Available online at www.sciencedirect.com

ScienceDirect

Procedia Social and Behavioral Sciences 9 (2010) 373–377

Procedia
Social and Behavioral Sciences

WCLTA 2010

Measure school learning through Rasch Analysis: the interpretation of results

Riccardo Di Nisio [a]

[a]*Department of Quantitative Methods and Economic Thory of Univerity of G.d'Annunzio, Viale Pindaro 42, 65127 Pescara – Italy*

Abstract

From the beginning of the docimology (1922), the assessing of the learning has become increasingly important in the interpretation of study and, in general, in evaluation of the quality in education. Several measurement instruments have been offered such as: distribution pentenaria, t scores, item analysis. Recently, in the context of item response theory (IRT) it is emerging a new measurement technique based on Rasch Analysis in order to obtain an objective measure that does not dependent on either the characteristics of the measuring instrument and does not depend on the skills of individuals.
Although Rasch Analysis allows of offering some interesting research, it remains open issues for the effective implementation of the instrument. The main problems are: the construction of a test and the interpretation of results. In this paper we address the second issue: the interpretation of results. At this regard it is proposed a conversion of logit scores into an evaluation scale more comprehensible and more closed to the measurements of assessing adopted by teachers typically expressed on the scale from 0 (very bad) to 10 (very good).
In detail, after having briefly presented the technique of Rasch Analysis, it will be presented an application of the method on 34 students with the purpose to analyze the ability in learning of mathematics. A conversion of the logit output in to a new measurement scale from 0 to 10 has been proposed to make more easy the interpretation of the score

Keywords: Rasch Analysis, Item Response Analysis, docimology;

1. Introduction

The problem of learning evaluation has increased of interest and it has been object of important debates that have criticized the characteristics of a lot of tools used for recognition and assessment of the knowledge acquired by students. At this purpose since 1922 a new approach has been arisen. It is the docimology and its development is due to the increasing social attention to the problems of education and learning assessment.

Particular importance was given to the tools of learning measurement. The tendency of many government institutions is to formulation a new approach based on a "objective measure" less affected by subjective errors of assessment that are often the basis for many known problems as: "pigmaglione effect", "halo effect", "contrast effect", etc.

The purpose of this study is to investigate, from a quantitative point of view, the validity and reliability characteristics that determine the reliability of assessment measures with specific attention to the sphere of school education. In this regard we present the results obtained by Rasch Analysis computing the data of a test concerning ability in learning of mathematics.

doi:10.1016/j.sbspro.2010.12.167

2. The concept of reliability in assessments of learning.

The reliability of a measure is closely linked to its ability to provide objective comparisons that do not change their meaning along the continuum of the variable (Rasch, 1967). Starting from this definition, in the assessment of learning, Rasch 1967 proposed to detect the level of learning through the following interaction

$$\theta_i - \beta_j$$

where θ_i and β_j are, respectively, the score of i-th student (*ability*) and the score of the j-th item (*difficulty*).

For dichotomous data the probabilistic model is:

$$P\left(X_{ij} = 1 \middle| \theta_i, \beta_j\right) = \frac{e\left(\theta_i - \beta_j\right)}{1 + e\left(\theta_i - \beta_j\right)} = p_{ij} \qquad [1]$$

In this model data are collected in the raw score matrix, with n rows (one for each subject) and J columns (one for each item), whose values are equal to 0 or 1. The sum of each row $r_i = \sum_{j=1}^{J} x_{ij}$ represents the total score of the subject i for all the items, while the sum of each column $s_j = \sum_{i=1}^{n} x_{ij}$ represents the score given by al the subjects to the item j. These scores are given to a metric that, being nonlinear, produces some conceptual distortion when looking to compare the row and column totals. In this instance, it is necessary to change these scores according to a metric that is founded on the conceptual distances between subjects and items [25]. The transformation takes place through the logit

$$\log \frac{p_{ij}}{1 - p_{ij}} \qquad [2]$$

By substituting equation (1), respectively with the numerator and the denominator of (2), it is possible to define the parameters θ_i and β_j in the same measurement unit of an interval scale. Consequently even the difference $\theta_i - \beta_j$ is gauged according to the same measurement unit.

The Rasch model possesses some important properties. The first is that the items measure only one latent feature (*one-dimensionality*) and this is an advantage in the assessment of learning where learning is typically one-dimensional. Another important characteristic is that the answers to an item are independent of answers to other items (*local independence*). In regard to parameters, for which no assumptions are made [22], by applying the logits previously described, θ_i and β_j can be expressed according to a common measurement unit on the same continuum (*parameters linearity*); the estimation of θ_i and β_j are respectively test and sample free (*parameters separability*); and the row and column totals on the raw score matrix are sufficient statistics for the estimation of θ_i and β_j (*sufficient statics*).

