Rapporto n. 1 2010
dmsia unibg.it

L. Bertoli-Barsotti,
C. Muschitiello and A. Punzo

Item Analysis of a Selected Bank from the Voluntary HIV-1 Counseling and Testing Efficacy Study Group

Serie Ricerca

Dipartimento
di Matematica, Statistica,
Informatica e Applicazioni
"Lorenzo Mascheroni"

UNIVERSITÀ DEGLI STUDI DI BERGAMO

Item Analysis of a Selected Bank from the Voluntary HIV-1 Counseling and Testing Efficacy Study Group

LUCIO BERTOLI-BARSOTTI[1]
Dipartimento di Matematica Statistica Informatica e Applicazioni
Università degli studi di Bergamo

CRISTINA MUSCHITIELLO
Dipartimento di Scienze Statistiche "Carlo Cecchi"
Università degli Studi di Bari

ANTONIO PUNZO
Dipartimento di Impresa, Culture e Società
Università degli Studi di Catania

[1]Corresponding Author:

LUCIO BERTOLI-BARSOTTI, Dipartimento di Matematica Statistica Informatica e Applicazioni, *Università degli studi di Bergamo, Via dei Caniana, 2 – 24127 Bergamo – Italy.*
Email: ***lucio.bertoli-barsotti@unibg.it***

Abstract

A 15-item questionnaire, that is part of the HIV-1 VCT dataset, has been considered in this paper, following a Rasch analysis – by means of the Partial Credit Model. We begin by reviewing the different methods of maximum likelihood estimation of the parameters of the model. We then focus on five among the most popular Rasch measurement softwares that implement these estimation methods. The questionnaire results to have the qualities of a good measure of attitude, in the population, towards the use of the condom- but a gender-based Differential Item Functioning has been detected for several items in the test.

K-words: Rasch model; Partial Credit Model; ML estimation methods; Rasch softwares; HIV prevention

Acknowledgments:
- *This work has been supported by* a research *grant* from MURST.
- The Voluntary HIV1 Counseling and Testing study was sponsored by UNAIDS/WHO, AIDSCAP/Family Health International and the Center for AIDS Prevention Studies, University of California, San Francisco.

Introduction

The Voluntary Counseling and Testing Efficacy Study was a randomized clinical trial conducted to test the efficacy of voluntary HIV-1 counseling and testing (HIV-1 VCT) in reducing sexual risk behavior. Methods and outcomes of the original clinical trial may be found in The Voluntary HIV-1 Counseling and Testing Study Group (2000a; 2000b). The whole dataset and other useful survey details are available from the website www.caps.ucsf.edu/tools/data/VCT. The study was conducted at three sites: Nairobi, Kenya; Dar es Salaam, Tanzania; Port of Spain, Trinidad.

To perform this study, 4292 persons were enrolled: 3120 individuals (1534 men, 1586 women) and 586 couples (586 men, 586 women). The dataset reported 1570 variables for each participant; among these, there was a bank of 15 questions – concerning opinions about condom use. Part of this questionnaire has been considered elsewhere for a Rasch Analysis by De Ayala (2003, 2009). Respondents were asked how much they agreed with a given statement on a 4-point Likert response scale, from "strongly disagree" (=1) to "strongly agree" (=4). Due to the negative wording of seven out of the fifteen items, the scores have been reversed for these questions (see Table 2). Missing values were present: a total of 3473 individuals had complete response patterns, 809 individuals had incomplete response patterns and 10 individuals omitted all the questions. Obviously, the only 4282 individuals with at least 1 response were considered in the analysis.

The present paper proposes the use of the Partial Credit Model (PCM; Masters, 1982) to measure the extent to which the subject is "in agreement" with the statement (in its positive wording - i.e. in favour of the use of the condom) by means of calibrations carried out, by comparative pourposes, with five different softwares –using four different maximum likelihood (ML) estimation approaches. The main interest of the analysis is on *item* calibration. Site, sex and individual/couple status were also taken in consideration as possibly useful person factors for the analysis.

In the next section we give a short summary about the ML approaches to the estimation of PCM parameters –with a view towards available softwares that implement these estimation methods. Then we show results of the application of PCM to the HIV-1 VCT 15-item questionnaire. Finally, in the Appendix, we provide a schematic tutorial on the typical steps that a user faces in using the estimation softwares considered in this paper.

Estimation methods, models and softwares: a brief overview

Fixed-score model – The JML and the CML estimation approaches

Let the scalar $x_{vi} = \sum_{h=0}^{m_i} h y_{vih}$ be the score of person v to item i, where $\boldsymbol{y}_{vi} = (y_{vi0}, y_{vi1}, ..., y_{vim_i})$ is a selection vector defined as follows: $y_{vih} = 1$ and 0 otherwise, if person v responds is in category C_h to item i. The $k-$dimensional vector $\boldsymbol{x}_v = (x_{v1}, x_{v2}, ..., x_{vk})$ represents the vector score (or response vector) of person v. By the local independence, in the *fixed-score* (or *fixed-effects*, or *functional*) model the probability of the vector score $\boldsymbol{x}_v$ is defined *for a given value* of person parameter θ_v,

$$\pi(\boldsymbol{x}_v) = \pi(\boldsymbol{x}_v | \theta_v) = \prod_{i=1}^{k} \pi(x_{vi} | \theta_v),$$

where $\pi(x_{vi} | \theta_v) = \dfrac{\exp \sum_{h=0}^{m_i} (\theta_v h + \beta_{ih}) y_{vih}}{\sum_{z=0}^{m_i} \exp(\theta_v z + \beta_{iz})} = \dfrac{\exp(x_{vi}\theta_v) \exp \sum_{h=0}^{m_i} \beta_{ih} y_{vih}}{\sum_{z=0}^{m_i} \exp(\theta_v z + \beta_{iz})}$. Then

$$\pi(\boldsymbol{x}_v) = \pi(\boldsymbol{x}_v | \theta_v) = D_v^{-1} \exp(t_v \theta_v) \exp \sum_{i=1}^{k} \sum_{h=0}^{m_i} \beta_{ih} y_{vih}$$

where $D_v = \prod_{i=1}^{k} \left(\sum_{z=0}^{m_i} \exp(\theta_v z + \beta_{iz}) \right)$, and where $t_v = \sum_{i=1}^{k} x_{vi}$ represents the raw score (total score) of person v. Now, by virtue of the sufficiency of T_v, the conditional distribution of $\boldsymbol{X}_v$ given $T_v = t_v$ does not depend on θ_v. Indeed one obtains, by summation over the set $I_v = \left\{ \boldsymbol{y}_{vi} \left| \sum_{i=1}^{k} \sum_{h=0}^{m_i} h y_{vih} = t_v \right. \right\}$ -i.e. the set of all the selection vectors that produce the total score t_v - the marginal probability distribution of T_v

$$\pi\left(t_v \left|\theta_v\right.\right) = \sum_{I_v} \pi\left(\boldsymbol{x}_v \left|\theta_v\right.\right) = D_v^{-1} \exp\left(t_v \theta_v\right) \gamma\left(t_v\right),$$

where $\gamma\left(t_v\right) = \sum_{I_v} \exp \sum_{i=1}^{k} \sum_{h=0}^{m_i} \beta_{ih} y_{vih}$, and then

$$\pi\left(\boldsymbol{x}_v \left|\theta_v\right.\right) = \pi\left(\boldsymbol{x}_v \left|t_v\right.\right) \pi\left(t_v \left|\theta_v\right.\right),$$

where the conditional probability $\pi\left(\boldsymbol{x}_v \left|t_v\right.\right) = \gamma\left(t_v\right)^{-1} \exp \sum_{i=1}^{k} \sum_{h=0}^{m_i} \beta_{ih} y_{vih}$ only depends on the item parameters, while $\pi\left(t_v \left|\theta_v\right.\right)$ depends on both the items and person parameters.

