

Effects of socioeconomic status on the use of written language: does it extend into Brazilian Higher Education?

Kaizô Iwakami Beltrão ^a
Mônica Cerbella Freire Mandarino ^b
Ricardo Servare Megahós ^c
Mônica Guerra Ferreira Pedrosa ^d

Abstract

Bourdieu and Passeron defended the thesis that the school was the main locus to legitimate and perpetuate class differences. This is reinforced by the multiple proficiency tests used to monitor public policies, which privilege the use of the formal language as part of the instruments and, therefore, penalizes participants with less mastery of the language. We adjusted two hierarchical models with Enade's results on standard Portuguese grades, using as covariates, indicators of students' socioeconomic status and economic independence and average values for these variables for the knowledge areas. The linguistic performance is disaggregated into three aspects: textual, orthographic and vocabulary/morphosyntactic. More affluent socioeconomic groups have greater proficiency in the Enade Portuguese Language component, even when controlling for the knowledge area average socio-economic level and financial autonomy of the students. The socioeconomic effect is not as strong as on lower educational level: university students constitute a rather homogeneous group. This reinforces Bourdieu's thesis that through the social, cultural and economic capital still prevails the domination of wealthier classes over more popular classes, reinforcing the inequality.

Keywords: Social Capital. Economic Capital. Social Inequality. Linguistic Performance. Enade. Public Policy. Hierarchical Models.

* An earlier version of this paper was presented at the Royal Statistical Society Conference – 2019 in Belfast.

^a Fundação Getúlio Vargas, Rio de Janeiro, RJ, Brasil./ Fundação Cesgranrio, Rio de Janeiro, RJ, Brasil.

^b Fundação Cesgranrio, Rio de Janeiro, RJ, Brasil.

^c Fundação Cesgranrio, Rio de Janeiro, RJ, Brasil.

^d Fundação Cesgranrio, Rio de Janeiro, RJ, Brasil.

Received: Sep. 29, 2020

Accepted: Mar. 15, 2021

1 Introduction

Bourdieu and Passeron (1970) defended the thesis that the school was the main locus to legitimate and perpetuate class differences. This is reinforced by the multiple proficiency tests used to monitor public policies, which privilege the use of the formal language as part of the instruments and, therefore, penalizes participants with less mastery of the language. We propose to use a hierarchical model with National Assessment of Student Achievement's (*Exame Nacional de Desempenho de Estudantes – Enade*) results on standard Portuguese grades, using as covariates, indicators of students' socioeconomic status (*SES*) and financial autonomy/independence (*AUT*).

According to *Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira* (Inep), a department of the Ministry of Education that conducts the exam, the Enade assesses undergraduate programs through an "... exam administered to students who are finishing courses in Higher Education institutions throughout Brazil. Programs are grouped in three representative areas and each year one group is assessed, meaning that each program is assessed every three years." Graduating students of all courses take the exam every third year.

These exams, besides a section on specific knowledge related to the professional area, pose questions of general interest in an introductory section. This section is composed of two written short-answer questions plus eight multiple-choices questions. Short-answer questions are graded for content and for use of standard Portuguese. The Portuguese grade takes into consideration three components: orthography, lexical and syntactical (morphosyntactic/vocabulary). Though questions contents in different exams are not strictly comparable, we hypothesized that the use of formal Portuguese in the answers is. Confirming Bourdieu and Passeron (1970) hypothesis, we find out that socioeconomic status do have an impact on language proficiency, but at the University level it is not as strong as measured at lower educational levels (FRITSCH *et al.*, 2019).

We show in this article that the more affluent socioeconomic groups have greater proficiency in the exams, in particular at Enade Portuguese Language component, even when controlling for the average socio-economic level of the knowledge area considered. This reinforces Bourdieu's (1966) thesis that through the social, cultural and economic capital still prevails the domination of wealthier classes over more popular classes, reinforcing the inequality.

2 Literature Review

Bourdieu and Passeron (1970) proposed a general theory of symbolic violence, based on empirical work, on the literate or common use of the university language and culture and on the economic and symbolic effects of examinations and diplomas. Foundations of this theory of symbolic violence are based on a set of 82 hierarchical and logically coordinated propositions. They define symbolic violence as any power which manages to impose meanings and to impose them as legitimate by concealing the power relations that are the foundation of its strength. In this imposition and inculcation of cultural meanings, pedagogical action plays a major role; and among the many instances of pedagogical action, the school holds a prominent place in society.

As part of the theory, the authors included the social conditions for the concealment of this violence. Academic institutions, under the aura of independence and academic neutrality, strongly contribute to the reproduction of the established order. The authors manage to uncover the processes and means by which academic institutions from all level of Education reinforce the structure of the distribution of cultural capital, and somehow expose the contradictions, which affect the educational systems. In this way, school proves to be the most effective mechanism for reproducing social inequalities, more effective because its own action and authority seem to owe nothing to social relations and those who practice and/or experience the symbolic violence of cultural arbitrariness are the first to give it total legitimacy.

In a previous text, Bourdieu (1966) asserted that the school system is understood as a factor of social mobility, consistent with the ideology of the “liberating school”, while evidence shows the opposite, since it is one of the most effective factors of social persistency. In fact, it provides the appearance of a legitimization to social inequalities and sanctions the cultural heritage, the social gift treated as a natural gift. Because the mechanisms of elimination acts throughout the curriculum, it is legitimate to grasp the effect at the highest levels of the school career. Nevertheless, the chances of access to Higher Education are the result of a direct or indirect selection, throughout the schoollife, which weighs with unequal rigor on the subjects of the different social classes.

The mastery of language, in particular of written language, favors the understanding of concepts and procedures at all levels of Education and in all areas. The act of reading/writing constitutes a memory rescue and forms a powerful network of meanings (CÂNDIDO, 2001). For Smole (2001), the difficulty of understanding the text of any subject matter lies in the interpretation of the subjacent meaning.

As the literature states, mastery of formal language, regardless of format, oral or written, is highly correlated with good performance in all subject matters.

In their critical synthesis, Dika and Singh (2002) cite several articles that address the issue of social capital and its direct relationship with educational achievement. It also highlights Bourdieu's disagreeing interpretation of Coleman (1966, 1988, 1992), who views positively differences in social capital and understands that it is the responsibility of families to adopt norms to increase the chances of success in the lives of their children.

In Brazil, several authors have already addressed the issue of intergenerational transmission of inequality (GONCALVES; FRANCA, 2008) through Education. Some consider specifically the mastery of the written language (LAROS *et al.*, 2012).

