be chosen close to zero for model 1a and values in the open interval $(0, R)$ for model 2a. In the Winbugs program, initial parameter values for the θ_j's are randomly generated from the standard normal distribution or from the uniform distribution, depending on the model under analysis.

5.3. Bayesian parameter estimation

Bayesian estimation methods to obtain the posterior parameter estimates require the specification of prior distributions for the parameters in the different proposed models. We assume the following prior distribution specification for all unknown parameters. For the difficulty parameters b_i's, a normal prior distribution, $N(0, 1000)$ (i.e., with mean 0 and variance equal to 1000), is assumed in model 1a, and a uniform prior distribution, $U(0, R)$, with $R = 5$, is assumed in model 2a.

Parameter estimates were obtained from 10000 Gibbs samples drawn, with every 5-th sample recorded, after a burn-in period of 2000 samples. WinBugs was used in these simulation procedures, see Spiegelhalter et al.(2003). Convergence of the Gibbs sampler algorithm was monitored using standard existing methods, such as the trace plots of the simulated samples and parallel chains starting from different initial values to be able to provide indication of chain stationarity. In all applications, the posterior samples showed the same behavior for all chains, providing a strong indication of convergence. That is, in all the models considered reported values for Convergence Diagnosis and Output Analysis (CDOA) belong to the interval $(0.99, 1.06)$, which supports the clear convergence of the chains.

Figure 3 shows the estimates of the difficulty parameter for model 1a (on the left panel), where the level of bullying is assumed to have a normal distribution and for model 2a (right panel), where the level of bullying is assumed to have a uniform distribution. For all models, the difficulty parameter estimates seem to follow the same structures, but they are understandably influenced by the initial distributional assumption for the bullying victimization in the model. If R is different from 5, for example $R = 1$, the parameter estimates of the θ_j's and b_i's are equivalent in the sense that the estimates in the $(0, R)$ scale are simply given by R times the corresponding estimates obtained from the $U(0, 1)$ uniform assumption. Figure 4 includes the histogram of the bullying victimization estimates for models 1a and 2a. The histograms clearly illustrate the influence of the initial distributional assumption for θ in the distribution of the parameter estimates. From the histograms in Figure 4, it seems that an initial uniform distributional assumption for θ is the best option to be able obtain model parameter estimates that can better describe the true shape of the bullying victimization distribution.

5.4. Anchor of the items and empirical selection of levels

We now present the results of the empirical procedure previously described for anchoring the items and selecting the levels for the bullying data, so that we can define the different bullying victimization levels. From the anchor plots included in Figure 5, four groups of items which define a bullying victimization scale can be clearly defined. These groups are characterized by the following specific situations:

1. **Group I**: Items 6, 11 and 22.

   "They criticize me and blame me for anything that I do or decisions that I make"; "They make fun of me and play tricks on me", and "They assign me disparaging nicknames." At this level the students did not report that often or sometimes "They take away my snacks" or "They force me to do things that endanger my physical integrity."

2. **Group II**: Items: 4, 8, 9, 14, 16, 17 and 18.

