Human Capital Growth for University Education Evaluation

Pietro Giorgio Lovaglio, Giorgio Vittadini[1]

Statistics Department, University of Milan-Bicocca, Italy

Summary. In this paper, we propose a method for the evaluation of relative and impact external efficiency of university studies as effects of education on the long-term income of graduates. In order to evaluate the "ceteris paribus" levels and growth we applied a multilevel longitudinal model with random effects. The variables used in our analysis refer to Human Capital studies. The Bank of Italy has collected the data used for the analysis in the years 1998, 2000 and 2002.

Keywords. External efficiency, Human capital, Multilevel growth model.

1. The external efficiency of university studies

The organizations that distribute services of public utility, as managers of collective resources, must account for their performances to the funding authorities, the community, and the customers (Hanushek, 1997).

Guidelines are being delineated, at national and international levels, for the accreditation of university education and research programmes, and services for the students. The university accreditation processes should not only meet the *ex-ante* requirements, but also verify the results.

To evaluate an educational institution, we adopt the outline put forward by Lockheed & Hanushek's (1994). The inputs of the system determine whether the terms "efficiency" or "effectiveness" should be used. The outputs of the system determine whether the descriptors "internal" or "external" are to be applied to efficiency and effectiveness (Table 1).

---

[1] This paper is the result of the joint effort of both the authors. G. Vittadini took care of the final editing of Section 1, whereas P.G. Lovaglio was responsible for all the other Sections.

Table 1. Criteria of evaluation for university education

	Internal	**External**
Effectiveness	INTERNAL EFFECTIVENESS: University/type of bachelor programme effect on the student achievement	EXTERNAL EFFECTIVENESS: University/ type of bachelor programme effect on employment Short term Human Capital
Efficiency	INTERNAL EFFICIENCY : Costs-benefit analysis of the investment in higher education	EXTERNAL EFFICIENCY University/ type of bachelor programme effect on the returns of investment in higher education Long term Human Capital

The internal effectiveness and efficiency can be used to self-evaluation of universities, while the external ones may be the most proper way for evaluating the universities' performances (Elias, 2002). As of today, there are very few studies of the analysis of external efficiency of education based on the evaluation of the earned income of graduates in the vital cycle. In the following, we will define and apply the criteria for the evaluation of organizations that distribute public utility services.

The investment in higher education may be evaluated in terms of the graduates' earned income. In other words, we can evaluate the so-called Human Capital (HC) of an individual in terms of his or her expected earned income related to the competences achieved through education (Becker, 1964).

In the following, the evaluation of the university education does not coincide with the estimation of HC, but with the following aims: to evaluate "ceteris paribus" the universities' investment in HC with returns in terms of earned income (relative efficiency), and if and how the university education is a real advantage in monetary terms (impact of efficiency).

2. Gain score, added value, and multilevel growth model

In order to overcome the problem of "selection bias" (Garen, 1984) due to lack of randomization of the individuals between treatments, the effectiveness and efficiency must be evaluated *ceteris paribus*, i.e. adjusting the outcome for the different individual characteristics, university resources, and local job markets (Fitz-Gibbon, 1997; Scheerens & Bosker, 1997).

To this aim, even if a linear model may be applied, the nature of the dependent variable suggests non-linear approaches. To this purpose, we introduce the concepts of initial performance or pre-test (y_1), final status or post-test (y_2), as well as the concept of improvement or gain (y_2-y_1).

In the *gain score* approach the "adjusted outcome" is simply the expected value of gain, $E(y_2-y_1)$, whereas the added value or *residual gain score* approach[2] is concerned with the quantity $y_2-E(y_2/y_1)$, i.e. the difference between the post-test and its expected value $E(y_2/y_1)$, a result of the regression of y_2 onto y_1.

