A Spatial Economic Perspective on Language Acquisition:
Segregation, Networking and Assimilation of Immigrants[*]

Raymond J.G.M. Florax [a,b] – Thomas de Graaff [a] – Brigitte S. Waldorf [c]

[a] Department of Spatial Economics, Free University
De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands
Phone +31 20 4446090, Fax +31 20 4446004
E-mail: rflorax@feweb.vu.nl, tgraaff@feweb.vu.nl

[b] Regional Economics Applications Laboratory (REAL)
University of Illinois at Urbana-Champaign
Urbana, IL, USA

[c] Department of Geography and Regional Development, University of Arizona
Harvill Building, Box #2, Tucson, Arizona 85721, USA
Phone: +1 520 621 7486, Fax: +1 520 621 2889
E-mail: bwaldorf@u.arizona.edu

ABSTRACT: Immigration and multiculturalism are at the heart of modern western societies. The issue of language acquisition of immigrants is intrinsically linked to immigration. We formally link language acquisition of immigrants to the relative size of the immigrant stock, employing a microeconomic trading framework. Our model allows for spatial interaction going beyond the immigrant's area of residence, and explicitly incorporates spatial segregation. In addition, behavioral differences of immigrants with respect to their level of assimilation into the host country as well as differences in networking within their own ethnic community are accounted for. We test our model for four non-western immigrant groups in the Netherlands using two different spatial scale levels. The empirical results reveal that there is only ambiguous support for the inverse relationship between size of the immigrant community and language acquisition or language proficiency in The Netherlands. We find instead, that there is strong support for language acquisition and understanding being positively influenced by assimilation to the host country's culture.

KEYWORDS: Immigration, segregation, networks, assimilation, language
JEL CODES: J15, J61, R12, R23

---

1. INTRODUCTION

Discussions surrounding immigration and multiculturalism – key characteristics of many western societies – inevitably involve the role of language acquisition. Knowledge of the official language of the host country is often seen as being associated with a positive pay-off (see Berman et al., 2003; Chiswick, 1998; Chiswick and Miller, 1999, 2002), and as a necessary condition for immigrants' economic success and social integration. Typically, the burden of language acquisition is placed on immigrants and is viewed as an important step towards assimilation into the host society.

Immigrants' propensity to learn the host society's language is not constant over time though. As immigration flows mature and the immigrant stock increases, migration costs decrease (Carrington et al., 1996). The result is a higher proportion of risk-averse immigrants relying heavily on network effects (Massey et al., 1993; Waldorf, 1996). Eventually the network effects contribute to persistent immigrant clustering, spatial segregation, occupational segregation, and the formation of ethnic neighborhoods. The literature has paid ample attention to the impact of spatial segregation on immigrants' economic performance, with many suggesting that an ethnic enclave is a positive externality that contributes to an immigrant's economic success (for example, Borjas, 1995; Edin et al., 2003). With respect to acquiring the host society's language, however, ethnic enclaves are expected to be less beneficial as the expected increase in revenues of language acquisition are smaller for immigrants living in an ethnic neighborhood as compared to those living elsewhere. Thus, under conditions of spatial segregation, clustering and ethnic neighborhoods, immigrants will only invest in language acquisition if the associated transaction cost is small.

The concept of linking language acquisition to the relative size of the immigrant stock has been formalized and empirically tested by Lazear (1999). We conceptually extend Lazear's line of reasoning in this paper by addressing three issues. First, we recognize that

immigrants' activity spaces are not confined to their own area of residence. Instead, revenue-inducing activities may involve interaction across a much wider spatial range. Thus, we explicitly account for the spatial arrangement of places with their different immigrant concentrations, thereby effectively differentiating between different levels of spatial segregation. As such, our approach is no longer topologically invariant and is capable of assessing the effects of spatial segregation on language acquisition. The key hypothesis states that highly segregated spatial patterns have a negative impact on language proficiency. From a methodological point of view, our approach is also innovative in that we account for the influence of spatial configurations by drawing on the rich spatial autocorrelation framework.

Second, we recognize that the nexus between language proficiency, the relative size of the immigrant stock as the key attribute describing the place of residence, and the spatial arrangement of these place attributes, is highly scale dependent. At large spatial scales, such as cities or counties, both the place attributes and their spatial arrangement are less informative about impediments to immigrant trading activities than at smaller spatial scales. As a corollary, we expect that the influence of spatial variables on language proficiency is more pronounced at smaller spatial scales such as neighborhoods than at larger spatial scales.

Third, we recognize that behavioral differences among immigrants strongly contribute to variations in language proficiency. We distinguish between two types of behaviors: network ties that are focused on interaction within the own ethnic community, and assimilation behaviors that are focused on interaction with members of the host society. Ties within the own ethnic community are hypothesized to decrease the expected revenue increase due to language acquisition, whereas the opposite is hypothesized for assimilation into the host society.

Empirically, we test the model for four non-western immigrant groups living in the Netherlands (Turks, Moroccans, Surinamese, and Antilleans), using data from the 1994

survey *Sociale Positie en Voorzieningengebruik van Allochtonen* (SPVA). Like many western countries, the Netherlands is effectively an immigration country. At present, foreign-born and second-generation immigrants make up 18 percent of the population and half of them originated in non-western countries. In contrast to traditional immigration countries like the Australia, Canada, and the United States, the Netherlands ceased permitting labor immigration in 1983. However, family reunification, family formation and asylum seekers have nevertheless led to an unprecedented increase in the immigrant stock even after 1983. The persistence of immigrant segregation and concentration that is so characteristic for many immigration countries (Bartel, 1989; LaLonde and Topel, 1991) can also be observed for the Netherlands (de Graaff et al., 2001; Musterd et al., 1998; Tesser et al., 1995). For example in 1998, 42 percent of the non-western ethnic population lived in the four largest Dutch cities (Tesser et al., 1999), with the lack of spatial integration often coinciding with a lack of economic and social adaptation (see Dagevos, 2001).

The paper is divided into six sections. Following this introduction, the second section formally presents Lazear's original theory of language acquisition. In the third section, we explicate the proposed extensions dealing with spatial segregation, spatial scale, networking and assimilation, and derive testable propositions for each of these extensions. In Section 4, we turn to the empirics, and discuss the data, and the measurement of segregation, networking and assimilation. Section 5 focuses on the model estimations, and includes a replication of the original Lazear model using the Dutch data, as well as models that account for the proposed extensions. All models are estimated for two different spatial scales, the city scale and the neighborhood scale. Finally, the sixth section summarizes the main theoretical and empirical conclusions and points to important model extensions and future research.

2. LAZEAR'S THEORY AND MODEL OF LANGUAGE ACQUISITION

Many nations in the western hemisphere are being rapidly transformed into multicultural societies. The notion of "culture" includes sharing similar values, mores, jargon, and rituals. Sharing the same culture increases the probability of exchange or trade, because no intermediaries have to be used, and trust among actors is enhanced due to common expectations and beliefs. In particular, a common language is of paramount importance, because it facilitates trading without interference of a mediator or translator. Lazear (1999) proposes a theory in the spirit of Gary Becker to explain the acquisition of the majority culture by immigrants. His model focuses on the probability of immigrants acquiring the language of the majority culture. It is based on a random encounter model, and assumes that trade among individuals occurs when people share a common language. Language acquisition, which is also an indicator of assimilation, is inversely related to the size of the minority group relative to the total population size in the host country.[1]

In formal terms, Lazear's language acquisition theory can be presented as follows.

Definition *A multicultural society comprises two cultures X: a majority culture A and a minority culture B.*[2]

Assumption 1 *Encounters among people are random, so the probability of encountering an individual from a specific culture X equals the proportion of the population belonging to that culture, $p_X = pop_X / pop$, where pop refers to the total population.*

---

[1] Assimilation is not explicitly incorporated in Lazear's original theory, although there is the ad hoc connotation that can be given to a minority's share of the population: the larger the relative immigrant stock, the lower the probability that immigrants decide to acquire the majority language, and hence the lower the level of assimilation. As detailed in Section 3, we explicitly model the effects of assimilation as an individual behavioral characteristic.

[2] The extension of the theory to a set of *n* disjoint minority cultures $B \in \{B_1, B_2, \ldots, B_n\}$ is straightforward. It is not included here as it complicates notation. Similarly, bi-cultural identities (i.e., immigrants who are fluent in their own language as well as in the majority culture language) can be easily accounted for (see Lazear, 1999).