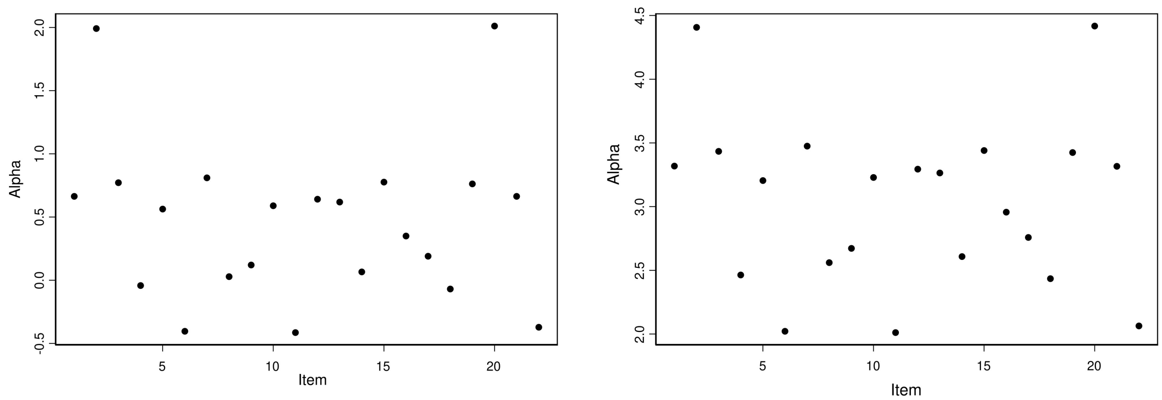

Figure 3: Difficulty parameter estimates for model 1a on the left panel (with normal bullying victimization) and model 2a on the right panel (with uniform bullying victimization). Here Alpha denotes the scale of the b_i parameters

.

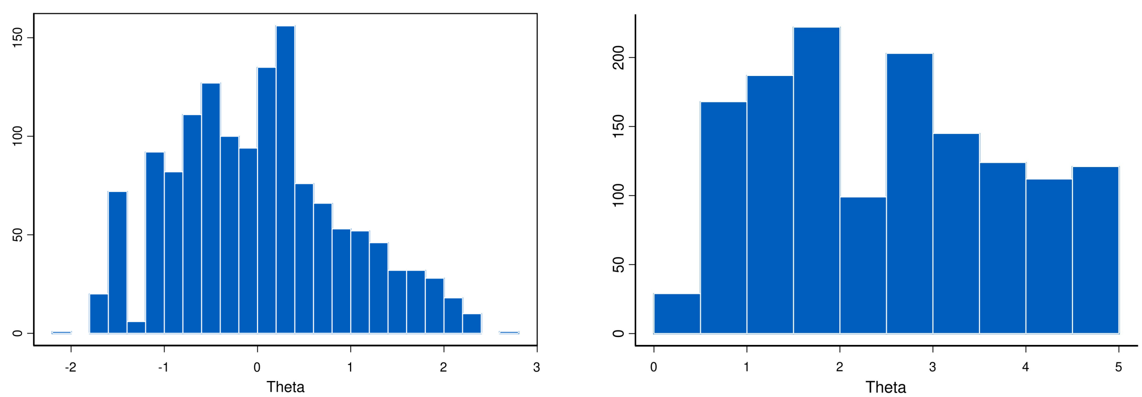

Figure 4: Histograms of the bullying victimization θ, for model 1a on the left panel (with normal distribution) and model 2a on the right panel (with uniform distribution).

"They don't talk to me"; "They don't count on me for any class activity"; "They maliciously distort all that I say or do"; "They continually interrupt me, not allowing me to express myself"; "They make cruel jokes about my physical appearance"; "They throw objects at me"; " They hide my possessions."

3. **Group III**: Items: 1, 3, 5, 7, 10, 12 13, 15, 19 and 21.

   "They force me to do things that I don't want to do"; "They persuade others not to talk to me"; "They despise my work, no matter what I do"; "They humiliate me and make fun of me in front of others"; "They blame me for everything bad that happens"; "They threaten me verbally or through gestures"; "They ignore me and exclude me"; "They prevent me from communicating with others"; "They damage my possessions"; and "They do things to try to get a violent reaction from me."

4. **Group IV**: Items 2 and 20.

   "They force me to do things that endanger my physical integrity" and "They take away my snacks." Students that are victims of bullying characterized by these two items are often victims of the bullying characterized by each of the remaining 20 items.

At the first level of harassment, the student is the victim of criticism, ridicule, jokes and derogatory nicknames. At the second one, s/he is the victim of isolation in class and phenomena of indirect physical and psychological aggression, which is also present in level 3, but with a greater level of intensity in isolation, where the student becomes a victim

