Cognitive skills and electoral participation: evidence from Chile

Pablo Argote

December 19, 2017

Abstract

In this paper, I study the relationship between cognitive skills and political participation using a rich administrative data of a given cohort of Chilean youngsters. Given that my design, I am able to disentangle the effects of years of education and cognitive skills, and also, I observe a behavioral measure of political participation. I found a consistent positive effect of verbal and history scores on the probability of enrolling in the electoral registers, which remains after controlling for socioeconomic variables and school-level heterogeneity. On the contrary, the effect of math scores disappears when introducing school fixed-effects. Moreover, I found that poor students are extra-benefited by an additional standard deviation of verbal scores. My results are consistent with Condon (2015), who also found a positive effect of verbal skills. This suggest that students who are able to understand abstract ideas and retain accurate facts are more likely to engage in politics.

1 Introduction

What is the effect of cognitive skills on political participation? Are more able students more likely to vote compared to less able students? Scholars in American Politics have long study the effect of education on political outcomes. Generally speaking, they have found a positive correlation between educational attain-

ment and political participation (Nie et al. 1996). The canonical explanation for this correlation relies on the idea that education confers information and cognitive skills. As a result, individuals would be more able to understand and make sense of their political environment (Verba et al, 1996; Delli Carpini & Keeter 1996; Delli Carpini 2000). Given that most of the variation on educational attainment occurs in post-secondary education, it is reasonable to find a positive effect of higher education on political participation. However, by using data of the Political Socialization Panel Study, Kam & Palmer (2008) found that higher education did not have an effect on education, calling into question the hypothesis that educational attainment causes political participation. Instead, the authors argue that higher education could be a proxy for a person who already is predisposed to participate in politics.

Condon (2015) also challenges the alleged relationship between education and political participation. She argues that the mentioned correlation usually utilize years of education as a proxy for human capital, omitting the most theoretically relevant variable, namely, cognitive skills. In this sense, research in American Politics has not been able to distinguish between attainment -measure through years of education- and achievement -measure through cognitive skills-. In particular, verbal skills could help people to acquire meaningful political information and to develop political ideas. By using the National Education Longitudinal Study of 1988 (NELS), she found a positive effect of verbal skills acquired during adolescence, suggesting that communicate abilities are crucial to explain participation in democratic life. Moreover, she did not found any effect of math scores (Condon 2015).

In Latin America, studies have usually used polling data to predict electoral participation, given the difficulty of observing individual voting. Indeed, Carreras & Castañeda 2014 found that demographic characteristics and insertion in mobilizing networks are strong predictors of self-reported electoral participation in 17 Latin American countries. However, the use of polling data could bias these results, due to the difficulties of measuring effective voter turnout.

In this paper, I investigate the relationship between cognitive skills and political participation using a rich administrative data of a given cohort of Chilean youngsters. With the data at my disposal, I am able to observe verbal, math and history scores, high-school grades, schools and a myriad of student's household characteristics. In this sense, I would be able to control for several covariates, to see whether cognitive skills have a significant effect on political participation.

My paper contributes to the cited literature in two ways. First, I would be able to clearly disentangle the effect of educational attainment and achievement. My sample corresponds to youngsters that already finish their high school. In this sense, everyone on the sample would have the same educational attainment, so any effect of skills would be controlled by years of education. Second, I have a behavioral measure of political participation, namely, enrollment in electoral registers. Probably, this measure is more accurate than self-reported electoral participation.

I found a positive effect of verbal and history scores on the probability of enrolling in the electoral registers, which remains after the incorporation of several covariates. The effect of math scores tend to disappear when accounting for school heterogeneity. Furthermore, I found that poor students are extra-benefited by an additional standard deviation of verbal scores and that performing well both in math and verbal enhance the chance of participating in politics. My results are consistent with Condon (2015), who also found a positive effect of verbal skills. I should point out that my results are not causal by any mean, since there is no chance of randomly assign cognitive skills. The evidence presented here should be interpreted as a solid correlational evidence using a rich individual-level administrative data set.

In the next section, I describe the data and the measures. Then, I explain the 10 statistical models that are estimated. In Section 4, I describe the main results, while in section 5, I present the main conclusions.

