ACADEMIC REPUTATION: A PROPOSAL FOR A NEW SCALE

Emma Zavarrone[1]

Department of Consumption, Behaviour and Corporate Communication "Giampaolo Fabris", IULM University, Milan, Italy

Abstract*. This paper deals with evaluation of a new scale, REPAC, for measuring academic reputation with an academic capitalism approach. The scale involves a reduced number of items to measure important aspects of academic reputation, such as academic performance, credibility, student career orientation, and external visibility. The REPAC scale is applied to data from two Italian universities, one public and the other private. Invariance of the construct measurements between universities is tested with statistical methods.*

Keywords*: Academic reputation, Measurement invariance, Scale validation*

1. INTRODUCTION

In the last few decades, several changes have taken place in the European university system, driving universities to adopt new paradigms of management. In Italy, this process is not fully complete, but some innovations have taken root, such as attracting external funding, reforming degree programmes and recruiting professors and lecturers. In other words, the management of Italian universities is becoming more like management in private markets, with students representing the customers, families and corporations the intermediate stakeholders, and universities the corporation itself. In the literature (Slaughter and Leslie, 1997), this approach is known as *academic capitalism*.

Italian universities have begun a virtuous cycle based on competition, in which teaching quality, student satisfaction, student retention, overall university performance (Agasisti and Dal Bianco, 2009) and academic reputation (AR) represent the most important aspects of effectiveness. AR might be analysed by adopting corporate paradigms, but the nature and definition of the structure still seems unclear (Delgado *et al.*, 2011). The lack of an accurate definition may be due to the multidimensional nature of university reputation, and, indeed, several dimensions are involved.

[1] Emma Zavarrone, email: emma.zavarrone@iulm.it.

In a recent literature review of strategic management, Barnett *et al.* (2006) proposed three definitions of reputation: reputation as a state of awareness, reputation as an assessment and reputation as an asset. From the academic perspective, only the definition of reputation as an assessment has been adopted, while the remaining two definitions have not been considered yet. AR should be considered also a strategic asset and measured using apposite tools.

For this purpose, either a formative or a reflective approach can be applied. The formative approach allows studying what reputation (or other latent constructs) is composed of, whilst the reflective approach is oriented to capture the observed variables governed by latent constructs (Bollen and Lennox, 1991; Diamantopoulos and Winklhofer, 2001).

This study aims to validate, through a reflective approach, a multi-item scale (hereinafter REPAC), as proposed in Zavarrone and Romenti (2012; hereinafter ZR), which is suitable for measuring academic reputation from the students' perspective in order to supply a simple instrument for university management.

The hypotheses of this study are:

H_1: Is the REPAC scale a valid scale for measuring AR?

H_2: Does reputation measurement depend on evaluators' characteristics?

The paper is organised as follows: Section 2 presents the literature review on academic reputation, Section 3 presents the data, Section 4 presents the main results of analyses, and Section 5 presents final considerations.

2. ACADEMIC REPUTATION AND STATISTICAL APPROACHES

From an academic market perspective, competition among universities is based on undergraduate reputation. Recruitment of the best faculty and staff, joint development of research projects with companies and selection of students are all keystones of academic reputation. Universities, in order to be appealing, must communicate their economic, teaching and research performance, but, in some cases, this communication is inadequate or even non-existent.

Given such asymmetric information, league tables can influence students' decision-making process. Rankings of universities can help students choose and can provide a benchmark for competing universities. However, rankings do not always reflect academic reputation because of arbitrariness of rules for their construction (Hand, 2004; Saisana *et al*., 2011). For example, in an exercise conducted by Filinov and Ruchkina (2002) ranking four universities, the removal of one university changed completely the ranks of the remaining three. This simple exercise suggests that ranking construction should be based on robust criteria (Noorda, 2011), should reflect the concepts to be measured (validity approach) and

should adopt methodologies appropriate to varying contexts (reliability approach).

Student characteristics, number and type of study programmes vary among universities, so they should not be measured by a static set of indicators, because if they were, some of those indicators would overestimate and some would underestimate the reputation for some universities.

