

Nota Tecnica

I fondi PNRR e il Capitale Umano

PNRR Lab - SDA Bocconi

Sommario

Il Pnrr ha fra i suoi obiettivi quello di generare sviluppo economico. La letteratura economica individua il capitale umano fra i principali driver della crescita economica, sia attraverso l'impatto positivo sulla produttività aggregata sia a livello individuale promuovendo migliori esiti occupazionali e di reddito nel medio-lungo termine. Dalla nostra analisi emerge che la distribuzione degli stanziamenti fra le regioni rispetta il criterio di assegnare maggiori fondi alle aree con maggiori problemi, mentre la distribuzione all'interno delle regioni non è correlata significativamente con covariate di tipo socio-economico associate al tema dell'istruzione.

1 Introduzione e rassegna della letteratura economica

Il capitale umano è uno dei principali driver della crescita economica. Diversi studi hanno dimostrato come lo sviluppo comparato delle nazioni sia da attribuirsi almeno in parte a differenze nel capitale umano, che generano migliori istituzioni e crescita economica (Glaeser et al., 2004). Studi più dettagliati mostrano come il livello di istruzione individuale sia collegato a numerosi esiti socioeconomici: una migliore formazione comporta causalmente minor disoccupazione, maggiore reddito, un minor utilizzo dei sistemi di supporto sociale e anche una migliore salute.¹

2 Dati e Metodologia

I dati utilizzati per questa analisi provengono da fonti di statistica ufficiale o da dataset di tipo amministrativo e sono stati rielaborati dal Pnrr Lab della SDA Bocconi. Per i dati relativi agli stanziamenti dei fondi Pnrr, la banca dati di riferimento è OpenCUP. Il 19 maggio 2022 OpenCUP (parte del Dipartimento di Programmazione Economica della Presidenza del Consiglio dei Ministri) ha pubblicato un dataset contenente informazioni riguardanti gli investimenti Pnrr. Il Soggetto titolare del progetto di investimento pubblico cui compete l'attuazione degli interventi emette un CUP (Codice Unico di Progetto). Il Soggetto titolare è – in genere – l'ultimo anello pubblico (può anche essere un soggetto privato) della catena di enti lungo la quale si muovono i finanziamenti pubblici prima di essere spesi sul territorio per la realizzazione del progetto. In sostanza, tutte le Amministrazioni Pubbliche che gestiscono investimenti Pnrr devono generare un codice specifico per il progetto, il quale viene incluso nella banca dati. Tuttavia, il dataset ha anche alcune limitazioni: anzitutto include anche progetti che, nelle intenzioni della PA che ha generato il CUP, dovevano essere finanziati attraverso Pnrr, ma che di fatto non sono risultati assegnatari di fondi. Inoltre, il dataset contiene un'indicazione sul finanziamento pubblico, ma non permette di distinguere il finanziamento Pnrr da quello attraverso altre fonti. Pertanto, assumeremo che il finanziamento pubblico equivalga al finanziamento Pnrr nel resto di questa analisi.² Il dataset permette, altresì, la localizzazione degli investimenti, poiché in fase di generazione del CUP, viene chiesto di indicare le aree geografiche su cui il progetto ha un impatto, *evitando di scegliere un livello territoriale eccessivamente elevato e generico*. In alcuni casi, tuttavia non vi

¹Per approfondire si veda Coelli et al. (2007); Hanushek and Woessmann (2020); Mezzanotte (2022) e i report OCSE Education at a glance.

²La discrepanza fra importo finanziato attraverso Pnrr (così come indicato nel Piano) e importo indicato nel dataset sembra essere sempre ragionevole, ad eccezione della Componente 1 - Investimento 1.3 dove il divario è significativo. I risultati sono robusti all'inclusione/esclusione di questo investimento.

