JIRSS (2002)
Vol. 1, Nos. 1-2, pp 111-126

On the Multivariate Rasch Model: Assessing Collaboration in Multiple Choice Tests

Yehuda Vardi, Cun-Hui Zhang

Department of Statistics, Rutgers University Piscataway, New Jersey 08854-8019, USA. (vardi@stat.rutgers.edu, czhang@stat.rutgers.edu)

Abstract. We examine the Rasch model for latent structure parameters in binary and multiple response questionnaires and develop methodologies and data-analytic tools for assessing collaboration/cheating in multiple choice tests.

1 Introduction

Data collection in the form of a questionnaire is a common practice in many fields of science and engineering. Applications could vary from product testing in reliability studies, to achievement tests in education. The collected data are used to assess individual's characteristics, such as student's ability based on his test score, or population's characteristics, such as product-reliability where test results of a sample are used to assess the population. Rasch (1960) proposed a latent structure model for analyzing binary response questionnaires.

Received: September 2002

Key words and phrases: Latent variable, logistic regression, maximum likelihood estimate, multiple choice test, number of matches, Rasch model.

The model introduces a "*difficulty*" parameter for each question and an "*ability*" parameter for each examinee, and uses the test results to estimate these parameters. The advantage of this approach, compared to estimating examinee's ability and question's difficulty by the corresponding marginal scores, is that the former calibrates the estimated abilities to the difficulties of the questions. Similarly, the estimated difficulties are calibrated to the abilities of the student who answered these questions. Two additional features of such a latent structure model are: (a) It provides a reasonable way of comparing across different tests, and (b) It provides a simple probability model for assessing more complicated quantities related to the performance of the examinees. This last point is demonstrated in a regression example later to assess collaboration in multiple choice tests.

For achievement tests with more complicated response options, the difficulty-of-question parameter could be a vector. Also, the test could be such that individual's ability is measured in several categories, so it too could be a vector parameter. To accommodate such high dimension situations Rasch (1961, Section 4) extended his original model and proposed the following class of probability functions, as a general statistical framework for analyzing a variety of achievement tests data.

$$P\{y|\alpha_i, \beta_j\} \propto \exp\{\phi(y)\alpha_i + \psi(y)\beta_j + \chi(y)\alpha_i\beta_j + \rho(y)\} \qquad (1.1)$$

Here y is the response, α_i and β_j are the ability and difficulty parameters, possibly vectors, and (1.1) is the probability that subject i chooses response y to question j, when the difficulty of the question is β_j and the ability of the subject is $\alpha_i; i = 1, \ldots, I, j = 1, \ldots, J$, $y \in \{\text{possible responses to question } j\}$. The functions ϕ, ψ, χ and ρ depend on the response, y, only. In the case of vector parameters, the products $\phi(y)\alpha_i$ and $\psi(y)\beta_j$ are interpreted as inner products of vectors, $\chi(y)$ is taken to be a matrix, and $\chi(y)\alpha_i\beta_j$ is a homogeneous bilinear form in the coordinates of α_i and β_j. The responses are assumed to be independent across the i's and j's.

Despite the fact that (1.1) is designed to handle vector parameters, most of the published literature still focuses on the binary response case, while the multivariate situation has not been explored in depth. In this paper we propose a model which is particularly useful and natural for looking at multiple choice test data. The model is a special case of (1.1) and can be labeled a "Rasch model for multiple choice tests." The essential feature is that the "difficulty-of-question" parameter, β_j, which was a one-dimensional parameter in the binary response case, is now an $M-1$ dimensional vector parameter where M is the number of choices for the question. The interpretation is changed accordingly from β_j being the difficulty of question j to β_{jk} being the appeal of the kth incorrect choice as a prospective response to question j; note that there are $M-1$ incorrect choices, as only one of the M is assumed to be correct.

In Section 2 we describe the details of the model and connect the interpretation of the parameters with the mathematical properties of the model. In Section 3 we discuss the likelihood function and the likelihood equations. We further show that the likelihood can be conveniently factorized into a conditional likelihood and a marginal likelihood, with the latter corresponding to a binary response Rasch model. This is used in the estimation procedure, which is discussed in Section 4. Specifically, Section 4 briefly reviews some of the known issues related to the consistency of the maximum likelihood estimator (MLE) and proposes a simple algorithm which converges monotonically, in the sense of increasing likelihood from iteration to iteration, to the MLE of the model's parameter. Section 5 describes an application of the model to a problem of assessing collaboration between two students who were accused of cheating on a test.