2 Data and measures

For this paper, I use two sources of data. First, I will use data from the University Admission Test (PSU) of 2007, a test that is extensively used as an admission tool across the university system. Indeed, admission to college heavily depends on the scores on this test. As a result, every student who plans to attend to college must take this evaluation. In this sense, the PSU data includes Chilean youngsters who finished their high school in that year, and planned to attend to College in the next year. It incorporates the actual scores of each student in math, verbal skills and history, grades in high school, school-level variables and household characteristics.

My second source are the actual electoral records of 2009, which is basically a list of all eligible people to vote for the 2009 presidential election. In that year, voting was mandatory, so people included in that list made a conscious decision of enrolling in the electoral registers in order to be eligible to vote. Moreover, enrollment involved a cost, because it required to show up in an office of the electoral services. Thus, I am fairly confident that being enrolled reveals a clear preference of participating in the 2009 election.

By merging both data sets through the national identifier, I am able to observe the number of youngsters who enrolled in the electoral registers as a fraction of the total of youngsters who took the PSU at the end of 2007. It is worth noting that not all the students included in the sample finished high-school in 2007. Indeed, there are also youngsters that finished high-school earlier, but decided to take the PSU in 2007. For example, a person might have finished high-school in 2006, enrolled in college for a year, and then decide to take the PSU again, in order to be admitted in a better school. To build a fairly homogeneous sample in terms of age, I excluded students that graduate from high-school before 2004.[1] Thus, I have a clean measure of electoral participation in a given cohort. In total, the sample size is 172,711 youngsters.

[1] In all the regression models, I controlled for year of graduation, since older people have had more time to enrolled in the electoral registers

3 Statistical models

The dependent variable of my study is to a dichotomous measure of political participation, which is equal to 1 if the student is enrolled and 0 otherwise. The main predictors of interest are four. First, I have a measure of high-school grades, which was homogenized in a single scale from 300 tom 800 for all students in the cohort. Then, I include three measures of math, verbal and history scores respectively. All the four variables are measured in the same scale from 300 to 800. I standardized these four variables, by subtracting the mean and dividing by the standard deviation. As a result, the variables have mean zero and standard deviation 1, and the marginal effects should be interpreted as the effect of an additional standard deviation increase. It is worth noting that this scores, together with high-school grades, strictly determined admission into college. Thus, they represent a very accurate measure of cognitive skills, because students have all the incentives to exert effort and perform to the best of their abilities. The reduced-form linear probability model (LPM) can be described as follows:

$$Enrolled_i = \beta_0 + \beta_1(Verbal) + \beta_2(Math) + \beta_3(History) + \beta_4(Grades) \tag{1}$$

Where the coefficient β_1 represents the marginal effect of an additional standard deviation of verbal scores on the probability of being enrolled in the electoral registers, controlling for math scores, history scores and high-school grades. I will also estimate a logistic regression model, given that my dependent variable is dichotomous. The model can be described as follow:

$$ln(\frac{F(x)}{1-F(x)}) = \beta_0 + \beta_1(Verbal) + \beta_2(Math) + \beta_3(History) + \beta_4(Grades) \tag{2}$$

Where $F(x)$ is equal to the probability that the dependent variable equals 1, given the predictors. Here, the marginal effect are measured in log-odds, so they cannot be interpreted directly. Moreover, the effect differs

at different units of the independent variables. To keep it simple, I will compute the effect of these predictors as the marginal effect at the mean, setting all the other variables also at their respective averages.

Moreover, equations 3 and 4 show similar models to the ones described in equation 1 ans 2, but with the inclusion of several individual-level control variables: gender, age, year of high-school graduation, whether the student had a paid-job, number of hours dedicated to study per week, socioeconomic status, ownership status of high-school, type of high-school father's education, mother's education, mother's occupational status, father's occupational status and region. For simplicity, in equations 3 and 4, I include all these covariates in the vector *cov*:

$$Enrolled_i = \beta_0 + \beta_1(Verbal)_i + \beta_2(Math)_i + \beta_3(History)_i + \beta_4(Grades)_i + \beta_5(cov)_i \quad (3)$$

$$ln(\frac{F(x)}{1-F(x)}) = \beta_0 + \beta_1(Verbal)_i + \beta_2(Math)_i + \beta_3(History)_i + \beta_4(Grades)_i + \beta_5(cov)_i + \beta_6(School)_s \quad (4)$$

