

**EJASA:DSS – Electronic Journal of Applied Statistical Analysis: Decision Support Systems and Services Evaluation**

[http://siba-ese.unisalento.it/index.php/ejasa\\_dss/index](http://siba-ese.unisalento.it/index.php/ejasa_dss/index)

e-ISSN: 2037-3627

DOI: 10.1285/i2037-3627v4n1p22

**Evaluation of performance at university with the Rasch Analysis**

By Piergiorgio Mossi, Sara Calogiuri, Paola Tondo

Published: 28 December 2013

This work is copyrighted by Università del Salento, and is licensed under a Creative Commons Attribution - Non commerciale - Non opere derivate 3.0 Italia License.

For more information see:

<http://creativecommons.org/licenses/by-nc-nd/3.0/it/>

# Evaluation of performance at university with the Rasch Analysis

Piergiorgio Mossi\*, Sara Calogiuri, and Paola Tondo

*University of Salento, Department of History, Society and Human Study  
Palazzo Parlangei, Via V.M. Stampacchia, 45, 73100, Lecce, Italy*

Published: 28 December 2013

In 2007/2008 the University of Salento has started a degree course in Psychology. Right from the start, the people in charge of the course have been focusing on preventing early dropouts by stimulating a better professional awareness and by ensuring the best possible course experience (Sánchez et al., 2001). To do so, the students were asked increasing performance standards to prevent the choice of university studies as mere alternative to work. This study refers specifically of controlling the training process through the analysis of performance of students (monitoring career). Normally a review of developments in the curriculum in relation to explanatory variables is performed using statistical inferential model where the dependent variable is the grade obtained in the different disciplines or even the credits acquired by the student. The particular structure of the activity since its inception allows to introduce the evaluation of the curriculum using effectively Rasch Analysis. The model involves the insertion of the exams outcomes analysed in terms of success / failure examination for each of the constituent disciplines of studies. This model establishes a more reliable criteria for the verification of differential variables useful for monitoring the training process (for example: final grade to high school diploma, sex, age, university registration renewal , etc.) and the subsequent management of the training process.

**keywords:** Education, Rasch Analysis, Many Facet Model, Performance of Academic career, career monitoring.

---

\*Corresponding author: [piergiorgio.mossi@unisalento.it](mailto:piergiorgio.mossi@unisalento.it)

## 1 Introduction

In 2007/2008 the University of Salento has started a degree course in Psychology. Right from the start, the people in charge of the course have been focusing on preventing early dropouts by stimulating a better professional awareness and by ensuring the best possible course experience (Venuleo et al., 2009). To do so, the students were asked increasing performance standards to prevent the choice of university studies as mere alternative to work.

Among the functional objectives that motivate the existence of a course of study we find that it is necessary to ensure a high standard of discipline for students while ensuring only a small number of drop-outs. In this regard, it must be said that generally the university implements various initiatives, like leveraging on the enlargement of the training and the implementation of initiatives to support the students to reduce early dropouts.

Various managing activities have been undertaken on this aspect, some (Salvatore and Scotto di Carlo, 2005) regarding the training experience (seminar approach to the studies<sup>1</sup>, debt recovery activities, tutoring, workshops based on the university experience). On the other hand it is equally useful to monitor the students' careers with the purpose of controlling the same training process (monitoring career). This has a dual utility: on one hand it helps the recruitment of students who are competent enough to benefit from the degree course, on the other hand it makes clear to the students themselves that the intervention is not reduced only to the operational dimension, but it also implies the dimension of organizational management of the training process (Sánchez et al., 2001). The ability to make predictions on the recruitment of students has aroused much interest and generated different research studies. Most of these studies take into account some independent variables, such as voting at secondary school and the type of studies in this cycle, or even the performance achieved in the admission test. These are then put in relation to a variable performance which is generally the credits or, more frequently, the average grade for career exams (Mouw and Khanna, 1993) (Sánchez et al., 2001). In this paper we will focus specifically on the evaluation of the variable measuring the quality of performance. To do this we used a Rasch model type multifacet. We will see how this model allows us to have more reliable measurements which provide, in return, a better reading of the weight taken by the independent variables in the analysis estimated by linear regression.

