AMERICAN SOCIOLOGICAL REVIEW

OFFICIAL JOURNAL OF THE AMERICAN SOCIOLOGICAL ASSOCIATION

ONLINE SUPPLEMENT
to article in

AMERICAN SOCIOLOGICAL REVIEW, 2007, VOL. 72 (JUNE:341–364)

Cultural Context, Sexual Behavior, and Romantic Relationships in Disadvantaged Neighborhoods

David J. Harding
University of Michigan

MEASURING HETEROGENEITY IN PREGNANCY FRAMES

Calculating neighborhood heterogeneity in pregnancy frames is complicated by the ordinal nature of the pregnancy frame measure, which makes the usual measure of variation for interval measures, the variance, inappropriate. I use an ordinal variation measure developed by Blair and Lacy (2000) that measures concentration:

$$l^2 = \frac{\sum_{i=1}^{k-1}(F_i - .5)^2}{(k-1)/4}$$

k is the number of categories (in this case $k = 5$), and F_i is the cumulative proportion for category i (i.e., $F_i = \sum_{j=1}^{i} p_j$ where p_j is the sample proportion for the jth of k categories). The numerator measures the difference between the observed distribution and a distribution with maximum dispersion, which occurs when responses are evenly divided between the two opposite extremes (the minimum dispersion occurs when all responses are in a single category; see Blair and Lacy 2000). The denominator normalizes by dividing by the maximum possible value of the numerator, so that l^2 varies from 0 to 1. Because some neighborhoods have small numbers of adolescents, Blair and Lacy's small sample bias-adjusted l^2 is required:

$$l_u^2 = l^2 - \frac{1-l^2}{N-1}$$

I take $1-l_u^2$ as my measure of neighborhood heterogeneity because l_u^2 is a measure of concentration.[1] Finally, the regression models also require controls for central tendency of the distribution of frames in neighborhoods, for which I use the mean.[2]

One complication in constructing the neighborhood heterogeneity measure by aggregation of individual respondents to the neighborhood level in the Addhealth data is the small number of respondents in some neighborhoods. For about half of the tracts in which at least one Wave I Addhealth respondent lives

[1] Ideally, one would construct gender specific measures of neighborhood cultural heterogeneity of both pregnancy frames and relationship scripts. Unfortunately, doing so with these data would cut in half the number of individuals per neighborhood who contribute to the neighborhood-level measures, significantly reducing the reliability of such measures.

[2] Strictly speaking, the mean is not an appropriate measure of central tendency for an ordinal variable because it assumes that the distances between categories carry information, whereas with an ordinal variable, only information about the order of the categories is present. However, other potential measures of central tendency, such as the mode or median, result in little variation across neighborhoods, and so would not serve as effective controls for central tendency when measuring the impact of neighborhood heterogeneity.

there are no other Addhealth respondents, and therefore it is impossible to measure heterogeneity for such tracts. These tracts are dropped from the analysis. However, since most Addhealth respondents live in tracts with many other respondents, this restriction results in the loss of only about 6 percent of respondents from the multilevel models predicting individual outcomes. For another 25 percent of tracts, there are fewer than five respondents per tract, leading to low reliability of neighborhood level measures created by aggregation. I weight models by the reliability of the mean to account for the low reliability of neighborhood measures when they are constructed from small numbers of respondents.[3]

MEASURING NEIGHBORHOOD RACIAL DIVERSITY

To measure neighborhood racial diversity, I use Simpson's Interaction Index (White 1986, cited in Reardon and Firebaugh 2002). This index is constructed from the percent of neighborhood residents in four racial groups: White (*W*), Black (*B*), Asian (*A*), and Other (*O*):

Racial Diversity Index =

$$100\left(\frac{W}{1-W}+\frac{B}{1-B}+\frac{A}{1-A}+\frac{O}{1-O}\right)$$

This index varies from zero (entire neighborhood population from one group) to 75 (neighborhood population evenly divided between the four groups).[4]

MULTIPLE IMPUTATION OF MISSING VALUES

Several individual, family, and school control variables have missing values. Rather than dropping cases with missing values on control variables, I impute missing values using multiple imputation, currently believed to be among the best methods for dealing with missing data (see Acock 2005 for a nontechnical discussion and references therein for more technical material, such as Little and Rubin 2002; Allison 2002). Multiple imputation involves creating multiple full datasets via MICE (multiple imputation by chained equations), estimating a model using each full dataset, and then combining results across datasets in a way that takes into account the variance in imputed values across datasets.[5] Here I use 10 imputed datasets.