Moreover, given that I have data on the high-school that each student attended, I am able to introduce school-fixed effects. This implies I will be able to control for any heterogeneity across schools, including the quality of instruction, the effect of peers, the quality of the facilities, and so forth. It should be noted that for the models described in equations 5 and 6, the source of the variation is across-students within-schools.

$$Enrolled_i = \beta_0 + \beta_1(Verbal)_i + \beta_2(Math)_i + \beta_3(History)_i + \beta_4(Grades)_i + \beta_5(cov)_i + \beta_6(School)_s \quad (5)$$

$$ln(\frac{F(x)}{1-F(x)}) = \beta_0 + \beta_1(Verbal)_i + \beta_2(Math)_i + \beta_3(History)_i + \beta_4(Grades)_i + \beta_5(cov)_i + \beta_6(School)_s \quad (6)$$

Finally, I included several interactions in both the LPM and the logit models. The first set of interactions correspond to the main predictors among themselves. For example, in the LPM that includes the interaction coefficient of verbal and math scores, the effect of verbal scores should consider both the direct effect and the interaction term. The total effect can be interpreted as the marginal effect of an additional standard deviation of both math and verbal scores. I also includes three-way interactions among the predictor of interests and an interaction between socioeconomic status and the four main independent variables. Moreover, I included interactions between the four measures of cognitive skills and the main measure of household socioeconomic status. For simplicity, in equations 7 and 8, I summarize the vector of interactions as *int*:

$$Enrolled_i = \beta_0 + \beta_1(Verbal)_i + \beta_2(Math)_i + \beta_3(History)_i + \beta_4(Grades)_i + \beta_5(cov)_i + \beta_6(int)_i \quad (7)$$

$$ln(\frac{F(x)}{1-F(x)}) = \beta_0 + \beta_1(Verbal)_i + \beta_2(Math)_i + \beta_3(History)_i + \beta_4(Grades)_i + \beta_5(cov)_i + \beta_6(int)_i \quad (8)$$

Finally, equations 9 and 10 includes both the interaction coefficients and the school fixed effects:

$$Enrolled_i = \beta_0 + \beta_1(Verbal)_i + \beta_2(Math)_i + \beta_3(History)_i + \beta_4(Grades)_i + \beta_5(cov)_i + \beta_6(School)_s + \beta_7(int)_i \quad (9)$$

$$ln(\frac{F(x)}{1-F(x)}) = \beta_0+\beta_1(Verbal)_i+\beta_2(Math)_i+\beta_3(History)_i+\beta_4(Grades)_i+\beta_5(cov)_i+\beta_6(School)_s+\beta_7(int)_i \quad (10)$$

4 Results

Table 1 shows a cross-tabulation with the average difference of the four standardized measures of cognitive skills by enrollment status. We observe that on average, enrolled students in the electoral registers have 0.14 standard deviation higher grades, 0.26 higher verbal scores, 0.21 higher math scores and 0.16 higher in history scores. Thus, we observe a pretty substantive gap in favor of enrolled students in the descriptive measures.

Meanwhile, table 2 shows the results of six of the regression models presented in the previous section. There are several consistent patterns. The coefficient of verbal scores is statistically significant in all the specifications. For example, in equation 3, we observe that an additional standard deviation of verbal scores increases the probability of enrollment in the electoral registers by 0.01 percentage points. With regards to the history scores, we observe a pretty similar result, since the coefficient is significant in every specification, and even higher in magnitude compared to verbal scores. Here, the effect of one standard deviation increase is generally associated with an increase of 0.02 percentage points in the probability of being enrolled. Concerning the math scores, we observe a smaller effect in magnitude, which is not always statistically significant. For instance, in equation 4, we observe that an additional standard deviation increase in math scores correlates with only 0.004 percentage point increase in the probability of enrollment on the electoral registers. Regarding high-school grades, we observe a small effect, which is significant in some of the specifications.

