ISSN: 1139-613X

1

THE RASCH APPROACH TO "OBJECTIVE MEASUREMENT" IN THE PRESENCE OF SUBJECTIVE EVALUATION FROM "JUDGES"

(APROXIMACIÓN AL MODELO DE "MEDICIÓN OBJETIVA" DE RASCH EN PRESENCIA DE LA EVALUACIÓN SUBJETIVA POR JUECES)

Enrico Gori
Michela Battauz
Università di Udine

RESUMEN

Algunas de las actividades humanas -deporte, educación, economía, investigación, desarrollo profesional, alimentación- requieren la participación de jueces en la evaluación de aspectos que son difíciles de medir de forma directa. Como estas evaluaciones pueden tener consecuencias relevantes para los sujetos examinados, es necesario investigar la máxima objetividad del proceso evaluador. El modelo de Rasch es el único procedimiento estadístico que asegura la medición objetiva, incluso en el proceso evaluador de jueces. Este artículo repasa la teoría del modelo de Rasch y propone una aplicación de datos relativos a la evaluación de proyectos financiados. Se examinan también las alteraciones significativas que se detectan cuando se emplean los modelos de las mediciones de Rasch.

ABSTRACT

A variety of human activities - sport, education, finance, research, professional development, feeding - require the participation of judges in order to evaluate aspects that are difficult to be measured directly. As this evaluations may have important consequences on the examined subjects, it is necessary to research the maximum objectivity in the evaluation process. The Rasch model is the unique statistical model that assures the construction of objective measurements, even in the presence of judges evaluations. This paper reviews Rasch models theory and proposes an application to data concerning the evaluation of projects presented for a funding competition. The serious alterations that arise from the use of the rough scores instead of the Rasch measures are explored.

1. INTRODUCTION

A variety of human activities - sport, education, finance, research, professional training, feeding - involve the judgement of aspects that are difficult to be measured directly, and then called "latent traits"[1]. Such situations are characterised by the presence of three sets: the set A of subjects to be judged with respect to some characteristics, the set B of tests able to provide useful informations on the measure of the latent trait, the set C of judges who observe how the subjects perform in the various tests and give them a judgement. The judgement of subject $a \in A$ on test $b \in B$ submitted to the evaluation of judge $c \in C$ provides the result *r=r(a,b,c)*. R is the set of all possible results. The collection of this 4 sets constitute the (three factors) reference system $F=\{A, B, C, R\}$ for the measure of the latent trait (Rasch, 1977)[2]. As this evaluations may strongly influence career, success and income of the examined subjects, it is necessary to research the maximum objectivity in the evaluation process. The objectivity degree of the evaluation depends on two fundamental factors: a) the possibility of eliminating the subjectivity in the measure of the latent trait, which arises from the subjective judgement of the evaluators, and the weights assigned to each test; b) the content validity, that means the tests appropriateness with respect to the latent trait to be measured[3] (Bond, 2003). In order to solve such problems, there exists a general methodology. It is based on a mathematical model which attributes the evaluations received to three factors: the subject ability[4], the test difficulty, and the judge severity[5]. The construction of such model is based on the research developed by the Danish mathematician Georg Rasch (1960). Reasoning about what characterises the superiority of natural sciences respect human sciences, Rasch concluded that the concept of "science" is related to the possibility of developing methods for transforming observations into measurements, according to rules that satisfy the specific objective principle. Intuitively, such principle means that the measurement method provide a measure of some latent trait of the subject independently of other subject features, other subjects or characteristics of the tool use for measuring. To this end, it is necessary modelling the observations in the reference system $F=\{A, B, C, R\}$ according to the Rasch model family (Gori, et al., 2005). It is important distinguishing the statistical approach, that tries to find the model that better fits data, from the Rasch approach, that requires the data to fit a model developed on the basis of the specific objectivity principle. Some authors state that the principle *"there is nothing more practical than a good theory"* is a necessary condition for a good research, both in experimental and observational studies (Embretson and Hershberger, 1999; Masters and Keeves, 1999; Rowe and Cilione, 2001; Wilson and Engelhard, 2000; Wright and Mok, 2000).

2. THE RASCH MODEL FAMILY

The Rasch model family is funded on three assumptions (Hambleton and Swaminathan, 1985):

A1. Unidimensionaliy. There exists an unidimensional latent trait θ_n, called latent ability, associated to a generic subject n, that determines his capacity of succeed the test submitted to him; the tests are related to this unique dimension[6] and are characterised by a difficulty δ_i=1, 2, K, I; the judges are characterized by a parameter Υ_j, j=1, K, J called severity.

A2. Monotonicity. X_{nij} represents the evaluation obtained by subject n, on test i, from judge j and constitutes a random variable which satisfies the condition that $P(X_{nij} > t|\theta_n, \delta_i, \Upsilon_j)$ is a monotonic function of the ability θ_n, for each i and t. Subjects with higher abilities have a greater probability of getting an higher evaluation. This assumption allows for utilizing the vector of observations on subject n in the various tests, $\mathbf{X}_n=\{X_{n1j}, X_{n2j}, K, X_{nIj}\}$, as repeated measures on the same subject.

A3. Local independence.
$P(\mathbf{X}_n|\theta_n,\delta_1,\delta_2\cdots\delta_I,\gamma_1,\gamma_2,\ldots\gamma_J)=\prod_{i=1}^{I}\prod_{j=1}^{J}P(X_{n\,ij}|\theta_n,\delta_i,\gamma_j)$, that is: conditioning on subject ability, test difficulty and judge severity, the random variables $\mathbf{X}_n=\{X_{n1j}, X_{n2j}, K, X_{nIj}\}$ are independent.

2.1 The binary case (absence of judges)

Every test submitted to subjects presents a binary response (correct/wrong, succeed/fail, ecc.). In this case, the Rasch model is

$$(1)\quad P(X_{ni}=1)=\frac{e^{\theta_n-\delta_i}}{1+e^{\theta_n-\delta_i}}$$

and it is the only IRT model that satisfies the *specific objectivity* condition. Such model is derived from the condition

$$(2)\quad \ln\frac{P(X_{ni}=1)}{P(X_{ni}=0)}=\theta_n-\delta_i$$

that, referred to two subjects m and n, and to any test i, allows to express the difference between the person parameters as function of the probabilities

$$(3)\quad \ln\frac{P(X_{mi}=1)}{P(X_{mi}=0)}-\ln\frac{P(X_{ni}=1)}{P(X_{ni}=0)}=(\theta_m-\delta_i)-(\theta_n-\delta_i)=\theta_m-\theta_n\ ,$$

that does not depend on the item parameter δ_i[7].

2.2 The partial credit and rating scale models (absence of judges)

In the case that responses are of ordinal type (i.e. on a Likert scale), the binary model is extended to the *partial credit* model (Masters, 1982) or to the *rating scale* model (Andrich, 1978a):

(4) $P(X_{ni}=k):\ln\frac{P(X_{ni}=k)}{P(X_{ni}=k-1)}=\theta_n-(\delta_i+\tau_{ik})$, $k=$ 0,1,2K K_i *(partial credit)*;

(5) $P(X_{ni}=k):\ln\frac{P(X_{ni}=k)}{P(X_{ni}=k-1)}=\theta_n-(\delta_i+\tau_k)$, $k=$ 0,1,2K K *(rating scale).*

(Both these models result unidentified, and require the constrains $\sum_{k=1}^{K_i}\tau_{ik}=0,\forall i$ or $\sum_{k=1}^{K}\tau_k=0$, to be estimated.)

The parameter δ_i represents the average difficulty of test *i*, and τ_{ik} is the additional difficulty of attain level *k* in test *i*. The *rating scale* model (particular case of the *partial credit* model) assumes such parameters constant across tests. In empirical applications it is possible the presence of tests with different responses types, so it is necessary the specification of a mixed response model, where some tests present the binary Rasch specification and other tests present the *partial credit* or *rating scale* specification. For example, a recent research aimed to evaluate students knowledge level in history (Irer, 2005) uses different evaluation criteria for the items (cfr. tab. 3). For example, the evaluations of some items are expressed on a Likert scale with 3 levels, while the evaluations of other items are expressed on a Likert scale with 4 levels: the choice is guided by the minimisation of erroneous classifications and by the suitability with the item.

2.3 The multifacet model

When the tests are evaluated by judges, the model that satisfies the specific objectivity principle was developed by Linacre e Wright (1997) and is called *multifacet* (or *many facets*). Denoting by X_{nij} the response given by judge *j* to subject *n* with respect to test *i*, the model takes the following form

(6) $P(X_{nij}):\ln\frac{P(X_{nij}=1)}{P(X_{nij}=0)}=\theta_n-\delta_i-\gamma_j$

In this version judge *j* establish if subject *n* has failed (X_{nij}= 0) or not (X_{nij}=1) test *i*. This is then a binary model with an additional parameter Υ_j that can be interpreted as judge severity. Here it is important to highlight that often results natural (but not necessary) administering all the tests to all the persons, but it is rather impossible get the evaluations of each judge for every subject in every test. This is the case, for example, of the evaluation of projects by a couple of evaluators chosen in a larger set of evaluators. Model (6) is straightforward to extend to the case of ordinal response items:

(7) $P(X_{nij}):\ln\frac{P(X_{nij}=k)}{P(X_{nij}=k-1)}=\theta_n-\delta_i-\gamma_j-\tau_k$

(items and judges have the same thresholds),

$$(8)\quad P\left(X_{nij}\right):\ln\frac{P\left(X_{nij}=k\right)}{P\left(X_{nij}=k-1\right)}=\theta_n-\delta_i-\gamma_j-\tau_{ik}$$

(every item has different thresholds),

$$(9)\quad P\left(X_{nij}\right):\ln\frac{P\left(X_{nij}=k\right)}{P\left(X_{nij}=k-1\right)}=\theta_n-\delta_i-\gamma_j-\tau_{jk}$$

(every judge has different thresholds)

$$(10)\quad P\left(X_{nij}\right):\ln\frac{P\left(X_{nij}=k\right)}{P\left(X_{nij}=k-1\right)}=\theta_n-\delta_i-\gamma_j-\tau_{ijk}$$

(every judge has different thresholds for each item)

where the parameters $\tau_{\bullet k}$ can be interpreted, likely the *partial credit* and *rating scale* models, as the additional difficulties to attain level *k*. The first two models (7) and (8) correspond to the *rating scale* and *partial credit* versions in presence of judges; model (9) assumes that the thresholds can be different according to the judge, while model (10) presents thresholds that vary across judges and items (the constrains $\sum_{k=1}^{K}\tau_{ki}=0$, $\sum_{k=1}^{K}\tau_k=0$, and $\sum_{k=1}^{K}\tau_{jk}=0$ $\sum_{k=1}^{K}\tau_{ijk}=0$, helps in interpreting δ_i and Υ_j as average difficulties and severities, Linacre, 1998). Also in this cases, the models satisfy the specific objectivity condition. However, "*...allowing each judge his own rating scale weakens inference because it lessens the generality of the measures obtained. Were a new judge included, it would be necessary to estimate not only his level of severity but also his own personal manner of using the rating scale*"[8]. We can then conclude that a major objectivity is attained by model (7), as it does not require additional parameters that reduce estimation efficiency and imply measurement scale with particularities that should be avoided for a better comparability over time and space.

Interactions between judges and items, or between judges and subjects are not admitted in the model as they compromise the *specific objectivity* property: interactions between judges and subjects imply favouritism of judges respect to some subjects, interactions between judges and items imply disagreement between judges about the importance of items. This interactions produce bias in the measurement process.

2.4. Analysis of the presence of bias

Lynch and McNamara (1998) propose a method for assessing the presence of judge bias with respect to items or persons. To this aim, an interaction term is included in the model (for example in model (7)).

$$(11)\quad P\left(X_{nij}\right):\ln\frac{P\left(X_{nij}=k\right)}{P\left(X_{nij}=k-1\right)}=\theta_n-\delta_i-\gamma_j-\tau_k+\left\{\begin{array}{l}C_{nj}[1]\\ C_{ni}[2]\\ C_{ij}[3]\\ C_{nij}[4]\end{array}\right. .$$

Where:

- in case [1] there is an interaction between subjects and judges that allows for detecting the judge equality with respect to subjects ($C_{nj} = 0 \; \forall \; n$), or the assignment of higher or lower scores to some subject compared to the evaluations given to the others;

- in case [2] there is an interaction between subjects and items that allows for detecting the equal functioning of the item with respect to subjects ($C_{ni} = 0 \; \forall \; n$), or the major difficulty of the item for some persons;

- in case [3] there is an interaction between items and judges that allows for discorvering the judge equality with respect to items ($C_{ij} = 0 \; \forall \; i$), or a different behaviour of the judge in some item;

- in case [4] there is an interaction between items, judges and subjects that allows for detecting the equal behaviour of a couple item-judge with respect to subjects, or the presence of higher or lower evaluations from some couple item-judge.