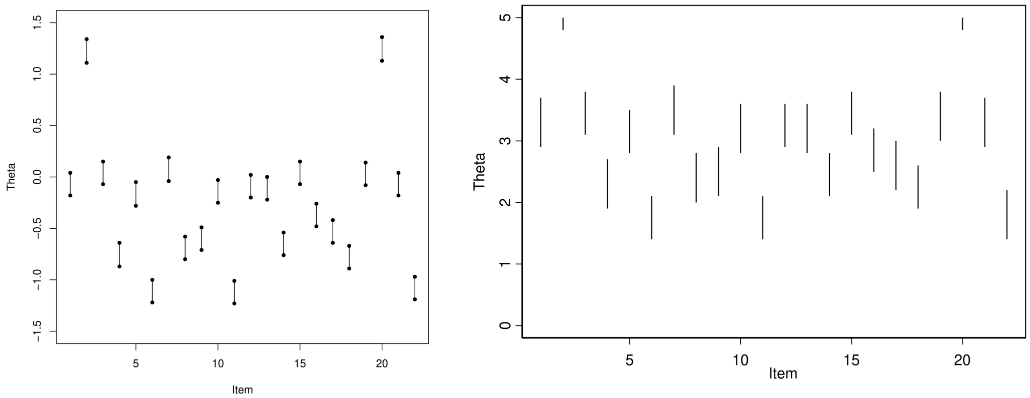

Figure 5: Anchor plots for items in the bullying dataset for model 1a with normal abilities (left panel) and model 2a with uniform abilities (right panel).

of humiliation and threats. At this level, besides facing violence, the victim presents a higher level of submission. Finally, at the fourth level, the victim is totally subjected and may be forced to put his/her physical integrity at risk. In conclusion, these levels characterize a scale where physical and psychological aggression intensifies with the level of the victim submission.

5.5. IRT regression models

We now present the results of fitting the proposed alternative IRT regression models to the bullying data, where we have assumed that:

$$h(\theta_j/R) = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} - b_i + e_j, \;\; e_j \sim N(0, \sigma^2), \tag{8}$$

where $R = 1$ and h is the identity function in model 1a. Moreover, h is the logit, probit or log-log function in model 2a. In this application, the covariate x_1 is defined to be equal to 1 for boys and 0 for girls, and the covariate x_2 to be equal to 1 for sixth, seventh and eighth grades, and 0 for ninth, tenth and eleventh grades. For the analysis presented here, a normal $N(0, 100)$ prior distribution for the regression parameters, having mean equal to 0 and variance equal to 100, and a gamma distribution, $G(0.01, 0.01)$, prior distribution with shape and rate parameters both equal to 0.01, for the variance parameter σ^2. For the difficulty parameters b_i's, standard normal prior distributions are assumed for Model 1b and uniform prior distributions are assumed under the uniform IRT model 2b. In these cases, the submodels fitted to the bullying dataset are the models given by equation (9) under the normal bullying assumption and by equation (10) under the uniform case. Detailed results from the statistical analyses for the aforementioned models where the posterior chains convergence was achieved are as follows:

1. **Model 1b**: We assume the proposed IRT normal regression models with uniform bullying victimization. That is,
$$\text{logit}\,[P(u_{ij} = 1|\beta, b_i] = \beta_1 x_{1j} + \beta_2 x_{2j} - b_i, \tag{9}$$
where β is the vector of the regression parameters and the b_i's are the item victimization impact corresponding parameters. Therefore, if we assume standard normal $N(0, 1)$ prior distributions for $b_i,\ i = 1 \ldots, 22$, and flat normal $N(0, 100)$, with variance equal to 100, prior distributions for $\beta_k,\ k = 1, 2$, the posterior regression parameter estimates, means with their corresponding standard deviations in parenthesis, are given by $\beta_1 = -0.0339(0.0236)$ and $\beta_2 = -0.0562(0.0235)$. Moreover, for this specific model, the deviance information criterion is given by $DIC = 40320.0$.

2. **Model 2b**: We assume the proposed IRT uniform regression model with uniform bullying victimization. That is,

$$\text{logit}\,[P(u_{ij}=1|\beta,b_i]=\beta_0+\beta_1x_{1j}+\beta_2x_{2j}-b_i, \tag{10}$$

where β and the b_i's are as before. In order to specify a general model, if we assume uniform $U(0,1)$ prior distributions for $b_i,\ i=1,\ldots,22$, and flat normal $N(0,100)$, with variance equal to 100, prior distributions for $\beta_k,\ k=0,1,2$, the posterior regression parameter estimates, means with their corresponding standard deviations in parenthesis, are given by $\beta_0=0.0724(0.0295)$, $\beta_1=-0.0286(0.0237)$, $\beta_2=-0.0504(0.0248)$. Moreover, for this specific model, the deviance information criterion is given by $DIC=40630.0$.

