Aldashev, Alisher und Danzer, Alexander M.:

Economic Returns to Speaking the Right Language(s)? Evidence from Kazakhstan's Shift in State Language and Language of Instruction

Economic Returns to Speaking the Right Language(s)? Evidence from Kazakhstan's Shift in State Language and Language of Instruction

Alisher Aldashev[*]

International School of Economics, Kazakh-British Technical University of Almaty

Alexander M. Danzer[†]

Ludwig-Maximilians-University, Munich, IOS Regensburg, IZA Bonn and CReAM London

Abstract

This paper investigates the economic returns to language skills and bilingualism. The analysis is staged in Kazakhstan, a multi-ethnic country with complex ethnic settlement patterns that has switched its official state language from Russian to Kazakh. Using two newly assembled data sets, we find negative returns to speaking Kazakh and a negative effect of bilingualism on earnings while Russian was the official state language in the 1990s. Surprisingly, the Kazakh language continues to yield a negative wage premium 13 years after it has been made official state language. While we do neither find evidence for an ethnically segmented labor market nor for reverse causality, the low economic value of the Kazakh language can be explained by the comparatively poor quality of schools with Kazakh as language of instruction. Based on PISA data, we illustrate that scholastic achievements are substantially lower for pupils taught in Kazakh, despite the official support for the titular language. Our results suggest that switching the official state language without appropriate investments in school resources is unlikely to cure the economic disadvantage of a previously marginalized language.

Keywords: Bilingualism, returns to language skills, wage premium, language policy, language of instruction

JEL Classification: J24, I21, P23, O15

---

[*]Alisher Aldashev is associate professor at the International School of Economics, Almaty. Address: Kazakh-British Technical University, International School of Economics, 59 Tole bi Str., Almaty 050000, Kazakhstan.

[†]Alexander M. Danzer is junior professor at Ludwig-Maximilians-University, Munich, IOS Regensburg, IZA Bonn and CReAM London. Address: Ludwig-Maximilians-University Munich, Department of Economics, Geschwister-Scholl-Platz 1, 80539 Munich, Germany.

1. Introduction

Language is an important component of the nation-building process and hence a potentially important policy variable. Changing the state language and/or language of instruction at school might suit governments in their attempt to create a national identity, yet it also affects workers in their productive capacity. So, what are the returns to speaking a newly introduced official language as opposed to speaking the out-dated one? And do bilinguals earn more, less or about the same as monolingual workers?

The previous literature on the economic returns to language knowledge and to knowing more than one language delivers mixed results and the ambiguity of the evidence may partly stem from institutional differences between the analyzed countries (Belgium, Bolivia, Canada, Switzerland, Wales). A methodological challenge faced by all studies on the effects of language skills stems from the potential self-selection of individuals into learning languages or into bilingualism. While most previous studies cannot deal with this problem, some predominantly bilingual countries offer rare policy interventions which exogenously switch the official state language.

In this paper we study the economic returns to the knowledge of different languages, including bilingualism. Our case-study is staged in Kazakhstan which has been characterized by widespread bilingualism, realtively equal ethnic composition between two major population groups (Kazakhs and Russians) as well as a switch in the official state language. Importantly, the country offers two highly comparable labor force surveys which provide language information and, hence, allow comparing the returns to language knowledge before and after the language shift. The paper adresses two questions: Firstly, we are interested in the relative wages of Kazakh- and Russian-language speakers. Given the widespread usage of the Russian language and its role as business language one might expect a wage differential with Russian-language speakers earning more on average. However, given the political support that the Kazakh language enjoys since its nomination as official state language in the 1990s and given the ever-growing share of population educated in the Kazakh language the differential might have reversed or at least decreased recently. Secondly, we would like to assess whether the individuals who are proficient in both languages, Kazakh (the official state language) and Russian

1

(the language most widely spoken and understood), enjoy a wage premium over workers who are fluent in only one of the two languages.

The contribution of our paper is twofold: First, we use two comparable surveys from different point in times (1996 before and 2010 after Kazakh has been made official state language) in order to investigate the economics returns to language knowledge and to bilingualism. We choose two distant points in time for our analysis since the language shift was implemented slowly; yet it has been completed by the year 2010. We find—quite surprisingly—that for both points in time the knowledge of Kazakh yields a negative economic premium—as does bilingualism. Second, we shed light on the quality of schooling in order to explain this persistent disadvantage of the titular language. We use standardized test scores from the Programme for International Student Assessment to qualify the differences in school outcomes and school resources. Differences in the quality of human capital seem to be the main explanation for the observed wage patterns.

The paper is structured as follows: Section 2 gives background information on the case-study of Kazakhstan. Section 3 reviews the data set and the methodology, while Section 4 provides empirical results for the returns to language skills. Section 5 discusses possible explanations for the observed language-wage pattern. Section 6 concludes.

2. Background

Kazakhstan is an ideal laboratory to study the economic effect of bilingualism. On the one hand, two languages (Kazakh and Russian) are widely spoken throughout the country. On the other hand, the government initiated a shift in the official state language (from Russian to Kazakh) in the year 1997 so that the importance of the Kazakh language has been growing in many life domains. The "Law on Languages of the Republic of Kazakhstan" (1997) states that every citizen should be able to master the Kazakh language. In order to reach this objective, the government supports the Kazakh language in the public sector work environment, in schools and in the media. For instance, the amount of media content in Kazakh should be no less than the joint amount of media content in other languages. While Russian at first remained important as "language of interethnic communication", the implementation of the "State Program for the Development of Languages for 2001-2010" (2001/6) put greater emphasis on the Kazakh language. As we will illustrate, this process has been accompanied by an increasing fraction of pupils being educated in the Kazakh language. Nevertheless, the Russian language remains important in the Kazakhstani society, even beyond the realm of interethnic communication.