2.5 The centrality of the goodness of fit

In the contest of the Rasch models, the goodness of fit indexes have a fundamental role, as the researcher does not search a model that better fits data, but requires the data to fit the model. Consequently, if some items do not fit the model these have to be eliminated or reformulated (Bond, 2003); if subjects give responses different from model predictions these have to do the test again or they have to be eliminated from the analysis; if judges presents biases with respect to subjects or items, they have to be substituted or adequately trained.

The fit indexes proposed and implemented in the most common software depend on the model used. They can then regard subjects or items, but also judges in the multifacet version. Such indexes (Wright and Masters, 1982) are based on the differences between observed and expected values, divided by the standard deviation (both computed under the hypothesis that the model is adequate), then

$$(12)\quad z_{ni}=\frac{x_{ni}-E\left(X_{ni}\right)}{\sqrt{V\left(X_{ni}\right)}}\ ,$$

where

$$(13) \quad \hat{E}(X_{ni})=\sum_{k=1}^{k_i} k \,.\, \hat{P}(X_{ni}=k), \quad \hat{V}(X_{ni})=\sum_{k=1}^{k_i} (k-\hat{E}(X_{ni}))^2 \,.\, \hat{P}(X_{ni}=k),$$

and $\hat{P}(X_{ni})$, is the probability specified by the Rasch model, computed with the estimated parameters. There are two types of indexes: *Infit* and *Outfit*, and they are reported in Table 1. The first one is more sensible to large differences in theoretical values around 0.5 (that present a larger variance), while the second one is more sensible to large differences in theoretical values around zero and one.

Tab. 1 - Fit indexes in the Rasch model

Index	Person	Item
Infit	$I_n = \frac{\sum_{i=1}^{I} V(X_{ni}) \cdot z_{ni}^2}{\sum_{i=1}^{I} V(X_{ni})}$	$I_i = \frac{\sum_{n=1}^{N} V(X_{ni}) \cdot z_{ni}^2}{\sum_{n=1}^{N} V(X_{ni})}$
Outfit	$O_n = \frac{1}{I}\sum_{i=1}^{I} z_{ni}^2$	$O_i = \frac{1}{N}\sum_{n=1}^{N} z_{ni}^2$

Values of the indexes greater than 1 indicate the presence of a larger variability than expected from the model (this happens when the responses to a test are given by chance); values smaller than 1 indicate a dependence in the data major than that hypothesized[9]. When data fit the model, these indexes have an expected value of 1 and, using the transformations of Wilson-Hilferty (Wright and Masters 1982), they can be approximated with a standard normal random variable, under the null hypothesis that the true model is the Rasch one. The goodness of fit of an item or a subject to the model, using this transformation, can then be performed referring to the standard interval (-2,+2) for a significance level of about 5%. When, instead, the indexes of infit and outfit are used, it is possible to refer to the practical rules reported in literature (tab. 2) (Bond and Fox, 2001). As Linacre highlights[10], however, keeping in the analysis the observations (subjects, items or judges) that present low values of the goodness of fit indexes would not alter the meaning of the measure, but it would reduce the precision increasing the standard errors of the estimates.

Tab. 2 - Invervals for the Infit and Outift indexes
(Bond and Fox, 2001)

Type de test	**Interval**
Multiple responses (1)	0.8 - 1.2
Multiple responses (2)	0.7 - 1.3
Rating scale (Likert scale)	0.6 - 1.4
Clinical observations	0.5 - 1.7
Judges presence	0.4 - 1.2

(1) the exam has important consequences for the student
(2) the exam is aimed to research

It is important to recall here that some authors (Nickerson and McClelland, 1984) showed, through simulation studies, that these indexes tend overestimate the goodness of fit. This is attributed to the computation of the theoretical probabilities on the basis of the same data used for the computation of the indexes. Recently, some alternative and more powerful indexes have been proposed (Karabatsos, 2001), especially for verifying assumptions A1-A2-A3. However, these are very difficult to compute and they are not yet implemented in standard softwares. Curtis (2004) also highlighted the necessity of developing more sensible indicators[11].

3. THE PROBLEM AND THE DATA AVAILABLE

3.1 Projects evaluation

The application described in the following regards the evaluation of projects presented to an Italian region for a funding competition. Projects selection was composed by two phases:

1. A preliminary investigation aimed to verify the formal correctness and the completeness of the application, of the documentation and the coherence of the projects with the objectives established in the announcement.

2. Evaluation of the contents and projects ranking on the basis of criteria defined in the announcement.

An evaluation committee performed the second phase, that was composed by three steps, each requiring the assignement of a score to several criteria:

1. Technical and scientific evaluation, aimed to verify the technical and scientific quality of the project, the competence and the operative capability of the proponents, the quality of the plan for exploiting and transferring the results, the coherence of the finance plan. In order to reach the successive step, the project must attain a minimum score in each of these aspects.

2. Evaluation of the regional priority elements defined in the announcement; in particular, the priority of the specific objective chosen, the involvement of other subjects interested in the results, the co-funding of other subjects interested in the research, the transferability of the results to public technical services, the annual length of the project.

3. Evaluation of the coherence with the regional programs, referring also to the economical/social importance of the area interested.

The preliminary investigation selected 123 projects for the content evaluation. Using rough scores the committee produced a ranking of these projects and the first 85 projects were considerer suitable for being financed. The funding availability allowed for finance 36 of theme.

3.2 The technical and scientific evaluation process

Judges evaluations were given by assigning to each project a score in each of the 14 criteria reported in table 3 and grouped in four groups :

1. technical and scientific quality and originality of the project,
2. possibility of transferring and exploiting the results,
3. competence and operative capability of the proponents,
4. coherence and management of the resources.

Tab. 3- Description of the criteria utilized for evaluating the projects

Cod.	Criteria description
v11	Description of the state of the art and analysis of the needs
v12	Clarity and practicability of the project objectives
v13	Scientific quality and innovative level of the research
v14	Suitability of the methodological approach and the operative plan
v15	Quality of the costs/benefits analysis
v21	Presence of indicators on the result and their coherence
v22	Quality of the program about informative initiatives and transfering of the results
v23	Utility of the results and time necessary for using them
v31	Competence of the proponents (on the basis of the curriculum)
v32	Suitability of structures and equipment available for the project
v33	Presence of all the necessary competences (also as partners or consultants)
v41	Suitability of the management system of the project and of the partnership
v42	Suitability of the length with respect to the objectives
v43	Suitability of the financial resources

Scores were expressed on a scale from 0 to 5, where the values have the following meaning:

- 0 = unacceptable;
- 1 = seriously insufficient;
- 2 = insufficient;
- 3 = sufficient;
- 4 = good;
- 5 = optimum.

It was allowed to assign intermediate scores, so the scale resulted formed by 11 values: 0, 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5. Scores were given independently by at least two evaluators for each project; one of them has the role of coordinator, the other is the support.

3.3 Preliminary analysis of the judgements received from a subset of projects

From the 123 projects to be evaluated, 89 projects have been selected for the application of the *multifacet* model. For each project, the scores given by the judges (they were 44) were collected and transformed on a scale from 0 to 10. Generally, all the evaluators gave a score to all the criteria and missing data are less than 2,5%, with the exception of criteria 1.5 that presents 6.1% of missing data. Anyway, this do not constitute a problem for the estimation of the Rasch model, provided that an evaluation in any criteria is available for each project.

Figure 1 represents the average, minimum and maximum score obtained by each project in all the criteria from all the judges. The figure shows important divergences between the evaluators that, in some cases, goes from 6 to 9.

Fig. 1 - Minimum, maximum and average rough score received by each project

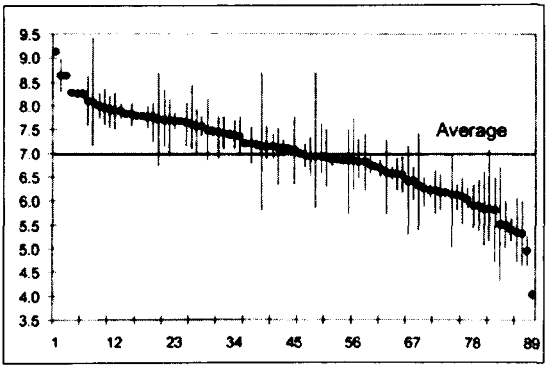
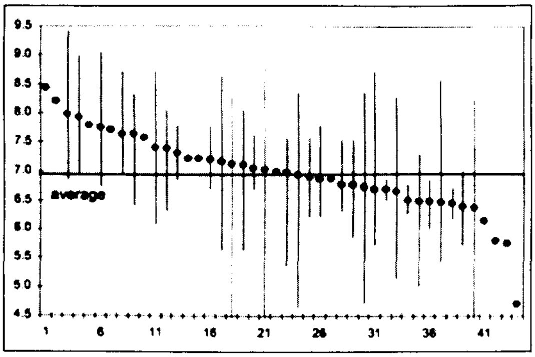


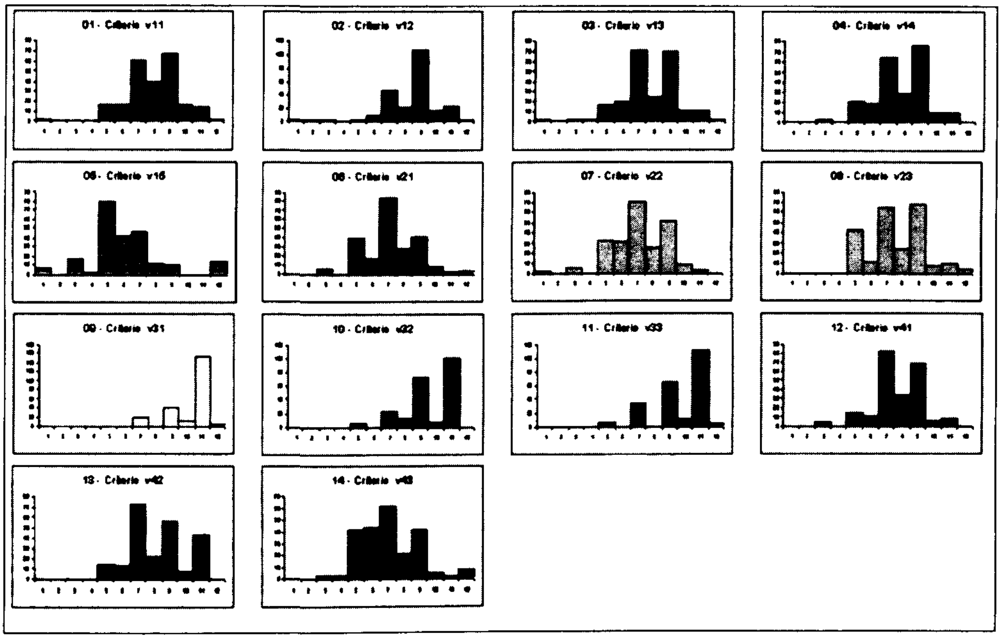
Figure 2 represents, instead, the minimum, maximum and average score given by each judge: in the hypothesis that projects were randomly assigned to the evaluators, the figure shows a large difference in severity between judges. These discrepancies could favour or not some project, and this will be further showed by the application of the *multifacet* model.

Fig. 2 - Minimum, maximum and average rough score assigned by each evaluator



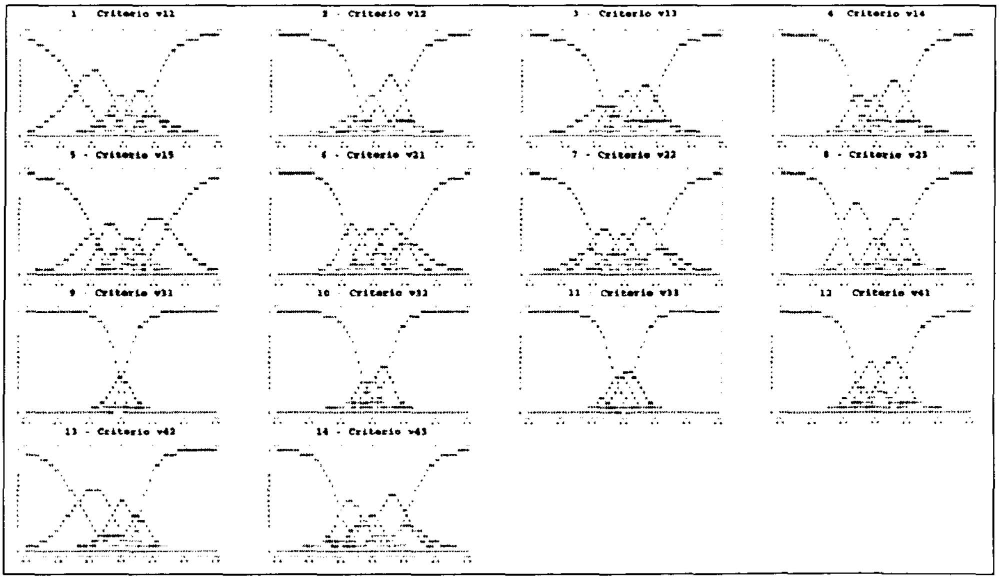
Finally, Figure 3, that represents scores distribution (from 0 to 10 + the missing data) for the different criteria, shows that not all the values were utilized and that a scale with 2 or 3 values would frequently be sufficient. This reduces the errors that judges commit in giving evaluations.