From the fitting of all of these models, the same practical conclusions and interpretation can be obtained. That is, there is no bullying victimization difference between girls and boys and the probability of bullying in grades six, seven and eight is larger than that in grades nine, ten and eleven. In our view, it is worth mentioning here that the DIC value for model 2b is slightly larger than that for model 1b. However, we should keep in mind that these models are not nested and, in addition, they have different metrics and, therefore, they are different models and cannot really be directly compared by using this goodness-of-fit criterion. However, we have to point out that, in this specific application, the convergence is better and faster for the uniform regression models. In any case, extensions for future work in this area include the use of simulation processes in order to be able to study and better illustrate the performance of these models. Finally, another possibility for fitting these models is to consider a two-stage approach where the item parameters are fixed at values obtained from the application of the previous IRT models. However, this two-stage approach ignores uncertainty in the item parameters and, in addition, it may also understate uncertainty in the regression parameters.

6. Concluding Remarks

We have proposed new and alternative item-response models by assuming that the abilities range from 0 to a positive real number R. The second interval should be more appealing for the analysis of the school bullying data as it seems to provide us with a more clearly defined scale to explain the bullying intensity parameter distribution. In our view, these new models present new theoretical and applied research possibilities in the area. In school bullying studies, it is commonly considered that tests have items with at least three response categories: frequently, sometimes and never. Therefore, a possible extension of this work would be to be able to propose IRT models for gradual response to the analysis in this type of data, by assuming that the latent variable ranges in a bounded interval. This work is out of the scope of this paper and it is part of current work in progress.

Finally, the empirical procedure presented here is also useful to be able to interpret educational quality tests, in order to understand their theoretical settings, allowing for the proposal of governmental educational policies, for example. The application of the proposed models can be extended to multiples areas of social research, where the settings of the latent variable cannot be assumed to take values from $-\infty$ to ∞. For example, in education quality evaluation, cognitive mathematical achievement or communicative development can be assumed to be latent variables ranging in a bounded interval. These models would then better capture the latent variable distribution and would also facilitate a better interpretation and understanding of the results obtained from the analysis.

Funding Information

This research has been partially funded by Ministerio de Ciencia e Innovación (MCIN, Spain), Agencia Estatal de Investigación (AEI/10.13039/501100011033/) and Fondo Europeo de Desarrollo Regional (FEDER) “Una manera de hacer Europa” under the I+D+i research grant PID2020-112951GB-I00 and by the Department of Education of the Basque Government (UPV/EHU Econometrics Research Group) under research grants IT1508-22 and IT1927-26.

References

1. Andrade, D., Tavares, H., and Cunha, R. (2000). *Teoria da Resposta ao Item: Conceitos e Aplicacões*. Associação Brasileira de Estatística, Sao Paulo, Brazil.