Both approaches present some drawbacks. Some researchers highlight the technical problems associated with the residual gain score approach (Cronbach & Furby, 1970; Scheerens & Bosker, 1997; Hanushek, 1997, 2002; Willett, 1988; Thum, 2002). First, it is reasonable to expect that the same set of factors affecting the pre-test would affect the post-test as well. Like the post-test, the pre-test is measured with error, so that a regression of post-test on a pre-test score violates the basic regression assumption that covariates are error-free. This implies that the residual gain score is an inefficient and inconsistent estimator of true change. Secondly, because the pre-test typically explains a large quota of the post-test outcome, this drastic reduction of residual variance limits the explicative power of other factors, for example the resources of the organization, its quality and so on.

Other researchers (Rogosa *et al.*, 1982; Rogosa & Willet, 1985; Rogosa, 1995; Singer & Willet, 2003) dispel the notion that gain scores are inherently unreliable, so demonstrating that the correlation between gain and initial status (the major drawback of the gain score approach) is merely an artefact of the design. The nature of the correlation is the outcome of the parameterization we choose for the growth factors and, as a consequence, the meaning itself we attach to them.

Thum (2002) and Bryk *et al.* (1998) propose to measure the outcomes in time series in a longitudinal framework, an approach that allows insertion of covariates related to the institutions that can affect the performances of the micro units (students, classes). The repeated measurement approach we adopt in this paper places the pre-test and the post-test on an equal footing in that both serve as outcomes in our models. The change with respect to gain scores is evident in recent research (Collins, 1996; Maris, 1998; Mellenbergh, 1999; Williams & Zimmerman, 1996).

The second development, known as "hierarchical", "multilevel", "growth" or "mixed-effects" modelling, provides the methodological complement for a proper, flexible, treatment of growth data in stratified sampling designs that are common in educational research settings (repeated measures on students over time, students nested with schools in the district, etc.).

---

[2] Its origins can be traced to notions of goods, value and wealth in political economy (Saunders, 1999); value-added analyses address a model for comparison (Goldstein & Spiegelhalter, 1996) adjusted for the factors out of control for the researchers. Researchers have frequently applied the regression analysis of the post-test score on the pre-test score, motivating this choice with the notion of "fairness" in order to flatten the playing ground (Good *et al.* 1975).

Multilevel models facilitate the study of the effects of covariates of development at every level of nesting (Laird & Ware, 1982; Bryk & Raudenbush, 1992; Goldstein, 1995). The advent of multilevel models helped accountability agencies to avoid the misleading picture conveyed by aggregate scores, such as the school or district means, and focus on student level and/or change (Meyer, 1996). Many current accountability systems adopt multilevel models (Bryk *et al.*, 1998).

A Multilevel Growth Model (MGM) analyses random effects in the sense that intercept and slopes describing individual trajectories over time vary between individuals (Bryk & Raudenbush, 1992). Such a model requires a specification of the covariance structure between observations at different times on the same subject, because it is plausible to assume that observations on the same unit are correlated (Willett, 1988).

There are two fixed effect factors, treatment and time. Random effects result from variation between subjects and from variation within subjects. Measures on the same subject at different times are usually correlated, with measures taken close in time being more highly correlated than measures taken far apart. Observations on different subjects are assumed independent, although the validity of this assumption depends on the study design. Multilevel models are used with repeated measures data to accommodate the fixed effects of treatment and time and the covariation between observations on the same subject at different times.

In a MGM, we will conceptualize the repeated measures as observations 'nested' within individuals. To start, we must again specify our model for the individual trajectory, which we will refer to as the Level 1 (or within-person) model for intra-individual change:

$$y_{tij} = \pi_{0ij} + \pi_{1ij}\, time_{tij} + r_{tij} \ , \tag{1}$$

where: y_{tij} represents the *educational outcome* at time t ($t=1, .., k$) for the i-th subject ($i = 1,…, n_j$) nested in the j–th institution (university, school, type of programme) ($j = 1,…, m$); $time_{tij}$ is the variable time having cardinality equal to the number of instants; π_{0ij} and π_{1ij} are the intercept (mean of outcome at $time_{tij}=0$) and slopes (rate of increase over time) for subject ij. In equation (1), where for the sake of simplicity we did not insert covariates varying over time, it is assumed that the outcome follows a linear trend.