Then, the total log-likelihood function may be decomposed into two parts, the *conditional* and the *marginal* log-likelihood functions, l_C and l_M , as follows:

$$l\left(\boldsymbol{\theta}, \boldsymbol{\beta}\right) = \sum_{v=1}^{n} \log \pi\left(\boldsymbol{x}_v \left|\theta_v\right.\right) = \sum_{v=1}^{n} \log \pi\left(\boldsymbol{x}_v \left|t_v\right.\right) + \sum_{v=1}^{n} \log \pi\left(t_v \left|\theta_v\right.\right) = l_C + l_M \; .$$

Hence:

$$l_C\left(\boldsymbol{\theta}\right) = -\sum_{v=1}^{n} \log \gamma\left(t_v\right) + \sum_{i=1}^{k} \sum_{h=0}^{m_i} \beta_{ih} y_{\bullet ih} \; ;$$
$$l_M\left(\boldsymbol{\theta}, \boldsymbol{\beta}\right) = -\sum_{v=1}^{n} \log D_v + \sum_{v=1}^{n} \log \gamma\left(t_v\right) + \sum_{v=1}^{n} t_v \theta_v \; ;$$

where $y_{\bullet ih} = \sum_{v=1}^{n} y_{vih}$. Note that this decomposition may be obtained equivalently by adding and subtracting, from total log-likelihood function $l = \sum_{v=1}^{n} t_v \theta_v + \sum_{i=1}^{k} \sum_{h=0}^{m_i} \beta_{ih} y_{\bullet ih} - \sum_{v=1}^{n} \log D_v$, the function $\sum_{v=1}^{n} \log \gamma\left(t_v\right)$. Hence, the log-likelihood functions l, l_C and l_M are obtained on the basis of three different linear combinations

$$l = g_1 + g_2 + g_3$$

$$l_C = g_2 - g_4$$

$$l_M = g_1 + g_3 + g_4$$

of the same 4 functions:

$$g_1(\boldsymbol{\theta}) = \sum_{v=1}^{n} t_v \theta_v$$

$$g_2(\boldsymbol{\beta}) = \sum_{i=1}^{k} \sum_{h=0}^{m_i} \beta_{ih} y_{\bullet ih}$$

$$g_3(\boldsymbol{\theta}, \boldsymbol{\beta}) = -\sum_{v=1}^{n} \log D_v$$

$$g_4(\boldsymbol{\beta}) = \sum_{v=1}^{n} \log \gamma(t_v)$$

The JML estimate is simply defined as $\left(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}\right) = \underset{(\theta,\beta)}{\arg\max}\, l$ (the usual ML estimation method). The JML approach was implemented in the LOGOG computer program (for mainframe) due to Kolakovski & Bock (1973). JML is/was also used in estimation programs such as LOGIST (Wingersky et al., 1999), PARMATE (Burket, 1995), BIGSTEPS (Linacre & Wright, 2006), FACETS (Linacre, 2009), WINSTEPS (Linacre, 2009), QUEST (Adams & Khoo, 1998), and RASCAL (Assessment System Corporation, 2008).

The CML estimation approach may be viewed as an approximation of the JML estimate. Inference concerning the *whole* parameter $(\boldsymbol{\theta}, \boldsymbol{\beta})$ are usually obtained with a *two-step* procedure:

i) The estimate of item parameter $\boldsymbol{\beta}$, say $\hat{\boldsymbol{\beta}}_C$ (so called CML estimate of item parameters), is computed as the maximum of l_C, then

ii) an estimate of $\boldsymbol{\theta}$ is obtained as the maximum of the function $l\left(\boldsymbol{\theta}, \hat{\boldsymbol{\beta}}_C\right)$ - i.e. by maximization of the total log-likelihood function with the CML estimate $\hat{\boldsymbol{\beta}}_C$ as fixed constant.

Obviously, for the same dataset, we should expect different estimates for the JML and the CML procedures. When the interest is *solely* in $\boldsymbol{\beta}$ and the latent abilities are regarded as nuisance parameters the CML procedure (giving $\hat{\boldsymbol{\beta}}_C$) is usually preferred to the JML method, because the CML estimates are proved to be consistent, and asymptotically normal, as $n \to \infty$ (if there are no constraints on the distribution of person parameters; see Pfanzagl, 1994) – while the JML estimator of $\boldsymbol{\theta}$ is inconsistent (although letting both *n* and *k* approach infinity at suitable rates produce consistency; see Haberman, 1977 and Andersen, 1973).

The CML approach was implemented in several computer programs such as: LPCM-Win by Fischer & Ponocny-Seliger (1998); RSP (Glas & Ellis, 1993); OPLM (Verhelst et al., 1995); WINMIRA (von Davier, 1994, 2001) (available at the web store at www.scienceplus.nl); eRm (Mair & Hatzinger, 2007). The *pairwise conditional* ML (PCML) is sligthly different from CML method, because it produces item parameter estimates by considering the items two at time - in all possible pairs. The pairwise estimation is conditional estimation in the sense that the person parameters are eliminated while the item parameters are estimated (person parameters are conditioned out in estimating the item parameters). This method was implemented in RUMM2020 (Andrich et al., 2003; see also Andrich & Luo, 2003, and Zwinderman, 1995).