2 The Model

The model we propose for multiple choice tests can be formally derived from (1.1) as follows. Denote the possible responses ("choices") for each question by $\delta_1, \ldots, \delta_M$, where $\delta_k = (\delta_{k1}, \ldots, \delta_{kM})$, $k = 1, \ldots, M$,

$$\delta_{km} = \begin{cases} 1, & \text{if } k = m \\ 0, & \text{if } k \neq m \end{cases} \tag{2.1}$$

Without loss of generality assume that the correct response to each of the questions is δ_1. Let β_j be a vector parameter of length M whose first coordinate is 0,

$$\beta_j = (0, \beta_{j2}, \ldots, \beta_{jM}), \quad j = 1, \ldots, J$$

and let α_i be a one-dimensional parameter. In (1.1) set $\chi(y) \equiv 0$, $\rho(y) \equiv 0$, $\phi(y) = 1[y = \delta_1]$ and $\psi(y) = y$. The resulting model can then be described as follows: Let Y_{ij} be the response of subject i to question j, $i = 1, \dots, I$ and $j = 1, \dots, J$. Then

$$P_{ijk} \equiv P\{Y_{ij} = \delta_k\} = \begin{cases} \frac{e^{\alpha_i}}{e^{\alpha_i} + \sum_{m=2}^{M} e^{\beta_{jm}}}, & \text{if } k = 1, \\ \frac{e^{\beta_{jk}}}{e^{\alpha_i} + \sum_{m=2}^{M} e^{\beta_{jm}}}, & \text{if } k = 2, \dots, M \end{cases} \tag{2.2}$$

Alternatively:

$$P_{ijk} \equiv P\{Y_{ij} = \delta_k\} = \frac{e^{\lambda_{ijk}}}{\sum_{m=1}^{M} e^{\lambda_{ijm}}} \tag{2.3}$$

$$i = 1, \dots, I, \; j = 1, \dots, J, \; k = 1, \dots, M$$

with

$$\lambda_{ijk} = \alpha_i I[k = 1] + \beta_{jk} I[k \neq 1] = \begin{cases} \alpha_i, & \text{if } k = 1, \\ \beta_{jk}, & \text{if } k = 2, \dots, M \end{cases} \tag{2.4}$$

The responses are assumed to be independent across questions, for any fixed individual, as well as across individuals.

Interpretation and Observations

(i) α_i are measures the ability of subject i. The probability of correct response from subject i to question j, $P\{Y_{ij} = \delta_1\}$, is a monotone function of α_i approaching 1 as $\alpha_i \to \infty$ and 0 as $\alpha_i \to -\infty$.

(ii) The conditional probability of a *particular* incorrect response to question j, given an incorrect response, is independent of the subject's ability parameter:

$$P\{Y_{ij} = \delta_k | Y_{ij} \neq \delta_1\} = \frac{\exp\{\beta_{jk}\}}{\sum_{m=2}^{M} \exp\{\beta_{jm}\}}$$

(iii) β_{jk} measures the appeal of the incorrect choice δ_k as a plausible choice for question j. For example, if (1) Albany, (2) New York City, (3) Buffalo and (4) Miami, are the four choices in a question about the capital of New York State, δ_2 ("New York City") appears more plausible than δ_3 or δ_4. Note that the probability that subject i chooses δ_k for problem j is a monotone function of β_{jk} which approaches 1 if $\beta_{jk} \to \infty$ and 0 if $\beta_{jk} \to -\infty$. Furthermore, the probability that subject i chooses the correct response to problem i, $P\{Y_{ij} = \delta_1\}$, is a monotone decreasing function in each of the β_{jk} $(k = 2, \dots, M)$ and approaches 0 as $\beta_{jk} \to \infty$ for some k.

(iv) The assumption that the number of possible choices, M, is the same for all questions was made only to simplify the notation. The statistical procedures we derive are easily modified to accommodate tests where different questions may have a different number of choices.