We also specify a Level 2 (or between-person) model for inter-individual differences. The Level 2 model decomposes the individual intercept and slopes (π_{0ij} and π_{1ij}) into the mean values of the parameter institutions across individuals (β_{00j} and β_{10j}), in the effect of a set of individual characteristics (x_{sij} with α_{sj} associated parameter) of subject ij and the individual effect (u_{0ij} and u_{1ij}).

Finally, the Level 3 model decomposes the mean effect of institutions into the mean values of the parameters across all institutions (γ_{000} and γ_{100}), in the

effect of a set of characteristics (g_{rj} with associated parameter λ_r) of institution j and institution effect (m_{00j} and m_{10j}):

$$\pi_{0ij} = \beta_{00j} + \Sigma_s \alpha_{sj} x_{sij} + u_{0ij} \quad \pi_{1ij} = \beta_{10j} + \Sigma_s \alpha_{sj} x_{sij} + u_{1ij}, \tag{2}$$

$$\beta_{00j} = \gamma_{000} + \Sigma_r \lambda_r g_{rj} + m_{00j} \quad \beta_{10j} = \gamma_{100} + \Sigma_r \lambda_r g_{rj} + m_{10j}. \tag{3}$$

Each equation presents errors at all nested levels: r_{tij} is the random effect of time t on the trajectory of outcome evolution for subject ij; u_{0ij} and u_{1ij} are the random effects of subject ij with respect to the mean trajectory for the subjects nested in institution j; m_{00j} and m_{10j} are the random effects of institution j with respect to the mean trajectory for all institutions.

The distribution of the time specific disturbances, individual trajectory parameters and institutions' specific parameters are assumed normally distributed, with zero expectation, and mutually independent:

$$\begin{bmatrix} \mathbf{r_{tij}} \\ \mathbf{u_{\cdot ij}} \\ \mathbf{m_{\cdot\cdot j}} \end{bmatrix} \sim \mathbf{N} \begin{pmatrix} 0 \\ 0, \\ 0 \end{pmatrix} \begin{bmatrix} \mathbf{\Sigma} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{G} \end{bmatrix} \tag{3}$$

where $\mathbf{\Sigma}$ is the (k,k) covariance matrix of the time-specific disturbances (r_{tij}), $\mathbf{T}$ the $(2,2)$ covariance matrix of the individual trajectory parameters ($\mathbf{u_{\cdot ij}}$) and $\mathbf{G}$ the $(2,2)$ covariance matrix of the trajectory parameters between institutions residuals ($\mathbf{m_{\cdot\cdot j}}$). Even if MGMs neglect the specification of the covariance between observations at different times on the same subject, a special attention must be paid to $\mathbf{\Sigma}$ otherwise we risk incorrect conclusions on fixed effects (Rogosa, 1995; Laird & Ware, 1982; Cnaan *et al.*, 1997).

The model specified with formulas (1), (2) and (3) is applicable for estimating the specific effect of institutions on the level (in a temporal instant, m_{00j}) and on the rate of outcome increase over time (m_{10j}). The model can be extended (Lovaglio, 2004) to the case of the evaluation of the efficiency among institutions (relative evaluation) and to that of the distributed service (impact evaluation).

In the latter case, the model allows to check if an institution that distributes a service (higher education) obtains better performances in comparison with a (control) group of subjects with similar characteristics to whom the service was not distributed. To this aim, a dummy variable I_{ij} – which assumes value 0 for subject i of the control group and 1 otherwise – is inserted in equations (2) and (3) for the estimation of the impact level (κ_0) and slope (κ_1):

$$\boldsymbol{\pi}_{0ij} = \beta_{00j} + \Sigma_s \alpha_{sj} x_{sij} + \boldsymbol{\kappa_0} \, \boldsymbol{I_{ij}} + u_{1ij} \qquad \boldsymbol{\pi}_{1ij} = \beta_{10j} + \Sigma_s \alpha_{sj} x_{sij} + \boldsymbol{\kappa_1} \, \boldsymbol{I_{ij}} + u_{1ij} \tag{5}$$

$$\beta_{00j} = \boldsymbol{\gamma}_{000} + \Sigma_s \boldsymbol{\lambda}_r g_{rj} + I_{ij} \, m_{00j} \qquad \beta_{10j} = \boldsymbol{\gamma}_{100} + \Sigma_s \boldsymbol{\lambda}_r g_{rj} + I_{ij} \, m_{10j} \tag{6}$$

