

Progetto n. 1

Valore di segnalazione del voto di diploma e *grading standard* nelle scuole secondarie superiori

versione finale (22 gennaio 2011)

Gruppo di ricerca:

Daniele Checchi (Coordinatore), Massimiliano Bratti, Antonio Filippin

Sommario:

1. Le scuole secondarie da analizzare	2
2. La provenienza dei dati sulle immatricolazioni universitarie	8
3. Il problema dell'autoselezione, e possibile rimedi per ridurlo	14
4. Accountability e indicatori di performance per le scuole superiori.....	18
5. La metodologia statistica utilizzata.....	22
6. La probabilità di proseguire nell'istruzione terziaria.....	24
7. Risultati dell'analisi sul numero di crediti acquisiti e sul voto medio degli esami di profitto	28
8. <i>Grading standard</i> delle scuole e valore informativo del voto di diploma e dei test all'ingresso	34
9. Conclusioni.....	43
Appendice A. Descrizione delle variabili.....	48
Riferimenti bibliografici.....	49

Il presente rapporto di ricerca dal titolo “Valore di segnalazione del voto di diploma e *grading standard* nelle scuole secondarie superiori” fa parte di un progetto più ampio su “La qualità del sistema d’istruzione lombardo. Misurazioni, confronti internazionali e proposte”, commissionato al Dipartimento di Studi del Lavoro e del Welfare (DSLW) e al Centro Interdipartimentale di ricerca su Lavoro, Formazione e Welfare (WTW) dell’Università di Milano, che comprende anche i sotto-progetti su: 2) La produttività e l’eccellenza scientifica delle università lombarde; 3) Fondazioni e promozione dell’eccellenza nel sistema educativo: esperienze di successo nel Regno Unito e negli USA. 4) Laboratorio sulla qualità delle istituzioni universitarie.

1. Le scuole secondarie da analizzare

Il primo passaggio di una ricerca che intenda provare ad analizzare la qualità delle scuole lombarde è quello di rispondere alla domanda: quante sono le scuole secondarie in Lombardia? Difficile fornire una risposta precisa. Noi abbiamo analizzato quattro fonti archiviali diverse, ottenendo risposte anche molto divergenti. La prima fonte utilizzata è quella che ci è pervenuta dal Ministero dell’Istruzione, dell’Università e della Ricerca (MIUR) riportanti i codici fiscali dei maturati nelle scuole statali della Lombardia nel giugno 2008.¹ Secondo tale fonte vi sono 573 istituzioni scolastiche secondarie statali, articolate in 2467 classi e 50487 studenti (al netto di 116 abbandoni e 142 trasferimenti – questo elimina anche alcuni casi di ripetizione dei codici degli studenti). Da questa fonte non conosciamo nessun dato identificativo della scuola se non il codice della scuola. Per esempio MIRA03601A, che abbiamo decrittato come segue:

- 1) le prime due lettere identificano la PROVINCIA di localizzazione della scuola
- 2) le seconde due lettere identificano l’ORDINE DI SCUOLA
- 3) le quattro cifre successive indicano lo stesso indirizzo. Questo lo abbiamo verificato utilizzando un diverso archivio (proveniente dalla terza fonte illustrata di seguito). In questa fonte, oltre che nome ed indirizzo, compaiono due variabili identificative delle scuole: “Codice” e “Codice ist. scol. di rif.” che è identico quando due istituti diversi si trovano nello stesso edificio. Ovviamente questo codice si ripete per provincia. Per esempio:

MILANO	IST PROF PER L'AGRICOLTURA E L'AMBIENTE	MIRA03601A	I.P.A.A. (SEZ.AS.)	VIA LITTA MODIGNANI, 55	20161	081	MIIS03600A	SEZ. ASS.
MILANO	LICEO SCIENTIFICO	MIPS03601R	"VILFREDO FEDERICO PARETO"	VIA LITTA MODIGNANI, 55	20161	081	MIIS03600A	SEZ. ASS.
MILANO	ISTITUTO TECNICO INDUSTRIALE	MITF03801E	GIUSEPPE LUIGI LAGRANGE	VIA LITTA MODIGNANI,65	20161	081	MIIS038002	SEZ. ASS.
MILANO	IST PROF PER I SERVIZI ALBERGHIERI E RISTORAZIONE	MIRH038016	"G. BRERA "	VIA LITTA MODIGNANI	20161	081	MIIS038002	SEZ. ASS.

A questo punto combinando questi tre elementi siamo in grado, anche senza conoscerne il nome, di identificare per ogni scuola la provincia di localizzazione, la tipologia di scuola e gli istituti con cui condivide gli edifici.

In assenza di informazioni più dettagliate noi tratteremo ciascuna tipologia di scuola come se si trattasse di una istituzione indipendente, anche se il fatto di condividere l’edificio solleva il sospetto che possa trattarsi in taluni casi (per esempio nel caso delle scuole paritarie) di gestioni comuni. Per contro, la prassi che si è diffusa nelle scuole statali di concentrare istituti diversi in centri omnicomprensivi contribuisce a rafforzare questo fenomeno. Per avere un’idea dell’entità del fenomeno di aggregazione, si veda la Tabella 1.

¹ Si tratta del file Vanno_lombardia.dta ricevuto il 29/12/2009 – 11472212 k, di seguito indicato come “fonte Barbieri”.

Tabella 1 – Distribuzione nelle scuole per sedi – fonte Barbieri

con quante scuole condivide l'edificio	numero scuole	% numero scuole	quante scuole contiene la sede	numero sedi	% numero sedi
1	95	16.58	1	95	37.85
2	196	34.21	2	88	35.06
3	105	18.32	3	35	13.94
4	44	7.68	4	11	4.38
5	50	8.73	5	10	3.98
6	24	4.19	6	4	1.59
7	35	6.11	7	5	1.99
8	24	4.19	8	3	1.20

La distribuzione per tipologia di scuola e provincia è riportata in Tabella 2, mentre quella dei maturati viene riportata in Tabella 3. Questi sono gli studenti di cui andremo a cercare di seguire i percorsi nell'eventuale primo anno di carriera universitaria, attraverso l'incrocio dei codici fiscali con gli immatricolati universitari. E quindi questo è l'insieme delle scuole secondarie che saremo in grado di "valutare". Tuttavia questo comporterebbe la non valutazione delle scuole paritarie, in quanto esse non riportano al Ministero della Pubblica Istruzione (MPI) i dati anagrafici dei loro maturati. Vedremo però in seguito come sia possibile recuperare qualche informazione in merito, accorpando i dati degli immatricolati universitari che non trovino un riscontro in questa fonte, perché appunto provenienti da scuole paritarie.

Tabella 2 - Distribuzione delle scuole per provincia e tipologia di scuola – fonte Barbieri

		BG	BS	CO	CR	LC	LO	MI	MN	PV	SO	VA	Totale
PC	liceo classico	3	2	1	3	1	0	10	1	3	1	4	29
PM	istituto magistrale	6	2	2	1	1	1	7	2	1	1	1	25
PS	liceo scientifico	11	16	5	2	3	2	40	5	4	5	9	102
RA	I.P. agricoltura	0	5	0	3	0	1	1	2	1	0	0	13
RC	I.P. Commerciale	7	6	1	3	1	1	20	3	4	1	8	55
RF	I.P. servizi sociali	1	0	0	0	0	0	0	1	0	0	0	2
RH	I.P. alberghiero	2	5	0	0	0	0	2	0	0	2	0	11
RI	I.P. artigianato	9	9	4	3	2	1	22	2	6	4	5	67
SD	Ist. arte	0	1	2	0	0	0	2	2	1	0	0	8
SL	liceo art	3	1	0	2	1	1	5	0	1	1	2	17
TA	I.T. agrario	2	2	0	1	0	1	3	2	1	1	0	13
TD	I.T. comm. e geometri	15	20	6	5	6	5	44	6	6	5	16	134
TE	I.T. att. sociali	0	0	0	0	0	0	1	1	0	0	0	2
TF	I.T. industriale	10	11	4	3	1	3	25	3	4	1	12	77
TL	I.T. geometri	2	1	1	2	1	0	1	2	2	1	2	15
TN	I.T. turismo	0	0	0	0	0	0	3	0	0	0	0	3
Totale		71	81	26	28	17	16	186	32	34	23	59	573

Tabella 3 - Distribuzione dei maturati per provincia e tipologia di scuola – fonte Barbieri

		BG	BS	CO	CR	LC	LO	MI	MN	PV	SO	VA	Total
PC	liceo classico	328	310	184	380	71	0	1468	122	235	45	596	3739
PM	istituto magistrale	754	453	303	130	224	215	1107	195	231	86	180	3878
PS	liceo scientifico	1148	1751	682	321	402	294	4631	369	673	298	1255	11824
RA	I.P. agricoltura	0	179	0	49	0	42	26	53	56	0	0	405
RC	I.P. Commerciale	593	521	94	345	93	122	1279	144	301	65	545	4102
RF	I.P. servizi sociali	68	0	0	0	0	0	0	12	0	0	0	80
RH	I.P. alberghiero	169	288	0	0	0	0	281	0	0	102	0	840
RI	I.P. artigianato	335	441	233	120	136	44	886	81	183	172	353	2984
SD	Ist. arte	0	13	142	0	0	0	185	115	10	0	0	465
SL	liceo art	261	63	0	128	88	75	604	0	79	20	246	1564
TA	I.T. agrario	165	218	0	51	0	82	162	71	112	28	0	889
TD	I.T. comm. e geometri	1384	1710	536	454	473	343	3960	251	452	299	1463	11325
TE	I.T. att. sociali	0	0	0	0	0	0	132	138	0	0	0	270
TF	I.T. industriale	895	730	315	452	148	333	2475	182	427	105	729	6791
TL	I.T. geometri	150	207	60	64	69	0	61	82	129	42	105	969
TN	I.T. turismo	0	0	0	0	0	0	362	0	0	0	0	362
Total		6250	6884	2549	2494	1704	1550	17619	1815	2888	1262	5472	50487

La seconda fonte che abbiamo utilizzato ci è stata fornita dalla Direzione Regionale della Lombardia.² Essa contiene informazioni relative anche alle scuole paritarie e parificate, oltre che ai dati aggregati sulle tipologie di diplomati (interni o esterni, italiani o stranieri) ed alcune informazioni sulla distribuzione dei voti di maturità (numero di studenti in intervalli di 10 punti). Questo secondo file contiene informazioni relative a 845 scuole, di cui 231 paritarie e 6 parificate (che non comparivano quindi nel file precedente). Ci sono inoltre 35 scuole statali presenti in questo file che non compaiono nel precedente, e 49 scuole che viceversa compaiono nel primo file e non nel secondo.³ Tranne che in un caso, usando la scomposizione del codice identificativo sopra illustrata, sappiamo che si tratta di scuole localizzate in edifici differenti dai precedenti. Riusciamo tuttavia a vedere che sono in media scuole più piccole, con un numero inferiore di maturati (vedi Tabella 4). Riscontriamo invece una sufficiente corrispondenza tra il numero dei nominativi di cui disponiamo ed il numero dei diplomati interni della scuola.⁴ Tuttavia la prima fonte riporta una media di 2.82 diplomati in più per scuola (con un intervallo di oscillazione delle differenze compreso tra -11 al primo decile e +14 al nono decile). Ci devono evidentemente essere alcuni casi di diversa classificazione delle scuole, perché si evidenziano almeno un paio di casi in cui il numero dei maturati differisce di più di 50 studenti (si confronti Tabella 3 con Tabella 7). Per coerenza con gli incroci successivi, terremo buono come numero di maturati quello della prima fonte, ricorrendo alla seconda solo nei casi in cui manchino le informazioni.

² Diplomati_a.s. 2007_08 BAGGIO.xls ricevuto il 4/5/2010 – 789504 k, di seguito indicata come “fonte Baggio”.

³ Abbiamo verificato che tali scostamenti sono imputabili ad una diversa modalità di classificazione delle scuole nei due archivi, per cui riteniamo che alcune scuole ospitate nei medesimi edifici siano state classificate in modo diverso dai due archivi.

⁴ Il coefficiente di correlazione tra il numero di maturati per scuola ed il numero dei diplomati interni è 0.97.

Tabella 4 – Numero medio di maturati/diplomati per scuola

	maturati per scuola (Barbieri)	diplomati interni (Baggio)	maturati o diplomati
solo fonte MIUR (Barbieri)	40.1		32.14
solo fonte DR (Baggio)		32.14	40.1
entrambe le fonti	92.6	89.77	92.6

Tabella 5 - Distribuzione delle scuole per provincia e tipologia di scuola –
solo scuole statali – fonte Baggio

secon2		BG	BS	CO	CR	LC	LO	MI	MN	PV	SO	VA	Totale
PC	liceo classico	3	2	1	3	1		11	1	3	1	4	30
PM	istituto magistrale	6	2	2	1	1	1	7	2	1	1	1	25
PS	liceo scientifico	11	17	5	2	3	2	42	5	4	5	9	105
RA	I.P. agricoltura		4		3		1	1	2	1			12
RC	I.P. Commerciale	7	7	2	2	1	1	22	3	3	1	7	56
RF	I.P. servizi sociali	1							1				2
RH	I.P. alberghiero	2	5					3			2		12
RI	I.P. artigianato	9	9	3	3	2	1	21	2	6	4	5	65
SD	Ist. arte		1	2				2	2	1			8
SL	liceo art	3	2		2	1	1	4		1	1	2	17
ST	conservatorio							1					1
TA	I.T. agrario	2	2		1		1	3	2	1	1		13
TD	I.T. comm. e geometri	13	17	6	4	6	4	43	5	6	5	15	124
TE	I.T. att. sociali							1	1				2
TF	I.T. industriale	9	10	3	3	1	2	28	3	3	1	10	73
TL	I.T. geometri	1	1	1	2	1			2	1	1	1	11
TN	I.T. turismo							3					3
Totale		67	79	25	26	17	14	192	31	31	23	54	559

Tabella 6 - Distribuzione delle scuole per provincia e tipologia di scuola –
solo scuole paritarie e parificate – fonte Baggio

		BG	BS	CO	CR	LC	LO	MI	MN	PV	SO	VA	Totale
PC	liceo classico	4	2	1	2			13	1	1		1	25
PL	liceo linguistico	5	2	4	2	2		10		2		4	31
PM	istituto magistrale	2	4	2		1		3		2		3	17
PQ	istituto magistrale	1						3			1	1	6
PS	liceo scientifico	10	7	2		1	2	20	1	2		6	51
RA	I.P. agricoltura			2									2
RC	I.P. Commerciale		1			2						1	4
RF	I.P. servizi sociali		1	1	1	1						1	5
RH	I.P. alberghiero	1		1								3	5
RI	I.P. artigianato	1						2				2	5
SD	Ist. arte	1	1	1				1		1			5
SL	liceo art		1	3				4		1		1	10
TB	I.T. aeronautico	1	1	1				1				2	6
TD	I.T. comm. e geometri	6	2	4				10	1	2		5	30
TE	I.T. att. sociali			1						1			2
TF	I.T. industriale	1	4					5	1			2	13
TL	I.T. geometri	4	2	2		1		3	1	1		2	16
TN	I.T. turismo	1	1	1								1	4
Totale		38	29	26	5	8	2	75	5	13	1	35	237

Tabella 7 - Distribuzione dei diplomati interni per provincia e tipologia di scuola – solo scuole statali – fonte Baggio

		BG	BS	CO	CR	LC	LO	MI	MN	PV	SO	VA	Total
PC	liceo classico	319	297	184	375	70		1493	119	213	45	588	3703
PM	istituto magistrale	727	423	294	125	213	212	1069	188	224	80	165	3720
PS	liceo scientifico	1121	1737	666	314	420	284	4533	360	657	290	1224	11606
RA	I.P. agricoltura		140		38		33	98	53	84			446
RC	I.P. Commerciale	549	644	101	303	85	101	1474	130	250	55	470	4162
RF	I.P. servizi sociali	65							16				81
RH	I.P. alberghiero	152	223					296			89		760
RI	I.P. artigianato	291	381	180	106	106	30	836	87	152	162	291	2622
SD	Ist. arte		13	121				232	108	5			479
SL	liceo art	273	100		164	101	66	552		74	31	265	1626
ST	conservatorio							26					26
TA	I.T. agrario	139	206		39		80	179	60	111	28		842
TD	I.T. comm. e geometri	1310	1561	512	426	536	271	3984	253	444	267	1350	10914
TE	I.T. att. sociali							120	132				252
TF	I.T. industriale	791	655	307	430	130	315	2784	174	393	98	635	6712
TL	I.T. geometri	140	170	58	57	64			81	110	32	68	780
TN	I.T. turismo							338					338
Totale		5877	6550	2423	2377	1725	1392	18014	1761	2717	1177	5056	49069

Tabella 8 - Distribuzione dei diplomati interni per provincia e tipologia di scuola – solo scuole paritarie e parificate – fonte Baggio

		BG	BS	CO	CR	LC	LO	MI	MN	PV	SO	VA	Totale
PC	liceo classico	55	39	26	44			715	16	16		10	921
PL	liceo linguistico	91	34	87	51	38		432		25		58	816
PM	istituto magistrale	52	81	38		54		106		44		63	438
PQ	istituto magistrale	19						110			21	19	169
PS	liceo scientifico	337	202	85		21	40	801	26	38		203	1753
RA	I.P. agricoltura			29									29
RC	I.P. Commerciale		9			44						18	71
RF	I.P. servizi sociali		11	13	6	38						9	77
RH	I.P. alberghiero	5		23								51	79
RI	I.P. artigianato	30						28				22	80
SD	Ist. arte	33	9	2				6		17			67
SL	liceo art		31	51				280		11		9	382
TB	I.T. aeronautico	45	10	13				27				35	130
TD	I.T. comm. e geometri	83	21	67				479	12	51		99	812
TE	I.T. att. sociali			22						22			44
TF	I.T. industriale	31	68					261	6			48	414
TL	I.T. geometri	81	39	37		13		88	16	45		66	385
TN	I.T. turismo	10	12	11								16	49
Totale		872	566	504	101	208	40	3333	76	269	21	726	6716

La terza fonte utilizzata è l'archivio ufficiale distribuito dal Ministero dell'Istruzione, dell'Università e della Ricerca (MIUR) per la determinazione degli organici.⁵ In questa fonte, oltre ai codici identificativi delle scuole statali, si trovano i dati sugli alunni (con la distinzione tra alunni normodotati e diversamente abili) per classe e il numero dei docenti (distinti tra cattedre interne ed esterne), da cui si può quindi inferire qualche informazione sulle risorse disponibili. In questo caso le scuole presenti nell'archivio

⁵ Ci riferiamo al CD del Ministero dell'Istruzione, dell'Università e della Ricerca – Direzione Generali per i sistemi Informativi – La scuola statale in Italia – I dati di organico di diritto AS.2007/8, di seguito indicato come “fonte MPI”.

sono 680, che è l'insieme più ampio tra quelli disponibili, e che includere tutti gli altri (ad eccezione di 3 scuole), come si vede dalla Tabella 9.

Tabella 9 – Fonte dei dati sulle scuole

origine dei dati	statale	paritaria/ parificata	Total
solo archivio MPI	75	0	75
solo archivio DR Lombardia (Baggio)	2	237	239
solo dati sui maturati (Barbieri)	1	0	1
solo MPI+Barbieri	48	0	48
solo MPI+Baggio	33	0	33
presenti in tutti gli archivi	524	0	524
Total	683	237	920

I dati sulla distribuzione delle scuole per provincia e tipologia di scuola sono riportati in Tabella 10. Confrontandola con Tabella 2 gli istituti per i quali vi è il maggior scostamento sono gli Istituti Professionali Commerciali, Alberghiero e per i Servizi Sociali, oltre che gli Istituti d'arte. Come indicatore indiretto del numero di diplomati in uscita prendiamo la somma degli alunni iscritti al quinto anno, pari complessivamente a 53749 studenti.⁶

Tabella 10 - Distribuzione delle scuole per provincia e tipologia di scuola –
solo scuole statali – fonte MPI

		BG	BS	CO	CR	LC	LO	MI	MN	PV	SO	VA	Total
PC	liceo classico	4	2	1	3	1		11	1	3	1	6	33
PM	istituto magistrale	8	2	2	1	1	1	8	2	1	1	1	28
PS	liceo scientifico	11	19	5	2	4	2	51	5	4	5	10	118
RA	I.P. agricoltura		5		3		1	1	3	1			14
RC	I.P. Commerciale	7	9	2	3	1	1	31	3	6	1	10	74
RF	I.P. servizi sociali	1				1			1				3
RH	I.P. alberghiero	2	7	2				3			3		17
RI	I.P. artigianato	12	12	5	3	2	1	24	3	7	4	7	80
SD	Ist. arte		1	2				2	2	1			8
SL	liceo art	4	3		2	1	1	12		1	1	2	27
TA	I.T. agrario	2	2		1		1	3	2	1	1		13
TD	I.T. comm. e geometri	16	21	7	5	7	5	49	7	9	6	16	148
TE	I.T. att. sociali							1	1				2
TF	I.T. industriale	11	12	4	3	2	3	33	3	7	1	12	91
TL	I.T. geometri	2	3	2	2	1		1	2	4	1	2	20
TN	I.T. turismo							4					4
Total		80	98	32	28	21	16	234	35	45	25	66	680

⁶ Il numero degli alunni in quinta classe è correlato con il numero dei maturati o dei diplomati interni al 0.98.

