
CONTEXTUAL EFFECTS ON EDUCATIONAL ATTAINMENT IN INDIVIDUALIZED NEIGHBORHOODS; DIFFERENCES ACROSS GENDER AND SOCIAL CLASS

EVA ANDERSSON & BO MALMBERG
DEPT. OF HUMAN GEOGRAPHY, STOCKHOLM UNIVERSITY

ABSTRACT

This paper analyzes if a multi-scale representation of geographical context based on statistical aggregates computed for individualized neighborhoods can lead to improved estimates of neighborhood effect. Our study group consists of individuals born in 1980 that have lived in Sweden since 1995 and we analyze the effect of neighborhood context at age 15 on educational outcome at age 30 controlling for parental background. A new software, Equipop, was used to compute the socio-economic composition of neighborhoods centered on individual residential locations and ranging in scale from including the nearest 12 to the nearest 25,600 neighbors. Our results indicate that context measures based on fixed geographical sub-divisions can lead to an underestimation of neighborhood effects. A multi-scalar representation of geographical context also makes it easier to estimate how neighborhood effects vary across different demographic groups. This indicates that scale-sensitive measures of geographical context could help to re-invigorate the neighborhood effects literature.

Key words: contextual effects, neighborhood effects, context, Equipop, adolescents, education, Sweden

CONTEXTUAL EFFECTS ON EDUCATIONAL ATTAINMENT IN INDIVIDUALIZED, SCALABLE NEIGHBORHOODS; DIFFERENCES ACROSS GENDER AND SOCIAL CLASS

INTRODUCTION

In 2012, two books with strikingly different ideas about the role of neighborhood processes in urban development were published. On the one hand, Robert Sampson's *Great American City: Chicago and the Enduring Neighborhood Effect*, argued that neighborhood processes are of fundamental importance for the working of a modern city. On the other hand, *Neighbourhood Effects Research: New Perspectives* edited by Van Ham et al., essentially argued that neighborhood effect studies are at an impasse and that further progress will require both radical rethinking of theories, and changes in research methodology. Together, these two books reflect the current state of neighborhood effect studies. They bear witness to the continuing interest in neighborhood effects, but they also make it clear that this is a field characterized by considerable controversy.

One possible reason for the controversy is that in spite of a strong theoretical argument—backed up by considerable qualitative evidence—there is mixed quantitative evidence for an influence of neighborhood context on life outcomes. This discrepancy is clearly frustrating.

In this paper, we will argue that one reason for the lack of clear-cut results could be problems associated with the measurement of neighborhood context. Up to now the main approach in

this field of study has been to measure context using aggregate values for administratively defined areas. This implies that neighborhood effect studies have, to a considerable degree, ignored the argument put forward by Openshaw (1984) and others that such aggregate measures will be plagued by indeterminacy. Certainly researchers have been aware that their measures have been far from perfect (see e.g. Putnam (2007)), but there seems to have been a widespread belief that values that have been aggregated using fixed areal units can serve as good approximation, given a lack of feasible alternatives. Statistical theory, however, says that measurement errors in explanatory variables will have strong negative effects on one's ability to obtain good estimates of the parameters of a statistical model. Therefore, it is possible that disappointing results in neighborhood effect studies are simply a reflection of weakness in the empirical design (Galster, 2008; Sampson, Morenoff and Gannon-Rowley, 2002).

The solution that we propose in this paper is to measure neighborhood context using aggregates for individualized, egocentric neighborhoods. These neighborhoods will be constructed as buffers around the residential location of the individuals under study in such a way that the buffer for each location will include the same number of nearest neighbors. In this way, the modifiable areal unit problem will be circumvented since the measurement of context will become independent of any statistically given areal subdivision. In addition to addressing the modifiable areal unit problem, such individualized, egocentric neighborhoods also offer a possible way to handle another challenge for contextual measurement: the problem of uncertain geographic context discussed by Kwan (2012). More specifically, individualized neighborhoods based on buffers of with different population counts can be used to obtain contextual measures for different neighborhood scales.

The question we will address in this paper is how educational achievement at adult age is influenced by the neighborhood context in early youth. We will use Swedish register data and

our basic design is similar to that used by Andersson (2004), Andersson and Subramanian (2006), Sundlöf (2008) and Bygren and Szulkin (2010). However, whereas these studies use aggregates for fixed geographical sub-divisions to measure context, we take advantage of the availability of individual level data with geo-coordinates to construct aggregates for individualized neighborhoods with fixed population counts. A similar approach to contextual measurement has previously been used by (Bolster et al., 2007; Chaix et al., 2005; Macallister et al., 2001). In our case we used the Equipop software developed by John Östh to extract contextual information from geo-coded, individual level register data. With Equipop it is easy to obtain aggregate information for neighborhoods that vary in scale, with scale measured by the by the population count. Neighborhoods can be defined to include only a handful of individuals, but it is also possible to compute aggregates for neighborhoods with population counts similar to those of medium-sized cities. This flexibility forces researchers to explicitly consider at which geographical scale different neighborhood effects are likely to operate.

Studies measuring contextual/neighborhood effects (Ainsworth, 2002; Andersson, 2004; Andersson and Subramanian, 2006; Crane, 1991; Evans, Wells and Moch, 2003; Immergluck, 1998; Ludwig, 1999; South, Baumer and Lutz, 2003) share a concern for data quality and an interest in determining if contextual effects are significant for the specified outcomes. However, there is less agreement about the mechanisms behind such contextual effects (Galster, 2012; Galster and Santiago, 2006).

Different mediating processes discussed among researchers include social control, collective socialization, social capital, and institutional characteristics, see Ainsworth (2002) and Galster (2007). However, what is not discussed at length is how such contextual/neighborhood effect mechanisms and processes work at different geographical scales (Andersson and Musterd, 2010; Östh, Malmberg and Andersson, forthcoming). We believe our multi-scalar approach to

contextual measurement can stimulate interest in theories about how scale is important for contextual effects.

Consider, first, neighborhood level *social control*; that is the monitoring and sanctioning of deviant behavior of youths and others. If there are fewer adults around and if they do not spend time with youths, youths may shape their own norms (Ainsworth, 2002). Here it could be argued that small-scale environments can be of special importance. Adult monitoring is stronger for pre-school children who play in local streets and playgrounds, see Jacobs, (1993). Likewise, children in lower primary school will typically gather in the vicinity of local schools with relatively restricted catchment areas. In Sweden, an average primary school is attended by 82 children in grades 1-3 (Swedish National Agency for Education, 2013).