3 Data base

To investigate the issue, data from three years of Enade were used (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA, 2015). The database involved 77 Higher Education areas, 22.594 courses in public and private institutions, and 1,092,875 graduating students of these courses. Emulating Beltrão and Mandarino (2014) for the 2004-2012 period, but extending the analysis up to 2017, two composite indicators were estimated for each student: *SES* and *AUT*. These indicators were constructed from the students' profile questionnaire using Optimal Scaling and Principal Components. *SES* was based on four variables: family income; type of high school attended (all in public schools, mostly in public schools; half and half public/private; mostly in private schools; and all in private schools); father's Schooling; and mother's Schooling. *AUT* was based on: student's working load (not working; working occasionally; at most 20 hours per week; between 20 and 40 hours per week; 40 hours or more per week); and family responsibilities (not working and totally dependent on family; working but receives family allowance; working and self-sufficient; working and contributing to family budget; and working and family mainstay).

The linguistic performance score of each student (Y_{ij}) was used as a dependent variable (i -th student in the j -th knowledge area) and the independent variables were: the knowledge area of the student's major (j), socioeconomic level (SES_{ij}) and financial autonomy (AUT_{ij}) of the student. Variables SES_{ij} and AUT_{ij} were considered at the student level and at the area level, SES_j and AUT_j , as the average of all graduating students in the area. *SES* variables combine information of

what Bourdieu classified as economic capital, as well as cultural capital. *AUT* is somehow an indicator of socioeconomic level, not everybody has to work while studying, and the literature shows that it jeopardizes the chances of success. Being the family mainstay also indicates that the student is not at the begging of the family life-cycle and is most possibly an older student who did not have the opportunity to enter university right after completion of high school.

4 Hierarchical model

Since the research and debate on school effectiveness began almost 50 years ago, new, more comprehensive sources of data and new, more sophisticated statistical models have been developed that have improved school effectiveness studies. In particular, the development of multilevel models and the computer software to estimate them have given researchers more and better approaches for investigating school effectiveness (RUMBERGER; PALARDY, 2004, p. 236).

Hierarchical modeling is widely used in the educational area to explain the proficiency achieved by students based on student's characteristics and higher level information (classroom, school etc.). In the case of this study, it was investigated whether factors associated with the socioeconomic profile explain the linguistic performance of graduates of Higher Education courses in all areas, over the three years considered in the analysis.

One further point to be taken into consideration is whether the students' linguistic performance can be attributed to their majors. For example, all other things kept constant, a Portuguese Language student should fare better than a Math student. It was already pointed out (BELTRÃO; MANDARINO, 2014) the high correlation of socioeconomic status and undergraduate major area.

Hierarchical linear models (HLM) take into account the data natural grouping structure, combining information from variables of different levels and reducing information loss. Unlike classical regression models, which have a fixed intercept and slope coefficient for all observations, in hierarchical models the intercept and slope could be random and dependent on the higher hierarchical level, in this case: knowledge area.

In this text two HLM were used, both with two levels: the one using explanatory variables only at student level and the other incorporating the explanatory variables also at area level.

The analysis began with a null model, with the purpose of calculating the proportion of the variance of the dependent variable (linguistic performance) among and within areas. A model with explanatory variables at the student as the first level was estimated. This model included the indicator factors of the student profile, *SES* and *AUT*. Area variables were also included, with the purpose of modeling both the intercept (controlled mean of *SES* and *AUT* of the knowledge area) and the coefficient that expresses the effect of the student's socioeconomic level on their performance.

In the first model the two factors obtained in the Principal Component Analysis (*SES* and *AUT*), and the interaction between them, were used as covariates. Dependent variables were the overall language performance score and the scores on each of the three linguistic aspects. Index *i* represents a given undergraduate student, level 1, in a specific knowledge area, index *j*, level 2. The hierarchical model will then have the following general expression:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j}SES_{ij} + \beta_{2j}AUT_{ij} + \beta_{3j}AES_{ij} \times AUT_{ij} + e_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \mu_{1j}$$

$$\beta_{2j} = \gamma_{20} + \mu_{2j}$$

$$\beta_{3j} = \gamma_{30} + \mu_{3j}$$

$$\text{Two level Model: } Y_{ij} = \gamma_{00} + \gamma_{10}SES_{ij} + \gamma_{20}AUT_{ij} + \gamma_{30}SES_{ij} \times AUT_{ij} + \mu_{1j}SES_{ij} + \mu_{2j}AUT_{ij} + \mu_{3j}SES_{ij} * AUT_{ij} + \mu_{0j} + e_{ij} \quad (1)$$

where Y_{ij} is the linguistic performance of the undergraduate student *i* of knowledge area *j*; β_{0j} is the intercept of the model, defined as a random variable; β_{1j} , β_{2j} and β_{3j} are the slope coefficients associated with the socioeconomic level of area *j* (SES_{ij}), financial autonomy level of area *j* (AUT_{ij}), and the interaction between socioeconomic level and financial autonomy ($SES_{ij} \times AUT_{ij}$) respectively, also defined as random variables; γ_{00} , γ_{10} , γ_{20} and γ_{30} are fixed parameters to be estimated; μ_{0j} is the individual effect of the knowledge area associated with the intercept; μ_{1j} , μ_{2j} and μ_{3j} are the random error component of the knowledge area level associated, respectively with the slope coefficient of the socioeconomic level, the financial autonomy slope coefficient and the slope coefficient of the interaction between socioeconomic level and financial autonomy; e_{ij} is the random error component

associated with graduating students, representing the residual of the student performance measure not explained by the model.