Russian is still predominantly used in business communication. So, the first research question relates to the economic returns to Russian vs. Kazakh before and after Kazakhstan's shift of state language. Furthermore, the coexistence of two widespread languages and the importance of Russian in the business sphere should make bilingualism a valuable asset in the labor market. As second research question we ask whether individuals who are proficient in both languages earn higher wages than individuals who speak only one language.

Historically, the Russian language gradually replaced Kazakh as the official language after Kazakhstan had become part of the Soviet Union. This process was accompanied by a shrinkage in the number of schools using Kazakh as language of instruction. Kazakh also slowly disappeared in familial communication, especially in urban areas. Hence, many ethnic Kazakhs grew up without Kazakh-language instruction and without proper Kazakh language skills.[1] Since independence in 1991, the situation has changed. The share of schools with Kazakh language of instruction grew slightly while the number of schools with Russian language of instruction dropped dramatically (see table 1).

Table 1: Number of schools by the language of instruction

Year	Russian	Kazakh	Mixed
1999-2000	2514	3474	2107
2005-2006	2100	3723	2125
2013-2014	1395	3817	2156

Source: National Statistical Agency of Kazakhstan

This apparent language shift becomes even stronger once we look at the number of pupils by language of instruction. As evident from table 2, the number of pupils educated in Kazakh remained relatively constant over the past 15 years whereas the number of pupils educated in Russian decrease by whopping 44 percent.

---

[1]Similar processes of language absorption have so far only been analyzed for immigrant groups (Rumbaut *et al.*, 2006).

Table 2: Number of pupils by the language of instruction

Year	Russian school[a]	Kazakh school[b]	Mixed school[c] Russian language	Mixed school[d] Kazakh language	Total[e] Russian language	Total[f] Kazakh language
1999-2000	834,538	1,080,460	606,963	612,223	1,441,501	1,692,683
2005-2006	580,190	1,039,775	522,034	560,431	1,102,224	1,600,206
2013-2014	350,496	1,090,664	454,946	526,939	805,442	1,617,603

Source: National Statistical Agency of Kazakhstan

a - number of pupils in schools with Russian as the only language of instruction

b - number of pupils in schools with Kazakh as the only language of instruction

c - number of pupils in mixed schools educated in Russian language

d - number of pupils in mixed schools educated in Kazakh language

e - total number of pupils educated in Russian language

f - total number of pupils educated in Kazakh language

In cities, where most of the enterprises are concentrated, the population share understanding Russian is 96.7% whereas the share of people understanding Kazakh is only 67.7%.[2] In rural areas the respective numbers are 91.4% and 82.0%. The difference between urban and rural areas is even more pronounced if we look at the population shares of those fully proficient in the language.[3] The share of population being fluent in Russian is 90.2% in urban and 77.9% in rural areas. The share of population being fluent in Kazakh is 54.3% in urban and 71.8% in rural areas.[4]

Most of the work on the economic effect of languages focuses on the language ability and earnings of immigrants in monolingual societies (see Chiswick and Miller, 1999; Koussoudji, 1988; McManus *et al.*, 1983). The prevailing evidence suggests that immigrants benefit from the proficiency in the language of the host country. However, whether the command of a second language yields a wage premium in the labor market is a different matter. In fact, the empirical evidence on the return to commanding a second language is mixed. For example, Shapiro and Stelcner (1981) find little return to bilingualism in Canada. However, in their later work Shapiro and Stelcner (1997) find positive return to English proficiency for the Francophones but no return to French for the Anglophones. Pendakur and Pendakur (1998) find higher earnings for bilinguals in Canada if the two languages are official languages (English and French). Knowledge of a second language other than English or French is associated with lower wages in most cases.

[2]2009 Census
[3]Can fluently read and write.
[4]2009 Census

Henley and Jones (2005) find positive return to bilingualism in Wales. Chiswick *et al.* (2000) find that bilinguals in Bolivia face a wage penalty compared to Spanish monolinguals. The cross-country analysis conducted by Williams (2011) reveals an interesting picture. In Belgium with its three official languages (Dutch, French and German) the return to the usage of a second langauge is positive if the second language is either German or Dutch. The return to the usage of French as a second language is insignificant. In Luxembourg—with the official languages Luxembourgish, French and German—the return to the usage of German or French as a second language is positive. In Switzerland, Grin and Sreddo (1998) find that speakers of Italian as a first language are disadvantaged in French- and German-speaking regions, whereas speaking German or French as a second language yields benefits.

Since the choice of bi- or multilingualism is endogenous, the literature has also turned to estimating the effect of a change in language of instruction on returns to schooling. Angrist and Lavy (1997) find a singificantly negative effect of giving up the widely established second language French at school in Morocco on earnings. Aspachs-Bracons *et al.* (2008) and Clots-Figueras and Masella (2013) illustrate that language of instruction at school has a causal effect on national identity in Catalonia.

