Available online at www.sciencedirect.com
ScienceDirect
Economics of Education Review 26 (2007) 349–360
Economics of Education Review
www.elsevier.com/locate/econedurev

Language skills and earnings: Evidence from a pre-industrial economy in the Bolivian Amazon

Ricardo Godoy[a,*], Victoria Reyes-García[a], Craig Seyfried[a], Tomás Huanca[a], William R. Leonard[b], Thomas McDade[b], Susan Tanner[b], Vincent Vadez[a]

[a]*Heller School for Social Policy and Management, Brandeis University, Waltham, MA 02454-9110, USA*
[b]*Department of Anthropology, Northwestern University, Evanston, IL 60208, USA*

Received 12 September 2005; accepted 5 December 2005

Abstract

Among linguistic minorities of industrial nations proficiency speaking the dominant national language increases earnings and wages, but do similar results apply to autarkic linguistic minorities of developing nations? We contribute to studies of the returns to language skills by applying the human-capital approach to a society of hunters, gatherers, and farmers in the Bolivian Amazon (Tsimane'). We use a panel consisting of five consecutive quarters to: (a) estimate the returns to language skills while controlling for schooling, math and writing skills, and other confounders and (b) explore the paths through which language skills might affect earnings. Fluent speakers of Spanish and the local language earned 36.9–46.9% more than monolingual speakers of the local language. Moderate fluency in Spanish bore no strong association with earnings. Spanish-Tsimane' bilingualism bore a positive association with earnings partly because bilingualism bore a positive association with credit access, use of modern production technologies, and labor productivity.

JEL classification: C23; J510; J24; J610; J15; F22

Keywords: Economic development: Human capital; Rate of return; Economic impact

1. Introduction

Among linguistic minorities of industrial nations proficiency speaking the dominant national language increases earnings and wages (Chiswick, 1998; Dustmann & Fabbri, 2003; McManus, 1998), but do results also apply to linguistic minorities of developing nations? Theory and empirical studies yield unclear predictions. Linguistic minorities in developing nations resemble immigrants of industrial nations because they often live in enclaves and lack the human capital to compete with success in the labor market. Proficiency speaking the national language should therefore yield private market benefits. Indeed, in a study of indigenous peoples in Bolivian cities, Chiswick, Patrinos, and Hurst (2000) found that monolingual speakers of Bolivia's national language, Spanish, earned 23–28% more than bilingual speakers of Spanish and an American Indian

*Corresponding author. Tel.: +1 781 7362784/7362770; fax: +1 781 7362774.
E-mail address: rgodoy@brandeis.edu (R. Godoy).

0272-7757/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.econedurev.2005.12.001

language. However, in rural areas of developing nations with more autarkic populations who speak a local language, the private market benefits of proficiency in the national language are less clear because daily activities take place in the local language. Proficiency in the national language might confer prestige, but might not produce significant private market returns if there is weak demand for proficiency in the national language and if contact with the national society is sporadic and tenuous rather than strong and continuous.

Indigenous populations in the tropical lowlands of Latin America provide an apt case study to explore whether proficiency speaking the national language among highly autarkic populations yields significant private market returns. In countries such as Bolivia, Peru, Guatemala, and Chile, a large share of the population speaks an American Indian language as their first tongue. Many studies have documented the economic disadvantages and discrimination of being Indian in Latin American nations with a significant share of an American Indian population, and the earnings premium that accrues to native peoples from investing in schooling, academic skills, and in other forms of modern human capital (Mc Ewan, 2004; Patrinos, 1994; Patrinos & Psacharopoulos, 1992; Psacharopoulos & Patrinos, 1994). Most such studies have come from indigenous populations in urban centers, where the returns to acquiring modern human capital are clear. To our knowledge, little of this research has taken place among highly autarkic rural indigenous populations.

Besides contributing to the literature on human capital, assessing the returns to proficiency speaking the national language among linguistic minorities of developing nations is important for reasons of public policy. As anthropologists have long taught, language, culture, and values intertwine (Maffi, 2001). Linguistic minorities in rural areas of developing nations hold much of the world's traditional ecological knowledge and cultures, and thus embody humanity's heritage and diversity. Large positive private market returns to proficiency in the dominant national language might provide incentives to linguistic minorities to acculturate into mainstream society and give up their traditional language and, in doing so also give up their traditional knowledge, culture, and values. If so, and if society values cultural diversity, then public policies might be needed to bring private and social interests into alignment.

In this article we assess whether proficiency speaking the national language by an ethnic minority in a developing nation bears a positive association with private market returns, and do so by following two steps. First we estimate the private market returns to proficiency speaking Spanish, the national language of Bolivia, among adults in a native Amazonian society of hunters, gatherers, and farmers in Bolivia, the Tsimane', who are in the early stages of continuous contact with the market economy and who are mainly monolingual in the Tsimane' language. Second, we identify the paths through which language skills might affect earnings.

2. Estimation strategy

We adapt the human-capital earnings function (Mincer, 1974) to estimate the parameters of the following expression for a panel consisting of five consecutive quarters:

$$\ln E_{ihvt=5} = \alpha + \beta L_{ihvt=1} + \gamma S_{ihvt=1} + \delta P_{ihvt} + \zeta H_{hvt=1} + \theta C_{vt} + \varepsilon_{ihvt}.$$

$\ln E_{ihvt=5}$ stands for the logarithm of earnings of adult i, household h, village v at follow up or during the fifth or last quarter ($t = 5$). $L_{ihvt=1}$ are two dummy variables for skills speaking the national language (Spanish)–either with difficulty (*Some Spanish*) or fluently (*Fluent Spanish*)–with monolingual speakers of Tsimane' as the excluded category. S stands for a vector of variables related to the modern human capital of the person. S includes math and writing skills, and the maximum school attainment of the person; the latter reflects other dimensions of schooling besides academic and language skills (e.g., discipline, patience). P is a vector of observed variables for the adult (e.g., age, sex), which directly affect earnings. H stands for household size. C includes dummy variables for villages to control for village characteristics, such as the size of the language enclave in the village (Lazear, 1999; McManus, 1998). ε is a random error term that reflects stochastic shocks. All explanatory variables refer to the first quarter or baseline. We use ordinary-least squares (OLS) with clustering of people by village because people are nested in villages.

