Rethinking the Cultural Context of Schooling Decisions in Disadvantaged Neighborhoods:

From Deviant Subculture to Cultural Heterogeneity

Sociology of Education

David J. Harding, University of Michigan

ONLINE APPENDICES

Appendix A: Descriptions of Individual, Family, and School Control Variables

(All measured at Wave 1 unless indicated otherwise)

Individual Controls:

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.

Age: A set of indicators for each age year. Age 15 is the omitted category.

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.

High School Graduation: Whether a respondent graduated from high school is used as a control in the models of college enrollment to ensure that the analysis captures neighborhood effects on processes related to post-secondary enrollment. Whether the respondent has ever graduated from high school is measured at wave three (when respondents were 18-25).

Family Controls:

Home Language not English: An indicator variable for those adolescents whose family does 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 step parent families). Married is the omitted category.

Except for education, 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 US.

Mother and Father Education: A set of indicator variables for each resident parent's completed level of education: less than High School, High School Graduate, Some College or Trade School, and College Graduate. An additional variable indicates when each parent is not present in the household. Less than high school is the omitted category for each set of parent education indicators.

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 Controls:

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 (400-1000), and Large (> 1000). 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 non-public schools.

Percent White: Indicators for quartiles of percent of the student body that is white. The highest quartile is the omitted category.

Note also that the models control for neighborhood disadvantage, which is likely correlated with school quality, and for school mean college goals and school heterogeneity of college goals. For students attending a middle or junior high school during wave 1, the characteristics of the high school into which their current school feeds are used. This is necessary to deal with the cross-classification in the Addhealth data discussed above, but has the potential to misclassify feeder school respondents who did not attend the corresponding high school. Of the respondents in my analytic sample who attended middle or junior high schools in Wave 1 and

did not attend that same school the following year (about 13% of the overall sample as recorded in wave 2), 65% attended the corresponding high school into which their school normally feeds, 30% attended a different school, and 5% did not attend school at all. However, compared to those who did attend their corresponding high school, those who did not were no more likely to enroll in college, net of the wave 1individual, family, and school control variables.

Neighborhood Controls:

In models that examine the relationship between neighborhood disadvantage and neighborhood cultural heterogeneity, it is necessary to control for other potential sources of neighborhood heterogeneity. Social organization theory points to three other potential structural sources of heterogeneity: immigration, residential instability, and ethnic and racial diversity. To control for immigration, I use the percent of neighborhood residents who are foreign born. To control for residential instability, I include percent of housing units that are owner occupied and percent of housing units that have housed the same household for the last five years. To control for ethnic diversity, I include percent Hispanic and percent Hispanic squared. Finally, to control for racial diversity, I use Simpson's Interaction Index (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):

$$\text{Racial Diversity Index} = 100\left(\frac{W}{1-W} + \frac{B}{1-B} + \frac{A}{1-A} + \frac{O}{1-O}\right)$$

This index varies from zero (entire neighborhood from one group) to 75 (neighborhood evenly divided between four groups).

4

Appendix B: Additional Tables

Table B1: Inter-item Correlations for Variables Included in the Neighborhood Disadvantage Scale (for census tracts that contain Addhealth in-home respondents at wave 1)

	Sign	Item-rest correlation	Average inter-item correlation	Alpha
Family Poverty Rate	+	0.7756	0.5759	0.8716
Female Headed Household Rate	+	0.6557	0.6179	0.8899
Percent College Grads	-	0.7811	0.5740	0.8708
Male Unemployment Rate	+	0.6679	0.6134	0.8881
Percent Managerial/Professional Occs	-	0.7673	0.5787	0.8729
Percent Families with income > $75K	-	0.6925	0.6047	0.8844
Scale			0.5941	0.8978

n = 2436 (all tracts represented among Addhealth Wave 1 In-home respondents)

Table B2: Descriptive Statistics for Variables in Table 2

	Mean	SD	Min	Max	% Imputed
College Goal Heterogeneity	0	1	-1.89	1.23	0
Nhood Disadvantage Scale	0	1	-3.20	5.59	0
% Hispanic	13.66	21.96	0.00	96.27	0
% Hispanic Squared	668.37	1739.95	0.00	9267.46	0
Racial Diversity Index	26.82	20.73	0.00	74.36	0
% Foreign Born	11.34	16.16	0.00	86.90	0
% Owner Occupied	59.97	22.78	0.00	97.84	0.2%
% Units Occupied 5 Years	0.48	0.14	0.10	1.00	0.2%
Intergenerational Closure	0	1	-3.77	3.93	2.6%
Social Cohesion	0	1	-3.09	4.15	0
Disorder	0	1	-4.24	2.27	3.0%
Reliability (weight)	0.22	0.20	0.04	0.91	0

n = 1403 (only tracts included in analytic sample)