Table 1: Cognitive skills by enrollment status

	Not enrolled	Enrolled	Mean differences
High school grades	-0.03	0.11	0.14***
Verbal scores	-0.07	0.19	0.26***
Math scores	-0.05	0.17	0.21***
History scores	-0.03	0.13	0.16***

Table 2: Regression estimates of the effect of four measures of cognitive skills on electoral participation

	LPM Eq. 1	Logit Eq. 2	LPM Eq. 3	Logit Eq. 4	LPM Eq. 5	Logit Eq. 6
High school grades	0.00	0.00	0.01	0.01	0.01	0.01
	0.001	0.001	0.001***	0.001***	0.001***	0.002***
Verbal scores	0.02	0.02	0.01	0.01	0.01	0.02
	0.001***	0.001***	0.001***	0.001***	0.001***	0.002***
Math scores	0.01	0.01	0.003	0.004	0.000	-0.001
	0.001***	0.001***	0.001***	0.001***	0.001	0.01
History scores	0.02	0.02	0.02	0.02	0.02	0.03
	0.001***	0.001***	0.001***	0.001***	0.001***	0.002***
Covariate-adjusted	No	No	Yes	Yes	Yes	Yes
School fixed effects	No	No	No	No	Yes	Yes

p-value < 0.1,* < 0.05,** < 0.01***. Robust standard errors are displayed below the regression coefficients. For the logit models, marginal effects are estimated at the mean.

It is worth to mention that after the inclusion of school-fixed effects, the coefficients of both verbal and history scores remain statistically significant. Actually, there is an increase of the effect of history scores. Moreover, the effect of math scores completely disappears in the fixed-effects specification.

To analyze the interaction terms, I will first present a table with the main effects plus the interaction coefficients between the cognitive skills variables. In table 3, we observe the main effects of the cognitive skills in the first four rows, and the interaction coefficients in the next eight rows. All the coefficients are covariate-adjusted, while equations 9 and 10 includes school-fixed effects. We observe that both the verbal and the history scores are statistically significant in all these specifications. Moreover, the only interaction term with a positive effect is the interaction between verbal and math scores. For instance, if we consider equation 10, we can conclude that a student who performed well in both verbal and math had 0.016 percentage points higher probability of enrolling in the electoral registers comparing to an student who only performed well in

Table 3: Regression estimates and interaction terms

	LPM Eq. 7	Logit Eq. 8	LPM Eq. 9	Logit Eq. 10
High school grades	0.01	0.01	0.01	0.01
	0.002**	0.001***	0.002***	0.002***
Verbal score	0.01	0.01	0.01	0.03
	0.002***	0.001***	0.002***	0.003***
Math score	0.000	0.003	-0.001	0.006
	0.002	0.002**	0.002	0.003**
History score	0.02	0.02	0.02	0.04
	0.003***	0.001***	0.002***	0.002***
High school * Verbal	0.000	-0.001	0.000	-0.002
	0.001	0.001	0.001	0.002
High school * Math	-0.002	-0.002	-0.002	-0.003
	0.001	0.001	0.001	0.002
High school * History	0.001	0.001	0.001	0.001
	0.001	0.001	0.001	0.002
Verbal * Math	0.008	0.007	0.006	0.010
	0.001***	0.001***	0.001***	0.002***
History * Verbal	0.004	0.001	0.004	0.00
	0.001	0.001	0.0013***	0.00
History * Math	-0.004	-0.004	-0.005	-0.01
	0.001***	0.001***	0.0014***	0.003***
High school * Verbal * History	0.003	0.000	0.001	0.001
	0.010	0.001	0.001	0.002
Verbal * Math * History	0.005	0.001	0.000	-0.002
	0.007	0.001	0.001	0.002
Covariate-adjusted	Yes	Yes	Yes	Yes
School fixed effects	No	No	Yes	Yes
Interaction terms	Yes	Yes	Yes	Yes

p-value < 0.1,* < 0.05,** < 0.01***. Robust standard errors are displayed below the regression coefficients. For the logit models, marginal effects are estimated at the mean.

verbal.[2] Strikingly, we observe that there is a small and negative effect of the interaction coefficient between history and math scores, meaning that performing well in both tests actually decreases the probability of enrolling in the electoral registers compared to only performing well in history.

Table 4 shows the interaction coefficients between socioeconomic status and the four measures of the cognitive skills for equations 7 and 9[3] The measure of socioeconomic status is categorical variable from 1 to 8, where 1 means poor and 8 means rich. In the regressions, the omitted category is the rich. As expected, we observe that poorer students are 0.03 - 0.04 percentage points less likely to enrolled in the electoral registers.