Random-score model - The MML estimation approach

If we are interested in the item parameter only, the person parameter may be treated as a nuisance parameter by assuming that $\theta_1, \theta_2, \ldots, \theta_n$ are realizations from unknown distributions. Under the *random-score* (or *structural*) approach, the probability of the vector score $\boldsymbol{x}_v$ is defined for a randomly selected individual, provided by its own ability distribution F_v. Then the total log-likelihood function is

$$l = l(\boldsymbol{\beta}, F) = \sum_{v=1}^{n} \log \pi(\boldsymbol{x}_v) = \sum_{v=1}^{n} \log \int \pi(\boldsymbol{x}_v | \theta)\, dF_v(\theta) = l_C + l_P ,$$

where

$$l_P = l_P(\boldsymbol{\beta}, F) = \sum_{v=1}^{n} \log \int \pi(t_v|\theta)\, dF_v(\theta) = \sum_{v=1}^{n} \log \gamma(t_v) + \sum_{v=1}^{n} \log \int D^{-1} \exp(t_v\theta)\, dF_v(\theta)$$

,

where D is the function $\prod_{i=1}^{k}\left(\sum_{z=0}^{m_i} \exp(\theta z + \beta_{iz})\right)$. This formula introduces a fourth kind of log-likelihood function, l_P, called *population* log-likelihood function. The difference between *marginal* and *population* log-likelihood functions is that the former makes no assumptions on the variation of θ_v within the population. Note that, by virtue of its independence of $\boldsymbol{\theta}$, the term l_C coincides with that of the functional model. Then, by substituting $l_C = -\sum_{v=1}^{n} \log \gamma(t_v) + \sum_{i=1}^{k} \sum_{h=0}^{m_i} \beta_{ih} y_{\bullet ih}$, we may write also

$$l = l_C + l_P = \sum_{i=1}^{k} \sum_{h=0}^{m_i} \beta_{ih} y_{\bullet ih} + \sum_{v=1}^{n} \log \int D^{-1} \exp(t_v\theta)\, dF_v(\theta).$$

Several approaches are possible, depending on the constraint on F_v (see Table 1). Possible choices are:

a) $F_v = F$, i.e. all individuals are sampled from the same distribution;

b) F_v not assumed to be equal.

Moreover, the distribution F_v may be supposed to be

1) completely known;
2) known to belong to a parametric family depending to a real parameter $\boldsymbol{\varepsilon}$;
3) completely unknown.

Table 1

Approaches to ML estimation

	1) $F_v = F$	2) F_v not equal
a) F_v known	little interest	little interest
b) F_v belongs to a parametric family	MML	little interest
c) F_v *unknown*	NMML	JML/CML

In its turn, it is intended that all these methods may be applied by maximizing the log-likelihood function over $(\boldsymbol{\beta}, F)$ all at once, say *one-stage* procedure, or, alternatively, by using a *two-stage* procedure, i.e.

i) computing *first* the CML estimate $\hat{\boldsymbol{\beta}}_C$ of item parameters

ii) and *then* completing the estimation process by maximizing $l_P\left(\hat{\boldsymbol{\beta}}_C, F\right)$ over the class of the distributions F_v.

When F_v is completely unknown – for each person *v* (case 2(c)) it is easy to see that the optimum F_v are one-point distributions, placing mass one to a single point, say θ_v. Then random-score model reduces to the fixed-score model. Hence, for the random-score model, the unrestricted maximization of the log-likelihood function $l = l_C + l_P$ over $(\boldsymbol{\beta}, F)$ *reduces to the JML case or CML* case depending on the type of maximization adopted, (i) one-step (JML) or (ii) two-step (CML). These three different ML methods (JML, CML and MML) are strongly related each other.

While the cases 1(a) and 1(b) have little practical interest, and the case 2(b) suffer the drawback of not reducing the number of person parameters, cases 2(a) and 3(a) are the most interesting ones. The former approach is usually referred to as MML (*marginal maximum likelihood*) estimation method, while the latter is known as NMML (*nonparametric marginal maximum likelihood,* or also *semiparametric marginal maximum likelihood*) estimation method. An example of an application of a two-stage MML estimation approach is given by Andersen and Madsen (1977) (see also Mislevy, 1984).

Let n_t be the number of person with total score *t* and let $A = \sum_{i=1}^{k} m_i$ the maximum possible score on the instrument. By the Jensen inequality and the strict concavity of the logarithmic function it follows

$$n_t^{-1} \sum_{\{v|t_v=t\}} \log \int \pi\left(t|\theta\right) dF_v\left(\theta\right) \le \log \sum_{\{v|t_v=t\}} n_t^{-1} \int \pi\left(t|\theta\right) dF_v\left(\theta\right) = \log \int \pi\left(t|\theta\right) dF_t\left(\theta\right)$$

where $F_t\left(\theta\right) = n_t^{-1} \sum_{\{v|t_v=t\}} F_v\left(\theta\right)$. Then -as in the fixed-score model-, all individuals with the same total score *t* (even if with different response patterns) correspond to a unique

individual estimable distribution F_t; in other words, they are indistinguishable from an inferential point of view.
We may also write:

$$l_C = -\sum_{t=0}^{A} n_t \log \gamma(t) + \sum_{i=1}^{k} \sum_{h=0}^{m_i} \beta_{ih} y_{\bullet ih} ;$$
$$l_P = \sum_{t=0}^{A} n_t \log \gamma(t) + \sum_{t=0}^{A} n_t \log \int D^{-1} \exp(t\theta)\, dF_t(\theta)$$

Note that for $t = 0$ and $t = A$ there is no contribution to l_C; this means that l_C does not depend on extreme patterns.
When F is assumed to belong to a given parametric family of distributions the MML estimate, say $\left(\hat{\boldsymbol{\beta}}_M, \hat{\boldsymbol{\varepsilon}}_M\right)$, is obtained by maximizing the log-likelihood in the form

$$l = l(\boldsymbol{\beta}, \boldsymbol{\varepsilon}) = \sum_{i=1}^{k} \sum_{h=0}^{m_i} \beta_{ih} y_{\bullet ih} + \sum_{t=0}^{A} n_t \log \int D^{-1} \exp(t\theta)\, dF_t(\theta|\boldsymbol{\varepsilon}).$$

An advantage of the MML method is that it furnish finite estimates even for extreme patterns (then such patterns have not to be removed from the dataset). On the other hand, the MML method is not consistent if the true distribution of $\boldsymbol{\theta}$ does not belong to the hypothesized parametric family (Zwinderman, 1991, Ch.4, proved that normal MML estimates can be highly biased if the prior person parameter distribution differed from normal).
The major drawback of the NMML is the lack of identifiability of F in the class of all distribution on the Borel class with positive and continuous Lebesgue density; since t assumes only a finite number of different values (at most A - limited by the number of items and categories in the test) the distribution will be known only through a finite number of conditions on its moments. The information per person is structurally limited by the length of the test; then, even if n goes to infinity, the distribution of $\boldsymbol{\theta}$ cannot be determined in great detail. A solution to this problem is to assume an additional condition on F. F must be as simpler as possible (*canonical*) – that is a step function, with the minimum possible number of steps. Under these conditions it may be proved that the number of support points ("knots") of F is at most $(A+2)/2$ if A is even and

$(A+1)/2$ if A is odd. Note that in the NMML approach the function to maximize becomes

$$l = l(\boldsymbol{\beta}, \boldsymbol{\vartheta}, \boldsymbol{\omega}) = \sum_{i=1}^{k} \sum_{h=0}^{m_i} \beta_{ih} y_{\bullet ih} + \sum_{t=0}^{A} n_t \log \sum_{j=1}^{r} \omega_j D_j^{-1} \exp t\vartheta_j$$

where $D_j = \prod_{i=1}^{k} \left(\sum_{z=0}^{m_i} \exp\left(\vartheta_j z + \beta_{iz} \right) \right)$.