USING SEQUENCE ANALYSIS TO MEASURE DIFFERENCES IN ROMANTIC RELATIONSHIP SCRIPTS

If we think of the ordered elements that constitute a relationship script as a sequence of events, then an optimal matching algorithm can be applied to measure the difference between any two relationship scripts.[6] Optimal matching provides a method for comparing two sequences based on the insertions, deletions, and substitutions required to transform one sequence into the other. An optimal matching algorithm measures the difference between two sequences by determining the minimum combination of insertions, deletions, and substitutions that are required to change one sequence into another. More extensive discussions of optimal matching and sequence analysis can be found elsewhere (Abbott 1990, 1995; Abbott and Hrycak 1990; Sankoff and Kruskal 1983).[7] This discussion is based on those sources and the summary in Stovel, Savage, and Bearman (1996).

An example helps to illustrate the idea. The menu of elements that respondents can select and order in the ideal relationship scripts is:

A: We would go out together in a group
B: I would meet my partner's parents
C: I would tell other people that we were a couple
D: I would see less of my other friends so I could

[3] Supplemental analyses (not shown) indicate that results in Table 3 are not sensitive to the exclusion of tracts with low measurement reliability.

[4] I also experimented with Thiel's Entropy Index (Thiel 1972, cited in Reardon and Firebaugh 2002). This index sums $p(\ln(1/p))$ for each of the four groups. The Simpson index is a better predictor of neighborhood heterogeneity.

[5] I use Royston's (2004) "ice" command in Stata to generate the imputed datasets and HLM6's multiple imputation capabilities to estimate the multiple models and combine results across models.

[6] I thank Hannah Brueckner and Peter Bearman for suggesting the use of sequence analysis methods and for advice on their implementation.

[7] Previous sociological applications using sequence analysis have investigated sequences of jobs within careers (e.g., Abbott and Hrycak 1990; Stovel et al. 1996). These applications involve sequences that explicitly take into account time. In the Addhealth relationship scripts there is no information about the elapsed time between events in the sequence, but there is nothing inherent to the optimal matching algorithm that requires information on elapsed time. The measures constructed here thus focus particularly on the ordering of events in a romantic relationship. This analysis also differs in the purpose of the sequence analysis. Previous applications typically calculate the difference between all possible pairs of sequences and then use cluster analysis to determine a set of typical sequences. Here I am using optimal matching only to measure particular characteristics of individuals and neighborhoods.

spend more time with my partner
E: We would go out together alone
F: We would hold hands
G: I would give my partner a present
H: My partner would give me a present
I: I would tell my partner that I loved him or her
J: My partner would tell me that he or she loved me
K: We would think of ourselves as a couple
L: We would talk about contraception or sexually transmitted diseases
M: We would kiss
N: We would touch each other under our clothing
O: We would have sex
P: My partner or I would get pregnant
Q: We would get married

"EAFMGBHKCLNOIJQP" is an example relationship script; "ABEDGMCJILMNOHKPQ" is another. To change the second script into the first, we can begin by inserting E at the beginning of script two, then substitute F for B, and then substitute M for D. This will align the first four elements of the scripts so they both start with EAFM. Alternatively, we could substitute E for A and then A for F, and then insert an F and then an M. Which of these two sets of operations is optimal (or whether some other is optimal) as well as the cost of the optimal set of operations depends on the relative costs of insertions, deletions, and substitutions. The optimal matching algorithm computes the minimum cost of all possible sets of operations that change one sequence into another given defined costs of insertions, deletions, and substitutions. This provides a quantitative measure of the difference between two relationship scripts. I do not discuss the mechanics of the algorithm here. Sankoff and Kruskal (1983) provide an extensive discussion of algorithms.