Fig. 3 - Use of the 11 values for the evaluation of the criteria



On the basis of this representation and of a first application of the model, a transformation on a new scale results appropriate. The model that was fitted is the partial credit model (8), estimated with the program FACETS (Linacre, 1998). Figure 4 represents the probabilities of receiving each of the 11 values on the basis of the model estimates. The large overlap of the curves indicates that the use of a big number of values increases the presence of errors in the evaluation process. This is confirmed by the indexes (Figure 7) for the interactions between evaluators and criteria and between projects and criteria and by the presence of a large number of misfitting indexes (Figure 6) for all the aspects (evaluators, criteria and projects).

Fig. 4 - Probability curves of receiving a score using the original scale on 11 values (*partial credit* model (8): every criteria has different thresholds)



A transformation on a different scale, with less values, was individuated through various and successive attempts, in order to attain a better separation of the probability curves and a better fit of the model. Figure 5 reports the final choice, that uses 3 values (like insufficient, sufficient, good) for all the criteria with the exception for criteria 9 (that is v31) that was expressed on a binary scale (unsuitable, suitable). The figure shows that curves present now a good separation, reducing in this way the errors that judges committed assigning scores.

The final model presents different thresholds for different groups of criteria, leading to a special case of model (8). The groups are represented with different colours in Figure 3 and the transformation applied to the scores is reported in Figure 5.

The goodness of fit results quite improved. Figure 6 compares the infit and outfit indexes for the three aspects (evaluators, projects and criteria) before and after the transformation:

- *Evaluators:* using a scale with 11 values, 11 infit indexes and 9 outfit indexes were out the limits; after the transformation on a scale with 3 values these are respectively 6 and 5;
- *Projects:* before the transformation there were 15 infit indexes and 14 outfit indexes out of the limits; after the transformation these are 9 and 10;
- *Criteria:* before the transformation there were 2 infit indexes and 3 outfit indexes out of the limits, after transformation there these are both 0.

Figure 7 represents the bias due to interactions between evaluators and projects, between evaluators and criteria, and between projects and criteria, before (left side) and after (right side) the transformation. The figure shows that after the transformation there are less interactions that result significative.

The final model presents good reliability indexes, that are equal to 83% for the evaluators, to 92% for projects (that is the measure of major interest) and to 97% for criteria.

Fig. 5 - Probability curves of receiving a score after the scale transformation (special case of model (8) with threshold constant within groups of criteria)

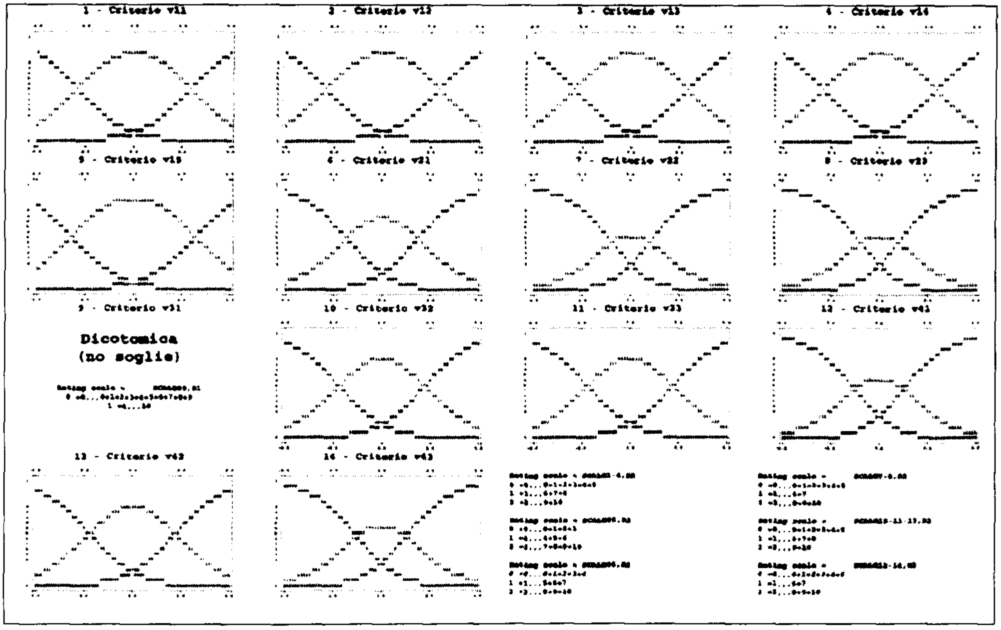


Fig. 6 - Standardized Infit and Outfit indexes before and after the scale transformation

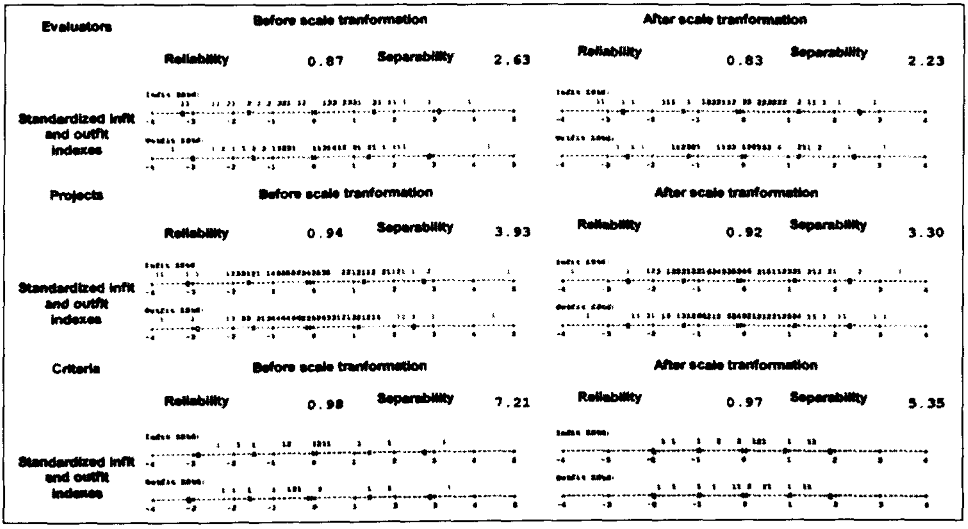
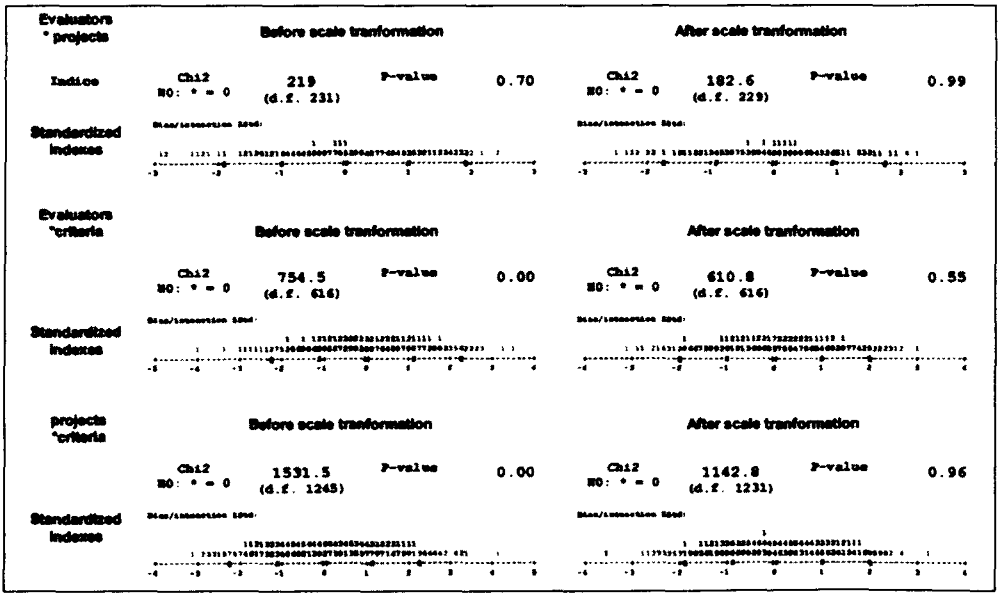


Fig. 7 - Indexes for detecting interaction bias, before and after the transformation of the scale



Evaluators * projects	Before scale tranformation				After scale tranformation			
Indice	Chi2 H0: * = 0	219 (d.f. 231)	P-value	0.70	Chi2 H0: * = 0	182.6 (d.f. 229)	P-value	0.99
Standardized indexes	Bias/interaction 1Std:				Bias/interaction 1Std:			
Evaluators *criteria	Before scale tranformation				After scale tranformation			
	Chi2 H0: * = 0	754.5 (d.f. 616)	P-value	0.00	Chi2 H0: * = 0	610.8 (d.f. 616)	P-value	0.55
Standardized indexes	Bias/interaction 1Std:				Bias/interaction 1Std:			
projects *criteria	Before scale tranformation				After scale tranformation			
	Chi2 H0: * = 0	1531.5 (d.f. 1245)	P-value	0.00	Chi2 H0: * = 0	1142.8 (d.f. 1231)	P-value	0.96
Standardized indexes	Bias/interaction 1Std:				Bias/interaction 1Std:			

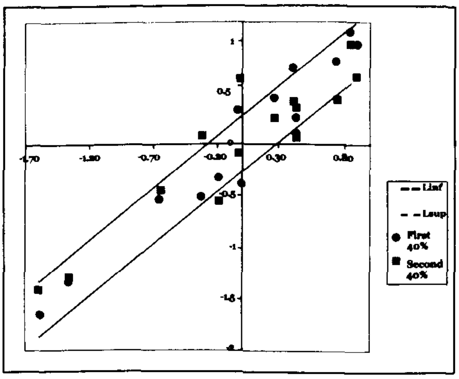
In order to further evaluate the goodness of fit of the model about the stability of the estimates of the criteria difficulties, the model was applied separately to the first 40% and the last 40% of the projects (ranked on the basis of their goodness). Figure 8 represents the comparison of the difficulties estimated on all the projects with those estimates on the two groups and shows a good similarity between them, with the exception of only one criteria (the 13th, that is v42).

Fig. 8 - Estimates of the criteria difficulties on all the projects and on the fist and last 40%

estimates

	Difficulties estimates			Standard errors			Test for the difference with the total	
Criteria	Total	First 40%	Second 40%	S.E.	S.E.1	S.E.2	z test first	z test second
1 - v11	0.43	0.13	0.09	0.15	0.23	0.26	-1.09	-1.13
2 - v12	-0.17	-0.32	-0.51	0.15	0.23	0.23	-0.55	-1.24
3 - v13	0.76	0.80	0.44	0.15	0.22	0.27	0.15	-1.04
4 - v14	0.91	0.98	0.66	0.15	0.22	0.28	0.26	-0.79
5 - v15	0.43	0.24	0.34	0.16	0.26	0.29	-0.62	-0.27
6 - v21	0.27	0.44	0.24	0.12	0.2	0.2	0.73	-0.13
7 - v22	0.42	0.73	0.39	0.11	0.19	0.17	1.41	-0.15
8 - v23	-0.01	0.32	-0.10	0.11	0.18	0.18	1.56	-0.43
9 - v31	-0.64	-0.54	-0.46	0.16	0.23	0.28	0.36	0.56
10- v32	-1.35	-1.32	-1.28	0.13	0.2	0.22	0.13	0.27
11- v33	-1.60	-1.65	-1.40	0.14	0.2	0.22	-0.20	0.77
12 - v41	-0.31	-0.52	0.07	0.11	0.17	0.18	-1.04	1.80
13 - v42	0.00	-0.39	0.54	0.13	0.2	0.2	-1.63	2.26
14 - v43	0.87	1.10	0.98	0.11	0.21	0.17	0.97	0.54

Estimates on the first and the second 40% and confidence intervals on the total



Data available for the analysis, after the trasformation of the scale, present an optimal fit to the Rasch model, that can be then utilized to obtain estimates of projects goodness, judges severities, and criteria difficulties.