Tabella 11 - Distribuzione dei diplomati interni per provincia e tipologia di scuola – solo scuole statali – fonte MPI

		BG	BS	CO	CR	LC	LO	MI	MN	PV	SO	VA	Total
PC	liceo classico	333	310	195	387	72		1525	122	213	48	607	3812
PM	istituto magistrale	767	458	307	131	217	211	1116	194	243	84	182	3910
PS	liceo scientifico	1148	1818	707	323	425	292	4787	375	677	308	1277	12137
RA	I.P. agricoltura		173		64		43	28	54	112			474
RC	I.P. Commerciale	621	725	123	382	100	119	1993	143	286	73	577	5142
RF	I.P. servizi sociali	68				0			15				83
RH	I.P. alberghiero	174	277	0				403			114		968
RI	I.P. artigianato	341	478	240	123	152	38	1063	104	198	170	359	3266
SD	Ist. arte		10	147				187	112	10			466
SL	liceo art	213	64		145	94	44	596		78	20	288	1542
TA	I.T. agrario	167	204		50		82	204	73	113	31		924
TD	I.T. comm. e geometri	1404	1693	564	429	594	276	4377	286	487	338	1451	11899
TE	I.T. att. sociali							148	139				287
TF	I.T. industriale	894	734	339	454	146	337	3157	192	432	101	758	7544
TL	I.T. geometri	142	176	56	73	79		69	95	104	40	92	926
TN	I.T. turismo							369					369
Total		6272	7120	2678	2561	1879	1442	20022	1904	2953	1327	5591	53749

Da ultimo, poiché le prime due fonti non riportano il nome dell'istituto o lo riportano in modo molto approssimativo, abbiamo fatto ricorso ad una ulteriore banca dati, utilizzata dalle segreterie dell'Università di Milano e fornita dal CINECA. Al netto della presenza di doppie imputazioni, tale banca dati contiene 1263 indicatori di scuola, che coprono quasi completamente gli archivi precedenti (vedi tabella 12). In realtà molte di queste scuole risultano a posteriori ridondanti, perché non si riesce a rintracciare dati sulle immatricolazioni universitarie provenienti da esse.

Tabella 12 – Provenienza delle informazioni

	banca dati istituti superiori lombardi	di cui archivio CINECA	di cui archivio MPI	di cui in archivio maturati (Barbieri)	di cui in archivio DR Lombardia (Baggio)	di cui presenti in MPI+ maturati
statale	954	884	680	573	559	545
parificata	28	28			10	
paritaria	302	295			227	
inf.assente	56	56				
Total	1340	1263	680	573	796	545

2. La provenienza dei dati sulle immatricolazioni universitarie

Al fine di poter analizzare la transizione dei maturati delle scuole secondarie di secondo grado alle Università lombarde, è stata inviata nel mese di novembre 2009 a tutti i rettori e pro-rettori alla didattica delle stesse università una richiesta di rilascio di dati relativa ai nominativi degli immatricolati nell'anno accademico 2008-09.⁷

⁷ Si riporta a titolo indicativo il testo della lettera inviata “Magnifico Rettore, Le scrivo su indicazione del prof.??? per una richiesta di dati. Fondazione Cariplo ha recentemente finanziato un progetto di ricerca nell'ambito dei progetti speciali rivolti alla promozione della valutazione e del merito nell'ambito

A tale richiesta hanno dato seguito il Politecnico di Milano (30/11/2009), l'Università di Pavia (21/12/2009), l'Università degli Studi di Milano (29/12/2009), l'Università di Brescia (23/2/2010). A seguito di nostra partecipazione ad incontro di rappresentanti degli Atenei Lombardi e di lettera di sollecito inviata dall'Ufficio Scolastico Regionale (9/6/2010), anche l'Università dell'Insubria (24/5/2010) e l'Università Cattolica⁸ (4/11/2010) fornivano i dati sui loro iscritti. Ulteriori tentativi informali rivolti ai Rettori delle Università di Milano-Bicocca, Università Bocconi, IULM e LIUCC si rivelavano infruttuosi, riscontrando sospettosità nei confronti dell'iniziativa che, a detta dei riceventi la richiesta, avrebbe potuto procurare un vantaggio competitivo a favore dell'Università degli Studi di Milano, sede dei ricercatori autori della ricerca. Qualora si ipotizzasse di replicare iniziativa analoga, si suggerisce di riunire preventivamente i responsabili di ateneo, al fine di offrire condivisione degli obiettivi e rassicurazione sulla non competitività di iniziative di ricerca analoghe.

In sintesi abbiamo ottenuto i dati sugli immatricolati nelle università lombarde per l'anno accademico 2008-09 per sei atenei, riportati in Tabella 14, mentre l'universo degli immatricolati (inclusivi di studenti provenienti anche da altre regioni) è riportato in Tabella 13.

dell'istruzione. L'obiettivo della ricerca che coordino, insieme a due colleghi, Bratti e Filippin, è quello di valutare la capacità "segnalica" dei voti (ed in particolare di quello di maturità) nell'indicare il reale livello di competenze possedute da uno studente in uscita dalla scuola secondaria di secondo grado, oltre che nel predire la futura carriera universitaria in caso di iscrizione all'università.

Questo progetto richiederebbe idealmente di poter avere i dati di tutti gli studenti usciti dal sistema scolastico secondario della Lombardia (Cariplo ha richiesto questa delimitazione territoriale) in un anno specifico (2007-08), con il corrispettivo bagaglio di voti (quantomeno il voto di maturità) per andare poi a rintracciare una parte degli stessi studenti che si siano riscritti in un Ateneo (lombardo) osservando il loro eventuale punteggio conseguito al test di ammissione e durante il primo anno di carriera. Stiamo lavorando con la Direzione Regionale dell'Istruzione per costruire la banca dati di partenza dei maturati, ma vorremmo altresì costruire una anagrafe degli immatricolati al primo anno (2008-09) di tutti i corsi universitari localizzati in Lombardia, in modo da ricostruire indirettamente le scelte di iscrizione/non iscrizione, come anche l'autoselezione in specifici atenei/facoltà. Per questo chiederemmo alla sua Università di poter ottenere un file degli immatricolati al primo anno di tutti i corsi di studio triennali ivi impartiti nell'anno 2008-09 contenente le seguenti informazioni:

- 1) nome, cognome e codice fiscale (per creare le corrispondenze con la scuola secondaria di provenienza) - dal codice fiscale desumiamo altresì genere, età e luogo di nascita - una volta realizzato il match tra le due banche dati gli identificativi personali vengono rimossi.
- 2) cap di residenza
- 3) scuola secondaria di provenienza (meglio il codice Ministero Istruzione se disponibile)
- 4) tipo di diploma di maturità posseduto
- 5) voto di maturità conseguito
- 6) eventuale punteggio ottenuto a test di ammissione o a test di accertamento competenze
- 7) classe di laurea cui è iscritto
- 8) data di iscrizione (se disponibile)
- 9) fascia di reddito cui viene assegnato
- 10) numero di crediti conseguiti al 1/11/2009 e votazione media ottenuta (ponderata con i crediti)

Resta inteso che la banca dati che viene creata anche grazie all'eventuale conferimento di dati da parte del Politecnico resterà a disposizione della Fondazione Cariplo. Analoga richiesta è stata inviata agli altri atenei lombardi, per i quali disponessimo di un contatto (Cattolica, Bocconi, Bicocca, Brescia, Bergamo, Pavia, LIUC-Castellanza, Insubria). In attesa di un suo riscontro in merito, cordiali saluti."

⁸ Quest'ultima solo in forma molto parziale, dato che molte delle variabili richieste risultavano missing.

Tabella 13 – Immatricolati anno 2008-09 – Atenei lombardi – Fonte: MIUR

	immatricolati totali	immatricolati con maturità scientifica	immatricolati con maturità classica	immatricolati nati nel 1982 e precedenti	immatricolati con voto di maturità 90-100	immatricolati con crediti all'entrata
Milano - Università commerciale "Luigi Bocconi"	2429	1138	688	1	1258	0
Milano-Bicocca - Università degli studi	5664	2009	389	380	947	297
Castellanza – Università "Carlo Cattaneo"	318	144	54	4	67	5
Varese - Università dell' Insubria	1833	484	94	162	302	235
Milano - Università degli studi	11314	3981	1789	883	2070	1380
Milano – Politecnico	6174	3540	494	106	1939	0
Milano - Università Cattolica del "Sacro Cuore"	6876	2283	1270	264	1350	0
Milano - Libera Università di Lingue e Comunicazione (IULM)	777	216	91	10	88	179
Milano - Università Vita-Salute San Raffaele	320	159	88	8	167	1
Bergamo - Università degli studi	2560	653	118	196	477	66
Brescia - Università degli studi	2675	1035	147	190	535	0
Pavia - Università degli studi	3996	1622	550	225	1142	74
Totale Lombardia	44936	17264	5772	2429	10342	2237

* Sono esclusi gli iscritti alla facoltà di Medicina con sede a Roma

Tabella 14– Immatricolati anno 2008-09 – Atenei lombardi – Fonte: nostro archivio

	dati iniziali	eliminati perché ricorrenti più di una volta	eliminati perché provenienti da scuole non lombarde	campione finale	di cui nota l'informazione sulla scuola di provenienza
unimi	8757	52	0	8705	8705
polimi	6129	30	0	6099	0
brescia	1720	6	81	1633	1619
pavia	3996	17	1140	2839	1997
cattolica*	4433	12	0	4421	4421
insubria	1470	23	0	1447	1425
Total	26505	140	1221	25144	18167

* esclusi gli studenti iscritti nella sede staccata di Piacenza (economia e giurisprudenza)

Tabella 15 – Immatricolati anno 2008-09 – caratteristiche primo anno di carriera –
Fonte: nostro archivio

uni	N	media voto maturità	media media esami	media cfu conseguiti
unimi	8705	76.05	25.12	34.68
polimi	6099	81.06	24.04	40.31
brescia	1633	77.22	23.89	34.53
pavia	2839	78.56	24.31	30.34
cattolica*	4421			44.72
insubria	1447	75.63	24.39	38.11
Totale	25144	77.93	24.49	42.00

* Il numero di crediti degli iscritti alla Cattolica è stimato moltiplicando il numero degli esami del primo anno × 6

Tabella 16 – Immatricolati anno 2008-9 per tipologia di riscontro in archivi maturati

	iscritti con riscontro codice fiscale maturati - archivio Barbieri	iscritti senza riscontro codice fiscale maturati ma con indicazione scuola provenienza	iscritti senza riscontro codice fiscale e senza indicazione scuola provenienza, quindi non utilizzabili	totale iscritti
unimi	6493	2212		8705
polimi	3254		2845	6099
brescia	1405	214	14	1633
pavia	1829	399	611	2839
cattolica	2426	1995		4421
insubria	1234	196	17	1447
Totale	16641	5016	3487	25144

Utilizzando quindi i dati di Tabella 16, lavoreremo con 21657 immatricolati (pari al 25% dei maturati nelle scuole pubbliche), che non sembrano rappresentare un sottocampione distorto dell'insieme degli immatricolati che ci sono stati forniti. Questi immatricolati sono riconducibili a 868 scuole, e sono così distribuiti: il 10% nell'Università Statale di Milano, 5.2% nell'Università Cattolica (sede di Milano o di Brescia, esclusa la sede di Piacenza), 3.7% nel Politecnico di Milano, 2.6% nell'ateneo pavese, 1.8% in quello bresciano e 1.6% in quello dell'Insubria.

Quando vengono incrociati con l'archivio sulle scuole, si incontrano 866 scuole⁹, di cui 526 nell'archivio di cui possediamo i dati individuali dei maturati (archivio Barbieri) e 340 scuole di cui possediamo indicazione del nome della scuola e, per un sottoinsieme di esse, solo il totale dei maturati o addirittura solo gli iscritti in classe quinta. È quindi possibile calcolare un "tasso di transizione" rapportando gli iscritti all'università per scuola al numero dei maturati. Eliminando 8 scuole per le quali questo tasso eccede l'unità¹⁰ e 144 scuole per le quali non si possiede neppure il totale degli iscritti, ci riduciamo a lavorare con 722 scuole, che sono potenzialmente ordinabili sulla base del tasso di transizione (ovviamente parziale, in quanto riferito ad un sottoinsieme degli atenei – vedi Tabella 17). La tabella presenta le diverse tipologie di istituti in ordine di probabilità di transizione, e ci permette di valutare il grado di distorsione del nostro campione a partire dalla parziale copertura degli atenei. Da un lato l'ordinamento delle tipologie di scuola è coerente con le aspettative teoriche: licei classici e scientifici assicurano il massimo tasso di transizione, mentre gli istituti professionali sono associati ai livelli minimi di probabilità.

⁹ Le due scuole che non trovano riscontro nell'archivio scuole sono caratterizzate dai codici MIPCX00005 e VATD190014 e riguardano rispettivamente 4 e 1 immatricolati.

¹⁰ Si tratta delle scuole contrassegnate dai codici BGPC04500R, BGPL055009, BSPM025008, COTD055005, MIPC18500P, MIPM04500G, MITA01901C e VAPC59500P, per un totale di 192 studenti immatricolati. Sette di esse non esistono nell'archivio originario (Barbieri) e vengono recuperate attraverso i dati sugli immatricolati: è quindi altamente probabile che non si tratti di scuole reali, ma di codici diversi usati dalle segreterie studenti delle università.

Tabella 17 – Tipologia delle scuole nel campione e immatricolazioni all'università
(2008-09)

tipo scuola	numero scuole	dimensione media	tasso transizione
liceo classico	48	94.33	0.53
liceo scientifico	147	91.48	0.53
liceo linguistico	28	27.61	0.48
istituto magistrale	37	114.38	0.41
IT turismo	6	65.83	0.38
IT att. sociali	3	97.33	0.37
IT agrario	11	76.64	0.33
liceo comm./magistrale	6	28.17	0.32
conservatorio	1	26.00	0.31
liceo artistico	25	72.64	0.30
IT comm.geometri	152	79.64	0.24
IT industriale	84	88.11	0.24
IP servizi sociali	4	16.50	0.22
IT geometri	24	49.42	0.19
IP agricoltura	13	30.00	0.17
IT aereonautico	5	21.40	0.17
ist. d'arte	10	50.40	0.15
IP commerciale	56	80.71	0.14
IP alberghiero	12	71.33	0.12
IP artigianato	50	48.60	0.11
Totale	722	77.68	0.32

Dall'altra, siccome ci aspettiamo che i diplomati di licei classici e scientifici tendano a proseguire al massimo grado, i dati della tabella ci dicono che dovremmo osservare tassi di transizione vicini al 100%; il riscontrare tassi molto più bassi ci suggerisce che in questa fascia di scuole perdiamo circa la metà degli studenti, che plausibilmente si iscrivono agli atenei lombardi di cui non disponiamo di informazioni (Milano-Bicocca, Bocconi, IULM, LIUC, Bergamo e san Raffaele) o ad Atenei fuori regione. Questo è parzialmente confermato dai dati riportati in Tabella 18, dove sono le province ad essere ordinate secondo i tassi di transizione. Poiché conosciamo dall'evidenza nazionale che la mobilità territoriale degli studenti iscritti ai corsi triennali è limitata, ci aspettiamo tassi di transizione maggiori nei capoluoghi di provincia per i quali abbiamo l'informazione dei corrispondenti atenei (Brescia, Pavia e Varese) rispetto ai capoluoghi i cui corrispondenti atenei non hanno fornito i dati (Bergamo). Milano ovviamente fa caso a sé, data la presenza di più atenei nel suo territorio.

Tabella 18 – Provincia di localizzazione delle scuole e immatricolazioni all'università
(2008-9)

provincia	numero scuole	dimensione media	tasso transizione
MI	252	84.10	0.36
PV	39	78.26	0.36
VA	80	74.32	0.36
CO	49	61.59	0.33
SO	22	56.41	0.32
LC	23	84.87	0.31
BS	97	76.34	0.29
LO	18	88.33	0.28
CR	30	83.30	0.27
BG	87	74.98	0.22
MN	25	66.72	0.13
Totale	722	77.68	0.32

Se si escludono le scuole che abbiano meno di 10 iscritti negli atenei di cui possediamo le informazioni, perdiamo ulteriori 307 scuole, riducendoci a 416. Queste sono le scuole su cui concentreremo la nostra analisi nelle pagine che seguono.

A ragioni di massima chiarezza ricordiamo che questo campione di scuole non può essere considerato come pienamente rappresentativo della realtà lombarda, in quanto nella sua formazione intervengono tre elementi di distorsione. Il primo è la partenza da un file di maturati in scuole statali, che rende impossibile valutare le transizioni all'università degli studenti provenienti dalle scuole paritarie (ma che non impedisce poi di ricostruirne la carriera una volta che essi siano iscritti all'università). Il secondo è che non copriamo tutti gli atenei lombardi, con significative esclusioni di alcune aree disciplinari. Per esempio le facoltà di economia sono particolarmente sotto-rappresentate, in quanto molto presenti negli atenei che non hanno fornito i dati (Bocconi, Bicocca, LIUC e Bergamo). Se i diplomati che si iscrivono nelle facoltà di economia sono significativamente diversi dagli studenti che si iscrivono in altre facoltà, noi otteniamo dei risultati che non sono rappresentativi dell'universo delle iscrizioni. Il terzo è che chi si iscrive all'università non è lo studente medio, ma generalmente quello con capacità di apprendimento superiori alla media. Di questo terzo elemento di distorsione discutiamo più ampiamente nel prossimo paragrafo. Auspichiamo comunque che in un prossimo futuro questi problemi di rappresentatività dei dati possano essere risolti da un intervento di autorità centrali che possano mettere a disposizione della ricerca informazioni più complete.

3. Il problema dell'autoselezione, e possibile rimedi per ridurlo

Il campione di studenti osservati è distorto dal problema dell'autoselezione, nel senso che per ogni scuola noi osserveremo soltanto gli studenti che:

- 1) hanno deciso di proseguire gli studi
- 2) fanno parte del campione di università per le quali disponiamo dei dati.

Questo problema se non affrontato inficia la bontà dei risultati dell'analisi sia per quanto riguarda la transizione dei diplomati verso l'istruzione terziaria sia per quanto riguarda il ranking delle scuole.

Probabilità di transizione

Nel valutare la probabilità di transizione noi osserviamo solo gli studenti che soddisfano entrambe le condizioni elencate sopra, e non è possibile distinguere tra coloro che non s'iscrivono e coloro che si iscrivono in atenei di cui non disponiamo dei dati. Illustriamo il problema con un esempio pratico: prendiamo due scuole identiche la cui distribuzione di voti di diploma è pari a quella esposta nella colonna (1) di Tabella 19, di cui abbiamo le informazioni. Supponiamo che le frequenze di transizione siano crescenti per fascia di voto e pari a quelle riportate in colonna (2). Ipotizziamo infine che gli studenti delle due scuole abbiano degli sbocchi diversi a livello di università in modo tale che per la scuola **A** osserviamo tutti gli studenti (colonna 3) mentre per la scuola **B** ne osserviamo solo una parte (colonna 4). Basandoci sul campione disponibile stimeremmo quindi una probabilità di transizione corretta e pari al 61% per la scuola **A**, mentre per la scuola **B** otterremmo una stima del 42% distorta dal fatto che gli altri studenti transitati si sono iscritti ad università di cui non disponiamo dei dati.