Following Stovel and colleagues (1996), substitution costs are defined using observed transition probabilities in the data, such that events frequently following one another in the set of observed sequences are assigned lower transition costs than those rarely observed following one another. For example, the event "going out together alone" more frequently follows "going out together in a group" than it follows "having sex," so substituting "going out together alone" for "going out together in a group" has a lower substitution cost than substituting "having sex" for "going out together in a group." I derive the transition cost matrix from all observed ideal relationship scripts in the Wave 1 Addhealth data. Substitution costs are defined as the inverse of the transition probability. To reduce extreme values, substitution costs are then logged using the natural logarithm. The resulting substitution costs range from .70 to 6.97 in the ideal relationship transition data. The substitution cost matrix is displayed in Table S1. Following Stovel and colleagues (1996), insertion and deletion costs are defined as the maximum substitution cost (6.97). Following Abbott and Hrycak (1990), the difference between two scripts is normalized by dividing by the length of the longer script, so the final difference measure is the average cost per event in the longer script.

To measure the difference between a particular respondent's ideal and actual relationship, I use the optimal matching algorithm. For construction of this measure, substitution costs are also calculated based on the set of all observed ideal relationship scripts in the Wave 1 Addhealth data. One difficulty arises in this comparison because, as described in the main text, the Wave 1 ideal relationships and Wave 2 actual relationships have different sets of events from which the respondent can choose. The only practical solution is to remove events from the scripts that are not common to both scripts prior to comparison. A second issue is how to deal with an actual relationship that has fewer events than a respondent's ideal relationship. An actual relationship may end due to decisions made by one's partner, or the relationship may still be ongoing at the time of the Wave 2 interview. I shorten the ideal relationship to the number of events contained within the actual relationship by removing events at the end of the ideal relationship script.

Neighborhood heterogeneity in relationship scripts is undefined for neighborhoods with only one Addhealth respondent (or only one Addhealth respondent with complete ideal relationship script data).

If n_j is the number of respondent pairs in neighborhood j, σ_u^2 is the within neighborhood variance, and σ_e^2 is the between neighborhood variance, the reliability of the neighborhood mean is $\sigma_u^2/(\sigma_u^2+\sigma_e^2/n_j)$. Supplemental analyses (not shown) indicate that models in Table 4 are not sensitive to the exclusion of tracts with low measurement reliability.

This analysis uses the implementation of optimal matching in the software Transition Data Analysis (Rohwer and Poetter 2005).

Table S1. Substitution Cost Matrix for Ideal Relationship Scripts

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
A	.00	1.71	2.32	3.18	1.47	1.85	3.11	3.47	3.79	3.86	2.42	3.69	3.30	4.47	5.70	6.68	5.14
B	2.17	.00	1.83	3.11	1.87	2.44	2.59	2.86	3.21	3.19	2.33	3.01	3.27	3.93	5.15	6.87	4.29
C	2.47	2.01	.00	2.38	2.06	1.94	2.70	3.01	3.22	3.21	2.19	3.52	2.90	4.13	5.68	6.92	5.28
D	3.00	2.60	2.68	.00	1.66	2.57	2.36	2.80	2.85	3.07	2.73	2.58	3.08	2.95	4.40	4.82	3.80
E	2.80	2.73	2.86	3.24	.00	1.17	2.52	3.05	3.25	3.45	2.37	3.45	2.15	4.03	5.64	6.75	5.26
F	3.08	2.80	2.72	3.97	2.26	.00	2.12	3.11	3.03	3.49	2.24	4.06	1.13	4.80	6.25	6.97	6.01
G	3.94	3.30	3.99	3.90	3.53	3.36	.00	.61	2.93	2.37	3.47	3.15	3.27	3.76	5.24	6.18	5.04
H	3.74	3.13	3.82	3.63	3.41	3.47	1.08	.00	1.79	2.52	2.82	2.73	2.94	3.38	4.84	5.99	4.58
I	4.34	3.36	3.99	4.06	3.71	3.62	3.06	2.92	.00	.69	3.12	2.45	2.92	3.08	4.20	6.02	4.22
J	4.12	3.37	3.80	3.94	3.68	3.48	2.90	3.36	.94	.00	2.03	2.37	2.56	3.08	4.50	6.13	4.43
K	3.17	2.78	.97	3.82	2.99	2.50	3.19	3.49	3.24	3.18	.00	2.54	2.25	4.27	5.67	6.52	5.18
L	3.80	3.37	3.81	3.55	3.44	3.65	3.47	3.43	3.25	2.85	3.22	.00	1.98	1.21	1.96	4.79	3.06
M	3.04	2.34	2.74	3.44	2.70	2.85	2.64	2.83	2.73	2.73	2.08	2.44	.00	2.15	3.89	6.14	3.84
N	4.36	3.85	4.53	4.07	4.81	5.06	3.76	3.98	3.50	3.42	4.31	1.94	3.90	.00	.57	3.56	3.08
O	3.53	3.11	3.82	3.04	4.30	4.67	3.16	3.26	3.32	3.07	3.68	3.51	4.41	2.61	.00	1.29	1.37
P	3.34	3.55	4.10	2.86	4.30	4.70	3.72	3.72	3.93	4.19	4.14	3.16	4.30	2.92	2.01	.00	.70
Q	3.73	4.05	4.10	3.00	4.19	4.22	3.82	3.89	3.78	4.19	4.08	2.35	3.44	1.79	1.37	1.53	.00