è una corrispondenza univoca fra progetto (e importo finanziato) e comune su cui il progetto ha un impatto. In questi casi suddividiamo l'importo finanziato fra i diversi comuni utilizzando dei pesi di ripartizione corrispondenti alla media ponderata dello share di popolazione residente al primo gennaio (dell'anno in cui è stato generato il CUP) e all'area del comune al primo gennaio 2022. Tutti questi dati hanno come fonte Istat.³ In termini matematici, consideriamo il progetto con un importo finanziato pari a Y . Questo progetto impatta sull'insieme di comuni $C = c_i$ (ad esempio i comuni della Regione Lombardia). I comuni hanno una superficie al primo gennaio 2022 descritta da a_i , e una popolazione, nell'anno t in cui il CUP è stato generato, pari a $p_{i,t}$. L'importo finanziato per il comune j è ricostruito essere pari a:

$$Y_j = \frac{1}{2} \left(\frac{p_{j,t}}{\sum_{i:c_i \in C} p_{i,t}} + \frac{a_j}{\sum_{i:c_i \in C} a_i} \right) \cdot Y$$

Un'ultima cautela riguarda le date di creazione dei codici CUP: sono infatti presenti nel dataset progetti approvati ben prima dell'approvazione di Next Generation EU e quindi del Pnrr. Al fine di evitare questi problemi, vengono considerati solamente i progetti approvati a partire dal 2020.

Per comprendere se vi sia una relazione tra spesa del Pnrr per istruzione e funzionamento del sistema scolastico a livello locale, sono necessari altri dati socio-economici relativi al capitale umano. Utilizziamo quindi dati provinciali e regionali sulla percentuale di laureati fra i 25 e 39 anni sul totale della popolazione della medesima fascia di età. Questi dati fanno riferimento al 2021 e sono tratti dal BES dei Territori di Istat. L'analisi è robusta nelle conclusioni utilizzando i dati regionali sul titolo di studio della popolazione con età maggiore o uguale a 15 anni ricavati da Istat (Rilevazione Continua della Forza Lavoro) e la percentuale di laureati sul totale della popolazione riferita al 2020. Per analizzare la dispersione scolastica implicita costruiamo un indicatore basato sul livello degli apprendimenti degli studenti e delle studentesse in quinta superiore (grado 13). Il livello è misurato utilizzando gli Open-Data Invalsi riferiti al 2018-19 e aggregati a livello di Sistema Locale del Lavoro. Infine, l'analisi dei dati sull'edilizia scolastica è eseguita partendo dagli Open-Data del Ministero dell'Istruzione su elenco e localizzazione degli edifici scolastici, età degli stessi e accorgimenti adottati per il risparmio energetico.

Le dinamiche di tipo socio-economico raramente vengono racchiuse all'interno dei confini di unità amministrative, come i comuni o le province. L'analisi è quindi svolta al livello dei Sistemi Locali del Lavoro (SLL), 610 aree geografiche definite dall'Istat nel 2011 che coprono tutto il Paese e all'interno delle quali la maggioranza della popolazione risiede, lavora e studia. Cerchiamo di chiarire quest'ultimo punto: sarebbe un errore analizzare i fondi a livello comunale poiché le scuole in alcuni comuni servono anche studenti residenti nei comuni limitrofi. Anche l'utilizzo delle province come unità di analisi non sarebbe ottimale a causa dei flussi di pendolarismo fra una provincia e quelle contigue. Usiamo quindi la suddivisione del territorio italiano in SLL, in grado di catturare il flusso di persone fra i diversi comuni per attività lavorative.