The data and results likely contain biases from selection into the modern economy. In the early stage of the panel study (1999–2000) we studied two villages–one remote and one close to the market–in

detail. We surveyed the villages again in 2004 and found that attriters or people who had moved out of the research sites tended to have more schooling and higher earnings than those who remained in the sample. If so, then the sample used in this article might contain a selection bias because people with more modern human capital who earn higher earnings might have left the sample to move closer to market towns. Since we did not have identifying instruments to control for selection into the modern economy or for the endogeneity of Spanish fluency, our estimates may contain biases. In the balance of the article we speak of associations between language skills and earnings because we cannot infer causality from the information collected or from the analysis done.

3. The Tsimane' and schooling

The Tsimane' are a native Amazonian society of ~8000 people living in ~100 villages, department of Beni, Bolivia. They live along riverbanks and logging roads in villages of ~24 households. Permanent out-migration is rare; only 7.15% of the Tsimane' over 15 years of age live outside their territory (Instituto Nacional deEstadística, 2003). Subsistence centers on hunting, gathering, fishing, and shifting slash-and-burn farming. Tsimane' have lived in continuous contact with Westerners only since the 1940s and remain highly autarkic. Goods bought in the market account for only 2.70% of the total value of household consumption. Mean annual personal income from cash earnings and from the imputed value of farm and forest goods consumed from their own fields and forests reaches $332, or $0.90/day. Though highly autarkic, the Tsimane' need cash to buy food, tools, clothing, and school supplies. To earn cash, Tsimane' work as unskilled laborers in logging camps, cattle ranches, and in the homesteads of colonist farmers. They also sell and barter rice and thatch palm. Employers prefer to hire workers who can speak Spanish to facilitate communication.

The first recorded contact of Tsimane' with Westerners dates to the seventeenth century, but continual exposure to Westerners dates only to the 1940s when Protestant missionaries from the USA entered the department of Beni for the first time (Chicchón, 1992). The work of missionaries took off during the 1950s when the Bolivian Government gave missionaries the responsibility of schooling remote lowland native Amazonian populations such as the Tsimane' (Castro, 1997). The agreement lasted from 1954 until 1985. As part of the agreement, in 1955 missionaries set up a center in the town of Tumichuco, several days away from the Tsimane' territory to train Tsimane' to become bilingual school teachers and to translate the Bible into the Tsimane' language. Protestant missionaries offered scholarships to promising Tsimane' young men so they could attend Tumichuco for three months a year to work as informants for missionary linguists. In Tumichuco missionaries taught the Tsimane' academic and practical skills (e.g., modern hygiene) and the Scriptures so Tsimane' could proselytize in the Tsimane' language once they returned to their villages. After 27 years of operating in Tumichuco, missionaries transferred their training to the outskirts of the market town of San Borja (population ~19,000), next to the main Tsimane' territory. After receiving training, schooled Tsimane' returned to their villages, where they worked as lay missionaries and teachers using instructional materials in the Tsimane' language prepared by missionaries. In 1985, when the agreement with the Government of Bolivia ended, the Government of Bolivia took over the responsibility of schooling the Tsimane', which meant keeping Tsimane' as school teachers and paying their salary. To this day, missionaries produce the textbooks used in Tsimane' classrooms, run training seminars for Tsimane' teachers, and offer training courses in reading and writing twice a year for Tsimane' adults. Most of today's top Tsimane' political leaders received their training from Protestant missionaries.

At present about 40% of Tsimane' villages have a primary school covering the first five grades. No village has a middle school or a high school. Four villages close to the town of San Borja have an education program for adults where Tsimane' with primary schooling can earn a high-school degree by attending classes one week a month.

From the founding of the first village school to this day, the Tsimane' have been able to select the type of human capital (modern versus traditional) they want for themselves or for their offspring since school attendance is, in practice, voluntary. Because the area has yet to experience economic growth, demand for schooling and academic skills remains modest (Foster & Rosenzweig, 1996). Parents take out their children from school when they need them to do chores in the house or farm. Young adolescents often leave school to work for loggers,

ranchers, and migrant farmers. Tsimane' typically are 15 years of age by the time they complete the fifth grade, so they spend about 2 years for each school grade.

4. Data and variables

The information presented here comes from a panel study that started in 1999 and continues. We designed the panel to assess the effect of integration to the market and modernization on the culture, economy, and lifestyle of a highly autarkic society as it gains a stronger foothold in the market economy. For the empirical analysis we use only a part of the panel. In particular, we use information from five consecutive quarters (May 2002–August 2003). We estimate the association between proficiency speaking Spanish at baseline or during the first quarter (May 16, 2002–August 15, 2002) and earnings at follow up or during the fifth quarter (May 16, 2003–August 15, 2003). We do so to reduce biases from possible reverse causality that might arise if we used data from only one survey wave or from five contemporaneous measures of outcome and explanatory variables. The procedure should not introduce a bias because there is little reason to think that language skills would change significantly over only five quarters among our target subjects—adults or people over 16 years of age; variability in language skills over five quarters among adults would likely reflect measurement error rather than true change in language competence. This said, in the analysis of robustness we use contemporaneous measure of outcome and explanatory variables over the five quarters and find essentially the same results as when using earnings for the fifth quarter and lagged explanatory variables.

