

Appendix B.

Let's Stick Together: Peer Effects in Secondary School Choice and Variations by Student Socio-economic Background.

Dieuwke Zwier^{1*}, Sara Geven¹, Thijs Bol¹, Herman G. van de Werfhorst¹

19-05-2022

¹ University of Amsterdam, Department of Sociology, Amsterdam, The Netherlands

Table B.1. Conditional logit models, secondary school choice, full (starting) sample.

	% actual peers		Dev. actual-predicted peers	
	(1a)	(2a)	(1b)	(2b)
Main effects				
% actual peers	0.826*** (0.007)	0.939*** (0.012)		
Deviation actual-predicted peers			0.306*** (0.002)	0.337*** (0.005)
Distance (km)	-0.329*** (0.005)	-0.308*** (0.005)	-0.530*** (0.005)	-0.530*** (0.005)
Denomination match (<i>ref.</i> = <i>public</i>)				
No match	-0.099*** (0.010)	-0.098*** (0.010)	-0.261*** (0.015)	-0.261*** (0.015)
Match	-0.003 (0.013)	-0.001 (0.013)	0.300*** (0.018)	0.300*** (0.018)
School quality (index)	-0.037*** (0.006)	-0.037*** (0.006)	-0.046*** (0.006)	-0.046*** (0.006)
% high-SES	0.005 (0.012)	-0.012 (0.012)	0.152*** (0.012)	0.152*** (0.012)
% non-western	0.106*** (0.014)	0.094*** (0.014)	0.168*** (0.017)	0.167*** (0.017)
Interactions parental education (<i>ref.</i> = <i>low</i>)				
Medium * % actual peers		-0.120*** (0.010)		
High * % actual peers		-0.131*** (0.012)		
Medium * Deviation actual-predicted peers				-0.029*** (0.005)
High * Deviation actual-predicted peers				-0.038*** (0.006)
Medium * Distance (km)	0.078*** (0.004)	0.056*** (0.005)	0.056*** (0.004)	0.056*** (0.004)
High * Distance (km)	0.123*** (0.005)	0.098*** (0.005)	0.085*** (0.006)	0.085*** (0.006)
Medium * School quality (index)	0.015* (0.006)	0.015* (0.006)	0.016* (0.006)	0.016* (0.006)
High * School quality (index)	0.072*** (0.007)	0.072*** (0.007)	0.062*** (0.007)	0.062*** (0.007)
Medium * % high-SES	0.065*** (0.011)	0.083*** (0.011)	0.099*** (0.010)	0.099*** (0.010)
High * % high-SES	0.615*** (0.013)	0.632*** (0.013)	0.544*** (0.014)	0.543*** (0.014)
Medium * % non-western	-0.180*** (0.011)	-0.168*** (0.011)	-0.217*** (0.012)	-0.216*** (0.012)
High * % non-western	-0.037* (0.015)	-0.028 (0.015)	-0.204*** (0.018)	-0.203*** (0.018)
Loglikelihood	-785396.1	-785118.4	-903506.5	-903446.8

Source: Netherlands Cohort Study on Education (NCO).

Notes: $N = 4,949,635$, $n = 623,402$. All continuous variables are mean-centered and standardized ($SD = 1$), except for distance (in km). Cluster-robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.2. Conditional logit models, secondary school choice, sample selection check parental education.

	% actual peers		Dev. actual-predicted peers	
	(1a)	(2a)	(1b)	(2b)
Main effects				
% actual peers	0.854*** (0.009)	0.883*** (0.011)		
Deviation actual-predicted peers			0.315*** (0.005)	0.326*** (0.006)
Interactions parental education (ref. = low/medium)				
High * % actual peers		-0.063*** (0.013)		
High * Deviation actual-predicted peers				-0.025** (0.009)
Controls	Yes	Yes	Yes	Yes
Loglikelihood	-95672.59	-95652.16	-110953.6	-110946.7

Source: Netherlands Cohort Study on Education (NCO).

Notes: $N = 612,988$, $n = 77,247$. The models control for distance, distance * parental education, denomination match, school quality, school quality * parental education, % high-SES students, % high-SES students * parental education, % non-western students, % non-western students * parental education. All continuous variables are mean-centered and standardized ($SD = 1$), except for distance (in km). Cluster-robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.3. Conditional logit models, secondary school choice, student choice sets based on 10 km distance radius.