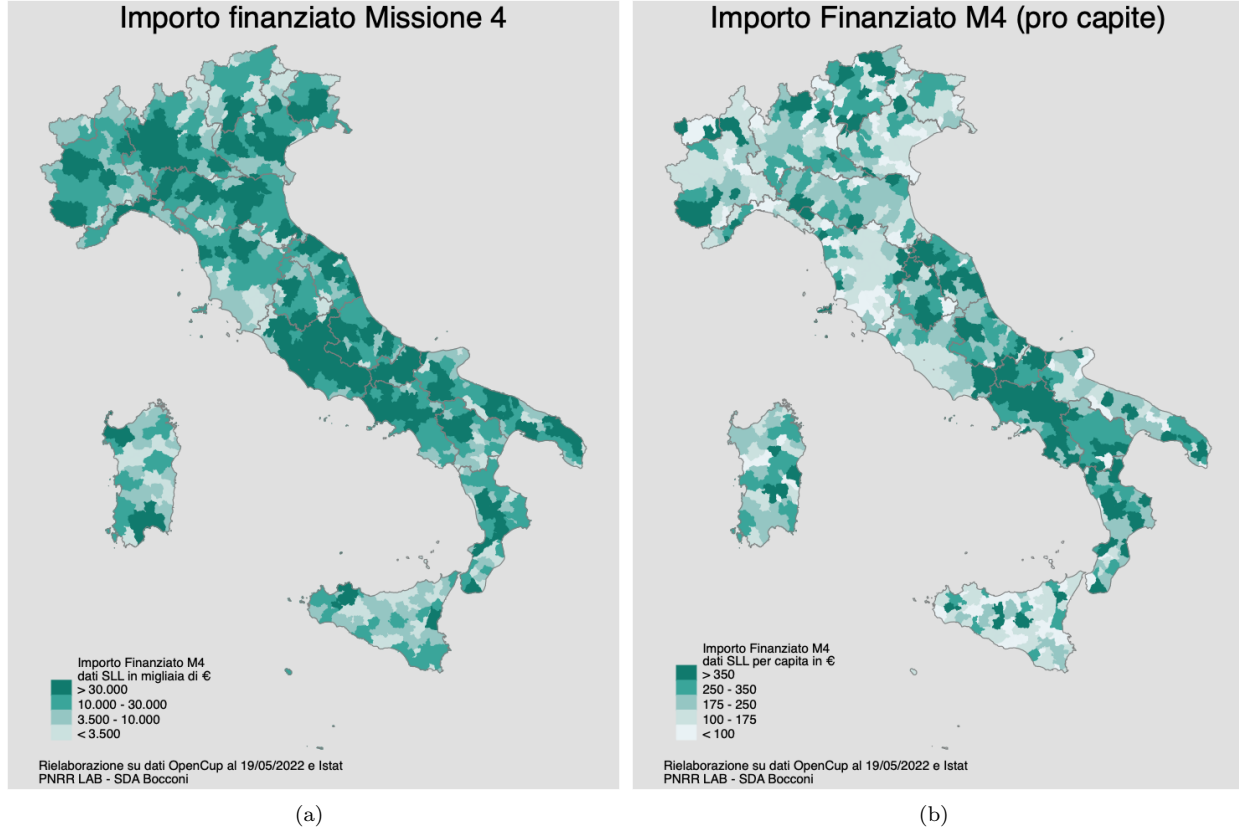
3 Analisi dei dati

Come primo passaggio dell'analisi, filtriamo i dati OpenCUP per missione e consideriamo solo la Missione 4. I finanziamenti inclusi nel dataset sono quindi €12,5 miliardi.⁴ In Figura 1 è possibile osservare la distribuzione dei fondi totali (a) e pro capite (b) fra i diversi Sistemi Locali del Lavoro. Fra gli obiettivi della Missione 4 vi è quello di colmare i divari territoriali esistenti nella sfera dell'istruzione. Questo obiettivo può essere raggiunto agendo su ogni tassello dell'istruzione ed infine sul finanziamento della ricerca, influenzando quindi sia domanda che offerta di capitale umano.

³Non sempre la popolazione residente è disponibile: in questi casi, è stata utilizzata solo la superficie del comune per costruire il peso.

⁴Questo totale è minore della somma di tutte le voci OpenCUP attinenti alla Missione 4 del PNRR. La differenza è dovuta all'esclusione della maggioranza degli interventi dell'investimento 1.3 della Componente 1 (Infrastrutture sportive scolastiche). Gli interventi registrati su OpenCUP per questa voce includono infatti tutte le domande fatte dalle scuole per questi fondi (oltre 3 miliardi di euro). Utilizzando i dati sull'assegnazione dei fondi del Ministero dell'Istruzione, abbiamo mantenuto soltanto le domande ammesse (ad oggi circa 200 milioni di euro, su 300 stanziati dal PNRR), a cui dovrebbe corrispondere un effettivo esborso.

Figura 1: Importo totale finanziato per sistema locale del lavoro



Di per sé è quindi interessante analizzare l'associazione fra tasso di laureati e fondi del Pnrr. L'analisi regionale e provinciale della correlazione fra questi dati e gli importi finanziati non porta mai a risultati statisticamente significativi (Tabella 1). Con questa premessa, analizziamo alcune correlazioni. Anzitutto possiamo verificare che le regioni con la minor percentuale di laureati ricevono leggermente più risorse pro capite della Missione 4 (indice di correlazione -0.32 tra percentuale di laureati e finanziamenti pro capite, non statisticamente significativo). Guardiamo anche alle province. Le province con meno laureati ricevono leggermente più fondi pro capite sia guardando alla variabilità complessiva che a quella intra-regionale (si veda l'Appendice per la metodologia del calcolo della correlazione intra-regionale). Anche in questo caso le correlazioni rilevanti non sono statisticamente significative. Questo risultato è quindi solo parzialmente coerente con l'obiettivo di colmare le differenze territoriali. Naturalmente, l'analisi svolta non evidenzia nessi causali.

