The Endogeneity between Language and Earnings: International Analyses

Barry R. Chiswick, *University of Illinois at Chicago*

Paul W. Miller, *University of Western Australia*

This study is concerned with the determinants of dominant language fluency, its effects on earnings, and its endogeneity with earnings among immigrants. Dominant language fluency is hypothesized to be a function of three fundamental variables: exposure to the language, efficiency in second language acquisition, and economic benefits from language fluency. Conceptual variables with empirical counterparts are developed. Earnings are hypothesized to be a function of language skills, among other variables. Ordinary least squares, instrumental variables, and sample selection bias techniques are used to estimate the equations for Australia. Comparisons are made with analyses for the United States, Canada, and Israel.

I. Introduction

It has become recognized in recent years that an important aspect of the overall adjustment of immigrants in the labor market is their linguistic

We appreciate the comments received on an earlier version of this article from Michael Kidd and at seminars presented at Flinders University, Hebrew University, Queen's University, Tel Aviv University, and the University of Illinois at Chicago, as well as at the European Science Foundation Conference on Migration and Development, Davos, Switzerland, 1992, the International Regional Science Association Annual Meeting, Chicago, 1992, and the Australian Conference of Economists, Melbourne, 1992.

[*Journal of Labor Economics*, 1995, vol. 13, no. 2]

0734-306X/95/1302-0006$01.50

adjustment.[1] Linguistic adjustment refers to the process by which immigrants who are not fluent in the destination's dominant language improve their fluency. Linguistic adjustment influences labor market outcomes, such as earnings. Previous research studies in this area have used ad hoc models and have not adequately addressed the issue of the extent to which language skills and earnings are mutually determined for immigrants.

This article addresses these deficiencies in the literature. First, it develops a systematic model based on economic principles for analyzing the determinants of immigrant fluency in the dominant language. Second, various techniques are used to analyze the effect on earnings of dominant language fluency. Third, the analysis explicitly examines the endogeneity between language fluency and earnings, in particular, whether those who anticipate greater earnings if they were to become fluent are in fact more likely to become fluent.

The empirical analysis is performed in detail for Australia, using the data from the 1981 and 1986 Australian Censuses. These results are then examined in a comparative framework with the findings obtained elsewhere for the United States, Canada, and Israel (Chiswick and Miller 1992; Chiswick 1993). There are several advantages to analyzing the Australia data. Australia has a large proportion foreign-born in its population (21% in 1986), compared to the United States (9% in 1990). The immigrants to Australia come from a diverse set of countries that differs from the major source regions for the United States. Among the foreign-born in Australia in 1981, 37% were born in Britain and Ireland, 43% are from various other parts of Europe (primarily southern Europe), 12% from Asia and Africa, 4% from New Zealand, and 3% from the Western Hemisphere. Unlike Spanish in the United States (which accounts for over half of all immigrants who speak a language other than or in addition to English at home), there is no single non-English language that plays a dominant role in Australia. Italian, Greek, and Chinese are the three most common immigrant languages. Illegal aliens, who constitute a population that tends to avoid census enumeration, are very few in number in Australia, and again, unlike those in the United States, they play a trivial role in the low-skilled immigrant labor market. Thus, the problems of census undercount that may be important among low-skilled immigrants, particularly Hispanics, in the U.S. Census are of lesser importance in Australia. Moreover, the Australian question on language fluency is comparable to the question used in the

[1] See, e.g., Cox (1975); McManus, Gould, and Welch (1983); Evans (1986); Stromback (1986); Chiswick and Miller (1988, 1992); Kossoudji (1988); Robinson (1988); Tainer (1988); and Chiswick (1991), among others. Research on this issue has used U.S., Canadian and Australian survey and census data. Most of the literature has analyzed fluency rather than literacy in the destination language because data on speaking skills are more abundant than data on reading/writing skills.

U.S. 1980 Census. Finally, the study of the 1981 and 1986 Censuses of Australia permits a comparative analysis over time that is a test of robustness.

Section II develops the model of the determinants of English language fluency. The data used to test the model, the Australian Censuses of 1981 and 1986, are described in Section III. The empirical analysis of the determinants of English language fluency among immigrants is presented in Section IV. In Section V the effect of language skills on earnings is studied using alternative econometric techniques to ascertain the magnitude of the effect and whether there is endogeneity between language and earnings. These analyses are highly instructive even though it is not possible to demonstrate causality. Comparative analyses of the findings for Australia with parallel analyses performed on census data for the United States (1980), Canada (1981), and Israel (1983) are reported in Section VI. This international comparative analysis demonstrates the robustness of the model. Section VII closes this article with a summary and conclusion.

II. Determinants of Language Fluency: The Theory

Language skills are an important form of human capital. They satisfy the three basic requirements for human capital: they are embodied in the person; they are productive in the labor market and/or in consumption; and they are created at a sacrifice of time and out-of-pocket resources.

Skills in one's "mother tongue" are acquired when young. At this stage the investments are made largely by parents or caregivers. This is a time in the life cycle when the human mind is especially efficient in creating language capital (Harley 1986; Long 1990). Among school-age children, language capital is acquired when other forms of human capital (e.g., physical maturation, schooling) are being acquired. Thus, their acquisition of spoken language skills in the mother tongue seems almost effortless.

Among immigrants, however, the acquisition of language capital relevant for the destination can be very costly and certainly does not appear to be an effortless process when this language differs sharply from the mother tongue.[2] Dominant language fluency among immigrants can then be expressed as a function of three conceptual variables: economic incentives, exposure, and efficiency.

As with other forms of human capital, *economic incentives* can be expected to be an important determinant of language capital acquisition. The economic incentives arise in part from the increment in the market wage rate, a higher rate of employment, and the decrease in the cost of consumption

[2] This explains, in part, the preference on the part of international migrants for destinations with the same or similar mother tongue. See, e.g., Chiswick and Miller (1994).

(including search costs) associated with a higher level of fluency. This suggests an endogeneity between earnings and language skills.

The economic incentives for language acquisition are also related to the expected future duration in the destination. Immigrants expecting to return to their country of origin would have a shorter expected future duration in the destination and, therefore, less of an incentive to make language investments specific to the destination and more of an incentive to make investments that retard the depreciation of language-of-origin skills. Thus, other variables being the same, destination-language fluency would be lower for immigrants from countries where there is a higher incidence of return migration.

Exposure refers to the learning by doing and the formal instruction aspects of acquiring fluency in the destination language, as well as formal language training. It includes the extent to which others, whether in person or through the media, use the destination language in one's presence and the extent to which the person himself or herself utilizes it. Exposure can be thought of as having three components: exposure prior to immigration, time units of exposure in the destination, and the intensity of exposure per unit of time in the destination.

Exposure to the destination language prior to immigration is greater the smaller the "linguistic distance" between this language and the immigrant's mother tongue. The linguistic distance between Spanish and Italian, for example, is smaller than that between Spanish and Korean. Therefore, including language (or country) of origin in an analysis of fluency in the destination language measures, in part, the effects of linguistic distance.

For immigrants from multilingual countries of origin (e.g., India), a measure of the linguistic distance between the immigrant's mother tongue and the destination language may be misleading. Exposure to the destination language while still in the country of origin, whether it is used as a lingua franca (e.g., English in India) or because of the presence of foreign nationals (e.g., overseas U.S. military bases), would enhance destination language skills. Preimmigration exposure may also include instruction in school in the destination language. This would result in a positive effect of schooling on destination-language skills.

Duration in the destination, as measured by the number of years since migration, is also an index of exposure. Other variables being the same, destination-language fluency would be expected to increase with duration, up to the asymptote of full fluency.

The intensity of exposure per unit of time in the destination is smaller for those living in an environment in which more people communicate in the immigrant's mother tongue. Thus, the greater the extent to which a minority language is spoken in the area in which the immigrant lives,

whether by immigrants or natives, the poorer will be the fluency in the destination language.[3]

Perhaps the most important language environment is in the home. Language skills emerge in large part through the linguistic interaction of those living together. Marriage to a spouse from the same linguistic origin will detract from destination-language exposure and thereby reduce fluency, compared to marriage to a native speaker of the destination language.[4] Children in the family, particularly those born in the destination, are more likely to be fluent in the dominant language because of the effects of age on language acquisition and their enrollment in school. Thus, to the extent that immigrant parents acquire fluency from their children, adult immigrant-language fluency is, on the one hand, enhanced by the presence of children. On the other hand, rather than serving as teachers, children may serve as parental interpreters to the world outside the immigrant-language enclave. If so, children would inhibit destination-language fluency on the part of adult immigrants. The role, if any, of children as interpreters is likely to be less important in the labor market than in the market for consumer goods and services.

An important aspect of exposure per unit of time to the destination language is formal instruction in this language (e.g., English as a Second Language programs). The determinants of enrollment and its consequences for destination-language fluency are clearly important issues. Unfortunately, the data available for the study of language fluency generally lack information on participation in language-training programs.

Efficiency refers to the extent to which a given amount of destination-language exposure produces language fluency. The very young have an impressive ability (efficiency) to acquire language skills, even in more than one language simultaneously. With age, however, this facility appears to diminish.[5]

Efficiency in language acquisition may be enhanced by a higher level of education. This may arise because the more educated have a greater mastery of their mother tongue and are more efficient in learning new concepts and new terminology.[6] Furthermore, those with schooling in the destination

[3] This approach is similar to the neighborhood effects model used by Case and Katz (1991) in their study of family and neighborhood peer effects on the behavior of inner-city youths in the United States.

[4] This is consistent with Evans's (1986, p. 234) hypothesis that "members of large immigrant groups with more inward-looking friendship networks and more in-marriage will be less skilled in the host country's language."

[5] For detailed analyses of this issue, including a survey of the literature, see Harley (1986) and Long (1990).

[6] As noted above, those with higher levels of schooling acquired in the origin may also have been exposed to the destination language in school if this was part of the curriculum.

would be expected to be more fluent in the destination language as fluency may be a prerequisite for school enrollment and the destination schooling itself would enhance fluency.

Refugees may appear to have a lesser efficiency in acquiring dominant language skills than economic migrants from the same linguistic origin for two reasons (Chiswick 1978). First, refugees are less intensively self-selected for the characteristics that enhance a successful adjustment in the destination. Factors other than successful adjustment play a larger role, if not a dominant role, in their decision to migrate. Second, refugees could be expected to have invested in less preparation for the move, particularly if sudden changes in political events are responsible for their refugee status. But to the extent that refugees have a lower probability of return migration, they would make greater investments specific to the destination. Ceteris paribus, this implies a lower initial level of fluency and a steeper rise in language fluency with duration among refugees than among economic migrants.

This analysis suggests the following conceptual equation:

$$\text{LANG} = f(\text{economic incentives, exposure, efficiency}),$$

where LANG is a measure of the immigrant's fluency in the dominant language. The empirical counterpart of this conceptual equation is

$$\begin{aligned}\text{LANG} = f[&\text{expected wage increment for language fluency (+)},\\ &\text{expected future duration (+), duration in destination (+)},\\ &\text{married to native of destination (?)},\\ &\text{married to native of origin (−), children (?)},\\ &\text{minority language concentration (−), destination language}\\ &\text{instruction (+), linguistic distance (−)},\\ &\text{age at migration (−), education (+), refugee status (−)}],\end{aligned}$$

where the expected partial effect (positive or negative) is indicated in parentheses.

Not all of these variables can be measured on an individual basis. There are no data on individual variations in expected increments in wages due to improved language fluency. It is, therefore, not possible to estimate directly the partial effect of expected wage increases due to fluency on the acquisition of fluency.

