# Evaluation of Human Capital in an Education-Based Perspective

Emma Zavarrone and Anna Simonetto

**Abstract** This paper considers one of the intangible aspect of human capital: the university knowledge accumulation. It is relevant both for academic management and for recruitment world. In the former case it can be an useful guide to identify the characteristics of clever students, while in the latter case it can be applied to worker selection. Because of the velocity of credit acquisition is not sign of cleverness, it becomes important to analyze different aspects of university human capital accumulation. We will investigate it through latent growth modeling on administrative data come from an Italian university.

**Key words:** University human capital, Latent growth model, Gompertz curve

## 1 Introduction

Since the beginning of modern economic thought, the human capital (hereafter *HC*) and its implication on growth process have been widely analyzed. The *HC* definition can be considered an expression of its time, so it has undergone several extensions and adjustments from its first appearance. Currently, OECD [10] introduces, in the *HC* definition, the non-market activities, outlining a set of "knowledge, skills, competencies and attributes embodied in individuals that facilitate the creation of personal, social and economic well-being". The OECD definition appears the most complete one and it would seem to indicate the successful completion of the *HC* theory both in terms of national accountability and individual analysis. However some methodological problems are still unresolved, i.e. which kind of metric has to

Emma Zavarrone
Department of Communication, Behaviour and Consumption "Giampaolo Fabris", IULM University, e-mail: emma.zavarrone@iulm.it

Anna Simonetto
Department of Economics and Management, University of Brescia, e-mail: simonett@eco.unibs.it

be adopted for *HC* investments. This absence can be justified by the nature of *HC* measurement: *HC* is formed by intangible and tangible aspects so a unique measure leads to the underestimation of some components. Several types of measurement models have been proposed, based on different approaches.

We adopt the *education-based approach*, in order to quantify the training or educational paths pursued. In details, we analyzed the process of accumulation of intangible aspects of *HC* during the university using latent models. We configure a sub-dimension of *HC* given by the difference between students' *HC* before entering university and the *HC* held at graduation. We refer to the definition proposed by Civardi and Zavarrone [3] (hereafter CZ), where this sub-dimension of University *HC*, called *UHC*, can be defined as the improvement of knowledge, skills and attitudes attained thanks to the educational activities, the use of didactic structures, passed exams and social interaction with fellow students. In CZ, *UHC* has been measured through exams, considering their marks and ECTs (European University Credits). These proxies, adequately multiplied and normalized, have been cumulated for each slice of time. CZ applied this measure at five courses of Economics at the Bicocca University in Milan, and they observed that the growth of *UHC* was nonlinear. The use of variance component model has highlighted that the accumulation process of knowledge was characterized by different velocity according to discipline courses analyzed. On the same data, analogous results have been achieved using the three way analysis [6]. Bianconcini and Cagnone [1] estimated the students' performances over time using a multivariate latent growth approach. They linked some students' covariates (as gender, type and degree of Bachelor) to the characteristics of the achieved exams (number and mark),and used as manifest indicators of student performances. Their results show the effective presence of significant differences in the latent variable of performance: lowest mark at bachelor level imply worse performance in University.

We want to extend the analysis of *UHC* considering different ways in which the students acquire knowledge during the time they spent at university. Since the latent trajectory of *UHC* is not linear, it becomes important to quantify the knowledge accumulation velocity. We use a set of variables to explain these different velocities, thus constituting different clusters of characteristics of students.

Section 2.1 is devoted to the methodology, in Section 2.2 we explain the used data and report the results in Section 2.3. The conclusions and further developments are in the last section.

## 2 Analysis

### *2.1 Methodology*

In social and behavioral sciences, the change over time can be studied through the growth curves approach on repeated observations. It allows to explain the changes inter-individuals through linear and/or nonlinear functions. The literature disentangles growth curves models via multilevel and via latent variable methodologies. In the multilevel approach, a simple way to study the change over time is a two-level model in which level-1 refers to the individual growth and level-2 to the variation of growth parameters as random effects [11]. In the latent variables methodologies, the manifested change is the consequence of a latent variable change, expressed in terms of latent intercept and latent slope. The confirmatory factor analysis represents the starting point for these estimates [7]. $Y$ is the vector of repeated measures, $\Lambda$ is the matrix of factor loadings, $\eta$ is the vector of latent factors and $\varepsilon$ is the vector of residuals: $Y = \Lambda\eta + \varepsilon$. In linear cases, the latent factors ($\eta$) are intercepts ($\alpha_i$) and slopes ($\beta_i$):