Consider, second, the scale at which *collective socialization* could be thought to influence ideas and traditions of education among children and adolescents. Collective socialization is a process in which youths are exposed to role models among adults, and adapt to those models to varying degrees. The importance of such role models has been questioned (Joseph, Chaskin and Webber, 2007) but in relation to educational aspiration one could imagine a context wherein homework and studying are not seen as the 'coolest' things. Or the opposite: a neighborhood where homework and reading were taken seriously by most parents and children. To the extent that attitudes towards education are formed in early youth, the spatial scale of such influence could be less restricted than the space in which smaller children move. Using Swedish data, Andersson and Subramanian (2006), for example, show effects for administrative areas with on average 970 individuals as influencing years of education.

Third, it can also be argued that *social capital* or social networks that exist in a given community (Putnam, 1993) will be found on a larger scale than social control. Children and adolescents living in advantaged neighborhoods are more likely to be exposed to helpful social

networks. In advantaged neighborhoods, children are also more likely to meet adults who can provide positive recourses in the form of information (or job opportunities, help with advanced homework etc.). If information about educational opportunities, encouragement and support become more important for the educational decisions of individuals when they are in upper secondary school this would allow contextual influences from environments that are more extended than those that have importance for children in primary school.

Fourth, it can be argued that neighborhood processes that are linked to institutional mechanisms operate at scales that, in the Swedish context, can transcend local neighborhoods and can encompass an entire urban district or an entire non-metropolitan municipality. Such institutional mechanisms are discussed by a number of authors (Ainsworth, 2002; Galster and Santiago, 2006; Sampson, Morenoff and Gannon-Rowley, 2002) and relate, for example, to the quality of institutions and the availability of institutions such as health centers, schools, universities, hospitals and job centers (for a discussion of school effects see e.g. (Brännström, 2008; Sellstrom and Bremberg, 2006; Sykes and Musterd, 2011; Östh, Andersson and Malmberg, 2013).

Clearly, the above discussion gives support to Sampson's (2012) argument that there is a need for spatial flexibility when it comes to measuring contextual influences. According to Sampson, what is required is not a "search for the 'best' or 'correct' operational definition of neighborhood" (Sampson, 2012). Instead, in his view "there are multiple scales of ecological influence and possibilities for constructing measures, ranging from micro level street blocks (or street corners) to block groups to neighborhood clusters to community areas of political and organizational importance to spatial 'regimes' and cross-cutting networks that connect far-flung areas of the city" (p. 361-2). Based on this argument, we propose an approach to contextual measurement that explicitly allows for multiple scales of influence. This is achieved

by computing aggregate values for individualized neighborhoods of twelve different sizes ranging from very small to large. Compared to traditional measures this gives a richer and more composite description of the socio-spatial environment of individuals. Below we will demonstrate that estimates of neighborhood effects based on this more composite description will be greater than those obtained with traditional measures based on fixed areal subdivisions.

It can, however, be questioned whether *neighborhood* effects is an appropriate term for describing the influence of environmental factors that are measured for such a broad range of geographical scales. Can the term *neighborhood* be used both for an area that encompasses the twelve nearest neighbors and areas that include the 25,600 nearest neighbors? Given that Equipop computes aggregate statistics for areas defined using population counts, the term *neighbor* effect could be used. However, in this paper we will use, interchangeably, the well-established concepts neighborhood effects and contextual effects.

EMPIRICAL DESIGN, DATA AND METHODS

The purpose of the empirical study presented below is to test if contextual measurements based on individualized neighborhoods can lead to improved estimates of neighborhood effects. The outcome variable will be educational achievement at age 30 for a cohort born in 1980, and we will analyze the effect of neighborhood exposure during the age range of 14-18 years. Moreover, we will use two sets of contextual data: one set based on individualized neighborhoods ranging in scale from areas including the 12 nearest neighbors to areas including the 25,600 nearest neighbors, and, to enable a comparison, one set based on aggregates for fixed geographical subdivisions, the Swedish SAMS areas. For both these data sets we will use the same socio-economic indicators, described below in Table 2. The data

originates from PLACE, a database delivered by Statistics Sweden located at Uppsala University.

In the contextual effect estimations we will not, however, use statistical aggregates obtained for individualized neighborhood and SAMS areas directly. Instead we will use factor scores resulting from a factor analysis carried out separately for the SAMS-based and individualized-neighborhood based data. With this setup, our claim that the individualized neighborhood provides a better basis for contextual analysis can be evaluated by comparing the estimates for SAMS-based and individualized-neighborhood-based context measures.

INDIVIDUAL LEVEL, COHORT AND HOUSEHOLD DATA

To be included in our study individuals had to have stayed in the same geographical location between the ages of 14 and 18 years. This non-mobility criterion, which reduced the sample by 15%, was imposed in order to ensure that they have been affected by the same surroundings during the exposure period. Individuals with missing data on parental background have also been excluded. Of 102,592 individuals born in 1980, 74,648 individuals are included in our final sample.

Excluding movers had some effect on the composition of the sample since movers have a higher share of unemployed, lower educated, foreign-born, low income, and visible minority parents. The dependent variable of education was measured by the existence of a university or university college degree, as shown in Table 1 below.

INSERT TABLE 1. HERE

To account for individual level influence on educational achievement our statistical model will include six indicators of parental background: visible minority parents, foreign-born parents, parents with tertiary education, parents with social allowance, non-employed parents, and

parents' disposable income in the top decile of parental incomes. In addition, we include an indicator for gender and for living in a single mother household. Descriptive statistics for these variables are given in Table 1.

CONTEXTUAL MEASUREMENT

With respect to the use of individual background variables the empirical design of this study is conventional. This is not the case, though, with our approach to context measurement. Here, instead, our study introduces two important novelties: first, and most importantly, we introduce contextual measures that are based on individually defined and scalable neighborhoods. Second, we introduce a factor-analysis based representation of the spatial variation in a socio-demographic context as a means to manage the wealth of information resulting from scalability. Last in this section, and as means of comparing this work with earlier studies, we present the often used Swedish areal division of so called SAMS areas.