For the second model, with two hierarchical levels, it was necessary to disaggregate the two factors obtained into four. First, we calculate averages of the two factors for all knowledge areas, and then the difference between the original score of the students and the average for the corresponding area. In this second model, the new factors, in addition to the interactions between them, were used as covariates and, again, the overall language performance score and the scores in each of the three aspects used as dependent variables. Considering, again, that each graduating student is represented by index i , and that index j stands for the knowledge area, the hierarchical model is:

$$\text{Level 1: } Y_{ij} = \beta_{0j} + \beta_{1j}SES^*_{ij} + \beta_{2j}AUT^*_{ij} + \beta_{3j}SES^*_{ij} \times AUT^*_{ij} + e_{ij}$$

$$\text{Level 2: } \beta_{0j} = \gamma_{00} + \gamma_{01}SES_j + \gamma_{02}AUT_j + \mu_{0j}$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}SES_j + \gamma_{12}AUT_j + \mu_{1j}$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}SES_j + \gamma_{22}AUT_j + \mu_{2j}$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}SES_j + \gamma_{32}AUT_j + \mu_{3j}$$

$$\begin{aligned} \text{Two level model: } Y_{ij} = & \gamma_{00} + \gamma_{01}SES_j + \gamma_{02}AUT_j + \gamma_{10}SES^*_{ij} + \gamma_{20}AUT^*_{ij} + \\ & + \gamma_{30}SES^*_{ij} \times AUT^*_{ij} + \gamma_{11}SES^*_{ij} \times SES_j + \gamma_{12}SES^*_{ij} \times AUT_j + \\ & + \gamma_{21}AUT^*_{ij} \times SES_j + \gamma_{22}AUT^*_{ij} \times AUT_j + \gamma_{31}SES^*_{ij} \times AUT^*_{ij} \times SES_j + \\ & + \gamma_{32}SES^*_{ij} \times AUT^*_{ij} \times AUT_j + \mu_{1j}SES^*_{ij} + \mu_{2j}SES^*_{ij} + \\ & + \mu_{3j}SES^*_{ij} \times AUT^*_{ij} + \mu_{0j} + e_{ij} \end{aligned} \quad (2)$$

where Y_{ij} represents the linguistic performance of the student i of knowledge area j ; β_{0j} is the intercept of the model, defined as a random variable; β_{1j} , β_{2j} and β_{3j} are the slope coefficients in area j , associated with SES^*_{ij} (difference between the socio-economic level of the student and the average of the knowledge area j , SES_j), AUT^*_{ij} (difference between financial autonomy and average of knowledge area j , AUT_j) and \times denotes interaction. Defined as fixed parameters to be estimated associated with the intercept, we have γ_{00} , γ_{10} , γ_{20} , γ_{30} , γ_{01} , γ_{11} , γ_{21} , γ_{31} , γ_{02} , γ_{12} , γ_{22} and γ_{32} ; μ_{0j} is the individual effect of the knowledge area associated with the intercept; μ_{1j} , μ_{2j} and μ_{3j} are the random error component of the knowledge area level associated, respectively with the slope coefficient of the socioeconomic level, the financial autonomy slope

coefficient and the slope coefficient of the interaction between socioeconomic level and financial autonomy; e_{ij} is the random error component associated with graduating.

In this analysis, we will first implement the null model, just with intercept, so that we can compare the influence of the other variables on the variance when they are incorporated into the models. As figures of merit we used the intraclass correlation coefficient (ICC), the Akaike information criterion (AIC) and Schwarz Bayesian criterion (BIC).

The null model is:

$$Y_{ij} = \beta_{0j} + e_{ij} \quad (3)$$

$$\beta_{0j} = \gamma_{00} + \mu_{0j}$$

$$Y_{ij} = \gamma_{00} + \mu_{0j} + e_{ij}$$

where Y_{ij} is the linguistic performance of the undergraduate student i of knowledge area j ; β_{0j} is the intercept of the model, defined as a random variable; γ_{00} , is the general intercept (overall average of language performance); μ_{0j} is the individual effect of the knowledge area associated with the intercept; and e_{ij} is the random error component associated with graduating students, representing the residual of the student performance measure not explained by the model, equal to the variance within each area of knowledge.

5 Mastery of the Portuguese language

The assessment of the Portuguese language in the exam considers three aspects, which will be analyzed separately: orthographic, textual and vocabulary/morphosyntactic (INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA, 2018).

With respect to the **orthographic** aspect, the performance of the participants revealed a very large difference in the two aspects analyzed in this competence low deviation from the spelling norm, but high deviation from the use of accent marks. There is often a complete absence of graphic accentuation in most of the words in the text, perhaps motivated by the habits related to social networks and the lack of clarification about the decisions of the 1990 Portuguese Language Orthographic Agreement.

Regarding the domain of the conventions related to the spelling of words, sporadic deviations were observed, mainly driven by casual orality, replacing letters with similar sounds, like “s”, “c” or “ç”, “g” and “j”, among others. It is also worth noting that there were found no abbreviations related to the use of social networks and e-mails.

For some of the undergraduates the **textual** aspect proved to be the most problematic, given the numerous problems observed, accumulated deviations throughout the school educational system. They are: juxtaposed sequence of ideas without syntactic fittings, leading to a reduction in subordinate structures, along with an increase in the frequency of coordinate and absolute structures; fragmented sentences that compromise the logical-grammatical structure; sentences formed only by subordinate clause, without main clause; reduction in the use of connectors to express logical relations essential to the construction of the text as a consequence of the change of phrasal structure; inappropriate use of relative pronoun (omitting or using an inappropriate relative pronoun), reflecting oral habits; absence of referencing features, such as substitution of terms for synonyms, hyponyms, hypernym, nominalizations, metaphorical expressions.

These problems reveal difficulties in relation to the formal structure of the text produced, which is worrisome. In some texts, a minimum of textuality and mastery of the standard language register is lacking. Regarding the use of punctuation marks, their absence was observed in most of the analyzed texts.

Regarding **morphosyntactic** aspects, the most frequent deviation was the preposition assignment to verbs (improper use or its absence) and before relative pronoun, a generalized process in the oral language modality, in casual situations. Despite the possibility that this change of conduct could be generalized in the written standard of the Portuguese language, as is already happening even in journalistic texts, the non-use of the preposition was considered as deviation in this evaluation process.

Another problem is gender and number agreement among verbs, determiner, nouns and adjectives. Verbal agreement and nominal agreement showed some very frequent deviations. An extra problem in Portuguese is the fact that the indicator of plural in some verb forms is just a circumflex in the “e” vowel. As for gender agreement, several cases were observed, usually within long noun phrases, where the adjective is far from the noun.

Regarding **vocabulary** aspects, some types of inadequacy were observed: oral expressions; vocabulary selection incompatible with context, leading to lack of

intelligibility; lack of mastery of a more complex vocabulary, essential for the development of the text. The main aspect observed was the excessive repetition of certain words, such as the term “person”, revealing limitation of vocabulary repertoire.

The text grading suggests that the written mode, at least among this group of university students, has the tendency to be clearly simplified, approaching the characteristics of the Portuguese language spoken register. In the case of essay-based and argumentative-based texts, the distance between the two modalities is even greater, which causes recurrent deviations in all three aspects analyzed. In this evaluation, it is worth mentioning the textual impairment, through fragmented and/or truncated structures, breaking the expected syntactic complexity in the formal pattern.