2. Bartholomew, D. (1980). Factor analysis for categorical data. *Journal of the Royal Statistical Society - Series B*, 42(3):293–321.
3. Bartholomew, D. (1984). The foundations of factor analysis. *Biometrika*, 71(2):221–232 doi:10.2307/2336238.
4. Beaton, A. and Allen, N. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics*, 17(2):191–204, doi:10.2307/1165169.
5. Beaton, A. and Johnson, E. (1992). Overview of the scaling methodology used in the national assessment. *Journal of Educational Measurement*, 29(2):163–175, doi:10.1111/j.1745–3984.1992.tb00372.x.
6. Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F. and Novick, M., editors, *Statistical Theories of Mental Test Score*, pages 395–479. Addison-Wesley Educational Publishers Inc., New York, USA.
7. Bock, R. and Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An applications of a EM algorithm. *Psychometrika*, 46(4):443–459, doi:10.1007/BF02293801.
8. Bock, R. and Lieberman, M. (1970). Fitting a response model for n dichotomously scored item. *Psychometrika*, 35(2):179–197, doi:10.1007/BF02291262.
9. Cepeda-Cuervo, E., Pacheco-Durán, P., García-Barco, L., and Piraquive-Peña, C. (2008). Acoso escolar a estudiantes de educación básica y media. *Revista de Salud Pública*, 10(4):517–528.
10. Cervantes, V., Cepeda-Cuervo, E., and Camargo, S. (2008). Una propuesta para la obtención de niveles de desempeño en los modelos de teoría de respuesta al ítem. *Avances en Medición*, 6(1):49–58.
11. Curtis, S. (2010). BUGS code for item response theory. *Journal of Statistical Software*, 36:1–34.
12. Fisher, G. (1995). Some neglected problems in IRT. *Psychometrika*, 60(4):449–487, doi:10.1007/BF02294324.
13. Goodboy, A. and Martin, M. (2015). The personality profile of a cyberbully: Examining the dark triad. *Computers in Human Behavior*, 49:1–4, doi:10.1016/j.chb.2015.02.052.
14. Hambleton, R., Swaminathan, H., and Rogers, H. (1991). *Fundamentals of Item Response Theory*. Sage Publications Inc., London, UK.
15. Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter irt models. *Educational Measurement: Issues and Practice*, 8(1):35–41, doi:10.1111/j.1745–3992.1989.tb00313.x.
16. Lazarsfeld, P. (1855). Recent developments in latent structure analysis. *Sociometry*, 18(4):391–403, doi:10.1098/rsta.1895.0010.
17. Lord, F. (1950). A theory of test scores. Psychometric Monographs, 7, University of Chicago Press, Chicago, USA.
18. Lord, F. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Lawrence Erlbaum Associates, New Jersey, USA.
19. Ludlow, L. and Haley, S. (1995). Rasch model logits: Interpretation, use, and transformation. *Educational and Psychological Measurement*, 55(6):967–975, doi:10.1177/0013164495055006005.
20. Muñiz, J. (1997). *Introducción a la Teoría de Respuesta a los Items*. Ediciones Pirámide, Madrid, Spain.
21. Oñate, A. and Piñuel, I. (2005). *Informe Cisneros VII: Violencia y Acoso Escolar en Alumnos de Primaria, ESO y Bachiller*. Instituto de Innovación Educativa y Desarrollo Directivo, Madrid, Spain.
22. Rasch, G. (1980). *Probabilistic Models for Some Intelligence Attainment Tests*. University of Chicago Press, Chicago, USA.
23. Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2003). Winbugs User Manual.
24. Van der Linden, W. and Hambleton, R. (1997). Item response theory: Brief history, common models, and extensions. In der Linden, W. V. and Hambleton, R., editors, *Hanbook of Modern Item Response Theory*, pages 1–28. Springer, New York, USA.
25. Wilson, M. (1989). Saltus: A psychometric model of discontinuity in cognitive development. *Psychological Bulletin*, 105(2):276, doi:10.1037/0033–2909.105.2.276.

Appendices

Appendix 1. Bullying Survey

The bullying victimization test is composed of 22 items, each of which has three categories related with frequency of the student's victimization described by each of the items. These categories are: *Never-Sometimes* and *Frequently*. The lowest category *Never* is related to absence of bullying.

1. They force me to do things that I don't want to do.
2. They force me to do things that endanger my physical integrity.
3. They persuade others not to talk to me.
4. They don't talk to me.
5. They despise my work, no matter what I do.
6. They criticize me and blame me for anything that I do or decisions that I make.
7. They humiliate me and make fun of me in front of others.
8. They don't count on me for any class activity.
9. They maliciously distort all that I say or do.
10. They blame me for everything bad that happens.
11. They make fun of me and play tricks on me.
12. They threaten me verbally or through gestures.
13. They ignore me and exclude me.
14. They continually interrupt me, not allowing me to express myself.
15. They prevent me from communicating with others.
16. They make cruel jokes about my physical appearance.
17. They throw objects at me.
18. They hide my possessions.
19. They damage my possessions.
20. They take away my snacks.
21. They do things to try to get a violent reaction from me.
22. They assign me disparaging nicknames.