Assumption 2 *Trade occurs when two individuals from the same culture meet. The pay-off value of trade is normalized to one.*

The definition of different cultures and the assumptions regarding encounter probabilities and trade lead to the following corollary.

Corollary *The expected gains of trade that accrue to members belonging to culture X are p_X, and the expected revenues of acquiring a new language for an individual from culture X are $1 - p_X$.*

Subsequently, the cost of acquiring a new language is introduced by the following assumption:

Assumption 3 *The variation in the population's talent to learn a different language is inversely related to a random cost variable T. The distribution of the cost of acquiring a new language is captured by the probability density function g(t), and the associated distribution function G(t).*

This leads to the central Proposition 1, stating that the probability of acquiring the language of the majority culture depends on the transaction cost being lower than a certain threshold, which is determined by the relative size of the majority culture.

Proposition 1 *The probability of type X individuals choosing to learn the language of a different culture, represented by a binary variable Y, is determined by the transaction cost T being smaller than the expected pay-off:* $\Pr(Y = 1) = \Pr(T < 1 - p_X) = G(1 - p_X)$.

Since $\partial G(1 - p_X) / \partial p_X < 0$, the proportion of a minority assimilating to the majority culture is decreasing in the population proportion that represents the relative size of the minority

group. Analogously, the proportion of the majority culture that learns the minority language is declining in the population proportion that speaks the majority language. Finally, assuming that $p_A > p_B$ guarantees that the proportion of minority culture members learning the majority language is greater than the proportion of majority members who learn the minority language (Lazear, 1999, pp. S.98–99).

Proposition 1 is easily tested by means of a discrete choice model where the dependent variable is a binary variable indicating whether an individual acquires the majority language, and the set of explanatory variables contains the population proportions of each culture, as well as several control variables (such as education, age, and duration of stay). However, the exact specification of the cumulative distribution function is unknown, except for the above property. A function that is linearly decreasing in p_X is a possibility, but not very likely because the associated probability density function $g(1 - p_X)$ is then uniformly equal to one. Instead, a nonlinear decline in p_X is much more likely, and occurs if the transaction costs are normally distributed.

In order to estimate $\Pr(T < 1 - p_X)$ Lazear (1999) uses an empirical model that controls for variation in personal characteristics of immigrants and contains a second-degree polynomial in the cost component p_X.[3] Variation in costs is obtained by distinguishing between different minority groups and by taking into account the spatial concentration of immigrants of a specific culture. The theory is then no longer stated in terms of random encounters per se, but the encounter probabilities depend on location. We therefore refer to this assumption as the regional random encounter model, using the term 'region' as a generic reference to a spatial unit. Effectively, this model changes Assumption 1 into:

---

[3] Note that the cumulative distribution function is then not necessarily strictly decreasing in p_X. It is hence only acceptable if the extreme value of the second-degree polynomial occurs outside the permissible range $0 < p_X < 0.5$. In a society with n disjoint minorities ($n > 1$) the bounds apply to the sum of the minority population proportions.

Assumption 1' *In a given spatial system* $\Re$*, consisting of spatial units r (= 1, 2, …, R), the probability of an individual of a specific culture X, living in region r, encountering an individual from culture X equals the proportion of the regional population belonging to that culture,* $p_X^r = pop_X^r / pop^r$*, where* pop^r *is the regional population.*

Subsequently, the location dependency carries over to the other elements of the theory. The condition for choosing to learn the language of the other culture in Proposition 1 turns into

$$\Pr(Y = 1 | X, r) = \Pr(T < 1 - p_X^r) = G(1 - p_X^r).$$

Empirically, Lazear models the probability of language acquisition using a logit model of the following form:

$$\frac{\Pr(l_X^{ir})}{1 - \Pr(l_X^{ir})} = \exp\left(\alpha_0 + \alpha_1 p_X^r + \alpha_2 p_X^{r\,2} + \sum_j \alpha_j (c_X^i)_j \right), \tag{1}$$

where the probability of fluency in the host society's language, $\Pr(l_X^{ir})$, for an individual i belonging to culture X and living in spatial unit r, is expressed as a function of the location specific minority percentage, p_X^r, and a set of control variables, $(c_X^i)_j$, indexed by j.

3. SPATIAL SEGREGATION, NETWORKING AND ASSIMILATION

In this section, we propose a model that builds on Lazear's framework by accounting explicitly for spatial segregation and behavioral differences. Moreover, instead of explaining language acquisition, represented by a binary variable, we rephrase the theory by using language understanding (or proficiency) as the variable to be explained.[4] Language understanding or proficiency allows for different degrees of familiarity with the majority

---

[4] Lazear (1999, p. S102) recognizes the deficiency of a binary variable, and indicates that the results of his analysis are consistent with an ordered logit model specification where the dependent variable measures various levels of proficiency in English.

language and should preferably be modeled as a continuous variable. From a modeling perspective, we thus focus on variations in the expected level of language proficiency.

3.1 Segregation and spatial scale

Spatial segregation of immigrants constitutes an important dimension of language and culture adoption (Bartel, 1989; de Graaff et al., 2001; Durlauf, 1994; Lalonde and Topel, 1991; Muster et al., 1998; Tesser et al., 1995). However, despite its importance, it has not yet been addressed in formal models of language acquisition. Following Lazear, the decision to learn the majority language is influenced by the expected revenues resulting from trade with members of the majority culture. The expected revenues are small if the probability of encountering a member of the majority culture is small. In the presence of spatial interaction across boundaries, encounter probabilities are not simply determined by the proportion of minorities in the own region, but by the distribution of minorities across the entire system of spatial units making up a person's activity space. Clearly, incorporating spatial interaction across boundaries gains added significance at smaller spatial scales where cross-boundary interactions account for a smaller share of the total interactions.

The debate on measuring segregation dates back to the seminal work by Duncan and Duncan (1955). They proposed to measure segregation via the dissimilarity index, expressing the proportion of the minority that is required to move in order to achieve a uniform distribution of minority and majority populations. Subsequently, a variety of segregation measures have been proposed (see Massey and Denton, 1989; and Waldorf, 1993, for an overview). Invariably, these measures are global in character in that they describe the level of segregation for the entire system of spatial units. In the context of language acquisition, however, members of the minority culture base their decision to learn the majority language on the level of segregation as experienced at their own place of residence. Thus, a local

measure of segregation is needed that responds to variations in encounter probabilities based on the proportion of minorities in the own region, the proportion of minorities in neighboring regions, and assumptions about the interaction patterns across boundaries.

To derive such a measure, we distinguish between three broad types of spatial settings. First, a person may live in a region with a high proportion of members from the own immigrant group and is surrounded by regions with similarly high proportions of immigrants from the same culture. This is the setting in which segregation from the other culture is most severe. Second, a person lives in a region with a low proportion of members from the own group and is surrounded by regions with a similarly low proportion. Under these conditions, the person experiences a low level of segregation from the majority culture and consequently will have ample opportunity to trade with their members as well as have a strong incentive to learn the majority language. A third type of spatial setting refers to a medium level of spatial segregation, applying to persons living in areas with high (low) proportions of their own culture and being surrounded by areas with low (high) proportions.

Such a typology is compatible with the spatial autocorrelation framework (see Anselin, 1988; Cliff and Ord, 1981; Cressie, 1991), in particular the concept of local spatial autocorrelation. We will thus use a local spatial autocorrelation measure to define location-specific spatial segregation and represent the three types identified above.

Anselin (1995, 1996) suggests measuring local spatial autocorrelation via local Moran's Is:[5]

$$I_X^r = \frac{\widetilde{p}_X^r \sum_{\substack{s \\ s \neq r}} w_{rs} \widetilde{p}_X^s}{\sum_s \left(\widetilde{p}_X^s\right)^2}. \tag{2}$$

---

[5] Anselin (1995) shows that, up to a scaling factor, the global Moran's I pertaining to the entire spatial system is equal to the sum of the local Moran's Is. In our case, the global variant can be seen as a system-wide indicator of spatial segregation.

where w_{rs} is a set of weights indicating whether area r is spatially related to area s, and the tilde indicates that the proportions are measured in deviations from the mean. For each region r, the numerator refers to the product of the proportion in the own region and the spatially lagged proportions. An important *a priori* decision is concerned with the specification of the weights. In the context of encounter probabilities of meeting trading partners throughout the spatial system, we define the weights as inverse distances, thus assuming linear distance decay in spatial interaction behavior.