3. The external efficiency of Italian universities

The proposed methodology was applied to the evaluation of the external efficiency of Italian university education using the data collected by the Bank of Italy with the Survey on Household Income and Wealth (Banca d'Italia, 2002) for the years 1998, 2000 and 2002.

The dependent variable is the earned income for each individual in a time series and the covariates are individual factors (social-demographic-economic background), university covariates (input, resources, intangible aspects, personnel) and job market variables (region, province, etc. where the subject is addressed).

Several indicators concur to identify the individual human capital and household socio-economical status: Age, Gender, Region of residence, Marital status, Educational level, Employment status, Type of job, Economic activity sector, Years of schooling, Age of entrance in the labour market, Number of children; Years of full-time work, Years part-time work, Household total wealth, Household total debt, Household income, and Parents' educational level.

The outcome of the analysis is the partial disposable earned income (from here on *earned income*) composed of earnings of employed and self-employed people, pensions, transfers and economic assistance.

The sample was composed of subjects belonging to the labour force with positive earned income for all three years; we identified the graduates as individuals who had achieved a university bachelor or master degree before 1998.

The panel of income earners is composed of less than 2000 subjects, whereas the group of graduates of 559 individuals (Table 2). Because of the reduced number of panel graduates, it is not possible to deepen our analyses on the external efficiency by university, or type of degree. The analysis is therefore limited to the effect of the higher education *tout court* on the income dynamics (impact analysis that compares the level and the growth rates of graduates and non-graduates earnings in Italy)[3].

Table 2. Composition of the Bank of Italy panel sample

Period	Families	Individuals	Income earners	Income earners Graduated after 1998
1998	7147	20901	7635	764
2000	8001	22268	8140	825
2002	8011	21148	7313	793
1998 - 2000 – 2002	2591	6953	1978	559

---

[3] Such data moreover prevent the estimation of the short-term university education effect, because we must select the graduates who have recently acquired a degree, typically going back three to five years from the first survey.

Table 3. Means of the earned income (in Euro) by degree level and time

Period	Graduate	Non Graduate
1998	17,959	13,108
2000	20,700	14,269
2002	22,024	15,429

In order to estimate income temporal trend and the between-group differences, a MGM is specified for evaluating:

- a) the temporal correlations for every subject (covariance structure);
- b) the existence of temporal dynamics for earned income trajectories (time effect);
- c) if such a trend differs between graduates and non-graduates (bachelor effect);
- d) the individual characteristics that explain the levels and the growth rates over time.

The specified longitudinal model is therefore a two-level model composed of equation (1) and (5) describing, respectively, Level 1 (time effect) and Level 2 (subject effect). We highlight that both level and growth rate for graduates are larger than for non-graduates (Table 3).

We can state that the hypotheses on which the classical approaches of repeated measures (Anova and Contrast Analysis) are based (constancy of covariance between couples of measures when the temporal lag varies, and constancy of variances in time) are too rigid. A correct modelling of the covariance structure Σ - representing it as a function of a relatively small number of parameters - is fundamental for valid inference and estimation of fixed effects (impact of the bachelor study programme on the level and growth rates of earned income). For most of the structures, the covariance between two observations on the same subject depends only on the lag between measurements with constant variance.

Estimates of between-graduate variances (diagonal elements) and correlations within groups at each period are shown in Table 4, referring to the MGM specified with the effects of treatment, time, treatment-by-time interaction, and other significant covariates (Table 5). The estimated parameters show that variances between graduates within each group are approximately equal across times and correlations decrease with increasing interval between observational times.