The MML estimation approach was implemented in GradeMap (Wilson et al., 2001); ConQuest (Wu et al., 2007; see also Adams et al., 1997); BILOG-MG (Zimowski et al., 2005); PARSCALE (Muraki, E., & Bock, 1997); MULTILOG (Thissen et al., 2002). While GradeMap is free, ConQuest, BILOG-MG, PARSCALE and MULTILOG are distributed from Assessment Systems Corporation (see www.assess.com).

Elements for a comparative analysis of the HIV-1 VCT dataset

For a complete comparative evaluation of all these methods of estimation, five different software packages were utilized for estimation throughout this study: i) WINSTEPS (JML); ii) eRm (CML); iii) RUMM2020 (PCML); iv) ConQuest (MML); v) GradeMap (MML). It is to be noted that all these programs handle omitted responses routinely, by considering the missing as “missing at random” (MAR). In the present case, the causes of missing data are not known (though one possible reason for not answering an item could be a sort of “uncomfortableness” with the question), but the amount of missingness is not very large (3,5%), then the MAR assumption may be tenable.

In our model, $m_i = 3$ for every item. Hence the item locations, say α_i , correspond to the values $-\beta_{im} / m$, $i = 1,...,15$. The parameter α_i represents the extent to which the *i*-th statement (in its positive wording) is agreed by the respondents, in the following sense: more positive, more difficult to agree. Then, by following De Ayala (2009), we speak of an “Attitude Towards Condoms Scale”. By default, all the softwares considered the constraint for identifiability $\sum_i \alpha_i = 0$, with the exception of eRm. By constraining parameters to satisfy the same constraint, we obtain the estimates reported in Table 3 (and depicted in Figure 1): on the whole, estimates result very similar.

Threshold parameters $\delta_{ih} = -(\beta_{ih} - \beta_{ih-1})$ are systematically disordered (see Table 4); this is due to the fact that respondents tend to make little use of the intermediate categories. Indeed, most participants (74,8%) selected either the lower or the higher response category. All the items have a strong tendency to behave in a dichotomous way.

Table 2. Items measuring attitudes towards condom

	Item wording (abbreviated)	*reverse scoring*	*# Missing*
I1	using condoms good protection from stds		*19*
I2	sex not as good when you use a condom	yes	*182*
I3	embarrassing to buy condoms	yes	*42*
I4	using condoms good pregnancy prevention		*34*
I5	embarrassing put on condom/ or on a man	yes	*63*
I6	frnds think use condoms incldng w/spouse		*135*
I7	condoms often break or slip	yes	*299*
I8	if sex partner wants condom I suspect	yes	*52*
I9	friends use condoms w/new partner		*334*
I10	easy to buy condoms in my area		*137*
I11	friends think that condoms uncomfortable	yes	*246*
I12	friends thnk alwys use condom new person		*212*
I13	easy to get free condoms in my area		*141*
I14	condoms cost too much	yes	*322*
I15	most people your age using condoms now		*197*

Table 3. Item location estimates

	Winsteps	**eRm**	**RUMM**	**ConQuest**	**GradeMap**
I1	-0,73	-0,680	-0,674	-0,675	-0,680
I2	0,29	0,264	0,256	0,263	0,278
I3	-0,09	-0,081	-0,096	-0,079	-0,115
I4	-0,68	-0,627	-0,599	-0,623	-0,632
I5	-0,16	-0,144	-0,165	-0,142	-0,180
I6	0,26	0,242	0,253	0,243	0,253
I7	0,45	0,414	0,402	0,409	0,452
I8	0,42	0,388	0,378	0,390	0,416
I9	-0,06	-0,055	-0,040	-0,055	-0,069
I10	-0,11	-0,097	-0,093	-0,099	-0,134
I11	0,66	0,606	0,601	0,604	0,682
I12	-0,3	-0,279	-0,269	-0,279	-0,302
I13	0,44	0,406	0,389	0,408	0,439
I14	-0,28	-0,256	-0,251	-0,265	-0,290
I15	-0,11	-0,100	-0,093	-0,100	-0,117

Figure 1. Dotplot of Item location estimates

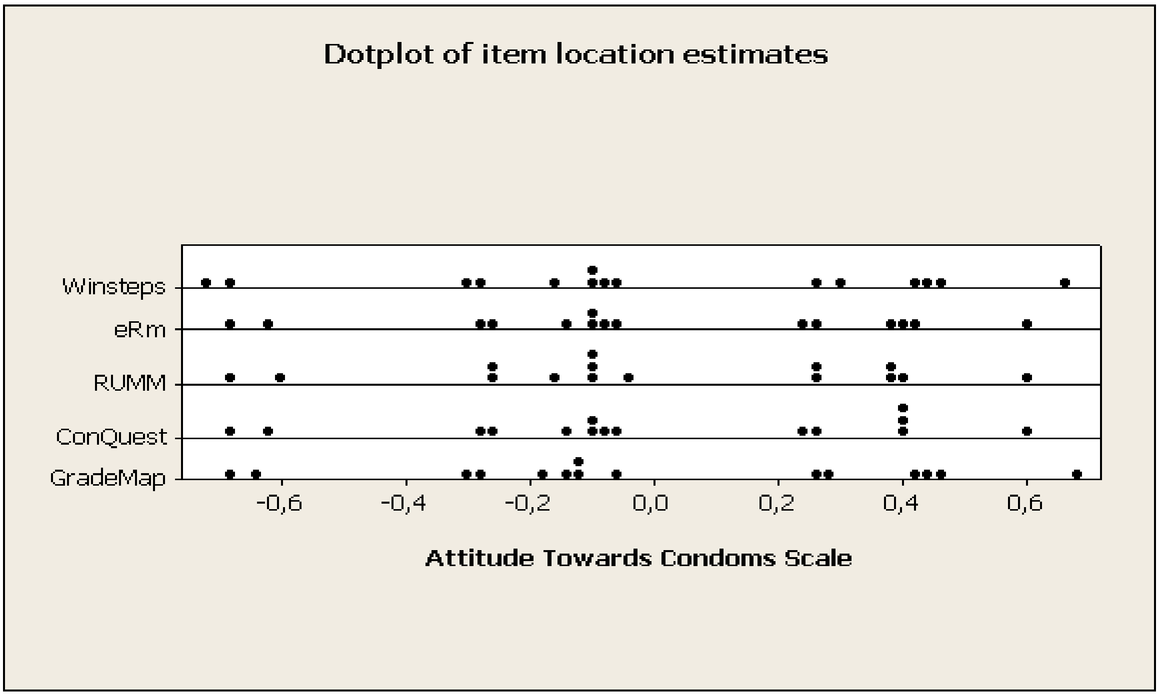


Table 4. Uncentralized threshold estimates (by RUMM2020)

	Thr.1	*Thr.2*	*Thr.3*
I1	0,561	-1,063	-1,519
I2	0,588	0,521	-0,342
I3	0,797	0,149	-1,235
I4	0,454	-1,080	-1,171
I5	0,774	0,032	-1,300
I6	0,604	0,263	-0,107
I7	0,530	0,497	0,180
I8	1,295	0,417	-0,577
I9	0,308	-0,349	-0,078
I10	0,928	-0,021	-1,186
I11	0,731	0,818	0,254
I12	0,304	-0,602	-0,510
I13	1,322	0,610	-0,765
I14	0,443	-0,400	-0,795
I15	0,233	-0,204	-0,309

Measuring the goodness of fit of the model

Item fit, at the item level, can be assessed using the weighted mean-square statistic (*infit*), a residual-based fit statistic. Weighted *infit* statistics for item parameters are determined, by default, by all the softwares considered. The Fit Graph furnished by GradeMap is depicted in Figure 2; the vertical bands at 0,75 and 1,33 represent theoretical boundaries defining (heuristically) acceptable values for infit mean-squares. The items seem to behave in a fashion consistent with the model. Besides, Pearson-type χ^2 statistics furnished by all the softwares suffer of the problem to be overly sensitive to large sample size -which is particularly important in this case study-, then they are not of great value.