Instead of using a ranking approach, academic reputation can be analysed from a subjective point of view (intangible models). Astin (1970) assimilated reputation into the prestige function of educational programme quality (positively) and external visibility (negatively). Gruning (1997) highlighted the effect of student selection (lower acceptance rates) on academic reputation. Arpan *et al*. (2003) proposed a scale oriented to measure the global image of a university. The scale thus constructed showed a high internal reliability as measured by the Cronbach's α coefficient. Academic and athletic ratings along with news coverage were reputation drivers for students, while news coverage, education level and general sports support (last two with negative signs) were drivers for adults. The relevant result of this study was that academic reputation dimensions increased as the number of stakeholders increased.

Helm (2007) achieved similar results for the corporate sector. Ressler and Abratt (2009) proposed a theoretical framework aimed at identifying beliefs, attitudes, intentions and behaviours for each class of stakeholders referred to in the Arpan scale. Yet, Van Vught (2008) used the concept of image (of quality, of influence, of trustworthiness) for stakeholders to identify reputation in higher education stating that reputation is the subjective reflection of the various actions an institution undertake to create an external image. However, external image does not indicate the quality of the higher education system, even if these dimensions are inter-related.

Alessandri *et al.* (2006) borrowed concepts and methods from corporate reputation literature to analyse the link between visual image and academic reputation among undergraduate students. These researchers defined the AR as collective representation held over time by multiple constituents of the university and found that the quality of internal and external performance and people's emotional engagement are dimensions of academic reputation.

In the last decades, university ranking has attracted the attention of various stakeholders as well. A causal link can exist between ranking and AR, as shown by Rindova *et al.* (2005) in a study on a business school. The university ranking influenced perceived quality and prominence, two dimensions of reputation. Boyd *et al.* (2010) proposed a new model in which the prominence construct is separated from the reputation one. In Boyd's study, prominence referred to the degree of

visibility of an organization and reputation referred to an assessment of being good or bad. Boyd's findings showed that AR could be influenced by academic prominence and could guarantee graduates a place in the job market with a chance at a good salary.

3. DATA

ZR proposed the following as dimensions that might influence AR evaluation from the students' point of view: academic prestige, credibility, external visibility, career guidance, service quality, governance and image. A new scale, named REPAC, also was proposed. REPAC's construction was based on the first five steps of Churchill's paradigm (1979): a) specifying domain of context; b) generating a sample of items; c) collecting data; d) refining measurement through several techniques; e) collecting new data. REPAC's latest version uses 39 items with six semantic labels together with scores from 0 to 10. The labels overlap in conveniently clustered scores, creating an ordinal scale.[2] The REPAC scale was administered to a sample (n = 1628) of students in the first level of the Economics degree programme at Milano-Bicocca University (MBU) between October 2009 and January 2010. Data cleaning operations operations reduced the sample size to 613 students. The sample size was divided into three smaller samples, B1, B2 and B3, corresponding to the first, second and third year of university courses, respectively. First-year courses had 262 subjects, with a male rate of 55%. The second year students were 180 with a higher presence of male students (65%). Third-year courses also had the same male rate (65%). Student allocation to each sample was based on date of birth and year of registration at the university. The same questionnaire was administered to second-year students in the first level of the degree programme at Libera Università di Lingue e Comunicazione (IULM) in Milan (n = 250, 42% male). This university was chosen for the following reasons:

1. Different age (IULM was instituted in the 1960s while MBU was instituted in the 1990s).
2. Different population of students (high percentage of females at IULM).
3. Different type of university (MBU is a public, whilst IULM University is private).

---

[2] Relevance is measured with the following scale.

Not at all	Very low		Low		Enough		High		Very high	
0	1	2	3	4	5	6	7	8	9	10

4. Different curriculum (there is a high level of quantitative programmes in Economics at MBU and a low level of quantitative programmes at IULM).

4. RESULTS

4.1 H1: VALIDATING REPAC SCALE

The first hypothesis test concerned REPAC scale validation in according to the reflective paradigm of the scale development. This approach seemed to be appropriate from the perspective of effectiveness evaluation. To test this hypothesis item analysis and covariance structures (Jöreskog, 1978) were applied. Item analysis reduced the number of items proceeding from the discrimination coefficient to the total correlation coefficient and Cronbach's alpha. The covariance structure identified the latent structure of the data using an exploratory factor analysis (EFA). For validation, confirmatory factor analysis (CFA) was used.