Most of the data presented comes from interviews, but to estimate the relation between language skills and labor productivity we did scans or spot observations, which we describe later. We interviewed all people over 16 years of age in 13 villages along the Maniqui river (304 women and 323 men; 257 households). We selected 13 villages at different distances from market towns to capture cross-sectional variation in market exposure. We limit the analysis to people over 16 years of age because younger people are less likely to work for wages or to sell goods since they still depend on their parents.

4.1. Dependent variable: Earnings at follow up (fifth quarter)

During quarterly interviews we asked people to list all sources of earnings for the 14 days before the day of the interview. Sources of earnings included sale of farm and forest goods, wage labor, and the value of all goods received in barter. We added the value of all monetary earnings to the value of goods received in barter to arrive at a total value of earnings and took the logarithm of the total. Since the recall period was short (2 weeks), we have reason to suspect low random measurement error from poor recall, whether recalling monetary earnings or barter transactions.

38.34% of women, 21.45% of men, and 29.98% of the total sample reported no earnings. Among women, the bulk of earnings came from sales (67%), followed by barter (19%), and wage labor (13%). Among men, most earnings came from wage labor (54%), followed by sales (37%), and barter (7%). We tested whether demographic attributes of the household, such as number of children, predicted entrance into the market economy, but not the intensity of earnings, and found no convincing identifying instrument.

4.2. Explanatory variable: proficiency speaking Spanish

During the interview, interviewers judged a person's ability to speak Spanish. We used answers to create two dummy variables: *Some Spanish* refers to people we classified as speaking Spanish with difficulty (1 = with difficulty; 0 = without difficulty) and *Fluent Spanish* refers to people we classified as speaking fluent Spanish (1 = fluent; 0 = not fluent). Monolingual speakers of Tsimane' is the excluded category. 98% of people said Tsimane' was their mother tongue, so the variables *Some Spanish* and *Fluent Spanish* reflect degrees of bilingualism in Spanish and Tsimane'. We found no monolingual speaker of Spanish.

4.3. Explanatory variables: other dimensions of human capital—math and writing

Math and writing skills should matter in market outcomes. For example, people with greater math ability will have greater skills in protecting themselves when selling or bartering goods with traders. People with writing skills should be able to write

market transactions, such as days worked, or loans owed to traders.

We gave tests to assess academic math and writing skills. The math tests had four questions, each of which required a respondent to add, subtract, multiply, or divide. We had several versions of equal difficulty of the math test and chose one at random for each person. We assigned a one to each correct answer, so total scores for the math test ranged from zero to four. Psychologists draw a distinctions between practical and academic skills (Grigorenko & Sternberg, 2001; Sternberg, Lipka, Newman, Wildfeuer, & Grigorenko, 2002; Sternberg et al., 2001). Academic skills refer to theoretical or passive knowledge of a person, whereas practical skills refer to the practical use of academic knowledge in daily life. The two need not overlap. People might do well in a driving test, but drive poorly. Sternberg et al. (2002) have shown that in rural areas children display math competence when asked to solve problems with meaning in their lives, not when asked to solve academic problems. Our math tests were abstract (e.g., "how much is 2 + 3?") rather than practical, so they might not have picked up true math competence relevant to the daily life of the Tsimane'. We later show a relatively high correlation (0.72) between math skills and years of schooling, suggesting that our measure of math skills likely measures material covered in the classroom. We assessed writing skills by asking people to sign their name in a white piece of paper under broad daylight during the interview. Interviewers coded answers to writing skills as follow: 0 = unable, 1 = with difficulty, 2 = well.

4.4. Controls

Control variables included age, sex, household size, days ill during the 14 days before the day of the interview, logarithm of body-mass index (BMI; weight in kilograms/physical stature in meters^2), and days worked in wage labor during the 14 days before the day of the interview. We next describe the rationale behind the use of the control variables.

We opted to use age instead of experience because of the recall bias from trying to remember the age at which a person first entered school. Also, the human-capital approach assumes a continuous commitment to the labor market (Chiswick & Miller, 1999). Those conditions do not apply here where people slip in and out of the labor market. The variable for days ill captures perceived illness. Elsewhere we describe the health and nutritional status of Tsimane' (Foster et al., 2005; Tanner, 2005; Burhop, 2003; Byron, 2003; McDade et al., 2003, 2005). Recent analysis of fecal and blood samples and anthropometric studies suggest that the most common illnesses relate to growth stunting, anemia, and parasitic infections. BMI is a canonical anthropometric indicator of short-run nutritional status. Only three people had a BMI below 18.5 (mild to moderate under-nutrition), only 17% of the sample was over-weight (BMI $\geqslant$ 25), and only two people were obese (BMI $\geqslant$ 30). The variable for days worked captures only the days worked in wage labor, not the days spent producing farm goods or extracting forest goods for sale, so it underestimates the true amount of labor that went into earnings. Table 1 contains definition and summary statistics of the variables used in the regressions.

4.5. Scans: labor productivity

To measure labor productivity we did scans. The scan is a standard method used in cultural anthropology and human ecology to measure the allocation of time in pre-industrial populations. In a classic scan, researchers walk the research site (e.g., village) and note what people are doing at the instant when they first see the person. We did a modified scan: we did a classic scan, but approached the person after noting their activity and asked them what they had done during the 24 h before we saw them. Thus, our scan contains elements of a classic scan and a quick survey.

During scans we visited each household in the village once/week on a day chosen at random. Visits took place during different blocks of time of the day (7–10 am, 10 am–1 pm, 1–4 pm, or 4–7 pm), also chosen at random. After arriving at the household, we recorded the activity of each person at the moment we first saw the person and the number of adults present in the household. We asked each person about their activities for the previous 24 h, and about the type and amount of forest plants, game, fish, and crops brought into the household for the 24 h preceding our meeting. We scanned people an average of six times each quarter (SD = 3.59). Since we asked people about their production for the previous 24 h, scans allow one to proxy labor productivity.