3 The Likelihood

Define the following statistics

$$\begin{cases} Y(i,j) = k \quad \text{iff} \quad Y_{ij} = \delta_k, \\ X_{ij} = I[Y_{ij} = \delta_1], \\ S_i = X_{i+} \equiv \sum_j X_{ij}, \\ C_{ij} = \sum_i I[Y_{ij} = \delta_k] = \sum_{i=1}^I \delta_k Y'_{ij}, \\ C_j = \sum_{k=2}^M C_{jk} = C_{j2} + \cdots + C_{jM} = \sum_i (1 - X_{ij}), \\ T_j = I - C_j = X_{+j} = \sum_i X_{ij} \end{cases} \tag{3.1}$$

Lower case letters indicate the observed values of the corresponding statistics. The interpretation is as follows:
$y(i,j)$ is the numerical-label of individual i's response to question j;
x_{ij} is 1 if subject i answers correctly question j; otherwise x_{ij} is 0;
s_i is the total test score of subject i;
c_{jk} is the total count for response k to question j;
c_j is the total count of incorrect responses to question j;
t_j is the total count of correct responses to question j;

Clearly $t_j + c_j = I$. For example, if individual number nine marked the third choice as the correct one for question number seventeen, then $y(9, 17) = 3$, etc.

Define also the following function of the parameters $\alpha_i, \beta_{j2}, \ldots, \beta_{jM}$,

$$\begin{cases} a_i = e^{\alpha_i}, \;\; b_j = e^{\beta_{j2}} + \cdots + e^{\beta_{jM}}, \;\; i = 1, \ldots, I, \;\; j = 1, \ldots, J, \\ a = (a_1, \ldots, a_I), \;\; b = (b_1, \ldots, b_J), \\ \eta_{jk} = e^{\beta_{jk}}/b_j = P\{Y_{ij} = \delta_k | Y_{ij} \neq \delta_1\}, \;\; k = 2, \ldots, M \end{cases} \tag{3.2}$$

From (1.3), the liklihood function is

$$L = \prod_{ij} \frac{\exp\{\lambda_{ijy(i,j)}\}}{\sum_{m=1}^{M} \exp\{\lambda_{ijm}\}} = \frac{\exp\{\sum_{ij} \lambda_{ijy(i,j)}\}}{\prod_{ij}(a_i + b_j)} \tag{3.3}$$

Simple algebraic manipulations give

$$\sum_{ij} \lambda_{ijy(i,j)} = \sum_i \alpha_i s_i + \sum_{jk} \beta_{jk} c_{jk} \tag{3.4}$$

and so the log-likelihood is

$$\ell \equiv \log L = \sum_i \alpha_o s_i + \sum_{jk} \beta_{jk} c_{jk} - \sum_{ij} \log(a_i + b_j) \tag{3.5}$$

The likelihood equations are

$$\frac{\partial \ell}{\partial \alpha_i} = s_i - \sum_{j=1}^{J} \frac{e^{\alpha_i}}{a_i + b_j} = 0 \tag{3.6a}$$

$$\frac{\partial \ell}{\partial \beta_{jk}} = c_{jk} - \sum_{i=1}^{I} \frac{e^{\beta_{jk}}}{a_i + b_j} = 0 \tag{3.6b}$$

Summing (3.6b) over k gives us

$$c_j - \sum_{i=1}^{I} \frac{b_j}{a_i + b_j} = 0 \tag{3.7}$$

Using this, and the fact that (3.6b) can be rewritten as

$$c_{jk} - \eta_{jk} \sum_{i=1}^{I} \frac{b_j}{a_i + b_j} = 0 \tag{3.6b'}$$

We conclude that

$$\hat{\eta}_{jk} = c_{jk}/c_j \quad (0/0 \equiv 0) \tag{3.8}$$

Plugging this into (3.6b′) shows that (3.6b) can be replaced by (3.7) which is independent of k (the choice of the incorrect response), $k = 2, \ldots, M$. Thus, we can combine (3.6a), (3.7) and (3.8) (with an obvious modification to (3.7)) as an equivalent form of the likelihood equations:

$$\sum_j \frac{a_i}{a_i + b_j} = s_i, \quad i = 1, \ldots, I \tag{3.9a}$$

$$\sum_i \frac{a_i}{a_i + b_j} = t_j, \quad j = 1, \ldots, J \tag{3.9b}$$

$$\hat{\eta}_{jk} = c_{jk}/c_j, \quad j = 1, \ldots, J,\ k = 2, \ldots, M \tag{3.10}$$

An equivalent form of (3.9) is

$$EX_{i+} = x_{i+},\ i = 1, \ldots, I; \quad EX_{+j} = x_{+j},\ j = 1, \ldots, J \tag{3.11}$$

That is, each row-margin of the matrix $\{a_i/(a_i + b_j)\}$ is equal to the corresponding subject's test score, while each column-margin is equal to the corresponding question's total correct count.