We conducted a preliminary analysis by applying item analysis and item normality investigation on the B2 sample. The item analysis identified those items that had not been screened sufficiently or that were correlated. This analysis allowed reduction of the ZR items from 39 to 15 (Tab. 1).

Table 1: Descriptive statistics for selected items in B2 sample

Items	Label	Mean	SD	Item.total correlation
dom1	Presence of faculty celebrities	5.772	2.564	0.603
dom2	Presence of important scientific scholars	6.244	2.527	0.609
dom3	International/ National awards	6.577	2.368	0.629
dom12	Library services efficiency	7.433	1.661	0.584
dom14	Orientation services efficiency	7.361	1.928	0.688
dom15	Scholarship granting efficiency	7.577	2.196	0.685
dom16	Qualified bachelors instruction	8.750	1.790	0.760
dom17	Market oriented curricula	8.455	1.874	0.765
dom18	Competitive advantage compared to other universities	8.222	2.206	0.680
dom29	Faculty staff availability	7.811	1.786	0.699
dom30	Students assessing transparency	7.977	1.739	0.705
dom31	Chosen path continuity insurance, regardless of reforms	7.805	1.960	0.733
dom37	Stakeholder information communication about organizations and business conventions and partnerships	6.388	2.017	0.601
dom38	Stakeholder information communication about foreign universities conventions and partnerships	7.016	1.906	0.664
dom39	Stakeholders communication about training provisionn	6.988	1.960	0.656

The normality of item distribution was investigated using the rules of Curran *et al*. (1996): if kurtosis is greater than ± 7 and skewness is greater than ± 2, then the distribution is not normal. Tab. 2 reports extreme values of kurtosis and skewness. It is evident that Curran's rules are satisfied for all samples.

Table 2: Range of kurtosis and skewness values for the samples

Sample	Kurtosis		Skewness	
	Lower limit	Upper limit	Lower limit	Upper limit
B2	-0.736	3.272	-1.785	0.097
Iulm	-1.293	1.385	-1.196	0.442
B1	-0.649	2.189	-1.363	-0.011
B3	-1.127	1.195	-1.221	0.100

EFA, resulting from application of maximum likelihood method and an oblique rotation[3], explains 63% of the total variability for four latent dimensions (Tab. 3).

Table 3: Rotated factor loadings in B2 sample

Item	Factor1	Factor3	Factor2	Factor4	Communalities	Uniqueness
dom1	-0.18	-0.05	0.89	0.15	0.74	0.26
dom2	0.28	0.13	0.51	-0.17	0.45	0.55
dom3	0.03	-0.06	0.99	-0.10	0.87	0.13
dom12	0.25	0.03	0.04	0.31	0.32	0.68
dom14	0.54	-0.08	-0.05	0.32	0.56	0.44
dom15	0.45	0.16	0.02	0.12	0.44	0.56
dom16	0.86	0.04	0.01	-0.05	0.71	0.29
dom17	0.93	-0.03	-0.06	0.00	0.78	0.22
dom18	0.79	-0.06	0.00	-0.01	0.55	0.45
dom29	0.06	0.02	-0.03	0.74	0.63	0.37
dom30	0.03	-0.08	-0.04	0.92	0.76	0.24
dom31	0.00	0.16	0.11	0.60	0.58	0.42
dom37	-0.04	0.83	-0.07	0.00	0.61	0.39
dom38	-0.11	0.83	0.10	0.03	0.69	0.31
dom39	0.08	0.91	-0.12	-0.07	0.75	0.25

[3] We applied oblique rotation because we expected the latent dimensions to be correlated.

Twelve items were selected using the guidelines suggested by Comrey and Lee (1992) for item selection. They recommend choosing factor loadings higher than 0.50 for a good item selection. The four latent dimensions are: students career orientation (SCO), which includes three items (16, 17 and 18); external visibility (EV), which contains three items (37, 38 and 39); academic prestige (AP), which refers to three items (1, 2 and 3), and credibility (CR), with three items (29, 30 and 31). The goodness of fit was investigated through the residual root mean square (RMSR = 0.02), which indicates an almost perfect fit. These results were obtained with the software product PSYCH (Revelle, 2010) implemented with R software.

The four-factor structure identified with the EFA in the B2 sample was tested using confirmatory factor analysis (CFA) both on the B2 and on the IULM samples.