Borrowing further from the spatial autocorrelation framework, one can utilize the Moran scatterplot (Anselin, 1996) to visualize the typology of spatial settings that can be encountered by the minority population. In a Moran scatterplot, the regional minority population proportions ($\tilde{p}_X^r$) are given on the horizontal axis, and the distance weighted minority population proportion in all other regions ($\sum_s w_{rs} \tilde{p}_X^s$) is measured on the vertical axis. Because both variables are measured in deviations from the mean, the first quadrant represents HIGH-HIGH combinations and corresponds to the highest level of segregation from the other culture. The third quadrant represents LOW-LOW combinations and matches up with the lowest level of spatial segregation from the other culture. Quadrants 2 and 4 include the LOW-HIGH and HIGH-LOW combinations, respectively, and represent medium levels of spatial segregation.

Incorporating the location-specific segregation and keeping the switch to a continuous language proficiency variable in mind, Lazear's Proposition 1 may be rephrased as follows:

Proposition 2 *The probability of type X individuals choosing to learn the language of a different culture is determined by the transaction cost being smaller than the expected pay-off. The expected level of proficiency in the language, $\mathrm{E}(L\,|\,p_X^r, s_X^r)$, depends on the*

condition $T < c$, where c is a cost function defined in terms of $(1 - p_X^r)$, and on the level of

spatial segregation experienced by a minority X living in region r, (s_X^r).

Utilizing the concept of local Moran's *I*, we measure local segregation experienced by a

minority member living in region *r* as the product of minority proportion in the own region

and the spatially weighted minority proportions in the neighboring regions, $s_X^r = p_X^r p_X^{wr}$,

where $p_X^{wr} := \sum_{s, r \neq s} w_{rs} p_X^s$. Operationally, variations in language proficiency can then be

captured in a regression model of the following form:

$$L_X^{ir} = \alpha_0 + \alpha_1 p_X^r + \alpha_2 (p_X^r p_X^{wr}) + \sum_j \alpha_j (c_X^i)_j + \varepsilon_X^i, \tag{3}$$

where language proficiency, L_X^{ir}, for an individual *i* belonging to culture *X* and living in

region *r*, is modeled as a function of the location specific minority percentage p_X^r, the

location specific level of spatial segregation $p_X^r p_X^{wr}$, and a set of control variables.

Since $\partial L / \partial p_X^r = \alpha_1 + \alpha_2 p_X^{wr}$, the impact of the relative size of the minority population

on language proficiency now also depends on the proportion of members of that same

minority in neighboring regions. The impact, which we hypothesize to be negative, is thus

also influenced by the spatial context of the region of residence, and the model is therefore no

longer topologically invariant.

3.2 Networking and assimilation

The extension presented above proxies the processes and behaviors influencing the costs of

language acquisition via the proportion of an immigrant group in the area of residence and all

other areas of the spatial system. While this is an elegant and parsimonious formulation of

variations in immigrants' language acquisition probabilities taking into account spatial

segregation, there is a need to make the effects of underlying behaviors more explicit. In this section, we focus on the behavioral mechanisms at work.

Lazear's theory is based on the crucial assertion that encounters among people are random (Assumption 1). While this assumption may be reasonable in the aggregate, at an individual level it suggests that a person's interaction behavior is void of preferences. Relaxing this assumption and allowing encounters to be governed by preferences, we distinguish between two types of behaviors. The first type includes behaviors geared towards maintaining contacts to the own immigrant community, such as belonging to an ethnic organization. We refer to these behaviors as *networking*. Networking behaviors affect language acquisition in two different ways. First, networking implies that trade occurs disproportionately with members of the own culture. Thus, the probability of encountering a member of the own culture is greater than p_X and, consequently, the expected revenues of acquiring the majority language are smaller than $1 - p_X$. This decrease in expected revenues will, in turn, lower the probability of choosing to learn the language. Second, networking reduces exposure to the majority language and is thus an impediment to learning the language.

The second type refers to behaviors that establish contacts with members of the host society, such as spending leisure time with members of the majority culture. We refer to these behaviors as *assimilation* behaviors. Seeking contact to members of the majority culture effectively increases the expected revenues of acquiring a new language, thus increasing the probability of choosing to learn the majority language. Moreover, unlike networking, these assimilative behaviors increase exposure to the majority language and thus foster language acquisition and language proficiency.

The incorporation of individual behavioral differences in networking within the ethnic community and assimilation into the host society, leads to the following testable proposition.

Proposition 3 *The expected level of proficiency in the language of the majority of type X individuals* $\mathrm{E}(L)$ *is determined by the transaction cost being smaller than the expected pay-off, which is influenced by the proportion of minorities in the own region* (p_X^r), *location-specific segregation* $(p_X^r p_X^{wr})$, *the level of networking within the ethnic community* (n_X^i) *and the assimilation into the majority culture* (a_X^i).

The operational regression model then takes on the form:

$$L_X^{ir} = \alpha_0 + \alpha_1 p_X^r + \alpha_2 p_X^r p_X^{wr} + \alpha_3 n_X^i + \alpha_4 a_X^i + \sum_j \alpha_j (c_X^i)_j + \varepsilon_X^i, \tag{4}$$

where, in addition to the expectation of $\alpha_1 + \alpha_2 p_X^{wr}$ being negative, we expect α_3 to be negative, and α_4 to be positive.

4. DATA, MEASUREMENT, AND OPERATIONALIZATION

In this section, we will use the Dutch SPVA survey data mentioned in the introduction to estimate model (4) for two spatial scales, the level of cities and the level of neighborhoods.

4.1 Data

The data are taken from the 1994 Dutch survey *Sociale Positie en Voorzieningengebruik van Allochtonen* (SPVA). The survey, commissioned by the *Institute for Sociological-Economic Research* (ISEO) and conducted by the *Amsterdam Bureau for Statistics and Research* (O+S), aimed at investigating to which extent minority groups in the Netherlands participate in education and the labor market. Interviewers administered a questionnaire of 255 items and obtained the response of 4,096 Dutch and immigrant household heads (60 percent response

rate) residing in 13 cities in the Netherlands.[6] Fifty-four percent of the immigrant population resides in these 13 cities, underscoring the notion that immigration in the Netherlands is primarily an urban phenomenon. In fact, as shown in Figure 1, the spatial distribution of the main immigrant groups (specifically, Turks, Moroccans, Surinamese, and Antilleans) is centered in the main urban areas of the so-called *Randstad*, whereas more peripheral regions to the north and the south have low immigrant proportions. These four groups represent two different types of immigrants. Turks and Moroccans originally migrated to the Netherlands as part of the guest-worker program sponsored by the Dutch government and private companies in response to labor shortages in the 1960s. Following the restrictions on labor migration in the early 1980s, Turkish and Moroccan migration has been mostly due to family unification. Immigration of Antilleans and Surinamese, in contrast, is primarily linked to colonial ties, with large flows of migration being triggered by the independence declarations of (parts of) the Dutch Antilles and Surinam. In addition, immigration was also induced by severe economic recessions in the Dutch Antilles and Surinam.

— FIGURE 1 about here —

From the pool of respondents, we select Turkish, Moroccan, Surinamese and Antillean first-generation immigrants, yielding a sample of 1,727 observations with complete demo-economic, behavioral and social information as well as information about the respondent's city of residence. The availability of data on minority proportions at the neighborhood scale is somewhat more restricted, and necessitates us to limit the sample to 1,387 immigrants.[7]

---

[6] Respondents are sampled from the following cities: Amsterdam, Rotterdam, The Hague, Utrecht, Eindhoven, Enschede, Almere, Alphen aan de Rijn, Bergen op Zoom, Hoogezand-Sappemeer, Delft, Dordrecht and Tiel. These cities include the six largest cities in the Netherlands.

[7] Effectively, the sample is then limited to the cities of Amsterdam, Rotterdam, The Hague, Utrecht, Eindhoven, and Enschede.

Table 1 summarizes the descriptive statistics for the sampled immigrants and for the spatial characteristics at the city and the neighborhood scale. The table shows that the distribution over the different immigrant groups is rather even, with a slight under-representation of Antilleans. The information clearly reveals sharp differences between Turks and Moroccans on the one hand, and Surinamese and Antilleans on the other hand. Compared to Turks and Moroccans, Antilleans and Surinamese household heads are six times more likely to be female, and more likely to be single. They are also more educated than Turks and Moroccans. The average length of stay in the Netherlands is equal among Turks and Moroccans, clearly higher for Surinamese, and lower for Antilleans. The four groups do not, however, differ substantially with respect to their average age, and neither do they differ substantially in terms of their networking and assimilation behaviors.