One approach would be to use observed earnings as a proxy for the expected wage increment variable. Then equations for language fluency and earnings could, in principle, be estimated simultaneously. This requires

identifying instruments for both the language and earnings equations. The limitations inherent in determining appropriate identifying instruments, particularly for identifying the language equation, preclude this approach. Alternative procedures, to be discussed below, are used to address the relationship between language fluency and earnings.

In addition, in the data to be studied explicit information is not available for refugee status, linguistic distance, and expected future duration. Dichotomous variables for country of birth are used as proxy measures for these variables. Data are also not available for destination-language instruction. Data files that include one or more of these variables are deficient in other, far more serious, dimensions.

III. The Data

The detailed empirical analyses reported in Sections IV and V below use data for Australia and test the robustness of the model by comparing findings for the 1981 and 1986 Censuses of Australia. Further tests of robustness are presented in Section VI through comparative analyses with findings for the United States, Canada, and Israel. This section presents a detailed discussion of the data used in the Australian analysis. The data for the United States and Canada are discussed in Chiswick and Miller (1992) and for Israel in Chiswick (1993).

The empirical analysis uses the microdata files (1/100 sample) released from the 1981 and 1986 Australian Censuses of Population and Housing. For most immigrant labor market analyses, the 1981 Census is far superior to the 1986 Census. The 1981 Census reports income in 14 categories, rather than the 8 available in the 1986 data, and period of immigration in single years (up to an upper open-end interval of 35 or more years), rather than the 5 broad categories used in presentation of the 1986 data. An advantage of the 1986 Census, however, is that it includes information on any second language spoken in the home by the respondent and thereby permits the creation of a better minority language concentration index. Thus, the empirical analysis of language fluency in Australia uses both Censuses so as to exploit as fully as possible the available data and test for robustness.

The statistical analysis is limited to foreign-born males aged 25–64 who were employed at the time of the Census. The 1981 Census asked if the respondent spoke a language other than English at home, and, if yes, also asked for the degree of fluency in spoken English—"very well," "well," "not well," and "not at all." In the 1986 Census, the respondent was also asked if a language other than English was spoken at home. This time, however, if the response was "yes," in addition to the question on fluency in English, the respondent was asked to identify the other language. Twelve minority languages, representing 74% of immigrants reporting a minority language, are coded on the data file.

It is useful to collapse the four-category English language proficiency variable into two categories for the statistical analysis: those who speak only English or speak it very well versus all other groups. The Australian Bureau of Statistics cautions that the "not well" and "not at all" categories will underestimate the extent of English-language deficiency (Australian Bureau of Statistics 1982, p. 1). Furthermore, preliminary tests indicated that among immigrants from non-English-speaking countries there is no difference in earnings between those who reported they spoke only English at home and those who spoke another language but spoke English very well, in contrast to the other three groups for which earnings were significantly lower by about 12%.[7]

Hence, in the analysis reported here, immigrants fluent in English are defined as those who speak only English at home and those who also speak another language but speak English very well. Using this definition, the fluency rate is 76.0% for adult male immigrants and 58.2% for those from non-English-speaking countries.[8] The fluency rate varies sharply across birthplace regions. It is relatively low for southern Europe (39.9%), other Asian countries (i.e., excluding Vietnam and South Asia, 46.9%), Vietnam (16.7%) and South and Central America (38.8%) but much higher for northern Europe (excluding the British Isles, 87.3%) and South Asia (87.8%).

The model in Section II above suggests the importance of a minority-language concentration variable in the analysis of language fluency. Specific minority languages are not identified in the 1981 census, but there is substantial detail on country of birth. Linguistic country groups were created by combining countries with the same language (e.g., combining Portugal, Brazil, and Timor). Then the proportion of the population aged 15–64 in the region in which the immigrant lives that is of the same minority linguistic-country group as the immigrant is assigned to the respondent.

[7] Using the 1981 Census data for immigrants from non-English-speaking countries, controlling for education, experience, duration, location, and citizenship, the earnings differentials from the benchmark (speaking only English at home) are

Speak	Coefficient	*t*-Ratio
Very well	−.020	−.84
Well	−.125	−4.91
Not well	−.128	−3.97
Not at all	−.111	−.88

[8] The English-speaking countries for this purpose and in the regression analysis include Britain, Ireland, Canada, the United States, British West Indies, and New Zealand.

The countries in which English is the mother tongue or a major lingua franca are assigned a value of zero.

In the 1986 census, the 12 most important minority languages are specifically identified. The minority-language concentration variable is defined as the proportion of the population aged 15–64 in the region in which the respondent lives that reports the same minority language. Those who speak only English and the minority languages with so few speakers that they are not separately identified are assigned a value of zero.

It should be noted that, while the dependent variable is the respondent's level of English-language fluency, the concentration measure is not the mean level of English fluency in the person's region. In one instance (1981) it is the proportion of immigrants from the same country group, and in the other (1986) the proportion of immigrants and natives who speak the same non-English language at home. This mitigates, if not avoids, the endogeneity problem raised by Case and Katz (1991).

The relevant characteristics of the Australian census, the variables discussed above, and the other variables used in the statistical analysis are defined in detail in the Appendix, where the means and standard deviations of the variables for Australia are also reported.

IV. Analysis of Language Fluency: Australia

The results of alternative specifications of the model for English language fluency are reported in table 1. The model includes the exposure and efficiency variables discussed above. In the absence of data on the increment in wages, the effect of economic incentives is analyzed below through an analysis of residuals. The substantive findings in the ordinary least squares (OLS) and logit analyses are virtually identical, so only the former are discussed explicitly.

The analysis in table 1 indicates that English-language fluency is related to education (years of schooling), duration of residence, age (which measures the effect of age at migration when duration is held constant), current marital status, whether married overseas, number and age of children, size of place, and dichotomous variables for country of birth. The minority birthplace concentration measure is included in columns 2–5. Columns 1–3 are OLS regressions, while column 4 is a logit equation and column 5 is OLS excluding immigrants from the major English-speaking source countries (Britain, Ireland, Canada, United States, British West Indies, and New Zealand).

The estimates for the baseline specification presented in table 1, column 1, reveal that each additional year of education is associated with an increase in the fluency rate of 2.5 percentage points, or about 3.6 percentage points for those from non-English-speaking countries. Educational attainment therefore has a pronounced effect on language skills: there is, for example, a difference of 12.5 percentage points between the language fluency rates

of individuals with the mean level of education (of around 11 years) and those who hold bachelor degrees, other variables being the same. This rises to 18.0 percentage points for non-English-origin immigrants.

Age at immigration is also an important consideration, with language proficiency declining the greater the age at migration. For example, immigrants who arrived in Australia at age 25 are predicted to have English fluency rates 10 percentage points greater than immigrants who arrived at 45 years of age, ceteris paribus, but this is 20 percentage points among the sample of non-English-origin immigrants.

Language skills improve rapidly with years in Australia. Each additional year of residence is associated with almost a 1 percentage point improvement in the language fluency rate among all immigrants and a 1.7 percentage point increase for those from non-English-speaking countries. The differences in the language fluency rates across duration of residence categories are impressive. For example, the language fluency rate of immigrants who have been in Australia for the mean period of residence (= 18.5 years) would be around 17 percentage points higher than for the most recent arrivals, while it would be 31 percentage points higher among those from non-English-speaking countries.

There is a distinctive pattern by marital status. The language proficiency of individuals who married in Australia does not differ significantly from the rate of individuals who are single. The rate for individuals who married prior to migration, however, is around 4 percentage points (8.7 percentage points for those who are not from the major English-speaking source countries) lower than for the remainder of the group. Marriage prior to migration—in general, marriage to a person of the same origin country—appears to reduce opportunities to acquire fluency in English.

The other family-related variables in the equation are the presence and age of children. These variables are generally statistically insignificant, although they are at the margin of statistical significance in the table 1, column 3, specification that includes limited interaction terms.[9] Moreover, the coefficients are typically negative, although they are less negative or more positive when there are at least two children, one preschool age and the other school age. More children may result in English language interaction among the children, and hence greater parental fluency.

What role, therefore, do children play in parental language attainment? One possibility is associated with the desire to preserve the language of origin, perhaps because of an expectation of return migration, to maintain a cultural/national identity, or to maintain ties with relatives in the origin. This means teaching the minority language to children at home or in

[9] Alternative specifications of the children variables, including analyses within birthplace groups, also result in negative, but generally statistically insignificant, coefficients.

Table 1
Regression Estimates of English-Language Fluency among Adult Foreign-born Men, Australia, 1981 (Dependent Variable: LANG)

Variable	Total Sample				Non-English-Speaking
	OLS (1)	OLS (2)	OLS (3)	Logit (4)	OLS (5)
Constant	.814	.819	.475	5.766	.292
	(30.24)	(30.53)	(9.94)	(10.44)	(5.71)
Education	.025	.024	.031	.230	.036
	(16.35)	(16.11)	(12.09)	(13.35)	(15.23)
Age	−.005	−.005	.001	−.068	−.010
	(10.95)	(10.91)	(1.59)	(11.23)	(11.33)
Years since migration (YSM)	.009	.009	.028	.116	.017
	(17.24)	(17.79)	(11.68)	(16.98)	(18.33)
Married	−.011	−.009	−.004	−.104	−.011
	(.91)	(.75)	(.37)	(.74)	(.51)
Married overseas	−.037	−.039	−.048	−.410	−.087
	(3.42)	(3.59)	(4.40)	(3.59)	(4.71)
Child < 6 years only	−.022	−.022	−.026	−.224	−.038
	(1.57)	(1.60)	(1.92)	(1.42)	(1.63)
Child 6–17 years only	−.007	−.007	−.020	−.068	−.015
	(.73)	(.80)	(2.12)	(.66)	(.96)
Children < 6 and 6–17 years	.004	.002	−.012	.059	−.002
	(.33)	(.15)	(.91)	(.41)	(.11)
Small urban location	.032	.011	.010	.045	.008
	(3.01)	(1.08)	(.93)	(.28)	(.36)
Rural location	.026	.006	.001	−.055	.007
	(1.97)	(.43)	(.11)	(.30)	(.26)
Minority birthplace concentration	...	−.047	−.048	−.275	−.053
		(5.35)	(5.48)	(5.10)	(5.75)

Birthplace:					
Southern Europe	−.578	−.472	−.475	−7.068	. . .
	(53.02)	(20.26)	(20.36)	(15.02)	
Northern Europe	−.186	−.137	−.146	−5.304	.277
	(15.79)	(8.93)	(9.46)	(11.32)	(14.36)
Eastern Europe	−.405	−.388	−.374	−6.999	.035
	(17.43)	(16.52)	(16.10)	(14.82)	(1.12)
Arab countries	−.499	−.462	−.452	−6.742	.037
	(14.68)	(13.35)	(13.23)	(13.91)	(.98)
South Asia	−.145	−.137	−.139	−4.759	.326
	(7.39)	(7.05)	(7.37)	(9.67)	(10.41)
Philippines	−.127	−.119	−.098	−4.928	.386
	(1.75)	(1.64)	(1.31)	(6.35)	(4.91)
Vietnam	−.787	−.764	−.741	−8.325	−.232
	(16.94)	(16.40)	(16.23)	(13.15)	(4.79)
Other Asian countries	−.541	−.526	−.518	−7.414	−.057
	(14.45)	(14.05)	(13.95)	(14.95)	(1.37)
South and Central America	−.558	−.535	−.532	−6.953	−.013
	(10.49)	(10.07)	(10.12)	(13.04)	(.25)
Africa	−.197	−.170	−.173	−5.262	.296
	(7.55)	(6.50)	(6.66)	(10.71)	(8.91)
Remainder group	−.154	−.152	−.146	−4.801	.271
	(3.32)	(3.30)	(3.22)	(6.52)	(4.75)
Age*YSM/100	. . .	. . .	−.034	. . .	. . .
			(9.45)		
Education*YSM/100	. . .	. . .	−.044	. . .	. . .
			(3.20)		
N	7,288	7,288	7,288	7,288	4,166
R^2	.4472	.4512	.4568		.3520
McFadden's R^2				.5108	

SOURCE.—1981 Australian Census of Population and Housing, 1/100 Sample of the Foreign Born.