The data model is (Jöreskog, 1978):

$$\mathbf{X} = \tau_X + \Lambda_X \xi + \delta_X \,, \tag{1}$$

where **X** denotes the items matrix (*p* observed variables), τ_X is the intercept matrix, Λ_x is the factor loading matrix, ξ is the scores vector of exogenous latent variables and δ_X is the measurement error matrix on **X**. Assuming that $E\left(\xi, \delta_x\right) = 0$, the covariance structure (Σ_{XX}) of the observed and latent variables becomes:

$$\sum_{\mathbf{XX}} = \Lambda_{\mathbf{x}} \Phi \Delta_{\mathbf{x}}^{'} + \Theta_{\delta} \tag{2}$$

where $\mathbf{\Phi}$ is the variances-covariance matrix of exogenous latent dimensions and $\mathbf{\Theta_\delta}$ is the matrix of measurement errors on **X**. The procedure for the parameters estimate depends on the nature of the data. For the REPAC data, a maximum likelihood was chosen. The LAVAAN package (Rosseel, 2011) was used to estimate CFA models. The latent dimensions that emerged at the exploratory stage were confirmed by the CFA, but the number of items was modified. M1 refers to testing the exploratory structure, but the fit index was not adequate. For this reason, we checked another latent structure (M2) apt to reduce the items. M2 presented a fit index value better than M1 (Tab. 4), so we used the latter. Note that even after reduction in the number of items, the model is identifiable while applying the *t-rule* for two indicators (Bollen, 1989).

Table 4 : Goodness of fit indices in the B2 and the IULM samples

Model	Sample	RMSEA	SRMR	CFI	χ^2	df	*p-value*
M1	B2	0.064	0.061	0.971	83.707	48	0.001
	IULM	0.057	0.041	0.979	86.618	48	0.001
M2	B2	0.032	0.026	0.996	16.63	14	0.276
	IULM	0.051	0.028	0.988	23.34	14	0.431

Figg. 1 and 2 present the path diagrams obtained from the QPLOT package. Comparing the two standardised structures, the weights of the single dimensions have the same order of importance. It can be concluded that the factorial structure in B2 is well supported by the IULM sample.

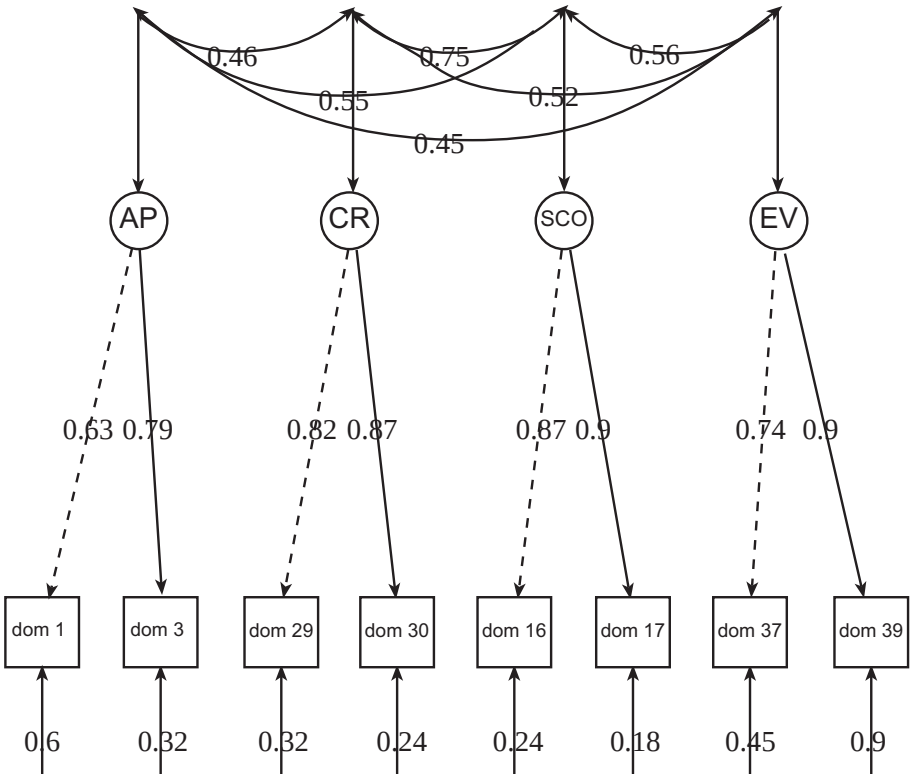

Figure 1: Path diagram of the B2 sample (standardized value)