Table 1
Definition and summary statistic of variables used in regression analysis for Tsimane' people over 16 years of age

Name	Definition	N	Mean	Std Dev
Dependent variable (fifth quarter)				
Earnings	Earnings from wage labor and sale of goods plus value of goods received in barter for 14 days before the day of the interview. Summary statistics are in 1000 bolivianos. 1 US dollar = 7.45 bolivianos during fieldwork). In regression, earnings entered in logarithms	509	.088	.168
Explanatory variables (first quarter): human capital				
Spanish	Ability to speak Spanish; in regression, "with difficulty" or "Some" and "Fluent" included			
	Unable (1 = unable; 0 = able)	627	.382	.486
	With difficulty or "Some" (1 = with difficulty; 0 = without difficulty)	627	.287	.452
	Well or "Fluent" (1 = well; 0 = unable or with difficulty)	627	.330	.470
	Overall (0 = unable; 1 = with difficulty; 2 = well)	627	.947	.843
Schooling	Maximum school grade completed by the person interviewed	627	1.968	2.347
Math	Score in math test; 4 questions. Range of score 0 to 4	627	1.059	1.485
Writing	Ability to sign own name; in regression, only one variable, "Overall", included			
	Unable (1 = unable; 0 = able)	626	.565	.496
	With difficulty (1 = with difficulty; 0 = without difficulty)	626	.123	.328
	Well (1 = well; 0 = unable or with difficulty)	626	.311	.463
	Overall (0 = unable; 1 = with difficulty; 2 = well)	626	.746	.902
Explanatory variables (first quarter): controls				
Male	Sex of person (1 = male; 0 = female)	627	0.515	.500
Age	Age of person in years	627	34.631	15.518
Health	Self-reported person-days ill from three main ailments during 14 days before interview	520	4.495	6.647
Days worked	Number of days worked in wage labor during 14 days before the interview	600	1.330	3.174
Household size	Household size measured during baseline with number of people in the household at time of interview	257	6.350	2.951
BMI	Body-mass index (kg/m^2); in regression entered in logarithms	564	23.105	2.495

5. Results

5.1. Correlation between variables related to modern human capital

Proficiency speaking Spanish correlated positively with writing (0.56) and math (0.47) skills and with schooling (0.42). Schooling and writing skills had a correlation coefficient of +0.70 and schooling and math skills had a correlation coefficient of +0.72; these correlation coefficients were higher than the correlation coefficients between language skills and other variables related to modern human capital (+0.42 to +0.56) because people learned writing and math mainly in school, whereas they could have learned Spanish inside or outside school. The low association between schooling and Spanish also has to do with the teachers and with the educational material. Recall that teachers in Tsimane' schools are Tsimane' and teach in the Tsimane' language, with school material prepared by Protestant missionaries in the Tsimane' language, so years of schooling proxies weakly for acculturation into mainstream Bolivian society.

5.2. Covariates of proficiency speaking Spanish

We adapt the model of language fluency of Chiswick and Miller to identify the covariates of fluency in spoken Spanish (Chiswick & Miller, 1999). Chiswick and Miller view fluency in the national language as a function of economic incentives, exposure to the national language, and efficiency. In Table 2 we show the results of an ordered logit regression to identify the covariates of fluency in spoken Spanish. The outcome variable

Table 2
Covariates of fluency speaking Spanish among Tsimane' over 16 years of age during the first quarter ($n = 237$)

Explanatory variables	Coefficient and standard error
Distance[a]	−0.017 (0.010)[b]
Age[c]	0.034 (0.014)[d]
Male[e]	2.397 (0.500)[f]
Children < 9	0.181 (0.122)
Schooling	0.440 (0.093)[f]
Pseudo R^2	0.255

Note: Regressions is ordered logit. Robust standard errors are shown in parenthesis. Monolingual = people who spoke only Tsimane'.

[a]Distance is measured by kilometers from village of birth to nearest market town.

[b]Significant at the 10% level.

[c]Age is measured in years.

[d]Significant at the 5% level.

[e]Male = 1 if person was a man, and zero if person was a woman.

[f]Significant at the 1% level.

took three values: 2 = speaks Spanish well, 1 = speaks Spanish with difficulty, 0 = unable to speak Spanish (monolingual in Tsimane'). Explanatory variables included age, sex (male = 1; female = 0), kilometers in a straight line from village of birth to the nearest market town, schooling, and the number of children under nine years of age living in the household. The distance variable picks up background characteristics of the household in which the person grew up. Since parents might learn the national language from their children, and children learn languages faster than adults, the number of children under nine years of age should bear a positive association with parental fluency speaking Spanish if parents do not use children as translators. If children have infrequent exposure to the national language, then the number of children under nine years of age living in the household should bear a weak association with the language skills of parents.

Age, being a male, and schooling were associated with a higher likelihood of speaking Spanish. The number of children in the household did not bear a significant association with fluency speaking Spanish. Distance from village of birth to the nearest market town bore a negative association with fluency speaking Spanish; people who had spent their childhood years in villages far from market towns were less likely to speak Spanish than people raised in villages closer to market towns.