Kazakhstan is an excellent case study for the analysis of bilingualism, language shift, and their effect on labor market outcomes. The country did not only experience a shift in the official state language, but has historically been characterized by regional ethnic variation. While the South has always been populated by a majority of Kazakhs, the North was dominated by ethnic Russians and Germans. There, the titular ethnicity gained demographic dominance only through the ethnically motivated emigration of Russians and Germans as well as with the help of the re-location of Kazakhstan's capital from the South (Almaty) to the North (Astana). We focus on returns to language skills in urban areas because this is where the communicative jobs are located (as opposed to agriculture) and where the language shift was strongly felt. More specifically, we will shed light on differences in the returns to languages between cities heavily engaged in older and declining industries (predominantly steel and coal in the north) versus cities dominated by more communication-based services and trading businesses. Kazakhstan is a dynamically emerging economy characterized by substantial economic growth accompanied by a sectoral re-allocation from mining and heavy industry towards energy and services. The annual real growth rate averaged 7.1% during the 2003-2010 period (UN, Development Policy and Analysis Division), and the per capita GDP PPP equals 13,000 USD (2011; IMF data), which is comparable to Brazil, Iran or Venezuela.

3. DATA AND METHODOLOGY

In order to analyse the effect of language knowledge and bilingualism on urban wages, we exploit two surveys: First, we use data from the nationally representative 1996 Living Standard Measurement Survey (LSMS) conducted by the Kazakhstan Statistical Committee in collaboration with the World Bank. The data set comprises individual level demographic information, language ability as well as employment status and wages. The overall sample contains 1,880 working-age individuals in urban areas (aged 21-59). Since some of the cities were labelled with wrong city names in the data set, we cannot distinguish exact locations and restrict our sample to large cities with at least 100,000 inhabitants in general. Second, we employ a novel data set that was collected by the Research Institute CIOM in cooperation with the Institute for East European Studies Regensburg, Germany, in 2010, using a multi-stage random sample (Dietz and Gatskova, 2011). The survey covers four major urban areas of Kazakhstan with more than 100,000 inhabitants: The old capital Almaty, the new capital Astana as well as Pavlodar and Karaganda. While the former two cities are important political and economic centers, the latter two are characterized by (heavy) industry. The data contain rich information on labor market relations at the date of the interview as well as in retrospect and provide a number of measures relating to language ability in Kazakh, Russian as well as further languages. The overall sample contains 2,000 individuals in the working age.

In both data sets we drop individuals who classify themselves as a native speaker of a third language (Ukrainian, German etc.)—less than 1% of the samples—in order to fully concentrate on the effect of bilingualism. We constrain our data set to include only individuals working fulltime. Moreover, we exclude individuals with missing information on any of the variables. Given the small share of urban population who do not speak Russian we exclude Kazakh monolinguals from the analysis.[5] This leaves us with final samples of 945 observations (1996) and 1,211 observations (2010), respectively. The two most important variables for our empirical estimation are wages and language abilities. Wages are measured as the natural logarithm of the monthly hourly net wage in the current job. For language ability we construct several dummies indicating the ability to speak Russian or Kazakh. Despite the fact that our samples reflect the ethnic composition of urban Kazakhstan quite well (39.6% Kazakh in 1996 and 42.7% Kazakh in 2010), 79.4% (1996) and 66.8% (2010) of all respondents report that they speak Russian at home. We define Kazakh fluency to be a binary variable equaling one if the person reports to be

---

[5]In our sample the share of Kazakh monolinguals is tiny. Thereby we lose 14 (1996) and 15 (2010) observations, respectively.

fluent both in spoken and written Kazakh. The Kazakh fluency variable is constructed from the following items of the questionnaire: 1) how well do you know Kazakh (speak)? and 2) how well do you know Kazakh (write)? Both items contain six possible answers ranging from "not at all" to "fluent". We also generate a variable for bilingualism for individuals claiming to speak/write Russian and Kazakh fluently.

To test the effect of language skills and bilingualism on wages with these two data sets we estimate the following year-specific equation: $w_i = \alpha \boldsymbol{Lang_i} + \beta \boldsymbol{X_i} + \epsilon_i$, where w is the logarithm of the wage rate, $\boldsymbol{Lang}$ is the matrix containing the language dummies, $\boldsymbol{X}$ is the matrix of socio-demographic characteristics like age, sex, education, occupation (white/blue collar) and industry information like economic sector dummies and enterprise legal form, and ϵ is the error term such that $\epsilon \sim N(0, \sigma^2)$. We account for heteroskedasticity by using the Huber-White-Sandwich estimator.

Our main explanation for the observed language-wage pattern will be illustrated using data from the Programme for International Student Assessment (PISA) conducted by the OECD; PISA was implemented for the first time in 2009 in Kazakhstan. We use the standardized test scores in reading, mathematics and science which are generated from two-hourly tests sat by 15 year-old pupils. We then combine these scholastic achievements with background information regarding the language of instruction as provided by the schools' principals.

4. Empirical Results

In the first step of our analysis, we investigate the economic effects of Kazakh language ability and bilingualism on earnings in 1996—some time before the official state language was switched from Russian to Kazakh. In the first regression we consider two groups of workers: Russian monolinguals and bilinguals, the latter being seperated into bilinguals with Russian as first spoken language, and bilinguals with Kazakh as first spoken language. The reference category are monolingual workers. Furthermore, we seperately focus on individuals who claim to speak Kazakh fluently. Table 3 reports results from this analysis: Bilingual speakers and Kazakh speakers had a highly significant wage penalty shortly after independence when Russian was still the dominant language. Importantly, bilinguals were much worse off if their first language was Kazakh, while the additional knowledge of Kazakh yielded an economic premium for those who spoke Russian as a first language.