Tabella 19 - Il problema dell'autoselezione nella valutazione dei tassi di transizione al sistema universitario

	(1)	(2)	(3)	(4)
voto	diplomati	% transiz.	Scuola A	Scuola B
60	6	0%	0	0
61-70	20	25%	5	4
71-80	24	50%	12	9
81-90	24	75%	18	15
91-99	20	100%	20	12
100	6	100%	6	2
tot	100		61	42

Nel caso riuscissimo ad ottenere i dati anche degli altri atenei saremmo in grado di identificare tutti gli studenti che si iscrivono in Lombardia, stimando in modo approssimativamente corretto la probabilità di transizione verso l'istruzione terziaria scuola per scuola.¹¹ In assenza di questi dati aggiuntivi non c'è modo di risolvere il problema dell'autoselezione e quindi di stimare correttamente le probabilità di transizione.

¹¹ La stima sarebbe un'ottima approssimazione ancorché non esatta perché perderemmo comunque gli iscritti ad atenei fuori regione.

La probabilità di transizione ragionevolmente è una funzione:

- a) del tipo di scuola: verosimilmente i tassi di transizione non sono gli stessi
- b) della distanza del comune di residenza (per approssimazione, comune della scuola secondaria) dalle sedi universitarie
- c) del voto di diploma

Non possiamo usare il campione che abbiamo per stimare queste probabilità. Per noi una scuola buona dovrebbe essere quella che fa transitare più studenti. Due scuole simili (come **A** e **B** nell'esempio precedente) risulteranno sempre valutate in modo distorto, con **B** valutata peggio di **A** solo per il fatto di mandare gli studenti in Atenei non censiti nel nostro dataset.

Riduzione della distorsione: valutare la rappresentatività del campione per fascia di voto, sotto l'ipotesi che la probabilità di transizione sia crescente.

Poiché l'effetto delle variabili sopra menzionate non è necessariamente lineare (ad esempio la distanza può incidere maggiormente sugli studenti peggiori e/o sulle scuole tecniche) proponiamo un metodo per individuare i casi in cui il problema dell'autoselezione è più severo. L'unica ipotesi sottostante è che la probabilità di iscriversi all'università sia crescente con il voto di diploma a parità di altre circostanze (tipo di scuola e distanza) dopodiché analizziamo per celle (tipo di scuola \times vicini-lontani) cosa succede tra scuole "simili". Possiamo ad esempio ragionevolmente ipotizzare che tutti gli studenti con un voto di diploma superiore ad 80 proseguano gli studi. Controllando nei dati quanti studenti con voto superiore ad 80 osserviamo nella nostra banca dati possiamo ottenere un'approssimazione di quanto il problema dell'autoselezione incida sulla stima dei tassi di transizione di ciascuna scuola.

Ranking

Il problema dell'autoselezione come descritto sopra si riflette anche quando si tenti di produrre un ranking delle scuole secondarie sulla base dei risultati ottenuti dai rispettivi studenti all'università. Un modo per ridurre questa distorsione è utilizzare come peso delle osservazioni l'inverso della loro frequenza. Se ad esempio dalla scuola **B** si sono diplomati 20 studenti con un voto tra 91 e 99 ma noi ne osserviamo solo 12, possiamo dare a ciascuna di queste osservazioni un peso pari a $20/12=1.66$. Ciò equivale a ipotizzare che tutti quelli che escono dalla stessa scuola con la stessa fascia di voto siano uguali tra loro, sia che si iscrivano agli Atenei del nostro campione sia che non lo facciano. Il che equivale a calcolare quello che sarebbe successo se tutti si fossero iscritti facendo in media le stesse scelte. Questa procedura ha il vantaggio che il peso rimane distinto per ciascuna fascia di voto di ciascuna scuola e non rende necessario aggregare le distribuzioni. D'altro canto, è un'ipotesi molto forte perché non è detto che le università di cui disponiamo dei dati siano mediamente uguali a quelle che non sono presenti nella nostra banca dati. I campioni potrebbero non necessariamente essere rappresentativi della popolazione delle rispettive scuole e pertanto eventuali *ranking* prodotti non sarebbero generalizzabili all'intera scuola. Ad esempio, potrebbe accadere che gli studenti più abili (con voti di diploma più alti) della scuola **A** sostengano un test insieme a quelli meno abili (con voti di diploma più bassi) della scuola **B**: evidenza di

risultati migliori nei test di ammissione per la scuola **A** sarebbe in questo caso il frutto di una diversa abilità media degli studenti che hanno sostenuto il test, e non sarebbe evidenza del fatto che la scuola **A** prepari meglio gli studenti della scuola **B**. Lo stesso discorso vale nel caso stessimo confrontando la performance degli studenti alla fine del primo anno di università.

Certamente è possibile tenere conto di questo problema dato che viene osservato il voto di diploma, ma questo a sua volta può non essere un indicatore oggettivo dato che gli standard possono differire da scuola a scuola (si veda il paragrafo 8). Ad esempio, se una scuola tende ad inflazionare i voti, avremo che il voto di diploma di uno studente da essa proveniente non sarà un buon indicatore del suo livello di abilità. Per ridurre il problema di autoselezione in questo caso possiamo collocare gli studenti all'interno di una distribuzione relativa di voti all'interno della rispettiva scuola. Purtroppo disponiamo solo della distribuzione di voti di maturità per fasce, per cui una simile approssimazione incorpora degli inevitabili errori di misurazione. Vediamo con un esempio numerico, supponendo di nuovo che la distribuzione dei voti di diploma sia quella rappresentata nella tabella seguente, in cui due scuole con la stessa distribuzione di abilità tra gli studenti praticano diverse politiche di valutazione.

Tabella 20 - Il problema delle politiche di valutazione: effetti sulle distribuzioni di voto di diploma

voto	Scuola A		Scuola C	
	diplomati	cumulata	diplomati	cumulata
60	6	6	0	0
61-70	20	26	6	6
71-80	24	50	20	26
81-90	24	74	24	50
91-99	20	94	40	90
100	6	100	10	100

Per ciascuna fascia di voto definiamo come “cumulata” la somma della percentuale di studenti diplomati con un voto uguale o inferiore a quello della fascia stessa (la cumulata di frequenze).¹² Confrontando la scuola **A** e la scuola **C** si nota come la funzione di ripartizione della scuola **C** sia caratterizzata da dominanza stocastica del prim'ordine rispetto alla funzione di ripartizione della scuola **A**. In altre parole, i voti della scuola **C** sono sistematicamente più alti che nella scuola **A**. Di conseguenza, in ciascuna fascia di voto il percentile risulta (debolmente) inferiore nella scuola **C**. Dalle statistiche descrittive riportate nella Tabella 21 (e riferite all'insieme più ampio di scuole di cui abbiamo informazioni – fonte “Baggio”) emerge come effettivamente ci siano distribuzioni di voti di diploma estremamente eterogenee. Risulta emblematico in proposito il caso della fascia 71-80 in cui sono presenti sia scuole che non hanno alcuno studente con voti superiori che scuole in cui tutti gli studenti hanno voti superiori.

¹² Questa misura verrà anche utilizzata nel paragrafo 8 per costruire il ranking relativo dello studente per scuola. Per uno studente con voto 82 (che cade nell'intervallo di voti 81-90) proveniente dalla Scuola A il ranking relativo è 74%, mentre se proviene dalla scuola B è 50%.

Tabella 21 - Il problema delle politiche di valutazione: percentili all'interno della stessa scuola

fascia di voto	N	Media	Std. Dev.	Min	Max
60	796	0.134	0.094	0.000	0.545
61-70	796	0.481	0.159	0.000	0.900
71-80	796	0.740	0.135	0.000	1.000
81-90	796	0.880	0.087	0.333	1.000
91-99	796	0.955	0.049	0.588	1.000
100	796	0.996	0.010	0.900	1.000
100 e lode	796	1.000	0.000	1.000	1.000

È possibile procedere ad un raffinamento della distribuzione dei voti relativi a patto di accettare delle ipotesi sulla distribuzione dei voti all'interno di ciascuna fascia. Ad esempio, se supponiamo che la distribuzione sia uniforme, sarà possibile assegnare un percentile diverso a seconda del singolo voto di diploma. Facciamo un esempio: supponiamo che ci siano due studenti che provengono dalla scuola **C** nella Tabella 20 uno con un voto di diploma pari a 72/100 e l'altro con 78/100. Al momento assegneremo ad entrambi un percentile pari a 26, ovvero la frazione di studenti che nella medesima scuola hanno conseguito un voto non superiore a quello della fascia 71-80. Se noi supponiamo che la distribuzione sia uniforme, possiamo assegnare i 20 studenti in tale fascia sui 10 voti disponibili in ragione di 2 per ogni voto. Di conseguenza il percentile dello studente con voto pari a 72/100 risulterebbe essere 10, mentre il percentile dello studente diplomato con 78/100 sarebbe 22. Così facendo siamo in grado di distinguere anche all'interno della distribuzione relativa studenti che hanno un voto di diploma diverso anche se appartenente alla stessa fascia.

Ovviamente non si può affermare che le differenze nelle distribuzioni dei voti relativi sia interamente dovuta a diverse politiche di valutazione. Nell'esempio della Tabella 20 la diversità tra le distribuzioni potrebbe riflettere il fatto che per problemi di autoselezione nella scuola **C** gli studenti siano effettivamente migliori che nella scuola **A**. Ciò nonostante, le informazioni sulle distribuzioni relative dei voti possono essere fruttuosamente utilizzate come controllo di robustezza, ovvero ripetendo le analisi sia con i voti in termini assoluti sia in termini relativi, e analizzando se emergono delle differenze. Se i *grading standards* sono diversi tra scuole, ci aspettiamo una correlazione positiva tra performance e posizione nella distribuzione relativa. Infatti, nelle scuole dove i voti sono inflazionati, a parità di voto di diploma dovremmo osservare una performance peggiore associata ad un percentile più basso nella distribuzione relativa poiché la distribuzione cumulata risulta sistematicamente più bassa. Questo, di nuovo, risulta chiaro nella Tabella 20: confrontando la scuola **A** e la scuola **C** vediamo come in quest'ultima, dove abbiamo ipotizzato esserci inflazione di voti, la cumulata risulti più bassa a parità di fascia. Poiché anche in presenza di *grading standards* diversi la distribuzione di voti risulta comunque troncata, verificheremo se l'effetto dell'inflazione dei voti è non-lineare. Tale effetto potrebbe ad esempio risultare decrescente all'aumentare del voto di diploma nel caso in cui le differenze tra le performances risultassero più marcate per studenti con voti di diploma medio-bassi.

4. Accountability e indicatori di performance per le scuole superiori

La recente enfasi sulla valutazione dell'operato delle amministrazioni pubbliche ha posto al centro dell'attenzione il problema della cosiddetta *accountability* delle stesse nei confronti degli *stakeholders*: essendo esse destinatarie di ingenti flussi di risorse pubbliche si pone il problema di valutarne l'efficienza e l'efficacia, al fine di dirigere tali flussi laddove producono i migliori risultati o di migliorarne la performance, qualora vi siano delle situazioni problematiche. Con lo stesso termine di *accountability* si indica spesso anche l'insieme delle tecniche con cui possono essere misurati e giudicati i risultati effettivamente conseguiti dalle amministrazioni e attribuite ad esse le relative responsabilità, rendendole oggetto a seconda dell'esito delle procedure di valutazione di premio o di sanzione (Martini, 2008). *Accountability* è allora sinonimo di "merito": le amministrazioni pubbliche che operano bene dovrebbero essere premiate e quelle che operano male sanzionate, dove il premio e la sanzione assumono la forma di maggiori o minori trasferimenti di risorse pubbliche, la cui allocazione dovrebbe seguire criteri di efficacia ed efficienza.

Per valutare l'efficacia di un'amministrazione pubblica, è necessario innanzitutto definirne la funzione, in particolare ciò che tale amministrazione dovrebbe produrre, e conseguentemente l'*output* che si intende misurare. Concentriamoci ora su una particolare tipologia di amministrazioni pubbliche: le istituzioni educative. In questo caso, diverse alternative sono possibili. Ci si può focalizzare sui processi interni all'istituzione, come il processo di apprendimento, e in questo caso l'*output* può essere rappresentato dai risultati in test standardizzati, come ad esempio i test somministrati dall'INVALSI¹³ o il test PISA,¹⁴ oppure ci si può concentrare su processi esterni alla istituzione educativa, come agli esiti degli studenti negli stadi successivi del processo educativo, ad esempio la performance nell'istruzione terziaria, o sul mercato del lavoro in termini di occupabilità o di redditi percepiti.¹⁵ Nel primo caso si parla di *indicatori di performance interna* e nel secondo di *indicatori di performance esterna*. In entrambi i casi una caratteristica centrale di un buon indicatore di performance è quello di misurare il contributo specifico dato dall'istituzione educativa all'*outcome* considerato, al netto delle caratteristiche individuali e di contesto che potrebbero influenzarlo. Un buon indicatore di performance dovrebbe allora fornire una misura del cosiddetto "valore aggiunto" dell'istituzione educativa (Ricci, 2008).

Qui ci contenteremo su un particolare indicatore di performance esterna, il contributo dato dalle scuole superiori al successo accademico dei loro studenti iscritti all'università. Il "successo" verrà misurato in termini di:

- a) probabilità di continuare negli studi,
- b) risultati conseguiti nei test d'ammissione all'università,
- c) ammontare di crediti conseguiti

¹³ Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e Formazione.

¹⁴ *Programme for international student assessment* (OCSE).

¹⁵ Ovviamente lo scopo della scuola non è limitato all'apprendimento, e alla preparazione per il mondo del lavoro, svolgendo essa altre importanti funzioni, tra cui quelle di socializzazione e di formazione di cittadini consapevoli, in grado di partecipare attivamente alla vita politica e sociale del paese.

d) ammontare di voto medio negli esami (media pesata per il numero di crediti) nel primo anno di corso.

Il focus sul primo anno di corso è motivato dal design dell'esercizio che andremo a svolgere, ovvero dalla volontà di considerare la coorte più recente di immatricolati per cui abbiamo a disposizione anche delle misure di performance universitaria. C'è poi da dire che la performance degli studenti universitari nel primo anno di corso è anche una buona proxy di quella negli anni successivi (Aina et al. 2010). Valutare la performance delle scuole secondarie rispetto alla carriera universitaria dei loro studenti ha pregi e difetti. In assenza di test standardizzati sulle competenze curriculari o di *literacy* degli studenti secondari, l'utilizzo di indicatori come il punteggio di diploma per valutare il valore aggiunto delle scuole dà sicuramente adito a più di un'obiezione. Infatti, da un lato si tratta di misure che non sono necessariamente comparabili, a causa dei diversi *grading standard* che potrebbero essere utilizzati dalle diverse scuole (come illustrato nel paragrafo precedente), e che sono soggette ad un elevato grado di manipolabilità da parte delle stesse scuole: essendo consapevoli che il successo lavorativo o la probabilità di iscriversi (essere accettati) ad (da) un Ateneo dipendono dal punteggio del diploma, esse potrebbero alzarlo (*grade inflation*).

Da questo punto di vista l'utilizzo come misura di performance della rapidità e qualità delle carriere universitarie, è senz'altro meno soggetto alla manipolazione delle scuole e rappresenta un indicatore più oggettivo di performance. Dall'altro lato occorre ricordare che preparare gli studenti per l'università è il ruolo fondamentale solo di certi tipi di scuola, principalmente i licei, per cui risulta difficile ed ha poco senso tramite quell'indicatore esprimere un giudizio sull'operato delle scuole i cui studenti solo in minoranza si iscrivono all'università, come gli istituti professionali ed in misura inferiore gli istituti tecnici. Questi ultimi andrebbero allora valutati utilizzando un altro metro, ad esempio l'occupabilità degli studenti che non decidono di continuare nell'istruzione terziaria, e che plausibilmente rappresentano la maggioranza, e questo deve essere tenuto ben presente quando nel prosieguo verranno presentati i *ranking* dei diversi tipi di scuola.

Un esercizio simile a quello che andremo a svolgere qui è stato recentemente fatto da De Simone et al. (2009) sulle scuole Piemontesi. In questo lavoro gli autori hanno costruito degli indicatori di performance delle scuole secondarie piemontesi sulla base delle carriere universitarie degli studenti iscritti nei tre atenei statali del Piemonte (Università degli Studi di Torino, Politecnico di Torino e Università degli Studi del Piemonte Orientale). Il lavoro di De Simone et al. (2009) rappresenta il nostro punto di partenza, rispetto al quale la nostra analisi presenterà tuttavia degli elementi significativi di differenziazione. Le principali conclusioni dell'analisi di De Simone e colleghi possono essere così riassunte:

a) La migliore performance delle scuole della provincia rispetto a quelle cittadine, che gli autori ricollegano alla maggiore dotazione di capitale sociale nelle piccole comunità rispetto ai grandi centri urbani o ad un possibile meccanismo di autoselezione, dove gli studenti della provincia, che devono sostenere maggiori costi per istruirsi, sono in media anche quelli più capaci;

- b) La buona performance degli istituti tecnici, dalla quale gli autori concludono che le scuole di qualità (ovvero con elevato valore aggiunto) non sono solo confinate ai licei. Anche in questo caso, tuttavia, l'effetto della qualità della scuola potrebbe in parte confondersi con un effetto di selezione, come gli autori osservano, dato che gli studenti degli istituti tecnici hanno migliori opportunità sul mercato del lavoro, il costo opportunità di proseguire negli studi universitari è più elevato, e solo gli studenti più abili e con i maggiori rendimenti attesi continueranno;
- c) La pessima performance degli istituti non statali, in parte razionalizzabile con il fatto che spesso queste scuole assolvono in Italia il ruolo di "scuole di recupero" ed accolgono gli studenti meno abili (Bertola et al., 2007).

Tuttavia De Simone et al. (2009) non analizzano un elemento importante nel valutare le scuole, che è l'incoraggiamento alla prosecuzione agli ordini di scuola successivi, in questo caso l'università. Se una scuola prepara bene i propri studenti, essa li orienta verso l'alto e li rende anche sicuri delle proprie capacità. Per questo noi partiremo dal valutare il contributo di ogni singola scuola alla probabilità di transizione agli ordini scolastici superiori, per poi passare a considerare i risultati durante il primo anno di università, ed arrivare a confrontare i potenziali ordinamenti ottenibili da ciascuna di queste analisi.

Un'altra caratteristica distintiva della nostra ricerca rispetto a quella di De Simone et al. (2009) è che analizzeremo, dove possibile, la performance degli istituti superiori anche in funzione dei risultati ottenuti dagli studenti nei test di ammissione all'università. Nel nostro dataset composto complessivamente da 25226 studenti iscritti all'università sono presenti 9961 studenti che hanno sostenuto uno o più test standardizzati all'inizio del loro percorso di istruzione terziaria. Le statistiche descrittive sono riportate nella tabella 22 dove non è presente l'Università Cattolica poiché non ha rilasciato dati sui test. Tali test sono eterogenei e non immediatamente confrontabili tra loro, per cui procediamo a costruire diverse variabili per ciascuna tipologia omogenea di test.

Tabella 22 – Numero di studenti osservati nei vari test d'ingresso

	studenti	totale test	medicina	ingegneria	architettura	altro
Politecnico	6094	5960		3902	2380	
Univ.Milano	8696	1744	523			1221
Univ.Pavia	2842	424	183			241
Univ.Insubria	1452	389	110			279
Univ.Brescia	1633	1444	322			1122
	20717	9961	1138	3902	2380	2863

Al Politecnico di Milano disponiamo delle informazioni sul test d'ingresso di 5960 studenti: 3902 di ingegneria e 2380 di architettura (la cui somma non corrisponde al totale perché 322 hanno svolto entrambi i test). Non abbiamo informazioni confrontabili con altre università pertanto generiamo con queste osservazioni due variabili separate: "test-ingegneria" e "test-architettura." Sono altresì presenti nel dataset studenti provenienti da 4 diversi atenei che svolgono il test per una o più delle 4 classi di laurea in professioni sanitarie. Poiché tali test sono standardizzati riuniamo i 1138 studenti in

un'unica variabile: “test-medicina.” Ciò consente di conservare e sfruttare l'eterogeneità che emerge tra atenei e tra le diverse classi di laurea nelle professioni sanitarie, in modo che ad esempio l'esito di uno studente nella classe SNT/1 (Professioni Sanitarie Infermieristiche e Professione Ostetrica) dell'Università di Brescia sia direttamente comparabile a quello di uno studente nella classe SNT/3 (Professioni Sanitarie Tecniche) dell'Università di Milano. Questo spiega perché nella tabella 22 la media dei risultati sia diversa da zero per tali università. Ci sono infine altri 2863 studenti per cui è disponibile un qualche tipo di test, ma in questo caso spesso essi non risultano confrontabili tra loro. Per gli studenti dell'Università di Brescia disponiamo del voto in una scala omogenea (tra 0 e 1) nelle varie classi di laurea. Per gli altri atenei abbiamo invece un valore assoluto senza tuttavia poter distinguere su quale scala esso sia misurato. Di conseguenza per questi 1741 studenti procediamo a standardizzare (media 0 e varianza 1) il risultato all'interno di ogni classe di laurea di ciascun ateneo.¹⁶ Questo implica la perdita di alcune informazioni poiché equivale ad assumere che tali classi di laurea siano omogenee, ma consente di conservare tali osservazioni che altrimenti non sarebbero sfruttabili.