Figure 2. Infit graph (by GradeMap)

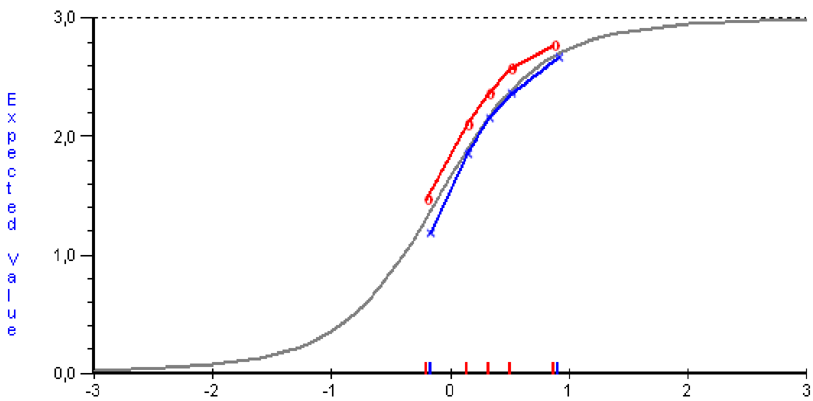
DIF Analysis

Sometimes item bias, or *differential item functioning* (DIF), is present. DIF occurs whenever respondents form two different population groups having the same amount of the underline trait measured by the test perform unequally on an item. To conduct an analysis of DIF one has to partition the data in some way, e.g. by employing suitable user-defined demographic variables. Site, gender and individual/couple status were taken in consideration as possibly person factors for this analysis. Interestingly (as to be expected), several items are functioning differentially across gender. The probability of agreement to a statement for males is sometimes significatively greater than that for females and vice versa. A (uniform) DIF has been detected in correspondence to the items as reported in Table 5. The last column report the gender with greater level of agreement.

Table 5. DIF Analysis (by RUMM2020)

	Item wording (abbreviated)	*DIF*	*most agreed by*
I1	using condoms good protection from stds	yes	F
I2	sex not as good when you use a condom		
I3	embarrassing to buy condoms		
I4	using condoms good pregnancy prevention	yes	F
I5	embarrassing put on condom/ or on a man	yes	M
I6	frnds think use condoms incldng w/spouse	yes	F
I7	condoms often break or slip		
I8	if sex partner wants condom I suspect		
I9	friends use condoms w/new partner	yes	M
I10	easy to buy condoms in my area	yes	M
I11	friends think that condoms uncomfortable	yes	F
I12	friends thnk alwys use condom new person		
I13	easy to get free condoms in my area		
I14	condoms cost too much	yes	F
I15	most people your age using condoms now	yes	M

Figure 3. Uniform DIF for item I10 (by RUMM2020)



Male= red; Female= blue

Appendix

Winsteps

To run Winsteps, the user must first create a TXT control file that specifies the model (here the PCM), scores (items # 2, 3, 5, 7, 8, 11, 14, have a reverse score), data structure and output format using a special program code. This control file is saved as a text file and then run from the Winsteps program. The data to be analyzed begin at the end of the control file text. Note that here the dot "." is used to denote the missing value. Table1A contains the command file for specifying the WINSTEPS calibration of the dataset.

Table 1A. Winsteps code in the control file .txt

```
&INST
TITLE="Attitude towards condoms"
ITEM1=1
NI=15
IREFER=ABBABABBAABAABA
CODES=1234
IVALUEB=4321
GROUPS=0
CHART=YES
&END
I1
I2
I3
I4
I5
I6
I7
I8
I9
I10
I11
I12
I13
I14
I15
ENDNAMES
4..............
4.4............
4..4...1.......
432............
```

GradeMap

Detailed instructions about the use of GradeMap are directly available from the Online Help web page http://bearcenter.berkeley.edu/wiki/index.php/ConstructMap User_Guide. GradeMap supports datasets in Excel format. In its simplest form, the dataset may be organized as shown in Figure 1A.

Figure 1A

Excerpt from the data file `.xls`

ID	**using condoms good protection from stds**	**sex not as good when you use a condom**	**embarrassing to buy condoms**	**...**
1	4	.	.	**...**
2	4	.	4	**...**
3	4	.	.	**...**
4	4	3	2	**...**
5	.	.	1	**...**
6	4	4	.	**...**
...	**...**	**...**	**...**	**...**

By selecting the command "File > Import Wizard", the file containing the response data will be imported into GradeMap. Clicking the "Answer Key" button, it is also possible to recode the original scores. The PCM is considered as default model, by the program, then by selecting "Estimation Tasks > Compute Item Parameters", the item parameters estimates, as well as population mean and population variance, are obtained. A pull-down bottom consent to select integration method and convergence criteria.

ConQuest

This program produces Marginal Maximum Likelihood estimates for the parameters of a wide variety of IRT models (including multidimensional models). In the following we explains the typical steps in obtaining item parameter estimates for a Partial Credit Model (for more details, see Wu *et al*. 2007).

- Excel data, without row(s) of item labels, have to be entered into a common text file manually saved as `.dat`, using one line per each subject. An extract of the data file is shown in **Figure 2A**.

  Here, a subject identification code has been entered in columns 1 through 4, and the subject's responses have been recorded in columns 5 through

19. The response to each item has been allocated in one column, and the above-mentioned codes 1,2,3 and 4 have been used to indicate which alternative the subject choses for each item. A point "." has been entered as missing response.

Figure 2A. Excerpt from the data file `.dat`

```
         1
1234567890123456789 (column numbers)
         .
         .
73263......4114.1.3
20864.4441..1..1..1
23674.121..2....112
24544.3421.4..4.3..
70413..3..4.14..441
70824.141..1.4...14
731044444....1..41.
74324..4.1.12..24.3
         .
         .
```

Here, a subject identification code has been entered in columns 1 through 4, and the subject's responses have been recorded in columns 5 through 19. The response to each item has been allocated in one column, and the above-mentioned codes 1,2,3 and 4 have been used to indicate which alternative the subject choses for each item. A point "." has been entered as missing response.