## 2 Access to course career

Before describing the method of data collection useful for measuring the performance, it is necessary to dwell briefly on the student recruitment system of the university. Please note that this system has been effectively used since the first year of the course and for the following three years, however, access to the first cohort that we used in this study

---

<sup>1</sup>The seminars on *methodology of the study* were targeted to students in educational debt subsequent to the result of the start tests. So, these training sessions for students in debt following the results of admission tests are set out to develop models, strategies and regulatory criteria capable to support the effectiveness of learning in the University.

did not include the closed number, while in later years a limitation was placed on access. The matriculation to the psychology undergraduate program was subordinated to an admission test, although non-selective. The paper-pencil text was composed as follows:

- the CIQ, the AIQ and the PIQ questionnaires, aimed to gather the dimensions of symbolization (DS) of the context (Venuleo et al., 2013). More precisely the Context Image Questionnaire (CIQ) with questions that concerned the social context (evaluation of the place where the student lives, evaluation of social structure and institutional reliability, such as Municipality, National Health System, Companies ...), the Academic Image Questionnaire (AIQ) with questions designed to facilitate the expression of perceptions/opinions/judgments concerning the academic system (opinions about main functions, problems, strategies of improvement, factors which influence the learning process, ...) and Professional identity Image Questionnaire (PIQ) designed to facilitate the expression of perceptions/opinions/judgments concerning professional identity (field of intervention, functions, users) <sup>2</sup>

- 4 subtest aimed to analyze the knowledge and the already achieved competence: a subtest of general culture, a subtest on English skill, a subtest on reading comprehension and a subtest on the logic competence. The questionnaire of the students who performed their matriculation (N=823 cases at first year, 656 at second Year and 529 at third year) compose data on which analysis was performed.

## 2.1 The database

The start of this course is especially marked by this type of monitoring initiative. In particular, for the purpose of monitoring all the teachers of subjects related to the degree course, the Presidency has repeatedly invited to provide the monitoring team of the results of various incurred exams. The data sent to the monitoring team gives the opportunity to record the rates of failure of each exam or, if successful, the vote reported in the examination. So the particular structure of the activity since its inception allows the acquisition of the results of a career not only in terms of quantitative results, or through the allocation of a vote or the subsequent derivation of the credits, but also by examining the reporting of failures recorded by the students during the academic career.

It was therefore established an electronic database that has collected data for each student career. The database contains information on socio-demographic characteristics of students, the grade obtained at the end of secondary education, the performance achieved in the university admission tests (see previous paragraph), the outcomes of

---

<sup>2</sup>In order to depict DS ad hoc questionnaires were constructed, by referring to a specific psychosocial methodology, elaborated by Carli (Carli and Paniccia, 2003) (Carli and Salvatore, 2001), and developed by Mossi and Salvatore (Mossi and Salvatore, 2011): ISO methodology. Broadly speaking, the method aims at singling out dimensions of symbolization of the context on which it is possible to differentiate a given population. The global questionnaire consists of 135 multiple-choice questions, constructed to facilitate the expression of perceptions / opinions / judgments on the three general areas of expertise (CIQ, AIQ and PIC). Three Multiple Correspondence Analysis were performed (ACM) (Benzécri, 1979) on the response matrix in order to detect the factorial dimensions that explain the entire variability of the matrix. The factorial coordinates were drawn from the three analyses.

the courses for the recovery of debts recorded under admission<sup>3</sup> and, finally, the results obtained for each session of exams in the disciplines covered by the plan of studies.

## 2.2 Toward the construction of a performance indicator

The end of the first year has provided us the first statistical processing monitoring: on one hand, descriptive statistics on changes in the degree course<sup>4</sup>, on the other statistical inferences on the extent of the relationship between the variables that characterize the initial preparation possessed by the students and results in examinations.

It was, however, soon realized that the classic indicators such as average grade and credit purchases did not take into account all available information on student performance (Kulatunga Moruzi and Norman, 2001). A more accurate indicator in fact, composed of the average rating, compared to the number of sessions dealt with to pass the exams, improved, albeit slightly, some correlation values obtained between the variables characterizing the students and the variables of performance exams (see Table 1).