Interestingly, we observe a sharp divide between the Moroccans and Turks on the one hand, and the Surinamese and Antilleans on the other in terms of language acquisition and language proficiency. This is obviously due to the colonial history of the Netherlands with Suriname and the Antilles.

A trough is also apparent when it comes to the relative size of the different immigrant groups and their spatial concentration. The Surinamese are the largest immigrant community, and they are highly spatially clustered/segregated, both at the city and the neighborhood scale. A similar pattern, although at a lower level, is apparent for Turks and Moroccans, who are very similar in terms of relative size and segregation patterns. The Antilleans are the smallest immigrant group, and they appear to be the least affected by spatial segregation.

4.2 Measurement and operationalization

Before we turn to the estimation of the models, we provide some more detail on the

measurement and operationalization of several key variables. The model suggested in (4) requires operational definitions of the level of segregation $(p_X^r\, p_X^{wr})$, the level of networking within the ethnic community (n_X^i), the degree of assimilation into the majority culture (a_X^i) and the dependent variable, language proficiency, L_X^{ir}.

The level of spatial segregation requires a specification of the spatial weights giving rise to $p_X^{wr} = \sum_{s,r \neq s} w_{rs} p_X^s$. For the city level analysis, we use inverse Euclidian distances between the midpoints of all municipalities in the Netherlands, with a cut-off point of 50 km, and standardize the resulting weights to one. Thus, for each region (city) r, the sum of weights representing distances to all neighbors within a 50 km range equals one. Moreover, all spatial units, not just those represented in the sample, are included in calculating the average spatially weighted minority proportion in the neighboring regions. For the neighborhood analysis, we use inverse Euclidian distances between all neighborhoods of the same city, and similarly standardize the resulting weights.

With the sets of spatial units described above, we utilize Moran scatterplots to classify the spatial units as belonging to a setting of high, medium, or low segregation (see Figure 2).[8] At the city level, the vast majority of immigrant locations is characterized by high segregation (first quadrant), whereas low segregation settings are not represented at all. At the neighborhood level, we find all three types of segregation levels represented. Interestingly, the low-segregation setting (third quadrant) is quite dominant, for Antilleans it even accounts for 55 percent of all neighborhoods represented in the sample.

— FIGURE 2 about here —

---

[8] One can compute the global Moran's *I* as an overall indicator of spatial segregation. Generally, Moran's *I* is equal to the slope of a trend line through the observations in the Moran scatterplot (see Anselin, 1996, for details). Note that this is not true with respect to the scatterplots in Figure 2, because they do not contain observations for all cities and neighborhoods of the entire spatial system.

To measure assimilation and network effects, a battery of survey items is entered into a factor analysis. Two factors nicely distinguish between the two types of behaviors (see Table 2). The first factor is labeled 'assimilation' and loads highly on three items describing the relationship with the Dutch population, namely 'meets Dutch people during leisure time', 'receives visits from Dutch friends', and 'values having Dutch friends'. The second factor is labeled 'networking' and loads highly on 'frequents ethnic places', and 'is member of an ethnic organization'. We use the scores associated with the assimilation and network factors as additional exogenous variables in the language proficiency model.

— TABLES 2 and 3 about here —

Finally, the language proficiency measure is derived from a factor analysis that uses eight language performance and language behavior measures (see Table 3). The primary factor explains 47 percent of the total variation and loads strongly negative on the variables that measure difficulties of understanding Dutch in various settings: reading newspapers, listening to radio/TV, and during the interview. Thus, factor scores of the primary factor are used to measure language proficiency.

5. MODELING RESULTS

We begin our analysis with a replication of the Lazear model, using the Dutch data for the municipality level and the neighborhood level respectively. The dependent variable of the logit model is the binary variable 'understanding Dutch'. As shown in Table 4, at both spatial scales the results are consistent with those of Lazear (1999) who employed 1990 US county data. That is, as the minority proportion (expressed in percentage and representing the encounter probability) of the immigrant's residential area increases, the probability of knowing Dutch decreases at a decreasing rate to a minimum, and increases at an increasing

rate thereafter. Extending this model by including age, gender, marital status, and education as additional control variables[9] does not influence the shape of the quadratic function. As shown in Figure 3, the minimum occurs at an encounter probability of $p_X^r = 4.5\%$ in the city model and at 10% in the neighborhood model.

— FIGURE 3 about here —

The second-degree polynomial specification is, however, not without problems. In a strict sense, Lazear's theory is corroborated only if the observed minority proportions are in the decreasing part of the function. The estimated minimum of the parabolic function is, however, within the range of the sample minority proportions, although the vast majority of cases are in the downward sloping part of the curve.[10] Immigrant concentrations to the right of the minimum are primarily found in the major cities; specifically, Turks in Rotterdam and The Hague, Moroccans in Amsterdam and Utrecht, and Surinamese in Amsterdam, Rotterdam and Utrecht. In light of this inconsistency, we estimated several other specifications, including a linear variant in p_X^r, and a logarithmic specification. These variants invariably led to insignificant estimates and/or estimates with the wrong sign.[11]

A comparison of the city and neighborhood models alludes to yet another issue that needs further exploration. The coefficient of the interaction term of the encounter probability and duration is significantly smaller than zero in the city specification, whereas it is not significant in the neighborhood specification. This points to an important difference in interpretation. The theory outlined so far is consistent with both a learning interpretation and

---

[9] Except for the gender coefficient, the estimated effects of the control variables are consistent across all models presented in this paper. Ceteris paribus, understanding Dutch increases with increasing length of stay (at a decreasing rate) and improved education, but decreases with age. Moreover, single immigrants are more likely to understand Dutch than non-single immigrants.

[10] In the city model, the percentage for minorities ranges from 0.4% to 9.2%, and six out of 52 minority proportions (12%) are in the upward sloping part of the city model. At the neighborhood scale, the minority proportions range from 0% to 29.1%, and 29 of 352 minority proportions (8%) are in the upward sloping part.

[11] The results of the linear and logarithmic variants are available from the authors upon request.

a sorting interpretation. The learning interpretation implies that the probability of acquiring the majority language is inversely related to the relative size of the minority culture. Most importantly, this interpretation suggests that the spatial location is exogenous. The sorting interpretation, on the other hand, implies that minorities who do not speak the majority language, choose to live in ethnic neighborhoods. Thus, spatial location becomes endogenous. Moreover, one can expect the effect of duration to be less pronounced and the interaction of minority proportion and duration not to be significantly different from zero (see Lazear 1999, pp. S103-105). This is exactly what we observe in the neighborhood specification, but not at the city scale.

— TABLE 4 about here —

The results of Table 4 thus show that at different spatial scales, the Dutch case provides only ambiguous support for Lazear's theory. Especially for cities or neighborhoods with the highest concentrations of ethnic minorities, the postulated negative effect of immigrant clustering on language acquisition is less than what can be expected if the relationship between size of the immigrant community and the probability of acquiring the majority language were strictly decreasing. Potentially this may be due to the higher probability of meeting members of the ethnic minority who already speak the majority language. This is not necessarily detrimental, but the theory should be further amended to account for positive externalities of clustering that facilitate exposure to the majority culture among ethnic minorities and/or reduce the cost of language acquisition.

In Table 5 we present the estimation results for models of language proficiency that account for a refined treatment of spatial context effects. Towards that end, we include the segregation variable $(p_X^r p_X^{wr})$, and allow its effect to vary across the three type of segregative settings, i.e., high segregation, medium segregation, and low segregation. Moreover, in these models we loosen the assumption of random encounters by allowing for behavioral

preferences for interaction within the own ethnic community (networking) or interaction with the Dutch population (assimilation). For both spatial scales, we present the robust OLS estimates as well as maximum likelihood estimates for a spatial autoregressive model with groupwise heteroskedasticity.

At the larger spatial scale of municipalities, the marginal effect of the encounter probability is negative,[12] thus confirming immigrants' decreasing language proficiency with increasing minority proportions. Moreover, segregation also has a negative effect and the effect is even stronger in settings classified as 'medium segregation' than in settings classified as 'high segregation'. This non-linearity in the segregation variable suggests a threshold beyond which a further increase in the size of the ethnic cluster no longer strengthens the impediments to learning Dutch. At the smaller neighborhood scale, the results are less supportive of our theory. The OLS results suggest that the marginal effect of encounter probabilities is negative if the minority proportions in neighboring areas are sufficiently small, as is the case in low-segregation settings. The segregation variables itself, however, have a positive effect on language proficiency thus hinting at clustering of minorities in space as a positive externality for learning Dutch. Interestingly, accounting for spatial autocorrelation at the neighborhood scale, renders all spatial variables insignificant.