- The item labels can be reported into a text file like that shown in **Figure 3A**; this file have to be saved as `.lab`.
  The first line contains the special symbol `===>` followed by the name of the variable to which the labels are to apply (in this case, `item`). The subsequent lines contain two pieces of information separated by one or more spaces. The first value on each line is the level of the to which a label is to be attached, and the second value is the label.

- A third text file, the most important one, have to be created and saved as `.cqc`. This file contains the commands needed to analyze the data; it is displayed in **Figure 4A**.

Figure 3A: Contents of the label file `.lab`

```
===> item
1  "using condoms good protection from stds"
2  "sex not as good when you use a condom"
3  "embarrassing to buy condoms"
4  "using condoms good pregnancy prevention"
5  "embarrassing put on condom/ or on a man"
6  "frnds think use condoms incldng w/spouse"
7  "condoms often break or slip"
8  "if sex partner wants condom i suspect"
9  "friends use condoms w/new partner"
10 "easy to buy condoms in my area"
11 "friends think that condoms uncomfortable"
12 "friends thnk alwys use condom new person"
13 "easy to get free condoms in my area"
14 "condoms cost too much"
15 "most people your age using condoms now"
```

Figure 4A. Command file for a Partial Credit Model

```
Title  Partial  Credit  Model:  Attitudes  towards  condoms
scale;
datafile 15items.dat;
format name 1-4 responses 5-19;
labels << 15items.lab;
codes 0,1,2,3;
recode (1,2,3,4) (0,1,2,3) !item(1);
recode (1,2,3,4) (3,2,1,0) !item(2);
recode (1,2,3,4) (3,2,1,0) !item(3);
recode (1,2,3,4) (0,1,2,3) !item(4);
recode (1,2,3,4) (3,2,1,0) !item(5);
recode (1,2,3,4) (0,1,2,3) !item(6);
recode (1,2,3,4) (3,2,1,0) !item(7);
recode (1,2,3,4) (3,2,1,0) !item(8);
recode (1,2,3,4) (0,1,2,3) !item(9);
recode (1,2,3,4) (0,1,2,3) !item(10);
recode (1,2,3,4) (3,2,1,0) !item(11);
recode (1,2,3,4) (0,1,2,3) !item(12);
recode (1,2,3,4) (0,1,2,3) !item(13);
recode (1,2,3,4) (3,2,1,0) !item(14);
recode (1,2,3,4) (0,1,2,3) !item(15);
model item + item*step;
estimate;
show !estimate=latent >> 15items.shw;
itanal >> 15items.itn;
```

- In line 1 a title for the analysis is given (if a title is not provided, the default line, "`ConQuest: Generalised Item Response Modelling Software`", will appear).
- In line 2, name and location of the data file is provided (any name that is valid for the operating system can be used here).
- In line 3 we have the `format` statement describing the layout of the data in the file `.dat`. It indicates that a field called `name` is

located in columns 1 through 4 and that the `responses` to the items are in columns 5 through 19 (the response block) of the file `.dat`.

- Line 4 is used to read the set of labels for the items from the file `.lab`.
- In line 5 the `codes` statement is used to restrict the list of codes that ConQuest will consider valid; any other codes for the items will be treated as missing-response data. It is important to note that the codes statement refers to the codes *after* the application of any recodes.
- From line 6 to line 20 we have fifteen `recode` statements, each of them consisting of three components, used in this analysis to recode the category indicators and to convert the seven "negative" items into a positive form. The first component is a list of codes contained within parentheses that will be found in the data file `.dat`, and these are called the *from* codes. The second component is also a list of codes contained within parentheses, these codes are called the *to* codes. Note that the length of the *to* codes list must match the length of the *from* codes list. When ConQuest finds a response that matches a *from* code, it will change (or recode) it to the corresponding *to* code. The third component gives the levels of the variables for which the recode is to be applied. Line 8, for example, says that, for item 3, 1 is to be recoded to 0, 2 is to be recoded to 1, 3 is to be recoded to 2 and 4 is to be recoded to 0. When ConQuest models these data, the number of response categories that will be assumed for each item will be determined from the number of distinct codes in the *from* codes list. Thus, all items in this analysis have four distinct codes (0, 1, 2 and 3), so four categories will be modelled.
- In line 21 we have the "core" `model` statement containing two terms (`item` and `item*step`) that reesults in the estimation of two sets of parameters. The term `item` results in the estimation of a set of item difficulty parameters, and the term `item*step` results in a set of item step-parameters that are allowed to vary

across the items. This is the partial credit model. If, for example, a rating scale model is chosen, the model statement will change in `item+step`.

  - In line 22 the `estimate` statement is used to initiate the estimation of the item response model.
  - In line 23 the `show` statement produces a display of the item response model parameter estimates and saves them into a file with extension `.shw`. The option `estimates=latent` requests that the displays include an illustration of the latent ability distribution.
  - In line 24 the `itanal` statement produces a display of the results of a traditional item analysis. As with the `show` statement, the results are redirected to a file `.itn`.

- After these preliminary phases, in which we have created - and saved in the same directory - the files `.cqc`, `.dat` and `.lab`, we can easily run the sample analysis launching the ConQuest program, opening the file `.cqc`, and finally choosing Run➔Run All. The results of the analysis will be displayed on the console, and saved (files `.shw` and `.itn`) in the same directory of the file `.cqc`.

eRm

eRm is a package of R. R is an open source statistical environment available through the CRAN family of Internet sites via http://CRAN.R-project.org. Once R is installed on the operative system, several packages can be downloaded too; among them, eRm (acronym of Extended Rasch Modeling; Mair & Hatzinger, 2007) is the available package to perform estimates for the family of Rasch models. In detail, this R-package produces Conditional Maximum Likelihood estimates for the parameters of a wide variety of (dichotomous/polytomous) unidimensional Rasch models. In order to install the package, open the R console and write `install.packages(eRm)`. Once the package is installed, it have to be loaded through the command `library(eRm)`.

Before going on, we suggest to save the data (deleting all columns that are not referred to items), contained into the excel file, in the `.csv` format, well-supported by the R environment. Moreover, the working directory have to be directed in the same directory containing the `.csv` file that we have called "`15items.csv`". Once these preliminary phases are made, the commands displayed below can be copied in the R console to produce the CML-estimates.

```
1.  X <- as.matrix(read.csv2("15items.csv"))
2.  n <- nrow(X)-1      # number of subjects
3.  k <- ncol(X)        # number of items
4.  X <- X[2:(n+1),1:k] # delete item labels
5.  for(v in 1:n){
6.       for(i   in   1:k)   ifelse(X[v,i]=="NA",,X[v,i]<-
as.numeric(X[v,i])-1)
7.  }
8.  X <- matrix(as.numeric(X),n,k)
9.  X[,2]  <- 3-X[,2]
10. X[,3]  <- 3-X[,3]
11. X[,5]  <- 3-X[,5]
12. X[,7]  <- 3-X[,7]
13. X[,8]  <- 3-X[,8]
14. X[,11] <- 3-X[,11]
15. X[,14] <- 3-X[,14]
16. library(eRm)
17. estimates <- PCM(X)
18. thresholds(estimates)
19. personestimates <- person.parameter(estimates)
20. summary(estimates)
21. itemfit(personestimates)
```