Descriptions of Individual, Family, and School Control Variables

(All Measured at Wave 1)

Individual Characteristics:

Race/Ethnicity: A set of indicator (0/1) variables for the adolescent's race and ethnicity. In Addhealth, the adolescent can self-identify as belonging to one or more categories, including White, Black, Native American, Asian, or other Race. White is the omitted category in models. I also include an indicator variable for those adolescents who choose more than one category. The adolescent can also choose to identify as Hispanic/Latino or not.
Immigrant: An indicator for those born outside the United States.
Low Birth Weight: An indicator for weighing less than 88 ounces (5.5 lbs.) at birth.
Mother's Age at Birth: The age in years of the mother when the adolescent was born.

Family Characteristics:
Home Language not English: An indicator variable for those adolescents whose families do not speak English at home.
Household Size: The number of persons living in the adolescent's household.
Household Type: A set of indicator variables for the family type: Married, Single Parent, and Other (which includes stepparent families). Married is the omitted category.

Parent variables are based on the primary residential parent who completed the parent questionnaire, usually the biological mother but sometimes the father or other caretaker.
Parent Immigrant: Primary parent not born in the United States.
Parent Education: A set of indicator variables for the primary parent's completed level of education: Less than High School, High School Graduate, Some College or Trade School, and College Graduate. Less than high school is the omitted category.
Parent Professional Occupation: Primary parent currently works in a managerial or professional occupation.
Parent Disabled: Primary parent is mentally or physically handicapped.
Parent Welfare Receipt: Primary parent currently receives welfare, either for self or for the adolescent.
Log Family Income: The natural logarithm of the household's total income in thousands of dollars, as reported in the parent questionnaire.

School Characteristics:
Urbanicity: A set of indicator variables for the location of the school: Urban, Suburban, or Rural. Suburban is the omitted category.
School Size: A set of indicator variables for the number of students at the school: Small (< 400), Medium (400–1,000), and Large (> 1,000). Medium is the omitted category.
Cumulative Dropout Rate: The proportion of students who begin the school in its lowest grade who complete its highest grade.
Percent College-Prep Program: The proportion of 12th graders who are enrolled in an academic or college-prep program.
Catholic School: An indicator for Catholic schools.
Private School: An indicator for all other nonpublic schools.