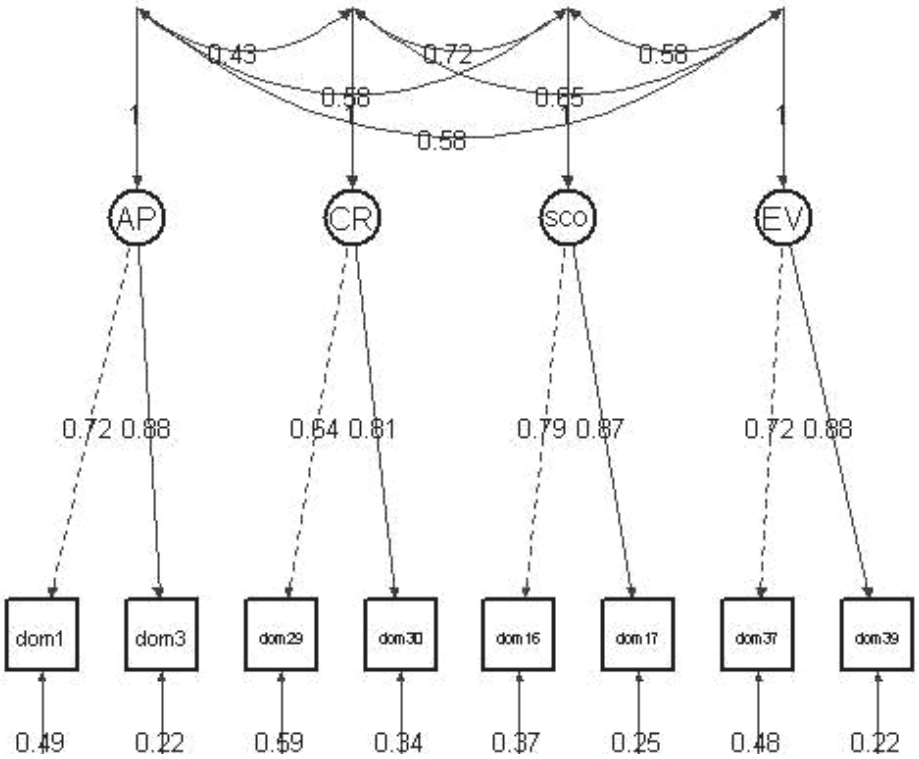

Figure 2: Path diagram of the IULM sample (standardized value)

According to Ping (2004), the validity of a scale can be determined through content and construct validity criteria. As is well known, construct validity includes forms of discriminant and convergent validity. The test of discriminant validity compares the AVE values for pairs of constructs to the squared correlation coefficient of the constructs. The AVE measures the amount of variance captured by a common construct (Walsh and Beatty, 2007). It is represented as:

$$AVE = \sum_{i=1}^{n} L_i^2 \Big/ n \quad , \tag{3}$$

where L_i indicates the standardised factor loadings and n the number of items. When the AVE index is equal to 0.50 or higher, convergent validity is proven (Fornell and Larcher, 1981). For the four REPAC dimensions, the AVE returns

values higher than 0.50 for both samples (Tab. 5). These results were confirmed by the construct reliability computed from the squared sum of the L_i and the sum of the error variance terms (*ei*) for each construct as:

$$CR=\left(\sum L_i\right)^2 \Big/ \left(\sum L_i\right)^2 + \left(\sum e_i\right)^2, \quad i=1,\ldots,4. \tag{4}$$

Table 5: AVE and reliability constructs for the B2 and the IULM samples

Constructs	B2 sample		IULM sample	
	AVE	**CR**	**AVE**	**CR**
AP	0.510	0.673	0.645	0.782
CR	0.728	0.836	0.538	0.696
SCO	0.787	0.881	0.690	0.816
EV	0.672	0.886	0.651	0.782

A rule of thumb indicates that for values of CR equal to or larger than 0.70 there is good construct reliability. Squared correlations for the B2 sample are in the range of 0.45 to 0.70, and for the IULM sample, they are in the range of 0.40 to 0.64. Hence, each construct truly is distinct from other constructs.