3.1 Standardization of the Model (Identifiability)

It is clear from (2.2) and (3.2) that the model (2.2) is invariant under the scale change

$$(a_i, b_j) \to (a_i, b_j)d \quad (d > 0)$$

so that the parameters are identifiable only up to a multiplicative constant. This is also apparent in the likelihood equations (3.6) and (3.9-10). The physical interpretation is that a matching scale-change to both units of measurement, that of ability of subjects (a_i) and that of difficulty of questions (b_j), does not effect the response probabilities. (Makes sense!) To eliminate this indeterminacy from our model, we set $a_i = 1$ for some $1 \leq i \leq I$. Excluding subjects who answered the questions all correctly or all incorrectly, for whom $\hat{\alpha}_i = 0$ or ∞, it is convenient to set the a of the lowest scoring subject to zero:

$$a_{i_1} = 1, \quad \text{where } i_1 = \arg\min\{s_i; i = 1, \ldots, I, s_i > 0\} \tag{3.12}$$

3.2 Decomposition of the Model

The likelihood equations (3.9-10) can be derived directly from the likelihood function upon observing that it can be decomposed as follows:

$$L = \left[\prod_{ij} \frac{a_i^{x_{ij}} b_j^{(1-x_{ij})}}{a_i + b_j}\right] \left[\prod_{ij} \eta_{jy(i,j)}^{1-x_{ij}}\right] = L_1 L_2 \qquad (3.13)$$

where

$$L_1 \equiv \prod_{ij} a_i^{s_i} b_j^{c_j} / \prod_{ij} (a_i + b_j) \qquad (3.14)$$

$$L_2 \equiv \prod_{jk} \eta_{jk}^{c_{jk}} \qquad (3.15)$$

L_1 is the likelihood of the data when we do not distinguish between different incorrect choices. This is the likelihood function of the standard, binary-response, Rasch model and the likelihood equations are those in (3.9). L_2 is the likelihood of the observed responses, conditionally on them being incorrect. From (3.2)

$$\sum_{k=2}^{M} \eta_{jk} = 1, \quad j = 1, \ldots, J$$

so that (3.10) gives the likelihood equations for L_2, or rather the MLE's for η_{jk} based on the likelihood L_2. The decomposition in (3.13) also shows that the statistics s_i, $1 \leq i \leq I$, c_j, $1 \leq j \leq J$ are sufficient for a_i, $1 \leq i \leq I$, b_j, $1 \leq j \leq J$, and the statistics c_{jk}, $1 \leq j \leq J$, $2 \leq k \leq M$, are sufficient for η_{jk}, $1 \leq j \leq J$.

4 Estimation

There are various likelihood-based methods for estimating the a_i's and b_j's, and from a practical point of view the appropriateness of different methods depends, among other things, on the relative values of I and J. For instance, if either I or J, but not both, is very large, the usual likelihood methodology fails due to a large number of incidental parameters, and conditional likelihood methods and mixture model techniques have been used to derive consistent estimators. (See Lindsay, Clogg and Grego, 1989, for a comprehensive discussion.) From a technical point of view the problem is the following: If, for instance, $b_1, \ldots, b_J$ are the model's "structural" parameters, and $a_1, a_2, \ldots, a_I$ are the "incidental" parameters (unknown constants), then the MLE of $b_1, \ldots, b_J$ is inconsistent fot fixed J when $I \to \infty$. On the other hand, if both $I \to \infty$ and $J \to \infty$, then under some mild conditions (Haberman, 1977) the MLE of $b_1, b_2, \ldots, b_J$ is consistent. In fact, when I *and* J approach infinity, both sets of parameters, $b_1, b_2, \ldots, b_J$ and $a_1, a_2, \ldots, a_I$ can be estimated consistently by the MLE.

The following is a very simple iterative procedure which converges monotonically (in the sense of increasing likelihood from one iteration to the next) to the solution of (3.9); i.e. to the MLE of a and b.

An Algorithm for Estimating a *and* b

The following iterative procedures solve (3.9).