The Rasch dichotomous model has been extended to the case of more than two ordered categories. The innovation of this approach is in the assumption that between each category and the next there is a threshold that qualifies the item's position and specializes the β_j as a function of the difficulty presented by every answer category. Thus the answer to every threshold h of the item j depends on the value $\beta_j + \tau_h$, where the second term represents the h-th threshold of the item j. The model of polytomous is

$$P\left(X_{ij} = x_{ij}\right) = \frac{e\left[k_{jx} + x_{ij}\left(\theta_i - \beta_j\right)\right]}{\sum_{h=0}^{m} e\left[k_{jx} + x_{ij}\left(\theta_i - \beta_j\right)\right]} \qquad [3]$$

Where X is the random variable which describes the answer of the subject i to the item j; $x_{ij} = 0,1,\mathrm{K}\ ,m$ is the number of ordered overtaken thresholds; k_{jx} are the coefficients of each category x for each item j and they can be estimatd by considering that: $k_{j0} = k_{jm} = 0$ (the first and the last parameters are equal to zero) and that: $k_{jx} = -\sum_{h=1}^{x} \tau_{jh}$ (the category coefficients are defined in terms of thresholds); τ_{jh} is the h-th ordered threshold of the item j.

3.The Rasch Model in practice.

The Rasch model has been applied to data collected during a test of mathematics. The questionnaire concerns 15 dicotomicous items that have been applied on 34 students.

The data have been processed using alghoritm PAIR by software RUMM 2020.

Table 1 highlights the summary test-of-fit statistics. The item-trait test of fit examines the consistency of all the item parameters across the subject measures: data are combined across all the items in order to give an overall test-of-fit. As it is shown in table 1

Table 1. Summary test of fit statistics for the Rasch model

Item trait interaction	Value
Total item χ^2	33,85
Total degree of freedom	30
Total χ^2 probability	0,28

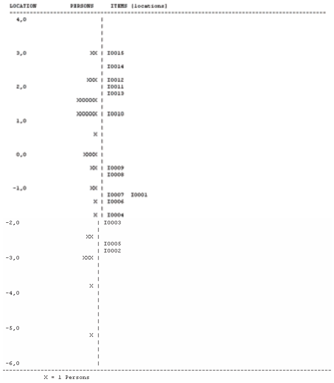
Reliability indices	Value
Separation Index	0,86

The Separation Index, which is the Rasch reliability estimate, computed as the ratio (true/(true+error)) variance whose estimates come from the model. A value of 1 indicates a lack of error variance, and thus full reliability. This index is usually very close to the classic Cronbach α .

Figure 1 shows the classical “Rasch ruler” (also called the “item map”) obtained from our data. The vertical dashed line represents the ideal less-to-more continuum of “ability”. Items and students share the same linear measurement units (logits, left column). Conventionally, the average item is set equal to 0. On the right of the dashed line, the ‘difficulty” items are aligned from easy to very difficult , starting from the bottom and the value represents the item number.

Along the same line, on the left, students are aligned in increasing order of ability from bottom to top. Each *X* symbol represents one student.

Students scores range from –5 to 3 logits. Thus, we observe a spread of more than 6 units for ability. The measurement of ability obtained from this set of items seems reliable, with the range being sufficiently wide.



```
LOCATION        PERSONS    ITEMS [locations]
================================================================================
  4,0                  |
                       |
                       |
                       |
                       |
  3,0               XX | I0015
                       |
                       | I0014
                       |
                   XXX | I0012
  2,0                  | I0011
                       | I0013
                XXXXXX |
                       |
                XXXXXX | I0010
  1,0                  |
                       |
                     X |
                       |
                       |
  0,0                XXXX |
                       |
                    XX | I0009
                       | I0008
                       |
 -1,0               XX |
                       | I0007  I0001
                     X | I0006
                       |
                     X | I0004
 -2,0                  | I0003
                       |
                    XX |
                       | I0005
                       | I0002
 -3,0              XXX |
                       |
                       |
                       |
                     X |
 -4,0                  |
                       |
                       |
                       |
 -5,0                  |
                     X |
                       |
                       |
 -6,0                  |
--------------------------------------------------------------------------------
          X = 1 Persons
```

Figure 1. Item map

The observed answer distribution is compared to the expected answer distribution, calculated with the logistic function distribution, by means of the χ^2 criterion. We examine the χ^2 probability (*p*-value) for whole item set; the null hypothesis is that there is no interaction between the response to the items and the locations of the subjects along the trait. In our case the overall χ^2 is 33,8 with 30 degree of freedom and the *p*-value is 0,28, so the null hypothesis is accepted.