NOTE.—*t*-statistics are in parentheses and have been derived using White's (1980) heteroscedasticity-consistent covariance matrix estimator. See the Appendix for variable definitions. The prediction success rate for the logit model is 86.72%.

school.[10] In his review of the position of various immigrant groups in Australia for the 1975 Australian Government Commission of Inquiry into Poverty, Cox (1975, p. 85) comments on Polish immigrants: "The resulting emphasis upon teaching Polish and utilizing it in the home had obvious implications upon the second generation. . . . It also affected the parents' degree of fluency in English." The negative coefficients could also indicate that children act as interpreters for their parents, thereby reducing the benefits from the development of dominant language fluency and thus reducing the incidence of this skill. While children can serve this role in the household, they can hardly be claimed to fulfill the same function in the workplace.[11]

The emergence of multiculturalism in Australia in the 1960s would have lowered the non-labor-market benefits from learning English. Prior to the 1960s, Australia's "immigrants were expected to assimilate largely unaided, that is, to embrace wholeheartedly the Australian way of life and deny and forget their origins. By the 1960s, . . . there developed a greater acceptance of the role of language and cultural maintenance in facilitating settlement. This integration model envisaged that immigrants would adapt to a core of Australian institutions and values while maintaining their cultural traditions" (Committee to Advise on Australia's Immigration Policies, 1987, p. 14). Consequently, immigrant parents would be more likely to speak their mother tongue at home with their children, and the positive effect of children on parental English fluency would diminish as a result.

Country of birth is another important determinant of English language proficiency. Compared to the benchmark—immigrants from the English-speaking countries—each of the birthplace coefficients is negative and statistically significant. Moreover, the estimated coefficients are generally quite large. In table 1, column 1, among the Europeans, there is a clear distinction between those from northern Europe (language fluency 18.6 percentage points lower than the reference group), southern Europe (57.8 percentage points lower fluency), and eastern Europe (40.5 percentage points lower fluency). Immigrants from Arabic-speaking countries have a language fluency rate 50 percentage points lower than the English-speaking country group, other variables being the same.

[10] In addition to limited opportunities in public schools, there exist private day schools and after-school and weekend programs for transmitting the country-of-origin culture and language to native-born children.

[11] If children serve as interpreters in household and community matters, their presence would tend to depress their parents' English language fluency and would be associated with a depressing effect on earnings, other variables being the same. Tests suggest, however, that children have an insignificant effect on the earnings of adult male immigrants in Australia. An analysis of gender differences in Australia indicates that children have a less positive or more negative effect on the English language fluency of their mothers than their fathers (Chiswick and Miller 1995).

The four Asian variables indicate that exposure to English prior to migration has an important effect on language attainment in Australia. Thus, for immigrants from the Philippines (a region of considerable U.S. colonial influence and military bases), the rate of language fluency is only 13 percentage points lower than for the benchmark group; for those from South Asia (a region of British colonial influence), it is 15 percentage points lower than for the benchmark; while for immigrants from Vietnam (nearly all post-1975 refugees) and other Asian countries, the rate of English-language proficiency is substantially lower, with the deficit being 78.7 and 54.1 percentage points, respectively. Immigrants from South and Central America also seem to have a substantial language deficiency compared to immigrants from English-speaking countries (55.8 percentage points lower fluency). Thus, the English-language proficiency of immigrants appears to decrease with the linguistic distance of the mother tongue, increase with exposure to English in the origin country, and be greater among economic migrants than among refugees.

The minority-linguistic/birthplace concentration variable is added to the equation in table 1, column 2. The inclusion of this variable is associated with moderate reductions in the effect of being of European origin, but with only minor changes in the estimated coefficients of the other birthplace dichotomous variables (cf. table 1, cols. 1 and 2). The estimated coefficient of the concentration measure is negative (−0.047) and highly statistically significant (t-ratio = 5.35). The estimated effect suggests that an increase in the linguistic/birthplace composition of the area favorable to an immigrant by 1 percentage point would be associated with a reduction in the language fluency rate of immigrants of around 5 percentage points.[12] This seems to be a quite powerful effect. It is not reflecting subtle effects of birthplace not captured by the dichotomous variables. As will be shown below, the effect persists when the equations are estimated within major birthplace regions.[13]

[12] This interpretation is consistent with the linguistic environmental effect discussed in Sec. II above. In another context, Manski (1993) argues that such an effect may be difficult to separate from effects deriving from the characteristics of the constituents of the region. The definition of the concentration variable used in table 1 (birthplace) makes identification of the separate effects difficult. This is less problematic for the concentration measure used for the 1986 census (see table 3 below) as it is based on the minority languages spoken by immigrant and native-born Australians.

[13] The analysis reported in Chiswick and Miller (in press) shows that the minority-language concentration measure reflects the effects of foreign language media (newspapers, radio, TV) and the presence of other relatives in Australia and, hence, with formal ethnic networks. This research also demonstrates that the minority-language concentration measure is a reasonable proxy for these formal ethnic networks when more direct measures are not available. This allays concern over the specification used here.

Two interaction terms are added to the estimating equation in table 1, column 3. The negative coefficient on the age-duration interaction variable indicates that the effect of duration on English-language fluency is weaker the older the age at migration. The negative coefficient on the education-duration interaction indicates that the effect of duration on fluency is weaker for the better educated than for the less well educated. In other words, the English proficiency gap by level of education diminishes with a longer period of residence.

English-language fluency models were estimated separately for the sample of those not born in the English-speaking source countries. This is shown in table 1, column 5, for the non-English-speaking source countries as a group and in table 2 for the major regions of origin. The effects of schooling, duration, age at immigration, and whether married overseas are all larger when those from English-speaking countries are deleted from the data.

For each of the major birthplace groups, language proficiency is positively related to educational attainment (table 2). The estimated effect varies from around 2 percentage points higher fluency per year of education for the most fluent groups (northern Europe, South Asia) to 6–7 percentage points higher fluency per year of education for immigrants from Arabic-speaking countries and South and Central America, who have relatively low levels of language attainment. Overall, the simple correlation coefficient between mean level of fluency and the partial effect of education is −0.68, and this is statistically significant at the 10% level.

Age at migration has a significant negative effect on language fluency among those from non-English-speaking countries overall and in seven of the eight major source regions (table 2, col. 2). It has an insignificant positive coefficient only for Central and South America, but this is based on a very small sample (67 observations).

Language fluency improves with duration of residence for all birthplace groups, and the partial effect varies from around 1 percentage point per year of residence for the northern Europeans to 2.6 percentage points and 3.5 percentage points, respectively, for immigrants from eastern Europe and other Asian countries (evaluated at a duration of 10 years). Note the steep effect of duration on fluency for the eastern Europeans who are predominantly a refugee group. While those with lower fluency have a stronger effect of duration, the simple correlation coefficient (−0.43) is not significant at conventional levels. It is also noted that for three birthplace groups (eastern Europe, South Asia, and other Asian countries) there is evidence of a curvilinear relationship between language attainment and duration in Australia; fluency rises but at a diminishing rate with a longer residence.

Minority-language concentration is significant and negative overall and exists in four out of eight individual birthplace regions (table 2). It is negative

Table 2
Selected Regression Coefficients for English-Language Fluency by Place of Birth, Adult Foreign-born Men, Australia, 1981

Birthplace	% Fluent	Education	Age	YSM	YSM²*	Minority Concentration	Married Overseas	*N*
All non-English-speaking	58.16	.036	−.010	.017	...	−.053	−.087	4,166
		(15.23)	(11.33)	(18.33)		(5.75)	(4.71)	
Southern Europe	39.88	.031	−.016	.024	...	−.048	−.053	1,921
		(7.80)	(11.55)	(14.34)		(4.27)	(1.89)	
Northern Europe	87.29	.022	−.005	.008	...	−.043	−.182	850
		(5.09)	(3.71)	(5.23)		(1.48)	(4.01)	
Eastern Europe	63.45	.024	−.010	.035	−.048	.012	−.216	394
		(2.81)	(3.10)	(3.39)	(1.88)	(.10)	(2.95)	
Arab countries	45.51	.061	−.001	.015	...	.048	−.211	178
		(6.51)	(1.50)	(3.17)		(.77)	(2.63)	
South Asia	87.85	.025	−.001	.027	−.042	−.201	.128	288
		(4.06)	(.20)	(3.94)	(2.81)	(2.66)	(2.40)	
Other Asian countries	46.90	.042	−.012	.048	−.065	−.358	.125	145
		(4.58)	(2.47)	(3.67)	(1.98)	(2.72)	(1.13)	
Central and South America	38.81	.066	.007	.013	...	−.444	−.609	67
		(3.44)	(1.23)	(1.96)		(1.43)	(4.29)	
Africa	81.60	.036	−.008	.008	...	−.121	−.023	212
		(3.42)	(2.75)	(2.59)		(2.67)	(.31)	

SOURCE.—1981 Australian Census of Population and Housing, 1/100 sample of the Foreign Born.

NOTE.—Additional control variables are age, whether married, child < 6 only, child 6–17 only, children < 6 and 6–17, small urban, rural. Equations for Vietnam, Philippines, and the remainder group are not presented owing to small sample size: Philippines (mean fluency rate of 86.36, sample size = 22), Vietnam (mean fluency rate of 16.67, sample size = 48), and the remainder group (mean fluency rate of 90.24, sample size = 41). See the Appendix for variable definitions. *t*-statistics are in parentheses and were derived using White's (1980) heteroscedasticity-consistent covariance matrix estimator.

* Variable was divided by 100.

and insignificant in two cases and positive and insignificant for the remaining two birthplaces.

Finally, the foreign-marriage variable also performs satisfactorily within the disaggregated analysis. It is significant and negative both overall and in five out of the eight individual birthplaces.[14]

The 1986 census provides the opportunity to construct a minority-language concentration variable that more closely matches the conceptual variable. Table 3 presents results from estimation of the language model using these data.[15] In spite of some minor changes in the definitions of a number of the other variables, comparison of tables 1 and 3 reveals that the estimated effects are virtually identical.[16] For example, the partial effects of education in column 1 of table 1 (1981 data) is 0.025, and that in column 1 of table 3 (1986 data) is 0.023. Even though the age and duration of residence variables for the analysis of the 1986 census data have been created from interval data, the partial effects are identical to those derived from the 1981 census, where the data were provided in single years.[17] The similarity of the two sets of results attests to the robustness of the model.

Some caution is warranted when comparing the birthplace effects in the two analyses, owing to the somewhat different groupings of countries. However, for the southern Europe, Arabic, Vietnam, and Africa groups, the variables are the same. The estimated coefficients are again remarkably similar: southern Europe (−0.578 in 1981, −0.577 in 1986), Arabic (−0.499 in 1981, −0.457 in 1986), Vietnam (−0.787 in 1981, −0.831 in 1986), and Africa (−0.197 in 1981 and −0.116 in 1986). The model, therefore, appears to be quite robust.

The minority language concentration measure is added in table 3, column 2. This is negative (−0.075) and highly significant (t = 13.38). The estimated

[14] The foreign marriage variable is positive and significant only in the case of South Asia. This finding is not unique to Australia. In comparable analyses for the United States and Canada, the foreign marriage variable for South Asia also had positive although not statistically significant effects on language fluency compared to those who were not married, although the effect was negative for all other source regions (Chiswick and Miller 1992). This seemingly South Asian effect is undoubtedly reflecting an unmeasured variable. One explanation is the much higher rate of arranged marriages among those from South Asia. More so than for other countries, other variables being the same, postmigration marriage may involve a wife with the same non-English mother tongue who has only recently arrived in the destination.