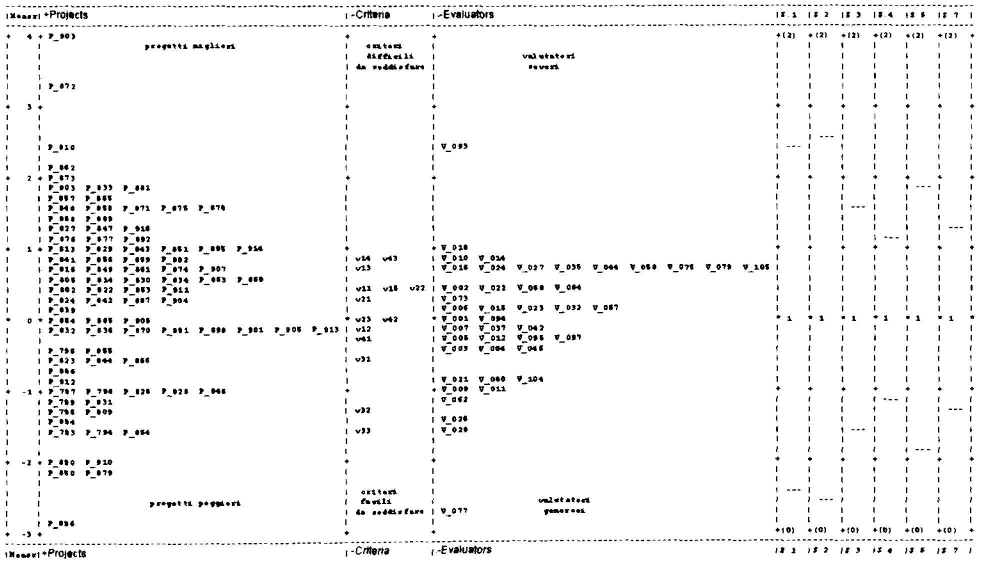
3.4 Estimation results of the multifacet model

The model utilized in the following is a special case of model (8), with thresholds kept constant within groups of criteria.

The model presents a good general fit, and the *Data log-likelihood chi-square* index is 4.45 with 3163 responses. An empirical rule, based on this index, for establishing the goodness of fit (Linacre, 1998) is the following: when the index is grater than [number of responses + 4*square root (number of responses)], then there is a bad general fit (significance level 5%). In this case, instead, 4.45 is largely smaller than such critical threshold.

Figure 9 represents some synthetic results of the estimation. On the left side there is the Rasch scale (from -3 to +4); then there are the projects collocated according to their goodness (the worse is project 896, the best is project 903); then there are the criteria (the most difficult are v14 and v43, the easiest is v33); on the next column there are the evaluators (the most severe is 93 and the most generous is 77); finally, on the right, there are the thresholds.

Fig. 9 - Map of the measures of the different aspects (Progects (P_), Criteria (v), Evaluators (V_)), and of the thresholds (S.)

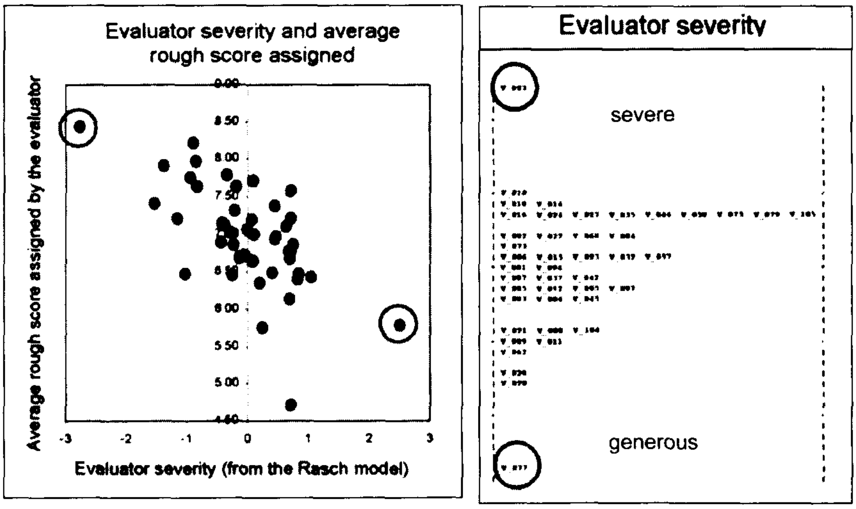


The appendix reports the estimated values of severities, goodness and difficulties. They all present acceptable infit and outfit indexes (between 0.8 and 1.2). The appendix reports also the thresholds estimates.

3.5 Bias induced by the use of rough scores

These analysis confirm the bias induced by the use of rough scores instead of the Rasch measures of projects goodness. Figure 10 shows that there is an inverse relation between the average score assigned by judges and their severity. The most severe is the 93 that gives on average scores between 5.5 and 6, while the most generous is the 77 who gives on average a score equal to 8.5.

Fig. 10 - Judges severity and scores assigned



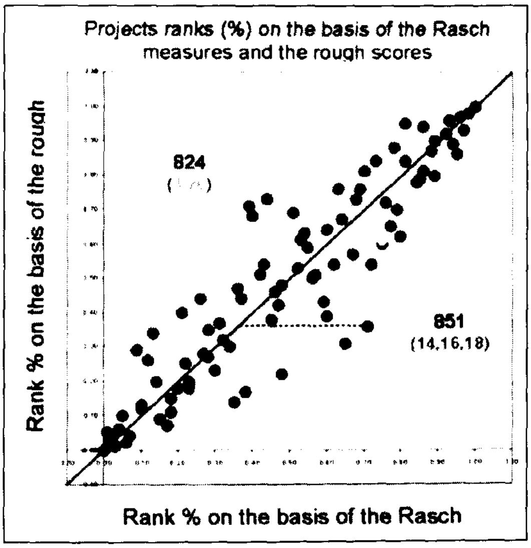
On the basis of the rough scores and the Rasch measures two different ranks for each project were obtained. This ranks were expressed on a percentual scale and the differences between the two percentual ranks for each project were computed. Table 4 reports the frequencies of such differences (grouped in classes). The 29% of the projects presents a gap grater than 10%, that means that, using the rough scores, 29 projects (on 100) unfairly overcome al least 10 projects (on 100).

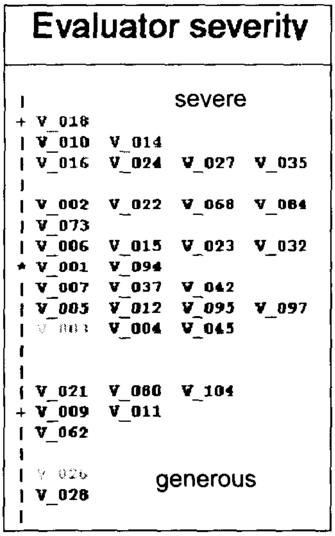
Tab. 4 - Discrepancies in the percentual rank of the projects on the basis of the rough scores and the Rasch measures

rank (%) difference	frequencies	decumulate
=0	7	1 . 00
0 . 00 < x < = 0 . 05	32	0 . 92
0 . 05 < x < = 0 . 10	24	0 . 56
0 . 10 < x < = 0 . 15	9	0 . 29
0 . 15 < x < = 0 . 20	7	0 . 19
0 . 20 < x < = 0 . 25	4	0 . 11
0 . 25 < x < = 0 . 30	3	0 . 07
0 . 30 < x < = 0 . 35	3	0 . 03
Total	**89**	

Figure 11 represents the plot of the percentual ranks on the projects in the two scales (rough scores and Rasch measures). Project 824 has a very good position (over 70%) when evaluated using the rough scores, but it has a percentual rank smaller than 40% using the Rasch measure. On the contrary, project 851 presents a percentual rank over 70% on the Rasch scale, but it is very penalized on the scores scale (about 35%). These discrepancies are due to the evaluators that judged the projects: the first one was evaluated by judges 3 and 26 that, on the severity scale, are collocated in the lower part (they are then more generous), while the second was evaluated by judges 14, 16 and 18 that are collocated in the upper part of the scale.

Fig. 11 - *Judges severity effect on the percentual ranks discrepancies*

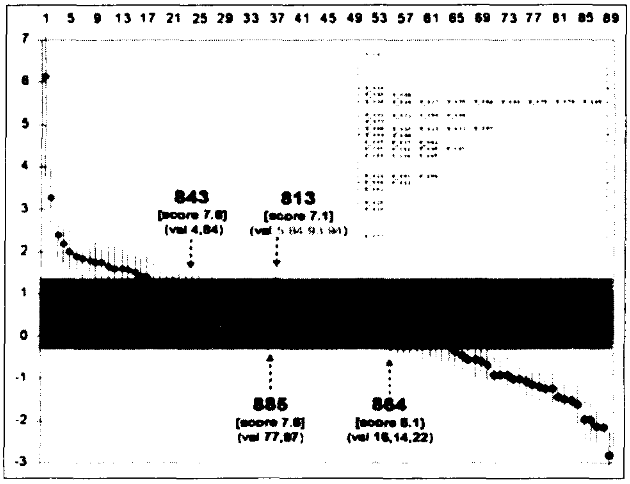



Evaluator severity

```
|              severe
+ V_018
| V_010  V_014
| V_016  V_024  V_027  V_035
|
| V_002  V_022  V_068  V_084
| V_073
| V_006  V_015  V_023  V_032
* V_001  V_094
| V_007  V_037  V_042
| V_005  V_012  V_095  V_097
| V_003  V_004  V_045
|
|
| V_021  V_080  V_104
+ V_009  V_011
| V_062
|
| V_026         generous
| V_028
|
```

Finally, considering that only 36 projects on 123 were financed, and assuming that the same proportion of financed ones is present between the 89 projects considered here, we can imaging that 26 project would be financed. Figure 12 represents the group of the 26 projects that are financed on the basis of the rough scores (on the left) and the 63 ones that are excluded (on the right). Inside the two groups the ranking is given be the Rasch measures, that are represented with the 95% confidence interval. The figure shows that there is an important overlapping zone (that we call litigation zone) between the financed projects and the excluded ones. In particular, the best of the excluded projects, the 813, presents the same goodness of the 843 (that was financed on the basis of the rough scores). This is due to the fact that the first was evaluated by severe judges, while the second by more generous evaluators. Similarly, the worst of the financed projects, the 885, presents the same goodness of the 864, that is excluded. The judges who evaluate the first one were more generous than the evaluators of the second one.

Fig. 12 - Analysis of the collocation of the projects that would be financed on the basis of the rough scores. Projects are divided in two groups and ordered on the basis of the Rasch measures



4. CONCLUSIONS

The analysis presented in this paper shows that the data concerning projects evaluations, in the particular contest considered, can be adequately utilized for constructing objective measures by means of a special case of the multifacet model (8), with thresholds kept constant within groups of criteria. The only condition required is the use of a transformation of the original scale on which the judge evaluations are expressed in a scale with only 3 values, that allows for a reduction of the errors produced by the large number of values than can be assigned.

The model so obtained presents good infit and outfit indexes, a good general fit and an optimal reliability of the estimates of the three aspects. It is not present bias due to interaction between them (see § 2.4), and in particular that between evaluators and projects which is the most preoccupant as it may hidden favouritisms.

The criteria seem well structured for constructing the scale, and successive analysis with new data could confirm their validity if the difficulties remain constant over time and space.

Only a couple of judges present extreme severity parameters and, probably, in the future they could be excluded or adequately trained, together with the judges who present misfitting indexes.

The use of the rough score, compared with the results what would produce the Rasch measures, highlights the serious alterations of the projects ranking and motivates a litigation that, in this moment, is excluded only because the rough scores and the methods utilized to obtain them have all the legal requirements, but they are certainly not scientific and objective measures that would be necessary in this contest.