The χ^2 value has been calculated also for each parameter. In table 2 it has been shown the values sorted by location parameter. The *p*-value suggest to remove such variable with a low value.

Table 2. Item sorted by item location parameters

	Location	Chi	Prob
I0002	-2,724	11,627	0
I0005	-2,437	0,933	0,617
I0003	-1,865	0,259	0,875
I0004	-1,772	2,527	0,264
I0006	-1,222	1,262	0,52
I0007	-1,188	3,412	0,16
I0001	-1,159	4,744	0,069
I0008	-0,536	0,956	0,61
I0009	-0,252	0,125	0,938
I0010	1,214	2,304	0,298
I0013	1,811	2,693	0,24
I0011	2,051	1,097	0,567
I0012	2,205	1,422	0,478
I0014	2,786	0,321	0,848
I0015	3,087	0,172	0,915

The table 2 also shows a ranking of items from the one with the low difficulty rating to the one with the high level. In our case the item easier is item 2 and item 5. The items more difficult are: item 14 and item 15. It is also possible to verify the goodness of single item by DIF analysis which allow to verify if subjects with differing levels of ability follow the Rasch model and also to measure if a generic item has a greater or lesser ability rating itself, within the various classes.

Rasch analysis also allows to estimate the location of individuals along their latent abilities. In this case the results have been sorted by location. The results are shown in table 3 between the correspondence score obtained through a suitable transformation in range 0-10.

Table 3. Student locations and scores

ID	Location	Score
30	-3,786	0
9	-2,903	1,291313
19	-2,903	1,291313
27	-2,903	1,291313
8	-2,292	2,184849
17	-2,292	2,184849
26	-1,781	2,932144
15	-1,31	3,620942
11	-0,848	4,296578
34	-0,848	4,296578
22	-0,375	4,988301
25	-0,375	4,988301
3	0,124	5,718046
5	0,124	5,718046
13	0,124	5,718046
32	0,124	5,718046
10	0,654	6,493127
1	1,207	7,301843
21	1,207	7,301843
23	1,207	7,301843
29	1,207	7,301843
31	1,207	7,301843
33	1,207	7,301843
2	1,776	8,133957
6	1,776	8,133957
12	1,776	8,133957
14	1,776	8,133957
20	1,776	8,133957
24	1,776	8,133957
7	2,373	9,00702
18	2,373	9,00702
28	2,373	9,00702
4	3,052	10
16	3,052	10

The transformation has been compute by using the following formula:

$$score = \frac{x_{\min} - x_i}{x_{\max} - x_{\min}} 10 \quad i = 1,2,\mathrm{K}\ ,34$$

Where x_i is the location of the *i*-th student and $x_{\min}$ and $x_{\max}$ are, respectively, the minimum and maximum value of score location. In this way it is been possible convert logit scores into an evaluation scale more comprehensible and more closed to the measurements of assessing adopted by teachers typically expressed on the scale from 0 (very bad) to 10 (very good).

4.Discussion.

In this paper we have shown the possibility to measure the learning ability by using an approach that allows to obtain an objective measure that increases the reliability of judgments. The results are computed in terms of logit. To make more easy the meaning of the data, these are been converted into a scale from 0 to 10 as shown in table 3.

In this way the Rasch Model produce an objective measure that combine ability of a student with the difficult of items making more reliable the measurement.

Rasch Model also solves two further problems: the quantification of ordinal data and the calibration of the test. Using the formula [3] it’s possible to extend the model also in polytomous variables.

References

Rasch G. (1960/1980). *Probabilistic Models for Some Intelligence and Attainment Tests*, (Copenhagen, Danish Institute for Educational Research), e with foreward and after word by B.D. Wright. Chicago: The University of Chicago Press.

Andrich D. (1978a). A rating formulation for ordered response categories. *Psychometrika*, XLIII(4), 561-573.

Andrich D., Sheridan B., Lyne A. & Luo G. (2000). *RUMM: A Windows-Based Item Analysis program Employing Rasch Unidimensional Mmeasurement Models*. Perth, Australia: Murdoch University.

Likert R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 595-639.