	% actual peers		Dev. actual-predicted peers	
	(1a)	(2a)	(1b)	(2b)
Main effects				
% actual peers	0.687*** (0.008)	0.798*** (0.020)		
Deviation actual-predicted peers			0.261*** (0.004)	0.295*** (0.011)
Interactions parental education (ref. = low)				
Medium * % actual peers		-0.120*** (0.021)		
High * % actual peers		-0.128*** (0.021)		
Medium * Deviation actual-predicted peers				-0.036** (0.013)
High * Deviation actual-predicted peers				-0.041** (0.012)
Controls	Yes	Yes	Yes	Yes
Loglikelihood	-90885.95	-90845.95	-102566.9	-102556.5

Source: Netherlands Cohort Study on Education (NCO).

Notes: $N = 814,120$, $n = 65,318$. The models control for distance, distance * parental education, denomination match, school quality, school quality * parental education, % high-SES students, % high-SES students * parental education, % non-western students, % non-western students * parental education. All continuous variables are mean-centered and standardized ($SD = 1$), except for distance (in km). Cluster-robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.4. Conditional logit models, secondary school choice, including students who pick a school outside distance radius.

	% actual peers		Dev. actual-predicted peers	
	(1a)	(2a)	(1b)	(2b)
Main effects				
% actual peers	0.888*** (0.009)	0.989*** (0.023)		
Deviation actual-predicted peers			0.323*** (0.005)	0.358*** (0.013)
Interactions parental education (ref. = low)				
Medium * % actual peers		-0.119*** (0.024)		
High * % actual peers		-0.110*** (0.023)		
Medium * Deviation actual-predicted peers				-0.039** (0.015)
High * Deviation actual-predicted peers				-0.042** (0.015)
Controls	Yes	Yes	Yes	Yes
Loglikelihood	-94956.64	-94931.09	-111841.9	-111833.8

Source: Netherlands Cohort Study on Education (NCO).

Notes: $N = 574,497$, $n = 68,814$. The models control for distance, distance * parental education, denomination match, school quality, school quality * parental education, % high-SES students, % high-SES students * parental education, % non-western students, % non-western students * parental education. All continuous variables are mean-centered and standardized ($SD = 1$), except for distance (in km). Cluster-robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.5. Results conditional logit models including secondary school FE.

	% actual peers		Dev. actual-predicted peers	
	(1a)	(2a)	(1b)	(2b)
Main effects				
% actual peers	0.875*** (0.009)	0.961*** (0.022)		
Deviation actual-predicted peers			0.293*** (0.005)	0.328*** (0.013)
Distance (km)	-0.358*** (0.011)	-0.342*** (0.011)	-0.579*** (0.010)	-0.580*** (0.010)
School quality (index)	-0.014 (0.017)	-0.014 (0.018)	-0.007 (0.017)	-0.006 (0.017)
% high-SES	-0.174*** (0.040)	-0.187*** (0.041)	-0.213*** (0.038)	-0.213*** (0.039)
% non-western	0.098* (0.042)	0.095* (0.042)	0.097* (0.040)	0.097* (0.041)
Interactions parental education (ref. = low)				
Medium * % actual peers		-0.087*** (0.023)		
High * % actual peers		-0.102*** (0.022)		
Medium * Deviation actual-predicted peers				-0.034* (0.015)
High * Deviation actual-predicted peers				-0.044** (0.014)
Medium * Distance (km)	0.080*** (0.012)	0.063*** (0.012)	0.062*** (0.011)	0.063*** (0.011)
High * Distance (km)	0.133*** (0.012)	0.112*** (0.012)	0.099*** (0.011)	0.100*** (0.011)
Medium * School quality (index)	-0.022 (0.019)	-0.022 (0.019)	-0.010 (0.019)	-0.010 (0.019)
High * School quality (index)	0.018 (0.019)	0.017 (0.019)	0.022 (0.018)	0.022 (0.018)
Medium * % high-SES	0.005 (0.032)	0.019 (0.033)	0.072* (0.030)	0.071* (0.030)
High * % high-SES	0.469*** (0.032)	0.482*** (0.033)	0.603*** (0.030)	0.601*** (0.030)
Medium * % non-western	-0.236*** (0.030)	-0.229*** (0.030)	-0.254*** (0.029)	-0.254*** (0.029)
High * % non-western	-0.256*** (0.033)	-0.252*** (0.034)	-0.313*** (0.033)	-0.313*** (0.033)
Secondary school FE	Yes	Yes	Yes	Yes
Loglikelihood	-73991.71	-73974.69	-84458.91	-84451.08

Source: Netherlands Cohort Study on Education (NCO).