Whereas the empirical support for the hypothesized influence of location is ambiguous, the results for the influence of assimilation and networking are consistent across spatial scales. Assimilative behaviors geared towards interaction with the Dutch population have a strongly positive impact on language proficiency. Networking is estimated to have the expected negative influence on language proficiency, but the effect is barely or not

---

[12] At the city scale, the partial derivative in a highly segregated setting is
$$\partial L / \partial p_X^r = 0.83 - 1.21 p_X^{wr} - 0.00008 Dur \leq -0.38008$$
in the robust OLS specification, and
$$\partial L / \partial p_X^r = 0.65 - 0.96 p_X^{wr} - 0.00011 Dur \leq -0.31011$$
in the ML-error model.

significant. Taken together, these results suggest that interaction with Dutch people is advantageous for language proficiency, whereas interaction with members of the own immigrant group is not necessarily disadvantageous. From a theoretical perspective, the results suggest that the assumption of random encounters that are void of preferences, is insufficient to account for variations in immigrants' language proficiency.

— TABLE 5 about here —

In view of the heterogeneity among different immigrant groups, it is appropriate to allow parameters to varying across the four different immigrant groups. It should be noted, however, that the variables measured at the regional scale vary over cities (or neighborhoods) and minority groups only. Consequently, allowing for variation over immigrant groups with respect to the constant term and the variables measured at the regional scale would distort the interpretation of these variables. We therefore allow for parameter variation over individual characteristics only. The results are presented in Table 6.

Using maximum likelihood estimation for a spatial autoregressive error model with groupwise heteroskedasticity, the model accounts for 53 percent of the variation in language proficiency at both spatial scales. One of the key results is a diminished importance of the spatial variables. Where immigrants live and thus their encounter opportunities with members of their own group is less relevant once we allow the effects of individual characteristics to vary across immigrant groups. Of particular interest are also the group-specific differences in the behavioral effects. The assimilation effect is significantly positive for all four immigrant groups. But it is especially strong for Turkish immigrants. In contrast, networking within the own ethnic community hinders language proficiency but, interestingly, it is insignificant for Moroccans. The effects of all other control variables by and large are consistent with the results from the models presented in Tables 4 and 5, with language proficiency being high for young ages, long stays abroad, extended education, and single immigrants.

6. CONCLUSIONS

This paper extends our understanding of language acquisition among immigrants. Inspired by and building upon Lazear's (1999) theory and empirical models, we maintain Lazear's basic argument that language acquisition is a function of transaction costs relative to the expected pay-off, and that transaction costs are a function of the relative size of the immigrant stock. Lazear's theoretical model – as well as his analytical results and subsequent interpretation – posits increasing transaction costs with increasing immigrant proportions, and thus decreasing levels of language acquisition in the immigrant population.

In this paper, we extend the theoretical arguments by emphasizing three issues. First, we explicitly account for immigrants' broader spatial setting by introducing a segregation variable. Effectively, this extension implies that we allow for trading interactions throughout the entire spatial system. Second, we draw attention to the importance of spatial scale. In particular, we argue that topological invariance and spatial constraints on trading interactions become questionable restrictions at small spatial scales. Third, we loosen the assumption of random encounters by accounting for individual preferences and behaviors that influence the selection of trading partners. Thus, we explicitly measure and model assimilation behaviors involving contacts with the host society and networking behaviors within the own immigrant community.

The model results – estimated for Turks, Moroccans, Antilleans and Surinamese living in the Netherlands, using two spatial scales – show that the influence of space on language proficiency is not as strong as suggested by our theory. For the city scale, albeit not for the neighborhood scale, we show that the relevance of location is not simply confined to where immigrants live but also involves the broader spatial setting. We thus maintain that,

conceptually, lifting the spatial constraint on trading activities is an important extension of Lazear's theory.

The empirical results do, however, provide convincing support of our assertion that behavioral differences play a key role in understanding immigrants' language proficiency. Assimilation behaviors involving contacts with the host society and, to a lesser extent, networking behaviors within the own immigrant community are important for language proficiency and overshadow the spatial effects. Assimilation behaviors take on a pivotal role and their power to foster language proficiency outweighs the networking induced hindrance of language acquisition. This result nicely fits arguments in favor of multiculturalism versus denial and intolerance towards immigrants' cultural roots.

Extensions of this research should further investigate the role of spatial effects. The Dutch data only allowed us to use rather broad spatial scales. However, measuring spatial interaction and spatial segregation at the level of municipalities and even neighborhoods may average out important variations at a smaller spatial scale, such as blocks or even individual apartment buildings. Exploring immigrants' language acquisition at such a fine spatial scale may shed further light on the role of segregation, networking and assimilation.

REFERENCES

Anselin, L., SpaceStat: A Program for the Analysis of Spatial Data, National Center for Geographic Information and Analysis, University of California, Santa Barbara, 1992 (see also: Anselin, L., SpaceStat Version 1.50: Revision notes, Research Paper No. 9428, Regional Research Institute, West Virginia University at Morgantown, 1994).

Anselin, L., Local Indicators of Spatial Association — LISA, *Geographical Analysis*, 27, pp. 93-115, 1995.

Anselin, L., The Moran Scatterplot as an ESDA Tool to Assess Local Instability in Spatial Association, in: M. Fischer, H. Scholten and D. Unwin (Eds.), *Spatial Analytical Perspectives on GIS*, Taylor and Francis, London, 1996.

Anselin, L., Spatial Econometrics, in: B. Baltagi (Ed.), *Companion to Econometrics*, Basil Blackwell, Oxford, 2000.

Anselin, L. and A.K. Bera, Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics, in: A. Ullah and D. Giles (Eds.), *Handbook of Applied Economic Statistics*, Marcel Dekker, New York, 1998.

Anselin, L., A.K. Bera, R.J.G.M. Florax and M.J. Yoon, Simple Diagnostic Tests for Spatial Dependence, *Regional Science and Urban Economics*, 26, pp. 77-104, 1996.

Bartel, A.P., Where do the New U.S. Immigrants Live?, *Journal of Labor Economics*, 7, pp. 371-391, 1989.

Borjas, G.J., Ethnicity, Neighborhoods, and Human Capital Externalities, *American Economic Review*, 85, pp. 365-390, 1995.

Carrington, W.J., E. Detrangiache and T. Vishnawath, Migration with Endogenous Moving Costs, *American Economic Review*, 86, pp. 909-930, 1996.

Chiswick, B.R., Hebrew Language Usage: Determinants and Effects on Earnings among Immigrants in Israel, *Journal of Population Economics*, 11, pp. 253-271, 1998.

Chiswick, B.R. and P.W. Miller, Language Skills and Earnings among Legalized Aliens, *Journal of Population Economics*, 12, pp. 63-89, 1999.

Chiswick, B.R. and P.W. Miller, Immigrants Earnings: Language Skill, Linguistic Concentrations and the Business Cycle, *Journal of Population Economics*, 15, pp.31-57, 2002.

Cliff, A.D. and J.K. Ord, *Spatial Processes: Models & Applications*, Pion, London, 1981.

Cressie, N., *Statistics for Spatial Data*, Wiley, New York, 1991.

de Graaff, T., C. Gorter and P. Nijkamp, Effects of Ethnic Geographical Clustering on Educational Attainment in the Netherlands, Tinbergen Institute Discussion Paper, TI 2001-028/3, Amsterdam, 2001.

Dagevos, J., Perspectief op Integratie, Werkdocumenten WRR, W 121, Den Haag, 2001.

Duncan O. and B. Duncan, A Methodological Analysis of Segregation Indexes, *American Sociological Review*, 20, pp. 210-217, 1955.

Durlauf, S.N., Spillovers, Stratification, and Inequality, *European Economic Review*, 38, pp. 836-845, 1994.

Edin P., P. Frederiksson and O. Aslund, Ethnic Enclaves and the Economic Success of Immigrants: Evidence from a Natural Experiment. *Quarterly Journal of Economics*, 118, pp. 329-357, 2003.