INDIVIDUALLY DEFINED AND SCALABLE NEIGHBORHOOD

We measure neighborhood population compositions using individual centered neighborhoods of fixed population size. Thus, we have used register data containing information of individual residential location to compute contextual variables based on the population composition among an individual's nearest 12, 25, 50, 100, 200, 400, 800, 1600, 3200, 6400, 12800, and 25600 neighbors for 1995 (for the population older than 25 years *and* the total population, depending on variables), see Table 2.

The Equipop software was developed by John Östh in order to address the modifiable areal unit problem (MAUP) in segregation measurement (Östh, Malmberg and Andersson, 2011). Traditional measures of segregation such as the isolation index are strongly dependent on the size of the statistical units for which the segregation index has been computed (Malmberg, Andersson and Östh, 2011). Recently, Equipop has also been used to analyze residential

segregation in the Los Angeles Metropolitan Area (Östh, Clark and Malmberg, 2013). In the Equipop software, the individualized neighborhoods are obtained by expanding a circular buffer around each residential location until the population encircled by the buffer corresponds to the population threshold chosen. When this threshold is reached, the program computes aggregate statistics on a selected socio-economic variable for the encircled population.

Equipop requires that the input data is geocoded on a detailed level. We have used data from the PLACE database of Uppsala University. This data contain register-based, individual level data for the population in Sweden from 1990 to 2010 with geocodes of the residential location by 100 meter squares. From this data, seven different socio-demographic indicators have been extracted and used as input for Equipop; see Table 2.

INSERT TABLE 2. HERE

SAMS AREAS AS CONTEXT

The residential differentiation according to the Small Area Market Statistics, SAMS, classification scheme is a national subdivision in homogenous residential areas (more than 9,000). SAMS was developed by Statistics Sweden in collaboration with the municipalities, and is a social division according to building characteristics and tenure form. Originally, for publicity purposes and municipal planning, SAMS was formed to include a certain target group. Some SAMS areas are uninhabited, and accordingly the number of inhabitants and sizes of SAMS vary. Our study makes use of 7,704 SAMS areas which vary in population size between 1 and 207 individuals.

FACTOR-ANALYSIS BASED REPRESENTATION OF CONTEXTUAL VARIATION: INDIVIDUALIZED
NEIGHBORHOODS

With seven different socio-demographic indicators and 12 different levels of neighborhood scale we obtain a total of 84 different contextual variables. Clearly, such a large number of contextual variables cannot be included, without problems, as explanatory variables in a regression of educational achievement. Moreover, many of the indicators are strongly correlated, for example, contextual indicators based on the same socio-economic indicator but computed for different neighborhood sizes. In order to make the analysis manageable we have subjected the contextual indicators to a factor analysis that compresses the 84 original indicators to 10 orthogonal factors that jointly capture more than 79% of the original variation. The factor analysis was based on covariances, and the number of principal components to be rotated was selected based on them having eigenvalues higher than one. The factors were rotated using the varimax method.

Some factors influence a small number of neighbors (k) as contextual variables, and other factors influence a large number of neighbors. This result of the factor analysis is clearly of interest since it provides an opportunity to analyze the scale dependence of contextual effects.

Figure 1 shows diagrams of what the different factors represent. This interpretation is important since we are going to include factor scores as explanatory variables in the logistic regression of educational achievement. Without an interpretation of the different factors it will be difficult to interpret the regression results. Our interpretation of the factors is given below.

Factor 1 **Elite areas**. High values for this factor in a location result in a high proportion of people with tertiary education and high disposable income, and a low proportion of unemployed.

Factor 2 **Low employment in adjacent areas**. High values for this factor result in a high level of non-employment at neighborhood scales, not among the closest neighbors, but beyond 1000 persons. The same areas are also characterized by having few inhabitants with high disposable income.

Factor 3 **Foreign-born**. High values for this factor imply a high proportion of foreign-born residents, and to some extent, social allowances.

INSERT FIGURE 1. HERE

Factor 4 **Marginal nearby**. High values for this factor result in low levels of single family housing, a high proportion of single mother households, households with social allowances, and foreign-born residents at neighborhood scales above 800 persons.

Factor 5 **Marginal intermediate**. Factor 5 is similar to Factor 4 with the difference that Factor 5 has an effect mainly on neighborhood scales below 800 persons.

Factor 6 **Single family housing**. This factor contributes to a high proportion of the population living in single-family houses, high disposable income and a low proportion of non-employed.

Factor 7 **Low employment, small-scale**. Factor 7 is similar to Factor 2 with the difference that Factor 7 has an effect mainly on neighborhood scales below 1000 persons.

Factor 8 **Low employment medium**. Factor 8 is similar to Factor 2 with the difference that Factor 8 has an effect mainly on neighborhood scales of around 1000 persons.

Factor 9 **Marginal medium**. Factor 9 is similar to Factor 4 with the difference that Factor 9 has an effect mainly on neighborhood scales of around 1600 to 6400 persons.

Factor 10 **Non-academic elite**. High values on Factor 10 are associated with high levels of disposable income but not with a high proportion of tertiary education.

FACTOR-ANALYSIS BASED REPRESENTATION OF CONTEXTUAL VARIATION: FIXED GEOGRAPHICAL SUB-DIVISIONS

The seven variables aggregated using SAMS areas are not as strongly correlated as the contextual variables resulting from using individualized neighborhoods, but to allow a clear-cut comparison we apply factor analysis on these variables too. In this case only three factors are needed in order to account for 80 % of the variation in the original variables. Factor 1 has high loadings for share of foreign-born, share with social benefit, and single mother share. Factor 2 has high loadings for single house share and high disposable income. Finally, Factor 3 has high loadings for high disposable income and share of people with tertiary education.

RESULTS

Below we present the estimates of four logit models with university education in 2010 as the dependent variable. Model 1 uses only individual level variables. Model 2 adds contextual variables based on individualized neighborhoods to the individual level variables. Model 3 is similar to model 2 but uses context variables based on administrative units, SAMS areas. Finally, model 4 is based on model 2 but adds individual-contextual interaction variables.

MODEL COMPARISON

As explained below, the *individual* level variables including parental characteristics are most important for predicting educational achievements.