6 Results

Table 1 presents the null model parameter estimates for language performance and its aspects as the dependent variable. The overall average language performance (d) is approximately 64.4 (on a 100 points scale) and 7.3% of the total variation in language performance can be explained by the knowledge area *j* in which the student participates.

Decomposing the linguistic performance in each of its three aspects, as can be seen in Table 1, the average linguistic performance in the vocabulary/morphosyntactic (a), textual (b) and orthographic aspects (c) are, respectively, 25.9 (on a 40 points scale), 24.3 (on a 40 points scale) and 14.2 (on a 20 points scale), adding to the 64.4 of the overall performance (d). The explained proportion of the total variation in language performance in each of the three aspect is 5.5%, 4.4% and 7.7%, respectively (see Table 2).

Table 1 - Fixed effects estimates – null model

Fixed Effects	Estimate	St. Error	df	t-stat	p-value
Intercept (γ_{00}) ^a	25.915375	0.144160	88.184	179.768	0.000
Intercept (γ_{00}) ^b	24.281967	0.127517	88.187	190.421	0.000
Intercept (γ_{00}) ^c	14.166924	0.094642	88.015	149.689	0.000
Intercept (γ_{00}) ^d	64.332438	0.359576	88.101	178.912	0.000

Dependent Variable: a. Morphosyntactic and vocabulary performance; b. textual performance; c. orthographic performance; d. Portuguese performance¹.

Source: Authors' calculations (2020)

¹ This note is the same for all tables presented.

Table 2 - Estimates of variance parameters – null model

	Variance Components	Estimate	St. Error	z-stat	p-value
a	Residual (e_{ij})	31.02791	0.044618	695.414	0.000
	Intercept (μ_{0j})	1.817419	0.275399	6.599	0.000
b	Residual (e_{ij})	31.05931	0.044663	695.414	0.000
	Intercept (μ_{0j})	1.419549	0.215469	6.588	0.000
c	Residual (e_{ij})	9.361674	0.013462	695.414	0.000
	Intercept (μ_{0j})	0.784778	0.118816	6.605	0.000
d	Residual (e_{ij})	144.072369	0.205650	700.572	0.000
	Intercept (μ_{0j})	11.325933	1.714249	6.607	0.000

Source: Authors' calculations (2020)

Table 3 presents the figures of merit (ICC, AIC and BIC) for the null model for the four dependent variables: (a) Morphosyntactic/vocabular; (b) textual; (c) orthographic; and (d) overall Portuguese.

Table 3 - Figures of merit– null model

Dependent variable	Figure of Merit	Value
a	ICC	0.0553327
	AIC	6,068,083.532
	BIC	6,068,118.878
b	ICC	0.0437068
	AIC	6,069,040.320
	BIC	6,069,075.667
c	ICC	0.0773450
	AIC	4,909,048.232
	BIC	4,909,083.579
d	ICC	0.0728832
	AIC	7,665,759.7382932
	BIC	7,665,795.1293858

ICC: Intraclass correlation coefficient, AIC: Akaike information criterion, BIC: Schwarz Bayesian criterion.

Source: Authors' calculations (2020)

The complete model in (1) has four fixed effects and four random effects. The first level represents the linguistic performance of the student i of knowledge area j as a function of the average linguistic performance of the knowledge area (β_{0j}), the socioeconomic level of the student (β_{1j}), the financial autonomy of the student (β_{2j}), the interaction between the socioeconomic level and the financial autonomy of the student (β_{3j}) and the random error at the student's (e_{ij}) level.

Table 4 and Table 5 present the estimates of the complete model with overall language performance as the dependent variable for model (1). The overall average language performance is approximately 64.3 and once the covariates are introduced, only 6.5% of the total variation in language performance can be explained by the knowledge area. Moreover, when comparing this model with the null model, the proportion of variance explained by the newly incorporated variables is 12.7% at the knowledge area level and 2.0% at the student level, the residual variance among knowledge areas decreased from 11.3 to 9.8, the internal residual variance of knowledge areas also decreased from 143.3 to 140.4. There are, as well, reductions in the figures of merit indicating a better model (see Table 6): ICC (from 0.07288 to 0.06551), AIC (from 7,665,759.738 to 7,528,881.779) and BIC (from 7,665,795.129 to 7,528,987.819). The complete model has a greater explanation power of the total variation of the phenomenon being measured.

Decomposing the linguistic performance by each of its three aspects, we have, as can be seen in the tables below, that the average linguistic performance in the vocabulary/morphosyntactic (a), textual (b) and orthographic (c) aspects are approximately equal to those of the null model. The explained proportion of the total variation in language performance in each aspect decreased to 5.0%, 3.7% and 7.2%, respectively. Still comparing with the null model: the proportion of variance explained by the new incorporated variables is 11.4% at the knowledge area level and 1.3% at the final level for the vocabulary/morphosyntactic aspect, 15.8%. and 1.2% for the textual aspect, and 9.9% and 2.2% for the orthographic aspect respectively. Residual variances for the vocabulary/morphosyntactic, textual and orthographic aspects among the areas of knowledge decreased from 1.8 to 1.6, 1.4 to 1.2 and 0.8 to 0.7, respectively. The internal residual variances for the vocabulary/morphosyntactic, textual and orthographic aspects of the knowledge areas also decreased, respectively, from 31.0 to 30.6, from 31.0 to 30.7 and from 9.4 to 9.1. We can also observe a reduction in ICC, AIC and BIC for each aspect. This indicates that the complete model also has a greater explanation of the total variation of the phenomenon for each of the aspects that make up the linguistic performance of the graduates.