Supplementary Tables

Table S2. Descriptive Statistics for Variables in Table 2

Pregnancy Frame Model (N = 1,322)	Mean	SD	Min.	Max.
Pregnancy Frame Heterogeneity	.5338	.3285	0	1
Nhood Disadvantage Scale	0	1	–3.0494	5.4209
Percent Hispanic	13.6931	22.0561	0	96.2677
Percent Hispanic Squared	673.6044	1750.7110	0	9267.4630
Racial Diversity Index	26.8420	20.8092	0	74.3581
Percent Foreign Born	11.3441	16.3284	0	86.9001
Percent Owner-Occupied	59.6721	22.7260	0	97.8407
Percent Units Occupied 5 Years	48.2259	13.8744	9.7990	97.5802
Reliability (weight)	.2523	.2153	.0559	.9345
Relationship Script Model (N = 1,367)	Mean	SD	Min.	Max.
Relationship Script Heterogeneity	0	1	–2.8475	6.3506
Nhood Disadvantage Scale	0	1	–2.9985	5.4274
Percent Hispanic	13.4554	21.9078	0	96.2677
Percent Hispanic Squared	660.6486	1735.4820	0	9267.4630
Racial Diversity Index	26.6161	20.7642	0	74.3581
Percent Foreign Born	11.1963	16.1379	0	86.9001
Percent Owner-Occupied	60.1739	22.5801	0	97.8407
Percent Units Occupied 5 Years	48.2388	13.9728	9.7990	100.0000
Reliability (weight)	.5076	.3476	.1090	.9998

Table S3. Descriptive Statistics for Variables in Table 3

Independent Variables	Mean	SD	Min.	Max.
Individual/Family Level Variables (N = 9,281)				
Sexual Activity Between Waves 1 and 2	.43	.49	0	1
Pregnancy Frame	4.21	1.00	1	5
Female	.50	.50	0	1
Age	16.25	1.04	15	21
Hispanic	.18	.39	0	1
Black	.21	.41	0	1
Native American	.04	.18	0	1
Asian	.08	.28	0	1
Other Race	.10	.30	0	1
Multi Race	.05	.21	0	1
Home Language Not English	.13	.34	0	1
Immigrant	.10	.30	0	1
Household Size	4.72	1.68	1	21
Single Parent Household	.24	.43	0	1
Other Household Type	.23	.42	0	1
Parent Immigrant	.22	.41	0	1
Parent Education – HS Grad	.29	.45	0	1
Parent Education – Some College	.27	.44	0	1
Parent Education – College	.23	.42	0	1
Parent Professional/Managerial Occ.	.32	.47	0	1
Parent Disabled	.05	.22	0	1
Family Welfare Receipt	.11	.31	0	1
Log Family Income	3.50	.87	0	6.91
Low Birth Weight	.11	.32	0	1
Mother's Age at Birth	25.5	5.40	5	53
Neighborhood Level Variables (N = 1,357)				
Neighborhood Disadvantage Scale	0	.70	–2.22	3.91
Pregnancy Frame Nhood Heterogeneity	.49	.30	0	1
Pregnancy Frame Nhood Mean	4.19	.61	1	5
Nhood Reliability (weight)	.30	.23	.06	.93
School Level Variables (N = 142)				
Urban	.30	.46	0	1
Rural	.13	.34	0	1
Small	.21	.41	0	1
Large	.31	.46	0	1
Cumulative Dropout Rate	7.80	11.2	0	68.52
Percent in College-Prep Program	40.44	34.15	0	100
Catholic School	.04	.20	0	1
Private School	.02	.14	0	1