Initialization. Start the algorithm with

$$a_{i_1}^{old} = 1 \text{ (for identifiability)} \tag{4.1a}$$

$$a_i^{old} = \begin{cases} 0 & \text{if } s_i = 0, \\ \infty & \text{if } s_i = J, \\ a_i^{start} & \text{if } 0 < s_i < J \quad \text{and } i \neq i_1 \end{cases} \tag{4.1b}$$

$$b_j^{old} = \begin{cases} 0 & \text{if } c_j = 0, \\ \infty & \text{if } c_j = I, \\ b_j^{start} & \text{if } 0 < c_j < I \end{cases} \tag{4.1c}$$

with i_1 defined in (3.12) and a_i^{start} and b_j^{start} being arbitrary finite positive numbers, $i = 2, \ldots, I$ and $j = 1, \ldots, J$.

Iteration. For $i = 1, \ldots, I$, a_i^{new} is the solution of

$$\sum_{j=1}^{J} \frac{a_i}{a_i + b_j^{old}} = s_i, \text{ if } 0 < s_i < J \tag{4.2a}$$

otherwise $a_i^{new} \leftarrow a_i^{old}$.

$$a_i^{old} \leftarrow a_i^{new} / a_{i_1}^{new}, \quad i = 1, \ldots, I \tag{4.2b}$$

(This rescales the a_i's and resets a_{i_1} back to 1, to maintain the identifiability condition.)

For $j = 1, \ldots, J$, b_j^{new} is the solution of

$$\sum_{i=1}^{\ell} \frac{b_j}{a_i^{old} + b_j} = c_j, \quad \text{if } 0 < c_j < I \tag{4.2c}$$

otherwise $b_j^{new} \leftarrow b_j^{old}$.

$$b_j^{old} \leftarrow b_j^{new}, \quad j = 1, \ldots, J \tag{4.2d}$$

$$\text{Return to (4.2a).} \tag{4.2e}$$

Necessary and sufficient conditions for the existence and uniqueness of the MLE were given in Fischer (1981). Fisher's conditions can be described as follow: Assume, without loss of generality, that the s_i's and t_j's are ordered, $s_1 \leq \cdots \leq s_I$ and $t_1 \leq \cdots \leq t_J$. Then the likelihood equation (3.9) has a unique solution if and only if there do *not* exist integers i_0 and j_0 such that

$$0 < s_{i_0} \leq s_{i_0+1} < J, \quad 0 < t_{j_0} \leq t_{j_0+1} < I \tag{4.3a}$$

$$\sum_{i=1}^{i_0} s_i + (I - i_0)(J - j_0) = \sum_{j=j_0+1}^{J} t_j \tag{4.3b}$$

Theorem 4.1. *Under the above conditions of Fischer, the algorithm (4.1)-(4.2) converges to the unique MLE of a and b.*

Proof. For the purpose of this proof, it is simpler to consider the likelihood as a function of the natural parameters. Because of the factorization (3.13) we can assume, without loss of generality, that the test consists entirely of binary response questions ($M = 2$). In this case $\beta_j \equiv \beta_{j2} \equiv \log b_j$, and the likelihood L_1 of (3.14) is log concave in α_i, β_j ($i = 1, \ldots, I,\ j = 1, \ldots, J$) and under Fisher's conditions has a unique point of maximum $(\hat{\alpha}, \hat{\beta})$. Furthermore, $(\hat{\alpha}, \hat{\beta})$ is the unique root of the likelihood equation (3.9) (Fisher 1981, Theorem 1). If we remove all subjects i with $s_i = 0$ or J (and guestions j with $c_j = 0$ or I), the values of $\hat{\alpha}_i$ and $\hat{\beta}_j$ and the values of the iterative sequence in the algorithm are unchanged for the remaining subjects and questions. Thus, we shall further assume $0 < s_i < J$

and $0 < c_j < I$ for all i, j. Combining the existence, uniqueness and finiteness of the MLE with the concavity and continuity of $l = \log L_1$, we find that it can be shown that the set

$$\Omega = \{(\alpha, \beta) : l(\alpha, \beta) \geq l(\alpha^{start}, \beta^{start})\}$$