Table 4 - Fixed effects estimates – model 1

	Fixed Effects	Estimate	St. Error	Df	t-stat	p-value
a	Intercept (γ_{00})	25.905361	0.135860	88.135	190.676	0.000
	SES (γ_{10})	0.623909	0.024304	84.739	25.671	0.000
	AUT (γ_{20})	-0.120678	0.014852	75.598	-8.125	0.000
	SES × AUT (γ_{30})	0.036721	0.010807	68.896	3.398	0.001
b	Intercept (γ_{00})	24.265565	0.117185	88.055	207.071	0.000
	SES (γ_{10})	0.576504	0.026093	85.862	22.094	0.000
	AUT (γ_{20})	-0.209221	0.016832	75.029	-12.430	0.000
	SES × AUT (γ_{30})	0.039465	0.010890	70.761	3.624	0.001
c	Intercept (γ_{00})	14.145597	0.089916	88.063	157.320	0.000
	SES (γ_{10})	0.453621	0.012621	80.820	35.941	0.000
	AUT (γ_{20})	-0.129830	0.011773	81.487	-11.027	0.000
	SES × AUT (γ_{30})	0.015704	0.006050	68.279	2.596	0.012
d	Intercept (γ_{00})	64.318747	0.335567	88.068	191.672	0.000
	SES (γ_{10})	1.649836	0.060929	86.388	27.078	0.000
	AUT (γ_{20})	-0.452088	0.040579	78.811	-11.141	0.000
	SES × AUT (γ_{30})	0.097533	0.025234	73.015	3.865	0.000

SES: Students' socioeconomic status, AUT: Financial autonomy/independence.
Source: Authors' calculations (2020)

Table 5 - Estimates of variance parameters – model 1

Variance Components	Estimate	St. Error	z-stat	p-value
Residual (e_{ij}) a	30.636012	0.044059	695.335	0.000
Intercept (μ_{0j}) a	1.610835	0.244661	6.584	0.000
SES (μ_{1j}) a	0.040693	0.007652	5.318	0.000
AUT (μ_{2j}) a	0.010919	0.002721	4.013	0.000
SES × AUT (μ_{3j}) a	0.003261	0.001212	2.690	0.007
Residual (e_{ij}) b	30.678797	0.044121	695.334	0.000
Intercept (μ_{0j}) b	1.194702	0.182089	6.561	0.000
SES (μ_{1j}) b	0.048303	0.008830	5.470	0.000

Continue

Continuation

AUT (μ_2) b	0.015782	0.003658	4.315	0.000
SES \times AUT (μ_3) b	0.003306	0.001213	2.725	0.006
Residual (e_{ij}) c	9.156087	0.013168	695.331	0.000
Intercept (μ_0) c	0.707239	0.107215	6.596	0.000
SES (μ_1) c	0.010698	0.002098	5.099	0.000
AUT (μ_2) c	0.009113	0.001807	5.043	0.000
SES \times AUT (μ_3) c	0.001049	0.000386	2.720	0.007
Residual (e_{ij}) d	140.423787	0.201952	695.331	0.000
Intercept (μ_0) d	9.844456	1.493201	6.593	0.000
SES (μ_1) d	0.271405	0.048355	5.613	0.000
AUT (μ_2) d	0.099996	0.021329	4.688	0.000
SES \times AUT (μ_3) d	0.020783	0.006819	3.048	0.002

SES: Students' socioeconomic status, AUT: Financial autonomy/independence.

Source: Authors' calculations (2020)

Table 6 - Figures of merit– model 1

Dependent variable	Figure of Merit	Value
a	ICC	0.049953262
	AIC	6,056,095.156
	BIC	6,056,201.196
b	ICC	0.037482609
	AIC	6,057,450.498
	BIC	6,057,556.538
c	ICC	0.071703891
	AIC	4,887,931.362
	BIC	4,888,037.402
d	ICC	0.0655126
	AIC	7,528,881.779
	BIC	7,528,987.819

ICC: Intraclass correlation coefficient, AIC: Akaike information criterion, BIC: Schwarz Bayesian criterion.

Source: Authors' calculations (2020)

For model (2), the null model represented in (3) will also be used for comparison. The structure and estimates of the null model are the same as those already presented in the previous model.

The complete model (2) has twelve fixed effects and four random effects. However, when estimating the model, three fixed effects were not significant at the 5% significance level in the fixed effects test and were therefore removed from the analysis. Table 7 provides this test.

Table 7 - Significance of fixed effects^d – model 2

Fixed Effects	df	F-stat	p-value
Intercept	88.039	40.885.673	0.000
SES_MEAN	87.781	16.026	0.000
AUT_MEAN	87.954	4.121	0.045
SES*	84.680	897.759	0.000
AUT*	71.184	153.567	0.000
SES_MEAN × SES*	77.573	20.084	0.000
SES_MEAN × AUT*	63.349	2.491	0.119
AUT_MEAN × SES*	83.426	0.083	0.774
AUT_MEAN × AUT*	78.030	4.288	0.042
SES* × AUT*	60.192	19.091	0.000
SES_MEAN × SES* × AUT*	58.290	18.579	0.000
AUT_MEAN × SES* × AUT*	84.020	1.174	0.282

SES: Students' socioeconomic status, AUT: Financial autonomy/independence.

Source: Authors' calculations (2020)

Table 8 and Table 9 present the estimates of the complete model with overall language performance as a dependent variable with non-significant variables excluded from the analysis, with nine fixed effects and four random effects. The overall average language performance is approximately 64.1 and only 5.7% of the total variation in language performance can be explained by the knowledge area in which the student is enrolled. Moreover, when comparing this model with the null model, the proportion of variance explained by the newly incorporated variables is 25.1% at the knowledge area level and 2.5% at the student level. The residual variance between knowledge areas decreased from 11.3 to 8.5, the

internal residual variance of the knowledge areas also decreased from 144.1 to 140.4, as did the ICC, AIC and BIC values.

Comparing now with model (1), the proportion of variance (Table 9) explained by the new variables is 13.8% at the knowledge area level and 0.0% at the student level, the residual variance among the knowledge areas decreased from 9.8 to 8.5, the internal residual variance of the knowledge areas remained practically the same, 140.4, though ICC presented a substantial decrease (see Table 10) and the value of AIC also decreased. However, the value of BIC increased. This indicates that this new complete model has a better explanation of the total variation of the phenomenon and, therefore, can be considered a better model to explain the phenomenon.

Table 8 - Fixed effects estimates^d – model 2

Fixed Effects	Estimate	St. Error	df	t-stat	p-value
Intercept (γ_{00})	64.107683	0.317055	88.039	202.197	0.000
SES_MEAN (γ_{01})	3.338720	0.834185	87.781	4.002	0.000
AUT_MEAN (γ_{02})	-1.601614	0.787206	87.951	-2.035	0.045
SES* (γ_{10})	1.670678	0.054981	84.703	30.387	0.000
AUT* (γ_{20})	-0.451524	0.037243	73.817	-12.124	0.000
SES_MEAN \times SES* (γ_{11})	0.106574	0.023080	66.420	4.618	0.000
AUT_MEAN \times AUT* (γ_{22})	-0.629557	0.135202	78.740	-4.656	0.000
SES* \times AUT* (γ_{30})	0.235995	0.090405	83.999	2.610	0.011
SES_MEAN \times SES* \times AUT* (γ_{31})	-0.223955	0.055118	64.800	-4.063	0.000

SES: Students' socioeconomic status, AUT: Financial autonomy/independence.