Table 1: Comparison between performance indexes: Variables that characterize the initial preparation possessed by the students and careers performance indexes

		Average Grade	Total credits earned	Number of Sessions	Average Rating
Mark of high school	Correlation	.326(**)	.276(**)	-.200(**)	.341(**)
	N	625	652	635	635
Logic	Correlation	.152(**)	.100(*)	-.164(**)	.153(**)
	N	626	653	636	636
Reading comprehension	Correlation	.149(**)	.104(**)	-.081(*)	.166(**)
	N	626	653	636	636
English skill	Correlation	.211(**)	.170(**)	-.158(**)	.199(**)
	N	619	646	629	629
General culture	Correlation	.259(**)	.115(**)	-.144(**)	.267(**)
	N	626	653	636	636

(\*\*) Correlation is significant at the 0.01 level (2-tailed).

(\*) Correlation is significant at the 0.05 level (2-tailed).

## 2.3 Problems of a performance index evaluation

Generally we find problems in the construction of an effective performance index. In fact, although they are often used in literature, the average grade and credits earned suffer from some of the following inaccuracies:

- Evaluation of the laboratories without grade: the average score does not take into account the results obtained by students in examinations for the outcome of which there is no vote. The issue is partially overcome by using the acquired credit;

<sup>3</sup>The seminars on *methodology of the study*, see note 1.

<sup>4</sup>These statistics are not reported here, but you can find them on the Internet at: [https://www.scienzedellaformazione.unisalento.it/c/document\\_library/get\\_file?uuid=2726af31-4031-4b4c-8b3a-d770d6b55d9f&groupId=886128](https://www.scienzedellaformazione.unisalento.it/c/document_library/get_file?uuid=2726af31-4031-4b4c-8b3a-d770d6b55d9f&groupId=886128)

- Evaluation of failures: the average grade that is not taken into account by the credits earned and that do not take into account the times when the students fail the exams;
- Evaluation of performances in time: classic indicators can generally be seen as 'static' i.e. assess performance at a given moment without taking into account the whole career of the student;
- Differences among professors' judgements when they assign the same grade: same grades by different teachers do not take the same meaning in value. Each teacher has its own yardstick fact;
- Possible missing data (the system is not automatically enforced): the collection system is not implemented through the study secretariat. Thus, there is the possibility that some calls are not recorded at all (data missing).

For this reason it was considered appropriate to deal differently with this measurement by using an estimate of the performance obtained with a logistic model. The way in which the results were collected made it possible to introduce the evaluation of the curriculum using Rasch Analysis (Salvatore et al., 2008). The model involves the insertion of the exams outcomes analysed in terms of success / failure or each of the constituent disciplines of studies. So by treating the exams as items of the analysis, the Rasch model allows to assign a value to each student and to estimate logit that defines the degree of performance of the career.

So, this model establishes more reliable criteria for the verification of differential variables characterizing the students and the subsequent management of the training process.

### 3 Analysis. Three Facet Model Dichotomous

To obtain an estimate of the performance it was decided to use a valuation model type Many Facet (Linacre, 1994) (Eckes, 2011). It is one of many other extensions to the Rasch model. This model expands the polytomic form of the classic dicotomic model. The mathematical features of the model are kept, but one of more extra component about the estimate situation are introduced.

$$\log \frac{P_{nst}}{1 - P_{nst}} = \beta_n - \delta_s - \theta_i \quad (1)$$

$$P_{nsi} = \frac{\exp(\beta_n - \delta_s - \theta_i)}{1 + \exp(\beta_n - \delta_s - \theta_i)} \quad (2)$$

In (1) and (2) is students' abilities, is the difficulty of disciplines and represents the assessment made in the years.

As in the dicotomic model, raw scores represent sufficient statistics for the evaluation of  $\beta_n$ ,  $\delta_s$  and  $\theta_i$ . The counts of observations in each category are the sufficient statistics for estimating the  $Fk$  that is the probability to be accounted in the k categories related to the

$k - 1$  one. As explained in the presentation of Facets (Linacre, 2011), this software that we are using to calculate the measures and estimate the model, also supports powerful quality control fit statistics for assessing the conformance of the data to the model. The model is robust against many forms of misfits, so that the typical perturbations in data tend to have little influence on the measure estimates. A further feature of the model is its robustness against missing data. Since the model is parameterized at the individual observation level, estimates are obtained only from the data that has been observed. There is no requirement to impute missing data or to assume the overall form of the distribution of parameters.