The first logistic regression (model 1) shows the strongest individual level effects for university education for adolescents with university-educated parents, and for girls; see Table 3. Significant but negative effects are found for adolescents with parents receiving social allowances and those with non-employed parents as well as those with single mothers. These results are supported by earlier research on Swedish data (Andersson and Subramanian, 2006).

Model 2 includes both individual as well as contextual level variables. Here too university-

educated parents are strongly and positively associated with an adolescents' university education in 2010. Also comparable with the model including individual level variables only, is negative effects from non-employed parents and parents receiving social allowances. In addition, being a girl or a boy strongly affects the likelihood of having a later university education. Finally for models 3 and 4, results from individual level variables are generally the same; having parents with university education, and whether one was a boy or a girl, are important factors in determining whether adolescents achieve a university education.

INSERT TABLE 3. HERE

Turning now to the contextual effect estimates we begin by comparing the log likelihood values across our four models. The most important finding here is that the use of a multi-scalar measurement of context based on individualized neighborhoods results in a drastic increase in estimated context effects on adolescents' educational achievements compared to the standard approach based on fixed geographical sub-divisions. This can be seen by comparing the full model log-likelihood values (see Table 3, bottom) for models 3 (SAMS based) and model 2 (Equipop), multi-scalar.

Compared to model 1 with individual level variables, the contextual level variables in model 2 add 273 to the full model log likelihood; see Table 3. Also taking into account that model 2 has more parameters than model 3, this addition is larger than an increase of 73, the result obtained using context measures based on SAMS areas (Busemeyer and Wang, 2000, p. 176). A likely explanation for the larger effect-size estimated for model 2 is that multi-scalar contextual measures provide a better representation of how socio-spatial conditions of relevance for individual educational careers vary between locations. If this is true, our finding of a larger effect size compared to a traditional approach based on SAMS areas aggregates is what should be expected.

CONTEXTUAL FACTORS

As shown in Table 3, model 2, it is the first three contextual factors that have the largest effect on educational achievement. The strongest effects are found for Factor 1 *Elite areas*. This large effect of growing up in an area with a high proportion of individuals with university education, high disposable income, and few non-employed fits with the idea that individuals' life choices are influenced by the choices of their peers and by norms that are present in their residential context.

The Elite areas are important in all counts of closest neighbors, from 12 closest up to the city scale of 25,600 closest neighbors. At the lowest scale this can be explained by collective socialization processes and social control. At the city scale, (25,600 neighbors) we suggest that institutional mechanisms such as the availability of universities, public role models and local newspapers are important in shaping the educational outcome.

Factor 3 *Foreign-born*, Factor 4 *Marginal Nearby*, and Factor 5 *Marginal intermediate* all have parameter estimates that are negative, with a strong effect for Factor 3 in particular. These three factors are all associated with a high proportion of foreign-born residents and a high proportion of households receiving a social allowance. Factor 4 and Factor 5 in addition are associated with a high proportion of single mothers and a low proportion of single-family housing. Factor 3 is associated with high loadings across all neighborhood scales, Factor 4 has high loadings only for large scale neighborhoods, and Factor 5 mainly for small- and medium sized neighborhoods. One reason for the negative effects of these factors on the probability of having a university education at age 30 could be that a high presence of marginal groups has negative effects on the school achievements (Sykes and Kuyper, 2009).

Table 3 reports a positive parameter estimate for Factor 6 *Single family housing*. This effect is significant but not strong in comparison to the effects of Factor 1, Factor 2 and Factor 3. A similar effect has been reported earlier by Andersson (2004). Bramley and Karley (2007) have also reported positive effects of home-ownership on educational achievement, and they provide a discussion of possible mechanisms for this positive effect.

The effect of Factor 2 *Low employment in adjacent areas* is about half as strong as the effect of Factor 1. Moreover, the effect of growing up with low income groups in adjacent areas is to increase the likelihood of getting a university degree by age 30. Other estimates that go in the same direction (but are much weaker) are obtained for Factor 8 *Low employment medium scale* and Factor 9 *Marginal medium scale*.

One explanation is that the positive effect on having a university education at age 30 from high values for Factor 2 is not the result of effects on aspiration but instead the results of differences in opportunity structure. In 1995, when our study cohort was 15 years of age, unemployment rates were still high in the aftermath of the early 1990s economic crisis in Sweden. This situation may have stimulated students in regions of high unemployment to consider academic studies as a more secure path to employment than a non-academic career. On the other hand, students living in regions with low unemployment and high labor demand may have been able to secure employment without the need for costly academic studies. The opportunity structure explanation is supported by the fact that Factor 2 is associated with high levels of non-employment, not in students' close neighborhood but mainly in neighborhoods of up to 25,600 people. In contrast, the effect of a low level of employment in the close neighborhood is negative (but weak), as shown by the estimate for Factor 7 *Low employment small-scale*.

Finally, the estimate for Factor 10 *Non-Academic Elite* is negative, a fact that is associated with high income but not with high levels of education of parents. This corroborates the view that high income *per se* does not imply that you have strong norms concerning the values of education in the Swedish case.

INTERACTIONS

In Table 3 we present model 4 where the projected effects for some of the contextual variables have been allowed to depend on the gender and parental education of the students. The analysis shows that the strength of the contextual effects varies according to gender and parental education.

The strongest effect of Factor 1 *Elite areas* is found for men with university-educated parents. The effect is weaker for women with university-educated parents. This is a group that, irrespective of context, has a high propensity to attain a university degree. But the effect is even weaker for men with parents lacking a university degree. This group, thus, is less influenced by an elite environment. One interpretation of this pattern is that elite areas can help to tip the balance for groups that are willing to consider the idea of a university education. This fits with the fact that women with parents lacking a university degree also experience a relatively strong effect of growing up in an elite area.

A tipping-the-balance pattern is also true for Factor 3 *Foreign-born*. Here it is again men with university-educated parents that experience the strongest effect of context, here with a clear negative effect on the probability of achieving a university degree. And again it is women with parents lacking a university degree that experience the second strongest negative effect of a high proportion of foreign-born residents in the neighborhood. For men with parents lacking a university degree, local context as measured by Factor 3 is of smaller importance.