$$\eta = \mu_\eta + \zeta$$

$$\begin{bmatrix}\alpha_i \\ \beta_i\end{bmatrix} = \begin{bmatrix}\mu_\alpha \\ \mu_\beta\end{bmatrix} + \begin{bmatrix}\zeta_{\alpha_i} \\ \zeta_{\beta_i}\end{bmatrix}$$

In this study, we will focus on the latent growth curves. These two parallel statistical perspectives, under restrictive hypothesis, lead to the same results [8]. We model the latent change in $UHC$ from one time to another and we use the person-parameter of the accumulation curve to identify a set of covariates that can help us to pinpoint students who accumulate knowledge more quickly and with better performances. We divide the analysis process in two stages: in the first one we estimate the latent Gompertz curve that models the acquisition of $UHC$ over time by individual students. In the second stage we use the $UHC$ estimates to cluster students. Following Grimm and Ram [4], we implemented the following model in Mplus 6.12 [9]:

$$\begin{aligned} Ti &= 1 * g_0 + Li * g_1 + \varepsilon_i \qquad && i = 1,2,...,t \\ Li &= e^{\left(-e^{\left(-\alpha*(i-\lambda)\right)}\right)} \qquad && i = 1,2,...,t. \end{aligned} \tag{1}$$

$Ti$ $(i = 1,2,...,t)$ are the $t$ temporal splits, $\alpha$ is the rate of change, $\lambda$ represents the time at which maximum growth rate occurs, $g_0$ is the lower asymptote, and $(g_0 + g_1)$ is the upper asymptotic value of the function. $\alpha$ e $\lambda$ are phantom variables, so they are unrelated to all other variables in the model, and have no variance. Their role is to constrain the slope loadings ($Li$) to follow the prespecified Gompertz functions.

## 2.2 Data

Data come from the statistical office of IULM University. The demo-social aspects (identification number, gender, date of birth, secondary school's type and mark) and university characteristics (enrollment year, course enrollment, marks, credits and time of examinations, types and marks of bachelor degree) are investigated for two cohorts of enrolled students at the five faculties of IULM. The first cohort refers to the students enrolled in 2006 ($N_1 = 864$); the second one to the students enrolled in the 2007 ($N_2 = 803$). In this application we analyze only students who have graduated by the end of 2012 (i.e. 67% and 63% of 2006 and 2007 cohort, respectively). Tab. 1 contains some statistical insights on demo-social structures of the cohorts.

**Table 1** Descriptive statistics for graduate students in the cohort 2006 and 2007[a]

| | **Cohort 2006** | **Cohort 2007** |
|---|---|---|
| Type of Bachelor degree (%) | | |
| Lyceum | 66.60 | 67.97 |
| Technical | 23.01 | 19.33 |
| Others | 10.38 | 12.70 |
| Mark of Bachelor degree | 79.61 (12.35) | 77.85 (11.84) |
| Age at degree | 23.46 (1.54) | 23.36 (2.11) |
| Degree's grade | 96.54 (8.80) | 96.49 (8.91) |

[a] In the brackets standard deviation values

In order to quantify the university knowledge accumulation, we use the $UHC$ measure based on the product between mark and credits for each examination [3]:

$$mc_j(i) = \sum_j mark_j(i) * credit_j(i)$$

$$UHC_j(T) = \frac{\sum_{i=1}^{T} mc_j(i)}{\max mc}.$$

$mc_j(i)$ is the $UHC$ acquired in the *i-th* time split by the *j-th* student, $\max mc$ equals 30 (maximum grade without considering the possible "laude") per 174 (number of formative credits to be achieved in order to access the degree, which is worth 6 credits), and $UHC_j(T)$ is the normalized cumulated $UHC$ acquired till the *T-th* time split. The observed trajectories of the $UHC_j(T)$ in the two cohorts suggest to analyze the knowledge accumulation using a nonlinear approach (see Fig. 1 and Tab.2). Following the CZ research, we use a Gompertz function to estimate the rate of accumulation.