Table 3: Bilinguality and wages in 1996, OLS estimates

variable	Urban Kazakhstan coef		Urban Kazakhstan coef		Urban Kazakhstan coef	
female	-0.53	***	-0.53	***	-0.53	***
high skilled	0.27	***	0.27	***	0.28	***
low skilled	-0.16		-0.18		-0.16	
bilingual	-0.18	***				
bilingual (Russian first)			0.16	*		
bilingual (Kazakh first)			-0.25	***		
fluent in Kazakh					-0.27	***
R^2	0.31		0.32		0.31	
N	945		945		945	

* significant at 10% level, ** at 5% level, *** at 1% level
The data are taken from the LSMS Kazakhstan survey 1996, collected by Goskomstat and the World Bank.

Robust standard errors. The regression also includes age, age squared, log of working hours, a dummy for supervisory tasks, a dummy for skill match and enterprise legal form (state, collective or private) and oblast dummies.

In the next step we turn to the year 2010, when Kazakh had become the official state language (table 4). Recall that the model is no longer estimated for all large cities but for four identifiable metropolitan areas: Almaty, Astana, Karaganda, and Pavlodar. The effect of socio-demographic variables on wages is unsurprising. Women earn 20-30% less than men, down from more than 50% in the mid 1990s. The highly skilled have a 20-30% earnings premium over the medium skilled.[6] Low skilled workers earn 11% less than medium skilled workers in Almaty and almost 20% less in Karaganda and Pavlodar albeit the effect is not statistically significant for Almaty and Karaganda.

Even after the language shift and an array of policies promoting the use of the Kazakh language, we still find a strong negative effect of bilingualism on wages in Kazakhstan's two biggest cities Almaty and Astana. In Almaty the wage penalty associated with bilingualism is in the magnitude of 20%. In Astana the penalty for bilinguals with Russian as the first spoken language is almost 30%. The penalty for bilinguals with Kazakh as the first spoken language is roughly 10%. Different from 1996, bilingualism is now economically detrimental even for workers with Russian as first spoken language. Bilingualism yields no significantly negative wage premium in the two industrial cities which are dominated by Russian speakers (Karaganda and Pavlodar).

---

[6]Pavlodar is the exception, where we do not observe any difference in wages between high and medium skilled.

Table 4: Spoken languages and wages, OLS estimates

variable	Almaty coef		Astana coef		Karaganda coef		Pavlodar coef	
gender: female	-0.16	**	-0.25	***	-0.32	***	-0.31	***
high skilled	0.31	***	0.20	***	0.23	***	0.08	
low skilled	-0.11		-0.01		-0.18		-0.18	***
bilingual (Russian first)	-0.23	*	-0.29	*	0.08		-0.01	
bilingual (Kazakh first)	-0.19	***	-0.11	**	0.07		-0.03	
R^2	0.30		0.26		0.12		0.33	
N	287		371		275		278	

* significant at 10% level, ** at 5% level, *** at 1% level
Robust standard errors. The regression also includes age, age squared, log of working hours, enterprise legal form
(state or private), white collar and economic sectors as additional variables

In the next stage of the analysis we estimate the effect of Kazakh language fluency on wages. The results are presented in table 5. The estimates indicate that the knowledge of the second language (Kazakh) lowers wages by 17 log points (or almost 19%) for Russian speaking workers in Almaty and 14 log points (or 15%) in Astana. No significant effect of the second language is observed in Karaganda and Pavlodar. Interestingly, the estimated wage discounts have decreased from 1996 (all cities) until 2010 (in Almaty and Astana).

Table 5: Fluency in official state language and wages, OLS estimates

variable	Almaty coef		Astana coef		Karaganda coef		Pavlodar coef	
sex	-0.17	**	-0.25	***	-0.24	**	-0.31	***
high skilled	0.37	***	0.22	***	0.23	***	0.08	
low skilled	-0.04		0.02		-0.19		-0.18	***
fluent in Kazakh (speak and write)	-0.17	*	-0.14	***	0.07		-0.05	
R^2	0.28		0.27		0.08		0.32	
N	248		361		281		282	

* significant at 10% level, ** at 5% level, *** at 1% level
Robust standard errors. The regression also includes age, age squared, enterprise legal form
(state or private), white collar and economic sectors as additional variables

Russian plays a significant role for business relations and trade. Quite differently, Kazakh has replaced the Russian language in the public administration and wide parts of the public sector (health, education). Since the former sectors have been growing much stronger than the state sector in Kazakhstan one might suspect that the differential wage effect between Russian

and Kazakh speakers simply reflects the selection of ethnicities into specific segments of the labor market. Since we control for economic sectors, this leads to a human capital explanation: The second language (in our case Kazakh) may be only rewarded in economic sectors where Kazakh is a scarce input. To test this hypothesis we interact the individual language ability with the fraction of Kazakh speakers in a person's sector (or more accurately with a dummy indicating that a person is employed in a sector which contains at least as many fluent Kazakh speakers as the median sector). The existence of economic rewards for scarce language knowledge would be indicated by a significantly negative effect on the interaction term, but we never find this empirically (see Table A.1 and Table A.2 in the Appendix). Still, sectors that employ more Kazakh speakers tend to pay lower wages on average; however, this effect is only highly significant in Astana and Almaty. Also, stratifying the sample into blue vs. white collar workers as well as according to services vs. trade-related jobs does not change the result that bilingual individuals suffer from a wage discount (These estimates are imprecise due to the small sample size and can be obtained from the authors upon request). Hence, Kazakh language fluency earns a negative reward despite the gradual "Kazakhization" of the country's elites. The occupational structure between ethnic Russians and ethnic Kazakhs (irrespective of language use) is skewed in our data set: Kazakhs are overrepresented in managerial and executive occupations, while Russians are overrepresented in machinery, electrical and mining occupations.