Table 4. Estimated correlations and variances of incomes over time

Time	1998	2000	2002
1998	0.34	0.50	0.39
2000	0.50	0.32	0.49
2002	0.39	0.49	0.39

Table 5. Effect of covariates on earned income at 1998 and on growth 1998-2002

Level (Income at 1998)	F Value	Pr > F		Growth rate	F Value	Pr > F
Intercept	173.61	<.0001		Intercept	123.56	<.0001
Gender	161.90	<.0001		t* Occupation	99.29	<.0002
Occupation	80.29	<.0002		t*Cl_age	66.34	<.0003
Bachelor	58.43	<.0003		t*Sector	23.56	<.0006
Full_job	14.65	<.0005		t*Bachelor	3.83	0.0218
Sector	8.51	<.0006		t*Gender	1.68	0.0932
Marital	4.43	0.0042		t*Full_job	1.65	0.1021
Area	3.85	0.0039		t*Area	1.15	0.1362
Cl_age	3.09	0.0150		t*Marital	1.12	0.1423

Akaike's information criterion and Schwarz's Bayesian criterion are the indices of relative goodness-of-fit chosen for comparing models with the same fixed effects but different covariance structures. The covariance structure assumed for Σ is the first order autoregressive one (the estimated correlation parameter for unitary lag is 0.373), which specifies homogeneous variance and covariances between observations that decrease toward zero with increasing lag. The test of heterogeneity indicates a common covariance structure for graduates and non-graduates.

Once Σ is structured, the longitudinal analysis requires the construction of the Unconditional Means Model (UMM), a model that does not require covariates at each level (level 1, level 2). It is useful to both estimate whether a systematic variability exists in the outcome, and show how the total variance of the outcome is decomposed between and within subjects, depending on the time or on individual factors that changed over time.

The estimated intraclass correlation (the ratio of the between and total variances) shows that the 43% of the total variability of income is attributable to the differences between the individuals (i.e. individual characteristics) and for the remaining part to the differences within the subjects, confirming the hierarchical structure (multilevel) of the dataset.

Inserting the variable time (as a continuous quantitative covariate) in the UMM we obtain the Unconditional Growth Model (UGM) that explains how much of the within subject variance is explained by time. We estimated that 28% of the within-subject variability is determined by time and the remaining part by individual characteristics that change over time (occupation, sector, etc).

Relative to the fixed effects, the overall growth rate of earned income is positive and highly significant, so indicating a strong dynamic effect on the income trajectories, while, relatively to random parameters, the variance of in-

come in 1998 (the reference time) and the variance of the growth rates - between the subjects- are significantly different from zero. The estimated covariance between the levels of incomes in 1998 and growth rates is not significant, and this implies unstable relations between slopes and initial income.

In order to explain the residual variability between the levels (earned incomes in 1998) and between the growth rates (yearly increment of income), we inserted individual covariates in the 2 level model (equation 5): some of which, for the sake of simplicity, are kept fixed in time and measured at the final year (2002).

The covariates and the significance levels of the associated parameters (fixed effects)[4] are shown in Table 5. The level of earned income at 1998 (left part of Table 5), gender, type of occupation at 2002 (Occupation), possession of a bachelor's degree (Bachelor), years of full-time job (Full_job), occupational sector at 2002 (Sector), area of residence at 2002 (Area), marital status at 2002 (Marital), and age at 2002 (Cl_age) are highly significant.

The factors that characterise the evolution of the growth rates are (right part of Table 5): type of occupation at 2002 (t*Occupation), age at 2002 (t*Cl_age), occupational sector at 2002 (t*Sector) and, to a lesser extent, the possession of a bachelor's degree (t*Bachelor), whereas the other covariates are not significant at 5% level. Relatively to the random parameters (not shown here), the residual variances in the 1998 level and growth rates remain significant.

We can conclude that income level at 1998 and income growth rates are greater and faster increasing (even though close to the limits of the customary significance levels) for graduates. The selected covariates explain a large quota (56%) of the UGM residual variance for the level and just a part (21%) of the between-subject growth rate variability, thus leaving open the search of other individual covariates that may affect the evolution of income over time. The trajectories estimated for each group of subjects (after adjustment for the selected covariates) give a mean growth annual rates of 1,051€ for a graduate and 426€ for a non-graduate.