The AR constructs matched for both universities. The comparison of the samples demonstrated that although the universities might have different images, the dimensions affecting reputation perception are the same.

4.2 H_2: MEASUREMENT INVARIANCE ACROSS UNIVERSITIES

The second hypothesis, H_2, was based on the absence of differences in the construct for academic reputation, irrespective of the examined universities or of students' characteristics. The hypothesis was evaluated on a CFA multi-group with a progressive limit of five models. If the null hypothesis for the five models were not rejected, it would confirm that academic reputation construction was not affected by instruments or by the universities concerned. Thus, the REPAC scale can be used to carry out effectiveness evaluations.

To test this hypothesis, the B1 and B3 samples were used. For these two samples, the factorial structure of the first hypothesis (H1) was demonstrated. Tab. 6 contains statistics covering goodness of fit according to this hypothesis.

Table 6: Comparison of the goodness of fit index for B1 and B3 samples

GOF	Lavaan	
	B1	**B3**
RMSEA	0.042	0.000
SRMR	0.027	0.018
CFI	0.991	1
χ^2	20.589	12.027
p-value	*0.113*	*0.604*
df	14	14

Measurement invariance (Meredith, 1993; Marsh, 1994) was assessed through multi-group factor analysis (Formulae (1) and (2)). The expectations of Equations 1 can be expressed as:

$$\boldsymbol{\mu} = \boldsymbol{\tau}_x + \boldsymbol{\Lambda}_x \boldsymbol{\kappa} \tag{5}$$

where $\boldsymbol{\kappa}$ denotes the mean vector of latent variables.

The complete covariance structure has five matrices: $\boldsymbol{\tau}_x, \boldsymbol{\Lambda}_x, \Phi, \Theta, \boldsymbol{\kappa},$ with

fixed, free and constrained parameters. Measurement invariance consists of applying this covariance structure across the samples (g), testing different hypotheses on the five classes of parameters. In other words, multi-group factor analysis presents the analytic form:

$$\Sigma_{XX}^{g} = \boldsymbol{\Lambda}_X^{g} \boldsymbol{\Phi}^{g} \boldsymbol{\Lambda}_x^{'} + \boldsymbol{\Theta}_x^{g} \tag{6}$$

$$\boldsymbol{\mu} = \boldsymbol{\tau}_X^{g} + \boldsymbol{\Lambda}_X^{g} \boldsymbol{\kappa}^{g} \tag{7}$$

where Σ_{XX}^{g} is the matrix of variances and covariances among the p items in the *g-th* group. The analysis was based on the following hypotheses:

1. The CFA was the baseline model, in which the same pattern of fixed and free factor loadings across groups were tested, but without equality constraints (*Configural invariance*).
2. The CFA model was tested by constraining the factor loadings of the same items. The cross-group relationship differences among observed variables were attributable to the cross-group relationship differences among latent dimensions (*Metric or weak invariance*).

3. Factor loadings and intercepts were constrained to be equal across groups. The covariance group differences among observed variables and in its means were attributable to covariance group differences and means on latent dimensions (*Strong invariance*).
4. Factor loadings, intercepts and residual variances were constrained to be equal across groups (*Strict invariance*).
5. Factor loadings, intercepts, residual variances and means were equal across groups (*Complete invariance*).

In the case of no rejection of the null hypothesis for all tested samples, the academic reputation construct was not affected by students' characteristics or by different universities; therefore such a scale can be used to assess effectiveness.

The results (Tab. 7), illustrate that all possible kinds of invariance are verified, starting from the simplest to the most complex form of invariance. All *p-values* connected to all kinds of considered models are insignificant as long as the RMSEA values are less than 0.05. Therefore, we conclude that hypothesis H_2 is verified by implication. In other words, we can state that the AR construct does not vary across universities.