5. DISCUSSION

We have investigated a series of explanations for the surprising persistence of the wage penalty for Kazakh in the 1990s and 2000s (Aldashev and Danzer, 2015): Particularily, we test whether the negative effect of Kazakh language knowledge is driven by discrimination (in the interview process), by the cost of learning a second language in dependence of the local language composition, or by reverse causality in the sense that individuals with poor labor market prospects learn a second language in order to improve their market value. First, discrimination running against the majority population is hardly plausible and we do not find evidence that the result are driven by Kazakh speakers who answer the survey in Russian (see Table A.3, Table A.4 and Table A.5 in the Appendix). Second, the cost of learning the Kazakh language would be much higher in cities which are dominated by Russian speakers (Lazear, 1999), hence implying poorer Kazakh language skills there. Consequently, the bilingualism penalty should be higher in cities like Karaganda or Pavlodar; yet, we find the opposite rendering the cost explanation implausi-

ble. Finally, we tackle the problem of reverse causality with two-stage least squares estimation using as instrumental variable the number of years an individual has spent in urban areas that are dominated by ethnic Kazakhs. This choice is based on findings by linguists that familiarity with Kazakh primarily depends on the time spent in Kazakh dominated areas (Fierman, 2012). Since our instrumental variables estimates are very close to the OLS estimates we conclude that reverse causality cannot explain the wage penalty of Kazakh language knowledge (Table A.6; for technical details see Aldashev and Danzer (2015)).

In this paper we suggest that the wage differentials stem from unobserved human capital differences between Russian monolinguals and bilinguals. We illustrate this in two steps: Initially, we show substantial and increasing quality differences between schools with Kazakh and Russian language of instruction. Then, we provide evidence that the language specific differences in the endowment with schooling inputs translate into differences in scholastic achievement across pupils with different language of instruction.

Despite Russian having been the only official state language during Soviet times, almost 30% of pupils in the academic year 1988-89 were taught in schools with Kazakh as language of instruction (Fierman, 2006). Virtually all Russian monolingual workers in our sample attended schools with Russian language of instruction while the group of bilinguals is likely to be comprised of two groups: those who attended schools with Russian language of instruction and those who attended schools with Kazakh language of instruction. Possibly, differences in schooling quality could lead to differences in acquired human capital and, hence, different rewards in the labor market.

Schools with Russian language of instruction have traditionally provided higher quality education than schools with Kazakh language of instruction. Examples include a wider choice of subjects, better textbook availability as well as teacher supply in Russian schools (Silova, 2002). Even when lectured in Kazakh, textbooks continue to be written in Russian—especially in technical subjects (Fierman, 2006, p. 113). Kazakh schools have experienced greater scarcity of instruction materials. We illustrate this point using comparable measures of schooling inputs at two points in time. Together with the aforementioned LSMS 1996 household survey, a community survey interviewed municipal school administration experts regarding the language of instruction and availability of school resources at each single school. And then in 2009, Kazakhstan participated for the first time in the OECD's Programme for International Student Assessment (PISA), in which school principals responded to very similar questions on the en-

dowment with school resources (for the exact wording of questions in both surveys, we refer the interested reader to the note of figure 1). In 1996, Kazakh speaking schools were less well equipped with textbooks and high quality teachers than Russian speaking schools, despite the fact that classes in Kazakh schools were smaller (as indicated by the positive gap of the student-teacher ratio in figure 1). The gap was modest in terms of teacher quality and motivation, but textbook shortages were serious. While relative class sizes have not changed at all until 2010, relative textbook shortages have become even more pronounced in Kazakh schools. In essence, the shift in school language has taken place for oral instructions only—without adequate modernization of textbook resources. More importantly, Kazakh schools face even greater difficulties to attract qualified and motivated teachers in 2010. This result mirrors the problem of hiring sufficiently large numbers of teachers who are able to teach in Kazakh. As a consequence, the expansion of Kazakh schools led to the lowering of educational requirements for teachers (Silova, 2002).
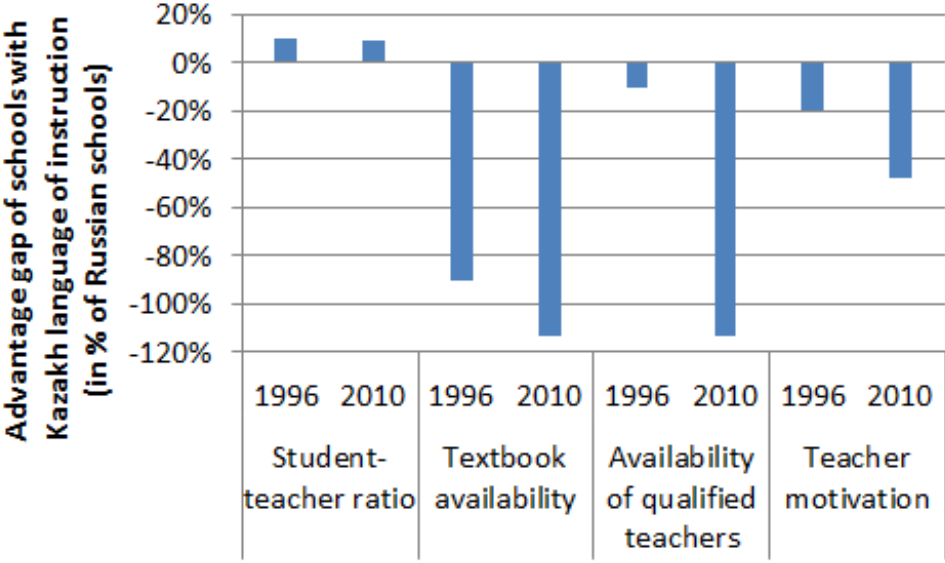