## 2.3 Results

In the first phase of the analysis we estimated the latent Gompertz curve that models the acquisition of $UHC$ over time by individual students. To identify the time instants of interest, we conducted an Exploratory Factor Analysis (EFA) on the exam sessions. As results of the EFA, for the cohort 2006 we used as splits: January, April, June, September from 2007 to 2009 and January 2010. We implemented a Gompertz model (see Eq. 1) with 13 temporal splits (henceforth defined "Long-time Gompertz model"). For the "Long-time Gompertz model", we obtain $\chi^2 = 3616.417$, $RMSEA = 0.290$, $CFI = 0.761$, and $TLI = 0.745$. As these fit statistics are decid-

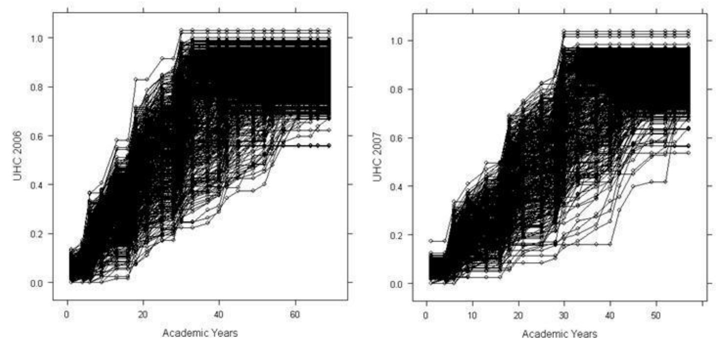


**Fig. 1** Observed trajectories for 2006 and 2007 cohorts

**Table 2** Mean of $UHC$ measures from 2007 to 2011

| Months | | Cohort 2006 | Cohort 2007 | Months | | Cohort 2006 | Cohort 2007 |
|---|---|---|---|---|---|---|---|
| 2007 | January | 0.04 (0.03) | – | 2010 | January | 0.74 (0.16) | 0.50 (0.14) |
| | April | 0.06 (0.04) | – | | April | 0.76 (0.15) | 0.54 (0.15) |
| | June | 0.16 (0.07) | – | | June | 0.79 (0.13) | 0.67 (0.18) |
| | September | 0.23 (0.07) | – | | September | 0.80(0.12) | 0.73(0.17) |
| 2008 | January | 0.29 (0.10) | 0.04 (0.03) | 2011 | January | 0.80 (0.11) | 0.75 (0.15) |
| | April | 0.30 (0.10) | 0.04 (0.03) | | April | 0.81 (0.10) | 0.78 (0.14) |
| | June | 0.42 (0.14) | 0.15 (0.03) | | June | 0.82 (0.09) | 0.80 (0.12) |
| | September | 0.48 (0.14) | 0.21 (0.07) | | September | 0.82 (0.08) | 0.81 (0.10) |
| 2009 | January | 0.54 (0.15) | 0.25 (0.09) | | | | |
| | April | 0.55 (0.15) | 0.27 (0.09) | | | | |
| | June | 0.65 (0.18) | 0.39 (0.13) | | | | |
| | September | 0.72 (0.17) | 0.46 (0.14) | | | | |

[a] In the brackets the standard deviation values

edly unpleasant, we decided to continue the investigation by testing new models. As well known, the significance of $\chi^2$ can be misleading if sample size is large or the assumption of multivariate normality has been violated. Both cases characterize our application then we can neglect the information come from this class of indexes. We can neglect RMSEA too: the index assumes high values when residual variances are quite small [2], as in our application. Hu and Bentler [5] suggestested to use alternative indexes as IFI, CFI, and SRMR for fit evaluating.

Focusing on these indices, we chose the model that considers a shorter period, ie the last 7 splits (2008: June, September; 2009: January, April, June, September; 2010: January, see Fig. 2). We summarized the parameters' estimates and the fit statistics of the "Short-time Gompertz model" in Tab. 3. Obviously for the cohort 2007, split times have been shifted to a year.The average growth rate is not constant between the two cohorts and it indicates that the cohort of 2007 performs better . This is reflected in a mean time-to-degree lower than the 2006's cohort. Obviously it is a partial comparison due to the fact that the observation time is different between the two cohorts. For both cohorts, starting at L4 (April 2009, see Fig. 2) we notice a change in the individual trajectories, embodied in the allocation of a predominant role to the disciplines addressed in the final phase. Among these exams, there are obviously those belonging to the third year of the curriculum, but also those that are faced as last because they are perceived as difficult.