Fleming, M., Techniques for Estimating Spatially Dependent Discrete Choice Models, in: L. Anselin, R.J.G.M. Florax and S.J. Rey (Eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Berlin: Springer Verlag, 2004 (in press).

LaLonde, R.J. and R.H. Topel, Immigrants in the American Labor Market: Quality, Assimilation, and Distributional Effects, *AEA Papers and Proceedings*, 81, pp. 97-302, 1991.

Lazear, E.P., Culture and Language, *Journal of Political Economy*, 107, pp. S95-126, 1999.

Massey, D. and N. Denton, Hypersegregation in U.S. Metropolitan Areas: Black and Hispanic Segregation along Five Dimensions, *Demography*, 26, pp. 373-391, 1989.

Massey, D., J. Arango, G. Hugo, G. Kouaouci, A. Pellegrino and J.E. Taylor, Theories of International Migration: A Review and Appraisal, *Population and Development Review*, 19, pp. 431-466, 1993.

Musterd, S., W. Ostendorf and M. Breebaart, Segregation, Polarization and Urban Policies in Amsterdam: An International Perspective, in: C. Gorter, P. Nijkamp and J. Poot

(Eds.), *Crossing Borders: Regional and Urban Perspectives on International Migration*, Ashgate, Aldershot, 1998.

Tesser, P.T.M., C.S. van Praag, F.A. van Dugteren, L.J. Herweijer and H.C. van der Wouden, *Rapportage Minderheden: Concentratie en Segregatie*, Sociaal en Cultureel Planbureau, Rijswijk, 1995.

Tesser, P.T.M., J.G.F. Merens and C.S. van Praag, *Rapportage Minderheden 1999*, Sociaal en Cultureel Planbureau, Rijswijk, 1999.

Waldorf, B., Segregation in Urban Space: A New Measurement Approach, *Urban Studies*, 30, pp. 1151-1164, 1993.

Waldorf, B., The Internal Dynamic of International Migration Systems, *Environment and Planning A*, 28, pp. 631-650, 1996.

TABLE 1: Summary statistics for different minority groups, and spatial characteristics at the city and the neighborhood scale (means with standard deviations in parentheses), and number of observations in the samples for the city and the neighborhood analysis.[a]

Characteristics	Turks	Moroccans	Surinamese	Antilleans	Total
Individual characteristics					
Sex (female = 1)	0.08	0.08	0.49	0.51	0.28
	(0.26)	(0.27)	(0.50)	(0.50)	(0.45)
Age (in years)	36.11	38.86	40.70	36.14	38.06
	(10.27)	(11.63)	(12.71)	(11.53)	(11.74)
Marital status (single = 1)	0.06	0.13	0.23	0.34	0.18
	(0.24)	(0.33)	(0.42)	(0.47)	(0.39)
Duration (in months)	179.09	188.66	210.73	148.09	183.72
	(88.33)	(105.97)	(88.59)	(108.26)	(99.67)
Education (four-point scale)	0.16	0.13	0.31	0.40	0.24
	(0.36)	(0.33)	(0.46)	(0.49)	(0.43)
Speaks Dutch (yes = 1)	0.28	0.32	0.85	0.69	0.53
	(0.45)	(0.47)	(0.36)	(0.46)	(0.50)
Networking[b]	4.26	4.11	3.83	3.78	4.00
	(0.96)	(0.94)	(1.04)	(1.02)	(1.01)
Assimilation[b]	4.01	3.79	4.03	4.24	4.01
	(0.91)	(0.90)	(0.94)	(1.00)	(0.95)
Proficiency in Dutch[b]	3.20	3.40	4.55	4.30	3.86
	(0.90)	(0.88)	(0.63)	(0.79)	(0.99)
Spatial characteristics city scale					
Size minority (%)	4.27	4.59	6.81	1.15	4.40
	(1.33)	(2.12)	(3.02)	(0.51)	(2.83)
High segregation[c]	1.24	2.70	3.96	0.28	2.15
	(1.75)	(1.48)	(2.15)	(0.17)	(2.14)
Medium segregation[c]	1.63	0.08	0.16	0.03	0.51
	(1.35)	(0.24)	(0.48)	(0.06)	(1.02)
Low segregation[c]	—	—	—	—	—
Spatial characteristics neighborhood scale					
Size minority (%)	6.73	6.11	10.45	1.92	6.65
	(6.92)	(4.98)	(9.86)	(1.69)	(7.47)
High segregation[c]	28.63	30.20	68.24	1.37	34.71
	(43.20)	(33.11)	(79.72)	(2.46)	(56.03)
Medium segregation[c]	7.03	5.26	5.22	0.94	4.82
	(12.05)	(12.77)	(11.37)	(1.87)	(10.95)
Low segregation[c]	1.17	0.58	3.61	0.34	1.55
	(3.81)	(1.67)	(10.44)	(0.42)	(6.12)
# Obs., city analysis	464	417	482	364	1727
# Obs., neighborhood analysis	363	340	397	287	1387

[a] The individual characteristics are reported for the sample used in the city analysis.
[b] The networking, assimilation and proficiency variables are continuous variables derived by means of factor analysis (see the main text for details). The variable ranges are [2.04,7.23] for assimilation, [1.61,6.35] for networking, and [1.71,5.35] for language proficiency.
[c] See the main text for the definitions of the segregation scores. The category 'Low segregation' does not occur in the cities included in our sample. Note that the descriptive statistics pertain to the entire (sub-)samples and therefore include zeros as well.

TABLE 2: Factor loadings for the assimilation and network variables (with factor rotation).[a]

Variables	Factors	
	Assimilation	Networking
Receives visits from Dutch friends	0.84	−0.02
Favors living in a clustered ethnic neighborhood	−0.31	0.20
Is member of an ethnic organization	0.05	0.73
Meets Dutch individuals in leisure time	0.87	−0.02
Values having Dutch friends	0.54	0.10
Partner from another ethnic group	−0.07	0.48
Is member of an organization	0.31	0.45
Frequents ethnic places (bars, coffee shops, etc.)	−0.12	0.78

[a] The cumulative percentage of total variance explained is 45 percent. The procedure used is principal components with VARIMAX rotation. The significance of Bartlett's test of sphericity is 0.00.

TABLE 3: Factor loadings for the language variable proficiency in Dutch (with factor rotation).[a]

Variables	Factor Understanding Dutch
Speaks another language	−0.11
Has difficulty writing letters in Dutch	−0.38
Migrant has children	−0.06
Has difficulty with Dutch during the survey	−0.74
Speaks Dutch with children	0.17
Has difficulty with Dutch in newspapers	−0.79
Speaks Dutch with partner	0.07
Has difficulty with Dutch on radio\TV	−0.85

[a] The cumulative percentage of total variance explained is 47 percent. The procedure used is principal components with VARIMAX rotation. The significance of Bartlett's test of sphericity is 0.00.

TABLE 4: Logit estimates of the probability of acquiring Dutch for Turkish, Moroccan, Surinamese and Antillean immigrants, in the Netherlands in 1995, replicating the Lazear specification, and extended specifications with personal characteristics included.[a,b]

	City scale		Neighborhood scale	
	Lazear	*Extended*	*Lazear*	*Extended*
Constant	0.41*	0.65**	–0.60***	–0.55*
	(0.22)	(0.32)	(0.20)	(0.30)
Pr(Encounter) (%)	–1.05***	–0.86***	–0.10***	–0.07**
	(0.08)	(0.09)	(0.03)	(0.03)
$Pr(Encounter)^2$	0.12***	0.10***	0.0052 ***	0.0047***
	(0.01)	(0.01)	(0.001)	(0.001)
Duration	0.01***	0.02***	0.007***	0.01***
	(0.002)	(0.002)	(0.002)	(0.002)
$Duration^2 \times 10^{-3}$	–0.02***	–0.01***	–0.01**	–0.01*
	(0.004)	(0.005)	(0.004)	(0.005)
Pr(Encounter) × *Duration* $\times 10^{-3}$	–0.48**	–0.48**	–0.05	–0.10
	(0.21)	(0.22)	(0.09)	(0.10)
Age	—	–0.05***	—	–0.04***
		(0.0065)		(0.007)
Gender (female = 1)	—	0.63***	—	1.20***
		(0.13)		(0.14)
Schooling (extended = 1)	—	0.96***	—	1.15***
		(0.15)		(0.16)
Marital status (single = 1)	—	0.33**	—	0.41**
		(0.16)		(0.16)
Log-Likelihood	–1039.30	–965.56	–921.28	–813.31
χ^2	308.13***	455.61***	75.73***	291.65***
Pseudo R^2	0.13	0.19	0.04	0.15
Akaike IC	1.21	1.13	1.34	1.18
n	1727	1727	1387	1387

[a] Estimated parameters are recorded with standard errors in parentheses. Significance at the 1, 5, and 10 percent level is indicated by ***, **, and *, respectively.
[b] We did not test and subsequently implement a correction for heteroskedasticity, so the *z*-values are potentially inflated. Another source of misspecification is the potential of dependence among individual observations, following a spatial pattern. The latter is not investigated as testing and estimation of spatial logit models is still not very well developed (see Fleming, 2004).