Moreover, we estimated the means of growth rates for the two groups for the more relevant age class. By comparing the growth rates of a graduate (1507€) and a non-graduate (916€) in the age class less of 30 years it is possible to quantify how the labour market rewards university education. The impact of a university degree appears highly significant in absolute terms, while in comparison with the non-graduate rates this difference remains significant, but constant in the older age groups (606€ for 31-40 year group; 584€ for 41-50 group).

Finally, we estimated the differences between group income means at every time and for the income growth rates. The contrast analysis shows that for all

---

[4] The application of QQ plot test for Level 1 and Level 2 residuals shows that normality is supported by data.
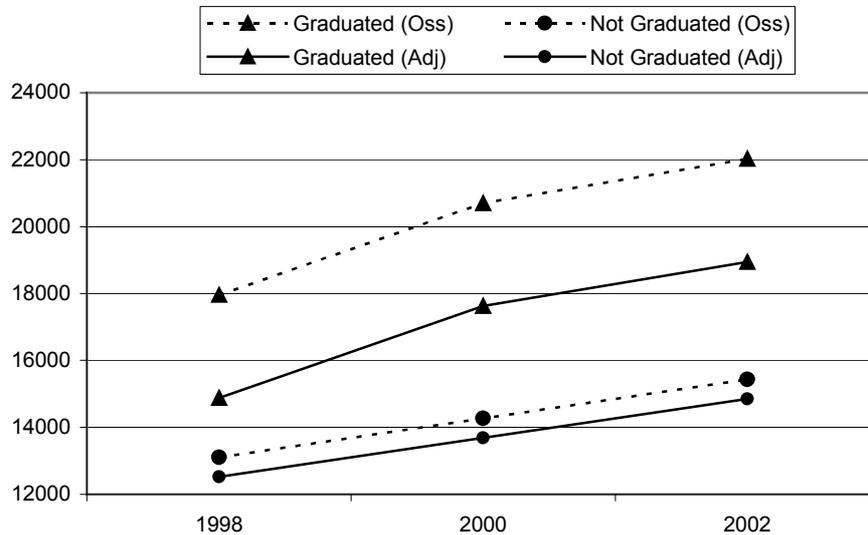
Figure 1. Observed and estimated income trajectories

three temporal moments the differences between income means are not only meaningful, but increasing in time (in particular, the difference is 1188€, 2015€ and 2118€, respectively for 1998, 2000 and 2002) and that the income growth rate of the graduates (estimate: 525€) is larger than that of non-graduates (p-value=0.0169).

Figure 1 shows the incomes trajectories for graduates and non-graduates, relatively to the average observed means (*Oss*) and to the average estimated means (*Adj*). The continuous lines, which show the income trends adjusted for the subjects' characteristics, confirm our previous results. The adjusted trajectories make it evident the "correction" effect (also called *risk-adjustment*).

4. Conclusions

The proposed methodology performs "ceteris paribus" comparisons, so allowing a better inference in the sense that the impact of the university degree is estimated net of the different group characteristics.

In fact, controlling for the possible factors of heterogeneity, the difference of the two trajectories is reduced, because graduates possess individual characteristics more coherent with the ability to produce high salaries (area of residence, household context, remunerative occupation, employment sector, etc.), as illustrated by the fact that the adjusted trajectory for graduates lowered drastically.

References

BANCA D'ITALIA (2002) I bilanci delle famiglie italiane, *Supplementi al Bollettino Statistico.*

BECKER G.S. (1964) *Human Capital*, Columbia University Press and NBER, New York.

BRYK A.S., THUM Y.M., EASTON J.Q., LUPPESCU S. (1998) Assessing school academic productivity: The case of Chicago School Reform, *Social Psychology of Education*, **2**: 103-142.

BRYK A.S., RAUDENBUSH S.W. (1992) *Hierarchical Linear Models: Applications and Data Analysis Methods*, Sage Publication, Newbury Park, CA.