Infine, generiamo una variabile “test” in cui aggregiamo le informazioni disponibili nei vari test (ingegneria, architettura, medicina, Brescia, altro, una volta che tutti sono stati normalizzati con media 0 e varianza 1). Naturalmente questo implica che stiamo confrontando risultati non omogenei, ma eventuali differenze sistematiche che potessero esserci (ad esempio se il test di ingegneria è più difficile di quello che qui abbiamo definito “altro” e quindi viene fatto da studenti mediamente più bravi) potranno essere catturate da effetti fissi.

In questo modo sarà possibile utilizzare come misura di output il risultato nel test di ingresso contemporaneamente per tutti quegli studenti per cui abbiamo qualche informazione anche relativamente alla maturità. Si noti che la standardizzazione viene effettuata sull'universo degli iscritti nel nostro dataset, non solo relativamente a coloro che riusciamo a rintracciare anche tra i maturati. Il motivo è che non è detto che le osservazioni che risulteranno assenti (*missing*) siano distribuite in modo casuale.

Infatti, si può notare nella tabella 23 come alla riduzione della numerosità dovuta al fatto che non tutti gli iscritti abbiano una controparte nel file dei maturati corrisponda una media nei test sensibilmente maggiore. Ciò è spiegato dal fatto che per il Politecnico e l'Università di Pavia non è disponibile il codice meccanografico della scuola e di conseguenza non è possibile recuperare alcuna informazione per questi studenti (provenienti prevalentemente da istituti superiori non statali), i cui risultati nei test sono mediamente peggiori. Similmente, abbiamo una ulteriore riduzione di numerosità concentrandoci sulle scuole che hanno almeno 10 studenti nel nostro file di iscritti. Di nuovo, si osserva un incremento nei risultati medi nei test a significare che gli studenti provenienti da scuole che non hanno molti iscritti hanno risultati mediamente peggiori.

¹⁶ Così facendo si perdono 5 osservazioni poiché in alcune classi ci sono osservazioni singole.

Tabella 23 – Numero di studenti e risultati medi nei vari test d’ingresso standardizzati

	Tutti		Controparte maturati		Scuole n>10	
	N	media	N	media	N	media
Politecnico	5969	0.00	3208	0.11	2778	0.18
Univ.Milano	1737	0.04	1737	0.04	1101	0.15
Univ.Pavia	424	-0.10	318	0.08	224	0.16
Univ.Insubria	391	-0.14	389	-0.14	285	0.01
Univ.Brescia	1444	0.02	1432	0.03	1047	0.16
	9965	0.00	7084	0.06	5435	0.16

5. La metodologia statistica utilizzata

Qui descriviamo in maniera più dettagliata il tipo di esercizio svolto. Definiamo con y la particolare misura di output della scuola che ci interessa (probabilità di iscrizione all’università, crediti o voti conseguiti negli esami universitari durante il primo anno). L’approccio che utilizzeremo in un primo stadio è la stima di una funzione di produzione di performance - una relazione di tipo lineare che lega y ad una serie di input educativi:

$$y_{ij} = a_0 + a_1 X_i + a_2 S_j + a_3 \text{CLASSE}_i + a_4 \text{ATENEO}_i + \varepsilon_{ij} \quad (1)$$

dove i e j sono i pedici per l’individuo e la scuola, rispettivamente, che di seguito ometteremo. X è un vettore di caratteristiche individuali che ricomprende il genere, l’età, la regione di nascita e la tipologia di iscrizione all’università (precoce, standard o tardiva) dell’individuo, tutte informazioni che possiamo desumere per l’insieme dei diplomati dai loro codici fiscali. Queste caratteristiche possono o influenzare direttamente la probabilità di iscrizione e/o la performance universitaria, o indirettamente per il tramite della qualità degli input educativi ricevuti in passato (ad esempio per la regione di nascita). S è un vettore di caratteristiche di scuola, che include il tipo di scuola (indirizzo), il tipo di gestione (scuola statale, paritaria o parificata) e la provincia dove la scuola è localizzata. La giustificazione per l’inclusione del tipo di indirizzo secondario è quella di tenere conto del fatto che esistono differenze sostanziali nello *student intake* delle diverse scuole, per cui non ha alcun senso confrontare la performance universitaria di studenti con diploma di scuola professionale, ad esempio, con quelli di liceo. I primi differiscono in maniera sostanziale dai secondi non solo in termini di abilità accademica e di motivazioni allo studio, ma anche in termini di aspirazioni educative, essendo per i primi la decisione di iscriversi agli studi universitari un’eccezione mentre per i secondi la regola. Per la stessa ragione abbiamo incluso il tipo di gestione, dato che molte scuole private assolvono la funzioni di recupero scolastico (*remedial schools*) come osservato da Bertola et al. (2007), mentre l’effetto fisso provinciale ha lo scopo di cogliere differenze territoriali nel mercato del lavoro, che potrebbero influire sui costi e sui rendimenti di iscriversi all’università. Infine **CLASSE** rappresenta la “classe di laurea” ed **ATENEO** l’ateneo di iscrizione. Queste due serie di effetti fissi servono per tenere conto del fatto che

diversi corsi di laurea o istituzioni universitarie potrebbero utilizzare diversi standard educativi (*grading standard*): in alcuni corsi di laurea i docenti potrebbero essere più inclini a dare voti alti agli studenti, o gli stessi corsi potrebbero essere oggettivamente meno impegnativi, ed in entrambi i casi osserveremmo delle performance migliori degli studenti (Bagues et al., 2008). Qualora gli studenti di certe scuole tendessero sistematicamente ad iscriversi a corsi di laurea più facili (difficili) e non controllassimo per la classe del corso di laurea tenderemmo ad osservare delle performance relativamente migliori (peggiori) per quelle scuole, che riflettono tuttavia in maniera del tutto spuria un processo di autoselezione degli studenti, e non il “valore aggiunto” delle scuole.¹⁷

L’equazione (1) verrà stimata col metodo dei minimi quadrati ordinari (*ordinary least squares*, OLS). Ottenuti i coefficienti dei regressori inclusi è possibile calcolare i residui stimati $\hat{\varepsilon}$ che rappresentano la parte della performance universitaria degli studenti che non è spiegata dalle variabile esplicative inserite nel modello. I residui stimati verranno utilizzati come variabile dipendente in un secondo stadio, in cui verranno regrediti su degli indicatori di scuola D :

$$\hat{\varepsilon}_{ij} = b_0 + b_j D_j + u_i \quad (2)$$

dove u è un termine di errore ortogonale agli indicatori di scuola. I nostri coefficienti di interesse sono i b_j , che rappresentano le differenze medie nella performance (aggiustata per i fattori contestuali del primo stadio) tra le singole scuole e la scuola di riferimento.¹⁸ Gli effetti fissi di scuola \hat{b}_j stimati dalla (2) con il metodo OLS verranno poi rappresentati graficamente insieme ai rispettivi intervalli di confidenza, per avere un’idea del *ranking* delle diverse scuole.¹⁹

Date le caratteristiche del dataset (vedi paragrafo 1) il campione oggetto dell’analisi escluderà l’Università Cattolica di Milano nelle analisi della performance (ma non dell’iscrizione universitaria), in quanto diverse informazioni sugli studenti immatricolati non sono state fornite. Al fine di ottenere delle stime affidabili degli effetti di scuola, le analisi riportate di seguito si riferiscono solo al campione di scuole per cui esistono almeno 10 studenti iscritti all’università e per cui i regressori che compaiono nella (1) sono non missing (De Simone et al. 2009). Per controllare la robustezza delle stime al variare del campione abbiamo anche stimato lo stesso modello per le scuole con almeno 20 studenti.

¹⁷ Dato che non rappresentano l’obiettivo principale dell’esercizio in questo capitolo, e per mantenere l’anonimità degli Atenei, non verranno riportati, né commentati, gli effetti degli Atenei e quelli delle classi di laurea.

¹⁸ Come noto, un effetto fisso di scuola deve essere omissivo. La scuola corrispondente rappresenta allora la “scuola di riferimento” ovvero il benchmark rispetto alla quale le differenze di performance sono espresse. La scuola di riferimento è un liceo scientifico, statale, situato nella provincia di Brescia.

¹⁹ La ragione per cui la regressione non è stimata in un solo stadio è che includendo gli effetti fissi di scuola direttamente nella (1) si pone un problema di multicollinearità con le variabili di scuola S . Nel primo stadio le osservazioni sono state clusterizzate per scuola e sono stati calcolati standard error robusti per l’eteroschedasticità.

6. La probabilità di proseguire nell'istruzione terziaria

Una prima dimensione sulla quale proviamo ad ordinare le scuole è data dalla capacità di spingere/motivare gli studenti a proseguire negli studi universitari. Date le poche informazioni di cui disponiamo a livello individuale (genere, età, regione di nascita) e a parità di indirizzo secondario frequentato (informazione che incorpora parte del condizionamento proveniente dall'ambiente familiare), identifichiamo come scuole più efficaci quelle che alzano la probabilità di iscrizione a parità di altre caratteristiche. Dai residui dei modelli stimati si procede poi ad una regressione di secondo stadio dei residui stessi su 667 effetti fissi di scuola; il contributo individuale di ciascuna scuola²⁰ viene rappresentato in Figura 1, all'interno del suo intervallo di oscillazione dato dalla sua significatività statistica. In Tabella 24 sono riportate le stime di due modelli, che fanno assunzioni diverse sulla distribuzione dei residui (non lineari e lineari). Nelle colonne (1) e (3) si considerano solo le informazioni anagrafiche, mentre nelle colonne (2) e (4) si introduce il tipo di scuola frequentata.

Dalla tabella si evince che le ragazze sono più inclini a proseguire con l'iscrizione universitaria (anche se questo effetto scompare statisticamente quando si tiene conto che le ragazze sono relativamente più frequenti in alcune tipologie di scuola). In svantaggio significativo troviamo gli studenti nati all'estero, in parte a causa del tipo di scuola secondaria che frequentano (in massima parte in Istituti professionali o tecnici) o quelli nati in regioni meridionali, ad indicazione che questa variabile potrebbe catturare una parte dell'effetto legato all'ambiente culturale della famiglia.

Quando teniamo anche conto dell'indirizzo scolastico secondario (misurato per differenza rispetto ai diplomati dei licei scientifici, tutti gli indirizzi (eccetto che nel caso del liceo classico) sono associati ad una penalizzazione in termini di probabilità, con valori negativamente più consistenti associati agli istituti professionali. Da ultimo la provincia di localizzazione delle scuole ci dà conferma della distorsione implicita nella costruzione del nostro campione: le province per le quali possediamo dati sulle immatricolazioni hanno coefficienti positivi, mentre l'opposto accade per quelle i cui atenei non sono stati disponibili a fornire informazioni.

²⁰ In riferimento ad una specifica scuola esclusa, il liceo scientifico di Brescia indicato con il codice BSPS11000A, individuato per avere un elevato numero di studenti iscritti all'università.

Tabella 24. Stime della probabilità di iscrizione all'università (scuole con almeno 10 studenti maturati)

	modello probit		linear probability model	
	(1)	(2)	(3)	(4)
donna	0.120*** (0.027)	0.017 (0.019)	0.046*** (0.010)	0.004 (0.006)
età	-0.041*** (0.007)	-0.008 (0.007)	-0.011*** (0.002)	-0.002 (0.002)
status di immigrato	-0.340*** (0.033)	-0.165*** (0.034)	-0.122*** (0.011)	-0.051*** (0.009)
regione ignota	1.443*** (0.559)	1.053* (0.631)	0.484*** (0.113)	0.281** (0.135)
Abruzzo	-0.010 (0.224)	-0.158 (0.246)	-0.006 (0.084)	-0.064 (0.080)
Basilicata	-0.407* (0.235)	-0.426 (0.266)	-0.137** (0.069)	-0.106* (0.059)
Calabria	-0.261*** (0.067)	-0.166** (0.070)	-0.093*** (0.023)	-0.055*** (0.020)
Campania	-0.336*** (0.066)	-0.299*** (0.068)	-0.119*** (0.021)	-0.086*** (0.018)
Emilia-Romagna	-0.139 (0.110)	-0.057 (0.087)	-0.052 (0.040)	-0.012 (0.027)
Friuli-Venezia-Giulia	0.630** (0.270)	0.553** (0.268)	0.242** (0.100)	0.197** (0.089)
Lazio	0.019 (0.114)	-0.065 (0.121)	0.006 (0.043)	-0.026 (0.040)
Liguria	0.385*** (0.132)	0.149 (0.149)	0.149*** (0.052)	0.054 (0.052)
Marche	0.340 (0.265)	0.275 (0.254)	0.127 (0.100)	0.089 (0.088)
Molise	-0.109 (0.432)	0.056 (0.445)	-0.039 (0.151)	0.007 (0.136)
Piemonte	-0.035 (0.131)	-0.281*** (0.097)	-0.013 (0.049)	-0.090*** (0.031)
Puglia	-0.220*** (0.076)	-0.144* (0.082)	-0.082*** (0.026)	-0.048* (0.025)
Sardegna	-0.350** (0.169)	-0.284 (0.187)	-0.124** (0.054)	-0.087* (0.049)
Sicilia	-0.184*** (0.058)	-0.114* (0.060)	-0.067*** (0.020)	-0.038** (0.018)
Toscana	0.357** (0.169)	0.015 (0.180)	0.137** (0.067)	0.009 (0.059)
Trentino-Alto Adige	-0.274 (0.175)	-0.154 (0.213)	-0.100* (0.060)	-0.051 (0.063)
Umbria	0.013 (0.397)	-0.366 (0.504)	0.008 (0.152)	-0.136 (0.163)
Valle d'Aosta	0.669 (0.762)	0.040 (0.728)	0.264 (0.280)	0.022 (0.257)
Veneto	-0.408*** (0.116)	-0.137 (0.111)	-0.140*** (0.036)	-0.027 (0.031)
liceo classico		-0.129 (0.087)		-0.047 (0.033)
istituto magistrale		-0.362*** (0.062)		-0.143*** (0.024)
IP agricoltura		-1.374*** (0.074)		-0.450*** (0.021)

IP commerciale	-1.338*** (0.069)	-0.451*** (0.019)		
IP alberghiero	-1.631*** (0.116)	-0.496*** (0.027)		
IP artigianato	-1.706*** (0.063)	-0.520*** (0.016)		
ist. d'arte	-1.341*** (0.154)	-0.446*** (0.049)		
liceo artistico	-0.629*** (0.072)	-0.241*** (0.026)		
IT agrario	-0.612*** (0.097)	-0.238*** (0.033)		
IT comm.geometri	-0.975*** (0.041)	-0.356*** (0.014)		
IT att.sociali	-0.409*** (0.115)	-0.175*** (0.055)		
IT industriale	-0.803*** (0.056)	-0.304*** (0.019)		
IT geometri	-0.999*** (0.104)	-0.363*** (0.031)		
IT turismo	-0.639*** (0.085)	-0.240*** (0.032)		
prov.scuola BG	-0.472*** (0.063)	-0.131*** (0.018)		
prov.scuola CO	0.263*** (0.064)	0.089*** (0.021)		
prov.scuola CR	-0.065 (0.090)	-0.017 (0.028)		
prov.scuola LC	-0.006 (0.097)	-0.001 (0.031)		
prov.scuola LO	0.125 (0.100)	0.044 (0.035)		
prov.scuola MI	0.177*** (0.049)	0.061*** (0.016)		
prov.scuola MN	-0.759*** (0.126)	-0.204*** (0.036)		
prov.scuola PV	0.455*** (0.075)	0.157*** (0.026)		
prov.scuola SO	-0.005 (0.077)	0.002 (0.023)		
prov.scuola VA	0.313*** (0.054)	0.107*** (0.018)		
Pseudo R ² /R ²	0.0107	0.1457	0.013	0.186
N	54505	53328	54505	53741

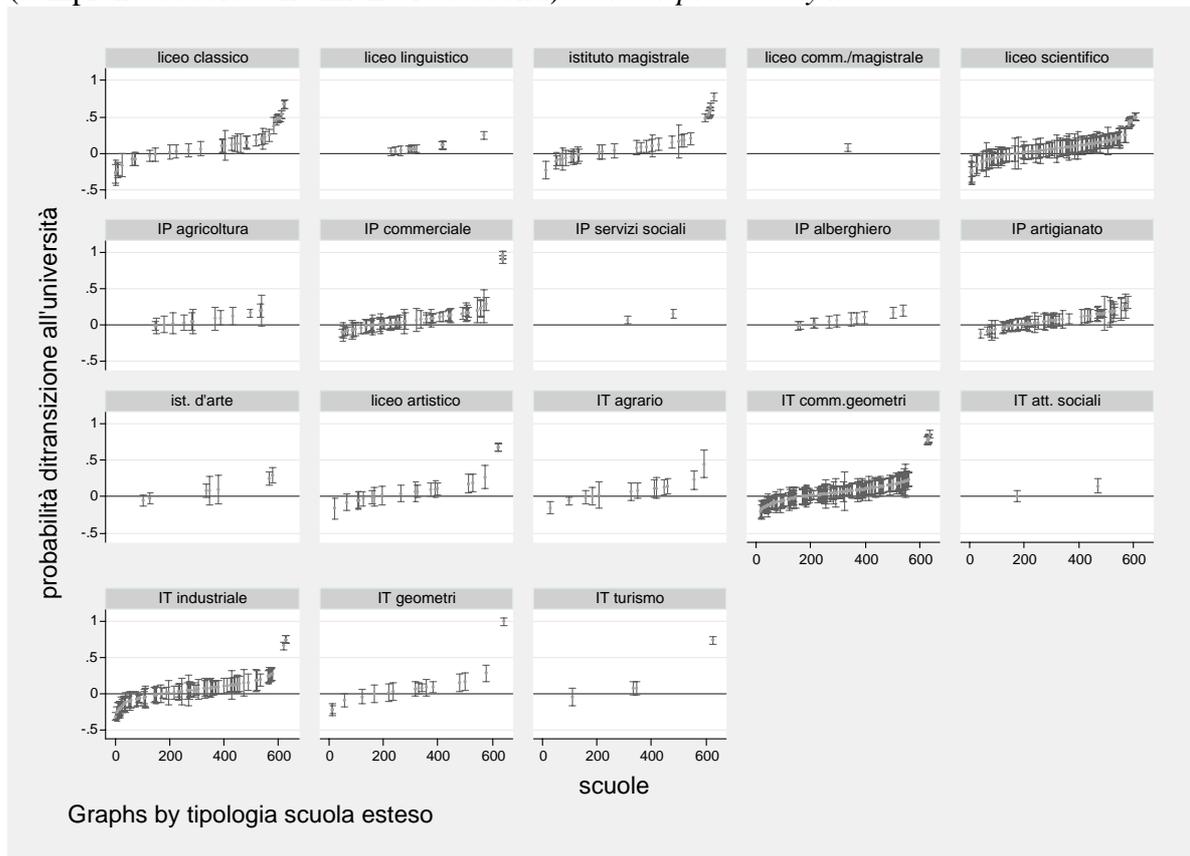
*, **, *** significativo al 10%, 5% e 1%, rispettivamente

Note: La tabella riporta i coefficienti (e gli errori standard tra parentesi – clusterizzati a livello di scuola) di due modelli (probit e *linear probability model*) in cui la variabile dipendente è risultare iscritto in uno dei sei atenei che ha reso disponibili i propri dati. Per ogni variabile dipendente sono riportate le stime sul campione delle scuole che includono almeno 10 studenti. Per le variabili categoriche la categoria di riferimento è: maschio/nato in Lombardia/diplomato in un liceo scientifico della provincia di Brescia

Una volta stimato il modello, i residui stimati vengono regrediti sugli effetti fissi di scuola, e i coefficienti vengono riportati per tipologia di scuola nella figura 1. La figura mostra come la massa delle scuole sia distribuita nell'intorno dello zero, indicando quindi l'assenza di differenziazione all'interno della stessa tipologia di scuola. Tuttavia

gli estremi (di eccellenza o di mediocrit ) sono chiaramente identificabili, e potrebbero essere identificati. Tuttavia, sappiamo dal terzo paragrafo che alcune scuole che risultano scarse nella loro capacit  di spingere i propri diplomati alla prosecuzione degli studi potrebbero in realt  essere scuole i cui studenti si iscrivono principalmente agli atenei che non osserviamo. Per questa ragione non riteniamo opportuno spingerci oltre nella interpretazione di questi risultati. Tuttavia il grafico della Figura 1   esemplificativo dei problemi che si presentano quando si voglia cercare di ordinare le scuole in riferimento ad un determinato criterio di risultato: mentre i risultati estremi sono facilmente identificabili, discriminare nella massa centrale   invece impossibile. Questo pu  rappresentare un risultato desiderabile dal punto di vista delle politiche scolastiche, perch  indica che il sistema   sufficientemente omogeneo da non discriminare tra studenti che abbiano scelto un certo ordine di scuola, una volta che si sia tenuto conto delle caratteristiche anagrafiche degli stessi.