To explore further the reasons for Spanish fluency, we draw on a later annual wave of the panel. During the 2004 annual survey wave we asked people over 16 years of age to report the maximum schooling, Spanish fluency, and writing ability of their mother and their father. We coded answers to questions about human capital in the same way we coded the questions for own human capital described earlier, and ran an ordered logit regression (not shown). The regression included Spanish fluency as the dependent variable, and the following as explanatory variables: age, sex, village of birth, and the maximum schooling, writing skills, and Spanish fluency of the mother and the father. We found that the strongest covariates of Spanish fluency were own schooling (coefficient 0.56, $p < 0.001$), and maternal and paternal Spanish fluency (maternal = 0.43, $p < 0.01$; paternal = 0.33, $p < 0.01$). The additional evidence therefore suggests that variability in Spanish fluency likely reflects parental fluency in spoken Spanish. It is also possible that variability speaking Spanish might relate to unobserved attributes of the person, such as a person's interest and propensity to learn a language, or their sporadic attachment (and need) to learn the national language. Lamentably, we do not have information to explore these possibilities.

5.3. Associations between language skills and earnings

Table 3 contains the main regression results. The result of regression [1] suggests that people who spoke fluent Spanish earned 46.90% more than monolingual speakers of Tsimane'. Moderate fluency in Spanish bore no strong association with earnings. In columns 2–4 we control for math (column 2) and writing (column 3) skills and for schooling (column 4) by adding each of the variables, one at a time, to a separate regression. Adding other dimensions of human capital lowers the estimated association between fluency speaking Spanish and earnings from 46.9% (column 1) to 36.9–37.4% (columns 2–4). After controlling for other dimensions of modern human capital, we find that fluency in spoken Spanish still bears a positive association with earnings, but moderate or some fluency still bears no strong association with earnings. In column 5 we add only schooling without any of the other variables related to modern human capital to estimate the returns to schooling. We find that an additional year of schooling is associated

Table 3
Regression results: Proficiency speaking Spanish and earnings among Tsimane' adults (16+ years of age)

Variables	[1]	[2]	[3]	[4]	[5]	[6]
Human capital						
Some Spanish	0.122 (0.185)	0.097 (0.189)	0.089 (0.185)	0.089 (0.189)	^	0.086 (0.196)
Fluent Spanish	0.469 (0.170)[a]	0.374 (0.165)[a]	0.369 (0.189)[b]	0.369 (0.197)[b]	^	0.391 (0.204)[b]
Math	^	0.11 (0.05)[a]	0.118 (0.054)[b]	0.118 (0.086)	^	0.076 (0.114)
Writing	^	^	0.020 (0.106)	0.020 (0.110)	^	−0.056 (0.171)
Schooling	^	^	^	−0.00008 (0.065)	0.056 (0.035)	0.075 (0.084)
Fluent*math	^	^	^	^	^	0.060 (0.153)
Fluent*writing	^	^	^	^	^	0.115 (0.174)
Fluent*schooling	^	^	^	^	^	−0.101 (0.074)
Other						
Age	0.002 (0.003)	0.006 (0.004)	0.006 (0.004)	0.006 (0.004)	0.004 (0.004)	0.007 (0.004)
Male	0.910 (0.161)[c]	0.874 (0.169)[c]	0.886 (0.170)[c]	0.886 (0.187)[c]	1.096 (0.127)[c]	0.875 (0.185)[c]
Days ill	−0.003 (0.009)	−0.004 (0.010)	−0.001 (0.009)	−0.001 (0.009)	−0.001 (0.008)	−0.001 (0.010)
Days worked	0.047 (0.018)[a]	0.033 (0.021)	0.029 (0.021)	0.029 (0.014)[b]	0.040 (0.014)[a]	0.034 (0.016)[b]
Log BMI	1.126 (0.741)	1.253 (0.754)	1.200 (0.730)	1.200 (0.727)	1.132 (0.736)	1.215 (0.716)
Household size	0.027 (0.021)	0.029 (0.022)	0.032 (0.021)	0.032 (0.021)	0.031 (0.021)	0.033 (0.021)
Constant	−7.746 (2.339)[c]	−8.279 (2.373)[c]	−8.212 (2.354)[c]	−8.212 (2.352)[c]	−7.845 (2.381)[c]	−8.298 (2.290)[c]
R^2	0.38	0.39	0.40	0.40	0.38	0.40
Sample size	290	290	289	289	290	289

Note: ^ variables not included in the regressions. Regressions = OLS with village dummies (not shown) and clustering by village.
Dependent variable is measured by the logarithm of earnings for the 2 weeks before interview.
First quarter data are used to measure the explanatory variables.
Fifth quarter data are used to measure the earnings.
[a]Significant at the 10% level.
[b]Significant at the 5% level.
[c]Significant at the 1% level.

with 5.6% higher earnings, but results were statistically insignificant at the 90% confidence level or higher ($p<0.143$). In column 6 we include three separate interaction terms: Fluent*math, Fluent*schooling and Fluent*writing, where Fluent stands for fluency in spoken Spanish. Results buttress the previous analysis. After conditioning for interaction effects we find that fluency in spoken Spanish is associated with 39.1% higher earnings relative to monolingual speakers of Tsimane' and some fluency continues to bear no strong association with earnings. The results of column 6 also suggest that the positive associations between Spanish fluency and earnings come directly from Spanish fluency, rather than from the interaction of Spanish fluency with other variables related to modern human capital. None of the interaction terms between fluency in spoken Spanish and variables related to modern human were statistically significant.

In sum, the results of Table 3 suggest that none of the variables related to modern human capital besides language skills mattered in earnings, and that language skills mattered only when people spoke fluent Spanish, not when they spoke it moderately well. The variable *Some Spanish* bore no strong association with earnings. Spanish fluency bore an independent positive association with earnings even after conditioning for years of completed schooling and interactions with other variables related to modern human capital.