is convex and compact. The algorithm is such that in each iteration we maximize the log-likelihood along a particular direction, so that the likelihood increases from one iteration to the next. Let Δ_j be the increment of the log-likelihood l at step (4.2c). Since the log-likelihood is increasing and bounded, the sequence of differences, $\{\Delta_j\}$, must diminish to zero as the iterations progress. That is, for any fixed j, $\Delta_j \to 0$ as the number of iterations goes to infinity. Since $\partial l/\partial \beta_j = 0$ at β_j^{new}, by the one-dimensional Taylor expansion of l as a function of β_j, about β_j^{new}, we get

$$-(\beta_j^{old} - \beta_j^{new})^2(\partial^2 l/\partial \beta_j^2)/2 = \Delta_j \to 0 \tag{4.4}$$

where the second derivative $\partial^2 l/\partial \beta_j^2$ is evaluated at $\alpha^{old}, \beta_1^{new}, \ldots,$ β_{j-1}^{new}, $\tilde{\beta}_j, \beta_{j+1}^{old}, \ldots, \beta_J^{old}, \tilde{\beta}_j$ for some between β_j^{old}, for some $\tilde{\beta}_j$ between β_j^{old} and β_j^{new} which lies in the set Ω. Since l is strictly concave in the direction β_j, $\partial^2 l/\partial \beta_j^2 < 0$. By the compactness of Ω and the continuity of the second derivative, $\partial^2 l/\partial \beta_j^2$ in (4.4) is uniformly bounded away from 0, so that necessarily $\beta_j^{old} - \beta_j^{new} \to 0$ as the iterations progress to infinity. By the same argument, we can also show that $\alpha_i^{old} - \alpha_i^{new} \to 0$. Hence, any point of convergence satisfies (3.9), and the theorem follows from the compactness of Ω and the uniqueness of the solution of (3.9).

5 An Example

The example we consider originated from our consulting work in connection with a test taken by thirty seven students, two of whom were accused of cheating by collaborating with each other during the exam. The test was made up of an essay and forty two multiple choice questions, with four choices for each question. A comparison of the two essays showed a strong similarity in choice of topics, key sentences, etc. The students attributed the similarity to the fact that they studied together. On the multiple choice part the two students also had a disproportionally high number of matches. In fact, their responses matched on 40 out of the 42 questions, which is very high, relative to their low overall scores of 28 and 29 (out of a possible maximum of 42) and taking into account the responses of other students. A complete discussion of the relevant statistical methodology and data analysis, for assessing whether the data provide significant evidence against the students, is outside the scope of this paper. Our purpose in bringing this example here, is to demonstrate how the extended Rasch model is used in estimating parameters from the multiple-choice test data.

Let $M_{ii'}$ be the total number of matched responses between students i and i', and $M^w_{ii'}$ be the total number of matched responses on wrong choices between students i and i'. If students i and i' collaborate during the test, there should be a high rate of matches in their responses. However, the number of matches are also statistically dependent on the scores. For example, the pair (i, i') tends to have large $M_{ii'}$ (and small M^w_{ii}) when both s_i and $s_{i'}$ are near J, although this dependency is not monotone in the scores throughout their range. Another example is that both $M_{ii'}$ and $M^w_{ii'}$ tend to be small when the score difference is large. To assess the significance of the matching pattern between a pair of students with equal or very similar over all scores, it is useful to plot $M_{ii'}$ and $EM_{ii'}$, or alternatively $M^w_{ii'}$ and $EM^w_{ii'}$, for all pairs (i, i'), $1 \le i \le i' \le I$, for which $|S_i - S_{i'}|$ is small (say $|S_i - S_{i'}| \le d$, where d is the score difference between the two suspected students) and $S_i + S_{i'} = s$ varies over the relevant range. Such plots can serve as proxies for the more conventional, but computationally complicated, regression plots,

$$E[M_{ii'} \,|\, |S_i - S_{i'}| \le d,\, S_i + S_{i'} = s]$$

and

$$E[M^w_{ii'} \,|\, |S_i - S_{i'}| \le d,\, S_i + S_{i'} = s]$$

The rational behind plotting $M_{ii'}$ and $EM_{ii'}$ as functions of the sum of the scores for all (i, i') with similar test scores is that it exhibits a well understood function between the matching level and the score level. Specifically, consider

$$EM_{ii'} = \sum_{j=1}^{J} \sum_{k=1}^{M} P_{ijk} P_{i'jk} \quad \text{and} \quad EM^w_{ii'} = \sum_{j=1}^{J} \sum_{k=2}^{M} P_{ijk} P_{i'jk}$$