Possiamo considerare solo i fondi relativi all'università e alla ricerca e ripetere l'esercizio iniziale di confronto con la percentuale di laureati.⁵ I risultati sono analoghi: le regioni con più laureati ricevono meno fondi pro capite e, a livello intra-regionale, le province con più laureati ricevono leggermente più fondi, ma nessuna di queste correlazioni è stasticamente significativa (Tabella 2). Un fattore che potrebbe aggiungere ulteriore rumore a questo dato è la difficoltà nel localizzare geograficamente l'impatto di un progetto di ricerca.⁶

⁵Consideriamo Componente 1 Investimenti 1.5, 1.6, 1.7, 3.4 e 4.1 e la Componente 2 nella sua interezza.

⁶Dalle FAQ del sito OpenCUP: *Fanno eccezione i progetti di ricerca per i quali va invece indicata l'area più estesa potenzialmente interessata dai risultati della ricerca; nel caso di progetti di interesse generale (es. PRIN), si sceglierà la localizzazione "ITALIA".*

Ci concentriamo ora sulla prima componente della Missione 4, ossia il potenziamento dell'Istruzione dagli asili all'università. Guardiamo agli investimenti diretti strettamente alla scuola.⁷ Nel dataset OpenCUP si tratta di circa €6,5 miliardi. Risulta interessante analizzare se questa misura è associata alla dispersione scolastica implicita. Non avendo a disposizione dati recenti e granulari su questa variabile, abbiamo costruito una proxy partendo dai dati Invalsi. L'indice di dispersione scolastica implicita è costruito come la percentuale minima di studenti ai livelli di competenza (minimi) 1 e 2 in italiano o in matematica o al livello 1 di inglese nelle prove Invalsi in quinta superiore. La dispersione scolastica implicita, costruita con i microdati Invalsi, è la percentuale di studenti che si trova contemporaneamente ai suddetti livelli minimi per le tre materie. A livello regionale, la dispersione scolastica implicita e i fondi pro capite destinati all'istruzione mostrano una leggera correlazione positiva (+0,33), non statisticamente significativa (Tabella 3 e Figura 2). A livello di sistema locale del lavoro, la correlazione fra dispersione scolastica implicita e importi finanziati pro capite è +0,08, non statisticamente significativa. Quindi, gli SLL con peggiori risultati educativi non hanno ricevuto più fondi in modo rilevante. Anche a livello intra-regionale, gli SLL con maggior dispersione hanno ricevuto solo marginalmente più fondi (correlazione +0,003, non statisticamente significativa).

Infine, ci concentriamo sui finanziamenti per l'edilizia scolastica (Componente 1, codici 1.3, 3.2 e 3.3). I territori con edifici scolastici a minor efficienza energetica hanno ricevuto più fondi? Abbiamo costruito un indice sull'efficienza energetica delle scuole partendo da sei indicatori binari disponibili negli Open-Data del Ministero dell'Istruzione. Questo indice è la prima componente principale (quindi, un indice sintetico) delle seguenti variabili: (i) presenza di doppi vetri; (ii) isolamento delle coperture; (iii) isolamento delle pareti; (iv) zonizzazione dell'impianto termico; (v) impianti solari termici; (vi) impianti fotovoltaici. Il resto dell'analisi è svolta come per la dispersione scolastica. Come nei casi precedenti, non ci sono correlazioni significative a livello di regioni, SLL e a livello intraregionale tra efficienza energetica e fondi per l'edilizia scolastica, anche controllando per la latitudine dei territori (Tabella 4). Di conseguenza, anche questi fondi non hanno raggiunto le regioni e i sistemi locali del lavoro con le scuole meno efficienti dal punto di vista energetico.