Figura 1. Ranking degli effetti fissi di scuola sulla probabilit  di iscrizione all'universit  (campione scuole con almeno 10 studenti) – *linear probability model*



Note. La figura mostra un cross-plot degli effetti fissi di scuola ed i relativi intervalli di confidenza ottenuti utilizzando come variabile dipendente i residui dalla regressione mostrata nella seconda colonna della Tabella 24.

7. Risultati dell'analisi sul numero di crediti acquisiti e sul voto medio degli esami di profitto

Le prime due colonne della Tabella 25 riportano le stime relative al numero di crediti conseguito nel primo anno di corso.²¹ Dalle stime ottenute nel campione delle scuole con almeno 10 studenti, emerge che le donne hanno carriere in media più rapide dei maschi (2.4 crediti), mentre la performance scende al crescere dell'età degli studenti. Le stime relative al tipo di scuola mostrano come solo i licei classici e quelli linguistici non abbiano performance statisticamente diverse da quelle dei licei scientifici. Tutte le altre scuole hanno invece performance peggiori, che variano dai circa -4 crediti per gli istituti tecnici industriali ai circa -14 punti per gli istituti d'arte.

Per quanto riguarda la performance delle scuole per provincia, Bergamo (+2.8), Como (+2.7), Lecco (+5.4), Monza (+7.2) e Varese (+2.3) hanno performance migliori delle scuole situate a Brescia, mentre le scuole di Pavia (-2.8) hanno performance in media peggiori. Gli studenti che si sono iscritti precocemente all'università hanno performance migliori di quelli iscritti nel periodo standard (+5.9 crediti) e di quelli che si sono iscritti in ritardo (+11.5 crediti). Infine, le scuole paritarie non hanno performance statisticamente diverse da quelle statali.

Se consideriamo le stime nel secondo campione, quello che include le scuole con almeno 20 studenti, le differenze principali sono che ora anche i licei linguistici hanno performance peggiori dei licei scientifici (-5.9 crediti), che la provincia di Sondrio ottiene risultati migliori di quella di Brescia (+3.6 crediti) e che gli studenti provenienti dalle scuole paritarie ora conseguono risultati migliori di quelli nelle scuole statali (+9.1). Quest'ultimo effetto potrebbe essere dovuto al fatto che nella colonna (1) considerando nel campione scuole paritarie in cui un basso numero di studenti continua nell'istruzione universitaria, è più probabile che stiamo considerando *scuole di recupero* ovvero scuole destinate al recupero degli anni scolastici ed in cui lo studente medio ha bassa abilità. Considerando invece scuole private dalle quali proviene un numero sufficientemente elevato di studenti universitari è plausibile che la qualità media del corpo studentesco migliori.

La terza e quarta colonna della Tabella 25 riportano invece i risultati delle analisi sul voto medio pesato per il numero di crediti acquisiti, un indicatore non solo della rapidità ma anche della qualità complessiva del percorso universitario. Partendo dalle stime sul primo campione, viene confermata la migliore performance delle donne (+0.4) e il gradiente negativo dell'età. Utilizzando questo secondo indicatore di performance, i licei classici vanno meglio di quelli scientifici (+0.2), mentre gli istituti magistrali (-0.5), gli istituti professionali per il commercio (-1), gli istituti professionali alberghieri (-0.8), gli istituti d'arte (-0.9), i licei artistici (-0.7), gli istituti tecnici agrari (-0.8), gli istituti tecnici commerciali e per geometri (-0.6), gli istituti tecnici per le attività sociali (-0.7), gli istituti tecnici industriali (-0.5), gli istituti tecnici per geometri (-1.1) e gli istituti tecnici per il turismo (-0.8) sono associati a carriere peggiori.

²¹ In realtà si tratta di un po' più di un anno riferendosi le variabili di outcome all'1/11/2009.

Per quanto riguarda le province, le scuole di Bergamo, Lecco e Monza, che ottenevano risultati migliori di Brescia anche riguardo al primo indicatore di performance, confermano la posizione di vantaggio rispetto alla provincia di riferimento (Brescia), mentre Pavia mantiene una posizione di svantaggio. Sotto la performance di Brescia si colloca anche quella media delle scuole di Cremona. Gli studenti con iscrizione precoce ottengono risultati migliori di quelli con iscrizione standard (+0.4) e con iscrizione tardiva (+1). Per questo secondo indicatore di performance non si rileva invece alcuna differenza delle scuole paritarie rispetto a quelle statali. Considerando il secondo campione i risultati rimangono pressoché invariati, da notare solo che si perde il vantaggio della provincia di Monza rispetto a quella di Brescia.

Tabella 25. Stime OLS per il numero di crediti ed il voto medio negli esami

	Numero crediti		Voto medio pesato	
	10 studenti (1)	20 studenti (2)	10 studenti (3)	20 studenti (4)
donna	2.352*** (0.360)	2.126*** (0.382)	0.367*** (0.055)	0.360*** (0.057)
età	-0.688*** (0.222)	-0.704*** (0.257)	-0.048** (0.021)	-0.067** (0.027)
<i>Tipo di scuola (liceo scientifico)</i>				
liceo classico	-0.435 (0.665)	-0.259 (0.672)	0.204** (0.098)	0.188* (0.099)
liceo linguistico	0.955 (3.121)	-5.929*** (2.272)	0.022 (0.444)	-0.262 (0.419)
istituto magistrale	-4.655*** (0.735)	-5.086*** (0.587)	-0.452*** (0.087)	-0.507*** (0.084)
IP commerciale	-4.918* (2.739)	-5.539 (4.926)	-0.953*** (0.268)	-1.074*** (0.317)
IP alberghiero	-5.886*** (0.935)	-	-0.775*** (0.181)	-
Ist. d'arte	-13.869*** (1.414)	-	-0.897*** (0.179)	-
liceo artistico	-7.740*** (1.086)	-8.529*** (1.118)	-0.699*** (0.119)	-0.826*** (0.126)
IT agrario	-5.128*** (1.127)	-4.754*** (1.112)	-0.786*** (0.221)	-0.495** (0.197)
IT comm. Geometri	-4.928*** (0.769)	-4.121*** (0.901)	-0.577*** (0.098)	-0.586*** (0.113)
IT att sociali	-6.071*** (1.555)	-4.642*** (0.597)	-0.684*** (0.135)	-0.623*** (0.100)
IT industriale	-3.922*** (0.775)	-3.494*** (0.823)	-0.521*** (0.098)	-0.412*** (0.098)
IT geometri	-9.147*** (1.160)	-7.643*** (0.986)	-1.061*** (0.170)	-0.947*** (0.127)
IT turismo	-6.480*** (1.264)	-6.266*** (1.260)	-0.848*** (0.130)	-0.954*** (0.130)
<i>Provincia scuola (BS)</i>				
BG	2.763*** (1.032)	3.324*** (1.088)	0.344** (0.152)	0.317* (0.171)
CO	2.672** (1.126)	2.923** (1.156)	0.143 (0.166)	0.152 (0.179)
CR	-0.414	-0.320	-0.503***	-0.548***

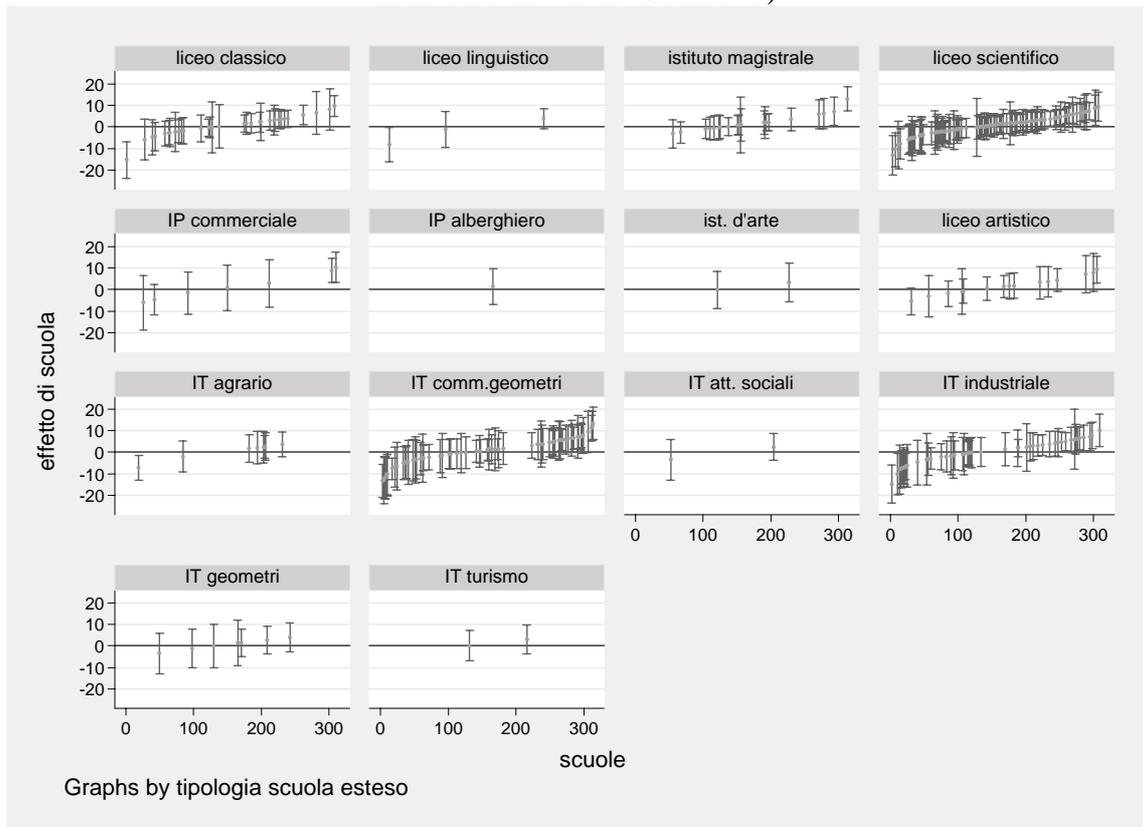
	(1.250)	(1.342)	(0.151)	(0.160)
LC	5.427***	5.738***	0.529**	0.578**
	(1.194)	(1.207)	(0.222)	(0.232)
LO	0.029	0.779	-0.295*	-0.325*
	(1.364)	(1.325)	(0.157)	(0.170)
MI	0.196	0.567	-0.118	-0.102
	(0.793)	(0.839)	(0.151)	(0.166)
MN	7.225***	6.723**	0.752**	0.678
	(2.192)	(2.685)	(0.348)	(0.426)
PV	-2.811**	-3.314**	-0.453**	-0.514***
	(1.301)	(1.351)	(0.181)	(0.195)
SO	2.296*	3.639***	0.278	0.384
	(1.344)	(1.187)	(0.342)	(0.364)
VA	2.268**	2.612***	-0.042	-0.068
	(0.944)	(0.987)	(0.165)	(0.180)
<i>Tipo di iscrizione (precoce)</i>				
standard	-5.931***	-5.895***	-0.411***	-0.451***
	(0.486)	(0.514)	(0.065)	(0.068)
tardiva	-11.470***	-11.126***	-1.046***	-1.032***
	(0.754)	(0.811)	(0.108)	(0.116)
<i>Tipo di gestione scuola (statale)</i>				
paritaria	-0.482	9.103***	-0.232	0.364
	(2.197)	(2.210)	(0.231)	(0.419)
R ²	0.215	0.217	0.263	0.267
N	12408	10924	11492	10130

*, **, *** significativo al 10%, 5% e 1%, rispettivamente

Note: La tabella riporta i coefficienti (e gli errori standard tra parentesi) di un modello lineare stimato con OLS in cui le variabili dipendenti sono il numero di crediti e il voto medio negli esami di profitto (pesato per il numero di crediti). I modelli includono anche effetti fissi per la classe di laurea, l'Ateneo di iscrizione, e la regione di nascita dell'individuo. Per ogni variabile dipendente sono riportate le stime sul campione delle scuole che includono almeno 10 studenti e su quello che include almeno 20 studenti iscritti agli Atenei del nostro campione. Per le variabili categoriche la categoria di riferimento è indicata tra parentesi.

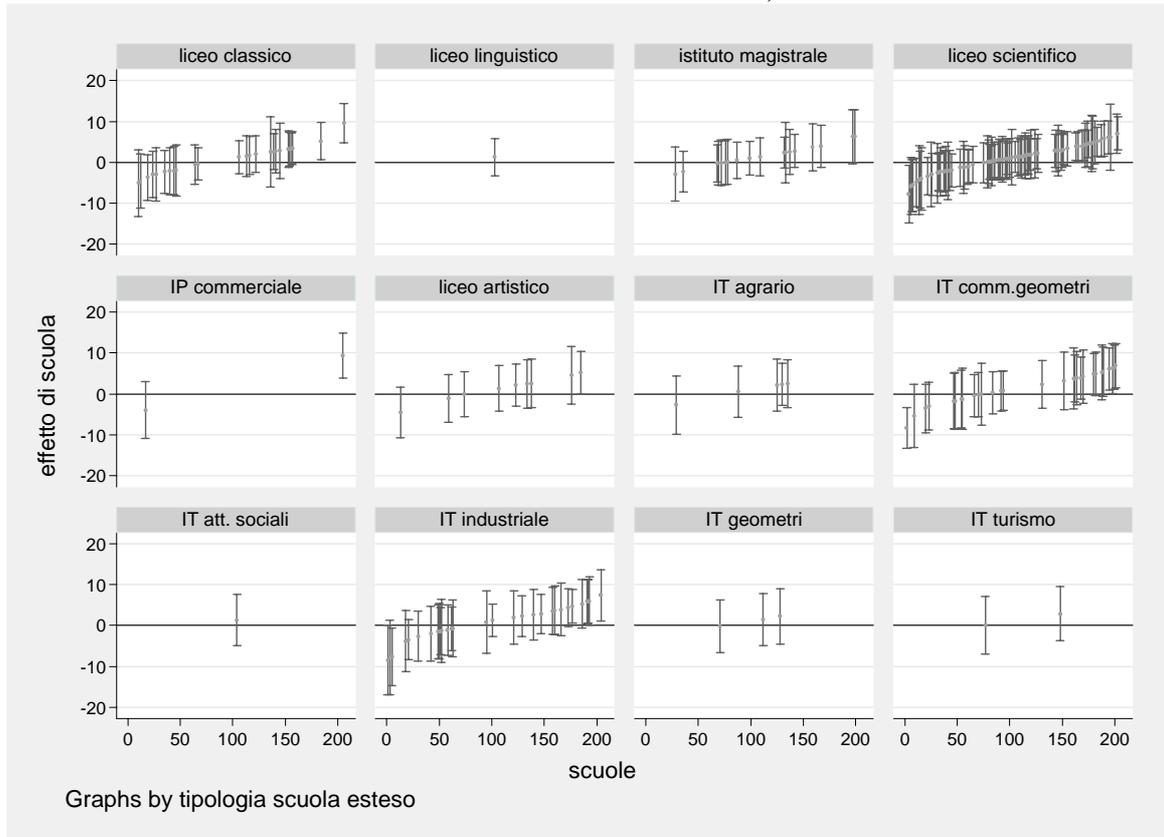
Le figure 2-5 riportano gli effetti fissi di scuola, ed i relativi intervalli di confidenza, stimati nel secondo stadio. La Figura 2 mostra i ranking ottenuti per il numero di crediti sul campione con 10 studenti. Da notare come vi sia una sostanziale eterogeneità tra scuole anche dello stesso tipo, e come posizioni di eccellenza nel ranking non siano appannaggio esclusivo dei licei. Vi è infatti un consistente numero di istituti tecnici e istituti commerciali e per geometri con performance molto elevate e con stime puntuali degli effetti fissi superiori a quelle di molti licei classici o scientifici. La Figura 3 mostra come queste conclusioni siano robuste rispetto alla selezione del campione.

Figura 2. Ranking degli effetti fissi di scuola sul numero di crediti acquisiti (campione scuole con almeno 10 studenti)



Note. La figura mostra un cross-plot degli effetti fissi di scuola ed i relativi intervalli di confidenza ottenuti utilizzando come variabile dipendente i residui dalla regressione mostrata nella prima colonna della Tabella 25.

Figura 3. Ranking degli effetti fissi di scuola sul numero di crediti acquisiti (campione scuole con almeno 20 studenti)

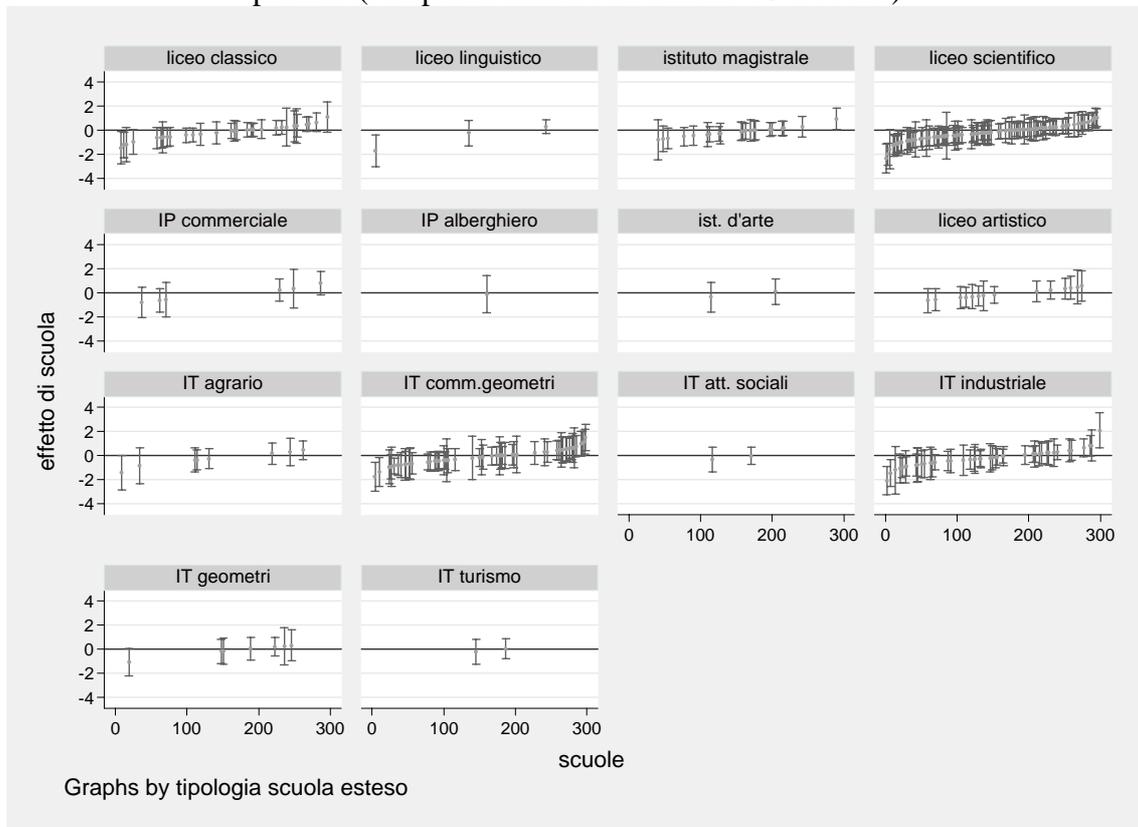


Note. La figura mostra un cross-plot degli effetti fissi di scuola ed i relativi intervalli di confidenza ottenuti utilizzando come variabile dipendente i residui dalla regressione mostrata nella seconda colonna della Tabella 25.