Table B3: Descriptive Statistics for Variables in Table 3

	Mean	*SD*	*Min*	*Max*	*% Imputed*
Individual/Family Level Variables (n=13,943)					
College Enrollment	0.48	0.5	0	1	0
College Goals	4.45	1.02	1	5	0
High School Graduate	0.83	0.37	0	1	0.1%
Female	0.53	0.5	0	1	0
Age	15.63	1.72	11	21	0
Hispanic	0.16	0.37	0	1	0
Black	0.22	0.41	0	1	0
Native American	0.03	0.18	0	1	0
Asian	0.08	0.27	0	1	0
Other Race	0.09	0.28	0	1	0
Multi Race	0.05	0.21	0	1	0
Home Language Not English	0.1	0.31	0	1	0
Immigrant	0.08	0.27	0	1	0
Household Size	4.62	1.62	1	18	0
Single Parent Household	0.23	0.42	0	1	0
Other Household Type	0.21	0.4	0	1	0
Parent Immigrant	0.19	0.39	0	1	0.7%
Mother Education – HS Grad	0.29	0.46	0	1	2.0%
Mother Education – Some College	0.27	0.45	0	1	2.0%
Mother Education – College	0.24	0.43	0	1	2.0%
Mother Education – No Mother Present	0.02	0.13	0	1	2.0%
Father Education – HS Grad	0.24	0.43	0	1	6.1%
Father Education – Some College	0.2	0.4	0	1	6.1%
Father Education – College	0.22	0.42	0	1	6.1%
Father Education – No Father Present	0.18	0.39	0	1	6.1%
Parent Professional/Managerial Occ	0.34	0.47	0	1	1.9%
Parent Disabled	0.05	0.21	0	1	1.7%
Family Welfare Receipt	0.1	0.29	0	1	2.4%
Log Family Income	3.54	0.84	0	6.91	24.2%
Low Birth Weight	0.11	0.31	0	1	17.3%
Mother's Age at Birth	25.78	5.36	5	53	26.4%
Neighborhood Level Variables (n=1,337)					
Neighborhood Disadvantage Scale	-0.04	0.8	-2.58	4.4	0
College Goals Nhood Heterogeneity	0.61	0.32	0	1	0
College Goals Nhood Mean	4.43	0.59	1	5	0
Intergenerational Closure Scale	0	0.2	-0.73	0.76	0
Social Cohesion Scale	0.02	0.35	-1.05	1.46	2.6%
Disorder Scale	0.04	0.96	-4.12	2.24	3.0%
Nhood Reliability (weight)	0.23	0.21	0.04	0.91	0
School Community Level Variables (n=89)					
Urban	0.31	0.47	0	1	0
Rural	0.16	0.37	0	1	0
Small	0.16	0.37	0	1	0
Large	0.47	0.5	0	1	0
Cumulative Dropout Rate	11.29	12.13	0	68.52	0
Percent in College Prep Program	58.53	27.29	0	100	3%
Catholic School	0.03	0.18	0	1	0
Private School	0.03	0.18	0	1	0
Percent White – Quartile 1	0.15	0.36	0	1	9%
Percent White – Quartile 2	0.28	0.45	0	1	9%

Percent White – Quartile 3	0.37	0.49	0	1	9%
College Goals School Heterogeneity	0.55	0.18	0	1	0
College Goals School Mean	4.38	0.36	2.5	5	0
School Reliability (weight)	0.82	0.2	0.09	0.99	0