In detail we have:

- In line 1 the command `read.csv2` reads the file `15items.csv` and put it into a matrix (via the command `as.matrix`) that we have denoted as `X`.
- In line 2 the number of effective rows of `X`, that is the number of subjects *n*, is computed by the command `nrow` (abbreviating of "number of rows") remembering the presence of the first item labels row.
- In line 3 the number of columns of `X`, that is the number of items *k*, is computed by the command `ncol` (abbreviating of "number of columns").
- In line 4 the row of item labels is deleted from `X`.
- From line 5 to line 7, a brief sequence of code is defined in order to shift the original categories 1,2,3, and 4, in 0,1,2 and 3, and in order to substitute the point symbol "." into the symbol `NA`, used by R as indicator of a missing value.
- From line 8 to line 15, in order to uniform the data, the negative-defined items are recoded as positive items.
- In line 16 the package eRm is loaded.

- In line 17 the item parameter estimates are produced, for the partial credit model, by the command `PCM`. These estimates are saved in `estimates`. Note that if, for example, one prefers the rating scale model, the alternative command `RSM` can be used.
- In line 18 also the threshold parameters for the PCM are computed by the command `thresholds`.
- In line 19 person parameters estimates are obtained and saved in `personestimates`, by applying the command `person.parameter`.
- In line 20 a summary, with further information, of the obtained estimates is visualized on the console by the command `summary` applied to `estimates`.
- Finally, in line 21, some item fit statistics are computed by the command `itemfit` applied to the person estimates `personestimates`.

RUMM2020

A *New Project* is created when attempting to analyse test data for the first time. This process involves three clearly defined stages within RUMM2020:

1. specifying the overall person-item test design
2. specifying the data format within the data file
3. specifying the test item structures

1. The first stage consists in defining test's Person Design and Item Design. The Person Design for a test relates to the structure of the sample according to some specific group membership. The group is known as a *Person Factor* and the category within the group to which a person can be assigned is designated a level. The Item Design for a test relates to the way a set of test items are replicated across two or more components. These components specify a *Facet* or *Factorial Design*. The most common item analysis structure is a single Item Factor design. The items comprising the basis are the levels within this single factor.
2. Specifying the *Data Format* involves three step: specifying the *Unique ID***,** the *Person Factors* and the *Item Responses*.
   - Step 1 : *Person ID*. Provides details on ONE individual identification field.
   - Step 2 : *Person Factors*. Provides details for up to NINE Person Factors.

- Step 3 : *Item Data Format*. Provides for item details. RUMM2020 will allow items to be selected as separate subsets of the total present if that is desirable. This procedure is an important requirement if the items are distributed across the record in separate or distinct blocks and, also, if they are grouped according to type, such as some *polytomous* and some *multiple choice* items; Each block of items must have the same settings in terms of data component, type, maximum number of characters required to specify a response for any item within the block, missing data character and , finally, nature of responses (*numeric* or, if *alpha*, whether *upper case* or *lower case)*

3. The third and final stage in creating a NEW Project involves specifying the Test Structure. This stage concentrates on the structure of the items as components of the test. The procedure is controlled from an *Item Specification Form* which reflects the details entered for the separate item blocks in step 3 of the data tructure specification. In this stage the item structure has to be determine for each block according to the items' type:

- For polytomous items name, number of response categories and the values and scores of each category have to be entered.
- For multiple choice items the key structure of the possible answers has to be specified.

Once the three stages of the New Project creation have been completed, is it possible to specify the Analysis characteristics and to run it.

References

Adams, R. J., & Khoo, S.T. (1998). *Quest: The interactive test analysis system* [Computer Software]. Camberwell, Australia: ACER Press.

Adams, R. J., Wilson, M. & Wang, W. C. (1997). The Multidimensional Random Coefficient Multinomial Logit Model."*Applied Psychological Measurement* 21(1), 1 - 23.

Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S. (2007). *ACER ConQuest 2.0:General item response modelling software* [computer program manual]. Camberwell, VIC: ACER Press.

Agresti, A. (2002). *Categorical Data Analysis*. New York: Wiley.

Andersen, E. B. (1970). Sufficiency and exponential families for discrete sample spaces. *Journal of the American Statistical Association, 65*, 1248-1255.

Andersen, E. B. (1973). Conditional inference for multiple choice questionnaires. *British Journal of Mathematical and Statistical Psychology, 26,* 31-44.

Andersen, E. B. (1980). Discrete statistical models with social sciences applications. Amsterdam: North-Holland.

Andersen, E. B. (1983). A general latent structure for contingency table data. In H. Wainer & S. Messik (Eds.), *Principals of modern psychological measurement* (pp. 117-138). Hillsdale, NJ: Lawrence Erlbaum.

Andersen, E. B., & Madsen, M. (1977). Estimating the parameters of a latent population distribution. *Psychometrika*, 42, 357-374.

Andrich, D., & Luo, G. (2003). Conditional Pairwise Estimation in the Rasch Model for Ordered Response Categories using Principal Component. *Journal of Applied Measurement*, *4*, 205-221.

Andrich, D., Lyne, A., Sheridan, B., & Luo, G. (2003). *RUMM 2020* [Computer Software]. Perth, Australia: RUMM Laboratory.

Assessment System Corporation (2008). *User's Manual for the RASCAL - Rasch Analysis Program.* St Paul: Author.

Baker, F. B., & Kim, S. H. (2004). *Item response theory. Parameter estimation techniques*. New York: Marcel Dekker.

Barndorff-Nielsen, O. (1978). *Information and exponential families in statistical theory*. New York: Wiley.

Bertoli-Barsotti, L. (2005). On the existence and uniqueness of JML estimates for the partial credit model. *Psychometrika*, *70*, 517-531.

Burket, G.R. (1995). *PARMATE* [Computer Software]. Monterey: CTB/McGraw-Hill.

Custer, M., Omar, M. H., & Custer, M. (2006). Vertical scaling with the Rasch model utilizing default and tight convergence settings with WINSTEPS and BILOG-MG, *Applied Measurement in Education*, *19*, 143-149.

De Ayala R.J. (2003), The effect of missing data on estimating a respondent's location using Ratings Data, *Journal of Applied Measurement*, 4(1), 1-9.

De Ayala, 2009, *The theory and the practice of item response theory*, NY: The Guilford Press.

de Boeck, P., & Wilson, M. R. (2004). *Explanatory item response models: A generalized linear and nonlinear approach*. New York: Springer-Verlag.

de Gruijter, D. N. M. (1988). Standard errors of item parameters estimates in incomplete designs. *Applied Psychological Measurement*, *12*, 109-116.

DeMars, C. (2002). Incomplete data and item parameter estimates under JMLE and MML estimation. *Applied Measurement Education*, *15*, 15-31.