## 4 Results

Table 2 shows some of the reliability indices obtained in the three years under investigation. One can observe an obvious increase in the number of discipline of study over the years. In the three models evaluated, both the separation indices and index of reliability are satisfactory.

Table 2: Indices of adequacy and reliability of the model used

<p><b>14 disciplines year 2008</b></p> <p>Without extremes, Model, Populn: RMSE .26 Adj (True) S.D. 2.29 Separation 8.90 Strata 12.21 Reliability .99                      Without extremes, Model, Sample: RMSE .26 Adj (True) S.D. 2.43 Separation 9.45 Strata 12.39 Reliability .99                      With extremes, Model, Fixed (all same) chi-square: 674.1 d.f.: 13 significance (probability): .00                      With extremes, Model, Random (normal) chi-square: 13.4 d.f.: 12 significance (probability): .34</p>
<p><b>26 disciplines year 2009</b></p> <p>Without extremes, Model, Populn: RMSE .50 Adj (True) S.D. 2.39 Separation 4.74 Strata 6.65 Reliability .96                      Without extremes, Model, Sample: RMSE .50 Adj (True) S.D. 2.46 Separation 4.87 Strata 6.82 Reliability .96                      With extremes, Model, Fixed (all same) chi-square: 1798.0 d.f.: 25 significance (probability): .00                      With extremes, Model, Random (normal) chi-square: 24.7 d.f.: 24 significance (probability): .42</p>
<p><b>30 disciplines year 2010</b></p> <p>Without extremes, Model, Populn: RMSE .43 Adj (True) S.D. 2.24 Separation 5.18 Strata 7.24 Reliability .96                      Without extremes, Model, Sample: RMSE .43 Adj (True) S.D. 2.29 Separation 5.30 Strata 7.40 Reliability .97                      With extremes, Model, Fixed (all same) chi-square: 2177.0 d.f.: 29 significance (probability): .00                      With extremes, Model, Random (normal) chi-square: 27.3 d.f.: 28 significance (probability): .50</p>

Figure 1 shows the measures in the vertical ruler during the 3 years analyzed in this study. These are stated as a possible example of the monitoring process. As you can see in the image, the measure changes in the third year of study. Although this change is not substantial, the performance turns out to be worse year after year. This result could be interpreted by a decrease in the required standard that follows an initial period of increased attention to the quality of university study. On the other hand, one can observe that while the distribution of students is nearly similar in the three years, disciplines seems to tend to a lower measure.

Table 3 lists some summary indices of linear regression models applied on two different variables used to measure students' performance. The indexes in question are compared in the three years of evaluation under investigation. As you can see the performance measures obtained with the Rasch measure, always leads to an improvement of the

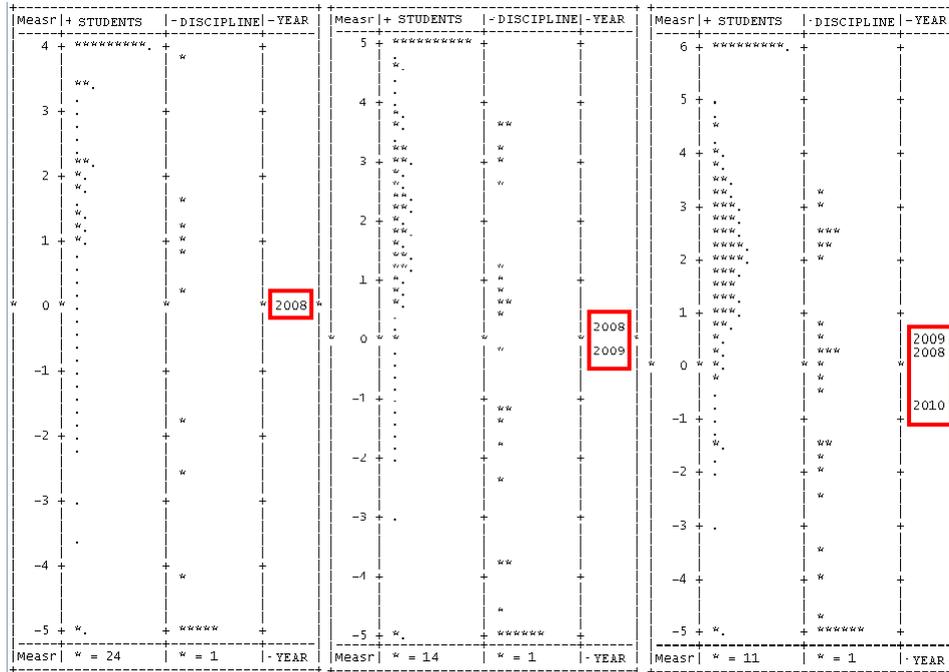


Figure 1: Vertical ruler over the years of study.