Appendix 2. Winbugs Code for the Uniform Rash Proposed Model

In the code provided below, $alpha[k] = b_k$.

```
model
 {
 # Calculate individual (binary) responses to each test from multinomial data
  for (j in 1 : culm[1]) {
   for (k in 1 : T) {
    r[j, k] <- response[1, k]
   }
  }
  for (i in 2 : R) {
   for (j in culm[i - 1] + 1 : culm[i]) {
    for (k in 1 : T) {
     r[j, k] <- response[i, k]
    }
   }
  }
 # Rasch uniform model
  for (j in 1 : N) {
   for (k in 1 : T) {
    logit(p[j, k]) <- log(theta[j]/5) -log( alpha[k]/5)
    r[j, k] ~ dbern(p[j, k])
   }
   theta[j] ~ dunif(0, 5)
  }
 # Priors
  for (k in 1 : T) {
   alpha[k] ~ dunif(0, 5)
   a[k] <-log( alpha[k]) - log(mean(alpha[]))
  }
  }
 Data
 list(N = 30, R = 30, T = 22,
  culm = c(1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,
27,28,29,30),
 response = structure(.Data = c(
0,0,0,0,0,0,0,0,0,0,1,0,0,0,0,1,0,0,0,0,0,0,
0,0,1,1,0,1,0,1,0,0,0,0,0,0,0,0,1,0,0,0,0,0,
1,1,1,0,1,1,1,1,1,1,1,1,0,1,1,1,0,1,1,1,1,1,
0,0,0,0,1,0,0,0,0,1,1,0,0,0,0,1,0,1,0,0,0,1,
0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,1,0,1,1,0,0,0,1,1,0,0,0,1,0,0,1,1,1,1,
1,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,0,1,0,0,0,1,0,0,0,0,0,0,0,0,0,0,0,0,
0,0,0,0,1,1,0,0,1,0,1,0,0,0,0,0,0,0,0,0,1,0,
0,0,0,1,0,1,0,0,0,0,0,0,0,0,0,1,0,0,1,1,0,1,
1,1,1,1,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
0,0,0,0,0,1,1,0,0,0,1,0,0,0,0,0,1,1,0,0,1,0,
1,0,0,1,1,1,1,1,1,1,1,1,1,1,0,1,1,1,1,0,1,1,
0,0,0,1,1,1,1,1,1,1,1,1,1,1,0,0,0,1,1,0,0,0,
0,0,0,0,0,1,0,1,0,0,0,0,1,1,1,0,0,0,0,0,0,0,
1,1,0,1,0,1,1,0,0,1,1,1,1,1,0,1,1,0,1,0,1,1,
1,0,0,1,1,1,0,0,0,0,1,0,1,0,0,1,0,1,0,0,0,1,
```

```
1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
0,0,0,1,0,1,0,1,0,0,1,0,0,0,0,0,1,0,0,0,1,0,
0,0,0,0,0,0,1,1,0,0,0,1,0,1,0,0,0,0,0,0,0,1,
0,0,0,1,1,1,0,1,1,1,0,1,1,1,0,1,1,1,0,1,1,1,
1,1,0,0,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,1,
1,0,0,0,0,1,0,0,1,0,1,0,0,1,0,1,0,0,0,0,0,1,
1,0,0,1,1,0,0,1,1,1,1,0,1,1,1,0,0,0,1,0,0,1,
0,0,0,1,0,1,0,1,1,0,1,1,1,0,0,0,1,1,1,0,0,1,
0,0,0,0,1,1,1,0,0,1,0,1,0,1,1,0,1,1,0,1,1,1,
0,0,0,0,0,1,0,0,1,0,1,0,0,0,0,0,0,0,0,0,0,1,
0,0,0,1,1,0,0,1,0,0,1,0,1,0,0,0,1,1,0,0,0,0,
1,0,0,0,1,1,0,0,1,0,1,0,0,1,1,1,0,1,1,0,0,1,
0,0,1,1,0,1,0,1,1,0,1,0,1,1,0,0,0,1,0,0,1,1), .Dim = c(30, 22)))
Inits
list(alpha = c(1,1,1, 1, 1, 1,1,1, 1, 1, 1,1,1, 1, 1, 1,1,1, 1, 1, 1,1))
```