However, for men with parents lacking a university degree Factor 2 *Low employment in adjacent areas* plays an important role. Indeed, the effect of Factor 2 is much stronger for this group than for any other group. This can be seen as favoring the opportunity structure argument for the positive effect of low local employment levels on the probability of getting a university education. The idea would be that the risk of becoming unemployed after school could help young men to overcome barriers to higher education that are linked to their gender and to parental education.

A MULTILEVEL APPROACH

To further analyze the finding of larger contextual effects we tested the same data in a multilevel model. The multilevel approach is common in neighborhood effect literature because it offers a way of analyzing data in hierarchical structures, for example, individuals in neighborhoods, in municipalities, in counties etc. The results can thereafter be interpreted as variance explained at different geographical levels. In this particular analysis the individuals in the 1980s cohort constitute the individual level and the SAMS areas constitute the second, contextual level (7 704 areas). Because the individualized neighborhoods are flexible in size they could not be used as a hierarchical level in the model. Instead, mean values from Equipop over SAMS areas were used.

In an empty model, no explanatory variables included, the unexplained variance of educational achievements for the 1980s cohort was 5.8% at the contextual level. The rest of the variance in educational achievement was attributed to the individual level. The level of variance of around 5% is found in other Swedish studies using multilevel approaches (Andersson and Subramanian, 2006; Bergsten, 2010), and in a study of the Oslo region a larger contextual level variance of 15% was found (Brattbakk and Wessel, 2013). Furthermore, studies in other

contexts, such as the United States, show higher contextual level proportions of variance for different outcomes. Thus, it is commonly believed that the welfare state regimes produce more equal societies with lower contextual effects (Sampson, 2012).

When we included individual level variables to explain university education in 2010 the contextual level variance was reduced to 2 percent. This remaining unexplained variance at the contextual level was then to be tested with both SAMS area variables and the individualized neighborhoods/factors (as means) to try to reduce the unexplained variance. Not only the log likelihood test above, but this test also showed individualized neighborhoods to be a better measure of context; it could explain more variance than the SAMS areas. Of the remaining unexplained variance at the contextual level the individualized (Equipop) measure captured 35% whereas the SAMS areas captured 8%. As stated above we consider the individualized and scalable measure from Equipop efficient in showing what neighborhood effects are expected to be.

INSERT TABLE 4. HERE

Because of the seemingly low remaining variance at the contextual level it is worth describing actual consequences in terms of the difference in the shares of individuals that achieved a university education in 2010; see Table 4. Note that for individuals living in the lowest 10th percentile of Elite area factor loadings, 45.6% were obtaining university education. This is significant when compared to Elite areas with the highest factor loadings, (90th percentile), where 72.3% had obtained a university education at 30 years of age. The difference of 26 percentage points should not be neglected as a contextual effect.

Factor 3, including foreign-born and to some extent parents with social allowances, has a negative association with university education. As a consequence, in areas with low factor 3 loadings (lowest 10th percentile), 63.7% obtained a university education. Adolescents living in

areas highly loaded with Factor 3 had a 9 percentage points lower probability of becoming university-educated. As for the factor 6 describing loadings of single family housing, the difference between the least loaded 10th percentile's proportions compared to the highest 90th percentile was smaller.

CONCLUDING DISCUSSION

In recent years, increasing criticism has been directed at the neighborhood effect studies (methods) and also the phenomenon as such (Hedman, 2011). Criticism has touched on: the stability of effects over time, subjects of cross-sectional studies, and the fact that measured effects are quite small or non-existent (Brännström, 2004; Hedman, 2011). In the same vein, different groups of inhabitants in a neighborhood might be influenced differently, which has not been researched sufficiently (Bergsten, 2010; Galster, Andersson and Musterd, 2010; Sykes and Kuyper, 2009). The policy of mixing the population and mixing housing (tenure forms) has been criticized for not being an effective policy against segregation because neighborhood effects are not satisfactorily assessed.

There are several reasons that neighborhood effects still remain a large research area and a matter of political interest and debate. One reason is that qualitative research, as well as lived experiences, show people that there is greater importance to where they have been growing up than has been proved scientifically. A second reason is that inequality of outcomes in e.g. education due to where adolescents live is against the expectations and goals set by welfare states and against national policies of education. In the US, for example, segregated schools were declared unconstitutional because of their detrimental effect on educational equality (Brown v. Board of Education, 1954; Clark, 1987; Coleman, 1966). A third reason is that mixing residential areas and schools is a direct policy and planning measure that is in constant debate.

Mixing strategies are questioned and need scientific support if continued in (especially) times of economic crisis (Galster, 2007; Holmqvist and Bergsten, 2009).

In this paper we have used Swedish register data to analyze contextual effects on educational achievement for a cohort born in 1980. For this cohort, neighborhood exposure was measured in 1995 (at age 15) and educational achievement was assessed in 2010 (at age 30). An important innovation in this study is that context is not measured using aggregate values for statistical areas. Instead, we have used statistics computed for individualized neighborhoods that have been expanded to include between 12 and 25,600 nearest neighbors. With this method—which departs significantly from the standard approach—we have obtained results that in many ways improve those obtained in earlier studies.

Our first finding is that the strength of the estimated contextual effects increases when statistics based on scalable individualized neighborhoods are used to measure context. Compared to traditional, area-based measures the effect is about three times stronger. These stronger effects are also tested with a multilevel approach.

Second, the stronger overall effect allows us to get significant estimates for several contextual indicators when they are used simultaneously in the same model. To avoid problems of multicollinearity, earlier studies of contextual effect on educational achievement have often included only one contextual variable per model. However, using individualized neighborhoods of varying size makes it possible to capture context at different scale levels: variation in the composition of the 50 nearest neighbors, variation in the composition of the 100 nearest neighbors, etc. This increases the amount of contextual variation that is used to estimate contextual effects, and with increased variance in the explanatory variables the problem of multicollinearity can be reduced. Hence, we have been able to show that high levels of

education, low levels of marginality, and a dominance of single family housing in the neighborhood all have separate, positive effects on educational achievement.

Third, the use of individualized neighborhoods has allowed us to explore how contextual effects are linked to scale. Most important here is the unexpected finding that low employment levels among the 1000+ nearest neighbors can have a positive effect on educational achievement.

Fourth, stronger overall contextual effects have allowed the estimation of interaction effects. Here our results indicate that the effects of a specific neighborhood context can be of great importance for one group but less important for a different group. As we see it, this finding provides a strong rationale for a new generation of neighborhood effect studies that focus less on diffuse overall neighborhood effects and more on how specific circumstances influence different groups.