variability explained by the regression model applied. In addition, the error of the estimate is more contained when it is used as the dependent variable the logit. The Durbin Watson test indicates the absence in all models of autocorrelation among the residuals (Dillon and Goldstein, 1984).

Table 4 shows, on the other hand, only the results of the linear regression performed on the measured performance in 2010 through the Rasch measures. The model is globally significant ( $F = 42.268$ ,  $df 7$ ,  $p < .000$ ). Too collinearity statistics indicate a substantial independence of the variables. As you can see from the standardized coefficients, maturity results seem to be the most effective predictors of a student's career.

The position expressed by each of the students in the questionnaires CIQ and AIQ proposed at the beginning of the study, was categorized, after analyzing the responses, in an attitude of trust or distrust with regard to the image of the Social Context and devaluation or investment with regard to the image of the Academic. Subsequently, through a contingency table, these categories of the questionnaires were linked with the performance measured by the Rasch analysis. In this case the students have been selected only with a low and a high skill (worst and best). The two figures (2 and 3) express the results of this analysis. The analysis shows that students that have the best performance at the end of the third year are also those who showed a larger trust in the social context (see figure 2). Instead, students with the worst performance are also those who showed a negative representation of the social context.

Similarly, students that at the end of the third year have a better performance are also

Table 3: Comparison between performance indexes: Summary indicators of adaptation and validity in linear regression models in the years 2008-2010, using as the dependent variable 2 different performance indices (average grade e Rasch measurement), and, as predictors, some variables that characterize students.

	<i>R</i>	<i>R</i> <sup>2</sup>	Adj. <i>R</i> <sup>2</sup>	S.E.	Durbin-Watson
Model Summary Performance Average grade in 2008(b)	.428(a)	.183	.175	1.883	1.989
Model Summary Facet Rasch measurement model in 2008(b)	.534(a)	.285	.276	.27058	1.976
Model Summary Performance Average grade in 2009(b)	.430(a)	.185	.176	1.905	2.005
Model Summary Facet Rasch measurement model in 2009(b)	.608(a)	.370	.361	.20400	1.986
Model Summary Performance Average grade in 2010(b)	.474(a)	.224	.216	1.896	1.936
Model Summary Facet Rasch measurement model in 2010(b)	.651(a)	.424	.415	.17383	2.116

(a) Predictors: (Constant) Subtest of general culture; Sex; Year of birth; Subtest of reading comprehension; Subtest of English skill; Mark of high school leaving qualifications; Subtest of logic competence; (b) Dependent Variable

Table 4: Linear regression model with Performance Facet Rasch measurement in 2010

	Unstandardized Coefficients		Standardized Coefficients		Sig.
	B	SE	Beta	t	
(Constant)	17.674	2.423		7.30	0.000
Sex	.003	.030	.004	.11	.914
Year of birth	-.009	.001	-.284	-7.58	.000
Mark of high school	.893	.079	.441	11.34	.000
Logic competence	.072	.033	.085	2.19	.029
Reading comprehension	.068	.031	.084	2.21	.027
English skill	.102	.031	.128	3.31	.001
general culture	.174	.032	.212	5.42	.000
	Correlations			Collinearity Statistics	
	Zero Order	Partial	Part	Tolerance	VIF
Sex	-.021	.005	.004	.917	1.090
Year of birth	-.230	-.343	-.277	.954	1.049
Mark of high school	.478	.480	.415	.885	1.130
Logic competence	.223	.105	.080	.883	1.132
Reading comprehension	.187	.106	.081	.924	1.082
English skill	.260	.158	.121	.896	1.116
general culture	.380	.253	.198	.877	1.141