5.4. Robustness analysis

Table 4 contains the results of additional analysis to ensure robustness of the main results. The regressions in Table 4 resemble the regression in column 4 of Table 3, except for the changes described in the last column of Table 4. The columns "Some" and "Fluent" in Table 4 stand for the coefficients of "Some fluency in spoken Spanish" and "Complete fluency in spoken Spanish" as described earlier. In row 1 we add a variable for distance from the village of birth to the nearest market town to control for childhood conditions. Conditioning for distance, the coefficient for the variable for Spanish fluency drops from 0.369 to 0.318, and becomes statistically insignificant

Table 4
Regression results: Robustness analysis

No.	Coefficient of Spanish proficiency:		Changes introduced to main model (column 4, Table 3):
	Some	Fluent	
	0.089 (0.189)	0.369 (0.197)[a]	Baseline from column 4, Table 3
1	0.065 (0.192)	0.318 (0.254)	+ Distance from village of birth to town
2	0.090 (0.190)	0.367 (0.201)[a]	+ Age squared
3	0.276 (0.225)	0.507 (0.258)[a]	+ Control for traditional human capital with measure of ethno-botanical knowledge
4	0.117 (0.228)	0.520 (0.296)[a]	+ Experience and experience^2 instead of age. Experience = age-schooling-entry age in school.
5	0.089 (0.173)	0.369 (0.205)[a]	Clustering by households instead of villages
6	0.091 (0.252)	0.652 (0.317)[b]	Household fixed effects; only baseline
7	0.26(0.09)[c]	0.305 (0.129)[b]	Household fixed effects; five quarters ($n = 1542$)
8	0.223 (0.083)[c]	0.312 (0.110)[c]	Attrition and quarter dummies added; five quarters ($n = 1542$)
9	0.23(0.07)[c]	0.29 (0.102)[c]	Random-effect regression with quarter and attrition dummies; five quarters ($n = 1542$)

Note: Similar to regression [4] in Table 3, except for changes described in the last column. Standard errors are in parenthesis.
[a]Significant at the 10% level.
[b]Significant at the 5% level.
[c]Significant at the 1% level.

($p < 0.235$) in part because the sample size drops from 289 to 214 observations, thereby inflating standard errors. The variable for distance from village of birth to the closest market town had many missing values. In row 2 we add a quadratic term for age and find that the coefficient for the variable for Spanish fluency remains virtually unchanged at 0.36 ($p < 0.09$). In row 3 we add a variable for traditional human capital. During the survey we asked people to identify the uses of 19 wild and semi-domesticated plants for food, crafts, apparel, medicines, and construction (Reyes-García et al., 2003). Controlling for traditional human capital raises the coefficient for the variable of Spanish fluency from 0.369 to 0.507 ($p < 0.07$). In row 4 we take out the age variable, and replace it with experience and with experience squared, where experience equals age minus maximum years of school grades completed minus age at which the person started school. The coefficient for fluency speaking Spanish increased from 0.369 to 0.520 ($p < 0.10$).

Clustering by households rather than by villages did not change results (row 5). In row 6 we use a regression with fixed effects for households and restrict the analysis only to the first quarter. The coefficient for Spanish fluency reaches 0.652 ($p < 0.04$). A fixed-effect model for households with contemporaneous information for all variables estimated over the five quarters with quarter dummies added (row 7) suggests that complete fluency in spoken Spanish bears a positive association with earnings (coefficient 0.305, $p < 0.018$), but so does some fluency in spoken Spanish (coefficient 0.268, $p < 0.005$).

In rows 8 and 9 we control for attrition bias. We do so by creating a dummy variable for 14% of the people we could not find during the fifth quarter because they had temporarily left to hunt and work for wages; the variable took the value of one if we could not find the person during the fifth quarter, and zero otherwise. We ran the regression with contemporaneous information for all five quarters plus the dummy variable for attriters and dummy variables for quarters. We estimate the returns to language skills conditioning for attrition by running two different types of regressions; row 8 contains the results of an OLS regression and row 9 contains the results of a household-level random-effect regression. The OLS regression suggests that the dummy variable for attriters was insignificant (coefficient 0.102; $p < 0.387$), but the two language variables were significant. In row 8 we see that some fluency and complete fluency in spoken Spanish were associated with 22.30% ($p < 0.007$) and 31.2% ($p < 0.005$) higher earnings relative to monolingual speakers of Tsimane'. Results of the random-effect household regression produced similar coefficients to those found in the OLS regression.

In sum, taken together, the results of Table 4 suggest that fluency speaking Spanish bears a positive association with earnings. As in the main regressions of Table 3, in the regressions testing for robustness we found that slight competence speaking Spanish was generally not associated with higher

earnings, suggesting a threshold of fluency beyond which Spanish competence pays off.

5.5. Paths

To identify the paths by which fluency in spoken Spanish might be associated with earnings, we used moderate and complete fluency in spoken Spanish as explanatory variables against the following dependent variables: (a) access to credit, (b) value of modern physical assets, (c) foraging productivity, and (d) self-perceived health (Table 5). The notes to Table 5 contain the other explanatory variables used in each regression.

The results in the first column of Table 5 suggest that people who spoke some Spanish (or who spoke it with difficulty) had 38.25 more bolivianos in credit ($p<0.074$) than monolingual speakers of Tsimane'; people who spoke fluent Spanish had 143.58 more bolivianos ($p<0.001$) in credit than monolingual speakers of Tsimane' (1 US dollar = 7.45 bolivianos during fieldwork). Results held up even after using a median regression (not shown) to control for the lower censoring of the credit variable (57.43% of people did not take out credit). To explore the topic further, we asked people whether they could borrow 100 bolivianos in an emergency. People who spoke Spanish with difficulty and those who spoke Spanish fluently were 8.74% ($p<0.009$) and 7.35% ($p<0.098$) more likely to have access to 100 bolivianos in credit in an emergency than monolingual speakers of Tsimane'.