Table B4: Control Variable Coefficients from Table 3

	Model 1	Model 2	Model 3	Model 4
Individual/Family Level Variables				
Female	0.3948*	0.3960*	0.3971*	0.3964*
	(0.0558)	(0.0562)	(0.0562)	(0.0558)
Age 11	-30.9603*	-31.0577*	-30.9566*	-30.3654*
	(0.3375)	(0.5645)	(0.3898)	(0.5520)
Age 12	0.3676*	0.3670*	0.3763*	0.3763*
	(0.1675)	(0.1672)	(0.1684)	(0.1688)
Age 13	0.3007*	0.3045*	0.3116*	0.3126*
	(0.0861)	(0.0864)	(0.0865)	(0.0868)
Age 14	-0.0459	-0.0405	-0.0376	-0.0386
	(0.0707)	(0.0714)	(0.0713)	(0.0713)
Age 16	0.0167	0.0182	0.0154	0.0170
	(0.0674)	(0.0678)	(0.0679)	(0.0681)
Age 17	-0.1134	-0.1115	-0.1109	-0.1096
	(0.0568)	(0.0566)	(0.0566)	(0.0569)
Age 18	-0.1289	-0.1288	-0.1288	-0.1264
	(0.0726)	(0.0724)	(0.0727)	(0.0727)
Age 19	-0.8222*	-0.8211*	-0.8144*	-0.8143*
	(0.1847)	(0.1855)	(0.1844)	(0.1841)
Age 20	0.9955*	1.0137*	1.0165*	1.0234*
	(0.3823)	(0.3782)	(0.3809)	(0.3818)
Age 21	0.9514*	0.9839*	0.9765*	0.9891*
	(0.8371)	(0.8654)	(0.8746)	(0.8813)
Hispanic	0.2016*	0.2095*	0.1954*	0.1978*
	(0.0863)	(0.0860)	(0.0844)	(0.0844)
Black	0.0541	0.0972	0.1001	0.1105
	(0.0798)	(0.0817)	(0.0802)	(0.0786)
Native American	-0.3138	-0.2977	-0.3067	-0.3011
	(0.1944)	(0.1916)	(0.1955)	(0.1950)
Asian	0.0047	0.0188	0.0127	0.0199
	(0.1557)	(0.1544)	(0.1484)	(0.1510)
Other Race	-0.1827*	-0.1705*	-0.1672*	-0.1583*
	(0.0793)	(0.0780)	(0.0767)	(0.0758)
Multi Race	-0.0347	-0.0577	-0.0554	-0.0613
	(0.1541)	(0.1529)	(0.1545)	(0.1548)
Home Language Not English	0.2682*	0.2762*	0.2648*	0.2652*
	(0.0857)	(0.0868)	(0.0852)	(0.0875)
Immigrant	0.0251	0.0294	0.0226	0.0223
	(0.0968)	(0.0957)	(0.0963)	(0.0951)
Household Size	-0.0339*	-0.0341*	-0.0339*	-0.0337*
	(0.0142)	(0.0141)	(0.0141)	(0.0141)
Single Parent Household	-0.2477*	-0.2421*	-0.2465*	-0.2434*
	(0.1246)	(0.1244)	(0.1243)	(0.1241)
Other Household Type	-0.4813*	-0.4790*	-0.4810*	-0.4796*
	(0.0523)	(0.0522)	(0.0521)	(0.0521)
Parent Immigrant	0.4653*	0.4609*	0.4533*	0.4545*
	(0.0848)	(0.0845)	(0.0845)	(0.0842)
Mother Education – HS Grad	0.2755*	0.2701*	0.2681*	0.2668*
	(0.1207)	(0.1206)	(0.1197)	(0.1193)
Mother Education – Some College	0.3925*	0.3851*	0.3815*	0.3804*
	(0.1046)	(0.1040)	(0.1029)	(0.1034)
Mother Education – College	0.7741*	0.7640*	0.7607*	0.7573*
	(0.1421)	(0.1419)	(0.1403)	(0.1408)