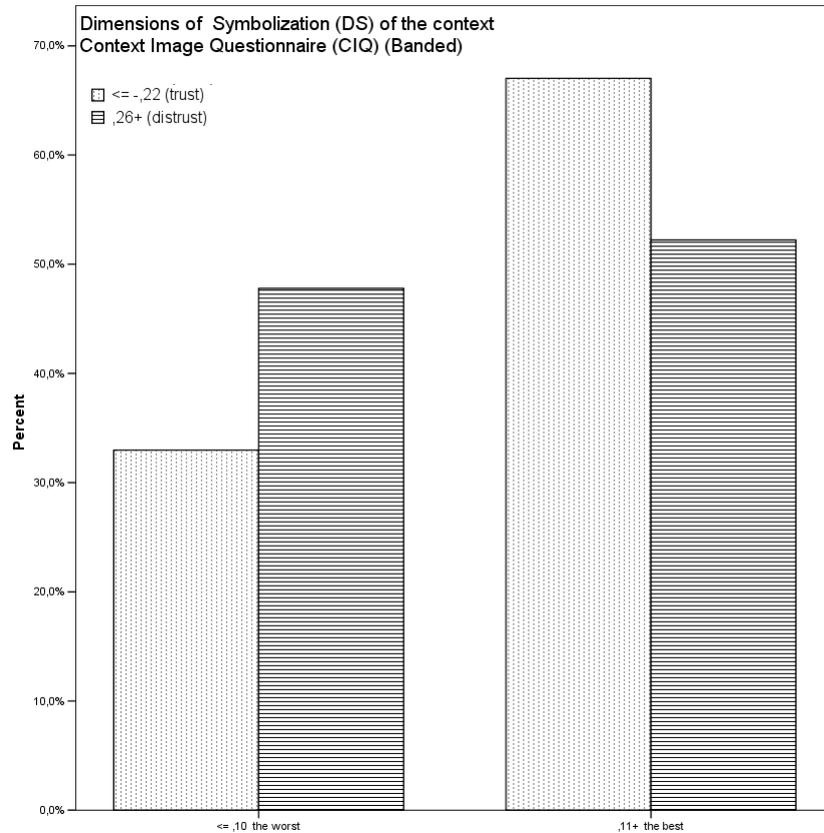


Figure 2: Contingency between Context Image Questionnaire (CIQ) with Summary Performance Facet Rasch measurement model in 2010.

the ones that had evaluated university as place for investments. Instead students with the worst performance are also those who showed a negative representation of University (see figure 3).

## 5 Discussion and conclusions

The career monitoring has a dual function: on one hand it allows to recruit competent students, on the other it makes clear to the same students that the intervention is not only the operative dimension, but involves also the management of the formative process organization (Sánchez et al., 2001). A more precise performance evaluation represents an important step for the graduation course activities. In this study we reported some results obtained by monitoring<sup>5</sup>. Having an effective model of performance measurement is an advantage when you want to expand the explained variability and decrease the error of interpretation. Moreover, the model Facets that was adopted for the evaluation, was

<sup>5</sup>For further information please refer to (Venuleo and Mossi, 2011) and (Venuleo et al., 2013).