APPENDIX

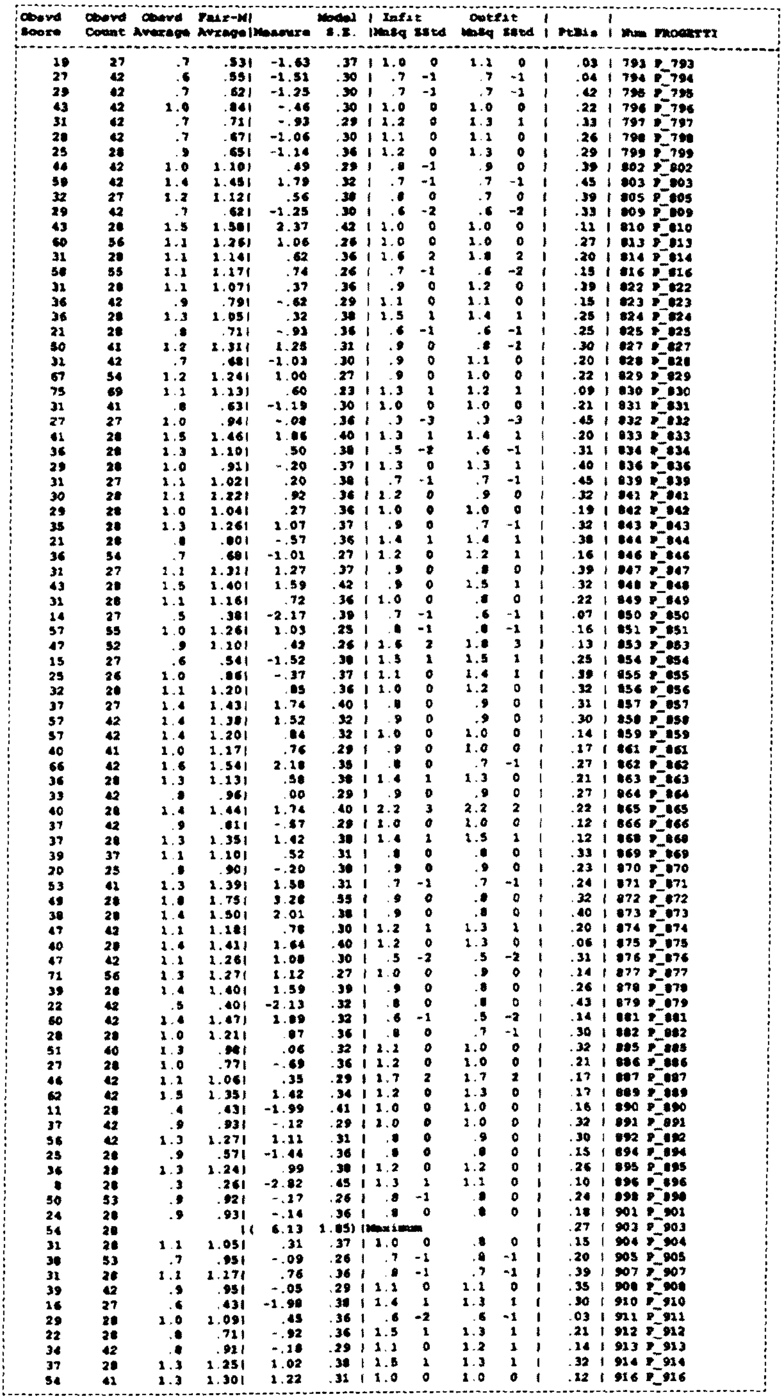
Tab. A1 - Estimates of evaluators severity (Υ_j = measure)
(ordered according to the identification number)

```
---------------------------------------------------------------------------------------------------------------
| Obsvd   Obsvd  Obsvd  Fair-M|          Model | Infit       Outfit    |        | Exact Agree. |               |
| Score   Count Average Avrage|Measure   S.E.  |MnSq ZStd  MnSq ZStd | PtBis  | Obs %  Exp % | Num VALUTATORI |
---------------------------------------------------------------------------------------------------------------
|   289    294    1.0   1.05|   -.03    .11 | 1.0    0    1.0    0  |   .36  |  59.6   54.9 |   1 V_001     |
|   293    291    1.0    .92|    .45    .11 |  .9   -1     .9   -1  |   .41  |  63.7   54.9 |   2 V_002     |
|   330    292    1.1   1.15|   -.38    .12 | 1.1    1    1.1    1  |   .36  |  67.4   55.7 |   3 V_003     |
|   214    194    1.1   1.16|   -.42    .14 |  .9    0     .9   -1  |   .40  |  81.4   55.9 |   4 V_004     |
|   233    222    1.0   1.12|   -.28    .13 | 1.2    1    1.1    1  |   .38  |  58.4   54.0 |   5 V_005     |
|    44     42    1.0   1.02|    .07    .29 |  .9    0     .9    0  |   .32  |  83.9   55.5 |   6 V_006     |
|    53     42    1.3   1.09|   -.19    .31 | 1.0    0     .9    0  |   .22  |  44.0   56.7 |   7 V_007     |
|    81     69    1.2   1.30|   -.95    .24 | 1.1    0    1.4    1  |   .42  |  72.6   54.1 |   9 V_009     |
|    55     67     .8    .81|    .83    .23 |  .9    0     .9    0  |   .24  |  65.5   51.5 |  10 V_010     |
|    43     51     .8   1.32|  -1.01    .26 |  .7   -1     .8   -1  |   .18  |  50.0   48.0 |  11 V_011     |
|    19     14    1.4   1.14|   -.34    .54 |  .7    0    1.0    0  |   .31  |  81.0   56.7 |  12 V_012     |
|    84     95     .9    .81|    .85    .19 | 1.2    1    1.1    0  |   .25  |  58.5   54.4 |  14 V_014     |
|    73     84     .9    .98|    .21    .23 | 1.1    0    1.1    0  |   .53  |  61.9   57.2 |  15 V_015     |
|    85     94     .9    .84|    .71    .20 |  .7   -2     .8   -1  |   .31  |  67.4   54.6 |  16 V_016     |
|    22     27     .8    .75|   1.05    .37 |  .9    0     .9    0  |   .19  |  73.1   54.7 |  18 V_018     |
|    69     54    1.3   1.27|   -.83    .28 | 1.1    0    1.0    0  |   .28  |  58.7   56.1 |  21 V_021     |
|    38     40    1.0    .93|    .42    .30 | 1.1    0    1.2    0  |   .12  |  63.5   54.6 |  22 V_022     |
|    43     42    1.0   1.01|    .10    .30 |  .9    0    1.0    0  |   .41  |  69.6   54.3 |  23 V_023     |
|    18     14    1.3    .84|    .72    .53 |  .6   -1     .5   -1  |   .12  |  78.6   56.4 |  24 V_024     |
|   119     84    1.4   1.41|  -1.38    .24 | 1.1    0     .9    0  |   .37  |  61.9   54.3 |  26 V_026     |
|    93     93    1.0    .86|    .68    .20 |  .9    0     .8   -1  |   .10  |  63.6   54.0 |  27 V_027     |
|    66     56    1.2   1.45|  -1.53    .27 | 1.1    0    1.3    1  |   .29  |  57.1   51.4 |  28 V_028     |
|    17     14    1.2   1.01|    .10    .52 | 1.0    0    1.4    0  |   .17  |  50.0   56.5 |  32 V_032     |
|    15     14    1.1    .84|    .71    .51 |  .2   -3     .2   -2  |   .32  |  78.6   56.4 |  35 V_035     |
|    34     28    1.2   1.10|   -.28    .37 | 2.0    2    2.3    3  |   .14  |  35.7   55.3 |  37 V_037     |
|    81     84    1.0   1.08|   -.14    .21 |  .9    0     .8   -1  |   .43  |  86.4   56.2 |  42 V_042     |
|   101     98    1.0    .86|    .65    .20 |  .9    0     .8   -1  |   .40  |  53.4   54.2 |  44 V_044     |
|    31     28    1.1   1.16|   -.44    .36 | 1.1    0    1.1    0  |   .13  |  61.4   54.2 |  45 V_045     |
|    15     14    1.1   1.01|    .11    .51 |  .5   -1     .5   -1  |   .37  |  75.0   56.0 |  57 V_057     |
|    28     28    1.0    .85|    .69    .36 | 1.0    0    1.1    0  |   .20  |  53.7   55.6 |  50 V_050     |
|    14     14    1.0   1.35|  -1.15    .50 | 1.2    0    1.7    1  |   .15  |  39.3   47.1 |  62 V_062     |
|    88     83    1.1    .91|    .40    .21 | 1.0    0    1.1    0  |   .28  |  57.6   55.8 |  68 V_068     |
|    23     28     .8    .97|    .24    .36 | 1.6    2    1.8    2  |   .22  |  33.9   53.9 |  73 V_073     |
|    17     28     .6    .85|    .70    .37 | 1.0    0    1.0    0  |   .18  |  66.1   51.7 |  75 V_075     |
|    23     14    1.6   1.71|  -2.76    .64 | 1.4    0    1.1    0  |   .17  |  38.5   42.0 |  77 V_077     |
|     7     14     .5    .84|    .72    .54 | 1.6    1    1.7    1  |  -.08  |  42.5   58.6 |  79 V_079     |
|    57     42    1.4   1.27|   -.84    .32 | 1.0    0     .7   -1  |   .39  |  50.0   51.7 |  80 V_080     |
|   111     96    1.2    .91|    .45    .20 | 1.1    0    1.1    0  |   .21  |  68.5   54.7 |  84 V_084     |
|     8     14     .6    .30|   2.58    .53 |  .9    0     .9    0  |   .07  |  45.2   37.5 |  93 V_093     |
|    98     93    1.1   1.05|    .00    .20 |  .6   -3     .6   -2  |   .19  |  59.9   53.1 |  94 V_094     |
|    81     84    1.0   1.11|   -.24    .22 |  .7   -2     .7   -2  |   .42  |  83.9   56.2 |  95 V_095     |
|    54     56    1.0   1.10|   -.22    .26 | 1.2    0    1.2    0  |   .37  |  71.9   52.3 |  97 V_097     |
|    21     14    1.5   1.28|   -.88    .50 | 1.2    0    1.3    0  |  -.03  |  70.4   53.5 | 104 V_104     |
|    15     14    1.1    .83|    .75    .51 | 1.5    1    1.1    0  |  -.05  |  76.2   53.9 | 105 V_105     |
---------------------------------------------------------------------------------------------------------------
| Obsvd   Obsvd  Obsvd  Fair-M|          Model | Infit       Outfit    |        | Exact Agree. |               |
| Score   Count Average Avrage|Measure   S.E.  |MnSq ZStd  MnSq ZStd | PtBis  | Obs %  Exp % | Num VALUTATORI |
---------------------------------------------------------------------------------------------------------------
|  75.1   71.9    1.1   1.04|    .00    .32 | 1.0   -.1  1.0    .0|    .25  |              | Mean (Count: 44) |
|  78.0   74.1     .2    .23|    .86    .14 |  .3   1.3   .4   1.2|    .14  |              | S.D.          |
---------------------------------------------------------------------------------------------------------------
           RMSE (Model)  .35 Adj S.D.   .78  Separation  2.23  Reliability  .83
           Fixed (all same) chi-square: 310.1  d.f.: 43  significance: .00
     Rater agreement opportunities: 2810  Exact agreements: 1795 = 63.9%  Expected: 1527.6 = 54.4%
---------------------------------------------------------------------------------------------------------------
```