Taken together, we would claim that the findings presented above suggest that a revised methodology that takes advantage of the possibilities offered by the use of individualized neighborhoods would not only provide neighborhood effect studies with a new lease of life, but would also help to make neighborhood effect studies a more central concern for social science research in general.

We acknowledge that the application of the individualized-neighborhood methodology can be difficult in circumstances where researchers do not have access to geo-coded individual level data. However, even if the computation of measures based on individualized neighborhoods requires data that is sensitive from an integrity point of view, this is not the case with the resulting aggregate measures. Thus, an important advantage with measures based on individualized neighborhoods is that they can provide very detailed geographical information

about the variation in a neighborhood context, and this information need not be sensitive since it is based on population aggregates.

REFERENCES

- AINSWORTH, J. W. (2002) Why does it take a Village? The Mediation of Neighborhood Effects on Educational Achievement, *Social Forces*, 81(1), pp. 117-152.
- ANDERSSON, E. (2004) From Valley of Sadness to Hill of Happiness - The Significance of Surroundings for Socio-economic Career, *Urban Studies*, 41(3), pp. 641-659.
- ANDERSSON, E. and SUBRAMANIAN, S. V. (2006) Explorations of neighborhood and educational outcomes for young Swedes, *Urban Studies*, 43(11), pp. 2013-2025.
- ANDERSSON, R. and MUSTERD, S. (2010) What scale matters? Exploring the relationships between individuals' social position, neighbourhood context and the scale of neighbourhood, *Geografiska Annaler: Series B, Human Geography*, 92(1), pp. 23-43.
- BERGSTEN, Z. (2010). *Better prospects through social mix? Mixed neighbourhoods and neighbourhood effects : An analysis of the purpose and effects of social mix policy*, Uppsala: Uppsala University.
- BOLSTER, A., BURGESS, S., JOHNSTON, R., JONES, K., PROPPER, C., and SARKER, R. (2007) Neighbourhoods, households and income dynamics: a semi-parametric investigation of neighbourhood effects, *Journal of Economic Geography*, 7(1), pp. 1-38.
- BRAMLEY, G. and KOFI KARLEY, N. (2007) Homeownership, poverty and educational achievement: School effects as neighbourhood effects, *Housing Studies*, 22(5), pp. 693-721.
- BRATTBAKK, I. and WESSEL, T. (2013) Long-term Neighbourhood Effects on Education, Income and Employment among Adolescents in Oslo, *Urban Studies*, 50(2), pp. 391-406.
- BROWN v. BOARD OF EDUCATION. (1954). In *US*, 347:483, Supreme Court.
- BRÄNNSTRÖM, L. (2004) Poor Places, Poor Prospects? Counterfactual Models of Neighbourhood Effects on Social Exclusion in Stockholm, Sweden, *Urban Studies*, 41(13), pp. 2515-2537.
- BRÄNNSTRÖM, L. (2008) Making Their Mark: The Effects of Neighbourhood and Upper Secondary School on Educational Achievement, *Eur Sociol Rev*, 24(4), pp. 463-478.
- BUSEMEYER, J. R. and WANG, Y.-M. (2000) Model comparisons and model selections based on generalization criterion methodology, *Journal of Mathematical Psychology*, 44(1), pp. 171-189.
- BYGREN, M. and SZULKIN, R. (2010) Ethnic environment during childhood and the educational attainment of immigrant children in Sweden, *Social Forces*, 88(3), pp. 1305-1329.
- CHAIX, B., MERLO, J., SUBRAMANIAN, S. V., LYNCH, J., and CHAUVIN, P. (2005) Comparison of a Spatial Perspective with the Multilevel Analytical Approach in Neighborhood Studies: The Case of Mental and Behavioral Disorders due to Psychoactive Substance Use in Malmö, Sweden, 2001, *American Journal of Epidemiology*, 162(2), pp. 171-182.
- CLARK, W. A. V. (1987) Demographic change, attendance area adjustment and school system impacts, *Population Research and Policy Review*, 6(3), pp. 199-222.
- COLEMAN, J. S. (1966). *Equality of Educational Opportunity, The Equality of Educational Opportunity Study (EEOS)* Washington, DC: United States Department of Education.
- CRANE, J. (1991) The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping Out and Teenage Childbearing, *American Journal of Sociology*, 96(5), pp. 1226-1259.
- EVANS, G., WELLS, N. M., and MOCH, A. (2003) Housing and Mental Health: A Review of the Evidence and Methodological and Conceptual Critique, *Journal of Social Issues*, 59(3), pp. 475-500.
- GALSTER, G. (2007) Should Policy Makers Strive for Neighborhood Social Mix? An Analysis of the Western European Evidence Base, *Housing Studies*, 22(4), pp. 523-545.
- GALSTER, G. (2008) Quantifying the effect of neighbourhood on individuals: Challenges, alternative approaches, and promising directions, *Schmollers jahrbuch*, 128(1), pp. 7-48.
- GALSTER, G., ANDERSSON, R., and MUSTERD, S. (2010) Who is affected by neighbourhood income mix? Gender, age, family, employment and income differences, *Urban Studies*, 47(14), pp. 2915-2944.
- GALSTER, G. C. (2012) The Mechanism(s) of Neighbourhood Effects: Theory, Evidence, and Policy Implications, in M. VAN HAM, D. MANLEY, N. BAILEY, L. SIMPSON and D. MACLENNAN (Ed.) *Neighbourhood Effects Research: New Perspectives*, pp. 23-56: Springer Netherlands.
- GALSTER, G. C. and SANTIAGO, A. M. (2006) What's the 'hood got to do with it? Parental perceptions about how neighborhood mechanisms affect their children, *Journal of Urban Affairs*, 28 (3), pp. 201-226.

