Age Effects in Primary Education: A Double Disadvantage for Second Generations

Antonio Abatemarco[1], Mariagrazia Cavallo[2], Immacolata Marino[3], Giuseppe Russo[4]

[1] University of Salerno
[2] University of Bristol
[3] University of Naples Federico II, CSEF
[4] Corresponding author. University of Salerno, CSEF, GLO. Dipartimento di Scienze Economiche e Statistiche, Via Giovanni Paolo II, 132 -84084 Fisciano (SA)- Italy

Abstract. A *double disadvantage* occurs when the interaction of two disadvantages generates an additional disadvantage. We show that second-generation immigrant children in the Italian primary school experience a double disadvantage that, relative to the average native, reduces scores in Italian by 17% and in Math by 20%. The double disadvantage stems from the interaction of the immigration background and age effects. On the basis of our results, we propose a simple class composition criterion that may help to counteract this double disadvantage.

Keywords: immigration · second generation · education · double disadvantage
JEL classification: I21, J01, J13, Z13.

1 Introduction

The *double disadvantage hypothesis* suggests that the interaction of two disadvantages may generate an additional disadvantage. We investigate the impact of two disadvantages on school performances of second-generation immigrant children in the Italian primary school. The first source of disadvantage is the immigration background. The second source of disadvantage, which is common to native children, is the vulnerability to age effects.

The immigration background usually puts a penalty on parental socioeconomic and cultural resources (Dustmann et al. (2012), Ochinata and Van Ours (2012)). As for age effects, many studies (Black, Devereux, and Salvanes (2011); Peña (2017)) find that age affects school performances through two main channels: the absolute age effect (henceforth AAE), and the relative age effect (henceforth RAE). The former occurs because older children can benefit from greater knowledge and maturity. The latter comes from peer effects that could, for instance, hinder self-esteem in pupils who feel weaker and less confident than their older peers in the classroom.[5]

In our case, a double disadvantage (henceforth DD) appears if the second-generation status reinforces the AAE or the RAE. We do find that these interactions create a DD both in Italian and Math scores. Though the literature on this topic is still thin, other authors report comparable findings. Lüdemann and Schwerdt (2013) show that the interaction of a migration background with less favorable socioeconomic status puts a DD on second generations in Germany at the transition to secondary school tracks. Dicks and Lancee (2018) find that RAEs and immigrant-specific disadvantages generate a DD in grade retention rates for 15 years old immigrant students in France. Lenard and Peña (2018) point out that part of the educational gap between minority and non-minority students in North Carolina is due to the higher frequency of *redshirting* (i.e. delayed school enrollment) in the majority group.

We contribute to the literature in three ways. First, we provide novel evidence that the interaction of the age effects with the immigration background causes a DD for *second-generation* children. Second, we assess the differential contribution of absolute and relative age to the DD. Third, we observe children at the crucial age of 10, when the basis of human capital accumulation is built; breaking this process may push the child onto different educational tracks and have lifetime consequences.

Our results bring simple policy implications in order to contrast the DD. First of all, they suggest that delaying enrollment (increasing the absolute age) is useless. On the other hand, criteria for class composition should take the RAE into account and possibly increase the relative age of second-generation children with respect to their native classmates.

2 Data

We use standardized test scores in Italian and Math administered by the Italian National Institute for the Evaluation of the Education System (INVALSI). The whole population of students in the 2nd and the 5th grade of the primary school is tested. We observe one cohort

[5] Relatively older children show higher self-esteem and leadership (Dhuey and Lipscomb (2008)).

in the school years 2012-13 (2nd grade) and 2015-16 (5th grade). Another cohort is observed in the school years 2013-14 and 2016-17. We rely on these waves because they contain the exact students' birthdate, which we need to disentangle RAEs and AAEs. The final sample includes 644 521 natives and 49 832 second-generation children.[6]

Given the longitudinal structure of the data, we can follow both cohorts of students from the 2nd grade to the 5th grade (2012-13/2015-16 and 2013-14/2016-17). As suggested by Peña and Duckworth (2018), the combination of information on children's birthdates and the longitudinal dimension of our data provides a way to decompose absolute and relative age. This because the heterogeneity in birthdates provides a variation useful to identify the relative age of pupils within the same class, while the availability of test scores at two different points in time gives a variation suitable to identify the absolute age effect for each pupil.

The data also include detailed information on family characteristics (like father's and mother's employment and education) and home possessions (e.g., the availability of computers, internet connections, quiet rooms, books, and so on) important to take into account other factors that also matter for children's school performance. These family components are summarized in the ESCS index, a synthetic index of economic, social and cultural status. Table 1 shows the summary statistics for the main variables used in the analysis.