Source: Authors' calculations (2020)

Table 9 - Estimates of variance parameters^d – model 2

Variance Components	Estimate	St. Error	z-stat	p-value
Residual (e_{ij})	140.424345	0.201952	695.335	0.000
Intercept (μ_{0j})	8.482566	1.285848	6.597	0.000
SES (μ_{1j})	0.211653	0.039063	5.418	0.000
AUT (μ_{2j})	0.076826	0.017585	4.369	0.000
SES \times AUT (μ_{3j})	0.012806	0.005124	2.499	0.012

SES: Students' socioeconomic status, AUT: Financial autonomy/independence.

Source: Authors' calculations (2020)

Table 10 - Figures of merit^d – model 2

Figure of Merit	Value
ICC	0.0569656
AIC	7,528,805.704
BIC	7,528,998.655

ICC: Intraclass correlation coefficient, AIC: Akaike information criterion, BIC: Schwarz Bayesian criterion.

Source: Authors' calculations (2020)

The same situation that occurs with linguistic performance, when fixed effects are not significant at the 5% level, also occurs with two of the three aspects evaluated: vocabulary/morphosyntactic and textual aspects. In both, the same variables that were eliminated from the linguistic performance analysis were also left out in the analysis of these two aspects. Regarding orthographic aspects, besides the same three non-significant variables, two other variables were also non-significant. However, one of them, the variable *AUT*, being considered essential for the analysis, was not removed from the model.

For the vocabulary/morphosyntactic aspects, comparing with the null model: the explained proportion of the total variation in language performance decreased to 4.4% (see Table 11 and Table 12); the proportion of variance explained by the newly incorporated variables was 23.3% at the knowledge area level and 1.7% at the student level; the residual variance among knowledge areas decreased from 1.8 to 1.4; and the internal residual variances reduced from 31.1 to 30.6. For these aspects, comparing with model (1): the proportion of variance explained by the newly incorporated variables is 13.2% at the knowledge area level and none at the student level; the residual variance between knowledge areas decreased from 1.6 to 1.4; and the internal residual variance of the knowledge areas remained the same.

Table 11 - Fixed effects estimates^a – model 2

Fixed Effects	Estimate	St. Error	df	t-stat	p-value
Intercept (γ_{00})	25.814163	0.128838	88.074	200.361	0.000
SES_MEAN (γ_{01})	1.256547	0.338899	87.736	3.708	0.000
AUT_MEAN (γ_{02})	-0.663630	0.319862	87.959	-2.075	0.041
SES* (γ_{10})	0.632281	0.021811	82.435	28.990	0.000

Continue

Continuation

AUT* (γ_{20})	-0.122297	0.013709	70.167	-8.921	0.000
SES_MEAN \times SES* (γ_{11})	0.040979	0.009736	63.892	4.209	0.000
AUT_MEAN \times AUT* (γ_{22})	-0.246724	0.053269	75.237	-4.632	0.000
SES* \times AUT* (γ_{30})	0.076643	0.033886	85.610	2.262	0.026
SES_MEAN \times SES* \times AUT* (γ_{31})	-0.107939	0.023219	62.968	-4.649	0.000

SES: Students' socioeconomic status, AUT: Financial autonomy/independence.
Source: Authors' calculations (2020)

Table 12 - Estimates of variance parameters^a – model 2

Variance Components	Estimate	St. Error	z-stat	p-value
Residual (e_{ij})	30.636021	0.044059	695.342	0.000
Intercept (μ_{0j})	1.398136	0.212279	6.586	0.000
SES (μ_{1j})	0.030880	0.006100	5.063	0.000
AUT (μ_{2j})	0.008245	0.002239	3.683	0.000
SES \times AUT (μ_{3j})	0.001677	0.000814	2.061	0.039

SES: Students' socioeconomic status, AUT: Financial autonomy/independence.
Source: Authors' calculations (2020)

For textual aspects, compared to the null model: the explained proportion of the total variation in language performance decreased to 3.0% (Table 13 and Table 14); the proportion of variance explained by the newly incorporated variables is 34.3% at the knowledge area level and 1.5% at the student level; the residual variance among knowledge areas decreased from 1.4 to 0.9; and the internal residual variances reduced from 31.1 to 30.7. Also, comparing with model (1): the proportion of variance explained by the new variables is 21.7% at the knowledge area level and none at the student level; the residual variance among knowledge areas decreased from 1.2 to 0.9; and the internal residual variance of the knowledge areas remained the same, 30.7.

Table 13 - Fixed effects estimates^b – model 2

Fixed Effects	Estimate	St. Error	Df	t-stat	p-value
Intercept (γ_{00})	24.165763	0.105598	88.042	228.846	0.000
SES_MEAN (γ_{01})	1.216493	0.277638	87.543	4.382	0.000

Continue

Continuation

AUT_MEAN (γ_{02})	-0.867506	0.262124	87.872	-3.310	0.001
SES* (γ_{10})	0.585540	0.023853	83.966	24.548	0.000
AUT* (γ_{20})	-0.207182	0.015769	68.242	-13.139	0.000
SES_MEAN \times SES* (γ_{11})	0.041927	0.009988	57.003	4.198	0.000
AUT_MEAN \times AUT* (γ_{22})	-0.258612	0.058486	77.407	-4.422	0.000
SES* \times AUT* (γ_{30})	0.093223	0.038529	79.539	2.420	0.018
SES_MEAN \times SES* \times AUT* (γ_{31})	-0.092064	0.023831	56.066	-3.863	0.000

SES: Students' socioeconomic status, AUT: Financial autonomy/independence.
Source: Authors' calculations (2020)

Table 14 - Estimates of variance parameters^b – model 2

Variance Components	Estimate	St. Error	z-stat	p-value
Residual (e_{ij})	30.678923	0.044121	695.335	0.000
Intercept (μ_{0j})	0.935776	0.142617	6.561	0.000
SES (μ_{1j})	0.038590	0.007321	5.271	0.000
AUT (μ_{2j})	0.012658	0.003182	3.978	0.000
SES \times AUT (μ_{3j})	0.001918	0.000940	2.040	0.041

SES: Students' socioeconomic status, AUT: Financial autonomy/independence.
Source: Authors' calculations (2020)

For orthographic aspects, comparing with the null model (see Table 15 and Table 16): the explained proportion of the total variation in language performance decreased to 6.7%; the proportion of variance explained by the new variables incorporated is 16.2% at the knowledge area level and 2.6% at the graduating level; the residual variance between the knowledge areas decreased from 0.8 to 0.7; and the internal residual variance decreased from 9.4 to 9.1.