Table S4. Control Variable Coefficients for Models in Table 3

Independent Variables	Model 1	Model 2	Model 3
Individual/Family Level Variables			
Female	.1697*	.1702*	.1682*
	(.0554)	(.0556)	(.0561)
Age	.3007*	.3001*	.3004*
	(.0230)	(.0231)	(.0232)
Hispanic	.1040	.0962	.0947
	(.0995)	(.0993)	(.0990)
Black	.1617*	.1498	.1392
	(.0816)	(.0825)	(.0823)
Native American	–.0662	–.0735	–.0836
	(.1144)	(.1138)	(.1141)
Asian	–.3709*	–.3787*	–.3768*
	(.1102)	(.1118)	(.1102)
Other Race	.0636	.0576	.0553
	(.0762)	(.0776)	(.0766)
Multi Race	.2394*	.2479*	.2539*
	(.1169)	(.1173)	(.1170)
Home Language Not English	–.3728*	–.3794*	–.3922*
	(.1130)	(.1144)	(.1137)
Immigrant	–.4069*	–.4113*	–.4065*
	(.0625)	(.0621)	(.0622)
Household Size	–.0351	–.0365	–.0373
	(.0151)	(.0152)	(.0152)
Single Parent Household	.3649*	.3535*	.3568*
	(.0903)	(.0921)	(.0918)
Other Household Type	.3977*	.3923*	.3921*
	(.0743)	(.0761)	(.0776)
Parent Immigrant	–.0241	–.0193	–.0166
	(.0822)	(.0825)	(.0824)
Parent Education – HS Grad	.1368	.1346	.1323
	(.0943)	(.0935)	(.0931)
Parent Education – Some College	.1233	.1258	.1221
	(.0901)	(.0902)	(.0900)
Parent Education – College	.0550	.0651	.0680
	(.1073)	(.1077)	(.1079)
Parent Professional/Managerial Occ.	–.0482	–.0499	–.0444
	(.0636)	(.0627)	(.0628)
Parent Disabled	–.1206	–.1298	–.1337
	(.1017)	(.0960)	(.0951)
Family Welfare Receipt	.0531	.0507	.0471
	(.0735)	(.0740)	(.0733)
Log Family Income	–.0168	–.0299	–.0252
	(.0437)	(.0394)	(.0395)
Low Birth Weight	–.1944*	–.1788*	–.1805*
	(.0848)	(.0819)	(.0817)
Mother's Age at Birth	–.0173*	–.0185*	–.0182*
	(.0063)	(.0067)	(.0066)
School Level Variables			
Urban	–.0968	–.0992	–.1134
	(.0865)	(.0870)	(.0878)
Rural	–.0389	–.0455	–.0616
	(.1014)	(.1006)	(.0988)
Small	–.0801	–.0773	–.1047
	(.1781)	(.1787)	(.1721)
Large	–.0253	–.0234	.0005
	(.0790)	(.0783)	(.0779)
Cumulative Dropout Rate	.0008	.0005	–.0001
	(.0034)	(.0034)	(.0033)

(continued on the next page)

Table S4. *(continued)*

Independent Variables	Model 1	Model 2	Model 3
Percent in College-Prep Program	.0003	.0006	.0008
	(.0015)	(.0015)	(.0015)
Catholic School	–.0153	–.0211	.0237
	(.1701)	(.1706)	(.1761)
Private School	–.2520	–.2342	–.1444
	(.3248)	(.3347)	(.3384)

Table S5. Descriptive Statistics for Variables in Table 4

Independent Variables	Mean	SD	Min.	Max.
Individual/Family Level Variables (N = 5,916)				
Ideal and Actual Relationship Script Difference	1.97	.60	0	5.9
Female	.51	.50	0	1
Age	15.41	1.56	11	20
Hispanic	.15	.36	0	1
Black	.21	.41	0	1
Native American	.04	.19	0	1
Asian	.06	.23	0	1
Other Race	.09	.28	0	1
Multi Race	.05	.22	0	1
Home Language Not English	.09	.28	0	1
Immigrant	.07	.25	0	1
Household Size	4.64	1.62	1	18
Single Parent Household	.25	.43	0	1
Other Household Type	.23	.42	0	1
Parent Immigrant	.16	.37	0	1
Parent Education – HS Grad	.30	.46	0	1
Parent Education – Some College	.30	.46	0	1
Parent Education – College	.24	.43	0	1
Parent Professional/Managerial Occ.	.34	.47	0	1
Parent Disabled	.05	.21	0	1
Family Welfare Receipt	.10	.30	0	1
Log Family Income	3.54	.85	0	6.91
Low Birth Weight	.11	.31	0	1
Mother's Age at Birth	25.48	5.34	12	53
Neighborhood Level Variables (N = 1,299)				
Neighborhood Disadvantage Scale	0	1	–3.14	3.29
Nhood Heterogeneity Ideal Relationship Script	0	1	–3.22	8.10
Nhood Reliability (weight)	.66	.33	.11	1
School Level Variables (N = 143)				
Urban	.31	.47	0	1
Rural	.13	.34	0	1
Small	.21	.41	0	1
Large	.31	.46	0	1
Cumulative Dropout Rate	7.74	11.18	0	68.52
Percent in College-Prep Program	41.44	33.43	0	100
Catholic School	.03	.18	0	1
Private School	.02	.14	0	1