Eriksson, N., Fienberg, S. E., Rinaldo, A., & Sullivant, S. (2006). Polyhedral conditions for the nonexistence of the MLE for hierarchical log-linear models. *Journal of Symbolic Computation*, *41*, 222-233.

Fienberg, S. E., & Rinaldo, A. (2007). Three centuries of categorical data analysis: Log-linear models and maximum likelihood estimation. *Journal of Statistical Planning and Inference*, *137,* 3430-3445.

Fischer, G. H. (1981). On the existence and uniqueness of maximum-likelihood estimates in the Rasch model. *Psychometrika*, *46*, 59-77.

Fischer, G. H. (1994). The linear logistic test model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 131-155). New York: Springer-Verlag.

Fischer, G. H. (1995). Some neglected problems in IRT. *Psychometrika*, *60,* 459-487.

Fischer, G.H., & Ponocny-Seliger, E. (1998). *Structural Rasch Modeling: Handbook of the Usage of LPCM-WIN 1.0*. ProGAMMA, Groningen.

Furlow, C. F., Fouladi, R. T., Gagné, P., & Whittaker, T.A. (2007). A Monte Carlo study of the impact of missing data and differential item functioning on theta estimates from two polytomous Rasch family models. *Journal of Applied Measurement*, *8,* 388-403.

Glas, C. A. W., & Ellis, J. L. (1994). *RSP: Rasch scaling program* [Computer software]. ProGAMMA, Groningen.

Haberman, S.J. (1977). Maximum likelihood estimates in exponential response models. *Annals of Statistics*, *5*, 815-841.

Kolakovski, D. & Bock, R.D. (1973). *LOGOG. Maximum likelihood item analysis and test scoring: Logistic model for multiple item responses.* Ann Arbor: National Educational Resources.

Lee, O. K. (1991). Convergence: Statistics or substance?, *Rasch Measurement Transactions*, *5*, 172.

Lee, O. K. (1992). Calibration matrices for test equating?, *Rasch Measurement Transactions*, *6*, 202-203.

Lee, O. K. (2003). Rasch simultaneous vertical equating for measuring reading growth. *Journal of Applied Measurement*, *4*, 10-23.

Lehmann, E. L., & Casella, G. (1998). *Theory of point estimation*. 2nd edition. New York: Springer-Verlag.

Linacre, J. M. (2004). Rasch model estimation: further topics. In E. V. Smith Jr. & R. M. Smith (Eds.), *Introduction to Rasch Measurement. Theory, Models and Applications* (pp. 48-72). Maple Grove: JAM Press.

Linacre, J. M. (2009). *WINSTEPS®. Rasch measurement computer program.* Beaverton, Oregon: Winsteps.com.

Linacre, J. M. (2009) *Facets. Rasch measurement computer program*, version 3.65.0. Chicago: Winsteps.com.

Linacre, J. M., & Wright, B.D. (2006). *A User's Guide to BIGSTEPS: A Rasch-Model Computer Program*. Chicago: MESA Press.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data*. New York: Wiley.

Little, R. J. A., & Schenker, N. (1995). Missing Data. In G. Arminger, C. C. Clogg & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 39-75). New York: Plenum.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale: Erlbaum.

Luo, G., & Andrich, D. (2005). Estimating parameters in the Rasch model in the presence of null categories. *Journal of Applied Measurement*, *6*, 128-146.

Mair P., & Hatzinger, R. (2007). Extended Rasch Modeling: The eRm Package for the Application of IRT Models in R. *Journal of Statistical Software*, 20(9), 1-20.

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

Mathworks Inc. (2007). *MATLAB 7.0* [Computer Software]. Natik, MA, USA.

Molenaar, I. W. (1995). Estimation of item parameters. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch models: Foundations, recent developments, and applications* (pp. 39-51). New York: Springer-Verlag.

Muraki, E., & Bock, R. D. (1997). *PARSCALE: IRT Item Analysis and Test Scoring for Rating-scale Data*. Chicago, IL: Scientific Software International.

Peugh, J. L., & Enders, C. K. (2004). Missing data in educational research: A review of reporting practices and suggestions for improvement. *Review of Educational Research*, 74, 525-556.

Rockafellar, R. T. (1970). *Convex analysis*. Princeton: Princeton University Press.

Rubin, D. B., & Thayer, D. (1978). Relating tests given to different samples. *Psychometrika*, *43*, 3-10.

Schafer, J. L., & Graham, J. W. (2002). Missing data. Our view of the state of the art. *Psychological Methods*, *7*, 147-177.

Thissen, D., Chen, W.-H., & Bock, R. D. (2002). *MULTILOG* [Computer software]. Lincolnwood, IL: Scientific Software International.

The Voluntary HIV-1 Counseling and Testing Efficacy Study Group (2000a). Efficacy of voluntary HIV-1 counselling and testing in individual and couples in Kenya, Tanzania, and Trinidad: a randomised trial. *Lancet, 256*, 103-112.

The Voluntary HIV-1 Counseling and Testing Study Group (2000b). The Voluntary HIV-1 Counseling and Testing Efficacy Study: Design and Methods . *AIDS and Behavior,* 4(1), 5-14.

Verhelst, N. D., Glas, C. A. W., & Verstralen, H. H. F. M. (1995). *One-parameter logistic model: OPLM* [Computer program]. Chicago: Scientific Software.

von Davier, M. (1994). *WINMIRA (A Windows program for analyses with the Rasch model, with the latent class analysis and with the mixed Rasch model)* [Computer software]. Kiel, Germany: Institute for Science Education.

von Davier, M. (2001). *WINMIRA 2001 manual:Windows software for discrete mixture distribution models like the (mixed) Rasch model, the latent class analysis, and the HYBRID model*. Unpublished manuscript.

von Davier, M., & Carstensen, C. H. (2007). *Multivariate and mixture distribution Rasch models. Extensions and applications*. New York: Springer Verlag.

Wilson, M., & Masters, G.N. (1993). The partial credit model and null categories. *Psychometrika*, 58, 87-99.

Wilson, M., Draney, K. & Kennedy. (2001). *Grade Map.* Berkeley: BEAR Center, University of California, Berkeley.

Wingersky, M.S., Patrick, R., & Lord, F.M., 1999, *LOGIST user's guide* (Version 7.1). Princeton: Educational Testing Service.

Wolfram Research, Inc. (2009). *Mathematica 7: Version 7.0.1*. Champaign: Wolfram Research.

Wright, B. D., & Masters, G. N. (1982). *Rating scale analysis.* Chicago: MESA Press.

Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2005). *BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items* [Computer software]. Lincolnwood, IL: Scientific Software International.

Wu M.L., Adams R.J., Wilson M.R. (2007), Haldane S.A. *ConQuest: Generalised item response modeling software,* ACER Press.

Zwinderman, A.H. (1991). *Studies of estimating and testing Rasch models*. (NICI Technical Report 91-02) (Doctoral Thesis) Nijmegen: University of Nijmegen.

Zwinderman, A, H. (1995). Pairwise estimation in the Rasch models. *Applied Psychological Measurement*, *19*(4), 369-375.