Under the extended Rasch model,

$$P_{ijk} = \begin{cases} \frac{a_i}{a_i+b_j}, & \text{if } k=1, \\ \frac{b_{jk}}{a_i+b_j}, & \text{if } k=2,\ldots,M \end{cases}$$

where $a_i = e^{\alpha_i}, b_j = \sum_{m=2}^{M} e^{\beta_{jm}}$ and

$$b_{jk} = b_j \frac{e^{\beta_{jk}}}{\sum_{m=2}^{M} e^{\beta_{jm}}} = b_j \eta_{jk}$$

It follows that

$$EM_{ii'} = \sum_{j=1}^{J} \frac{(\sum_{k=2}^{M} \eta_{jk}^2) b_j^2 + a_i a_{i'}}{(a_i + b_j)(a_{i'} + b_j)}$$

and

$$EM_{ii'}^{w} = \sum_{j=1}^{J} \frac{b_j^2 \sum_{k=2}^{M} \eta_{jk}^2}{(a_i + b_j)(a_{i'} + b_j)}$$

which depend on students i and i' only through their ability parameters a_i and $a_{i'}$. Now, it is clear from the estimating equations (3.9a) that students with the same overall scores have the same estimated abilities $\hat{a}$ (because their estimating equations are interchangeable) and so $\widehat{EM}_{ii'}$ depends on the pair (i, i') only through the pair of scores $(s_i, s_{i'})$. In other words, there exists a function, $h(u, v)$ such that $s_i = u$ and $s_{i'} = v$ imply $\widehat{EM}_{ii'} = h(u, v)$. Figure 1 plots the functions $h(u, v)$ against $u + v = s$ for $|u - v| = 0$ and 1. Specifically, it plots

$$\widehat{EM}_{ii'} = \sum_{j=1}^{J} \frac{(\sum_{k=2}^{M} \hat{\eta}_{jk}^2)\hat{b}_j + \hat{a}_i \hat{a}_{i'}}{(\hat{a}_i + \hat{b}_j)(\hat{a}_{i'} + \hat{b}_j)}$$

against the sum of the test-scores when the difference of the scores is constrained to be small. It also gives a scatter plot of the observed matches $M_{ii'}$ for all pairs with a small score difference (≤ 1) as a function of the sum of the test scores. Figure 2 gives the analog results calculated only for matches on wrong choices $M_{ii'}^{w}$. Both pictures flag out the pair of students accused of cheating as an extreme outlier. The estimation of the parameters for the plots was determined on the basis of the extended Rasch model as discussed in Section 4. As a remark we add also that we have seen the argument in the education-testing literature that only matches on wrong choices should be used to assess cheating. Note, however, that for binary response questions we have the identity

$$M_{ii'} = S_i + S_{i'} + 2M^w_{ii'} - J$$

so that inference based on $M_{ii'}$ is equivalent to inference based on $M^w_{ii'}$. Much of this relation is preserved in multiple-choice tests, because many of the questions are *affectively* binary responses.

References

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

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

Rasch, G. (1960/1980), Probabilistic Models for Some Intelligence and Attainment Tests. Expanded edition, Chicago: University of Chicago Press.

Rasch, G. (1961), On general laws and the meaning of measurement in psychology. Proc. Fourth Berkeley Symp. Math. Statist. Probab. **4**, 321-333.

Lindsay, B., Clogg, C. C. and Grego, J. (1989), Semi-parametric Estimation in the Rasch Model and Related Exponential Response Models. Including a Simple Latent Class Model for Item Analysis.