- HEDMAN, L. (2011) *Residential mobility and neighbourhood effects : a holistic approach [Elektronisk resurs]* Uppsala: Department of Social and Economic Geography, Uppsala University.
- HOLMQVIST, E. and BERGSTEN, Z. (2009) Swedish social mix policy: a general policy without an explicit ethnic focus, *Journal of Housing and the Built Environment*, 24(4), pp. 477-490.
- IMMERGLUCK, D. (1998) Neighborhood Economic Development and Local Working: The Effect of Nearby Jobs on Where Residents Work, *Economic Geography*, 74, pp. 170-187.
- JACOBS, J. (1993) *The Death and Life of Great American Cities* New York: Modern Library.
- JOSEPH, M. L., CHASKIN, R. J., and WEBBER, H. S. (2007) The Theoretical Basis for Addressing Poverty Through Mixed-Income Development, *Urban Affairs Review*, 42(3), pp. 369-409.
- KWAN, M.-P. (2012) The Uncertain Geographic Context Problem, *Annals of the Association of American Geographers*, 102(5), pp. 958-968.
- LUDWIG, J. (1999) Information and inner city educational attainment, *Economics of Educational Review*, 18(1), pp. 17-30.
- MACALLISTER, I., JOHNSTON, R. J., PATTIE, C. J., TUNSTALL, H., DORLING, D. F. L., and ROSSITER, D. J. (2001) Class Dealignment and the Neighbourhood Effect: Miller Revisited, *British Journal of Political Science*, 31(01), pp. 41-59.
- MALMBERG, B., ANDERSSON, E., and ÖSTH, J. (2011) To what extent does the level of segregation vary between different urban areas? Introducing a scalable measure of segregation. In *European Network for Housing Research Conference*. Toulouse, <http://www.enhr2011.com/sites/default/files/Paper-BoMalmberg-WS16.pdf>.
- OPENSHAW, S. (1984) The modifiable areal unit problem, *CATMOG (Concepts and Techniques in Modern Geography)*, *Geo Abstracts* 40.
- PUTNAM, R. D. (1993) The prosperous community: social capital and public life, *American prospect* 13, pp. 35-42.
- PUTNAM, R. D. (2007) E pluribus unum: Diversity and community in the twenty-first century the 2006 Johan Skytte Prize Lecture, *Scandinavian political studies*, 30(2), pp. 137-174.
- SAMPSON, R. J. (2012) *Great American City: Chicago and the Enduring Neighborhood Effect* Chicago: University of Chicago Press.
- SAMPSON, R. J., MORENOFF, J. D., and GANNON-ROWLEY, T. (2002) Assessing "Neighborhood Effects": Social Processes and New Directions in Research, *Annual Review of Sociology*, 28, pp. 443-78.
- SELLSTROM, E. and BREMBERG, S. (2006) Is there a "school effect" on pupil outcomes? A review of multilevel studies, *J Epidemiol. Community Health*, 60(2), pp. 149-155.
- SOUTH, S. J., BAUMER, E. P., and LUTZ, A. (2003) Interpreting Community Effects on Youth Educational Attainment, *Youth and Society*, 35(1), pp. 3-36.
- SUNDLÖF, P. (2008). *Segregation and career position: A study of the significance of the neighbourhood context for education, employment and income among young people in the Stockholm region*, Uppsala: Uppsala university.
- SWEDISH NATIONAL AGENCY FOR EDUCATION. (2013) Skolverket, Tabell 1: Skolor och elever läsåret 2012/13. skolverket.se, Swedish national agency for education, Utbildningsstatistiksenheten.
- SYKES, B. and KUYPER, H. (2009) Neighbourhood effects on youth educational achievement in the Netherlands: Can effects be identified and do they vary by student background characteristics?, *Environment and Planning A*, 41(10), pp. 2417 – 2436.
- SYKES, B. and MUSTERD, S. (2011) Examining Neighbourhood and School Effects Simultaneously, *Urban Studies*, 48(7), pp. 1307-1331.
- ÖSTH, J., ANDERSSON, E., and MALMBERG, B. (2013) School Choice and Increasing Performance Difference: A Counterfactual Approach, *Urban Studies*, 50(2), pp. 407-425.
- ÖSTH, J., CLARK, W., and MALMBERG, B. (2013). Measuring the scale of segregation using k-nearest neighbor aggregates, submitted November 2013.
- ÖSTH, J., MALMBERG, B., and ANDERSSON, E. (2011) Introducing Equipop. In *6th international conference on population geographies*.
- ÖSTH, J., MALMBERG, B., and ANDERSSON, E. (forthcoming) Analysing segregation with individualized neighbourhoods defined by population size, in C. D. LLOYD, I. SHUTTLEWORTH and D. WONG (Ed.) *Social-Spatial Segregation: Concepts, Processes and Outcomes*: Policy Press.

CAPTIONS

FIGURE 1. FACTORS AND LOADINGS. (TO REDUCE CLUTTER, THESE GRAPHS ONLY SHOW FACTORS THAT FOR AT LEAST ONE K-LEVEL HAVE A LOADING HIGHER THAN 0.2 OR LOWER THAN -0.2.)

TABLE 1. INDIVIDUAL LEVEL VARIABLES.

TABLE 2. CONTEXT VARIABLES RUN IN EQUIPOP FOR K NEAREST NEIGHBORS IN 1995.

TABLE 3. PARAMETER ESTIMATES FROM FOUR MODELS FOR UNIVERSITY EDUCATION IN 2010.

TABLE 4. AREA DESCRIPTION, EFFECT ON PROPORTION OBTAINING A UNIVERSITY EDUCATION.

TABLE 1. INDIVIDUAL LEVEL VARIABLES.

	N	Minimum	Maximum	Mean	Std. Deviation
Sex	74649	1	2	1.48	0.499
Single mothers	74649	0	1	0.15	0.354
Univ. education 2010	74649	0	1	0.51	0.500
Parent with univ. education	74649	0	1	0.41	0.492
Parent in visible minority	74649	0	1	0.02	0.145
Parent with social allowance	74649	0	1	0.10	0.299
Parent foreign-born	74649	0	1	0.17	0.378
Parent non-employed	74649	0	1	0.25	0.432
Disposable income decile	74649	0	100	50.78	28.465

TABLE 2. CONTEXT VARIABLES RUN IN EQUIPOP FOR K NEAREST NEIGHBORS IN 1995.