Figure 1: Qualitative disadvantages of Kazakh schools

The bars report the advantage (for positive values) or disadvantage (for negative values) of Kazakh-speaking schools, as percentage of Russian schools. Minus 20% indicates that Kazakh schools perform 20% poorer in terms of teacher motivation than Russian schools. Measures are not perfectly comparable: LSMS 1996 (PISA 2010) asks for textbook availability (teaching material availability), shortage of qualified teachers (shortage of qualified science, maths, English teachers), problems with teacher motivation (problems with teacher absenteeism, low teacher expectations and staff resisting change).
Source: Community LSMS 1996 and School questionnaire PISA 2010.

Differences in the quality of human capital across Kazakh vs. Russian schools seem to be a plausible explanation for the persistence of wage penalties, especially since the qualitative differences across school types have a long tradition and have survived the change in state language. Albeit suggestive, the significant endowment gaps between schools with Kazakh vs. Russian as language of instruction cannot prove that the observed wage differentials stem from quality differences in human capital. To shed more light on individual-level quality differences between schools with different languages of instruction we analyze pupil performance in the Programme for International Student Assessment (PISA) 2009 data. We restrict our data set to include only four cities: Almaty, Astana, Karaganda, and Pavlodar. The results in table 6 reveal that students perform better in schools with Russian language of instruction (column RUS). Albeit test scores for the year 1996 do not exist, all available qualitative evidence suggests that students already performed better in Russian-language schools during the 1990s (Silova, 2002; Fierman, 2006).

Table 6: PISA scores: Kazakh vs. Russian schools

	Almaty		Astana		Karaganda		Pavlodar	
	KAZ	RUS	KAZ	RUS	KAZ	RUS	KAZ	RUS
math	391.72	459.48	459.28	480.80	404.56	449.86	412.25	470.45
reading	360.52	475.62	423.37	501.70	384.78	452.96	394.00	455.40
science	381.87	477.44	438.54	466.97	387.68	459.36	414.75	446.53

Table 7: PISA scores: Bilinguals vs. monolinguals

	Almaty				Astana			
	I	II	III	IV	I	II	III	IV
math	392.97	462.51	439.53	351.11	456.06	483.45	519.97	500.14
reading	361.08	479.05	455.05	349.80	418.58	505.78	526.35	479.48
science	381.84	483.94	430.25	351.03	435.37	472.80	484.74	478.34
	Karaganda				Pavlodar			
	I	II	III	IV	I	II	III	IV
math	403.40	453.53	443.12	436.25	417.84	468.49	483.22	391.74
reading	383.79	457.21	452.78	411.85	397.04	456.68	446.38	382.86
science	386.84	466.24	435.07	410.71	426.34	449.55	423.94	372.23

I - language at home is Kazakh, language of instruction is Kazakh
II - language at home is Russian, language of instruction is Russian
III - language at home is Kazakh, language of instruction is Russian
IV - language at home is Russian, language of instruction is Kazakh

The PISA scores allow disentangling performance differences even further: PISA data provide the information on language spoken at home and the language of instruction at school. We would like to understand whether there are differences in academic performance between bilinguals who attended schools with Russian language of instruction and those attending schools with Kazakh language of instruction. Table 7 presents the test scores for four groups of students. Group I is comprised of students who report to speak Kazakh at home and attend schools with Kazakh language of instruction. Group II is comprised of students who report to speak Russian at home and attend schools with Russian language of instruction. Group III is comprised of students who report to speak Kazakh at home and attend schools with Russian language of instruction. Group IV is comprised of students who report to speak Russian at home and attend schools with Kazakh language of instruction. Given that the data neither provide information on the second spoken language nor on language proficiency we are unable to strictly identify monolingual and bilingual students. Therefore, group I is our best approximation to Kazakh monolinguals (given data restriction). Group II are Russian monolinguals. Groups III and IV are thus bilingual students.[7] The results reveal the following pattern: the highest scores are achieved by groups II and III and the test scores of group III are generally higher than the scores of group IV.

This result suggests that students attending schools with Russian language of instruction outperform their peers being taught in Kazakh.[8] If better test performance implies higher skills and if workers are paid their marginal product we would expect students with Russian as language of instruction to command higher wages in the future. Given that the test scores of group III are higher than the ones of group IV we conclude that among bilingual students those with Russian language of instruction perform best. Furthermore, as the test scores of groups II and III are comparable, human capital differences between Russian monolinguals and bilinguals instructed in Russian seem negligible. This suggests that for academic success language of instruction is more important than the language spoken at home. These results are not fully in line with our previous result suggesting that bilinguals earn less than Russian monolinguals. Still, our wage models lack the information on the language of instruction in school and thus the given group of bilinguals in our sample contains bilinguals from Russian and from Kazakh

---

[7]Given that for groups III and IV languages at home and in school are different these two groups are most likely to be the genuine bilinguals. However, groups I and II might still contain bilingual students as using the same language at home and in school does not rule out the possibility that they are also fluent in some other language.

[8]Note, that demand for education cannot explain the differences between Russian and Kazakh speakers. In fact, the 1996 LSMS survey reveals that parents of children who spoke Kazakh at home actually had slightly higher educational aspirations than parents of children who spoke Russian at home. Also, expenditures on education per pupil did not differ between households speaking Russian and Kazakh.

schools. The larger the share of graduates from Russian schools within the sample of bilinguals, the smaller should be the wage differential between the Russian monolinguals and bilinguals. Hence, the reason why the wage penalty for bilingualism in Karaganda and Pavlodar is insignificant can be found in the fact that most of the bilinguals were taught at Russian schools in those Russian-dominated cities.