Still for the orthographic aspects, comparing with model (1): the proportion of variance explained by the new variables is 6.7% at the knowledge area level and 0.0% at the graduating level; the residual variance between knowledge areas decreased from 1.2 to 0.9; and the internal residual variance of the knowledge areas remained the same, 30.7.

Table 15 - Fixed effects estimates^c – model 2

Fixed Effects	Estimate	St. Error	Df	t-stat	p-value
Intercept (γ_{00})	24.165763	0.105598	88.042	228.846	0.000
SES_MEAN (γ_{01})	1.216493	0.277638	87.543	4.382	0.000
AUT_MEAN (γ_{02})	-0.867506	0.262124	87.872	-3.310	0.001
SES* (γ_{10})	0.585540	0.023853	83.966	24.548	0.000
AUT* (γ_{20})	-0.207182	0.015769	68.242	-13.139	0.000
SES_MEAN × SES* (γ_{11})	0.041927	0.009988	57.003	4.198	0.000
AUT_MEAN × AUT* (γ_{22})	-0.258612	0.058486	77.407	-4.422	0.000
SES* × AUT* (γ_{30})	0.093223	0.038529	79.539	2.420	0.018
SES_MEAN × SES* × AUT* (γ_{31})	-0.092064	0.023831	56.066	-3.863	0.000

SES: Students' socioeconomic status, AUT: Financial autonomy/independence.

Source: Authors calculations (2020)

Table 16 - Estimates of variance parameters^c – model 2

Variance Components	Estimate	St. Error	z-stat	p-value
Residual (e_{ij})	30.678923	0.044121	695.335	0.000
Intercept (μ_{0j})	0.935776	0.142617	6.561	0.000
SES (μ_{1j})	0.038590	0.007321	5.271	0.000
AUT (μ_{2j})	0.012658	0.003182	3.978	0.000
SES × AUT (μ_{3j})	0.001918	0.000940	2.040	0.041

SES: Students' socioeconomic status, AUT: Financial autonomy/independence.

Source: Authors' calculations (2020)

For the three aspects that make up language performance, when comparing model (2) with the null model, there was an increase in the proportion of variance explained with the incorporation of the new variables, a reduction in the residual variances between the knowledge areas and the internal residual variances, as well as AIC and BIC values (see Table 17).

Comparing to model (1), there was also an increase in the proportion of variance explained with the new variables, a reduction in residual variances between areas of knowledge and internal residual variances, a reduction in AIC values and an increase in BIC. This indicates that, although the BIC value increases, model (2), as well as the general linguistic performance, presents a greater

explanation of the total variation and, therefore, can be considered a better model to explain the phenomenon in each of the aspects that make up the overall language performance.

Table 17 - Figures of merit – model 2

Dependent variable	Figure of Merit	Value
a	ICC	0.0436452
	AIC	6,056,044.200
	BIC	6,056,209.152
b	ICC	0.0295994
	AIC	6,057,401.107
	BIC	6,057,566.058
c	ICC	0.0672264
	AIC	4,887,904.882
	BIC	4,888,058.052

Source: Authors' calculations (2020)

ICC: Intraclass correlation coefficient, AIC: Akaike information criterion, BIC: Schwarz Bayesian criterion.

7 Conclusions

As shown in the results, the first model has some explanatory power of the total variation of the phenomenon being measured above the null model. Decomposing the linguistic performance by each of its three aspects, we have that the average linguistic performance in the vocabulary/morphosyntactic (a), textual (b) and orthographic (c) aspects are approximately equal to those of the null model. This indicates that the complete model also has a greater explanation of the total variation of the phenomenon for each of the aspects that make up the linguistic performance of the graduating students. This is to say, that even controlling for the knowledge area, *SES* and *AUT* of the graduating students are both statistically significant for the performance in Portuguese, according to the grades in the exam. This assertion holds true for all the three aspects considered in language performance: textual; orthographic; and vocabulary/morphosyntactic.

The second hierarchical model has a slight better explanation of the total variation of the phenomenon and, therefore, can be considered a better model to explain

the phenomenon. But it is important to note that using the average *SES* and *AUT* of the graduating students in each area does not add much to the explanatory power, suggesting that the effect is more a knowledge area effect (e.g. Portuguese major as opposed to a STEM major) than a peer effect. For this model also, decomposing the language performance in its three aspects, when comparing model (2) with model (1), there was an increase in the proportion of variance explained, a reduction in the residual variances among the knowledge areas and the internal residual variances, as well as two of the figures of merit: ICC and AIC.

Confirming Bourdieu and Passeron (1970) hypothesis, we find out that socioeconomic status do have an impact on language proficiency, but at the University level it is not so strong as measured at lower educational levels, most possibly because some filtering has already taken place. As already stated, university students constitute a much more homogeneous group.

We show in this article that the more affluent socioeconomic groups have greater proficiency in the exams, in particular at the Enade Portuguese Language component, even when controlling for the average socio-economic level of the knowledge area considered. Our text shows that among Brazilian graduating students, socioeconomic indicators are correlated with the mastery of the formal language. This supports Bourdieu's (1966) thesis that through the social, cultural and economic capital still prevails the domination of wealthier classes over more popular classes, since the criteria of excellence are defined by the upper classes, reinforcing the inequality in order to maintain domination.

Efeitos do status socioeconômico no uso da linguagem escrita: estende-se ao Ensino Superior brasileiro?

Resumo

Bourdieu e Passeron defenderam a tese de que a escola era o principal locus para legitimar e perpetuar as diferenças de classe. Isso é reforçado pelos múltiplos testes de proficiência utilizados para monitoramento de políticas públicas, que privilegiam o uso da linguagem formal como parte dos instrumentos e, portanto, penalizam os participantes com menor domínio da língua. Ajustamos dois modelos hierárquicos com os resultados do Enade nas notas de desempenho linguístico (português), usando como covariáveis, indicadores da condição socioeconômica e autonomia financeira dos alunos e dos valores médios dessas variáveis para as áreas de conhecimento. O desempenho linguístico é desagregado em três aspectos: textual, ortográfico e vocabulário/morfossintático. Mostramos que os grupos socioeconômicos mais afluentes apresentam maior proficiência no componente linguístico do Enade, mesmo controlando pela média do nível socioeconômico e da autonomia financeira dos alunos da área. O efeito socioeconômico não é tão forte quanto em níveis educacionais mais baixos, mas universitários constituem um grupo social homogêneo. Isso reforça a tese de Bourdieu de que por meio do capital social, cultural e econômico ainda prevalece o domínio das classes mais ricas sobre as classes mais populares, reforçando a desigualdade.