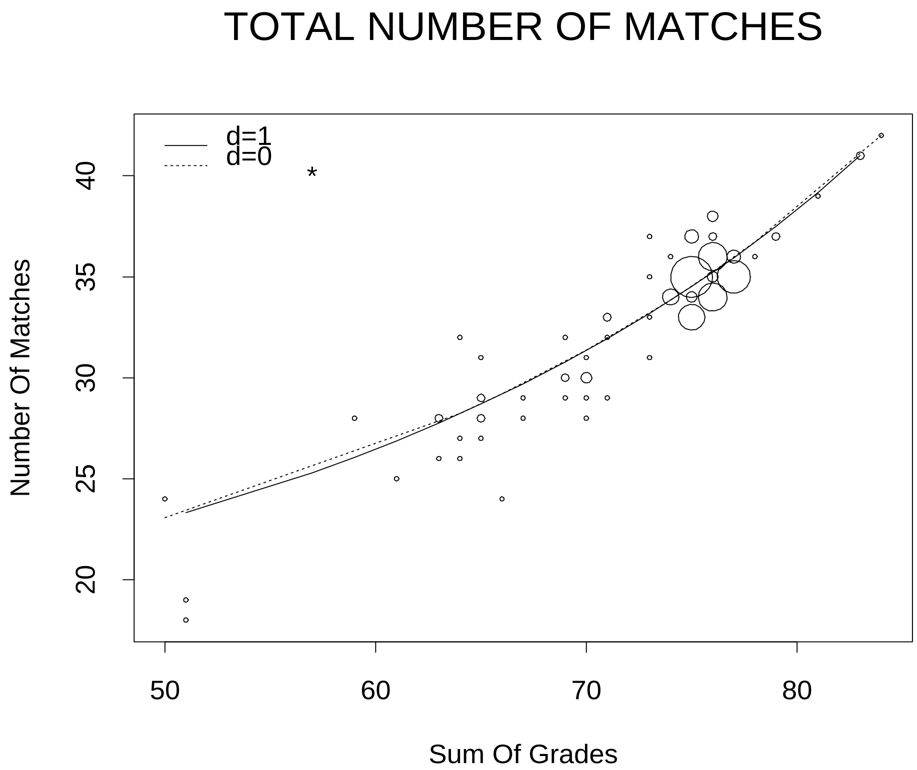

Figure 1: Total number of matches: The area of a circle is proportional to the number of observations at a point. The estimated expectations are connected by line segments into curves for $d = 0$ and 1, where d is the difference of scores.

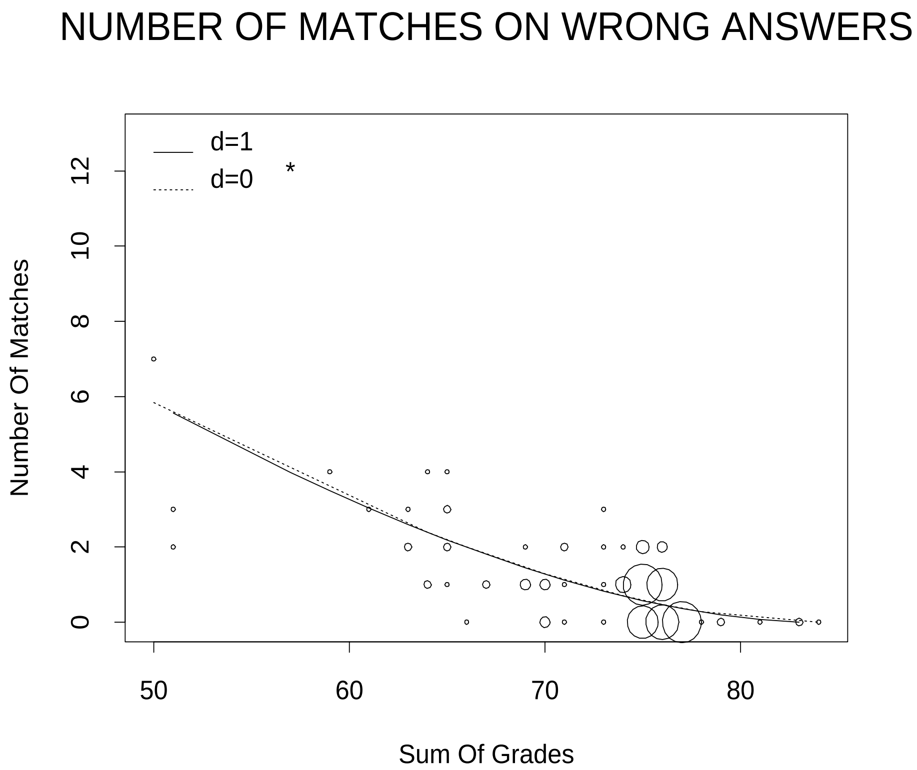

Figure 2: Number of matches on wrong answers: The area of a circle is proportional to the number of observations at a point. The estimated expectation are connected by line segments into curves for $d = 0$ and 1, where d is the difference of scores.