Tab. A2 - Estimates of the projects goodness (θ_n = measure)
(ordered according to the goodness)

```
-----------------------------------------------------------------------------------------------
| Obsvd   Obsvd  Obsvd   Fair-M|         Model | Infit      Outfit    |       |               |
| Score   Count Average Avrage|Measure  S.E.  |MnSq ZStd  MnSq ZStd  | PtBis | Num PROGETTI  |
-----------------------------------------------------------------------------------------------
|   54     28                 |(  6.13  1.85) |Maximum               |   .27 | 903 P_903     |
|   49     28     1.8    1.75|   3.28    .55  |  .9   0     .8   0   |   .32 | 872 P_872     |
|   43     28     1.5    1.58|   2.37    .42  | 1.0   0    1.0   0   |   .11 | 810 P_810     |
|   66     42     1.6    1.54|   2.18    .35  |  .8   0     .7  -1   |   .27 | 862 P_862     |
|   38     28     1.4    1.50|   2.01    .38  |  .9   0     .8   0   |   .40 | 873 P_873     |
|   60     42     1.4    1.47|   1.89    .32  |  .6  -1     .5  -2   |   .14 | 881 P_881     |
|   41     28     1.5    1.46|   1.86    .40  | 1.3   1    1.4   1   |   .20 | 833 P_833     |
|   59     42     1.4    1.45|   1.79    .32  |  .7  -1     .7  -1   |   .45 | 803 P_803     |
|   40     28     1.4    1.44|   1.74    .40  | 2.2   3    2.2   2   |   .22 | 865 P_865     |
|   37     27     1.4    1.43|   1.74    .40  |  .8   0     .9   0   |   .31 | 857 P_857     |
|   40     28     1.4    1.41|   1.64    .40  | 1.2   0    1.3   0   |   .06 | 875 P_875     |
|   43     28     1.5    1.40|   1.59    .42  |  .9   0    1.5   1   |   .32 | 848 P_848     |
|   39     28     1.4    1.40|   1.59    .39  |  .9   0     .8   0   |   .26 | 878 P_878     |
|   53     41     1.3    1.39|   1.58    .31  |  .7  -1     .7  -1   |   .24 | 871 P_871     |
|   57     42     1.4    1.38|   1.52    .32  |  .9   0     .9   0   |   .30 | 858 P_858     |
|   62     42     1.5    1.35|   1.42    .34  | 1.2   0    1.3   0   |   .17 | 889 P_889     |
|   37     28     1.3    1.35|   1.42    .38  | 1.4   1    1.5   1   |   .12 | 868 P_868     |
|   31     27     1.1    1.31|   1.27    .37  |  .9   0     .8   0   |   .39 | 847 P_847     |
|   50     41     1.2    1.31|   1.25    .31  |  .9   0     .8  -1   |   .30 | 827 P_827     |
|   54     41     1.3    1.30|   1.22    .31  | 1.0   0    1.0   0   |   .12 | 916 P_916     |
|   71     56     1.3    1.27|   1.12    .27  | 1.0   0     .9   0   |   .14 | 877 P_877     |
|   56     42     1.3    1.27|   1.11    .31  |  .8   0     .9   0   |   .30 | 892 P_892     |
|   47     42     1.1    1.26|   1.08    .30  |  .5  -2     .5  -2   |   .31 | 876 P_876     |
|   35     28     1.3    1.26|   1.07    .37  |  .9   0     .7  -1   |   .32 | 843 P_843     |
|   60     56     1.1    1.26|   1.06    .26  | 1.0   0    1.0   0   |   .27 | 813 P_813     |
|   57     55     1.0    1.26|   1.03    .25  |  .8  -1     .8  -1   |   .16 | 851 P_851     |
|   37     28     1.3    1.25|   1.02    .38  | 1.5   1    1.3   1   |   .32 | 914 P_914     |
|   67     54     1.2    1.24|   1.00    .27  |  .9   0    1.0   0   |   .22 | 829 P_829     |
|   36     28     1.3    1.24|    .99    .38  | 1.2   0    1.2   0   |   .26 | 895 P_895     |
|   30     28     1.1    1.22|    .92    .36  | 1.2   0     .9   0   |   .32 | 841 P_841     |
|   28     28     1.0    1.21|    .87    .36  |  .8   0     .7  -1   |   .30 | 882 P_882     |
|   32     28     1.1    1.20|    .85    .36  | 1.0   0    1.2   0   |   .32 | 856 P_856     |
|   57     42     1.4    1.20|    .84    .32  | 1.0   0    1.0   0   |   .14 | 859 P_859     |
|   47     42     1.1    1.18|    .78    .30  | 1.2   1    1.3   1   |   .20 | 874 P_874     |
|   40     41     1.0    1.17|    .76    .29  |  .9   0    1.0   0   |   .17 | 861 P_861     |
|   31     28     1.1    1.17|    .76    .36  |  .8  -1     .7  -1   |   .39 | 907 P_907     |
|   58     55     1.1    1.17|    .74    .26  |  .7  -1     .6  -2   |   .15 | 816 P_816     |
|   31     28     1.1    1.16|    .72    .36  | 1.0   0     .8   0   |   .22 | 849 P_849     |
|   31     28     1.1    1.14|    .62    .36  | 1.6   2    1.8   2   |   .20 | 814 P_814     |
|   75     69     1.1    1.13|    .60    .23  | 1.3   1    1.2   1   |   .09 | 830 P_830     |
|   36     28     1.3    1.13|    .58    .38  | 1.4   1    1.3   0   |   .21 | 863 P_863     |
|   32     27     1.2    1.12|    .56    .38  |  .8   0     .7   0   |   .39 | 805 P_805     |
|   39     37     1.1    1.10|    .52    .31  |  .8   0     .8   0   |   .33 | 869 P_869     |
|   36     28     1.3    1.10|    .50    .38  |  .5  -2     .6  -1   |   .31 | 834 P_834     |
|   44     42     1.0    1.10|    .49    .29  |  .8  -1     .9   0   |   .39 | 802 P_802     |
|   29     28     1.0    1.09|    .45    .36  |  .6  -2     .6  -1   |   .03 | 911 P_911     |
|   47     52      .9    1.10|    .42    .26  | 1.6   2    1.8   3   |   .13 | 853 P_853     |
|   31     28     1.1    1.07|    .37    .36  |  .9   0    1.2   0   |   .39 | 822 P_822     |
|   46     42     1.1    1.06|    .35    .29  | 1.7   2    1.7   2   |   .17 | 887 P_887     |
|   36     28     1.3    1.05|    .32    .38  | 1.5   1    1.4   1   |   .25 | 824 P_824     |
|   31     28     1.1    1.05|    .31    .37  | 1.0   0     .8   0   |   .15 | 904 P_904     |
|   29     28     1.0    1.04|    .27    .36  | 1.0   0    1.0   0   |   .19 | 842 P_842     |
|   31     27     1.1    1.02|    .20    .38  |  .7  -1     .7  -1   |   .45 | 839 P_839     |
|   51     40     1.3     .98|    .06    .32  | 1.1   0    1.0   0   |   .32 | 885 P_885     |
|   33     42      .8     .96|    .00    .29  |  .9   0     .9   0   |   .27 | 864 P_864     |
|   39     42      .9     .95|   -.05    .29  | 1.1   0    1.1   0   |   .35 | 908 P_908     |
|   27     27     1.0     .94|   -.08    .36  |  .3  -3     .3  -3   |   .45 | 832 P_832     |
|   38     53      .7     .95|   -.09    .26  |  .7  -1     .8  -1   |   .20 | 905 P_905     |
|   37     42      .9     .93|   -.12    .29  | 1.0   0    1.0   0   |   .32 | 891 P_891     |
|   24     28      .9     .93|   -.14    .36  |  .8   0     .8   0   |   .18 | 901 P_901     |
|   50     53      .9     .92|   -.17    .26  |  .8  -1     .8   0   |   .24 | 898 P_898     |
|   34     42      .8     .91|   -.18    .29  | 1.1   0    1.2   1   |   .14 | 913 P_913     |
|   20     25      .8     .90|   -.20    .38  |  .9   0     .9   0   |   .23 | 870 P_870     |
|   29     28     1.0     .91|   -.20    .37  | 1.3   0    1.3   1   |   .40 | 836 P_836     |
|   25     26     1.0     .86|   -.37    .37  | 1.1   0    1.4   1   |   .39 | 855 P_855     |
|   43     42     1.0     .84|   -.46    .30  | 1.0   0    1.0   0   |   .22 | 796 P_796     |
|   37     42      .9     .81|   -.57    .29  | 1.0   0    1.0   0   |   .12 | 866 P_866     |
|   21     28      .8     .80|   -.57    .36  | 1.4   1    1.4   1   |   .38 | 844 P_844     |
|   36     42      .9     .79|   -.62    .29  | 1.1   0    1.1   0   |   .15 | 823 P_823     |
|   27     28     1.0     .77|   -.69    .36  | 1.2   0    1.0   0   |   .21 | 886 P_886     |
|   22     28      .8     .71|   -.92    .36  | 1.5   1    1.3   1   |   .21 | 912 P_912     |
|   21     28      .8     .71|   -.93    .36  |  .6  -1     .6  -1   |   .25 | 825 P_825     |
|   31     42      .7     .71|   -.93    .29  | 1.2   0    1.3   1   |   .33 | 797 P_797     |
|   36     54      .7     .68|  -1.01    .27  | 1.2   0    1.2   1   |   .16 | 846 P_846     |
|   31     42      .7     .68|  -1.03    .30  |  .9   0    1.1   0   |   .20 | 828 P_828     |
|   28     42      .7     .67|  -1.06    .30  | 1.1   0    1.1   0   |   .26 | 798 P_798     |
|   25     28      .9     .65|  -1.14    .36  | 1.2   0    1.3   0   |   .29 | 799 P_799     |
|   31     41      .8     .63|  -1.19    .30  | 1.0   0    1.0   0   |   .21 | 831 P_831     |
|   29     42      .7     .62|  -1.25    .30  |  .7  -1     .7  -1   |   .42 | 795 P_795     |
|   29     42      .7     .62|  -1.25    .30  |  .6  -2     .6  -2   |   .33 | 809 P_809     |
|   25     28      .9     .57|  -1.44    .36  |  .8   0     .8   0   |   .15 | 894 P_894     |
|   27     42      .6     .55|  -1.51    .30  |  .7  -1     .7  -1   |   .04 | 794 P_794     |
|   15     27      .6     .54|  -1.52    .38  | 1.5   1    1.5   1   |   .25 | 854 P_854     |
|   19     27      .7     .53|  -1.63    .37  | 1.0   0    1.1   0   |   .03 | 793 P_793     |
|   16     27      .6     .43|  -1.98    .38  | 1.4   1    1.3   1   |   .30 | 910 P_910     |
|   11     28      .4     .43|  -1.99    .41  | 1.0   0    1.0   0   |   .16 | 890 P_890     |
|   22     42      .5     .40|  -2.13    .32  |  .8   0     .8   0   |   .43 | 879 P_879     |
|   14     27      .5     .38|  -2.17    .39  |  .7  -1     .6  -1   |   .07 | 850 P_850     |
|    8     28      .3     .26|  -2.82    .45  | 1.3   1    1.1   0   |   .10 | 896 P_896     |
-----------------------------------------------------------------------------------------------
```