TABLE 5: Robust OLS estimates, and maximum likelihood estimates for a spatial autoregressive error model with groupwise heteroskedasticity, explaining language proficiency in Dutch for Turkish, Moroccan, Surinamese and Antillean immigrants, in the Netherlands in 1995.[a,b]

	City scale		Neighborhood scale	
	Robust OLS	ML-error	Robust OLS	ML-error
Constant	2.68***	3.18***	2.97***	3.39***
	(0.17)	(0.17)	(0.19)	(0.18)
Pr(Encounter) (%)[c]	0.83***	0.65***	–0.08***	0.0003
	(0.07)	(0.14)	(0.02)	(0.023)
High segregation	–1.21***	–0.96***	0.013***	0.0009
	(0.10)	(0.22)	(0.002)	(0.0032)
Medium segregation	–1.36***	–1.14***	0.014***	0.00007
	(0.09)	(0.21)	(0.003)	(0.0047)
Low segregation	—	—	0.02***	0.001
			(0.004)	(0.005)
Duration	0.01***	0.00416***	0.0034***	0.0031***
	(0.001)	(0.0006)	(0.0008)	(0.0006)
$Duration^2 \times 10^{-3}$	–0.01***	–0.004***	–0.003	–0.003*
	(0.002)	(0.001)	(0.002)	(0.001)
$Pr(Encounter) \times Duration \times 10^{-3}$	–0.08	–0.11*	–0.03	–0.03
	(0.07)	(0.06)	(0.03)	(0.03)
Age	–0.02***	–0.02***	–0.02***	–0.02***
	(0.003)	(0.002)	(0.003)	(0.002)
Gender (female = 1)	0.18***	–0.03	0.45***	–0.01
	(0.05)	(0.04)	(0.05)	(0.05)
Schooling (extended = 1)	0.44***	0.30***	0.52***	0.27***
	(0.04)	(0.04)	(0.05)	(0.05)
Marital status (single = 1)	0.26***	0.18***	0.28***	0.13***
	(0.05)	(0.05)	(0.06)	(0.05)
Networking	–0.04**	–0.03	–0.06**	–0.0025
	(0.02)	(0.02)	(0.02)	(0.020)
Assimilation	0.23***	0.19***	0.22***	0.19***
	(0.02)	(0.02)	(0.02)	(0.02)
Spatial autoregressive parameter	—	0.68***	—	0.83***
		(0.04)		(0.03)
Group Variances	—		—	
- Turks		0.61***		0.61***
		(0.04)		(0.05)
- Moroccans		0.65***		0.62***
		(0.05)		(0.05)
- Surinamese		0.38***		0.33***
		(0.02)		(0.02)
- Antilleans		0.48***		0.51***
		(0.04)		(0.04)
R^2-adjusted or Buse	0.37	0.24	0.29	0.11
Log-likelihood	–2031.03	–1877.49	–1718.75	–1478.46
Akaike IC	4088.06	3780.99	3465.50	2984.92
n	1727	1727	1387	1387

[a] Estimated parameters are recorded with standard errors in parentheses. Significance at the 1, 5, and 10 percent level is indicated by ***, **, and *, respectively.
[b] The results are OLS estimates with White-adjusted variances, and maximum likelihood estimates for a spatial autoregressive error model with groupwise heteroscedasticity. The latter estimator is chosen on the basis of a Breusch-Pagan test for heteroskedasticity, and Lagrange Multiplier tests for spatial dependence (see Anselin et

al., 1996; and Anselin and Bera, 1998, for details). We test for autocorrelation following a spatial pattern based on inverse distances between individuals within 50 kilometers of each other. Because we do not know the exact location of individuals within zip code areas, we assign spatial coordinates to individuals randomly, within the minimum and maximum longitude and latitude of the zip code area. The estimation has been carried out using the SpaceStat software package (Anselin, 1992). Detailed test results are available from the authors upon request.

TABLE 6: Maximum likelihood estimates for a spatial autoregressive error model with groupwise heteroscedasticity, explaining language proficiency in Dutch for Turkish, Moroccan, Surinamese and Antillean immigrants, in the Netherlands in 1995.[a,b]

	City scale				Neighborhood scale			
	Turks	*Moroccans*	*Surinamese*	*Antilleans*	*Turks*	*Moroccans*	*Surinamese*	*Antilleans*
Constant		3.69***				3.63		
		(0.16)				(0.15)		
Pr(Encounter) (%)[c]		−0.28**				−0.013		
		(0.15)				(0.02)		
High segregation		0.47**				0.0017		
		(0.24)				(0.0028)		
Medium segregation		0.41				0.0013		
		(0.26)				(0.0036)		
Low segregation		———				0.0001		
						(0.0041)		
Duration × 10^{-2}	0.41***	0.63***	0.12	0.38***	0.44***	0.43***	0.19**	0.39***
	(0.16)	(0.15)	(0.09)	(0.12)	(0.16)	(0.16)	(0.09)	(0.14)
Duration² × 10^{-3}	0.006	−0.004	−0.001	−0.005**	0.006	−0.003	−0.002	−0.004
	(0.004)	(0.004)	(0.002)	(0.002)	(0.005)	(0.004)	(0.002)	(0.003)
Pr(Encounter) × Duration × 10^{-3}	−0.025	−0.143	0.077	0.730	0.034	0.041	−0.0077	0.207
	(−0.183)	(0.111)	(0.078)	(0.575)	(0.042)	(0.047)	(0.0281)	(0.210)
Age	−0.060***	−0.043***	−0.008***	−0.008**	−0.062***	−0.039***	−0.008***	−0.012**
	(0.005)	(0.005)	(0.002)	(0.004)	(0.005)	(0.006)	(0.002)	(0.004)
Gender (female = 1)	−0.06	0.15	−0.09	0.02	−0.059	0.32**	−0.067	0.011
	(0.13)	(0.15)	(0.05)	(0.07)	(0.15)	(0.16)	(0.058)	(0.085)
Schooling (extended = 1)	0.54***	0.40***	0.24***	0.27***	0.51***	0.40***	0.23***	0.24**
	(0.09)	(0.12)	(0.06)	(0.08)	(0.11)	(0.13)	(0.06)	(0.10)
Marital status (single = 1)	0.47***	0.23**	0.02	0.18**	0.25	0.22*	0.01	0.18*
	(0.14)	(0.12)	(0.07)	(0.08)	(0.16)	(0.12)	(0.07)	(0.10)
Networking	−0.09***	−0.03	0.05**	−0.06*	−0.08***	−0.003	0.064**	−0.05
	(0.03)	(0.04)	(0.03)	(0.04)	(0.04)	(0.040)	(0.026)	(0.04)
Assimilation	0.25***	0.15***	0.17***	0.16***	0.27***	0.13***	0.17***	0.15***
	(0.03)	(0.04)	(0.03)	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)
Spatial autoregressive parameter		0.15**				0.17**		
		(0.07)				(0.08)		
Group Variances	0.50***	0.59***	0.33***	0.45***	0.50***	0.57***	0.30***	0.48***
	(0.03)	(0.04)	(0.02)	(0.03)	(0.04)	(0.04)	(0.02)	(0.04)
R² (Buse)		0.53				0.53		
Log-likelihood		−1766.43				-1400.22		

n	1727	1387

[a] See the footnotes to Table 5.