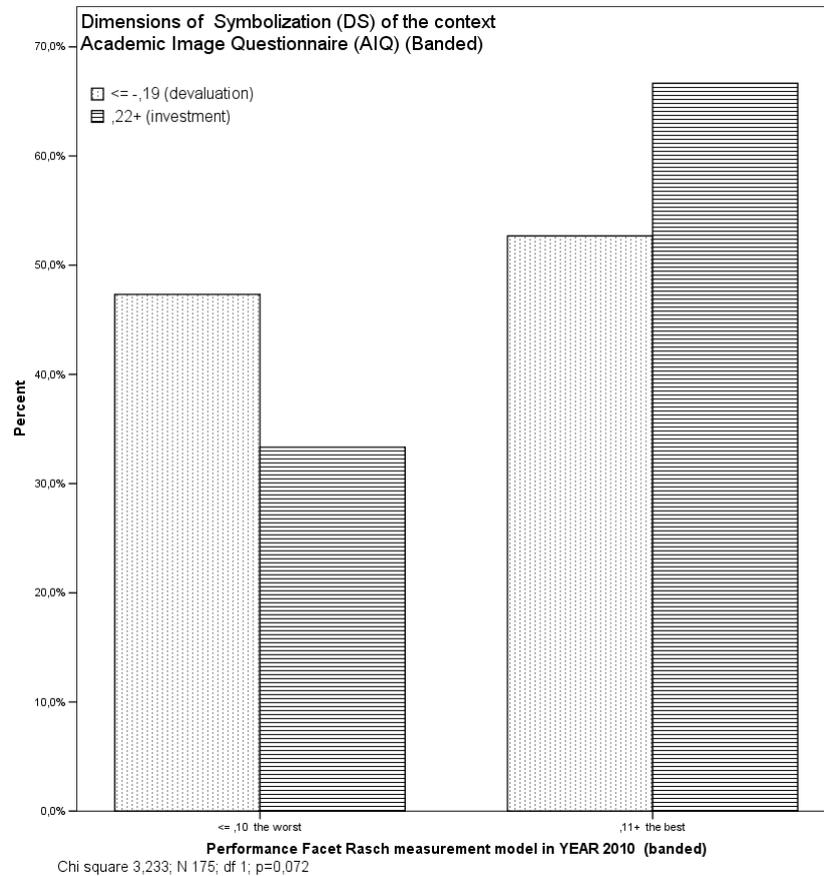


Figure 3: Contingency between Academic Image Questionnaire (AIQ) and Summary Performance Facet Rasch measurement model in 2010.

able to take account of student performance even under a dynamic profile. Estimates obtained in fact held into account the views expressed by teachers in several years of duration of the investigation.

The monitoring is still going on with the following courts. We will try to confirm with these ones the results obtained from the students with from this first detection.

## References

- Benzécri, J. P. (1979). Sur le calcul des taux d'inertie dans l'analyse d'un questionnaire. *Cahiers de l'Analyse des Données*, 4(3):377–379.
- Carli, R. and Panicia, R. M. (2003). *Psicologia della formazione*. Il Mulino.
- Carli, R. and Salvatore, S. (2001). *L'immagine della psicologia. una ricerca sulla popolazione del lazio*.
- Dillon, W. R. and Goldstein, M. (1984). *Multivariate analysis*. Wiley New York.

- Eckes, T. (2011). Introduction to many-facet rasch measurement. *Frankfurt: Peter Lang*.
- Kulatunga Moruzi, C. and Norman, G. R. (2001). Validity of admissions measures in predicting performance outcomes: the contribution of cognitive and non-cognitive dimensions. *Teaching and Learning in medicine*, 14(1):34–42.
- Linacre, J. M. (1994). *Many-facet Rasch measurement*. Mesa Press Chicago.
- Linacre, J. M. (2011). *Facets computer program for many-facet Rasch measurement, version 3.68.0*. Beaverton, Oregon: Winsteps.com.
- Mossi, P. and Salvatore, S. (2011). Psychological transition from meaning to sense. *European Journal of Education and Psychology*, 4(2).
- Mouw, J. T. and Khanna, R. K. (1993). Prediction of academic success: A review of the literature and some recommendations. *College student journal*.
- Salvatore, S., Mossi, P., Venuleo, C., and Guidi, M. (2008). Pregnanza e sensatezza della formazione. *La qualità dell'Università. Verso un approccio psicosociale*, pages 49–69.
- Salvatore, S. and Scotto di Carlo, M. (2005). *L'intervento psicologico nella scuola. Modelli, metodi, strumenti*. ICA Roma.
- Sánchez, M. M., Rejano, E. I., and Rodríguez, Y. T. (2001). Personality and academic productivity in the university student. *Social Behavior and Personality: an international journal*, 29(3):299–305.
- Venuleo, C., Guidi, M., Mossi, P., and Salvatore, S. (2009). L'istituzione generativa. l'impatto di un setting riflessivo sulle immagini anticipatorie del corso di laurea in psicologia. *Psicologia Scolastica*, page 197.
- Venuleo, C. and Mossi, P. (2011). Modelli di interpretazione dell'ambiente sociale e continuità degli studi. un caso studio. *Psicologia Scolastica*, 10(2):171–193.
- Venuleo, C., Mossi, P., and Salvatore, S. (2013). Symbolic positions and retentions study at university. a case study.