The objective of the second phase of the analysis is to identify homogenous groups of students according to their ability to accumulate $UHC$. Our interest is to define the characteristics of these groups. Using Gompertz growth curve model parameters and graduation marks, we perform a hierchical cluster analysis with Ward[1] linkage. We divided the graduates into three clusters, indicating 3 different levels of $UHC$: "A-fast and good students"; "B- good but not fast students"; "C- neither good nor fast". Tab. 4 summarizes the characteristics of these groups.

Looking at the cohort of 2006, the 3 clusters are sorted with respect to the estimates of $UHC$. For ease of reading, we reported only the average grade in each cluster and the corresponding standard deviation. With reference to the geographical residence, passing from the group of the fastest and good to that of less clever, lower the percentage of students from the northeast Italy, while increasing those coming from northwest and middle Italy. The percentage of students coming from southern Italy and abroad remains fairly constant. The distribution of the bachelor degree mark is similar to that of graduation mark, so we can say that students who enter the university with the best grades are those who have a vote of degree higher. Similarly, students with a lower degree mark, they also have a low-grade diploma. It is interesting to note that even the distribution of Bachelor degree varies in the different clusters. We find more students graduated at lyceum in the groups of the best students. In the third cluster we have a bit of technical graduates. Also the weight of the residual category (masterly, professional, foreign) considerably increases in the group of students with the lowest $UHC$.

Regarding the cohort of 2007, the considerations are entirely analogous. The only difference concerns the distribution of geographical residence, with a peak in cluster B of students from northwest and southern Italy.

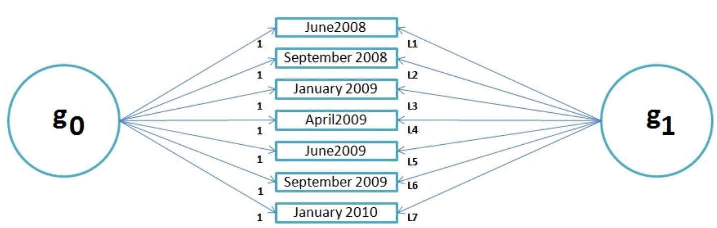


**Fig. 2** The SEM Short-time Gompertz model latent curves model

[1] The objective of this method is to minimize the total within-cluster variance. In general, it is regarded as very efficient, however, it tends to create clusters of small size.

**Table 3** Parameter's estimates and fit statistics for the "Short-time Gompertz model"

| Cohort | 2006 | 2007 |
|---|---|---|
| $g_0$: mean (SE) | 0.000 (0.012) | 0.000 (0.014) |
| $g_1$: mean (SE) | 0.000 (0.027) | 0.000 (0.026) |
| $\alpha$ | 1.442 | 1.829 |
| $\lambda$ | 4.602 | 4.305 |
| T1: intercept - L1 | 0.416 - 0.000 | 0.391 - 0.000 |
| L2: intercept - L2 | 0.479 - 0.000 | 0.460 - 0.000 |
| L3: intercept - L3 | 0.538 - 0.000 | 0.497 - 0.000 |
| L4: intercept - L4 | 0.552 - 0.092 | 0.538 - 0.174 |
| L5: intercept - L5 | 0.654 - 0.569 | 0.675 - 0.755 |
| L6: intercept - L6 | 0.729 - 0.875 | 0.739 - 0.956 |
| L7: intercept - L7 | 0.768 - 0.969 | 0.781 - 0.993 |
| $\chi^2$ | 479.225 | 273.082 |
| RMSEA | 0.224 | 0.177 |
| CFI | 0.947 | 0.963 |
| TLI | 0.931 | 0.951 |
| IFI | 0.947 | 0.965 |
| SRMR | 0.099 | 0.091 |