Using the number of school graduates (grades 10 and 11; reported by the National Statistical Agency of Kazakhstan) across regions we calculate the ratio of Kazakh to Russian graduates for the year 2002.[9] For Karaganda and Pavlodar the ratios are 39 and 41 percent respectively. For Almaty the ratio is 66 percent and for Astana 92 percent.[10]

Since within the sample of bilinguals the share of graduates from schools with Russian language of instruction is large in Karaganda and Pavlodar, the effect of Kazakh language proficiency turns insignificant. On the contrary, the share of graduates from Russian schools is small in Almaty and Astana. Therefore we observe the significant negative effect of Kazakh language proficiency on wages.

6. CONCLUSIONS

This paper analyzes the effect of bilingualism and fluency of a newly introduced state language on earnings. We exploit two extensive data sources from two different periods in Kazakhstan's nation building process: One survey was implemented before Kazakh was made the sole official state language which replaced Russian in public life and school, the other survey was collected 13 years after the official language shift. Individuals reporting to speak two languages (Russian and Kazakh) had and still have lower earnings than monolingual Russian speakers. Despite significant political efforts to promote Kazakh in several life domains, fluency in Kazakh (beside fluency in Russian which is the language commonly understood by everyone) yields a negative economic return. This is forcefully exemplified by the fact that the economic disadvantage of the Kazakh language (implying bilingualism) has remained very stable between 1996 and 2010, although the official state language was switched during this period. While policies that enact a new official state language do not necessarily benefit those who can switch between

---

[9]The number of graduates from schools where Kazakh is the only language of instruction divided by the number of graduates from schools where Russian is the only language of instruction.

[10]If we also include the mixed schools the share of graduates who were taught in Kazakh is roughly 37 percent for Almaty and Astana, 33 percent for Karaganda and 29 percent for Pavlodar. Mixed schools comprise groups being taught in Russian and groups being taught in Kazakh.

languages, the actual differences in schooling qualities matter significantly. An important policy implication of our research is that promoting the Kazakh language could be better achieved through improving the quality of schools with Kazakh as language of instruction rather than through general language legislation.

7. ACKNOWLEDGEMENTS

References

ALDASHEV, A. and DANZER, A. M. (2015). Bilingualism and economic performance. *Gazzola, Michele and Bengt-Arne Wickström (eds.) The Economics of Language Policy*, p. forthcoming.

ANGRIST, J. D. and LAVY, V. (1997). The effect of a change in language of instruction on the returns to schooling in Morocco. *Journal of Labor Economics*, **15** (1), S48–76.

ASPACHS-BRACONS, O., COSTA-FONT, J., CLOTS-FIGUERAS, I. and MASELLA, P. (2008). Compulsory language educational policies and identity formation. *Journal of the European Economic Association*, **6** (2/3), 434–444.

CHISWICK, B. and MILLER, P. (1999). Language skills and earnings among legalized aliens. *Journal of Population Economics*, **12** (1), 63–89.

—, PATRINOS, H. and HURST, M. (2000). Indigenous language skills and the labor market in a developing economy: Bolivia. *Economic Development and Cultural Change*, **48** (2), 349–367.

CLOTS-FIGUERAS, I. and MASELLA, P. (2013). Education, language and identity. *Economic Journal*, **123** (570), F332F357.

DIETZ, B. and GATSKOVA, K. (2011). *Migration and Remittances in Kazakhstan: Survey Technical Report*. Discussion Paper no. 303, Osteuropa-Institut Regensburg.

FIERMAN, W. (2006). Language and education in post-soviet Kazakhstan: Kazakh-medium instruction in urban schools. *The Russian Review*, **65** (1), 98–116.

— (2012). Reversing language shift in Kazakhstan. *Schiffman, Harold (ed.) Language Policy and Language Conflict in Afghanistan and Its Neighbors*, pp. 121–175.

GRIN, F. and SREDDO, C. (1998). Language-based earnings differentials on the Swiss labour market: Is Italian a liability? *International Journal of Manpower*, **19** (7), 520–532.

HENLEY, A. and JONES, R. E. (2005). Earnings and linguistic proficiency in a bilingual economy. *Manchester School*, **73** (3), 300–320.

KOUSSOUDJI, S. (1988). English language ability and the labor market opportunities of Hispanic and East Asian immigrant men. *Journal of Labor Economics*, **6** (2), 205–228.

LAZEAR, E. P. (1999). Culture and language. *Journal of Political Economy*, **107** (6), S95–S126.

McManus, W., Gould, W. and Welch, F. (1983). Earnings of Hispanic men: the role of English language proficiency. *Journal of Labor Economics*, **1** (2), 101–130.

Pendakur, K. and Pendakur, R. (1998). Speak and ye shall receive: Language knowledge as human capital. *Breton, A. (ed.) Economic Approaches to Language and Bilingualism*, pp. 89–121.

Rumbaut, R. G., Massey, D. S. and Bean, F. D. (2006). Linguistic life expectancies: Immigrant language retention in Southern California. *Population and Development Review*, **32** (3), 447–460.

Shapiro, D. and Stelcner, M. (1981). Male-female earnings differentials and the role of language in Canada, Ontario, and Quebec, 1970. *Canadian Journal of Economics*, **14** (2), 341–348.