8

Mother Education – No Mother Present	0.0474	0.0371	0.0395	0.0343
	(0.2498)	(0.2503)	(0.2492)	(0.2503)
Father Education – HS Grad	0.1475	0.1441	0.1491	0.1486
	(0.0896)	(0.0900)	(0.0904)	(0.0917)
Father Education – Some College	0.4086*	0.4002*	0.4009*	0.3978*
	(0.0821)	(0.0806)	(0.0808)	(0.0830)
Father Education – College	0.7579*	0.7460*	0.7404*	0.7347*
	(0.1025)	(0.1013)	(0.1024)	(0.1035)
Father Education – No Father Present	0.2452	0.2355	0.2368	0.2349
	(0.1618)	(0.1616)	(0.1622)	(0.1639)
Parent Professional/Managerial Occ	0.1843*	0.1821*	0.1827*	0.1821*
	(0.0494)	(0.0496)	(0.0495)	(0.0496)
Parent Disabled	-0.1655	-0.1640	-0.1637	-0.1591
	(0.1077)	(0.1069)	(0.1067)	(0.1065)
Family Welfare Receipt	-0.2894*	-0.2828*	-0.2875*	-0.2843*
	(0.0895)	(0.0892)	(0.0890)	(0.0895)
Log Family Income	0.2159*	0.2113*	0.2072*	0.2060*
	(0.0470)	(0.0469)	(0.0469)	(0.0470)
Low Birth Weight	-0.0740	-0.0744	-0.0726	-0.0720
	(0.0679)	(0.0679)	(0.0685)	(0.0680)
Mother's Age at Birth	0.0261*	0.0260*	0.0258*	0.0258*
	(0.0053)	(0.0052)	(0.0052)	(0.0052)
School Community Level Variables				
Urban	-0.1631	-0.1252	-0.0456	-0.0243
	(0.1073)	(0.1073)	(0.0981)	(0.0968)
Rural	-0.0914	-0.0542	0.0136	-0.0073
	(0.1052)	(0.1024)	(0.1141)	(0.1151)
Small	0.1738	0.1757	0.1241	0.1582
	(0.0922)	(0.0917)	(0.0930)	(0.0869)
Large	0.1552	0.1065	0.0624	0.0732
	(0.0926)	(0.0953)	(0.0919)	(0.0968)
Cumulative Dropout Rate	-0.0080	-0.0070	-0.0030	-0.0025
	(0.0037)	(0.0038)	(0.0031)	(0.0032)
Percent in College Prep Program	0.0077*	0.0075*	0.0046*	0.0046*
	(0.0020)	(0.0021)	(0.0019)	(0.0019)
Catholic School	-0.1929	-0.2503	-0.3483	-0.2916
	(0.1283)	(0.1386)	(0.1775)	(0.1709)
Private School	0.9298*	0.8505*	0.4665*	0.4886*
	(0.2071)	(0.2195)	(0.2591)	(0.2670)
Percent White Quartile 1	0.0601	0.1074	-0.0050	-0.0359
	(0.1336)	(0.1368)	(0.1711)	(0.1677)
Percent White Quartile 2	0.1324	0.1675	0.0735	0.0677
	(0.1554)	(0.1507)	(0.1349)	(0.1384)
Percent White Quartile 2	0.0584	0.0583	0.0093	-0.0184
	(0.1122)	(0.1067)	(0.1016)	(0.1080)

Appendix C:

Scales for Neighborhood Intergenerational Closure, Social Cohesion, and Disorder

The Intergenerational Closure scale measures the degree to which neighborhood parents act collectively to monitor and communicate with one another about children, based on reports from Addhealth parent interviews. The intergenerational closure scale is based on three measures from the Addhealth parent interview, all of which are converted to a five-point scale: (1) if the respondent saw a neighbor's child getting into trouble, would she tell the child's parent, (2) if a neighbor saw the respondent's child getting into trouble, would the neighbor tell the respondent, and (3) the number of parents of the adolescent's friends the parent has talked to in the last four weeks. The Social Cohesion scale measures the degree to which neighborhood residents know one another and look out for one another, based on reports from Addhealth adolescent respondents. The social cohesion scale is based on three true/false measures from the student interview: (1) "You know most people in the neighborhood," (2) "In the past month, you have stopped on the street to talk with someone who lives in your neighborhood," and (3) "People in this neighborhood look out for each other." The neighborhood Disorder scale measures the degree to which residents are able to control undesirable public behavior in the community. The disorder scale is based on two measures in the parent questionnaire: (1) "In this neighborhood, how big a problem is litter or trash on the streets and sidewalks?" and (2) "In this neighborhood, how big a problem are drug dealers and drug users?" Response categories include "no problem at all," "a small problem," and "a big problem."