3 Empirical Strategy

To empirically investigate the existence of a DD, we first disentagle the AAE and the RAE; then, we analyze their interaction with a second-generation dummy. We estimate the following model for student test scores:

$$Score_{ict\tau} = \beta_0 + \beta_1 AA_{it\tau} + \beta_2 RA_{ic} + \beta_3 Second_i + \beta_4 AA_{it\tau} * Second_i + \beta_5 RA_{ic} * Second_i + \\ + \mathbf{X}_{ict\tau}\rho + \lambda_c * \mu_\tau + \epsilon_{ict\tau} \quad (1)$$

where $Score_{ict\tau}$ is the test score (respectively, in Italian and Math) for student i, in class c, in year t, in cohort τ, $Second$ is a dummy for second-generation children, $\mathbf{X}$ is a vector of controls for socioeconomic status, $\lambda_c * \mu_\tau$ are class-by-cohort fixed effects, and $\epsilon_{ict\tau}$ is an error term capturing time varying idiosyncratic shocks or unobserved class characteristics.

The absolute age $AA_{it\tau}$ is defined as the age on the test day, measured in days and divided by 365.25. It captures the knowledge the child has accumulated and in general the child's maturity level. The relative age RA_{ic} is the difference between the oldest classmate and child's own age and it captures peer effects.[7]

Given this specification, we address two important problems well-known in the literature: 1) the collinearity between AAE and RAE; 2) the endogeneity of Age. An issue of collinearity between AAE and RAE naturally arises, for a student who is older at the moment of the test

[6] We define "second generation" as children born in Italy with both non-Italian parents. We define "natives" the children born in Italy with both Italian parents.
[7] Similar results considering the youngest, the average and the median age classmate are available on request.

is also older with respect to her classmates. As a consequence, the decomposition in *absolute* and *relative* age has rarely been achieved in the literature (Elder and Lubotsky (2009); Black, Devereux, and Salvanes (2011); Peña (2017); Peña and Duckworth (2018)).
Following Peña and Duckworth (2018), we take advantage of the information on children's birthdates and of the longitudinal dimension of our data to decompose absolute and relative age. Specifically, the heterogeneity in birthdates provides the variation to identify the relative age of each pupil within the same class. The availability of test scores at two different points in time provides the variation to identify the absolute age effect. The coefficients β_4 and β_5, plus β_3, measure the DD.

As for the endogeneity of Age, we know that its coefficients could be biased due to the endogeneity induced by grade retention or by the practice of anticipation in children's enrollment. The school year begins in September. The Italian law establishes that children born from May 1 to December 31 must be enrolled when they are 6. On the other hand, children born from January 1 to April 30 can be enrolled when they are either 5 or 6 depending on parents' decision. This possibility creates an important source of endogeneity. Therefore, Age might be correlated with unobservable factors in the error term and using OLS to identify the age effect on educational outcomes may yield biased estimates. Figure 1 shows the students distribution by date of birth in the two waves. As expected, only "regular"students born between May 1 and December 31 are uniformly distributed.[8]

To address this issue, we follow the literature (see, among others, Peña and Duckworth (2018)) and we estimate a Two-Stage Least Squares (2SLS) where we instrument absolute age with "expected" absolute age (${AA^e}_{it}$) and relative age with "expected" relative age (${RA^e}_{ic}$). We compute ${AA^e}_{it}$ assuming that the pupil enrolled as a "regular" student. We describe in Figure 2 the method used to assign the expected birthdate. Recall first that "regulars" are born between May 1 and December 31. For redshirting students, we shift forward the month of birth: those born in January are assigned to May, those born in February to June, and so on. For anticipating students, we shift backwards the month of birth: those born in April are assigned to December of the previous year, those born in March to November of the previous year, and so on. ${AA^e}_{it}$ is computed with respect to the assigned birthdate. ${RA^e}_{ic}$ is the relative age computed from the assigned birthdate.

Redshirting or anticipation could generate a violation of the monotonicity assumption, invalidating the instrumental variable strategy (Barua and Lang (2016)). For robustness, we estimate the model on the subsample of "regulars", confirming our findings.[9]

4 Results

In Tables 2 and 3, we report the OLS and IV estimates of Equation 1 for Italian and Math respectively[10]. Both the OLS and the IV estimates show a DD for second-generation children.

[8] These students can neither anticipate nor delay their enrollment.
[9] Estimates available upon request.
[10] Given the longitudinal structure of the data, we cluster by student to account for the dependency across observations at the individual level.