La Figura 4 riporta gli effetti fissi di scuola relativi al voto medio pesato negli esami sul campione delle scuole con almeno 10 studenti universitari. Anche in questa figura si rilevano le eccellenze di molti istituti industriali, commerciali e per geometri, ed anche in questo caso il confronto con la Figura 5 conferma la robustezza dei risultati ottenuti utilizzando il secondo campione.

Un elemento degno di nota è che, una volta aggiustato nel primo stadio per una serie di variabili di contesto, molte scuole non risultano statisticamente differenti dalla scuola di riferimento, ovvero gli intervalli di confidenza della maggior parte degli effetti fissi includono lo zero. Spesso non è possibile stabilire confronti statisticamente significativi tra coppie di scuole, dato che gli intervalli di confidenza relativi ai loro effetti fissi si “sovrappongono”. Ciò potrebbe essere tuttavia spiegato dalla sostanziale omogeneità delle caratteristiche delle scuole Lombarde in termini di input scolastici, ad esempio di livello di preparazione e motivazione degli insegnanti. Si riconferma pertanto l'utilità di questi esercizi non tanto in termini di stabilire delle classifiche tra scuole, ma di individuare dei casi di eccellenza o di problematicità, che potrebbero essere oggetto di ulteriori approfondimenti.

Figura 4. Ranking degli effetti fissi di scuola sul voto medio pesato negli esami di profitto (campione scuole con almeno 10 studenti)



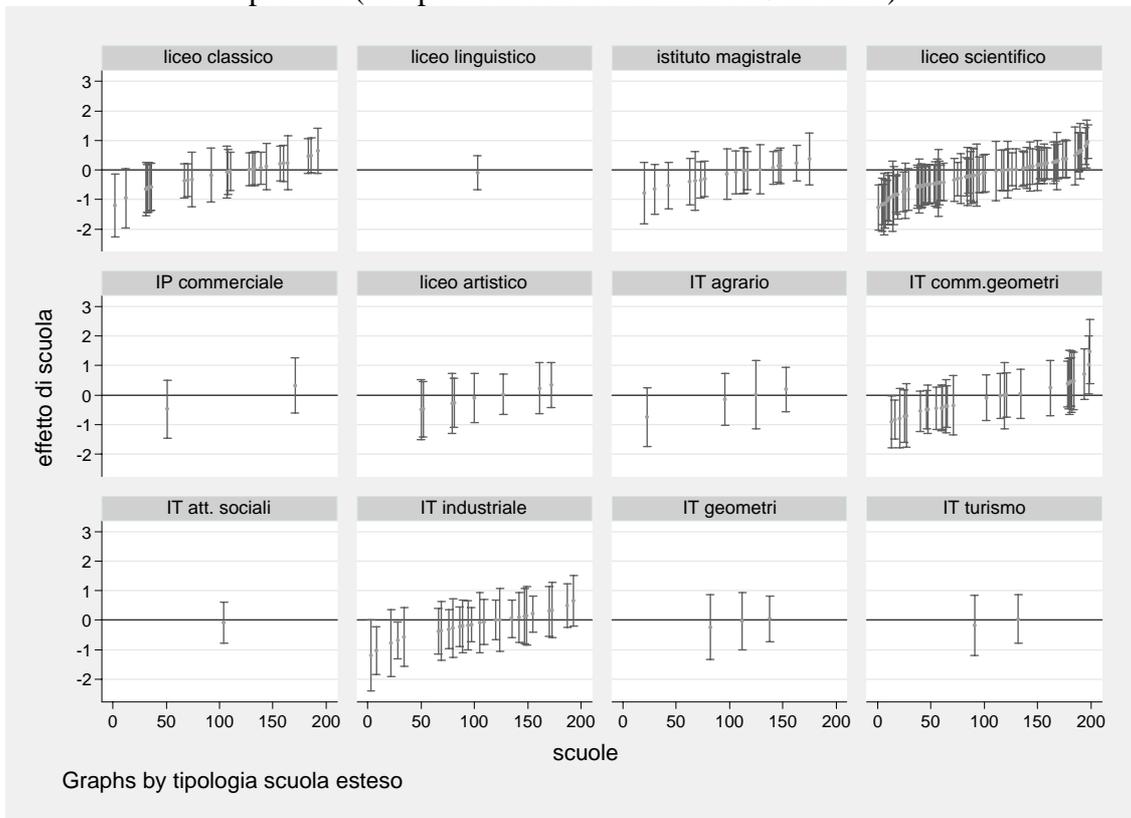
Note. La figura mostra un cross-plot degli effetti fissi di scuola ed i relativi intervalli di confidenza ottenuti utilizzando come variabile dipendente i residui dalla regressione mostrata nella terza colonna della Tabella 25.

Infine, vale la pena qui di evidenziare un limite principale del nostro esercizio, oltre ai problemi di rappresentatività del nostro campione di cui abbiamo già parlato nelle sezioni 2 e 3. Gli indicatori che abbiamo considerato sono prettamente “quantitativi”, indicano la rapidità nella carriera universitaria ed il livello dei voti negli esami di profitto, ma non considerano invece altri aspetti qualitativi altrettanto importanti. Nella nostra analisi ciò che conta per la buona performance di una scuola è che i suoi studenti siano rapidi e/o prendano voti alti negli esami di profitto indipendentemente dalla particolare classe di laurea scelta.²² Tuttavia, anche elementi qualitativi, che prendessero in considerazione la probabilità degli studenti provenienti da una certa scuola di scegliere certe classi di laurea o certi Atenei, ed il prestigio occupazionale o i redditi associati a queste classi di laurea o Atenei potrebbero essere utilizzati come indicatori indiretti della performance delle scuole.²³

²² La classe di laurea è stata inclusa solo come variabile di controllo, per rendere conto della potenziale difficoltà di certi corsi in cui la progressione potrebbe essere meno agevole. Tuttavia, ai fini degli indicatori qui elaborati per un buon posizionamento di una scuola nel ranking è indifferente avere studenti rapidi nell’acquisizione dei crediti nella classe di laurea **A** (es. Medicina) o in quella **B** (es. Scienze Politiche).

²³ Per la costruzione di indicatori di performance di questo tipo sarebbe tuttavia necessario avere dati relativi ai laureati. Per un’analisi dell’effetto dell’Ateneo (e di altri fattori) sui redditi per i laureati

Figura 5. Ranking degli effetti fissi di scuola sul voto medio pesato negli esami di profitto (campione scuole con almeno 20 studenti)



Note. La figura mostra un cross-plot degli effetti fissi di scuola ed i relativi intervalli di confidenza ottenuti utilizzando come variabile dipendente i residui dalla regressione mostrata nella quarta colonna della Tabella 25.

8. Grading standard delle scuole e valore informativo del voto di diploma e dei test all'ingresso

Lo scopo di questa sezione è quello di valutare il valore informativo del voto di diploma degli studenti come predittore delle loro carriere accademiche. Quello che faremo è valutare se, ed in quale misura, il voto di diploma una volta controllato per alcune caratteristiche degli individui sia positivamente correlato con indicatori di performance degli studenti universitari. In seconda battuta, valuteremo anche se il risultato relativo nei test all'ingresso (rispetto all'Ateneo e classe di laurea di iscrizione) e la posizione relativa in termini di voto di diploma di uno studente rispetto ai voti nella sua scuola superiore di provenienza abbiano pure un qualche valore informativo. Questi indicatori infatti potrebbero essere utili qualora (i) le scuole non utilizzassero gli stessi *grading standard* o, anche in presenza degli stessi standard, (ii) qualora i test all'ingresso

italiani si veda Brunello e Cappellari (2005) e Ballarino e Bratti (2009) per un'analisi dell'effetto del campo di studi sull'occupabilità dei laureati. Checchi e Salini (2010) in un altro rapporto alla Fondazione Cariplo analizzano l'influenza di vari fattori sugli esiti occupazionali degli studenti lombardi.

utilizzati dalle università fornissero una misura migliore delle competenze necessarie per completare con successo uno specifico corso di studi.

I modelli stimati in questa sezione includono tutti i controlli riportati nell'equazione (1). In aggiunta, essi includono a seconda dei casi anche;

- i) il voto di diploma standardizzato in modo da avere media zero e varianza uno nel nostro dataset;
- ii) il punteggio del test all'ingresso sostenuto dallo studente, normalizzato in modo da avere media 0 e varianza 1 (si veda la sezione 4 per la descrizione dei test);
- iii) la percentuale degli studenti provenienti dalla stessa scuola con voto di diploma minore o uguale a quello dello studente considerato, che fornisce una sorta di ranking relativo dello studente all'interno della scuola di provenienza.

La Tabella 26 mostra i risultati delle regressioni sul numero di crediti. La colonna (3) mostra come sia il punteggio di diploma che quello nel test di ammissione predicano in maniera statisticamente significativa la performance degli studenti. L'incremento di una deviazione standard nel punteggio di diploma è associato a 5.4 crediti in più, lo stesso incremento nel punteggio del test di ingresso è invece associato a 2.5 crediti in più. Il confronto con le colonne (1) e (2) mostra anche come il punteggio nel test e il voto di diploma siano fortemente correlati. Le stime sono molto robuste alla variazione del campione di stima.

Tabella 26. Stime OLS per il numero di crediti

	almeno 10 studenti per scuola			almeno 20 studenti per scuola		
	(1)	(2)	(3)	(4)	(5)	(6)
voto maturità	6.215*** (0.251)		5.442*** (0.261)	6.038*** (0.275)		5.314*** (0.286)
punteggio test		4.554*** (0.261)	2.485*** (0.263)		4.393*** (0.287)	2.322*** (0.291)
R ²	0.355	0.290	0.368	0.365	0.300	0.376
N	4791	4791	4791	3698	3698	3698

*, **, *** significativo al 10%, 5% e 1%, rispettivamente

Note. La tabella riporta i coefficienti (e gli errori standard tra parentesi) di un modello lineare stimato con OLS in cui le variabili dipendente è il numero di crediti. I modelli includono anche controlli per il tipo, provincia e tipo di gestione della scuola, e per genere, età, classe di laurea, Ateneo di iscrizione, tipo di iscrizione e regione di nascita dell'individuo. Per ogni variabile dipendente sono riportate le stime sul campione delle scuole che includono almeno 10 studenti e su quello che include almeno 20 studenti iscritti agli Atenei del nostro campione

La Tabella 27 mostra invece i risultati delle regressioni sul voto medio negli esami universitari pesato per il numero di crediti. Le conclusioni sono le stesse di quelle della Tabella 26: sia il voto di diploma che il punteggio del test di ingresso hanno valore informativo. Nel campione delle scuole con almeno 10 studenti, l'aumento di una deviazione standard nel voto di diploma è associato ad un aumento circa unitario nella variabile dipendente, mentre la stessa variazione nel punteggio del test di ammissione è associata ad un aumento di mezza unità nella variabile di performance. I risultati sono simili nel secondo sotto-campione. Come abbiamo visto, i risultati delle Tabelle 26 e 27

ovvero la significatività statistica del punteggio conseguito nel test di ammissione potrebbero segnalare differenti *grading standard* praticati dalle scuole, o una maggiore specificità della conoscenze misurate dai test di ammissione rispetto a quelle generiche acquisite nella scuola secondaria superiore.

Tabella 27. Stime OLS per il voto medio agli esami pesato per il numero di crediti

	almeno 10 studenti per scuola			almeno 20 studenti per scuola		
	(1)	(2)	(3)	(4)	(5)	(6)
voto maturità	1.221*** (0.036)		1.059*** (0.036)	1.274*** (0.041)		1.114*** (0.041)
punteggio test		0.924*** (0.045)	0.522*** (0.044)		0.945*** (0.053)	0.510*** (0.052)
R ²	0.396	0.305	0.418	0.413	0.311	0.434
N	4791	4791	4791	3698	3698	3698

*, **, *** significativo al 10%, 5% e 1%, rispettivamente

Note. La tabella riporta i coefficienti (e gli errori standard tra parentesi) di un modello lineare stimato con OLS in cui le variabili dipendente è il voto medio negli esami di profitto pesato per il numero di crediti. I modelli includono anche controlli per il tipo, provincia e tipo di gestione della scuola, e per genere, età, classe di laurea, Ateneo di iscrizione, tipo di iscrizione e regione di nascita dell'individuo. Per ogni variabile dipendente sono riportate le stime sul campione delle scuole che includono almeno 10 studenti e su quello che include almeno 20 studenti iscritti agli Atenei del nostro campione.

Al fine di approfondire le ragioni di questa evidenza empirica, la Tabella 28 mostra i risultati delle regressioni che includono come proxy di abilità degli studenti anche il ranking relativo degli stessi nel voto di diploma. Il ranking relativo nel punteggio di diploma fornisce la frazione di studenti provenienti dalla stessa scuola che lo studente si lascia “alla propria sinistra” nella distribuzione dei voti di diploma, è cioè una misura di abilità relativa.²⁴ Questa frazione dipende sia dall'abilità dello studente, che dall'abilità dei suoi pari e dai *grading standard* praticati dalla scuola. Una volta controllato per il voto di diploma individuale, maggiore è questa frazione più rigidi potrebbero essere considerati i *grading standard*. A parità di voto di diploma dello studente, avere una maggiore massa di probabilità alla propria sinistra nella distribuzione dei voti implica che: 1) l'individuo ha un più elevato numero di pari di abilità uguale o inferiore alla sua e/o che 2) i *grading standard* sono più “rigidi” (ovvero che a parità di conoscenze gli insegnanti tendono a dare voti bassi). L'effetto sulla performance che ci si dovrebbe attendere dal ranking relativo seguendo la prima interpretazione è negativo (approssimando la bassa qualità del *peer group*), nel secondo caso potrebbe essere positivo alla luce dell'evidenza che una certa “ridigita” degli insegnanti potrebbe avere un effetto positivo sull'accumulazione di competenze (sarebbe a dire un certo voto “x” vale più in termini di competenze reali quanto maggiore è la frazione di individui che hanno voti minori o uguali ad “x”).²⁵ A nostro parere dovrebbe prevalere il secondo

²⁴ In termini più rigorosi, si tratta della cumulata di frequenze per scuola degli studenti con voto minore o uguale a quello dello specifico studente di interesse (vedi sezione 3).

²⁵ Consideriamo due individui A e B con lo stesso voto di diploma, ad es. 80, A proviene da una scuola in cui la metà degli studenti ha voti minori o uguali ad 80 e B da una scuola in cui il 75% degli studenti ha voti minori o uguali ad 80. Un segno positivo della variabile ranking significa che B ha all'università una

effetto, dato che il primo incide solo indirettamente sul risultato finale. I risultati nella colonna (4) mostrano la significatività di tutte le proxy di abilità dell'individuo (voto di diploma, punteggio nel test e ranking nel voto) eccetto il ranking relativo. La non significatività del ranking relativo potrebbe essere interpretata come evidenza che le scuole praticano gli stessi *grading standard*, mentre risultano confermati i precedenti risultati circa il valore informativo sia del voto di diploma che del punteggio nei test di ammissione come predittori significativi della performance accademica degli studenti universitari.

Tabella 28. Stime OLS per il numero di crediti

	almeno 10 studenti per scuola				almeno 20 studenti per scuola			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
rank studente	7.486*** (0.553)		5.578*** (0.584)	-0.323 (0.618)	7.567*** (0.579)		5.644*** (0.627)	-0.525 (0.687)
punteggio test		4.554*** (0.261)	4.169*** (0.267)	2.489*** (0.264)		4.393*** (0.287)	3.980*** (0.295)	2.329*** (0.293)
voto maturità				5.489*** (0.283)				5.394*** (0.315)
R ²	0.259	0.290	0.301	0.368	0.274	0.300	0.312	0.376
N	4791	4791	4791	4791	3698	3698	3698	3698

*, **, *** significativo al 10%, 5% e 1%, rispettivamente

Note. La tabella riporta i coefficienti (e gli errori standard tra parentesi) di un modello lineare stimato con OLS in cui le variabili dipendente è il numero di crediti. I modelli includono anche controlli per il tipo, provincia e tipo di gestione della scuola, e per genere, classe di laurea, Ateneo di iscrizione, tipo di iscrizione e regione di nascita dell'individuo. Per ogni variabile dipendente sono riportate le stime sul campione delle scuole che includono almeno 10 studenti e su quello che include almeno 20 studenti iscritti agli Atenei del nostro campione.

performance attesa migliore di A. Questo effetto potrebbe dipendere dal fatto che prendere 80 nella scuola A e nella scuola B non significa la stessa cosa in termini di competenze.

Tabella 29. Stime OLS per il voto medio agli esami pesato per il numero di crediti

	almeno 10 studenti per scuola				almeno 20 studenti per scuola			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
rank studente	1.911*** (0.108)		1.537*** (0.103)	0.473*** (0.103)	2.090*** (0.121)		1.694*** (0.114)	0.508*** (0.123)
punteggio test		0.924*** (0.045)	0.818*** (0.044)	0.515*** (0.044)		0.945*** (0.053)	0.821*** (0.051)	0.503*** (0.051)
voto maturità				0.990*** (0.039)				1.037*** (0.046)
R ²	0.276	0.305	0.337	0.421	0.290	0.311	0.348	0.437
N	4791	4791	4791	4791	3698	3698	3698	3698

*, **, *** significativo al 10%, 5% e 1%, rispettivamente

Note. La tabella riporta i coefficienti (e gli errori standard tra parentesi) di un modello lineare stimato con OLS in cui le variabili dipendente è il voto medio negli esami di profitto pesato per il numero di crediti. I modelli includono anche controlli per il tipo, provincia e tipo di gestione della scuola, e per genere, classe di laurea, Ateneo di iscrizione, tipo di iscrizione e regione di nascita dell'individuo. Per ogni variabile dipendente sono riportate le stime sul campione delle scuole che includono almeno 10 studenti e su quello che include almeno 20 studenti iscritti agli Atenei del nostro campione.

Infine la Tabella 29 ripete l'analisi nella tabella precedente ma considerando come variabile dipendente il voto pesato negli esami. La principale differenza rispetto alla Tabella 28 è che il ranking dello studente risulta ora significativo considerando questa seconda variabile di outcome che tiene conto non solo della rapidità ma anche della qualità della carriera accademica. Una possibile interpretazione è che la posizione relativa dello studente nella distribuzione dell'abilità influenzi soprattutto la qualità del percorso formativo dello studente, ma non la sua velocità. Ad esempio, potrebbe accadere che studenti con migliori ranking relativi prediligano i risultati ottenuti nei singoli esami piuttosto che la rapidità delle carriere rispetto, o che la selettività dell'istruzione terziaria si manifesti soprattutto nella capacità degli studenti di conseguire voti elevati nei singoli esami.