Construction of the neighborhood intergenerational closure, social cohesion, and disorder scales is based on methods described in Raudenbush and Sampson (1999). These methods

provide a way to aggregate survey data collected from individual respondents to the neighborhood level. Each scale combines data from multiple indicators of the concept. There are three five-category ordinal measures for the intergenerational closure scale, three binary indicators for the social cohesion scale, and two three-category indicators for the disorder scale (these indicators are described in the main text). The scales are constructed using all wave 1 Addhealth cases, not just those respondents used in this analysis (which is limited to those followed through wave 3).

For each scale, the multiple indicators can be thought of as hierarchically nested in a three level model: items nested within individuals nested within neighborhoods. I modeled these items using three-level hierarchical models. The dependent variable in each model is the value of the particular indicator. The level 1 model includes a constant and dummy variables for each item (excluding one to allow for an omitted category). The level two model includes a constant with variance component and age and gender indicators. These age and gender variables remove age and gender variation from the scale. The level three model includes only a constant with a variance component. The estimated models used to construct the scales are displayed in Tables C1-C3.

After estimating this model, the predicted value of the constant for the neighborhood is the measure of the scale in the logit metric (known as empirical Bayes estimates). These values are the sum of the constant and the neighborhood-specific random effect. The variables are then standardized for easier interpretation. Coefficients on the item indicators can be interpreted as item "severity" relative to the omitted category. The more negative a coefficient, the rarer the indicator. The age and gender indicators capture differences by age and gender in the indicators, and they allow the resulting scales to be independent of differences across neighborhoods in the

age and gender of sampled individuals. An additional advantage of this framework is that individuals with missing data on some items do not need to be excluded from the model as long as they have data on at least one item.

Reliability of the neighborhood scale based on binary indicators can be calculated for neighborhood k as (Raudenbush and Sampson 1999, Equation 10):

$$\frac{\omega_{pp}}{\omega_{pp} + \dfrac{\tau_{pp}}{J_k} + \dfrac{1}{n_k J_k w_k}}$$

Where ω_{pp} is the between neighborhood variance from the model, τ_{pp} is the within neighborhood variance from the model, J_k is the number of individuals in neighborhood k, n_k is the mean number of items per individual in neighborhood k, and w_k is the average (over individuals in neighborhood k) of the product of the proportion of positive items (those coded one) times proportion negative items (those coded zero). Thus the primary determinants of neighborhood level reliability are the proportion of variance between neighborhoods and the number of individuals per neighborhood (number of items per individual varies little across neighborhoods). Reliability will also be higher when w_k is near its maximum, which occurs when the proportion of items that are positive is one half. When more of the variance is between neighborhoods, reliability is higher for all neighborhoods, and neighborhoods with more individuals have higher reliability. Reliability of the neighborhood social cohesion scale has a mean of 0.29 and a standard deviation of 0.23. Reliabilities are uncorrelated with neighborhood disadvantage and are largely a function of the small numbers of respondents in some neighborhoods. Though these reliabilities are somewhat lower than those of neighborhood scales in prior research (e.g. Sampson et al. 1999), models weight by the reliability of the neighborhood mean college goals

scale, essentially privileging neighborhoods with more Addhealth respondents, for which

measurement reliability is higher,.

Table C1: Multi-level Ordinal Logit Model Used in Construction of Neighborhood Intergenerational Closure Scale

Term	Coefficient (Standard Error)
Threshold 1	-2.773 (0.038)
Threshold 2	1.064 (0.015)
Threshold 3	2.243 (0.020)
Threshold 4	2.667 (0.025)
Item 1 (tell neighbor)	omitted
Item 2 (neighbor tell you)	0.731 (0.019)
Item 3 (friends parents spoken to)	2.410 (0.048)
Age 11	0.335 (0.484)
Age 12	-0.054 (0.067)
Age 13	-0.172 (0.040)
Age 14	-0.071 (0.037)
Age 15	omitted
Age 16	0.121 (0.037)
Age 17	0.202 (0.034)
Age 18	0.218 (0.037)
Age 19	0.433 (0.090)
Age 20	0.465 (0.200)
Age 21	0.277 (0.297)
Male	-0.056 (0.021)
Variance Components:	
Neighborhood	0.122
Individual	0.533
N items	52,108
N individuals	17,752
N neighborhoods	2,261

Table C2: Multi-level Logit Model Used in Construction of Neighborhood Social Cohesion Scale