The first disadvantage is captured by the second-generation dummy, which is negative and significant at 1% level. Being second generation reduces the normalized score by 3.5 points in Italian (5.5% relative to the average), and by 4.9 points in Math (8.4% relative to the average). The second source of disadvantage is given by the age effects. In principle, the AAE benefits more mature children, and the RAE benefits relatively older children in the class, thus a positive coefficient is expected for both. However, we find that, while the RAE has the expected sign, a tiny negative effect arises for the AAE. Black, Devereux, and Salvanes (2011) notice that this can happen as older pupils start school later, but learn more at school than at home. Nonetheless, since the RAE dominates the AAE, the overall age effect is still positive.

The differential age effect between natives and second generations is given by the estimated coefficients for the two interaction variables. These coefficients are 1% significant for both Italian and Math. The AAE is 5 percentage points higher for second generations in Italian, and 2 percentage points lower in Math. The RAE is 0.5 percentage points higher in Italian, and 0.4 percent points higher in Math. The *total* disadvantage experienced by second-generation children is about 11 points in Italian (17% relative to the average native) and 12 points in Math (20% relative to the average native).

Interestingly, the interaction reinforces both the AAE and the RAE only in Italian. In Math, instead, the coefficient of AA*Second is positive and significant at 1% level, contrasting the negative effect of AA. As Black, Devereux, and Salvanes (2011) suggest, staying longer at home seems to hinder Italian proficiency of second generations but not their Math performance.

5 Conclusions

Multiple disadvantages and their interactions threaten the integration of the second generations. In such cases, interventions focusing on only one source of disadvantage can mask persistent problems and hinder overall progress (Taş, Reimão, and Orlando (2014)). This is even more important in the childhood, since mechanisms like the dynamic complementarity and the self-productivity of skills outlined by Cunha and Heckman (2007) tend to reproduce and amplify early educational gaps, making it harder and harder to catch up with the natives.

Using Italian data on second-generation children in the primary school, we support existing evidence on the penalization originating from the immigration background and from age effects. Most importantly, we bring to light the existence of a DD in Italian and Math, showing that these disadvantages interact and reinforce each other. Policy implications are clear-cut: 1) the (slightly) negative effect of the AAE implies that children do not benefit from postponing school enrollment. This is specially relevant for the second generations' achievement in Italian; 2) criteria for class composition should account for the RAE and possibly include second-generation children older than their classmates. This would make second generations less vulnerable to age effects. Both these policies are cost-free.

Declaration of interests: None.

Acknowledgements: We thank Antonio Acconcia, Simon Burgess, Vincenzo Carrieri, Fausto Galli, Tullio Jappelli and Hans H. Sievertsen. This paper uses confidential data provided by the Italian Institute for the Evaluation of the Educational System. The views expressed in this paper are solely those of the authors and do not necessarily reflect those of the INVALSI.
Funding: this study was funded by the Italian Ministry of Research and University (PRIN 2017).

References

Barua, Rashmi and Kevin Lang (2016). "School Entry, Educational Attainment, and Quarter of Birth: A Cautionary Tale of a Local Average Treatment Effect". *Journal of Human Capital* 10.3, 347–376.

Black, Sandra, Paul Devereux, and Kjell Salvanes (2011). "Too Young to Leave the Nest? The Effects of School Starting Age". *The Review of Economics and Statistics* 93.2, 455–467.

Cunha, Flavio and James Heckman (2007). "The Technology of Skill Formation". *American Economic Review* 97.2, 31–47.

Dhuey, Elizabeth and Stephen Lipscomb (2008). "What makes a leader? Relative age and high school leadership". *Economics of Education Review* 27.2, 173–183.

Dicks, Alexander and Bram Lancee (2018). "Double Disadvantage in School? Children of Immigrants and the Relative Age Effect: A Regression Discontinuity Design Based on the Month of Birth". *European Sociological Review* 34.3, 319–333.

Dustmann, Christian et al. (2012). "Educational achievement of second-generation immigrants: an international comparison". *Economic Policy* 27.69, 145–185.

Elder, Todd and Darren Lubotsky (2009). "Kindergarten Entrance Age and Children's Achievement: Impacts of State Policies, Family Background, and Peers". *Journal of Human Resources* 44.3, 641–683.

Lüdemann, Elke and Guido Schwerdt (2013). "Migration Background and Educational Tracking: Is There a Double Disadvantage for Second-Generation Immigrants?" *Journal of Population Economics* 26.2, 455–481.

Lenard, Matthew and Pablo Peña (2018). "Maturity and minorities: the impact of redshirting on achievement gaps". *Education Economics* 26.6, 593–609.

Ochinata, Asako and Jan Van Ours (2012). "Young immigrant children and their educational attainment". *Economics Letters* 116.3, 288–290.

Peña, Pablo (2017). "Creating winners and losers: Date of birth, relative age in school, and outcomes in childhood and adulthood". *Economics of Education Review* 56, 152–176.