Sempre attraverso gli Open-Data del Ministero dell'Istruzione possiamo guardare all'età degli edifici scolastici, che tuttavia non considera eventuali ristrutturazioni o interventi edilizi, anche significativi.⁸ In ogni caso, con un approccio analogo a quello precedente, notiamo che le regioni con scuole più datate hanno ricevuto un'allocatione pro capite per l'edilizia scolastica solo leggermente più generosa (correlazione regionale pari a +0,12, non statisticamente significativa), e all'interno delle regioni vi è un'associazione minima fra finanziamenti pro capite ed età degli edifici (correlazione *within regions* pari a +0,08, non statisticamente significativa) (Tabella 5). In altre parole, anche in questo caso, le aree più bisognose non ha ricevuto significativamente più fondi. Riteniamo che un approfondimento del nostro esercizio di analisi possa essere utile anche oltre il Pnrr, visto che la fondazione Agnelli nel 2019 ha stimato in 200 miliardi di euro (il Pnrr ne stanziava circa 6) le risorse necessarie per rinnovare l'intero patrimonio di edilizia scolastica.⁹

4 Conclusione

In base a quest'analisi, le risorse del Piano Nazionale di Ripresa e Resilienza destinate all'Istruzione sembrano distribuite geograficamente a pioggia, non mostrando correlazioni significative con le variabili socio-economiche su cui dovrebbero operare (dispersione scolastica, laureati, stato degli edifici scolastici).

⁷Consideriamo la Componente 1 Investimenti da 1.1 a 1.4, 2.1 e da 3.1 a 3.3.

⁸Questi dati tuttavia sono molto grezzi: il 25% degli edifici scolastici, circa, non ha dati sull'età. In altri casi è disponibili un intervallo di anni, viene imputato l'anno mediano.

⁹<https://www.fondazioneagnelli.it/wp-content/uploads/2019/11/presentazione-Rapporto-edilizia-scolastica.pdf>

Riferimenti bibliografici

- Coelli, M. B., D. A. Green, and W. P. Warburton (2007). Breaking the cycle? the effect of education on welfare receipt among children of welfare recipients. *Journal of Public Economics* 91(7), 1369–1398.
- Glaeser, E. L., R. La Porta, F. Lopez-de Silanes, and A. Shleifer (2004, Sep). Do institutions cause growth? *Journal of Economic Growth* 9(3), 271–303.
- Hanushek, E. A. and L. Woessmann (2020). The economic impacts of learning losses. (225).
- Mezzanotte, C. (2022). The social and economic rationale of inclusive education. (263).

A Tabelle e Figure

Nelle tabelle che seguono, sono mostrati gli indici di correlazione esposti nel testo. Per ottenere le correlazioni intra-regionali e al netto di eventuali controlli sono stati svolti due passaggi. In primis, sono stati stimati i seguenti modelli:

$$X_{i,r} = \alpha_r^X + Z_{i,r} + \varepsilon_{i,r}^X \quad (1)$$

$$Y_{i,r} = \alpha_r^Y + Z_{i,r} + \varepsilon_{i,r}^Y \quad (2)$$

Dove X e Y sono le due variabili d'interesse (ad esempio, dispersione scolastica implicita e fondi pro capite destinati all'istruzione, Tabella 3), i indicizza l'unità geografica più granulare (provincia o SLL), r indicizza le regioni, α_r sono gli effetti fissi (FE) regionali, e Z gli eventuali controlli oltre ai FE (ad esempio latitudine, Tabella 4). In secondo luogo, è stata calcolata la correlazione tra i residui stimati, $\hat{\varepsilon}_{i,r}^X$ e $\hat{\varepsilon}_{i,r}^Y$. Quest'ultima quantità rappresenta la correlazione tra X e Y al netto della variabilità regionale, cioè la correlazione intra-regionale tra le due variabili.

Tabella 1: Correlazioni fra percentuale di laureati e importi finanziati dalla M4

Livello territoriale	Controlli	Fondi pro capite	Fondi in valore assoluto
Regioni	-	-0.3231 (0.1647)	+0.0431 (0.8567)
Province	-	-0.0678 (0.4879)	+0.1481 (0.1278)
Province	FE Regionali	-0.0151 (0.1102)	+0.3169*** (0.1151)

Note: Errori standard tra parentesi. ***: correlazione significativa al 99%. **: correlazione significativa al 95%. *: correlazione significativa al 90%. La percentuale di laureati si riferisce alla fascia di popolazione 25-39 anni. Gli importi finanziati sono i valori pro capite e totali a livello territoriale dell'intera M4 del PNRR. Le correlazioni con i controlli (FE Regionali) corrispondono alla correlazione tra i residui $\hat{\varepsilon}^X$ e $\hat{\varepsilon}^Y$ delle equazioni (1) e (2), stimate con OLS. Gli errori standard sono ottenuti tramite bootstrapping. Dati: OpenCup e Istat.