[15] The mean values of the dependent variable LANG are similar in the two Census samples. It is equal to 0.760 in the 1981 data and 0.769 in the 1986 data.

[16] See the Appendix for these differences in definitions.

[17] In the table 3, col. 3, specification the interaction between age and duration of residence is insignificant, whereas it was highly significant in table 1. This may be attributable to the fact that both variables are available only in interval form in the 1986 data.

coefficient implies that in a region with 1% of the population speaking the same non-English language as the respondent, language proficiency will be a massive 7.5 percentage points lower than in a region in which none speak the language.[18]

Inclusion of the appropriately defined minority-language concentration variable (cf. table 3, cols. 1 and 2) in the estimating equation results in slightly greater changes to the birthplace effects than recorded in table 1. For example, the partial effect of a southern European origin declines by 13.8 percentage points, from −0.527 to −0.389, that for a Vietnamese origin from −0.831 to −0.753, and that for a South American origin from −0.600 to −0.554. It should be noted that Italian, Greek, Vietnamese, and Spanish are among the 12 languages used in the construction of the minority-language concentration variable. Yet the coefficients still indicate that language distance, exposure to English, and refugee status appear to explain the pattern of birthplace coefficients.

Table 3, column 5, reports the estimated equation for immigrants from non-English-speaking countries using the 1986 census. The equations were also estimated separately for each of the nine non-English-speaking origins identified in table 3. In summary, overall and for *each* of the nine non-English origin regions, education and duration have a positive effect on English-language fluency while age at immigration and the minority language concentration variable have a negative effect.

Table 4 examines the interrelationships between the minority-language concentration variable and education, age, and period of residence. The negative influence of the concentration variable (col. 4) is greater for the less well educated, for immigrants who arrive as adults, and for immigrants who have been in Australia for only a short time. Thus, ceteris paribus, for the very immigrants who have the lowest levels of language skills, the recently arrived, less educated, and older immigrants, living in a minority language enclave has a greater retarding effect on their acquisition of English language fluency.

V. Analysis of Earnings: Australia

The analysis of earnings is based on the standard human capital earnings function modified for immigrant adjustment (Chiswick 1978). The struc-

[18] The estimated partial effect, −0.075, is four times the size of that estimated for Canada and five times that estimated for the United States (Chiswick and Miller 1992). There are two factors that could be responsible for this stronger effect in Australia. First, Australian immigration is more intensely characterized by "waves" of immigrants from specific source countries than immigration in North America. This would provide a strong basis for language maintenance. Second, it is possible that the stronger measured effect of the language environmental factor reflects the greater emphasis on multiculturalism in Australia. That is, the "ethnics" may be more ethnic in Australia than elsewhere. This is consistent with one of the arguments advanced earlier concerning the (marginally) negative effect of children on their father's dominant language fluency.

Table 3
Regression Estimates of English-Language Fluency among Adult Foreign-born Men, Australia, 1986 (Dependent Variable: LANG)

Variable	Total Sample				Non-English-Speaking OLS (5)
	OLS (1)	OLS (2)	OLS (3)	Logit (4)	
Constant	.781	.779	.592	4.110	.242
	(28.17)	(28.77)	(10.29)	(9.67)	(5.09)
Education	.023	.022	.024	.205	.033
	(15.69)	(14.66)	(7.98)	(13.02)	(14.97)
Age	−.005	−.005	−.001	−.054	−.009
	(11.55)	(10.40)	(.64)	(10.68)	(11.12)
Years since migration (YSM)	.009	.008	.018	.108	.017
	(20.48)	(21.20)	(7.24)	(20.23)	(22.51)
Married	−.011	−.009	−.009	−.130	−.018
	(.97)	(.88)	(.86)	(1.06)	(.95)
Child < 6 years only	−.001	.003	.003	.117	.023
	(.03)	(.17)	(.18)	(.69)	(.90)
Child 6–17 years only	−.014	−.011	−.016	−.055	.023
	(1.56)	(1.25)	(1.81)	(.58)	(.90)
Children < 6 and 6–17 years	−.001	.001	−.003	.017	−.016
	(.01)	(.02)	(.24)	(.12)	(1.03)
Small urban location	.011	−.012	−.013	−.119	−.016
	(1.08)	(1.22)	(1.26)	(.81)	(.74)
Rural location	.042	.015	.014	.298	.027
	(3.51)	(1.26)	(1.22)	(1.69)	(1.17)
Minority language concentration	...	−.075	−.074	−.362	−.070
		(13.38)	(13.26)	(11.73)	(12.60)

Birthplace:					
Southern Europe	−.527	−.389	−.390	−5.836	...
	(44.33)	(24.53)	(24.62)	(18.20)	
Northern Europe	−.151	−.129	−.128	−4.217	.241
	(11.15)	(9.76)	(9.66)	(12.51)	(12.42)
Other Europe	−.351	−.337	−.333	−5.474	.065
	(17.17)	(16.64)	(16.47)	(16.85)	(2.64)
Arab countries	−.457	−.393	−.391	−5.525	.040
	(14.45)	(12.37)	(12.37)	(16.13)	(1.21)
Other Asian countries	−.285	−.250	−.250	−4.779	.170
	(16.20)	(14.58)	(14.64)	(14.77)	(7.31)
Vietnam	−.831	−.753	−.736	−7.461	−.276
	(26.54)	(24.10)	(23.52)	(15.52)	(7.62)
South America	−.600	−.554	−.556	−6.117	−.095
	(12.42)	(11.56)	(11.59)	(15.37)	(2.00)
Other American countries	−.051	−.042	−.043	−2.155	.366
	(3.31)	(2.79)	(2.79)	(3.24)	(13.90)
Africa	−.116	−.084	−.087	−3.591	.343
	(5.05)	(3.67)	(3.80)	(9.62)	(11.59)
Age*YSM/100	...	...	−.016	...	...
			(1.19)		
Education*YSM/100	...	...	−.018	...	...
			(4.27)		
N	7,194	7,194	7,194	7,194	4,152
R^2	.3979	.4254	.4268		.3268

SOURCE.—1986 Australian Census of Population and Housing, 1/100 Sample of the Foreign Born.

NOTES.—*t*-statistics are in parentheses and are derived using White's (1980) heteroscedasticity-consistent covariance matrix estimator. See the Appendix for variable definitions. OLS = ordinary least squares.

Table 4
Selected Regression Coefficients for English Fluency Model with Minority Language Concentration Interaction Terms, Adult Foreign-born Men, Australia, 1986 (N = 7,194)

Variable	(1)	(2)	(3)	(4)
Education	.022	.015	.016	.016
	(14.66)	(10.09)	(10.58)	(10.39)
Years since migration (YSM)	.008	.009	.009	.007
	(21.20)	(21.19)	(21.19)	(16.55)
Age	−.005	−.004	−.003	−.003
	(10.40)	(10.10)	(8.12)	(6.95)
Minority language concentration (CONC)	−.075	−.153	−.057	−.131
	(13.38)	(10.67)	(2.12)	(4.83)
CONC*Education	...	.008	.006	.007
		(5.85)	(4.39)	(4.54)
CONC*Age	...	...	−.002	−.003
			(4.18)	(6.59)
CONC*YSM	...	...	...	.005
				(9.18)
R^2	.3979	.4301	.4324	.4428

SOURCE.—1986 Australian Census of Population and Housing, 1/100 Sample of the Foreign Born.
NOTE.—These are selected regression coefficients gained from adding interaction variables to the table 3 equations. In addition to the variables listed, all other control variables used in table 3 are included in these equations. See the Appendix for variable definitions. Partial derivatives [from col. (4)] evaluated at sample means are

$$\delta\text{LANG}/\delta\text{EDUC} = 0.016 + 0.007\ \text{CONC} = 0.020,$$

$$\delta\text{LANG}/\delta\text{YSM} = 0.007 + 0.005\ \text{CONC} = 0.010,$$

$$\delta\text{LANG}/\delta\text{Age} = -0.003 - 0.003\ \text{CONC} = -0.005,$$

and

$$\delta\text{LANG}/\delta\text{CONC} = -0.131 + 0.007\ \text{EDUC} - 0.003\ \text{Age} + 0.005\ \text{YSM} = -0.075.$$

tural model of the determinants of earnings among adult immigrant men assumes that, aside from current marital status, household characteristics (such as whether married prior to or after immigration and the number and age structure of the children) do not directly enter the earnings function and that, with competitive labor markets, wage rates for workers of a given level of skill are invariant with the size of an ethnic/linguistic enclave. These are implicitly standard assumptions in the immigrant adjustment literature.

As a result, the earnings equation specifies that the natural logarithm of earnings is a function of education, total potential labor market experience, duration in the destination, marital status, Australian citizenship, size of place, and country of birth. It is hypothesized that, with the exception of the birthplace variables, all of these partial effects are positive. These variables are defined, and their means and standard deviations are reported in the Appendix.

The 1981 Australian Census information on earnings was collected and released in the Public Use Sample File in 14 broad brackets, and in the case of the 1986 Census the data were aggregated to only eight intervals in the Household Sample File (Section of State). While there are statistical techniques that may be used to accommodate the grouping of income data (see Stewart 1983), a previous application of the technique using the 1986 census [Miller 1989]) reveals that there are few gains from doing so. Accordingly, the analyses in this section are based on a dependent variable formed from the midpoints of the income intervals and by using a value of 1.5 times the lower threshold for the open-ended upper limit. This is the procedure employed in previous research based on the 1981 Australia census by Chiswick and Miller (1985) and Stromback (1986).

Previous research (Chiswick and Miller 1985) suggests that the curvilinear relationship between earnings and duration of residence found in earnings functions for immigrants in other immigrant-receiving countries (e.g., the United States, Canada, and Israel) is not evident in the Australian data. Hence, only a linear duration of residence term is included in the estimating equations.

The basic regression equation is reported in table 5, column 1, for the adult foreign-born men in the 1981 Australian census. The partial effects are all statistically significant, have the hypothesized signs, and are consistent with other studies. Table 5, columns 2–5, address the issue of the effect of English-language fluency on the earnings of immigrants in Australia. The observed dichotomous English-fluency variable is added in column 2 and replaced in column 3 by a predicted measure using an instrumental variables (IV) approach.[19] In columns 4–7, the analyses are done separately for those fluent in English and those not fluent, where columns 4 and 6 are OLS equations, and columns 5 and 7 are equations corrected for the potential selectivity bias in such a dichotomy of the data.

Reading across the columns in table 5, it is apparent that earnings rise among immigrants in Australia by about 6% per year of schooling for immigrants as a whole, but the effect is larger (8%) for those fluent in English and smaller (about 2%) for those not fluent in English. Earnings increase, but at a decreasing rate with a rise in total potential labor market experience, where the increase in table 5, column 1, is about 1.3% per year for the first year and 0.8% when evaluated at 10 years.

The earnings of immigrants increase by about 0.4% per year in the country, and this effect is highly statistically significant.[20] Among those not

[19] The instruments are all of the explanatory variables in the language and earnings equations. The identifying instruments are whether married overseas, number and age of children, and the birthplace concentration variable. The predicted value has a much smaller standard deviation than the observed value, and this results in a much larger standard error of the estimate.

[20] This effect of duration in Australia on earnings is smaller than what is found

fluent in English there is no significant effect of duration. Note, however, that part of the effect of schooling, experience, and years since immigration is to raise the level of fluency of immigrants. Martial status also matters. Earnings are about 12% higher for those currently married, regardless of fluency in English.