Table 7: Summary of measurement invariance results

Model of equivalence	χ^2	df	*pvalue*	CFI	RMSEA
Configural	32.615	28	0.250	0.997	0.027
Weak	33.161	32	0.410	0.999	0.013
Strong	35.363	36	0.499	1.000	0.000
Strict	47.493	44	0.332	0.998	0.019
Complete	60.912	48	0.100	0.992	0.035

5. CONCLUSIONS

The reputation of a university can be considered a pillar of its academic management. Reputational importance tends to be limited by reduction of resource availability (public funds, number of enrolled students, lecturers and professors, academic staff and employees) and by social innovation (new universities, new rules for recruiting professors/students/employees and new jobs). Universities are becoming business units, and students are becoming customers, willing to move from one university to another, like unsatisfied shoppers. A good reputation can be appealing, while a bad reputation can cause failure; this means that academic reputation can start either a virtuous or a vicious circle. In this scenario, a bad

reputation can be a dangerous driver for a university, and it should be monitored.

There is no agreement on academic reputation measurement, so monitoring academic reputation becomes alchemy. Customers' evaluations have not been measured yet, so this dimension is neglected. Methodological choices for indicator construction cannot be suitable for all universities if they cannot capture the nature of university structures. For instance, it is not adequate to measure the quality of research in the humanities through number of patents.

The academic reputation framework drawn from the strategic-managerial literature presents strengths and weaknesses. Strengths are: a) opinion of the most reputation-sensitive stakeholder students; b) academic reputation, which is assimilated to a latent continuous variable; c) adoption of reflective models. All these aspects contribute to generalising 'client' evaluations and define a piece of the global model of the AR, which includes families, institutions and corporations. Weaknesses are: a) unshared methodological choice of AR construction; b) ranking systems in which validity and reliability questions are unsolved; c) similar definitions investigated in different dimensions (i.e. image *vs*. external visibility *vs*. global image); d) lack of a shared measurement scale; e) use of statistical procedures, which is not always suitable.

From a statistical perspective, analysis of the literature has pointed out some potential improvements:

1. Use of a sample size suitable to the hypothetical test formulation.
2. Application of probabilistic samples instead of convenient samples.
3. Test of sampling distributions to apply methods based on the multivariate Gaussian hypothesis.
4. Test of reliability based on new measures, not just on Cronbach's alpha, which underestimates reliability (see Smith, 1974).
5. Investigation of the validation scale

This study proposes a new scale that generalises known university reputation measures. The scale, called REPAC, has a reflective approach based on students' points of view, and it arises from the hypothesis that reputation is a latent, multidimensional construct inherent to each student. From this hypothesis, a validated scale of university reputation was proposed for comparing two Italian universities.

We showed that REPAC allows for identification of academic prestige, career student orientation, university credibility and external visibility as dimensions playing roles in academic reputation. This result underlined that evaluation of university reputation comes from the job market and the society at large and can be considered an alternative solution when evaluating effects of the educational

process. Another result of this study regards the fact that all four latent dimensions (SCO, EV, AP and CR) are common to all groups of students.

This study contains two important implications: the first one of a managerial-strategic nature and the second related to aspects of measurement. The managerial-strategic implication is the lack of difference between private and public universities and the necessity to evaluate reputation from the students' perspective. This indication is very important in university rivalries: competition will depend on how analysed dimensions are treated over time. For this reason, periodic investigation of students' opinions through the REPAC or a similar reputation scale could be very useful.

The measurement implication relates to the use of the four latent variables i.e.:

1) an evaluation map (heatmap style) could be an easy tool for the faculties and staff, or
2) a complex AR indicator based on second order factorial model for comparing the university's courses.

REFERENCES

Agasisti, T. and Dal Bianco, A. (2009). Reforming the university sector: Effects on teaching efficiency. Evidence from Italy, *Higher Education*, 57(4): 477-498.

Alessandri, S.W., Yang, S.U. and Kinsey, D.F. (2006). An integrative approach to university visual identity and reputation, *Corporate Reputation Review*, 9(4): 258-270.

Arpan, L.M., Raney, A.A. and Zivnuska, S. (2003). A cognitive approach to understanding university image, *An International Journal*, 8(2): 97-113.

Astin, A.W. (1970). How colleges are rated, *Change*, 2(1): 11-86.

Barnett, M., Jermier J.M. and Lafferty B.A. (2006). Corporate reputation: The definitional landscape, *Corporate Reputation Review*, 9(1): 26-38.

Boyd, B.K., Bergh, D.D. and Ketchen, D.J. (2010). Reconsidering the reputation - Performance relationship: A resource-based view, *Journal of Management*, 36(3): 588-609.