Term	Coefficient (Standard Error)
Constant	0.964 (0.038)
Item 1 (know most neighbors)	omitted
Item 2 (stopped to talk to someone on street)	0.484 (0.035)
Item 3 (neighbors look out for each other)	0.049 (0.038)
Age 11	-0.085 (0.504)
Age 12	0.097 (0.091)
Age 13	0.127 (0.052)
Age 14	0.105 (0.042)
Age 15	omitted
Age 16	-0.111 (0.042)
Age 17	-0.274 (0.041)
Age 18	-0.308 (0.044)
Age 19	-0.368 (0.097)
Age 20	-0.613 (0.226)
Age 21	-0.109 (0.433)
Male	0.236 (0.026)
Variance Components:	
Neighborhood	0.313
Individual	1.085
N items	60,947
N individuals	20,432
N neighborhoods	2,432

Table C3: Multi-level Ordinal Logit Model Used in Construction of Neighborhood Disorder Scale

Term	Coefficient (Standard Error)
Threshold 1	0.0271 (0.039)
Threshold 2	2.810 (0.039)
Item 1 (litter and trash)	omitted
Item 2 (drug users and dealing)	0.122 (0.037)
Parent Respondent Age 18-34	-0.163 (0.056)
Parent Respondent Age 35-44	omitted
Parent Respondent Age 45-54	0.149 (0.039)
Parent Respondent Age 55-64	0.301 (0.102)
Parent Respondent Age 65-89	0.414 (0.168)
Parent Respondent Age Missing	0.671 (0.292)
Parent Respondent Male	0.059 (0.065)
Parent Respondent Gender Missing	-0.096 (0.097)
Variance Components:	
Neighborhood	1.494
Individual	1.119
N items	34,413
N individuals	17,419
N neighborhoods	2,244

Appendix D: Additional Methodological Information

Another potential option for measuring peer college goals is the friendship network data in Addhealth, which allows researchers to link responses of up to five friends who are also Addhealth respondents to the respondent. However, many nominated friends are not Addhealth respondents (because they do not attend one of the Addhealth sampled schools in the community), and therefore their college goals would be missing from any peer group measure constructed from these data. For example, if we examine the same sex friendship nominations from the in school data and attempt to use the college expectations survey question (variable s45e) from the in school questionnaire, only about 42 percent of respondents can be successfully linked to their first nominated friend's response to this question. About 18 percent of respondents did not nominate a friend at all, eight percent nominated a friend who did not respond to the question, and 30 percent nominated a friend who could not be linked to an Addhealth respondent record. Across all same sex friend nominations, only about 30 percent of in-school respondents can be linked to all their nominated friend's responses to the college expectations question. See also Harding (2008), which shows that Addhealth in-home respondents also make many friend nominations that cannot be linked to other Addhealth respondent records, and that this problem is more common among youth from disadvantaged neighborhoods.

Neighborhood Disadvantage Scale

This scale measures the economic and social characteristics of the families that make up the neighborhood and which are thought to lead to negative outcomes for youth. Three of these variables (poverty, single mother households, and male unemployment) indicate the presence of disadvantaged families. The remaining three (college graduates, managerial and professional

17

workers, affluent families) indicate the absence of middle class families since their polarity is reversed. While some researchers (e.g. Brooks-Gunn et al 1993) have argued that the absence of middle class families is more important than the presence of disadvantaged families, there are high inter-item correlations across all six variables in these data. This suggests that these two sets of measures capture the same underlying neighborhood SES concept but simply focus on the presence of families at opposite ends of the SES distribution as indicators of a neighborhood's position in that distribution. A correlation matrix of these variables is available from the author upon request.

Appendix Table B1 shows the inter-item correlations of the variables that make up the scale. The item-rest correlation is the correlation between each item and the scale constructed without that item. Note that correlations are high for all three of the measures of presence of middle class families (college graduates, managerial/professional occupations, and family income above $75K), indicating that these variables are strongly correlated with the scale even when they are not included in it. The average inter-tem correlation is the average correlation of the items in a scale constructed without that single item. Note that there is not much variation in this column across items, indicating that no one item is pulling down the inter-item correlation in the overall scale. Finally, Cronbach's alpha is a measure of the reliability of the scale. The alpha column shows this value for the scale constructed without each individual item (and at the bottom for the scale produced with all the items). Note that the reliability of the scale is little affected by the removal of any one item and that it is lowered slightly if any one of the items measuring the presence of middle class families is removed.