Earnings vary systematically by size of place. In the full sample, earnings are about 6% lower in small urban areas and about 20% lower in rural areas compared to the large urban areas. Among those not fluent in English, however, living in a small urban area results in earnings about 15 percentage points lower than in the large urban area. The partial effects of the size of place variables may be reflecting equalizing wage differentials, unmeasured differences in the cost of living or in immigrant skill.

Australian citizenship, in contrast, provides no additional earnings, when other variables including country of origin are the same. This finding is consistent with previous research (see, e.g., Chiswick and Miller 1985) and the observation by the Committee to Advise on Australia's Immigration Policies (1988, p. 11) that "citizenship is of little material value."

It is premature to link the absence of a pecuniary return to citizenship to the low propensity among immigrants to naturalize. The mean citizenship rate in Australia (56%) exceeds that in the United States (48%), even though analyses for the United States indicate a 5 percent earnings premium associated with citizenship (Chiswick and Miller 1992). Perhaps the other factors to which the Committee to Advise on Australia's Immigration Policies (1988) drew attention may be responsible. These include the low symbolic value of citizenship in Australia, the absence of links between citizenship and welfare entitlements, the absence of special privileges granted citizens for sponsoring relatives (as is the case in the United States and Canada), and the special privileges granted British citizens.

The estimated birthplace effects are measured relative to the earnings of British immigrants. They fall into two groups. The first group comprises immigrants from the other English-speaking countries, for whom the estimated effects are not statistically significant. This result is not surprising for immigrants from Ireland, Canada, the United States, and New Zealand.[21]

---

in the United States and Canada but consistent with other studies of Australia (see Chiswick and Miller 1985, 1988, 1992).

[21] For the immigrants from the British West Indies, however, the finding contrasts with the evidence from Britain and North America (Chiswick 1980; Chiswick and Miller 1992), which indicates significantly lower earnings for (Black) West Indian immigrants. However, the sample of West Indian immigrants in Australia is very small (only six observations), and given the greater cost involved in migrating to Australia compared with the United States, Canada, and the United Kingdom, the West Indian immigrants in Australia are also likely to be a highly select group. Furthermore, the racial composition of these immigrants is not known.

The estimated partial effects of birthplace for the second group, the non-English speaking countries, are all negative, and, with the exception of the small and heterogeneous remainder group, they are all statistically significant. Moreover, the partial effects are essentially invariant with respect to whether an English-fluency variable is added to the equation. The differential is "small" (13% or smaller) for South Asia, Africa, other Asian countries, northern Europe, and eastern Europe. There is, however, a "large" differential (20% or more), other variables being the same, for those from the Philippines, South America, southern Europe, Arab countries, and Vietnam.[22]

The basic estimating equation is augmented with the measure of dominant-language fluency in table 5, column 2. English-language facility is associated with a statistically significant 5.3% higher earnings ($t = 2.54$). For the sample of immigrants from non-English-speaking countries, the effect of language fluency in a specification similar to that in column 2 is 6.4 percent ($t = 2.83$). Analysis of the limited income data released in the 1986 Unit Record Sample File reveals a partial effect of English-language fluency of 8.3% ($t = 4.75$) for the total sample and 9.3% ($t = 5.10$) for the portion of the sample from non-English-speaking countries. It appears there has been a rise in the premium for English language skills over the 5 years.[23] In other words, dominant-language fluency is a skill that is rewarded in the Australian labor market, and its importance appears to have increased in the early 1980s.

A test was performed to determine whether being in a minority/ethnic enclave has an effect on the earnings of immigrants independent of the person's own level of fluency. The birthplace concentration measure used in the language analysis and an ethnic concentration measure were each found to be statistically insignificant.[24]

[22] Among the Vietnamese, the earnings differential is quite marked; the estimated coefficient of −0.557 implies an earnings differential of about 43 percentage points. Nearly all of the Vietnamese, however, were in Australia for less than 5 years, and this effect may not have been fully controlled by the linear duration of residence variable. Refugees would be expected to have particularly low earnings during the initial period of adjustment.

[23] This difference in the returns to language fluency is not due to the different grouping of the income data. Repeating the analysis of the 1981 census data after deriving the dependent variable from 8 rather than 14 intervals results in only a slight increase in the estimated effect on the dominant language fluency variable, from 0.053 ($t = 2.54$) to 0.057 ($t = 3.40$).

[24] The ethnic concentration measure was based on father's country of birth. Birthplace regions were combined into 26 ethnic groups, e.g., Germanic (Austria and Germany), Arabic (Lebanon and other Middle Eastern countries), Spanish (Spain and Latin America), French, and Italian. The proportions of the population in each region with fathers of Germanic, Arabic, Spanish origins, etc., were computed, though in cases where the specific country was not identified (e.g., other

Table 5
Regression Estimates of Earnings Equations, Adult Foreign-born Men, Australia, 1981
(Dependent Variable: Natural Logarithm of Annual Income)

	Fluent Total Sample			Not Fluent in English		Fluent in English	
Variable	OLS (1)	OLS (2)	IV (3)	OLS (4)	Selectivity Corrected (5)	OLS (6)	Selectivity Corrected (7)
Constant	8.625	8.586	8.805	8.390	8.383	8.832	8.787
	(138.21)	(136.17)	(55.29)	(108.20)	(132.51)	(71.24)	(79.31)
Education	.061	.060	.066	.076	.077	.020	.011
	(19.66)	(19.14)	(14.03)	(19.40)	(23.23)	(4.02)	(1.94)
Experience	.013	.013	.013	.020	.020	.008	.012
	(4.44)	(4.48)	(4.44)	(5.37)	(6.08)	(1.32)	(1.84)
Experience^2/100	.027	−.027	−.029	−.039	−.039	−.016	−.017
	(4.92)	(4.82)	(5.37)	(5.52)	(6.53)	(1.77)	(1.69)
Years since migration	.004	.003	.006	.003	.004	.003	−.004
	(4.00)	(3.34)	(2.88)	(3.21)	(3.61)	(1.12)	(1.28)
Small urban	−.065	−.066	−.057	−.053	−.053	−.138	−.159
	(2.92)	(3.00)	(2.69)	(2.41)	(2.46)	(1.67)	(2.85)
Rural	−.196	−.197	−.190	−.222	−.221	−.009	−.031
	(5.98)	(6.01)	(7.68)	(6.46)	(8.64)	(.09)	(.48)
Married	.122	.123	.115	.117	.116	.147	.172
	(6.37)	(6.43)	(6.17)	(5.95)	(6.12)	(2.44)	(3.99)
Citizen	−.011	−.012	−.005	−.008	−.008	−.017	−.011
	(.72)	(.80)	(.31)	(.47)	(.48)	(.52)	(.38)
Ireland	−.049	−.049	−.048	−.053	−.053	...	...
	(1.63)	(1.64)	(1.15)	(1.77)	(1.29)		
Canada	−.076	−.074	−.086	−.104	−.105	...	...
	(.87)	(.85)	(.91)	(1.18)	(1.13)		
United States	.018	.019	.011	−.032	−.033	...	...
	(.18)	(.20)	(.18)	(.33)	(.53)		
New Zealand	.049	.048	.055	.039	.040	...	...
	(1.41)	(1.37)	(1.62)	(1.12)	(1.20)		

British West Indies	−.100	−.100	−.099	−.135	−.135	...	...
	(.63)	(.62)	(.45)	(.87)	(.62)		
Southern European	−.211	−.180	−.353	−.177	−.197	...	...
	(12.10)	(8.36)	(2.95)	(7.32)	(4.90)		
Northern Europe	−.101	−.091	−.148	−.103	−.110	.159	.029
	(4.20)	(3.67)	(3.27)	(3.87)	(4.20)	(2.61)	(.41)
Eastern Europe	−.131	−.109	−.231	−.130	−.145	.156	.124
	(4.06)	(3.13)	(2.59)	(2.83)	(3.30)	(4.20)	(2.47)
Arab countries	−.275	−.248	−.398	−.234	−.252	−.084	−.119
	(6.65)	(5.70)	(3.59)	(3.18)	(3.75)	(1.98)	(2.08)
South Asia	−.055	−.047	−.092	−.078	−.084	.153	−.039
	(1.94)	(1.65)	(1.99)	(2.53)	(2.23)	(2.06)	(.34)
Philippines	−.190	−.183	−.222	−.240	−.246	.302	.120
	(2.15)	(2.07)	(1.87)	(2.37)	(2.00)	(2.12)	(.39)
Vietnam	−.557	−.515	−.749	−.268	−.295	−.267	−.228
	(3.95)	(3.66)	(4.20)	(3.46)	(1.53)	(1.55)	(2.57)
Other Asian countries	−.121	−.092	−.255	−.163	−.182	.227	.217
	(2.30)	(1.66)	(2.12)	(1.72)	(2.52)	(4.16)	(3.42)
South and Central America	−.207	−.177	−.344	−.254	−.273	.123	.103
	(3.24)	(2.67)	(2.59)	(1.73)	(2.50)	(2.72)	(1.22)
Africa	−.084	−.073	−.134	−.075	−.083	.094	−.053
	(2.56)	(2.21)	(2.33)	(2.03)	(1.86)	(1.79)	(.53)
Remainder group	−.048	−.039	−.087	−.018	−.024	−.068	−.251
	(.68)	(.56)	(.95)	(.24)	(.27)	(.60)	(.94)
LANG	...	.053	−.243	...	...	...	...
		(2.54)	(1.20)				
λ	...	...	...	...	.025	...	−.180
					(.62)		(2.84)
R^2	.1479	.1487		.1460	.1459	.0484	.0523
N	7,288	7,288	7,288	5,540	5,540	1,748	1,748

SOURCE.—1981 Australian Census of Population and Housing.

NOTE.—t-statistics are in parentheses and are calculated using White's (1980) heteroscedasticity-consistent covariance matrix estimator. OLS = ordinary least squares; IV = instrumental variables. See the Appendix for variabe definitions.

The analysis with the instrumental variables measure of fluency is presented in table 5, column 3. The dominant language fluency effect is negative and statistically insignificant, the t being only 1.20. This method of estimation is sensitive to the choice of instruments, and there is often little to guide the choice of "good" instruments that will yield the minimum asymptotic variance. Comparison of the OLS and instrumental variables estimates listed in columns 2 and 3, respectively, reveals that a number of birthplace coefficients (in particular, those for southern Europe, Arab countries, Vietnam, other Asian countries, South and Central America) change considerably, indicating a pronounced widening of the inter-birthplace wage differential under the instrumental variables method.

An analysis of the limited income data available in the 1986 Census adds further insights. First, it suggests that the widening in the inter-birthplace wage differentials when the instrumental variables method is used to obtain the English language fluency variable (table 5, col. 3) is due to the use of a minority concentration variable based on birthplace (1981 census) as an identifying variable, rather than one based on the preferred measure of minority languages (1986 census).

Second, in the 1986 census analysis using instrumental variables, the dominant language fluency variable becomes positive (0.043), although it is not statistically significant (t-ratio = 0.52). Thus, as might be expected on the basis of econometric theory (see, e.g., Koutsoyiannis 1977), the instrumental variable results are unstable. This emphasizes the importance of using alternative data sets to test the effects of language on earnings to determine the pattern of bias introduced into OLS estimation by selectivity. This is done in Section VI below.

Analyses for the sample fluent in English are presented in table 5, columns 4 and 5. As the selectivity correction factor (λ) is statistically insignificant, the two sets of results are quite similar. They reveal that, among groups fluent in English, the partial effect of education on earnings is higher than reported for the pooled analysis (7.7% compared to 6.1%). Similarly, the partial effect of preimmigration experience is marginally higher for the group proficient in English than for the aggregate-level results (1.2% compared with 0.7% when evaluated at 10 years of experience). The effect of duration of residence in Australia is, however, the same for the various language-fluency groups under investigation.