Tab. A3 - Estimates of the projects goodness
(ordered according to the identification number)



```
---------------------------------------------------------------------------------------
|Obsvd   Obsvd  Obsvd  Fair-M|         Model | Infit       Outfit   |       |
|Score   Count Average Avrage|Measure  S.E.  |MnSq ZStd  MnSq ZStd  | PtBis | Num PROGETTI
---------------------------------------------------------------------------------------
|   19     27      .7    .53|  -1.63    .37 | 1.0    0   1.1    0  |   .03 | 793 P_793
|   27     42      .6    .55|  -1.51    .30 |  .7   -1    .7   -1  |   .04 | 794 P_794
|   29     42      .7    .62|  -1.25    .30 |  .7   -1    .7   -1  |   .42 | 795 P_795
|   43     42     1.0    .84|   -.46    .30 | 1.0    0   1.0    0  |   .22 | 796 P_796
|   31     42      .7    .71|   -.93    .29 | 1.2    0   1.3    1  |   .33 | 797 P_797
|   28     42      .7    .67|  -1.06    .30 | 1.1    0   1.1    0  |   .26 | 798 P_798
|   25     28      .9    .65|  -1.14    .36 | 1.2    0   1.3    0  |   .29 | 799 P_799
|   44     42     1.0   1.10|    .49    .29 |  .8   -1    .9    0  |   .39 | 802 P_802
|   59     42     1.4   1.45|   1.79    .32 |  .7   -1    .7   -1  |   .45 | 803 P_803
|   32     27     1.2   1.12|    .56    .38 |  .8    0    .7    0  |   .39 | 805 P_805
|   29     42      .7    .62|  -1.25    .30 |  .6   -2    .6   -2  |   .33 | 809 P_809
|   43     28     1.5   1.58|   2.37    .42 | 1.0    0   1.0    0  |   .11 | 810 P_810
|   60     56     1.1   1.26|   1.06    .26 | 1.0    0   1.0    0  |   .27 | 813 P_813
|   31     28     1.1   1.14|    .62    .36 | 1.6    2   1.8    2  |   .20 | 814 P_814
|   58     55     1.1   1.17|    .74    .26 |  .7   -1    .6   -2  |   .15 | 816 P_816
|   31     28     1.1   1.07|    .37    .36 |  .9    0   1.2    0  |   .39 | 822 P_822
|   36     42      .9    .79|   -.62    .29 | 1.1    0   1.1    0  |   .15 | 823 P_823
|   36     28     1.3   1.05|    .32    .38 | 1.5    1   1.4    1  |   .25 | 824 P_824
|   21     28      .8    .71|   -.93    .36 |  .6   -1    .6   -1  |   .25 | 825 P_825
|   50     41     1.2   1.31|   1.28    .31 |  .9    0    .8   -1  |   .30 | 827 P_827
|   31     42      .7    .68|  -1.03    .30 |  .9    0   1.1    0  |   .20 | 828 P_828
|   67     54     1.2   1.24|   1.00    .27 |  .9    0   1.0    0  |   .22 | 829 P_829
|   75     69     1.1   1.13|    .60    .23 | 1.3    1   1.2    1  |   .09 | 830 P_830
|   31     41      .8    .63|  -1.19    .30 | 1.0    0   1.0    0  |   .21 | 831 P_831
|   27     27     1.0    .94|   -.08    .36 |  .3   -3    .3   -3  |   .45 | 832 P_832
|   41     28     1.5   1.46|   1.86    .40 | 1.3    1   1.4    1  |   .20 | 833 P_833
|   36     28     1.3   1.10|    .50    .38 |  .5   -2    .6   -1  |   .31 | 834 P_834
|   29     28     1.0    .91|   -.20    .37 | 1.3    0   1.3    1  |   .40 | 836 P_836
|   31     27     1.1   1.02|    .20    .38 |  .7   -1    .7   -1  |   .45 | 839 P_839
|   30     28     1.1   1.22|    .92    .36 | 1.2    0    .9    0  |   .32 | 841 P_841
|   29     28     1.0   1.04|    .27    .36 | 1.0    0   1.0    0  |   .19 | 842 P_842
|   35     28     1.3   1.26|   1.07    .37 |  .9    0    .7   -1  |   .32 | 843 P_843
|   21     28      .8    .80|   -.57    .36 | 1.4    1   1.4    1  |   .38 | 844 P_844
|   36     54      .7    .68|  -1.01    .27 | 1.2    0   1.2    1  |   .16 | 846 P_846
|   31     27     1.1   1.31|   1.27    .37 |  .9    0    .8    0  |   .39 | 847 P_847
|   43     28     1.5   1.40|   1.59    .42 |  .9    0   1.5    1  |   .32 | 848 P_848
|   31     28     1.1   1.16|    .72    .36 | 1.0    0    .8    0  |   .22 | 849 P_849
|   14     27      .5    .38|  -2.17    .39 |  .7   -1    .6   -1  |   .07 | 850 P_850
|   57     55     1.0   1.26|   1.03    .25 |  .8   -1    .8   -1  |   .16 | 851 P_851
|   47     52      .9   1.10|    .42    .26 | 1.6    2   1.8    3  |   .13 | 853 P_853
|   15     27      .6    .54|  -1.52    .38 | 1.5    1   1.5    1  |   .25 | 854 P_854
|   25     26     1.0    .86|   -.37    .37 | 1.1    0   1.4    1  |   .39 | 855 P_855
|   32     28     1.1   1.20|    .85    .36 | 1.0    0   1.2    0  |   .32 | 856 P_856
|   37     27     1.4   1.43|   1.74    .40 |  .8    0    .9    0  |   .31 | 857 P_857
|   57     42     1.4   1.38|   1.52    .32 |  .9    0    .9    0  |   .30 | 858 P_858
|   57     42     1.4   1.20|    .84    .32 | 1.0    0   1.0    0  |   .14 | 859 P_859
|   40     41     1.0   1.17|    .76    .29 |  .9    0   1.0    0  |   .17 | 861 P_861
|   66     42     1.6   1.54|   2.18    .35 |  .8    0    .7   -1  |   .27 | 862 P_862
|   36     28     1.3   1.13|    .58    .38 | 1.4    1   1.3    0  |   .21 | 863 P_863
|   33     42      .8    .96|    .00    .29 |  .9    0    .9    0  |   .27 | 864 P_864
|   40     28     1.4   1.44|   1.74    .40 | 2.2    3   2.2    2  |   .22 | 865 P_865
|   37     42      .9    .81|   -.57    .29 | 1.0    0   1.0    0  |   .12 | 866 P_866
|   37     28     1.3   1.35|   1.42    .38 | 1.4    1   1.5    1  |   .12 | 868 P_868
|   39     37     1.1   1.10|    .52    .31 |  .8    0    .8    0  |   .33 | 869 P_869
|   20     25      .8    .90|   -.20    .38 |  .9    0    .9    0  |   .23 | 870 P_870
|   53     41     1.3   1.39|   1.58    .31 |  .7   -1    .7   -1  |   .24 | 871 P_871
|   49     28     1.8   1.75|   3.28    .55 |  .9    0    .8    0  |   .32 | 872 P_872
|   38     28     1.4   1.50|   2.01    .38 |  .9    0    .8    0  |   .40 | 873 P_873
|   47     42     1.1   1.18|    .78    .30 | 1.2    1   1.3    1  |   .20 | 874 P_874
|   40     28     1.4   1.41|   1.64    .40 | 1.2    0   1.3    0  |   .06 | 875 P_875
|   47     42     1.1   1.26|   1.08    .30 |  .5   -2    .5   -2  |   .31 | 876 P_876
|   71     56     1.3   1.27|   1.12    .27 | 1.0    0    .9    0  |   .14 | 877 P_877
|   39     28     1.4   1.40|   1.59    .39 |  .9    0    .8    0  |   .26 | 878 P_878
|   22     42      .5    .40|  -2.13    .32 |  .8    0    .8    0  |   .43 | 879 P_879
|   60     42     1.4   1.47|   1.89    .32 |  .6   -1    .5   -2  |   .14 | 881 P_881
|   28     28     1.0   1.21|    .87    .36 |  .8    0    .7   -1  |   .30 | 882 P_882
|   51     40     1.3    .98|    .06    .32 | 1.1    0   1.0    0  |   .32 | 885 P_885
|   27     28     1.0    .77|   -.69    .36 | 1.2    0   1.0    0  |   .21 | 886 P_886
|   46     42     1.1   1.06|    .35    .29 | 1.7    2   1.7    2  |   .17 | 887 P_887
|   62     42     1.5   1.35|   1.42    .34 | 1.2    0   1.3    0  |   .17 | 889 P_889
|   11     28      .4    .43|  -1.99    .41 | 1.0    0   1.0    0  |   .16 | 890 P_890
|   37     42      .9    .93|   -.12    .29 | 1.0    0   1.0    0  |   .32 | 891 P_891
|   56     42     1.3   1.27|   1.11    .31 |  .8    0    .9    0  |   .30 | 892 P_892
|   25     28      .9    .57|  -1.44    .36 |  .8    0    .8    0  |   .15 | 894 P_894
|   36     28     1.3   1.24|    .99    .38 | 1.2    0   1.2    0  |   .26 | 895 P_895
|    8     28      .3    .26|  -2.82    .45 | 1.3    1   1.1    0  |   .10 | 896 P_896
|   50     53      .9    .92|   -.17    .26 |  .8   -1    .8    0  |   .24 | 898 P_898
|   24     28      .9    .93|   -.14    .36 |  .8    0    .8    0  |   .18 | 901 P_901
|   54     28              |(  6.13   1.85)|Maximum              |   .27 | 903 P_903
|   31     28     1.1   1.05|    .31    .37 | 1.0    0    .8    0  |   .15 | 904 P_904
|   38     53      .7    .95|   -.09    .26 |  .7   -1    .8   -1  |   .20 | 905 P_905
|   31     28     1.1   1.27|    .76    .36 |  .8   -1    .7   -1  |   .39 | 907 P_907
|   39     42      .9    .95|   -.05    .29 | 1.1    0   1.1    0  |   .35 | 908 P_908
|   16     27      .6    .43|  -1.98    .38 | 1.4    1   1.3    1  |   .30 | 910 P_910
|   29     28     1.0   1.09|    .45    .36 |  .6   -2    .6   -1  |   .03 | 911 P_911
|   22     28      .8    .71|   -.92    .36 | 1.5    1   1.3    1  |   .21 | 912 P_912
|   34     42      .8    .91|   -.18    .29 | 1.1    0   1.2    1  |   .14 | 913 P_913
|   37     28     1.3   1.25|   1.02    .38 | 1.5    1   1.3    1  |   .32 | 914 P_914
|   54     41     1.3   1.30|   1.22    .31 | 1.0    0   1.0    0  |   .12 | 916 P_916
---------------------------------------------------------------------------------------
```

Tab. A4 - Estimates of criteria difficulties (δ_i = measure)
(ordered according to the difficulty)

```
--------------------------------------------------------------------------------------------
| Obsvd   Obsvd  Obsvd  Fair-M|          Model | Infit      Outfit    |         |              |
| Score   Count Average Avrage|Measure    S.E. |MnSq ZStd  MnSq ZStd  | PtBis   | Nu CRITERI   |
--------------------------------------------------------------------------------------------
|    202    229      .9   .90|     .91    .15 |  .8   -1     .8   -1 |     .39 |  4 v14       |
|    177    222      .8   .75|     .87    .11 | 1.1    1    1.1    0 |     .39 | 14 v43       |
|    207    227      .9   .92|     .76    .15 |  .9    0     .9    0 |     .41 |  3 v13       |
|    211    215     1.0   .98|     .43    .16 | 1.0    0    1.1    0 |     .30 |  5 v15       |
|    223    228     1.0   .98|     .43    .15 | 1.1    0    1.0    0 |     .32 |  1 v11       |
|    218    229     1.0   .93|     .42    .11 | 1.0    0    1.0    0 |     .37 |  7 v22       |
|    227    225     1.0  1.00|     .27    .12 | 1.0    0    1.0    0 |     .35 |  6 v21       |
|    247    229     1.1  1.07|     .00    .13 | 1.1    1    1.1    1 |     .21 | 13 v42       |
|    252    225     1.1  1.14|    -.01    .11 |  .9   -1     .9   -1 |     .45 |  8 v23       |
|    249    228     1.1  1.07|    -.17    .15 |  .9    0     .9   -1 |     .33 |  2 v12       |
|    278    229     1.2  1.24|    -.31    .11 | 1.0    0    1.0    0 |     .33 | 12 v41       |
|    152    224      .7   .71|    -.64    .16 | 1.0    0    1.1    0 |     .29 |  9 v31       |
|    325    228     1.4  1.44|   -1.35    .13 | 1.1    0    1.1    1 |     .29 | 10 v32       |
|    335    225     1.5  1.51|   -1.60    .14 | 1.0    0    1.0    0 |     .37 | 11 v33       |
--------------------------------------------------------------------------------------------
|  235.9  225.9     1.0  1.05|     .00    .13 | 1.0     .0  1.0    .0|     .34 | Mean (Count: 14)  |
|   49.1    3.7      .2   .22|     .74    .02 |  .1    1.0   .1   1.0|     .06 | S.D.              |
--------------------------------------------------------------------------------------------
            RMSE (Model)  .14 Adj S.D.   .73  Separation  5.35  Reliability  .97
             Fixed (all same) chi-square: 416.9  d.f.: 13  significance: .00
--------------------------------------------------------------------------------------------
```

Tab. A5 - Thresholds estimates (τ_{ik}= step) for the scales of the groups of criteria

Criteria from 1 a 4 (v11, v12, v13, v14)

```
-----------------------------------------------------------------------------------------------------------------
|     DATA              | QUALITY CONTROL |     STEP       |  EXPECTATION   | MOST   |.5 Cumul.| Cat|Response     |
| Category Counts  Cum.| Avge  Exp.  OUTFIT|CALIBRATIONS |  Measure at    |PROBABLE|Probabil.|PEAK|Category     |
|Score   Used   %    % | Meas  Meas   MnSq |Measure  S.E. |Category  -0.5 |  from  |   at    |Prob|  Name       |
-----------------------------------------------------------------------------------------------------------------
| 0       133  15%  15%| -1.49  -1.29   .9 |              |( -3.42)        |  low   |  low    |100%| 0 (0+1+2+3+4+5)|
| 1       677  74%  89%|  -.12   -.15   .9 | -2.36    .10 |   .00    -2.36| -2.36  |  -2.36  | 84%| 1 (6+7+8)    |
| 2       102  11% 100%|  1.15   1.09  1.0 |  2.36    .11 |(  3.44)   2.37|  2.36  |   2.35  |100%| 2 (9+10)     |
-------------------------------------------------------(Mean)---------(Modal)--(Median)---------------------------
```

Criteria 5 (v15)

```
-----------------------------------------------------------------------------------------------------------------
|     DATA              | QUALITY CONTROL |     STEP       |  EXPECTATION   | MOST   |.5 Cumul.| Cat|Response     |
| Category Counts  Cum.| Avge  Exp.  OUTFIT|CALIBRATIONS |  Measure at    |PROBABLE|Probabil.|PEAK|Category     |
|Score   Used   %    % | Meas  Meas   MnSq |Measure  S.E. |Category  -0.5 |  from  |   at    |Prob|  Name       |
-----------------------------------------------------------------------------------------------------------------
| 0        25  12%  12%| -1.11  -1.21  1.1 |              |( -3.64)        |  low   |  low    |100%| 0 (0+1+2+3)  |
| 1       169  79%  90%|  -.11   -.09  1.0 | -2.57    .23 |   .00    -2.57| -2.57  |  -2.57  | 87%| 1 (4+5+6)    |
| 2        21  10% 100%|  1.09   1.08  1.0 |  2.57    .25 |(  3.65)   2.57|  2.57  |   2.56  |100%| 2 (7+8+9+10) |
-------------------------------------------------------(Mean)---------(Modal)--(Median)---------------------------
```

Criteria 6 (v21)

```
-----------------------------------------------------------------------------------------------------------------
|     DATA              | QUALITY CONTROL |     STEP       |  EXPECTATION   | MOST   |.5 Cumul.| Cat|Response     |
| Category Counts  Cum.| Avge  Exp.  OUTFIT|CALIBRATIONS |  Measure at    |PROBABLE|Probabil.|PEAK|Category     |
|Score   Used   %    % | Meas  Meas   MnSq |Measure  S.E. |Category  -0.5 |  from  |   at    |Prob|  Name       |
-----------------------------------------------------------------------------------------------------------------
| 0        47  21%  21%| -1.01   -.94  1.1 |              |( -2.58)        |  low   |  low    |100%| 0 (0+1+2+3+4)|
| 1       129  57%  78%|   .06    .01    9 | -1.48    .18 |   .00    -1.58| -1.48  |  -1.52  | 69%| 1 (5+6+7)    |
| 2        49  22% 100%|   .97   1.02  1.0 |  1.48    .18 |(  2.59)   1.60|  1.48  |   1.51  |100%| 2 (8+9+10)   |
-------------------------------------------------------(Mean)---------(Modal)--(Median)---------------------------
```

Criteria from 7 to 8 (v22, v23)

```
-----------------------------------------------------------------------------------------------------------------
|     DATA              | QUALITY CONTROL |     STEP       |  EXPECTATION   | MOST   |.5 Cumul.| Cat|Response     |
| Category Counts  Cum.| Avge  Exp.  OUTFIT|CALIBRATIONS |  Measure at    |PROBABLE|Probabil.|PEAK|Category     |
|Score   Used   %    % | Meas  Meas   MnSq |Measure  S.E. |Category  -0.5 |  from  |   at    |Prob|  Name       |
-----------------------------------------------------------------------------------------------------------------
| 0       127  28%  28%|  -.89   -.84   .9 |              |( -1.97)        |  low   |  low    |100%| 0 (0+1+2+3+4+5)|
| 1       184  41%  69%|   .04    .05   .9 |  -.77    .12 |   .00    -1.12|  -.77  |   -.93  | 52%| 1 (6+7)      |
| 2       143  31% 100%|  1.05   1.00   .9 |   .77    .12 |(  1.98)   1.13|   .77  |    .92  |100%| 2 (8+9+10)   |
-------------------------------------------------------(Mean)---------(Modal)--(Median)---------------------------
```