The interaction coefficients in table 4 indicate the extent in which cognitive skills might counteract the effect of socioeconomic status on electoral participation. We observe that the only interaction coefficient with a positive sign is Verbal score * Poor, implying that for poor students, an additional standard-deviation of verbal scores is associated to an extra increase of electoral participation of 0.02 percentage points. Thus, when considering equation 7, for an average poor student, the total effect of verbal scores is $0.01+0.02 = 0.03$ percentage points increase.

5 Discussion and conclusion

Overall, there is consistent positive effect of verbal and history scores on the probability of enrolling in the electoral registers, which remains after controlling for socioeconomic variables and after introducing school-fixed effects. The coefficient of high-school grades was positive and significant only after controlling students and school heterogeneity, while math scores looses significance with these same set of controls. Moreover, there is clearly a positive effect of both performing well in verbal and in math, while there is

[2]To calculate this difference, I considered the coefficients of equation 10, and add verbal, math and the interaction term between verbal and math, and interaction between the two: $0.03 + 0.006 + 0.01 = 0.046$. Then, I subtracted this result to the coefficient of verbal scores. Therefore, the effect was $0.046 - 0.03 = 0.016$

[3]I was not able to run the fully interacted model in the logit specification with the school fixed-effects, due to constraints of computational power. However, I do not expect to find very different results, since the LPM and the logit models have usually show the same findings

Table 4: Interaction terms between socioeconomic status and cognitive skills

	LPM Eq. 7	LPM Eq. 9
Poor	-0.04	-0.03
	0.01***	0.01***
High school grades * Poor	-0.003	-0.01
	0.01	0.01
History score * Poor	-0.002	0.003
	0.00	0.00
Verbal score * Poor	0.02	0.02
	0.01**	0.01**
Math score * Poor	-0.02	-0.003
	0.01**	0.01
Covariate-adjusted	Yes	Yes
School fixed effects	No	Yes
Interaction terms	Yes	Yes

p-value < 0.1,* < 0.05,** < 0.01***. Robust standard errors are displayed below the regression coefficients.

not an interactive effect of both history and verbal skills. Finally, we observe that poor students are extra-benefited by an additional standard deviation increase of verbal skills.

My results are strikingly consistent with Condon (2015), who also found an effect of verbal skills, and not a persistent effect of math scores. These results suggest that students who learn how to communicate better and to understand abstract ideas are more likely to participate in politics, since they might be better able to process political information. Moreover, they could be more able to update their beliefs about the candidates, and discern which of the options fits better with their own ideologies. Certainly, these skills are captured in the verbal scores measure used in this paper.

I should mention the other relevant finding of my paper: the positive effect of history scores, which is even slightly higher than the effect of verbal scores. Here, the mechanism could be slightly different. Indeed, history scores reflects the extent in which a student is versed about the history of a country, which might be a good proxy of the capacity of retaining accurate information. In this sense, students who are more able to retain actual facts are also more able to participate. Thus, I have presented evidence of two mechanisms in play: first, the skill of understanding and processing abstract ideas and 2) the capacity of retaining accurate

facts. Both factors -and not any measure of cognitive skills, such as math scores- significantly explain the probability of participate in politics.

6 Bibliography

Carreras, M, Castañeda, 2014. Who Votes in Latin America? A Test of Three Theoretical Perspectives. Comparative Political Studies 2014, Vol. 47(8) 1079– 1104

Condon, Meghan. 2013 Voice Lessons: Rethinking the Relationship Between Education and Political Participation. Political Behavior (2015) 37:819–843

Delli Carpini, Michael X., and Scott Keeter. 1996. What Americans Know About Politics and Why It Matters. New Haven: Yale University Press.

Delli Carpini, Michael X. 2000. In search of the informed citizen: What Americans know about politics and why it matters, The Communication Review, 4:1, 129-164,

Kam, C. Palmer, C. 2008. Reconsidering the Effects of Education on Political Participation. The Journal of Politics, Vol. 70, No. 3, July 2008, Pp. 612–631

Nie, Norman H., Jane Junn, and Kenneth Stehlik-Barry. 1996. Education and Democratic Citizenship in America. Chicago: University of Chicago.

Verba, Sidney. 1996. The Citizen as Respondent. American Political Science Review 90 (1): 1–7.