Alla luce dei risultati ottenuti nella prima parte di questa sezione, in particolare dell'evidenza di possibili differenze nei *grading standard* tra scuole, nel prosieguo cercheremo di determinare la dimensione, la significatività statistica ed il pattern di queste differenze analizzandolo da un punto di vista disaggregato. Anche in questo caso optiamo per una specificazione a due stadi, iniziando con la stima dell'equazione (1) col metodo dei minimi quadrati ordinari (OLS). Ottenuti i coefficienti dei regressori inclusi è possibile calcolare i residui stimati $\hat{\epsilon}_i$ che rappresentano la parte della performance universitaria degli studenti che non è spiegata dalle variabile esplicative inserite nel modello. I residui stimati verranno utilizzati come variabile dipendente in un secondo stadio, in cui verranno regrediti sui voti di maturità scuola per scuola:

$$\hat{\epsilon}_{ij} = c_0 + c_1 \text{votomat} + c_j \text{votomat}_j + \eta_i \quad (3)$$

dove *votomat* è una variabile che contiene il voto di maturità per ciascuno studente di ciascuna scuola, mentre *votomat_j* contiene il voto di maturità separatamente per ciascuna scuola. Se i *grading standard* fossero omogenei il voto di maturità avrebbe lo

stesso valore in ciascuna scuola (al netto degli effetti delle variabili per cui abbiamo già controllato nel primo stadio), e i coefficienti c_j dovrebbero dunque risultare non significativi. Al contrario, se i nostri coefficienti di interesse c_j risultano diversi da zero, essi rappresentano le differenze medie tra i *grading standard* delle singole scuole e la scuola di riferimento.²⁶

I risultati del primo stadio ricalcano quanto già riportato nella Tabella 25. L'unica differenza da sottolineare è che la dummy “donna” ha ora un segno negativo e significativo, dovuto al fatto che nel campione per cui sono disponibili i test prevalgono gli studenti del Politecnico ed è un risultato già rilevato in letteratura che le donne tendono ad avere performances peggiori nelle materie quantitative. Gli effetti fissi di scuola c_j interagiti con il voto di maturità, così come stimati nella (3) con il metodo OLS sono rappresentati graficamente insieme ai rispettivi intervalli di confidenza nelle figure seguenti, per avere un'idea del *ranking* delle diverse scuole.

Sebbene l'ampiezza degli intervalli di confidenza dovuta alla scarsa numerosità impedisca di raggiungere risultati statisticamente significativi, si nota immediatamente come ci sia evidenza dell'eterogeneità con cui le performances dei vari istituti superiori correlino con il voto di maturità. Ciò emerge chiaramente per tutte le tipologie di istituto superiore per cui abbiamo un campione sufficientemente ampio soprattutto utilizzando come proxy per la performance il voto medio durante il primo anno di università. Le scuole con un coefficiente minore sono quelle i cui studenti mostrano performances peggiori a parità di voto di maturità e di altre variabili osservabili, indicando la tendenza ad inflazionare i voti. Evidenza simile si ottiene utilizzando i test standardizzati, anche se in questo caso il campione è molto meno numeroso e quindi difficilmente confrontabile, mentre nel caso del numero dei crediti l'evidenza è meno marcata.

²⁶ Come noto, un effetto fisso di scuola deve essere omesso. La scuola corrispondente rappresenta allora la “scuola di riferimento” ovvero il benchmark rispetto alla quale le differenze sono espresse. La scuola di riferimento è un liceo scientifico, statale, situato nella provincia di Brescia, per la quale il coefficiente del voto di maturità risulta c_1 . Per le altre scuole l'effetto complessivo va letto come somma tra c_1 e c_j .

Figura 6. Ranking delle interazioni tra effetti fissi di scuola e voto di diploma sui test normalizzati (campione scuole con almeno 10 studenti)

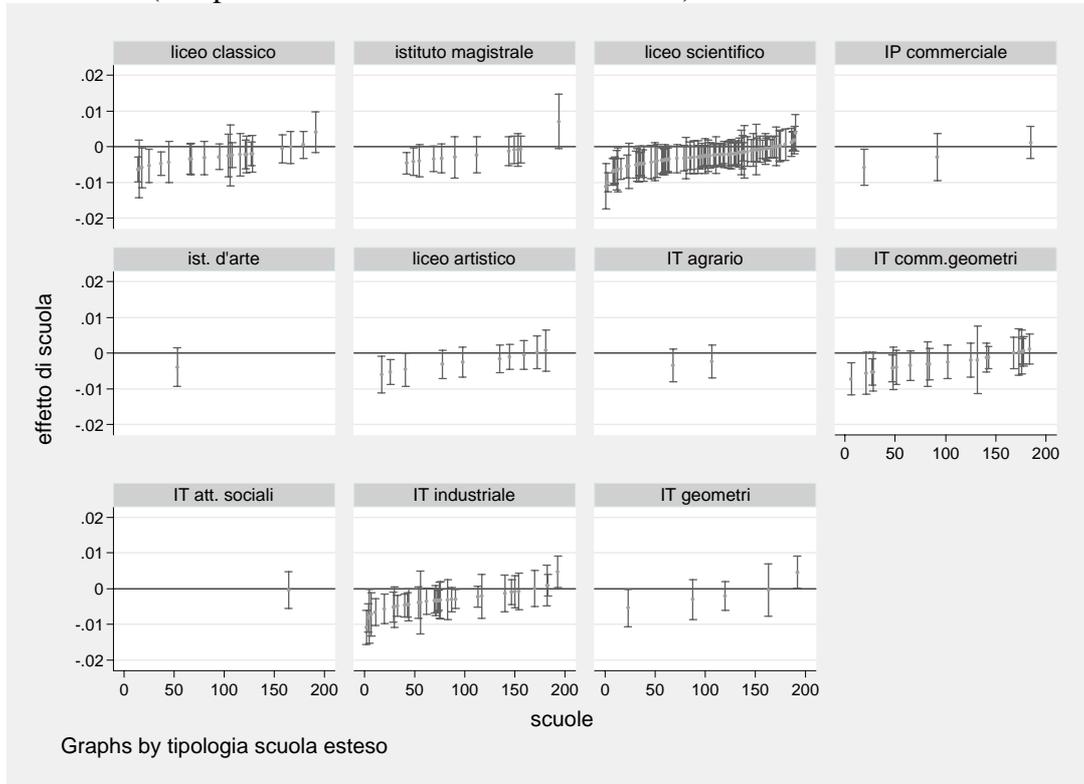


Figura 7. Ranking delle interazioni tra effetti fissi di scuola e voto di diploma sui test normalizzati (campione scuole con almeno 20 studenti)

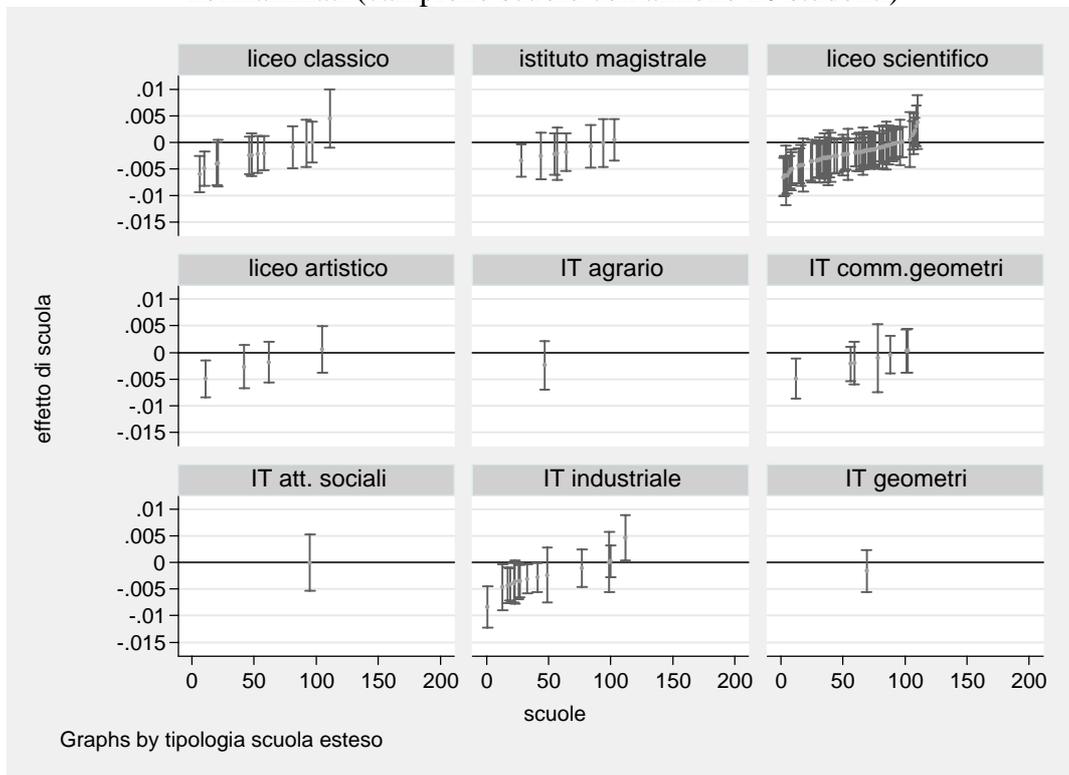


Figura 8. Ranking delle interazioni tra effetti fissi di scuola e voto di diploma sul numero di crediti (campione scuole con almeno 10 studenti)

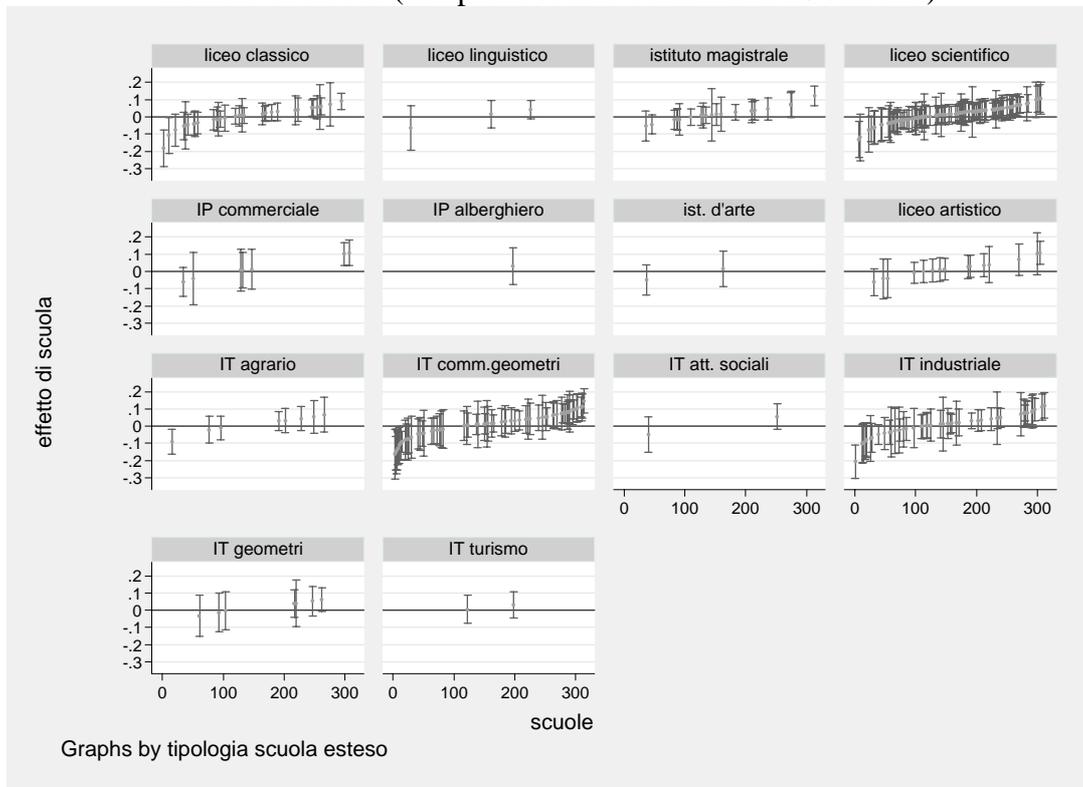


Figura 9. Ranking delle interazioni tra effetti fissi di scuola e voto di diploma sul numero di crediti (campione scuole con almeno 20 studenti)

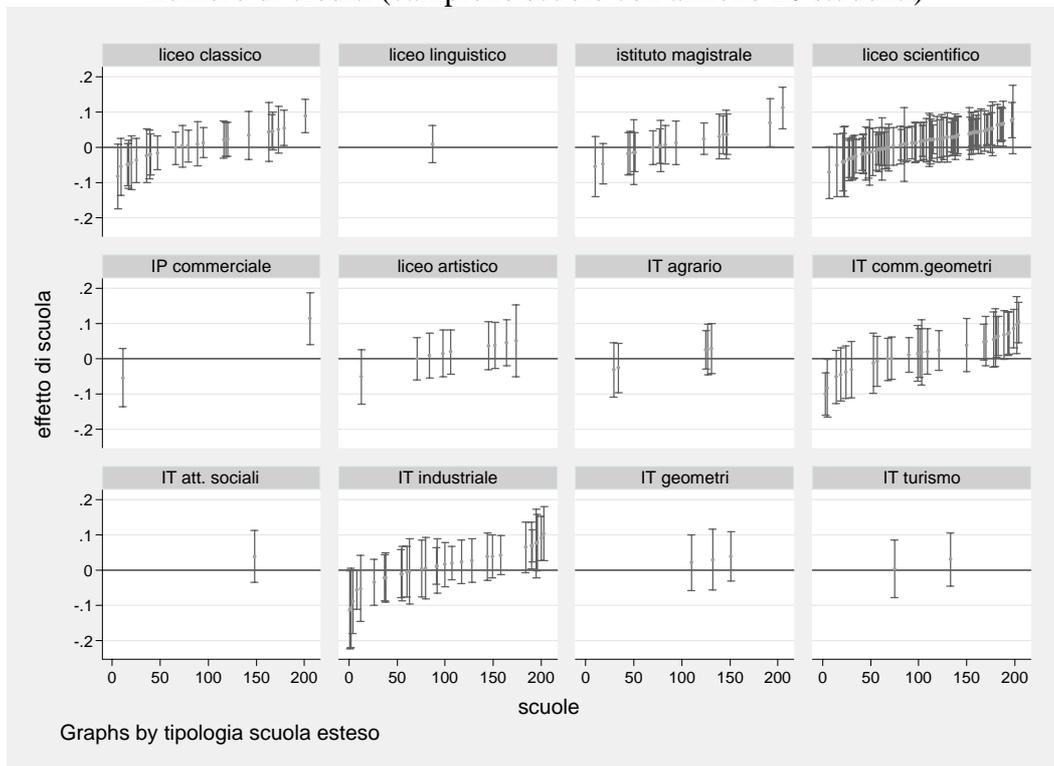


Figura 10. Ranking delle interazioni tra effetti fissi di scuola e voto di diploma sul voto medio pesato (campione scuole con almeno 10 studenti)

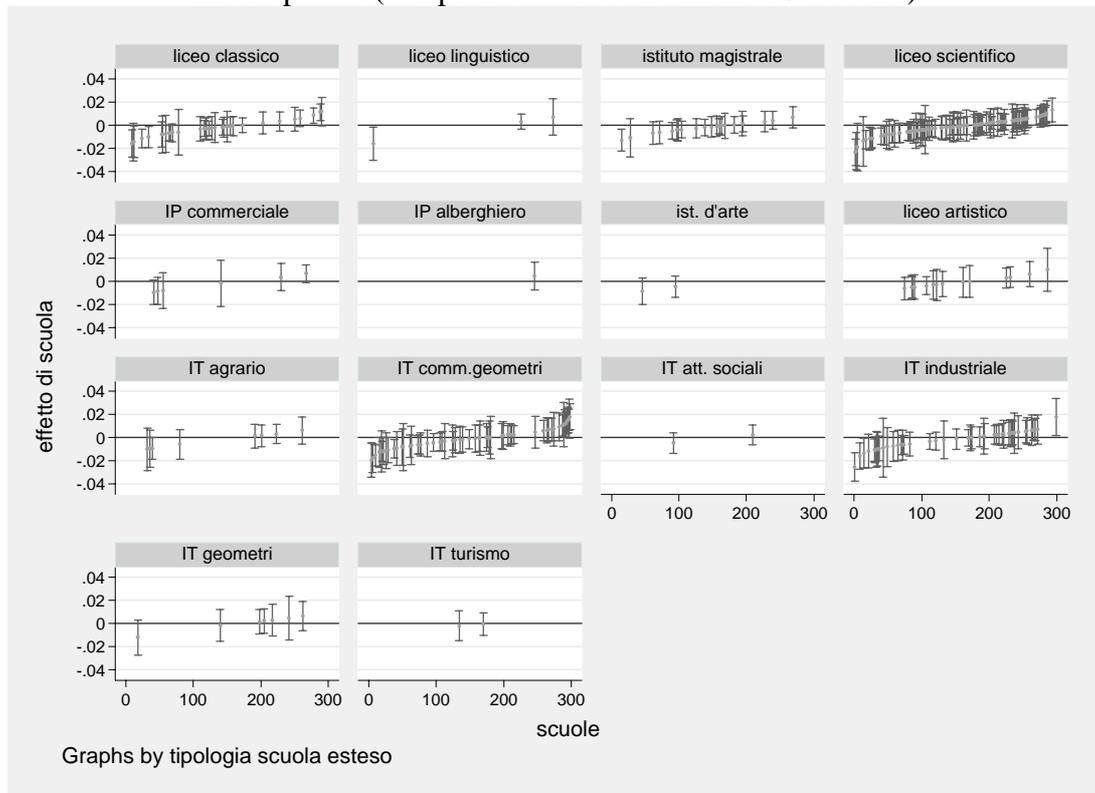
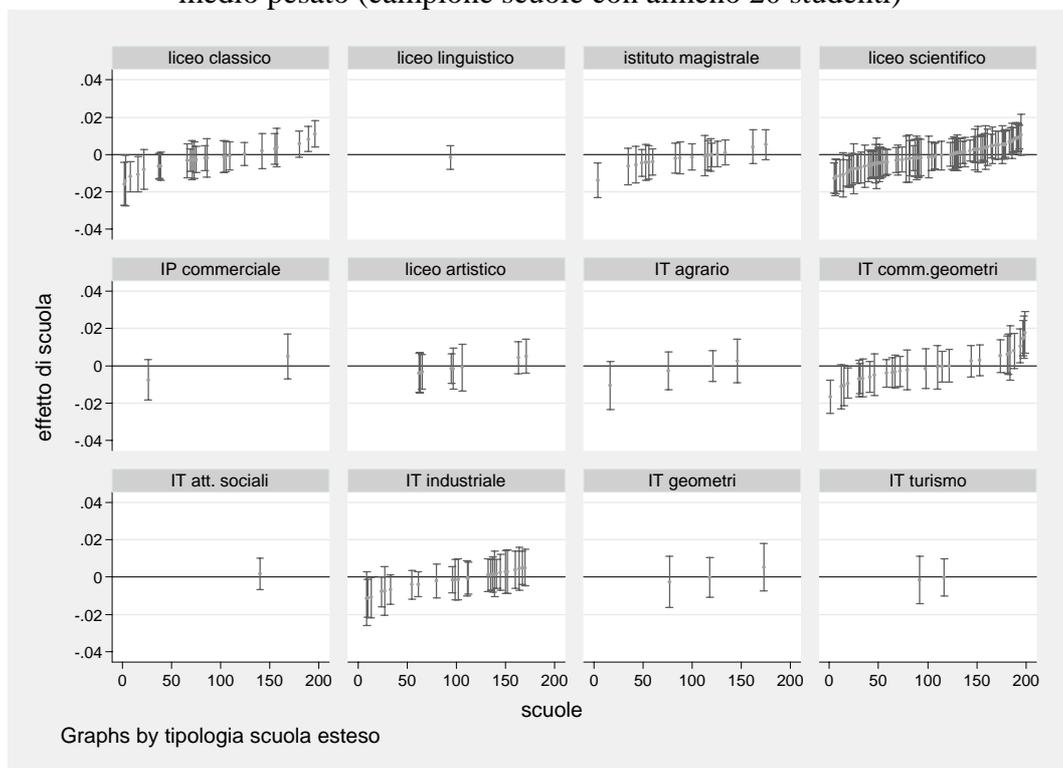


Figura 11. Ranking delle interazioni tra effetti fissi di scuola e voto di diploma sul voto medio pesato (campione scuole con almeno 20 studenti)



9. Conclusioni

In questo rapporto abbiamo analizzato la qualità delle scuole secondarie lombarde attraverso i risultati dei loro studenti: quanti di essi proseguono allo stadio successivo e che risultati conseguono (in termini di voti e di quantità di crediti acquisiti). Abbiamo anche studiato il segnale che questi studenti ricevono dalle loro scuole, in termini di votazione alla maturità, e se se questo sia un segnale veritiero delle loro capacità, una volta che essi debbano poi affrontare un test di ammissione.