Criteria 9 (v31)

```
-----------------------------------------------------------------
|     DATA              | QUALITY CONTROL  |Response             |
| Category Counts  Cum.| Avge  Exp.  OUTFIT|Category             |
|Score   Used   %    % | Meas  Meas   MnSq |  Name               |
-----------------------------------------------------------------
| 0        72  32%  32%|   .25    .20  1.1 | 0 (0+1+2+3+4+5+6+7+8+9)|
| 1       152  68% 100%|  1.29   1.31  1.0 | 1 (10)              |
-----------------------------------------------------------------
```

Criteria 10, 11 and 13 (v32, v33, v42)

```
-----------------------------------------------------------------------------------------------------------------
|     DATA              | QUALITY CONTROL |     STEP       |  EXPECTATION   | MOST   |.5 Cumul.| Cat|Response     |
| Category Counts  Cum.| Avge  Exp.  OUTFIT|CALIBRATIONS |  Measure at    |PROBABLE|Probabil.|PEAK|Category     |
|Score   Used   %    % | Meas  Meas   MnSq |Measure  S.E. |Category  -0.5 |  from  |   at    |Prob|  Name       |
-----------------------------------------------------------------------------------------------------------------
| 0        49   7%   7%|  -.20   -.41  1.1 |              |( -2.86)        |  low   |  low    |100%| 0 (0+1+2+3+4+5)|
| 1       359  53%  60%|   .88    .86  1.1 | -1.77    .16 |   .00    -1.83| -1.77  |  -1.80  | 75%| 1 (6+7+8)    |
| 2       274  40% 100%|  2.09   2.15  1.1 |  1.77    .09 |(  2.87)   1.84|  1.77  |   1.79  |100%| 2 (9+10)     |
-------------------------------------------------------(Mean)---------(Modal)--(Median)---------------------------
```

Criteria 12 and 14 (v41, v43)

```
-----------------------------------------------------------------------------------------------------------------
|     DATA              | QUALITY CONTROL |     STEP       |  EXPECTATION   | MOST   |.5 Cumul.| Cat|Response     |
| Category Counts  Cum.| Avge  Exp.  OUTFIT|CALIBRATIONS |  Measure at    |PROBABLE|Probabil.|PEAK|Category     |
|Score   Used   %    % | Meas  Meas   MnSq |Measure  S.E. |Category  -0.5 |  from  |   at    |Prob|  Name       |
-----------------------------------------------------------------------------------------------------------------
| 0       124  27%  27%| -1.03  -1.05  1.1 |              |( -2.15)        |  low   |  low    |100%| 0 (0+1+2+3+4+5)|
| 1       199  44%  72%|   .03    .01   .9 |  -.99    .12 |   .00    -1.25|  -.99  |  -1.11  | 57%| 1 (6+7)      |
| 2       128  28% 100%|  1.06   1.11  1.1 |   .99    .12 |(  2.16)   1.26|   .99  |   1.10  |100%| 2 (8+9+10)   |
-------------------------------------------------------(Mean)---------(Modal)--(Median)---------------------------
```

NOTES

1.- Latent trait can be defined as every charateristic of individuals or things which has not a measurement instrument. Note that event the height would be a latent trait if there exists no balances.

2.- http://www.rasch.org/memo18.htm.

3.- The content validity assure that only tests able to measure the latent trait are included in set B. This aspect is partially related with the weights associated to the different test, as a weight equal to zero implies that such test is not in set B.

4.- Like the ability of an athlete, the goodness of a project, the risk of a company, etc.

5.- According to Linacre (http://www.rasch.org/memo61.htm), a study aimed to evaluate the degree of concordance of judges in the optimal setting (expert evaluators, well structured evaluation criteria, and registration of behaviour of individuals with clearly different level) it reached only the 80% and the evaluations expressed by the judges resulted quite imperfect on the basis of the established criteria (Gruenfeld, 1981 p.12). Then, instead of believe that training strategies can remove such differences or pretend such differences do not exist, it would be much better try to account for them measuring their magnitude.

6.- A Mathematics test can not include latin questions, etc.

7.- An analogous property holds for the difference between two item parameters: $\ln\frac{P(X_{ni}=1)}{P(X_{ni}=0)} - \ln\frac{P(X_{nj}=1)}{P(X_{nj}=0)} = (\theta_n - \delta_i) - (\theta_n - \delta_j) = \delta_j - \delta_i$

8.- http://www.rasch.org/memo61.htm.

9.- http://www.rasch.org/rmt/rmt133m.htm.

10.- See also http://www.rasch.org/rmt/rmt82a.htm.

11.- http://ehlt.flinders.edu.au/education/iej/articles/v5n2/curtis/paper.pdf.

REFERENCES

Aerts, D., Gabora, L. (2005). A theory of concepts and their combinations: *I. Kybernets,* vol 34, n.1/2, 151-175.

Andersen, E. B. (1977). Sufficient statistics and latent trait models. *Psychometrika,* (42), 69-81.

Andrich, D. (1978a). A rating scale formulation for ordered response categories. *Psychometrika,* 43, 561-573.

Andrich, D. (1978b). Scaling attitude items constructed and scored in the Likert tradition. *Educational and Psychological Measurement,* (38), 665-680.

Andrich, D. (1978c). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement,* (2), 581-594.

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

Battauz, M., Bellio, R., Gori. E. (2005). *Combining interval and ordinal measures in the assessment of student achievement. CLADAG,* Parma.

Bertoli-Barsotti, L., (2002). On a condition for the existence of the maximum likelihood estimate for concave log-likelihood functions, in Frosin, B. V.; Magagnoli, U.; Boari, G. *Studi in onore di Angelo Zanella.* Milano: (Vita e Pensiero), 25-42.

Bertoli-Barsotti, L., (2003). An order-preserving property of the maximum likelihood estimates for the Rasch model, *Statistics and Probability Letters,* 61, 91-96.

Bertoli-Barsotti, L. (2004). Effects of Ordered Scores in Rasch Analysis, *Atti della XLII Riunione Scientifica SIS,* Bari 9-11 giugno 2004. Padova: Sessioni Spontanee, 743-746.

Bond, T. G. , Fox, C. M. (2001). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences.*

Bond, T. G. , (2003). Validity and assessment: a Rasch measurement perspective. *Metodología de las Ciencias del Comportamiento* 5(2), 179-194

Cronbach, L. J. (1949). *Essentials of Psychological Testing.* New York: Harper & Row. Quotations from the 1970 edition.

Curtis D. D. (2004). Person Misfit in Attitude Surveys: Influences, Impacts and Implications, *International Education Journal* Vol 5, No 2, 2004.

Embretson, S. E., & Hershberger, S. L. (1999). *The new rules of measurement: What every psychologist and educator should know. Mahwah,* NJ: Lawrence Erlbaum Associates.

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

Fischer, G. H. (1995). The Linear Logistic Test Model. In G. H. Fischer & I. Molenaar (Eds.), *Rasch models. Foundations, recent developements, and applications.* New York: Springer-Verlag, 131-156.

Gruenfeld, E. F. (1981). *Performance Appraisal: Promise and Peril.* Ithaca, New York: Cornell University Press.

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

Hambleton, R. K., Swaminathan, H. (1985). *Item response theory.* Boston: Kluwer-Nijhoff.

Henning, G. (1992). *Dimensionality and construct validity of language tests.* Language Testing, 9 (1), 1-11.

Holland, P. W., Wainer, H. (Ed.). (1993). *Differential Item Functioning.* Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Irer (2005). *Dalla differenza, l'equità. Misurare gli apprendimenti disciplinari nella scuola dell'autonomia: Rapporto finale,* ID Progetto FSE: 154633, Cod. IReR 2003C009, Irer, Milano.

Jacobsen, M. (1989). Existence and unicity of MLEs in discrete exponential family distributions. *Scandinavian Journal of Statistics,* 16, 335-349.

Karabatsos, G. (2001). The Rasch Model, Additive Conjoint Measurement, and New Models of Probabilistic Measurement Theory, *Journal of Applied Measurement,* 2(4), 389-423.

Linacre, J. M. (1989). *Many-facet Rasch measurement.* Chicago: MESA Press.

Linacre, J. M. (1998). *A user's guide to FACETS: A Rasch measurement computer program.* Chicago: MESA Press.

Linacre, J. M., Wright, B.D. (1997). *FACETS: Many-Faceted Rasch Analysis*. Chicago: MESA Press.

Lord, F. M. (1975). *Evaluation with artificial data of a procedure for estimating ability and item characteristic curve parameters*. (Research Report RB-75-33). Princeton: ETS.

Lynch, B. K., McNamara, T. F. (1998). Using g-theory and many-facet Rasch measurement in the development of performance assessments of the ESL speaking skills of immigrants. *Language Testing*, 15(2), 158-80.

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

Masters, G. N., Keeves, J.P. (Eds.) (1999). *Advances in Measurement in Educational Research and Assessment*. New York: Pergamon.

McNamara, T. (1996). *Measuring second language performance*. New York: Addison Wesley Longman.

Myford, C., Wolfe, E. (2000b). *Strengthening the ties that bind: Improving the linking network in sparsely connected rating designs* (TOEFL Research Report 65). Princeton, NJ: ETS.

Nickerson, C. A., McClelland, G. H. (1984). Scaling distortion in numerical conjoint measurement. *Applied Psychological Measurement*, 8, 183-198.

Popper, K. R. (1959). T*he Logic of Scientific Discovery*. Hutchinson, London.

Rasch, G. (1977). Danish Yearbook of Philosophy, Vol. 14, 1977, 58-93. *Proceedings of the Symposium in Scientific Objectivity held at "Rolighed", Vedbæk,* May 14-16. Copenhagen: Munksgaard. Disponible en: http://www.rasch.org/memo18.htm

Rasch, G. A. (1968). Mathematical theory of objectivity and its consequences for model construction. In *"Report from the European Meeting on Statistics Econometrics and Management Sciences"*, Amsterdam.

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: The Danish Institute of Educational Research (Expanded edition, 1980. Chicago: The University of Chicago Press).

Rasch, G. (1961). *On general laws and the meaning of measurement in psychology.* Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Theory of Probability (Vol. IV, 321-333). Berkeley: University of California Press.

Rowe, K. J., & Cilione, P. (2000). Data mining and neural network analysis. *ACSPRI Newsletter,* 42, 3-5.

Scheiblechner, H. (1995). Isotonic ordinal probabilistic models. *Psychometrika*, 60, 281-304.

Stocking, M. L. (1989). Empirical estimation errors in item response theory as a function of test properties. *(Research Report RR-89-5)*. Princeton: ETS.

Swaminathan, H. (1983). Parameter estimation in item response models. In R.Hambleton (Ed.), *Applications of item response theory*. Vancouver, BC: Educational Research Institute of British Columbia, 24-44.

Tesio, L. (1995). Independence: the core and currency of functional assessment in rehabilitation medicine. *Acta Gerontol,* 45:133-136.

Wilson, M., Engelhard, G. jr (Eds.) (2000). *Objective Measurement:* Theory into Practice-V. Stamford, CO: Ablex Publishing Corporation.

Wingersky, M. S. (1983). LOGIST: A program for computing maximum likelihood procedures for logistic test models. In Hambleton, R. K. Ed. *Applications of item response theory*. Educational Research Institute of British Columbia. Vancouver B.C. 45-56.

Wright, B. D. (1968). Sample-free test calibration and person measurement. In "Proceedings of the 1967 Invitational Conference on Testing Problems". Princeton, N. J.: Educational Testing Services.

Wright, B. D. (1977). Solving measurement problems with the Rasch model. *Journal of Educational Measurement*, 14, 97-116.

Wright, B. D., Masters, G. N. (1982). *Rating Scale Analysis: Rasch Measurement*. Chicago: MESA Press.

Wright, B. D., Stone, M. H. (1979). *Best Test Design: Rasch Measurement*. Chicago: Mesa Press.

Wright, B., & Mok, M. (2000). Rasch models overview. *Journal of Applied Measurement*, 1 (1), 83-106.

PALABRAS CLAVE

Jueces, medición objetiva, evaluación de proyectos, modelo de Rasch.

KEY WORDS

Judges, Objective measurement, Proyects evaluation, Rasch model.

PERFIL ACADÉMICO DE LOS AUTORES

Profesor Enrico Gori, Catedrático de Estadística (Universidad de Udine).

Doctora Michela Battauz, PhD degree en Estadística (Universidad de Padua) y Becaria Postdoctoral en la Universidad de Udine.

Dirección de los autores: Departament of Statistic - University of Udine
Via Treppo, 18
33100 Udine (Italia)

Fecha recepción del artículo: 27. marzo. 2006
Fecha aceptación del artículo: 24. mayo. 2006