Palavras-chave: Capital Social. Capital Econômico. Desigualdade Social. Desempenho Linguístico. Enade. Políticas Públicas. Modelos Hierárquicos.

Efectos de lo estatus socioeconómico en el uso de la lengua escrita: ¿se extiende a la Educación Superior brasileña?

Resumen

Bourdieu y Passeron defendieron la tesis de que la escuela era el locus principal para legitimar y perpetuar las diferencias de clase. Esto se ve reforzado por las múltiples pruebas de competencia utilizadas para el seguimiento de las políticas públicas, que privilegian el uso del lenguaje formal como parte de los instrumentos y, por tanto, penalizan a los participantes con menor conocimiento del idioma. Ajustamos dos modelos jerárquicos con los resultados de Enade en los puntajes de desempeño lingüístico (portugués), utilizando como covariables, indicadores del nivel socioeconómico e autonomía financiera de los estudiantes y valores promedio de estas variables para las áreas de conocimiento. Mostramos que los grupos de nivel socioeconómicosociales más altos tienen mayor dominio del componente lingüístico de Enade, aunque controlando por el promedio del nivel socioeconómico y autonomía financiera de los estudiantes de la área. El efecto socioeconómico no es tan fuerte como en los niveles educativos más bajos, pero debe tenerse en cuenta que los estudiantes universitarios son un grupo social mucho más homogéneo. Los

hallazgos refuerzan la tesis de Bourdieu de que, a través del capital social, cultural y económico, aún prevalece el dominio de las clases más ricas sobre las clases más populares, reforzando la desigualdad.

Palabras clave: *Capital Social. Capital Económico. Desigualdad Social. Desempeño Lingüístico. Enade. Políticas Públicas. Modelos Jerárquicos.*

References

- BELTRÃO, K.; MANDARINO, M. C. F. *Perfil socioeconômico dos concluintes de cursos superiores de 2004 a 2012*. Rio de Janeiro: Fundação Cesgranrio, 2014.
- BOURDIEU, P. L'école conservatrice: les inégalités devant l'école et devant la culture. *Revue Française de Sociologie*, Paris, v. 7, n. 3, p. 325-347, July/Sept.1966.
- BOURDIEU, P.; PASSERON, J. C. *La reproduction: éléments d'une théorie du système d'enseignement*. Paris: Minuit, 1970. (Collection le sens commun).
- CÂNDIDO, P. T. Comunicação em matemática. In: DINIZ, M. I.; SMOLE, K. S. (orgs.). *Ler, escrever e resolver problemas: habilidades básicas para aprender matemática*. Porto Alegre: Artmed, 2001. p. 15-29.
- COLEMAN, J. S. *Equality of educational opportunity*. Ann Arbor: Inter-university Consortium for Political and Social Research, 1966.
- COLEMAN, J. S. Social capital in the creation of human capital. *American Journal of Sociology*, Chicago, v. 94, n. Suppl., s95-s120, 1988.
- COLEMAN, J. S. Some points on choice in education. *Sociology of Education*, Thousand Oaks, v. 65, n. 4, p. 260-262, Oct. 1992. <https://doi.org/10.2307/2112769>
- DIKA, S. L.; SINGH, K. Applications of social capital in educational literature: a critical synthesis. *Review of Educational Research*, Thousand Oaks, v. 72, n. 1, p. 31-60, Mar. 2002.
- FRITSCH, R. *et al.* Percursos escolares de estudantes do ensino médio de escolas públicas do município de São Leopoldo, RS: desempenho escolar, perfil e características. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 27, n. 104, p. 543-567, July/Sept. 2019. <https://doi.org/10.1590/s0104-40362019002701306>
- GONCALVES, F. O.; FRANCA, M. T. A. Transmissão intergeracional de desigualdade e qualidade educacional: avaliando o sistema educacional brasileiro a partir do SAEB 2003. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 16, n. 61, p. 639-662, Oct./Dec. 2008. <https://doi.org/10.1590/S0104-40362008000400009>
- INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA - INEP. *Exame nacional de desempenho dos estudantes*: database. Brasília: Ministério da Educação, 2015.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS
ANÍSIO TEIXEIRA - INEP. *Relatório síntese de área: formação geral*.
Brasília: Ministério da Educação, 2018.

LAROS, J. A.; MARCIANO, J. L.; ANDRADE, J. M. Fatores associados ao desempenho escolar em português: um estudo multinível por regiões. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 20, n. 77, p. 623-646, Oct./Dec. 2012.
<https://doi.org/10.1590/S0104-40362012000400002>

SMOLE, K. C. S. Textos em matemática: por que não? In: DINIZ, M. I.; SMOLE, K. C. S. (orgs.). *Ler, escrever e resolver problemas: habilidades básicas para aprender matemática*. Porto Alegre: Artmed, 2001. p. 29-68

RUMBERGER, R. W.; PALARDY, G. J. Multilevel models for school effectiveness research. In: KAPLAN, D. (ed.). *The sage handbook of quantitative methodology for the social sciences*. Thousand Oaks: Sage, 2004. p. 235-258.



Information about the authors

Kaizô Iwakami Beltrão: Ph.D. in Statistics. Full Professor at Escola Brasileira de Administração Pública da Fundação Getúlio Vargas. Chief Statistician of the Academic Department at Fundação Cesgranrio, Rio de Janeiro, RJ, Brasil. Contact: kaizo.beltrao@fgv.br

 <https://orcid.org/0000-0002-3590-8057>

Mônica Cerbella Freire Mandarin: Ph.D. in Education. Project Supervisor of the Academic Department at Fundação Cesgranrio, Rio de Janeiro, RJ, Brasil. Contact: monicamandarin@cesgranrio.org.br

 <https://orcid.org/0000-0003-4230-4662>

Ricardo Servare Megahós: Bachelor of Statistics. Statistician at Fundação Cesgranrio, Rio de Janeiro, RJ, Brasil. Contact: ricardo.servare@cesgranrio.org.br

 <https://orcid.org/0000-0002-7198-3134>

Mônica Guerra Ferreira Pedrosa: Master and Social Research. Statistician at Fundação Cesgranrio, Rio de Janeiro, RJ, Brasil. Contact: monicaguerra@cesgranrio.org.br

 <https://orcid.org/0000-0002-7174-0351s>