Peña, Pablo and Angela Duckworth (2018). "The effects of relative and absolute age in the measurement of grit from 9th to 12th grade". *Economics of Education Review* 66, 183–190.

Taş, Emcet O., Maira Emyand Reimão, and Maria Beatriz Orlando (2014). "Gender, Ethnicity, and Cumulative Disadvantage in Education Outcomes". *World Development* 64, 538 –553.

Figures and Tables

Fig. 1. Histogram

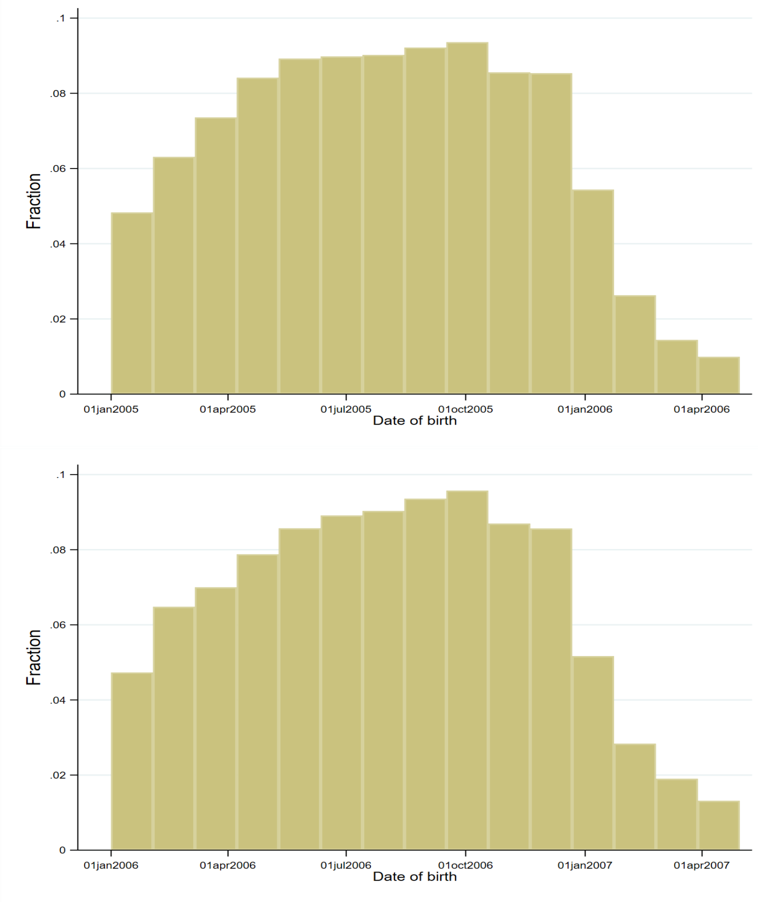

Fig. 2. Actual and assigned calendar month of birth

	Redshirting				Regular								Anticipation			
	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr
Birth month	-3	-2	-1	0	1	2	3	4	5	6	7	8	9	10	11	12
Assigned birth month	1	2	3	4	1	2	3	4	5	6	7	8	5	6	7	8

Table 1. Summary Statistics

	Mean	St.Dev.
Italian	63.4	18.2
Math	58.2	18.6
ESCS	0.086	0.95
Female	0.50	0.50
AA (Italian)	9.26	1.53
AA (Math)	9.26	1.53
RA	-0.46	0.30
Expected AA (Italian)	9.20	1.51
Expected AA (Math)	9.21	1.51
Expected RA	-0.26	0.18
N	1,383,030	
Second	7.2%	

Table 2. Italian

	(OLS)	**(IV)**
AA	-0.454***	-0.455***
	(0.00796)	(0.00796)
RA	5.189***	5.697***
	(0.0581)	(0.0693)
AA*Second	-0.226***	-0.218***
	(0.0290)	(0.0290)
RA*Second	1.356***	1.011***
	(0.233)	(0.279)
Second	-3.326***	-3.547***
	(0.312)	(0.321)
Obs	1,383,030	1,383,030
Controls	✓	✓
Class-by-cohort FE	✓	✓

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$
Standard Errors in Parenthesis Clustered by Student

Table 3. Math

	(OLS)	**(IV)**
AA	-0.339***	-0.339***
	(0.00775)	(0.00775)
RA	5.305***	5.749***
	(0.0592)	(0.0704)
AA*Second	0.0810***	0.0897***
	(0.0273)	(0.0272)
RA*Second	1.205***	0.751***
	(0.233)	(0.280)
Second	-4.599***	-4.883***
	(0.296)	(0.306)
Obs	1,383,030	1,383,030
Controls	✓	✓
Class-by-cohort FE	✓	✓

$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$
Standard Errors in Parenthesis Clustered by Student