Bollen, K.A. (1989). *Structural Equations with Latent Variables*. Wiley Series in Probability and Mathematical Statistics, Wiley, New York.

Bollen, K.A. and Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective, *Psychological Bulletin*, (110): 305-14.

Churchill, A.G. (1979). A paradigm for developing better measures of marketing constructs, *Journal of Marketing Research*, (11): 64-73.

Comrey, A.L. and Lee, H.B. (1992). *A First Course in Factor Analysis*. Hillsdale, New Jersey: Erlbaum.

Curran, P.J., West, S.G. and Finch, J.F. (1996). The robustness of test statistics to nonnormality and specification error in confirmatory factor analysis, *Psychological Methods*, 1(1): 16-29.

Delgado-Màrquez, B.L., Hurtado-Torres, N.E. and Bondar, Y. (2011). Internationalization of higher education: Theoretical and empirical investigation of its inßuence on university institution rankings, *Revista de Universidad y Sociedad del Conocimiento* (RUSC), 8(2): 265–284.

Diamantopoulos, A. and Winklhofer, H.M. (2001). Index construction with formative indicators: An alternative to scale development, *Journal of Marketing Research*, (38): 269-277.

Filinov, N. B. and Ruchkina, S. (2002). The ranking of higher education institutions in Russia: Some methodological problems, *Higher Education in Europe*, 27(4): 407-21.

Fornell, C. and Larcker, D.F. (1981). Evaluating structural equation models with unobservable variables and measurement error, *Journal of Marketing Research*, 18(1): 39-50.

Grunig, S.D. (1997). Research, reputation, and resources: The effect of research activity on perceptions of undergraduate education and institutional resource acquisition, *Journal of Higher Education*, 68(1): 17–52.

Hand, D.J. (2004). *Measurement Theory and Practice: The World through Quantification*. Edward Arnold, London.

Helm, S. (2007). One reputation or many? Comparing stakeholders, perceptions of corporate reputation, *Corporate Communications An International Journal*, 12(3): 238-254.

Jöreskog, K.G. (1978), Structural analysis of covariance and correlation matrices, *Psychometrika*, (43): 443-477.

Marsh, H.W. (1994). Confirmatory factor models of factorial invariance: A multi-faceted approach, *Structural Equation Modeling*, (1): 5-34.

Meredith, W. (1993). MI, factor analysis and factorial invariance, *Psychometrika*, (58): 525-43.

Noorda, S. (2011). Future business schools, *Journal of Management Development*, 30(5): 519-525.

Ping, R.J. (2004). On assuring valid measures for theoretical models using survey data, *Journal of Business Research*, (57):125-141.

Ressler, J. and Abratt, R. (2009). Assessing the impact of university reputation on stakeholder intentions, *Journal of General Management*, 35(1): 35-46.

Rindova, V.P., Williamson, I.O., Petkova, A.P. and Sever, J.M. (2005). Being good or being known: An empirical examination of the dimensions, antecedents, and consequences of organizational reputation, *Academy of Management Journal*, 48(6): 1033-1049.

Rosseel, Y. (2011). Lavaan: latent variable analysis. R package version 0.4-10, URL http://CRAN.R-project.org/package=lavaan, Last access: 31/01/2011.

Saisana, M., d'Hombres, B. and Saltelli, A. (2011). Rickety numbers: Volatility of university rankings and policy implications, *Research Policy*, 40 (1): 165-177.

Slaughter, S. and Leslie, L.L. (1997). *Academic Capitalism: Politics, Policies, and the Entrepreneurial University*. The Johns Hopkins University Press, Baltimore.

Smith, K.W. (1974). On estimating the reliability of composite indexes through factor analysis, *Sociological Methods & Research*, (2): 485–510.

van Vught, F. (2008). Mission diversity and reputation in higher education, *Higher Education Policy*, (21): 151-174.

Walsh, G. and Beatty, S.E. (2007). Customer-based corporate reputation of a service firm: Scale development and validation, *Journal of the Academy of Marketing Science*, 35(1): 127-143.

Zavarrone, E. and Romenti, S. (2012). Un approccio riflessivo alla misurazione della reputazione accademica: Il modello REPAC. In: Civardi M. (Ed.), *Modelli e metodi per valutare la reputazione di strutture formative*, Cleup, Padova: 69-117.