— and — (1997). Language and earnings in Quebec: trends over twenty years, 1970-1990. *Canadian Public Policy*, **23** (2), 115–140.

Silova, I. (2002). *The right to quality education: Creating child-friendly schools in Central Asia*. Almaty, Kazakhstan: UNICEF CARK.

Williams, D. (2011). Multiple language usage and earnings in Western Europe. *International Journal of Manpower*, **32** (4), 372–393.

A Appendix

Table A.1: Bilingualism, Ethnic Sector Composition and Wages, OLS estimates

	Almaty and Astana		Karaganda and Pavlodar	
variable	coef		coef	
sex	-0.19	***	-0.29	***
high skilled	0.29	***	0.11	**
low skilled	-0.06		-0.19	**
working in Kazakh sector	-0.22	**	-0.13	
bilingual (Russian first)	-0.25	**	0.04	
bilingual (Russian first)×Kazakh sector	0.07		-0.06	
bilingual (Kazakh first)	-0.12	**	-0.03	
bilingual (Kazakh first)×Kazakh sector	0.06		0.09	
R^2	0.27		0.15	
N	566		526	

* significant at 10% level, ** at 5% level, *** at 1% level
Robust standard errors. The regression also includes age, age squared, log of working hours, enterprise legal form (state or private), white collar and economic sectors as additional variables

Table A.2: Fluency in Official State Language, Ethnic Sector Composition and Wages, OLS estimates

	Almaty		Astana		Karaganda		Pavlodar	
variable	coef		coef		coef		coef	
sex	-0.16	**	-0.24	***	-0.22	**	-0.31	***
high skilled	0.38	***	0.24	***	0.24	***	0.08	
low skilled	-0.03		0.01		-0.17		-0.19	***
fluent in Kazakh (speak and write)	-0.23	*	-0.18	***	0.06		-0.05	
working in Kazakh sector	-0.21		-0.23	**	-0.21		-0.02	
fluent in Kazakh×Kazakh sector	0.12		0.10		0.06		-0.00	
R^2	0.29		0.29		0.08		0.32	
N	248		361		281		282	

* significant at 10% level, ** at 5% level, *** at 1% level
Robust standard errors. The regression also includes age, age squared, log of working hours, enterprise legal form (state or private), white collar and economic sectors as additional variables

Table A.3: Fluency in Official State Language and Wages, restricted sample

variable	Almaty coef		Astana coef		Karaganda coef		Pavlodar coef	
sex	-0.19	***	-0.27	***	-0.23	**	-0.34	***
high skilled	0.36	***	0.21	***	0.22	***	0.06	
low skilled	-0.00		0.03		-0.20		-0.17	***
fluent in Kazakh (speak and write)	-0.15	*	-0.15	***	0.09		-0.06	
R^2	0.27		0.27		0.07		0.30	
N	243		353		278		274	

Excluding respondents who report to be not fluent in Kazakh but chose the Kazakh questionnaire

* significant at 10% level, ** at 5% level, *** at 1% level
Robust standard errors. The regression also includes age, age squared, enterprise legal form
(state or private), white collar and economic sectors as additional variables

Table A.4: Fluency in Official State Language and Wages, restricted sample

variable	Almaty coef		Astana coef		Karaganda coef		Pavlodar coef	
sex	-0.18	**	-0.26	***	-0.23	**	-0.33	***
high skilled	0.38	***	0.21	***	0.22	***	0.06	
low skilled	-0.00		0.01		-0.20		-0.19	***
fluent in Kazakh (speak and write)	-0.16	*	-0.14	***	0.09		-0.06	
R^2	0.28		0.26		0.07		0.30	
N	245		361		278		276	

Excluding respondents whose mother tongue is Russian and who report to be not fluent in Kazakh but chose

the Kazakh questionnaire

* significant at 10% level, ** at 5% level, *** at 1% level
Robust standard errors. The regression also includes age, age squared, enterprise legal form
(state or private), white collar and economic sectors as additional variables

Table A.5: Fluency in Official State Language and Wages, restricted sample

variable	Almaty coef		Astana coef		Karaganda coef		Pavlodar coef	
sex	-0.24	***	-0.37	***	-0.17		-0.35	***
high skilled	0.33	***	0.19	***	0.18	*	0.07	
low skilled	0.19	*	-0.08		-0.26	*	-0.17	**
fluent in Kazakh (speak and write)	-0.21	**	-0.27	***	0.17		–	
R^2	0.40		0.33		0.05		0.30	
N	159		204		217		189	

Excluding (1) respondents whose mother tongue is Russian but chose the Kazakh questionnaire and (2)

respondents whose mother tongue is Kazakh but chose the Russian questionnaire

* significant at 10% level, ** at 5% level, *** at 1% level
Robust standard errors. The regression also includes age, age squared, enterprise legal form
(state or private), white collar and economic sectors as additional variables

Table A.6: Fluency in Kazakh and Wages, Second stage 2SLS estimates

variable	Almaty and Astana coef		Karaganda and Pavlodar coef	
Female	-0.23	***	-0.28	***
High skilled	0.27	***	0.12	**
Low skilled	-0.09		-0.18	**
White collar	0.17	***	0.08	
Fluent in Kazakh	-0.27	**	0.16	
F-test, first stage	16.4		21.6	
Instrument	0.05	***	0.07	***
R^2	0.29		0.10	
N	715		593	

* significant at 10% level, ** at 5% level, *** at 1% level
Robust standard errors. The regression also includes age, age squared, enterprise legal form
(state or private) and economic sectors as additional variables.