The findings for the 24% of the sample lacking dominant language fluency are presented in table 5, columns 6 and 7. The effect of education, other variables being the same, is quite low and indicates that each additional year of education is associated with only 1.1% higher earnings. This

European countries) the ethnic concentration index was assumed to be zero. The value of the ethnic concentration index was based on the respondent's birthplace and region in Australia.

low partial effect, and the contrast with the 7.7% effect for the group fluent in English, may be indicative of a complementarity between the skills represented by formal education and language. A similar interpretation may apply to the relatively low effects of preimmigration experience (0.86 percentage points, which is marginally lower than the 1.2 percentage point effect per year of experience estimated for the group possessing English-language fluency).

The duration of residence variable becomes statistically insignificant in table 5, columns 6 and 7. In other words, labor market experience in Australia does not attract an earnings premium if the individual is not fluent in English. This is consistent with one of the explanations generally offered for the positive relationship between earnings and duration of residence—acquiring skills relevant for and learning about the institutions of the labor market. This learning is presumably impeded by inadequate facility in English.

A final feature of the results in table 5, column 7, is that the coefficient on the selectivity correction term (λ) is negative and statistically significant ($t = 2.84$). The negative coefficient implies positive selection into the nondominant language fluency market.

To complete the study of the relation between language and earnings among immigrants, tests of endogeneity based on the selectivity corrected estimates were conducted. The first test was the conventional asymptotic t-test on the selection terms included in the table 5, columns 5 and 7, results. These tests did not reject exogeneity in the English-fluent group, but they did reject exogeneity in the case of the group lacking English fluency. In view of these mixed findings, a second test was performed. Following Robinson (1988, 1989), the tests based on the selectivity correction terms were expressed in terms of an F-test of the joint significance of the two sample selection terms in an equation estimated on data pooled across those fluent and those not fluent in English. The computed F-statistic was 5.33, which exceeded the critical value of 3.0. Thus there is evidence that dominant language fluency is endogenous to the process of earnings determination among immigrants.

VI. International Comparisons

This section compares the findings for Australia with parallel econometric analyses of census microdata for adult male immigrants (aged 25–64 years) in the United States (1980 census, 1/100 sample), Canada (1981 census, 1/50 sample) and Israel (1983 census, 1/20 sample). This comparison demonstrates the robustness of the model.

The language question in the U.S. 1980 census is closest to that used in Australia. The U.S. data are from the self-reported responses to the question, "How well does this person speak English?" (very well, well, not well, or

not at all). Those who speak only English or report speaking it well or very well are treated as fluent in English in this analysis.

The Canadian language question is, "Can you speak English or French well enough to conduct a conversation?" The responses are English only, French only, both English and French, and neither English nor French. Those who responded they could carry on a conversation in either of the two official languages are considered fluent. The U.S. analysis was also computed using a definition closer to the Canadian concept, where all but those who speak English "not at all" are considered fluent.

In the Israel 1983 census, respondents were asked, "What languages do you speak daily? Do not report a language that you know if you do not speak it daily." The response categories delineated a primary and, if one existed, a secondary language. Those who reported Hebrew as their only or primary language were counted as fluent.

While the detailed econometric analyses can be found elsewhere (Chiswick 1993; Chiswick and Miller 1992), table 6 compares the analysis of the determinants of destination-language fluency for the United States, Canada, and Israel with the findings from above for Australia.[25] The most striking feature is the similarity of the results—the coefficients are statistically significant, have the same hypothesized signs, and have very similar magnitudes.

Education raises fluency by 1.0–2.4 percentage points per year of schooling, except for Canada, where the smaller effect seems to be the result of a more lax definition of fluency. Age at migration (age when years since migration is held constant) decreases fluency by 0.4–0.6 percentage points per year of age, again except for the Canadian definition of fluency. Duration raises fluency (when evaluated at 10 years) by 0.5 percentage points per year in Canada, 0.8 percentage points per year in Australia, 1.4 percentage points per year in the United States, and in the predominantly refugee population in Israel by 2.6 percentage points per year. Those who married prior to migration have significantly lower levels of fluency in each of the four countries.[26]

For all four countries the minority language concentration measure is based on nondominant languages spoken in the person's region of residence. It is negative and significant in all instances, ranging from −0.014 (United States) to −0.075 (Australia, 1986).

[25] Beenstock (1993) uses panel data from the 1970s on immigrants in Israel for up to 3 years to analyze Hebrew fluency and literacy, as well as employment, unemployment, and occupational convergence. His data lack information on wages or earnings. Beenstock's findings are consistent with those reported here for Israel.

[26] The language acquisition model offered no hypothesis for those who married after migration compared to the unmarried. The partial effect of marriage after immigration compared to those who are not married is found to vary in sign and significance across the four countries.

Table 6
Partial Effects of Selected Variables on Dominant Language Fluency among Adult Male Immigrants, Australia, United States, Canada, and Israel

Variable	Australia		United States		Canada	Israel
	1981	1986	1980	1980, Canadian Definition	1981	1983
Education	.024	.022	.027	.009	.006	.010
Age	−.005	−.005	−.004	−.001	−.001	−.006
Years since migration[a]	.009	.008	.014	.006	.005	.026
Married	−.009*	−.009*	.012	.009	−.001*	.024
Married overseas	−.039	. . .	−.035	−.012	−.013	−.083
Minority language concentration	−.047	−.075	−.014	−.005	−.018	−.014
Proportion fluent[b]	.76	.77	.80	.95	.97	.84

SOURCES.—For Australia, see table 1 (col. 2) and table 3 (col. 2); for the United States/Canada, see Chiswick and Miller (1992), table 7; for Israel, see Chiswick (1993), table 4.
[a] Evaluated at 10 years.
[b] Proportion fluent in the dependent variable, based on country-specific language variables and definitions.
* Estimated effect not statistically significant at 5% level.

The Australian analysis indicates important interaction effects between the concentration measure and education, age and duration (table 4). Similar patterns emerge for the other countries (table 7). Living among others who speak the same language of origin has a greater depressing effect on fluency among those with less education, who immigrated at an older age, and who have arrived more recently.

The measures of the presence of children in the family differ among the countries. The findings also differ. Children are found to have no significant effect on adult male fluency in Australia or Canada,[27] but a significant

Table 7
Sign and Significance of the Minority Language Concentration Interaction Variables in Analyses of Language Fluency: Australia, United States, Canada, and Israel

	Australia	United States	Canada	Israel
CONC	−	−	−	−
EDUC*CONC	+	+	+	+
AGE*CONC	−	−	−	−
YSM*CONC	+	+	+	+

SOURCES.—For Australia, see table 4, col. 4, and unpublished table for 1986 census; for the United States and Canada, see Chiswick and Miller (1992); for Israel, see Chiswick (1993).
NOTE.— +: positive and significant; −: negative and significant. Education, age, duration, marital status, children, location, and country of origin are controlled for. See the Appendix for variable definitions.

[27] This finding for Canada is based on an analysis of the 1/100 Household and

positive effect for the United States, and a larger and very highly significant effect for Israel.[28]

The equations were also estimated separately for the major source regions of immigrants for the United States, Canada, and Israel, as was done for Australia. Although the analyses are not reported here, the basic patterns within major source regions for the effects of education, age, duration, whether married overseas, and the language concentration measure are the same. They tend to have the expected signs and are statistically significant.

Moreover, there are systematic differences in the degree of fluency by country of origin, when other variables are the same. Destination-language fluency varies inversely with the probability of return migration. Fluency rates are low, ceteris paribus, among Mexican immigrants in the United States and North American immigrants in Israel, two groups with very high rates of return migration. Fluency rates are lower among refugees than among economic migrants. For example, fluency is lower among Cuban, Vietnamese, and Chinese immigrants (who are disproportionately refugees) relative to South American and other Asian immigrants in the United States and Canada, and lower among Soviet refugees than among other European immigrants in Israel. Fluency rates are also lower the greater the linguistic distance of the immigrant's language of origin from the destination language. Indeed, this may explain the higher level of Hebrew-language fluency of North African Jews in Israel, whose country-of-origin language, Arabic, is linguistically closer to Hebrew than any of the other immigrant languages.

Table 8
Partial Effect on Earnings of Proficiency in the Dominant Language, OLS and IV Estimation, Australia, United States, Canada, and Israel

	Australia				
Method	1981	1986	United States, 1980	Canada, 1981	Israel, 1983
OLS	.053	.083	.169	.122	.110
	(2.54)	(4.75)	(12.52)	(2.43)	(12.66)
IV	−.243	.043	.571	.414	.351
	(1.20)	(.52)	(5.43)	(1.34)	(4.25)

SOURCES.—For Australia, see table 5 and unpublished table for 1986 Census (available from us on request); for the United States and Canada, see Chiswick and Miller (1992), tables 8 and 9; for Israel, see Chiswick (1993), table 8.

NOTE.—Schooling, experience, years since immigration, marital status, country of origin, region variables, and for the United States and Canada weeks worked are controlled for. *t*-ratios are in parentheses. OLS = ordinary least squares; IV = instrumental variables.

Family File of the 1981 census of Canada discussed in Chiswick and Miller (1992, p. 246).

[28] Perhaps this is confirmation of the Israeli myth that immigrant parents learn the nation's language (Hebrew) from their children.

The four country analyses of the effects of language fluency on earnings (OLS and IV) are summarized in table 8. In the OLS analysis the partial effect of the language variable is positive and statistically significant. The magnitude is lower in Australia (5% and 8% in the 1981 and 1986 Census data, respectively) than in Israel (11%), Canada (12%), and the United States (17%). The coefficient in the instrumental variables analysis shows considerable instability. It is positive and statistically significant in the United States and Israel, but the magnitudes are surprisingly large.

The responsiveness of dominant language fluency to these earnings gains was examined in the analyses for the United States and Canada using the sample selectivity tests reported above for Australia (table 9). For each country analysis, as with the Australian study, the assumption that the dominant language fluency variable in the wage equation is exogenous is rejected (Chiswick and Miller 1992).

Further support for the robustness of the model is provided in the analysis by Dustmann (1994) of fluency and earnings among southern European and Turkish immigrants in Germany using the 1984 wave of the German Socio-Economic Panel. The analysis for about 1,000 male immigrants is not strictly comparable to the analyses reported above because of differences in the questions asked in the German survey, the coding of variables, and the specification of the estimating equations. Nonetheless, the observed patterns are the same. A younger age at entry, a longer duration of residence, and a higher level of schooling are all associated with greater fluency in German among immigrants in Germany. Those who are married to a German wife are the most fluent, while those married to a non-German wife are less fluent than the single immigrants. School-age children have a marginally significant positive effect on the father's fluency. A minority-language concentration index could not be constructed with the German data.

By country of origin, ceteris paribus, those from Yugoslavia are the most fluent, followed by immigrants from Greece and Italy, then Spanish immigrants, and finally Turkish immigrants, who are the least fluent. Note that, among these five origins, Yugoslavia has had the greatest Germanic

Table 9
Significance of Selectivity Correction Terms for Analyses of Language Fluency: Australia, United States, and Canada

Sample	Australia	United States	Canada
Fluent in dominant language	no	yes	yes
Not fluent in dominant language	yes	no	yes
Pooled sample	yes	yes	yes

SOURCES.—For the United States and Canada, see Chiswick and Miller (1992); for Israel, see Chiswick (1993); for Australia, analyses are from unpublished tables from the 1981 and 1986 censuses.
NOTE.—yes = significant; no = not significant.

influence over the past century, including being part of the Austro-Hungarian Empire. Moreover, Turkish is the most distant linguistically from German, and those of Turkish origin are the most numerous of the immigrants.