COLLINS L.M. (1996) Is reliability obsolete? A commentary on \Are simple gain scores obsolete?, *Applied Psychological Measurement*, **20**: 289/292.

CNAAN A., LAIRD N.M., SLASOR P. (1997) Using the General Linear Mixed Model to analyse unbalanced repeated measures and longitudinal data, *Statistics in Medicine*, **16**: 2349–2380.

CRONBACH L.J., FURBY L. (1970) How should we measure "change" or should we?, *Psychological Bulletin*, **74**(1): 68-80.

ELIAS G. (2002) L'accreditamento del processo e la misura del risultato. In: Workshop "*Valutazione dell'Università, accreditamento del processo, misurazione del prodotto*", Franco Angeli, Università Milano Bicocca.

FITZ-GIBBON C.T. (1997) *The Value-Added National Project: Final Report. Feasibility Studies for a National System of Value-Added Indicators*, SCAA Publication, Hayes Middlesex.

GAREN J. (1984) The returns to schooling: a selectivity bias approach with a continuous choice variable, *Econometrica*, **52**, 23-54.

GOLDSTEIN H. (1995) *Multilevel Statistical Models, II ed.,* Edward Arnold, London.

GOLDSTEIN H., SPIEGELHALTER D.J. (1996) League table and their limitations: statistical issues in comparisons of institutional performances (with discussion), *Journal of Royal Statistical Society,* **159(5)**: 385-443.

GOOD T.L., BIDDLE B.J., BRODY J.E. (1975) *Teachers Make a Difference,* Holt, Rinehart and Winston, New York.

HANUSHEK E.A. (1997) Assessing the effects of school resources on economic performance, *Education Evaluation and Policy Analysis*, **19**(2): 141-164.

HANUSHEK E.A. (2002) Publicly provided education. In: AUERBACH A., FELDSTEIN M. (eds) *The Handbook of Public Economics*, IV.

LAIRD N.M. WARE J. H. (1982) Random effects models for longitudinal studies, *Biometrics*, **38**: 973-974.

LOCKHEED M.E., HANUSHEK E.R. (1994) Concepts of educational efficiency and effectiveness, *International Encyclopedia of Education, II Edition.*

LOVAGLIO P.G. (2004) Efficacia relativa e di impatto di iniziative nell'ambito dei servizi alla persona di pubblica utilità, *Statistica* (in press),

MARIS E. (1998) Covariance adjustment versus gain score-Revisited, *Psychological Methods*, **3**: 309-327.

MEYER D. (1996). *The Power of their Ideas: Lessons for America from a Small School in Harlem.* Boston: Beacon Press.

MELLENBERGH G.J. (1999) A note on simple gain score precision, *Applied Psychological Measurement*, **23(1)**: 87-89.

ROGOSA D.R. (1995) Myths about longitudinal research. In: GOTTMAN J.M. (ed), *The Analysis of Change,* LEA, Mahwah, NJ: 3-66.

ROGOSA D.R., BRAND D., ZIMOWSKI M. (1982) A growth curve approach to the measurement of change, *Psychological Bulletin*, **90**: 726-748.

ROGOSA D.R., WILLETT J.B. (1985) Understanding correlates of change by modeling individual differences in growth, *Psychometrika*, **50**: 203-228.

SAUNDERS L. (1999) A brief history of educational "value added": how did we get to where we are?, *School Effectiveness and School Improvement*, **10(2)**: 233-256.

SCHEERENS J., BOSKER R.J. (1997) *The Foundations of Educational Effectiveness*, Pergamon, Oxford.

SINGER J.D., WILLET J.B. (2003) *Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence*, Oxford University Press, Oxford.

THUM Y.M. (2002) Measuring progress towards a goal: estimating teacher productivity using a multivariate multilevel model for value-added analysis (http://www.mff.org/pubs/tchprodv3.pdf).

WILLETT J.B. (1988) Questions and answers in the measurement of change, *Review of Research in Education*, **15**: 345-422.

WILLIAMS R.H., ZIMMERMAN D.W. (1996) Are gain scores obsolete?, *Applied Psychological Measurement*, **20**: 59-69.