People who spoke Spanish with difficulty and those who spoke Spanish fluently had 64.73% ($p<0.001$) and 63.68% ($p<0.001$) more wealth than monolingual speakers of Tsimane'. Modern wealth included assets that protect health (e.g., mosquito nets) and that increase labor productivity (e.g., metal fishhooks). People who spoke Spanish

Table 5
Paths: relation between fluency in spoken Spanish and credit, modern assets, foraging productivity, and self-perceived health among Tsimane' adults (16+ years of age) over five quarters

Explanatory variables, Spanish fluency	Dependent variables				
	Access to credit		Log of value of modern assets	Foraging productivity	Self-perceived health
	Level [1]	Emergency [2]	[3]	[4]	[5]
Some	38.25[a] (21.41)	.087[b] (.035)	.647[b] (.146)	1.475[c] (.645)	.016 (.610)
Fluent	143.58[b] (27.79)	.073[a] (.047)	.636[b] (.183)	.825 (.839)	1.524[a] (.816)
Sample size	2116	1916	1739	1303	1703

Notes: Standard errors in parenthesis.
Explanatory variables common to all regressions: two dummies for different levels of proficiency speaking Spanish, sex, age, math, writing, schooling, distance from village of birth to nearest market town, household size/quarter, and dummies for quarters and villages. Below we list explanatory variables unique to each regression besides the ones just listed.
[1] Dependent variable = total credit obtained by person. Explanatory variables: logarithm of total wealth (modern and traditional assets). Regression is lowered-censored tobit.
[2] Dependent variable = 1 if person reported being able to access 100 bolivianos in an emergency, 0 otherwise. Explanatory variables same as [1] Regression is probit with probability estimated at mean values of explanatory variables and clustering by people.
[3] Dependent variable = logarithm of value of 13 modern physical assets owned by person (e.g., metal fishhooks, mosquito nets). Explanatory variables: age and sex-standardized height for age and logarithm of body-mass index. Regression is OLS with household-fixed effect (but no village dummies) and clustering by people.
[4] Dependent variable = kilograms of fish brought by person during 24 h before scans or spot observations (see text). Explanatory variables: total number of days with self-reported illness for 2 weeks before interview, and logarithms of BMI and total wealth (modern and traditional assets). Regression is lowered-censored tobit.
[5] Dependent variable = total number of days with self-reported illness for 2 weeks before interview. Explanatory variables: total number of days worked during the 2 weeks before the day of the interview, logarithms of BMI and total wealth (modern and traditional assets), and age and sex-standardized z scores of height for age and mid-arm muscle area. Regression is lowered-censored tobit.
[a]Significant at the 10% level.
[b]Significant at the 5% level.
[c]Significant at the 1% level.

with difficulty caught 1.47 more kilograms of fish/day than monolingual speakers of Tsimane' ($p<0.022$).

Last, bilingual speakers of Spanish and Tsimane' reported more self-perceived illness. People who spoke fluent Spanish reported feeling ill 1.52 more days during the 14 days before the day of the interview than monolingual speakers of Tsimane' ($p<0.062$). The positive association between fluency in Spanish and perceived illness may stem from the cultural inflation of self-perceived morbidity that arises with economic prosperity (Murray & Chen, 1992).

We did other analysis (not shown) to assess whether competence speaking Spanish bore an association with anthropometric indicators of short-run nutritional status, hunting (rather than fishing) productivity, frequency of contact with outsiders or travel to market towns, social capital or pro-social behavior, and self-perceived emotions (e.g., happiness), and found no strong, significant associations.

In sum, the analysis of possible paths suggests that increased skills speaking Spanish are associated with greater access to credit, accumulation of modern physical assets, improvements in some types of foraging productivity, and greater self-perceived illness. Except for greater self-perceived illness, the other paths might help to explain why proficiency speaking the national language bore a positive association with earnings.

6. Conclusions

We end by comparing our results with the results from other nations and pointing to a policy conclusion that might bear on other native populations of Latin America.

Immigrants in the USA who are proficient speaking English earned 15–33% more than immigrants who could not speak English well (Bleakley & Chin, 2004; Chiswick & Miller, 1999). The results of Table 3 suggest that fluency in spoken Spanish was associated with 36.9–46.9% higher earnings compared with the earnings of monolingual speakers of Tsimane'. Our estimates fall toward the upper range of estimates from industrial nations. The 36.9–46.9% premium is also higher than the 25% premium for bilingualism in Spanish and an American Indian language that Chiswick et al. (2000) found among indigenous people in the labor market of Bolivian cities. If OLS estimates produce conservative estimates of the returns to language skills (Godoy, Karlan, Rabindran & Huanca, 2005), then one might read the 36.9–46.9% return as a lower bound of the true magnitude.

In industrial nations, linguistic assimilation has typically stood for broader cultural assimilation and the irrevocable loss of competence in the mother tongue of immigrants' descendants. In contrast to the "linguistic graveyard" one finds in industrial nations, what we find among the Tsimane' is a case where a small-scale society of only ~8000 native Amazonians has become bilingual, captured the premium from speaking the majority national language, but has done so without absconding their traditional language and culture. This has come about through a school curriculum in the Tsimane' language taught by Tsimane' teachers in village schools, continued access to their traditional farming and foraging grounds, and the survival of their traditional social organization.

Acknowledgements

Research was funded by Grants from the programs of Biological and Cultural Anthropology of the National Science Foundation (0134225, 0200767, and 0322380). Thanks go to Lilian Apaza, Esther Conde, Johnny Dávila, Homero Rivas, Yerko Lobo, Lourdes Parada, Bernabé Nate, Paulino Pache, Evaristo Tayo, Santiago Cari, José Cari, Manuel Roca, Daniel Pache, Javier Pache, and Vicente Cuata for help collecting the information and for logistical support. Thanks go to Russ Bernard, John Robst, and an anonymous reviewer for commenting on an earlier version of the paper.