Tabella 2: Correlazioni fra percentuale di laureati e importi finanziati per università e ricerca dalla M4

Livello territoriale	Controlli	Fondi pro capite	Fondi in valore assoluto
Regioni	-	-0.1484 (0.5325)	+0.0198 (0.9339)
Province	-	-0.1195 (0.2539)	+0.1630 (0.1186)
Province	FE Regionali	+0.0580 (0.1167)	+0.2812** (0.1377)

Note: Errori standard tra parentesi. ***: correlazione significativa al 99%. **: correlazione significativa al 95%. *: correlazione significativa al 90%. La percentuale di laureati si riferisce alla fascia di popolazione 25-39 anni. Gli importi finanziati sono i valori pro capite e totali a livello territoriale riferiti a università e ricerca della M4 del PNRR. Le correlazioni con i controlli (FE Regionali) corrispondono alla correlazione tra i residui $\hat{\varepsilon}^X$ e $\hat{\varepsilon}^Y$ delle equazioni (1) e (2), stimate con OLS. Gli errori standard sono ottenuti tramite bootstrapping. Dati: OpenCup e Istat.

Tabella 3: Correlazioni fra dispersione scolastica e importi finanziati per la scuola dalla M4

Livello territoriale	Controlli	Fondi pro capite	Fondi in valore assoluto
Regioni	-	+0.3293 (0.1563)	+0.1101 (0.6440)
SLL	-	+0.0780 (0.0652)	-0.0773 (0.0677)
SLL	FE Regionali	+0.0027 (0.0371)	-0.0758*** (0.0158)

Note: Errori standard tra parentesi. ***: correlazione significativa al 99%. **: correlazione significativa al 95%. *: correlazione significativa al 90%. La dispersione scolastica è definita nel testo. Gli importi finanziati sono i valori pro capite e totali a livello territoriale riferiti alla scuola della M4 del PNRR. Le correlazioni con i controlli (FE Regionali) corrispondono alla correlazione tra i residui $\hat{\varepsilon}^X$ e $\hat{\varepsilon}^Y$ delle equazioni (1) e (2), stimate con OLS. Gli errori standard sono ottenuti tramite bootstrapping. Dati: OpenCup e Istat.

Tabella 4: Correlazioni fra efficienza energetica e importi finanziati per l'edilizia scolastica dalla M4

Livello territoriale	Controlli	Fondi pro capite	Fondi in valore assoluto
Regioni	-	-0.0295 (0.9017)	-0.3709 (0.1074)
SLL	-	-0.0464 (0.2620)	-0.0312 (0.4506)
SLL	FE Regionali	+0.0055 (0.0500)	-0.0305 (0.0208)
SLL	FE Regionali + latitudine	+0.0034 (0.0494)	-0.0312 (0.0206)

Note: Errori standard tra parentesi. ***: correlazione significativa al 99%. **: correlazione significativa al 95%. *: correlazione significativa al 90%. L'efficienza energetica è definita nel testo. Gli importi finanziati sono i valori pro capite e totali a livello territoriale riferiti all'edilizia scolastica della M4 del PNRR. La variabile latitudine corrisponde alla latitudine del centroide del sistema locale del lavoro. Le correlazioni con i controlli (FE Regionali e latitudine) corrispondono alla correlazione tra i residui $\hat{\varepsilon}^X$ e $\hat{\varepsilon}^Y$ delle equazioni (1) e (2), stimate con OLS. Gli errori standard sono ottenuti tramite bootstrapping. Dati: OpenCup e Istat.

Tabella 5: Correlazioni fra età degli edifici scolastici e importi finanziati per l'edilizia scolastica dalla M4

Livello territoriale	Controlli	Fondi pro capite	Fondi in valore assoluto
Regioni	-	+0.1223 (0.6073)	+0.1881 (0.4270)
SLL	-	-0.0171 (0.6801)	+0.0314 (0.4480)
SLL	FE Regionali	+0.0797 (0.0495)	+0.0422 (0.0341)

Note: Errori standard tra parentesi. ***: correlazione significativa al 99%. **: correlazione significativa al 95%. *: correlazione significativa al 90%. A livello regionale, l'età degli edifici scolastici è una media pesata sulla popolazione degli SLL. Gli importi finanziati sono i valori pro capite e totali a livello territoriale riferiti all'edilizia scolastica della M4 del PNRR. La variabile latitudine corrisponde alla latitudine del centroide del sistema locale del lavoro. Le correlazioni con i controlli (FE Regionali e latitudine) corrispondono alla correlazione tra i residui $\hat{\varepsilon}^X$ e $\hat{\varepsilon}^Y$ delle equazioni (1) e (2), stimate con OLS. Gli errori standard sono ottenuti tramite bootstrapping. Dati: OpenCup e Istat.