[b] Centered coefficients are restricted not to vary over immigrant groups.

|	
	0.000 - 0.010
	0.010 - 0.027
	0.027 - 0.056
	0.056 - 0.117
	0.117 - 0.216

Fi
mu
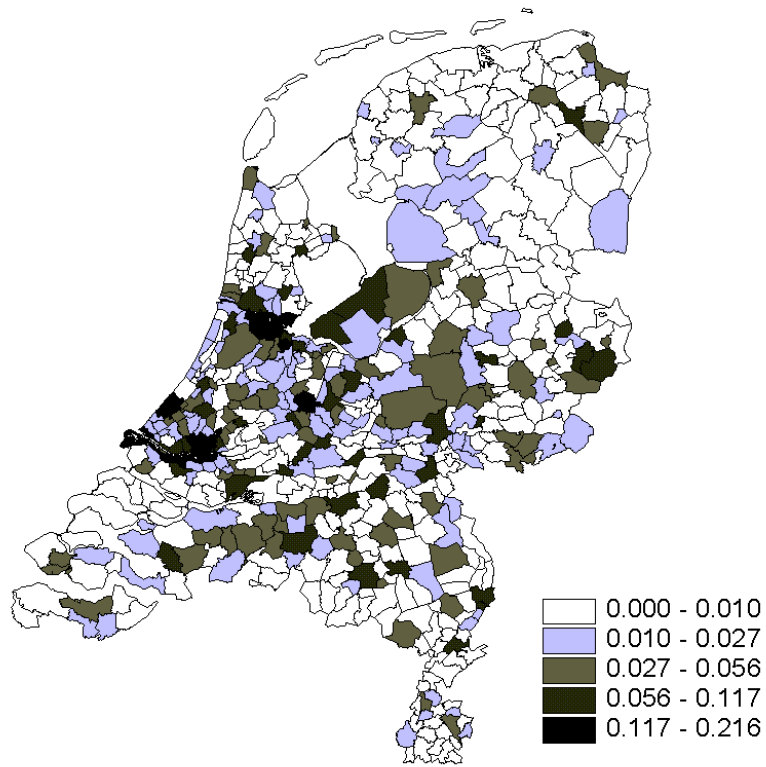
So
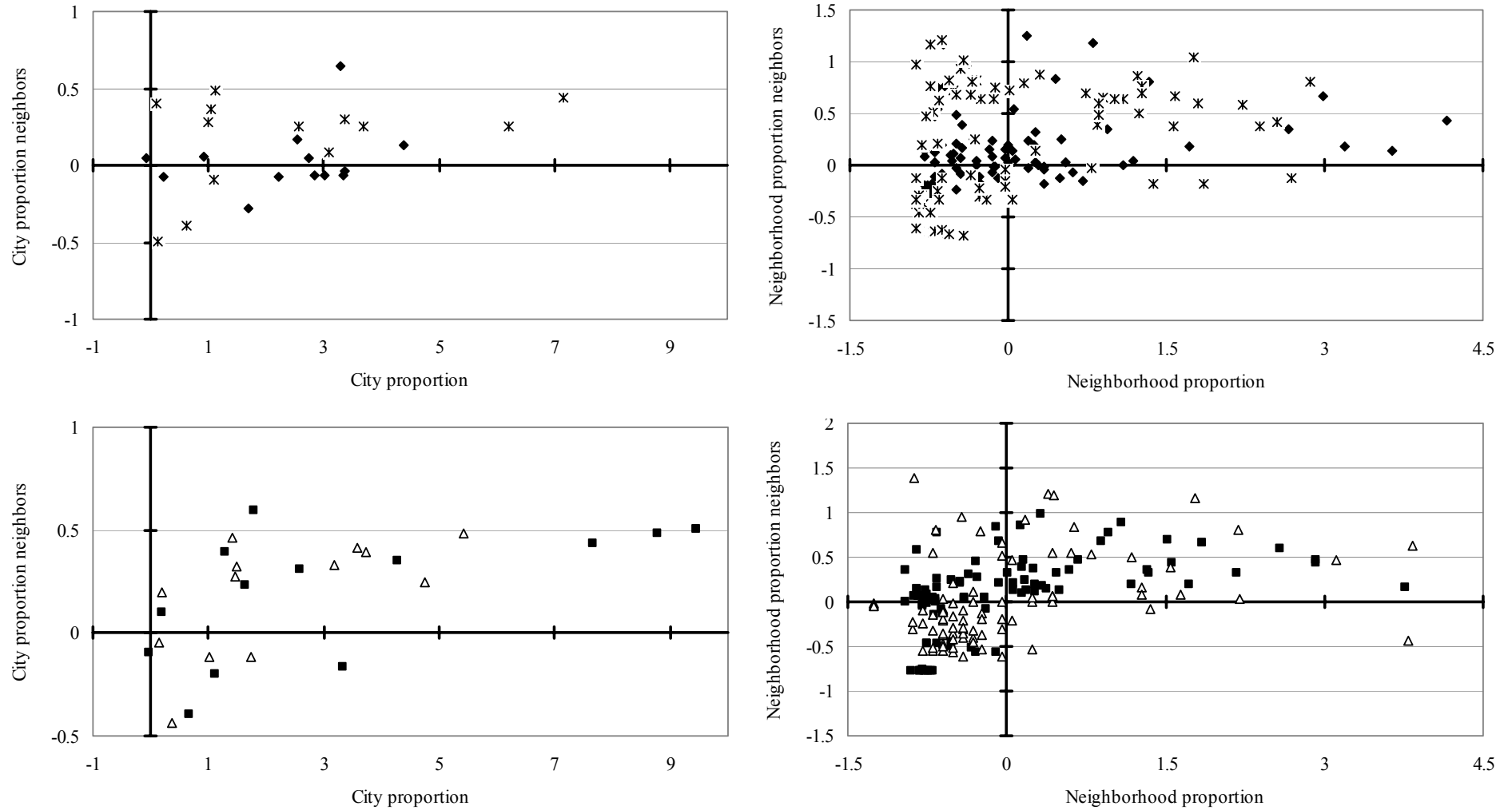
FIGURE 2: Moran scatterplot with minority population proportions on the horizontal axis, and distance weighted minority population proportions in neighboring regions on the vertical axis (both measured in deviations from the mean).

Note: the graphs are for the city level (left) and the neighborhood level (right). Turks (diamond) and Moroccans (asterisk) are presented in the top graphs, and Surinamese (square) and Antilleans (triangle) in the bottom graphs.
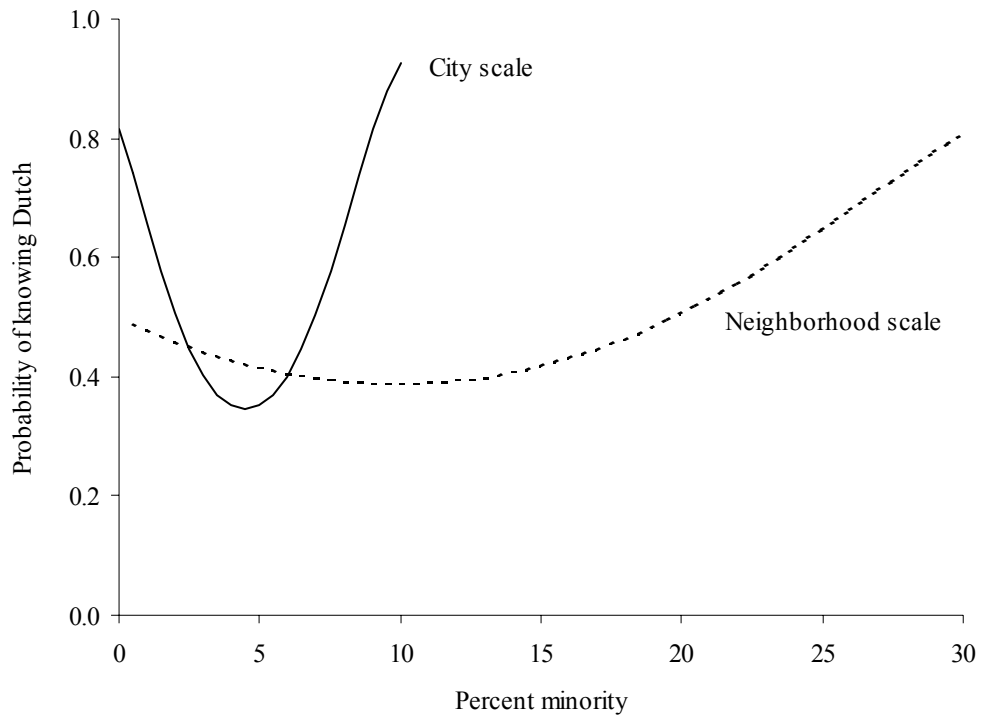
FIGURE 3: Probability of knowing Dutch evaluated for a 38-year old non-single man with average education, who has lived in the Netherlands for 15 years (sample means), at the level of cities and neighborhoods.