References

Bleakley, H., & Chin, A. (2004). Language skills and earnings: Evidence from childhood immigrants. *Review of Economics and Statistics*, *86*(2), 481–496.

Burhop, J.E.J. (2003). *Infection, C-reactive protein, and growth of Tsimane' Amerindian children from lowland Bolivia*. Ph.D. dissertation, Department of Neurobiology and Physiology, Northwestern University.

Byron, E. (2003). *Markets and Health: The Impact of Markets on the Nutritional Status, Morbidity, and Diet of the Tsimane' Amerindians of Lowland Bolivia*. Ph.D. dissertation, Department of Anthropology, University of Florida.

Castro, M. (1997). *La viva voz de las tribus: El trabajo del ILV en Bolivia, 1954–1980*. La Paz, Bolivia: Ministerio de Desarrollo Sostenible y Planificacion, Vice Ministerio de Asuntos Indigenas y Pueblos Originarios.

Chicchón, A. (1992). *Chimane resource use and market involvement in the Beni Biosphere Reserve, Bolivia*. Ph.D. dissertation, Department of Anthropology, University of Florida.

Chiswick, B. R. (1998). Hebrew language usage: Determinants and effects on earnings among immigrants in Israel. *Journal of Population Economics*, *11*, 253–271.

Chiswick, B. R., & Miller, P. W. (1999). Language skills and earning among legalized aliens. *Journal of Population Economics*, *12*, 63–89.

Chiswick, B. R., Patrinos, H. A., & Hurst, M. E. (2000). Indigenous language skills and the labor market in a developing economy: Bolivia. *Economic Development and Cultural Change*, *48*(2), 347–367.

Dustmann, C., & Fabbri, F. (2003). Language proficiency and labor market performance of immigrants in the UK. *Economic Journal*, *23*, 695–717.

Foster, A. D., & Rosenzweig, M. R. (1996). Technical change and human-capital returns and investments: Evidence from the Green Revolution. *American Economic Review*, *86*(4), 931–954.

Foster, Z., Byron, E., Reyes-García, V., Huanca, T., Vadez, V., Apaza, L., et al. (2005). Physical growth and nutritional status of Tsimane' Amerindian children of lowland Bolivia. *American Journal of Physical Anthropology*, *126*, 343–351.

Godoy, R. A., Karlan, D. S., Rabindran, S., & Huanca, T. (2005). Do modern forms of human capital matter in primitive economies? Comparative evidence from Bolivia. *Economics of Education Review*, *24*(1), 45–53.

Grigorenko, E. L., & Sternberg, R. J. (2001). Analytical, creative, and practical intelligence as predictors of self-reported adaptive functioning: A case study in Russia. *Intelligence*, *29*, 57–73.

Instituto Nacional deEstadística. (2003). *Bolivia Caracteristicas sociodemograficas de la poblacion indigena*. La Paz: INE.

Lazear, E. P. (1999). Culture and language. *Journal of Political Economics*, *107*, S95–S126.

Maffi, L. (2001). *On Biocultural Diversity. Linking Language, Knowledge, and the Environment*. Washington DC: Smithsonian Institution Press.

McDade, T. W., Leonard, W. R., Burhop, J., Reyes-García, V., Vadez, V., Huanca, T., et al. (2003). Acculturation, C-reactive protein, and child growth in lowland Bolivia. *American Journal of Human Biology*, *15*, 273–274.

McDade, T.W., Leonard, W.R., Burhop, J., Reyes-García, V., Vadez, V., Huanca, T., et al. (2005). Predictors of C-reactive protein in Tsimane' 2–15 year-olds in lowland Bolivia. *American Journal of Physical Anthropology*, in press.

Mc Ewan, P. J. (2004). The indigenous test score gap in Bolivia and Chile. *Economic Development and Cultural Change*, *53*, 157–190.

McManus, W. S. (1998). Labor market effects of language enclaves: Hispanic men in the United States. *Journal of Human Resources*, *25*(2), 228–252.

Mincer, J. (1974). *Schooling, Experience and Earnings*. Cambridge, MA: National Bureau of Economic Research.

Murray, C. J. L., & Chen, L. C. (1992). Understanding morbidity change. *Population and Development Review*, *18*(3), 481–503.

Patrinos, H.A. (1994). *The cost of discrimination in Latin America*. Washington DC: The World Bank; Report nr HROWP 45.

Patrinos, H.A., & Psacharopoulos, G. (1992). *Socioeconomic and ethnic determinants of grade repetition in Bolivia and Guatemala*. Washington DC: The World Bank: Technical Department, Latin America and the Caribbean, Report no. WpS 1028.

Psacharopoulos, G., & Patrinos, H. A. (1994). Indigenous people and poverty in Latin America. *Finance & Development*, *23*, 41–43.

Reyes-García, V., Godoy, R. A., Vadez, V., Apaza, L., Byron, E., Huanca, T., et al. (2003). Indigenous people share botanical knowledge. Estimates from an Amerindian society in Bolivia. *Science*, *299*, 1707.

Sternberg, R.J., Lipka, J., Newman, T., Wildfeuer, S., & Grigorenko, E.L. (2002). Triarchically based instruction and assessment of sixth-grade mathematics in a Yup'ik cultural setting in Alaska. *Contemporary Education Psychology* [under review].

Sternberg, R. J., Nokes, C., Geissler, P. W., Prince, R., Okatcha, F., Bunda, D. A., et al. (2001). The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence*, *29*, 401–418.

Tanner, S. (2005). *A population in transition: Health, culture change, and intestinal parasitism among the Tsimane' of lowland Bolivia*. Ph.D. dissertation, Department of Anthropology, University of Michigan.