Variable	Description	Year	Population	Number of neighbors (k)
Education	1 = university/college, 0 = not university/college	1995	>25 years	12, 25, 50, 100, 200, 400, 800, 1600, 3200, 6400, 12800, 25600
Social allowance	1 = social allowance	1995	all	12, 25, 50, 100, 200, 400, 800, 1600, 3200, 6400, 12800, 25600
Family type	1 = single mother	1995	>25 years	12, 25, 50, 100, 200, 400, 800, 1600, 3200, 6400, 12800, 25600
Disposable income	percentiles	1995	>25 years	12, 25, 50, 100, 200, 400, 800, 1600, 3200, 6400, 12800, 25600
Born abroad	1 = born abroad (not Sweden)	1995	all	12, 25, 50, 100, 200, 400, 800, 1600, 3200, 6400, 12800, 25600
Unemployed	1 = non employed 0 = employed	1995	>25 years	12, 25, 50, 100, 200, 400, 800, 1600, 3200, 6400, 12800, 25600
Housing	1 = Single owner occupied housing, 0 = other types of housing	1996	All	12, 25, 50, 100, 200, 400, 800, 1600, 3200, 6400, 12800, 25600

TABLE 3. PARAMETER ESTIMATES FROM FOUR MODELS FOR UNIVERSITY EDUCATION IN 2010.

Model 1. Parameter estimates individual level for tertiary education in 2010. Model 2. Parameter estimates individual level and contextual level for university education in 2010. Model 3. parameter estimates individual level and SAMS-area level for tertiary education in 2010. Model 4. Parameter estimates individual level, contextual level and interactions for university education in 2010.

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq	Estimate	Std Error	ChiSquare	Prob>ChiSq	Estimate	Std Error	ChiSquare	Prob>ChiSq	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	0.014	0.019	0.56	0.4555	0.012	0.019	0.36	0.5459	0.022	0.019	1.27	0.2588	0.011	0.019	0.35	0.5527
Parent with univ. education	0.634	0.009	5302.70	<.0001	0.605	0.009	4697.70	<.0001	0.625	0.009	5061.90	<.0001	0.604	0.009	4663.40	<.0001
Parent visible minority	-0.152	0.059	6.70	0.0096	-0.166	0.059	7.84	0.0051	-0.141	0.059	5.68	0.0172	-0.168	0.059	7.99	0.0047
Parent social allowance	-0.581	0.031	360.91	<.0001	-0.526	0.031	281.07	<.0001	-0.530	0.031	290.33	<.0001	-0.526	0.031	281.08	<.0001
Parent foreign born	0.016	0.023	0.47	0.4907	0.111	0.025	20.26	<.0001	0.042	0.024	3.08	0.0790	0.110	0.025	19.99	<.0001
Parent non employed	-0.242	0.020	141.75	<.0001	-0.222	0.020	117.47	<.0001	-0.230	0.020	126.88	<.0001	-0.222	0.020	117.66	<.0001
Disposable income decile	0.007	0.000	553.27	<.0001	0.007	0.000	518.70	<.0001	0.006	0.000	399.60	<.0001	0.007	0.000	513.94	<.0001
Single mothers	-0.427	0.024	322.62	<.0001	-0.387	0.025	247.71	<.0001	-0.390	0.024	256.76	<.0001	-0.386	0.025	246.45	<.0001
Sex	0.391	0.008	2337.9	<.0001	0.394	0.008	2350.30	<.0001	0.392	0.008	2342.90	<.0001	0.397	0.008	2337.20	<.0001
1. Elite areas					0.006	0.000	304.21	<.0001					0.006	0.000	285.41	<.0001
2. Low employment in adjacent areas					0.003	0.000	166.40	<.0001					0.003	0.000	154.60	<.0001
3. Foreign born					-0.003	0.000	159.77	<.0001					-0.003	0.000	149.66	<.0001
4. Marginal nearby					-0.001	0.000	32.66	<.0001					-0.001	0.000	30.22	<.0001
5. Marginal intermediate scale					-0.001	0.000	17.15	<.0001					-0.001	0.000	16.64	<.0001
6. Single family housing					0.003	0.001	17.85	<.0001					0.003	0.001	19.00	<.0001
7. Low employment. small scale					-0.002	0.001	8.33	0.0039					-0.002	0.001	8.39	0.0038
8. Low employment medium scale					0.002	0.000	24.74	<.0001					0.002	0.000	24.51	<.0001
9. Marginal medium scale					0.001	0.000	62.07	<.0001					0.001	0.000	59.91	<.0001
10. Non-academic elite					-0.002	0.001	12.78	0.0004					-0.002	0.001	11.38	0.0007
SAMS_Factor1									-0.046	0.009	26.83	<.0001				
SAMS_Factor2									0.045	0.009	24.21	<.0001				
SAMS_Factor3									0.083	0.009	89.80	<.0001				
Factor 1 -1.01825)*Sex[2]													-0.0003	0.000	0.88	0.3473
Factor 2 -5.74404)*Sex[2]													-0.0003	0.000	1.97	0.1606
Factor 3 -9.90044)*Sex[2]													-0.0003	0.000	2.18	0.1394
Factor 6 +3.4295)*Sex[2]													0.0011	0.000	5.83	0.0158
Factor 1 -1.01825)*Parent_Edu.[1]													0.0005	0.000	2.65	0.1036
Factor 2 -5.74404)*Parent_Edu.[1]													0.0001	0.000	0.54	0.4615
Factor 3 -9.90044)*Parent_Edu.[1]													-0.0004	0.000	3.00	0.0835
Factor 1 -1.01825)*Parent_Edu.[1]*Sex[2]													-0.0009	0.000	9.56	0.002
Factor 2 -5.74404)*Parent_Edu.[1]*Sex[2]													-0.0005	0.000	5.25	0.0219
Factor 3 -9.90044)*Parent_Edu.[1]*Sex[2]													0.0006	0.000	7.25	0.0071
-Log Likelihood, Reduced model	51717				51717				51717				51717			
-Log Likelihood, Full model	45282				45009				45210				44993			
-Log Likelihood, Difference	6434				6707				6507				6722			
Difference from individual model					273				72				288			

TABLE 4. AREA DESCRIPTION, EFFECT ON PROPORTION OBTAINING A UNIVERSITY EDUCATION.

Proportion higher education	Factor 1 Elite areas	Factor 3 Foreign-born	Factor 6 Single family housing
10th percentile mean	45.6%	63.7%	57.3%
mean	59.4%	59.4%	59.4%
90th percentile mean	72.3%	54.6%	61.2%