Table S6. Control Variable Coefficients for Models in Table 4

Independent Variables	Model 1	Model 2
Individual/Family Level Variables		
Female	–.0618*	–.0614*
	(.0142)	(.0142)
Age	.0221*	.0216*
	(.0061)	(.0061)
Hispanic	.0713*	.0712*
	(.0316)	(.0314)
Black	.1147*	.1031*
	(.0231)	(.0252)
Native American	–.0001	–.0045
	(.0552)	(.0554)
Asian	.0123	.0107
	(.0666)	(.0651)
Other Race	–.0068	–.0099
	(.0371)	(.0371)
Multi Race	–.0252	–.0188
	(.0630)	(.0638)
Home Language Not English	–.0903	–.0904
	(.0498)	(.0500)
Immigrant	.0302	.0331
	(.0350)	(.0354)
Household Size	.0008	.0007
	(.0048)	(.0048)
Single Parent Household	.0189	.0200
	(.0293)	(.0294)
Other Household Type	.0321	.0331
	(.0226)	(.0226)
Parent Immigrant	–.0052	–.0040
	(.0335)	(.0334)
Parent Education – HS Grad	–.0738*	–.0720*
	(.0250)	(.0252)
Parent Education – Some College	–.0860*	–.0822*
	(.0258)	(.0259)
Parent Education – College	–.0882*	–.0836*
	(.0347)	(.0349)
Parent Professional/Managerial Occ.	–.0059	–.0040
	(.0178)	(.0178)
Parent Disabled	.0096	.0093
	(.0381)	(.0382)
Family Welfare Receipt	.0214	.0208
	(.0305)	(.0305)
Log Family Income	–.0100	–.0065
	(.0146)	(.0147)
Low Birth Weight	–.0045	–.0047
	(.0300)	(.0299)
Mother's Age at Birth	–.0010	–.0008
	(.0020)	(.0020)
School Level Variables		
Urban	.0001	–.00405
	(.0194)	(.0194)
Rural	–.0290	–.03262
	(.0204)	(.0202)

(continued on next page)

Table S6. *(continued)*

Independent Variables	Model 1	Model 2
Small	–.0245	–.02956
	(.0255)	(.0245)
Large	–.0298	–.02323
	(.0201)	(.0211)
Cumulative Dropout Rate	.0009	.000608
	(.0007)	(.0008)
Percent in College-Prep Program	–.0003	–.00024
	(.0004)	(.0004)
Catholic School	.0253	.029547
	(.0453)	(.0443)
Private School	–.0555	–.03681
	(.0418)	(.0396)

REFERENCES

Abbott, Andrew. 1990. "A Primer on Sequence Methods." *Organization Science* 1:375–92.

———. 1995. "Sequence Analysis: New Methods for Old Ideas." *Annual Review of Sociology* 21:93–113.

Abbott, Andrew and Alexandra Hrycak. 1990. "Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians' Careers." *American Journal of Sociology* 96:144–85.

Acock, Alan C. 2005. "Working With Missing Values." *Journal of Marriage and Family* 67:1012–28.

Allison, Paul D. 2002. *Missing Data.* Thousand Oaks, CA: Sage.

Blair, Julian and Michael G. Lacy. 2000. "Statistics of Ordinal Variation." *Sociological Methods and Research* 28: 251–79.

Little, Roderick J. A. and Donald B. Rubin. 2002. *Statistical Analysis With Missing Data*. 2d ed. New York: Wiley.

Rohwer, Goetz and Ulrich Poetter. 2005. *Transition Data Analysis* (version 6.4). Computer software. Available from: http://steinhaus.stat.ruhr-uni-bochum.de/tda.html.

Royston, Patrick. 2004. "Multiple Imputation of Missing Values." *The Stata Journal* 4:227–41.

Sankoff, David and Joseph Kruskal. 1983. *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison.* Reading, MA: Addison-Wesley.

Stovel, Katherine, Michael Savage, and Peter Bearman. 1996. "Ascription into Achievement: Models of Career Systems at Lloyds Bank, 1890–1970." *American Journal of Sociology* 102:358–99.

Reardon, Sean F. and Glenn Firebaugh. 2002. "Measures of Multigroup Segregation." *Sociological Methodology* 32:33–67.

White, Michael J. 1986. "Segregation and Diversity Measures in Population Distribution." *Population Index* 52: 198–221.