Dustmann (1994) also analyzes the effect of speaking fluency on earnings in an OLS equation. He finds that adult male immigrants who report they speak German very well or well earn 6.3% more than those who speak it badly or not at all, while those who speak at an intermediate level earn only a marginally significant 3.8% more than the less fluent group.

VII. Summary and Conclusion

Our study was concerned with the determinants of dominant-language fluency among immigrants and the labor market effect of this fluency. The analysis explored the endogeneity between dominant language fluency and earnings. The empirical tests were conducted primarily for Australia, with international comparisons to the United States, Canada, and Israel.

We developed a theoretical model of language fluency, where fluency is a rising function of economic incentives for acquiring fluency, exposure to the destination language, and efficiency in achieving destination-language proficiency. *Economic incentives* include the increment in annual earnings with greater fluency and the expected duration in the destination. *Exposure* includes the preimmigration experience with the destination language, duration in the destination (learning by living), and the intensity of exposure per unit of time (i.e., the extent of postmigration training in the destination language), and destination language usage in the area and the household in which the immigrant lives. *Efficiency* refers to the extent to which exposure produces language fluency and is related to the level of other skills (e.g., schooling) and biological/maturational factors (i.e., age). Empirical counterparts are developed for the theoretical variables.

The model was tested, and the parameters were estimated for foreign-born men aged 25–64 years using large microdata files from the censuses of Australia (1981 and 1986), Canada (1981), Israel (1983), and the United States (1980). These international comparisons demonstrate the robustness of the model. The analyses were reported in detail for Australia using the 1981 and 1986 1/100 samples. These two data sources were used because of the somewhat different attractive features of the files and to test robustness.

The empirical analysis is very robust. In each of the four countries, the following patterns hold overall and when they are done separately by region of birth. Fluency in the destination language (English in Australia and the United States, English or French in Canada, and Hebrew in Israel) increases with exposure. It is greater the greater the use of the destination language in the country of origin, the longer the duration of residence, the smaller the proportion of people in the immigrant's area of residence

who speak his mother tongue, and if the spouse does not have the same mother tongue. Fluency increases with efficiency in language acquisition; it increases with the level of schooling and decreases with age at immigration. Other variables being the same, fluency rates are lower the greater the probability of return migration, the greater the linguistic distance between the origin and destination languages, and among refugees than among economic migrants.

Living among others who speak the same nondestination language as the immigrant retards the acquisition of destination-language fluency in a manner that varies with other characteristics. A minority-language enclave has a greater depressing effect on destination-language fluency among the immigrants with lowest levels of fluency, that is, those recently arrived, the less well educated, and those who immigrate at an older age.

The analysis of earnings uses the now standard human capital earnings function model augmented for immigrant adjustment. In the Australian 1981 census, English-language fluency is shown to be associated with a statistically significant 5.3% higher earnings, which increases to 6.4% for those from non-English-speaking countries. For the Australian 1986 census the effects are larger, 8.3% and 9.3%, respectively, suggesting an increase over time in the returns to English-language skills. The effects of fluency in the destination language are even larger in the United States (16.9%), Canada (12.2%), and Israel (11.0%).

These estimates of the labor market benefits of language fluency can be used to estimate the rate of return on the investment if there are also cost estimates. We can use as a guide the Israeli "ulpan" system for new immigrants that usually involves 6 months of intensive Hebrew language training to bring adults to a modest level of fluency and literacy (reading and writing). Suppose fluency for adults costs the equivalent of 6 months to 1 year of full-time potential earnings. Then, overall, for permanent immigrants (i.e., assuming a long working life) estimated rates of return based on labor market earnings would vary from about 9%–18% for Australia, 11%–22% for Israel, 12%–24% for Canada, and 17%–34% for the United States. The rate of return would, of course, vary by age; it would be higher for young adult permanent immigrants, as they learn language skills more rapidly and have a longer remaining work life, than for older immigrants. These estimates ignore the consumption benefits from acquiring fluency in the destination language. Investments in language fluency, therefore, appear to be very profitable for immigrants who are not fluent in the dominant language of the four destinations we have studied.

Comparisons of the coefficients on the schooling and experience variables in samples of those fluent and not fluent indicate the complementarity among skills. That is, the effects on earnings of schooling and labor market experience are much larger for those fluent in the destination language.

We implemented various procedures to test for the endogeneity of language skills. These tests indicate that those who anticipate higher earnings for unmeasured reasons if they were to become fluent are more likely to acquire destination language fluency.

Our model has been shown to be applicable for a wide range of source countries and immigrant receiving countries. The findings indicate that the determinants of destination language skills among immigrants can be studied using econometric techniques, that fluency responds to incentives (economic, exposure, and efficiency), that language skills have an important effect in the labor market, and, finally, that earnings and language fluency are determined jointly.

Appendix

Definitions of the Variables and Descriptive Statistics

I. Analysis of 1981 Australian Census of Population and Housing

The Australian Bureau of Statistics released two public use samples from the 1981 Census of Population and Housing: the 1/100 Persons Sample File and the 1/100 Households Sample File. These differ in terms of the target population (persons vs. households) and also in the extent of geographic information provided and the degree of detail contained in the classification scheme used for some characteristics. All geographic references other than for a code showing major urban/other urban/rural region of residence were removed from the Household Sample File to ensure that the confidentiality of individuals is protected. In the case of the Persons Sample File, some data at the state level are included, but at the cost of a reduction in the number of categories used when classifying birthplace (40 compared to 101) and period of residence (only five broad intervals for valid codes). Some of the birthplace classifications in the Persons File are less useful for analysis of linguistic effects. For example, Indonesia is grouped with the Philippines; South Africa is coded with Egypt. Furthermore, information on the duration of marriage of males and the number of children is unavailable in the Persons File. For these reasons the Households File is used in this study.

A. Definitions

Population.—These are foreign-born men aged 25–64 who were employed as wage, salary, or self-employed workers (excluding unpaid helpers) at the time of the census. The analysis is restricted to individuals living in private dwellings and who were members of the primary family in such dwellings (i.e., all single-family noninstitutional households and the primary family in multifamily households).

Earnings (LNEARN).—Respondents were asked to report the gross income (including pensions and/or allowances) that they usually receive each week from all sources. The answer was given simply by placing a checkmark in a box corresponding to 14 weekly income categories. This

was converted to a usual yearly income by the Australian Bureau of Statistics by multiplying the weekly figures by 52. Hence the data are standardized for weeks worked. For the open-ended upper-income interval (over $26,000 per year) a mean value of $39,000 is imposed.

Years of Education (EDUC).—This variable records the total years of full-time education. It has been created from the census "Age Left School" and "Qualifications" variables. Years of education is calculated as "Age Left School Minus 5." Individuals who stated a school-leaving age of 19 or more years were assigned 13 years of education. Individuals who possess a diploma have been assumed to have the equivalent of 15 years of full-time education, individuals who possess a bachelors degree the equivalent of 16 years of full-time education, individuals who possess a graduate diploma have been assumed to have the equivalent of 17 years of full-time education, and individuals who have a higher degree (master's, Ph.D.) have been coded as having 19 years of education.

Years of Experience (EXP).—This is computed as age minus years of education minus 5 (i.e., EXP = AGE − EDUC − 5). A quadratic specification is used.

Years since Migration (YSM).—For individuals with fewer than 35 years of residence in Australia, information on years since migration is available in single years. The open-ended category of 35 years or longer is assigned a value of 40 years.

Birthplace.—The following birthplace regions were formed from the 99 valid country codes available in the original data: Britain, Ireland, northern Europe, southern Europe, eastern Europe, Arab countries, Philippines, Vietnam, South Asia (which primarily comprises the regions of British influence), other Asian countries, Canada, the United States, British West Indies, South and Central America, Africa, New Zealand, Other. For the study of language proficiency, immigrants from Britain, Ireland, Canada, the United States, British West Indies, and New Zealand are used as the control group, whereas for the study of earnings, the omitted category is restricted to immigrants from Britain. The remainder group includes other countries and country not specified.

English Language Proficiency (LANG).—LANG is set to one for individuals who speak only English at home, or if a language other than English is spoken in the home, speak English "very well." The LANG variable is set to zero where a language other than English is spoken in the home and the respondent speaks English either "well," "not well," or "not at all." No information was collected in the 1981 Australian Census of Population and Housing on second languages spoken. Census pretests indicate that the Census language information will provide only a broad indication of the level of proficiency in English. In particular, the Australian Bureau of Statistics notes: "Testing prior to the census compared responses to the question with assessments of language proficiency for a sample of respondents interviewed for the purpose. These results showed that those who responded 'not well' or 'not at all' were correctly identified as prospective 'clients' for English language tuition. However, some who responded 'well' or 'very well' did not rate

highly in the interview assessment. Responses to the 1981 Census question therefore may underestimate the number of people who were not proficient in English" (Australian Bureau of Statistics 1982, p. 1).

Citizenship (CITIZEN).—This is a dichotomous variable, set to one for immigrants whose country of citizenship is recorded as Australia and set to zero for individuals stating a non-Australia country of citizenship and also for those who are classified as stateless.

Marital Status (MARRIED).—This is a binary variable, defined to equal one for individuals who are married (spouse present) and defined to equal zero for all other marital states. Information on whether the individual is married is derived from the census question on marital status. Information on whether the spouse is present is derived from the family structure of the public use sample file.

Married Overseas (FORMAR).—This binary variable is constructed using information on the duration of the current marriage and duration of residence in Australia. Information on duration of marriage is derived from the family structure of the file. Individuals having a duration of current marriage in excess of their duration of residence in Australia are assumed to have married in the country of origin and are coded with FORMAR equal to one. The variable is set to zero for all other individuals.

Children.—Three variables were constructed from the family structure of the public use sample file to parallel the variables included in the analysis of the U.S. census presented in Chiswick and Miller (1992). The first of these records whether one or more children aged less than 6 years were living in the family and there were no older children. The second records whether one or more children aged between 6 and 17 years inclusive were living in the family and there were no younger children. The third variable records the presence of children aged less than 6 years and between 6 and 17 years.

Location.—The only geographic information contained in the Households public use sample distinguishes individuals living in major urban areas, other urban areas, and rural areas. On this basis, two dichotomous variables were formed, the first for individuals living in "other urban areas" and the second for individuals living in "rural" areas.

Minority Language Concentration (CONC).—This variable is defined as the percentage of the respondent's linguistic-country group (see following definition) that resides in the same sized locality (metropolitan area, other urban area, and rural area) as the respondent. It is set equal to zero for the English-speaking majority group.

B. Linguistic-Country Groups

The following groupings are used in the study: English (the United Kingdom and Ireland, Canada, British West Indies, the United States, Bangladesh, Burma, India, Pakistan, Sri Lanka, Kenya, Malawi, Tanzania, Uganda, Zambia, Mauritius, Republic of South Africa, Christmas Islands, Cocos Islands, Oceania other than New Caledonia), Spanish (Spain, Argentina, Boliva, Colombia, Ecuador, Mexico, Paraguay, Peru, Uruguay,

Venezuela), Portugese (Portugal, Brazil, Timor), Arabic (Bahrain, Jordan, Saudi Arabia, Kuwait, Muscat and Oman, Qatar, Yemen, Iraq, Lebanon, Syria, Egypt), Scandinavian (Norway, Sweden, Denmark, Finland), Chinese (China, Hong Kong, Taiwan, Singapore), German (Austria, Germany, Switzerland), French (Belgium, France, Switzerland, New Caledonia), Russian (Estonia, Latvia, Lithuania, Ukraine, the Soviet Union), Italian (Italy), Greek (Greece, Cyprus), Slavic (Albania, Bulgaria, Czechoslovakia), Hungarian (Hungary), Dutch (the Netherlands), Polish (Poland), Romanian (Romania), Indonesian (Indonesia), Persian (Iran), Hebrew (Israel), Japanese (Japan), Cambodian (Kampuchea), Korean (Korea), Laotian (Laos), Malaysian (Malaysia), Tagalog (Philippines), Thai (Thailand), Turkish (Turkey), Vietnamese (Vietnam), Maltese (Malta), Serbian (Yugoslavia).