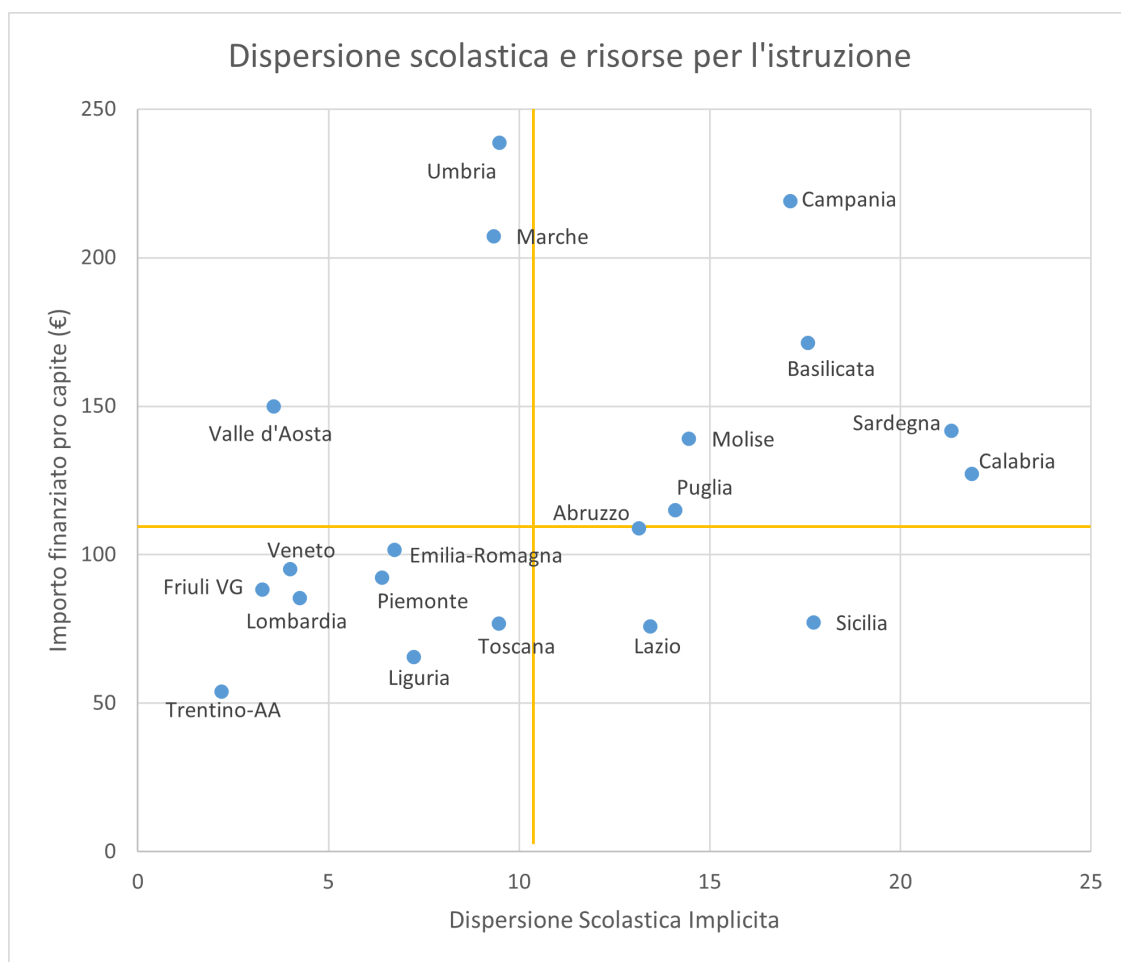


Figura 2: Scatter plot di indice di dispersione scolastica implicita e importo finanziato pro capite a livello regionale. Le linee gialle rappresentano le medie nazionali delle due variabili. Il coefficiente di correlazione a questo livello di aggregazione è $+0,2829$, non statisticamente significativo. Dati: OpenCup e Invalsi. Rielaborazione degli autori.