Nei primi due paragrafi illustriamo il complicato processo di individuazione della platea delle scuole da valutare, partendo dai dati individuali dei maturati nelle scuole statali nell'anno 2007-08 e incrociandoli con i dati sulle immatricolazioni universitarie in una parte degli atenei lombardi nel successivo anno accademico 2008-09. Poiché la creazione del dataset è esposta a rischi di non rappresentatività, abbiamo discusso ampiamente le fonti, le esclusioni legate agli incroci e la plausibilità dell'inferenza desumibile, in particolare con riferimento alla transizione dalla scuola secondaria all'università. Sotto le condizioni indicate, abbiamo proposto un primo rating basato sulla quota di studenti che transita all'università. Una buona scuola in questo caso è una scuola che fa transitare molti suoi alunni all'università. Tuttavia, poiché abbiamo pochissime informazioni sugli studenti (ed in particolare non sappiamo nulla della condizione socio-culturale della famiglia di provenienza, se non indirettamente attraverso la scuola secondaria frequentata) questa prima graduatoria è costruita principalmente a scopo metodologico, senza alcuna pretesa di definitività. Il ranking assegnato alle scuole è tuttavia misurato con un margine di incertezza, rappresentato dall'intervallo di confidenza che accompagna la stima. Qualora si tenga conto anche di questa dimensione, la maggioranza delle scuole risulta non ordinabile, e la graduatoria delle stesse può essere utilizzata solo per individuare le scuole agli estremi, quelle eccezionali e quelle molto scadenti.

Siamo poi passati ad analizzare la qualità delle scuole sulla base del contributo che esse forniscono alla carriera universitaria dei propri studenti che si iscrivono all'università. In questo caso una buona scuola è identificata sulla base del valore aggiunto che produce in termini di voto medio o di crediti accumulati attraverso gli esami del primo anno. Questa analisi è ovviamente condotta a parità di altre condizioni, da cui si evidenziano risultati noti e meno noti: le donne conseguono risultati migliori rispetto a quelli degli uomini; gli studenti dei licei hanno performance migliori di quelle degli studenti provenienti dagli istituti tecnici e professionali; gli studenti che si scrivono in anticipo all'università hanno risultati migliori rispetto agli iscritti tardivi (suggerendo l'importanza dell'attività di orientamento per gli studenti, in modo da assicurare una precoce percezione dei corsi che li interessano e da garantire un buon match tra le aspirazioni degli studenti e l'offerta formativa).

Al netto di questi effetti le scuole di provincia tendono ad ottenere risultati migliori delle scuole del capoluogo di regione, così come si individuano numerosi istituti tecnici (ad esempio industriali e commerciali) che ottengono performance eccellenti e superiori (in termini di stima puntuale) alla maggior parte degli studenti dei licei scientifici. Anche in questo caso il ranking delle scuole non permette di individuare graduatorie

univoche, data l'ampiezza degli intervalli di confidenza. Abbiamo interpretato questo risultato come esito della sostanziale omogeneità del sistema scolastico lombardo, ottenuta grazie alla omogeneità degli input scolastici (per esempio nella qualità degli insegnanti), data la forte centralizzazione delle politiche scolastiche in Italia.

Quando abbiamo analizzato la coerenza tra i ranking ottenuti secondo tre misure di risultato (studenti che si iscrivono all'università, voto medio conseguito dagli iscritti, numero di crediti conseguito dagli iscritti), troviamo tuttavia dei risultati abbastanza scoraggianti. Tabella 30 mostra la correlazione esistente tra queste tre misure: da essa si evince che una valutazione della scuola sulla base dei tassi di transizione è sostanzialmente scollegata a una valutazione basata sui risultati universitari (vedi anche figure 12 e 13, che presentano graficamente i valori stimati scuola per scuola). Maggiore correlazione vi è invece tra le graduatorie costruite sulla base del voto medio e della velocità di percorso (vedi figura 14). Tuttavia è anche possibile che questa correlazione dipenda dalla qualità degli studenti che frequentano le scuole: poiché anche in questo caso le informazioni sul background familiare sono ridotte o inesistenti, non possiamo escludere che gli studenti più brillanti si concentrino in alcune scuole. Se uno studente brillante è più veloce nel dare gli esami e nel contempo ottiene anche voti migliori, ecco che le scuole che riescono ad attrarre gli studenti migliori sono anche quelle che compaiono in alto in entrambe le graduatorie.

Tabella 30. Indice di correlazione di rango degli effetti fissi di scuola
(299 scuole con più di 10 studenti iscritti)

	iscrizione	voto	crediti
effetto sulla probabilità iscrizione	1.0000		
effetto sul voto	0.0491	1.0000	
effetto sui crediti	0.0851	0.6239	1.0000

Per questo motivo nell'analisi della qualità delle scuole (attraverso una sorta di analisi del valore aggiunto prodotto sulla carriera successiva) è fondamentale poter controllare per l'abilità degli studenti iscritti. Per coloro che si iscrivono all'università noi disponiamo dell'informazione relativa al voto di maturità, mentre per tutti gli altri conosciamo solo la distribuzione dei voti di maturità per intervalli. Per questo motivo abbiamo effettuato un approfondimento sulla capacità segnaletica dei voti di maturità, come indicatori delle reali capacità degli studenti (inclusive quindi sia delle abilità naturali sia delle capacità scolastiche acquisite). Da questa analisi emerge che i punteggi nei test di ingresso ed il voto di maturità hanno entrambi un contenuto informativo indipendente, in quanto risultano entrambi predittori significativi del percorso universitario degli studenti. Anche introducendo una ulteriore misura di posizionamento relativo dello studente all'interno della scuola di provenienza, quest'ultima variabile risulta anche essa un predittore statisticamente significativo della performance, solo nel caso in questa venga misurata in termini di voti medi pesati per il numero di crediti ottenuti. Lo stesso risultato non si ottiene però utilizzando come indicatore di performance il numero di crediti non pesato. Una possibile interpretazione di questi risultati è che le scuole superiori pratichino *grading standard* non omogenei, cosa su cui abbiamo riscontrato evidenza a supporto anche dal fatto che in molti casi varia tra scuola e scuola l'importanza giocata dal voto di diploma nello spiegare i successivi risultati degli studenti.

Figura 12. Incrocio tra effetti fissi di scuola relativi alla probabilità di iscrizione universitaria e numero di crediti acquisiti nel primo anno di immatricolazione (campione scuole con almeno 10 studenti)

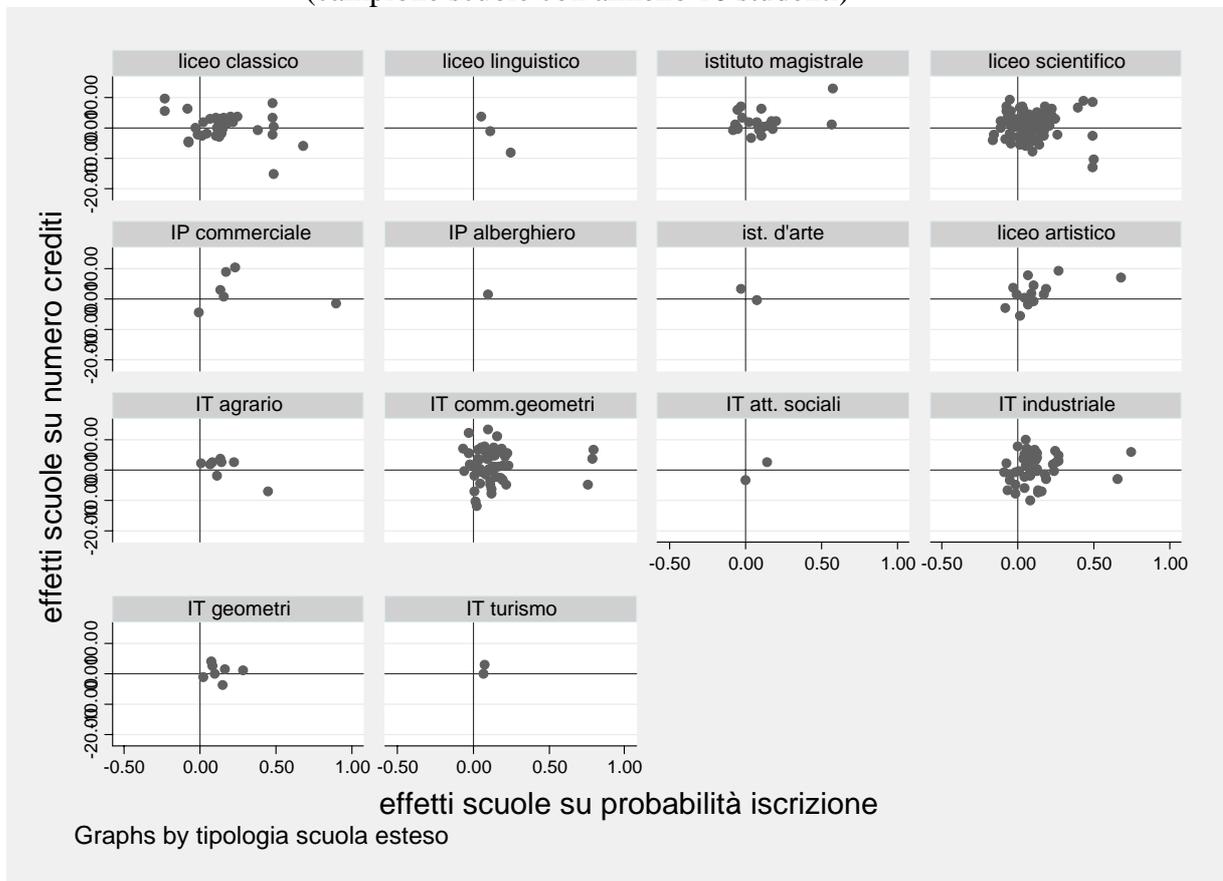


Figura 13. Incrocio tra effetti fissi di scuola relativi alla probabilità di iscrizione universitaria e voto medio conseguito nel primo anno di immatricolazione (campione scuole con almeno 10 studenti)

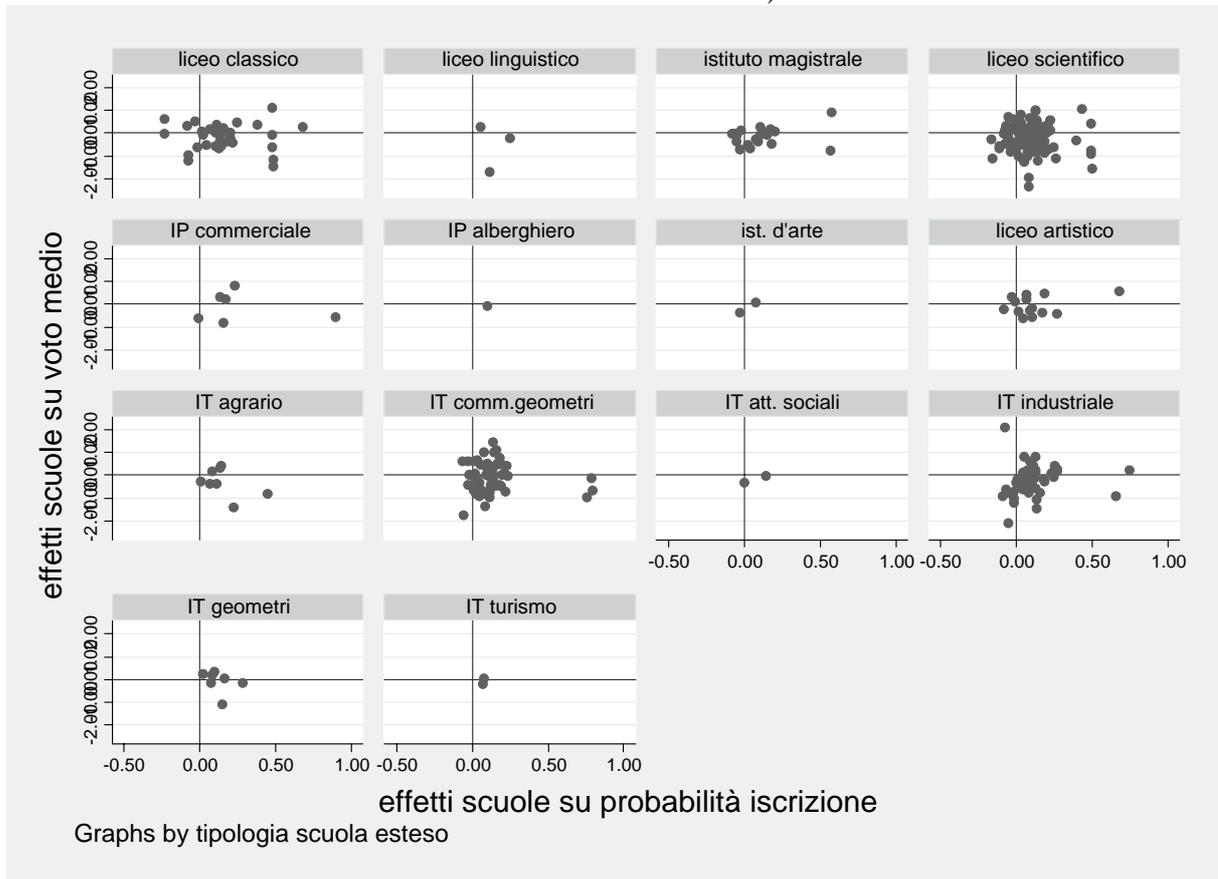
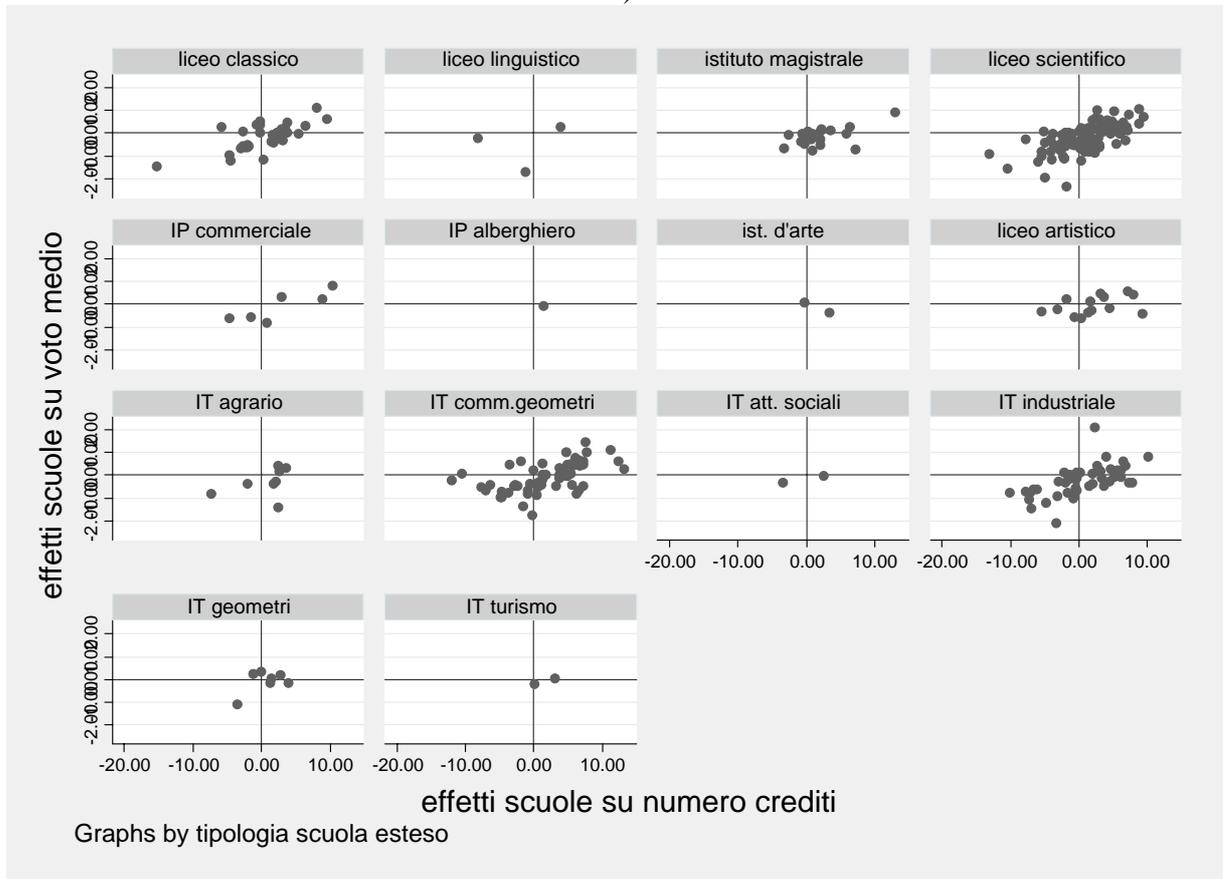


Figura 14. Incrocio tra effetti fissi di scuola relativi al voto medio e al numero dei crediti conseguiti nel primo anno di immatricolazione (campione scuole con almeno 10 studenti)



Appendice A. Descrizione delle variabili

Le variabili categoriche omesse sono indicate in corsivo.

DONNA. Variabile dummy pari ad uno se lo studente è una donna e zero altrimenti (*uomo*).

ETÀ. Età dello studente espressa in anni.

TIPO DI SCUOLA. Indirizzo di scuola secondaria: *liceo scientifico*, liceo classico, liceo linguistico, istituto magistrale, IP commerciale, IP alberghiero, Ist. d'arte, liceo artistico, IT agrario, IT commerciale e per geometri, IT attività sociali, IT industriale, IT per geometri, IT per il turismo.

PROVINCIA. provincia lombarda in cui la scuola secondaria è localizzata: Bergamo (BG), *Brescia (BS)*, Como (CO), Cremona (CR), Lecco (LC), Lodi (LO), Milano (MI), Monza (MN), Pavia (PV), Sondrio (SO), Varese (VA).

TIPO DI ISCRIZIONE. È il tipo di iscrizione dello studente all'università: *precoce (Luglio-Agosto)*, standard (Settembre), tardiva (Ottobre-Novembre).

REGIONE DI NASCITA. È la regione di nascita dello studente (*Lombardia*).

CORSO DI LAUREA. È il corso di laurea al quale lo studente è iscritto.

ATENEIO. È l'ateneo al quale lo studente risulta iscritto: Politecnico di Milano, *Università degli Studi di Brescia*, Università degli Studi dell'Insubria, Università degli Studi di Milano, Università degli Studi di Pavia.

VOTO DI MATURITÀ. È il voto di maturità dello studente normalizzato in modo da avere media 0 e varianza 1 nell'insieme degli studenti iscritti agli Atenei inclusi nel nostro dataset.

PUNTEGGIO NEL TEST DI INGRESSO. È il punteggio nel test di ingresso, normalizzato in modo da avere media zero e varianza 1 rispetto agli studenti che risultano iscritti nello stesso Ateneo e nello stesso corso di laurea.

RANKING DELLO STUDENTE. È la frazione degli studenti provenienti dalla stessa scuola dello studente con voto di diploma minore o uguale a quello dello studente (calcolata sulla distribuzione discreta degli studenti per classe di voto di diploma disponibile per scuola).

Riferimenti bibliografici

- Aina, C., Cappellari, L. e Francesconi, M. (2010), “Is well begun half job done? Educational production functions of university students”, mimeo.
- Bagues, M., Sylos-Labini, M. e Zinovyeva, N. (2008), “Differential grading standards and university funding: Evidence from Italy”, *CESifo Economic Studies*, 54(2): 149-176.
- Ballarino, G. e Bratti, M. (2009), “Field of study and university graduates’ early employment outcomes in Italy during 1995–2004”, *Labour*, 23(3): 421-457.
- Bertola, G., Checchi, D. e Oppedisano, V. (2007), “Private school quality in Italy”, *Giornale degli Economisti e Annali di Economia*, 66(3): 375-400.
- Brunello, G. e Cappellari, L. (2005), “The labour market effects of Alma Mater: Evidence from Italy”, *Economics of Education Review*, 27(5): 564-574.
- De Simone, G., Monastero, B. e Stanchi, A. (2009), “Un esercizio di valutazione esterna: come le Università piemontesi giudicano (indirettamente) le scuole della Regione”, FGA Working Paper n. 18, Fondazione Giovanni Agnelli (Torino).
- Martini A. (2008), “L’accountability nella scuola”, FGA Working Paper n. 8, Fondazione Giovanni Agnelli (Torino).
- Ricci R. (2008), “La misurazione del valore aggiunto nella scuola”, FGA Working Paper n. 9, Fondazione Giovanni Agnelli (Torino).