Note.—All variables for Australia are dichotomous except earnings, education, total experience, duration in the destination, and the minority concentration measure.

C. Language Question: 1981 Australian Census of Population and Housing

Q. 15 Does the person speak a language other than English at home?

No, only speaks English	☐
Yes	☐

How well does this person speak English?

Very Well	☐
Well	☐
Not Well	☐
Not at All	☐

II. Analysis of 1986 Australian Census of Population and Housing

The analyses of the 1986 Australian Census presented in this article are based on the 1986 Household Sample File (Section of State). Two data files were released from the 1986 census. The Section of State Household Sample File identifies "major urban areas," "other urban areas," and "rural areas." The State/Territory file identifies seven states/territories and "major urban" and "balance of State/Territory." The Section of State file is comparable to the 1981 Census Public Use Sample and hence has been used in these analyses. Because of differences in the way in which primary information has been coded in this and the 1981 census, it is necessary to change the definitions of some variables. The new definitions are listed below.

A. Definitions

Years since Migration (YSM).—The 1986 census data are released in categorical form, and only five broad categories are distinguished: 0–4

years, 5–9 years, 10–14 years, 15–19 years, and 20 years and over. A "continuous" duration of residence variable is created using the midpoints of the closed intervals, and a value of 30 years for the open-ended upper interval.

Birthplace.—Eleven broad birthplace regions are recognized in the study. They are the United Kingdom and Ireland, southern Europe, northern Europe (the Netherlands, Germany), other Europe (all other countries of Europe, including the Soviet Union), Arab countries, Vietnam, other Asian countries, South America, Other American countries (Canada, the Caribbean, El Salvador, Mexico, the United States, other American countries), New Zealand, other Oceania. The remainder group includes other countries and country not specified.

Foreign Marriage (FORMAR).—In the study of the 1981 census, the foreign marriage variable was constructed using information on duration of marriage and duration of residence in Australia. For most respondents this information was available in single years. However, as noted above, the duration of residence data in the 1986 census are in very broad intervals. The duration of marriage information is coded into 5-year intervals in the 1986 sample file. This method of presenting the primary data prevents the construction of a usable foreign marriage variable.

Age.—The age data are presented in 5-year intervals: 25–29, 30–34, 35–39, 40–44, 45–49, 50–54, 55–59, 60–64. The midpoints of these intervals have been used to create a "continuous" age measure.

Minority Language Concentration (CONC).—Twelve minority languages are coded in the Household Sample File. These are Arabic/Lebanese, Chinese, Dutch, French, German, Greek, Italian, Maltese, Polish, Serbian and Croatian, Spanish, and Vietnamese. The minority-language concentration variable (CONC) is constructed from these data as follows: each respondent is assigned a value equal to the percentage of the population aged 15–64 in the region (defined broadly using information on location) in which he lives that reports the same second (minority) language as the respondent. The percentage representation in each language group is displayed in Appendix table A2.

B. Language Question: 1986 Australian Census of Population and Housing

Q. 17 Does the person speak a language other than English at home?

☐ No, speaks only English
☐ Yes (Specify language) _____

[Answer question 18 for each person who speaks a language other than English at home]

Q. 18. How well does this person speak English?

Very Well	☐
Well	☐
Not Well	☐
Not at All	☐

III. Descriptive Statistics

Table A1
Means and Standard Deviations of Variables by Regions of Origin for Adult Foreign-born Men, Australia, 1981

Variable	Total Sample (N = 7,288)		English-Speaking (N = 3,122)		Non-English-Speaking (N = 4,166)	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Education	10.961	2.80	11.266	2.455	10.732	3.012
Age	41.808	10.221	41.405	10.429	42.110	10.054
Experience	25.847	11.146	25.138	11.166	26.378	11.103
YSM	18.578	9.790	17.008	9.948	19.755	9.503
Married	.839	.368	.819	.385	.854	.353
Married overseas	.306	.461	.358	.480	.266	.442
Child < 6	.124	.330	.118	.323	.129	.336
Child 6–17	.343	.475	.328	.470	.355	.479
Children < 6 and 6–17	.131	.338	.115	.319	.144	.351
Small urban	.113	.317	.149	.356	.087	.282
Rural	.076	.265	.098	.297	.060	.237
Minority language concentration	.650	1.891	.000	.000	1.138	2.388
Citizenship	.558	.497	.322	.467	.734	.442
Birthplace:						
Britain	.346	.476	.808	.394	...	
Ireland	.025	.156	.058	.234	...	
Canada	.005	.067	.011	.102	...	
United States	.011	.102	.025	.155	...	
West Indies	.001	.029	.029	.044	...	
New Zealand	.041	.199	.002	.296	...	
Southern Europe	.264	.441	...		.461	.499
Northern Europe	.117	.321	...		.204	.403
Eastern Europe	.054	.226	...		.095	.293
Arab countries	.024	.154	...		.043	.202
South Asia	.040	.195	...		.069	.254
Philippines	.003	.055	...		.005	.072
Vietnam	.007	.081	...		.012	.107
Other Asian countries	.020	.140	...		.035	.183
South America	.009	.095	...		.016	.126
Africa	.029	.168	...		.051	.220
Remainder	.006	.075	...		.010	.099
Earnings	14,855	7,808	16,347	8,426	13,737	7,108
Log (earnings)	9.475	.574	9.577	.538	9.399	.588
LANG	.760	.427	.998	.045	.582	.493

SOURCE.—1981 Australian Census of Population and Housing, 1/100 Sample of the Foreign Born.
NOTE.—The English-speaking regions include Britain, Ireland, Canada, United States, New Zealand, and the British West Indies. See above for variable definitions.

Table A2
Percentage Representation of Major Minority-Language Groups by Locality, Australia, 1986

Language	Metropolitan	Small Urban	Rural
Arabic/Lebanese	1.06	.04	.02
Chinese	1.55	.33	.09
Dutch	.61	.43	.55
French	.43	.25	.30
German	1.03	.63	.61
Greek	2.80	.37	.28
Italian	4.24	1.27	1.70
Maltese	.67	.09	.22
Polish	.74	.18	.19
Serbian, Croatian	.67	.06	.23
Spanish	.72	.07	.16
Vietnamese	.70	.07	.02
Other	4.94	1.62	2.50

SOURCE.—1986 Australian Census of Population and Housing.

References

Australian Bureau of Statistics. *Census 81-Language.* Catalogue No. 2152.0. Canberra: Australian Bureau of Statistics, 1982.

Beenstock, Michael. "Learning Hebrew and Finding a Job: Econometric Analysis of Immigrant Absorption in Israel." Paper presented at the Conference on the "Economics of International Migration: Econometric Evidence," Konstanz, Germany, February 1993.

Case, Anne C., and Katz, Lawrence F. "The Company You Keep: The Effects of Family and Neighborhood on Disadvantaged Youths." Photocopied. Discussion Paper no. 1555. Cambridge, MA.: Harvard Institute of Economic Research, 1991.

Chiswick, Barry R. "The Effect of Americanization on the Earnings of Foreign-born Men." *Journal of Political Economy* 86 (October 1978): 897–921.

———. "The Earnings of White and Coloured Male Immigrants in Britain." *Economica* 47 (February 1980): 81–87.

———. "Speaking, Reading and Earnings among Low Skilled Immigrants." *Journal of Labor Economics* 9, no. 2 (April 1991): 149–70.

———. "Hebrew Language Usage: Determinants and Effects on Earnings among Immigrants in Israel." Paper presented at the Conference on "Immigrant Absorption: The Interface Between Research and Policy Making," Technion, Haifa, Israel, June 1993.

Chiswick, Barry R., and Miller, Paul W. "Immigrant Generation and Income in Australia." *Economic Record* 61 (1985): 540–53.

———. "Earnings in Canada: The Roles of Immigrant Generation, French Ethnicity and Language." In *Research in Population Economics,* ed. T. Paul Schultz, pp. 183–228. Greenwich, CT: JAI, 1988.

———. "Language in the Labor Market: The Immigrant Experience in Canada and the United States." In *Immigration, Language and Ethnic Issues: Canada and the United States,* ed. Barry R. Chiswick, pp. 229–96. Washington DC: American Enterprise Institute, 1992.

———. "Language Choice among Immigrants in a Multilingual Destination." *Journal of Population Economics* 7, no. 2 (1994): 119–31.

———. "Language and Labor Supply: The Role of Gender among Immigrants in Australia." *Research in Income Inequality* 5 (1995): 153–89.

———. "Ethnic Networks and Language Proficiency among Immigrants." *Journal of Population Economics* (in press).

Committee to Advise on Australia's Immigration Policies. *Understanding Immigration.* Canberra: Australian Government Printing Service, 1987.

———. *Immigration: A Commitment to Australia.* Canberra: Australian Government Printing Service, 1988.

Cox, David. "The Role of Ethnic Groups in Migrant Welfare." In *Commission of Inquiry into Poverty: Welfare of Migrants,* ed. Ronald F. Henderson, pp. 1–150. Canberra: Australian Government Publishing Service, 1975.

Dustmann, Christian. "Speaking Fluency, Writing Fluency and Earnings of Migrants." *Journal of Population Economics* 7, no. 2 (1994): 133–56.

Evans, M. D. R. "Sources of Immigrants' Language Proficiency: Australian Results with Comparisons to the Federal Republic of Germany and the United States of America." *European Sociological Review* 2, no. 3 (1986): 226–36.

Harley, Brigit. *Age in Second Language Acquisition.* Avon: Multilingual Matters, 1986.

Kossoudji, Sherrie A. "English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men." *Journal of Labor Economics* 6, no. 2 (April 1988): 205–28.

Koutsoyiannis, A. *Theory of Econometrics.* 2d ed. Hong Kong: Macmillan, 1977.

Long, Michael H. "Maturational Constraints on Language Development." *Studies in Second Language Acquisition* 12, no. 3 (September 1990): 251–85.

Manski, Charles F. "Identification Problems in the Social Sciences." *Sociological Methods* 23 (1993): 1–56.

McManus, Walter; Gould, William; and Welch, Finis. "Earnings of Hispanic Men: The Role of English Language Proficiency." *Journal of Labor Economics* 1, no. 2 (April 1983): 101–30.

Miller, Paul W. "An Economic Analysis of Citizenship in Australia." Photocopied. Nedlands: University of Western Australia, Department of Economics, 1989.

Robinson, Chris. "Language Choice: The Distribution of Language Skills and Earnings in a Dual Language Economy." *Research in Labor Economics* 9 (1988): 53–90.

———. "The Joint Determination of Union Status and Union Wage Effects: Some Tests of Alternative Models." *Journal of Political Economy* 97 (1989): 639–67.

Stewart, Mark B. "On Least Squares Estimation When the Dependent Variable Is Grouped." *Review of Economic Studies* 50 (1983): 737–53.

Stromback, T. "The Earnings of Migrants in Australia." Conference Paper no. 46. Canberra: Bureau of Labour Market Research, 1986.

Tainer, Evelina. "English Language Proficiency and Earnings among Foreign-born Men." *Journal of Human Resources* 23, no. 1 (Winter 1988): 108–22.

White, H. "A Heteroskedasticity-consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity." *Econometrica* 48 (May 1980): 817–35.