**Table 4** Cluster analysis [a]

| Cohort | 2006 | | | 2007 | | |
|---|---|---|---|---|---|---|
| Cluster | A | B | C | A | B | C |
| # of students | 154 | 147 | 277 | 176 | 152 | 183 |
| graduation mark | 108.01 | 99.12 | 88.80 | 106.65 | 96.36 | 86.84 |
| | (2.92) | (2.03) | (4.31) | (11.54) | (10.95) | (8.71) |
| Area of residence | | | | | | |
| Northwest Italy | 58% | 63% | 67% | 66% | 64% | 68% |
| Northeast Italy | 25% | 16% | 13% | 19% | 18% | 11% |
| Middle Italy | 3% | 6% | 4% | 4% | 1% | 5% |
| Southern Italy | 13% | 12% | 14% | 10% | 16% | 13% |
| Abroad | 1% | 3% | 2% | 1% | 1% | 3% |
| Bachelor degree's mark | 89.67 | 80.23 | 73.56 | 85.41 | 76.98 | 70.97 |
| | (10.10) | (11.52) | (10.51) | (11.54) | (10.95) | (8.71) |
| Type of Bachelor degree | | | | | | |
| Lyceum | 70% | 61% | 53% | 74% | 61% | 48% |
| Technical | 20% | 22% | 26% | 12% | 20% | 25% |
| Other | 10% | 17% | 21% | 14% | 19% | 27% |

[a] In the brackets the standard deviation values

## 3 Discussion and future directions

The *HC* can be studied under several perspectives; in this application we have chosen to analyze the *HC* accumulated by students for achieving a degree. Following the definition proposed by CZ [3], we focused on University Human Capital ($UHC$) and we tested the velocity (or rate) of knowledge accumulation through a latent growth curve model. This model has been applied on two cohorts of students. Given the non-linear nature of these trajectories, we estimated a latent type Gompertz curve. In the second stage of analysis, we used estimates of the $UHC$ to divide students according to different modes of knowledge accumulation. We have thus obtained three different clusters: "A- good and fast students", "B- good but not fast students" and "C-not clever students". This information can be very useful in the process of university orientation. These results highlight that lyceum's students with high Bachelor degree's marks are the best and fastest ones. A hints for university management could be to orient their activity in upper schools in order to capture these typologies of students. Finally, the model can be used profitably by job placement support, providing information about the different capabilities of students. It is interesting to note that this analysis represents one of the first cases of application of the $UHC$ analysis in humanistic faculties. The cited studies, in fact, refer to scientific faculties. Despite this important difference, the results found are in line with those reported in the literature.

## References

1. Bianconcini S., Cagnone S.: A general multivariate latent growth model with applications to student achievement. Journal of Educational and Behavioral Statistics, **37**, 339-364 (2012)
2. Browne M.W., MacCallum, R.C., Kim, C., Andersen, B.L., and Glaser, R.: When fit indices and residuals are incompatible. Psychological Methods**7**,403-21(2002)
3. Civardi M., Zavarrone E.: Estimating University Human Capital through Education Evaluation In: Fabbris L. (eds.) Effectiveness of University Education in Italy. Physica-Verlag, New York, 369-380 (2006)
4. Grimm, K. J., Ram, N.: Nonlinear growth models in Mplus and SAS. Structural Equation Modeling.**6**, 676-701(2009)
5. Hu, L.T. Bentler, P.M.: Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives. Structural Equation Modeling.**6**(1),1-55 (1999)
6. Mariani, P., Mussini M., Zavarrone E.: Measuring Dissimilarity In Learning Capital Using Time Trajectories. accepted for publication by Stat. Appl. (2013)
7. McArdle J.J. and Epstein D:B: Latent growth curves within developmental structural equation models. Child Development, **58**(1)110-133(1987)
8. Muthén, B: Latent variable growth modeling with multilevel data. In M. Berkane (Ed.), Latent Variable Modeling with Applications to Causality, 149-161. Springer Verlag (1997)
9. Muthén, L. K., Muthén, B. O.: Mplus user's guide (6th ed.) [Computer software manual]. Los Angeles, CA: (1998-2010)
10. OECD: Social Capital,Human Capital and Health What is the Evidence? OECD-IRDES (2001) http://www.oecd.org/edu/research/45760738.pdf
11. Singer, J.: Using SAS Proc Mixed to fit multilevel models, hierarchical models, and individual growth curves. Journal of Educational and Behavioral Statistics. **24**, 4, 323-355(1998)