Notes: $N = 496,635$, $n = 62,455$. Schools that are only observed in one year are dropped. All continuous variables are mean-centered and standardized ($SD = 1$), except for distance (in km). These models do not control for denomination match given the lack of within-school variation over time. Cluster-robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Mixed logit models

The CLM makes the Independence of Irrelevant Alternatives (IIA) assumption. This means that outcome categories should be distinct, and adding or excluding alternatives should not affect the relative odds of choosing for certain outcomes. To relax this assumption, we compared our CLM estimates with those from a mixed logit model (MLM).

The MLM was similar to the CLM. However, instead of estimating point estimates, the MLM assumed coefficients to vary over the population, and to be drawn from a normal distribution. This made the MLM suitable for estimating heterogeneity in secondary school preferences across students. Conversely, downsides were that mixed logit models lack a closed form solution, resulting in computational limitations (simulated probabilities), and that one has to assume the appropriate mixing distribution. More formally stated, mixed logit probabilities were the integrals of “standard” logit probabilities L_{ist} ($= P_{ist}$ in Equation 2) at different values of β with density function $f(\beta)$ as mixing distribution (Train, 2009). Hence, the MLM choice probability was

$$P_{ist} = \int L_{ist}(\beta) f(\beta) d\beta = \int \left(\frac{e^{\mathbf{W}_{it}\beta + \mathbf{X}_{st}\beta + \mathbf{Z}_{ist}\beta}}{\sum_{k=1}^n e^{\mathbf{W}_{it}\beta + \mathbf{X}_{kt}\beta + \mathbf{Z}_{ikt}\beta}} \right) \phi(\beta | b, W) d\beta \quad (3)$$

where the density $\phi(\beta | b, W)$ was specified to be normal with mean b and covariance W . Results for the MLMs are presented in Table B.6.

Table B.6. Mixed logit models, secondary school choice.

	% actual peers		Dev. actual-predicted peers	
	(1a)	(2a)	(1b)	(2b)
Main effects				
% actual peers	1.077*** (0.014)	1.209*** (0.026)		
Deviation actual-predicted peers			0.660*** (0.013)	0.730*** (0.028)
Distance (km)	-0.336*** (0.011)	-0.315*** (0.011)	-0.557*** (0.012)	-0.558*** (0.012)
Denomination match (<i>ref.</i> = <i>public</i>)				
No match	-0.079*** (0.018)	-0.077*** (0.018)	-0.237*** (0.023)	-0.237*** (0.023)
Match	-0.006 (0.021)	-0.004 (0.021)	0.310*** (0.026)	0.310*** (0.026)
School quality (index)	-0.034* (0.016)	-0.034* (0.016)	-0.038* (0.017)	-0.038* (0.017)
% high-SES	0.009 (0.027)	-0.004 (0.028)	0.155*** (0.029)	0.156*** (0.030)
% non-western	0.083** (0.026)	0.073** (0.026)	0.130*** (0.030)	0.131*** (0.030)
Interactions parental education (<i>ref.</i> = <i>low</i>)				
Medium * % actual peers		-0.120*** (0.025)		
High * % actual peers		-0.167*** (0.024)		
Medium * Deviation actual-predicted peers				-0.049 (0.032)
High * Deviation actual-predicted peers				-0.100*** (0.030)
Medium * Distance (km)	0.074*** (0.012)	0.055*** (0.013)	0.056*** (0.013)	0.056*** (0.013)
High * Distance (km)	0.125*** (0.012)	0.098*** (0.013)	0.099*** (0.013)	0.100*** (0.013)
Medium * School quality (index)	-0.014 (0.019)	-0.015 (0.020)	-0.010 (0.021)	-0.010 (0.021)
High * School quality (index)	0.078*** (0.018)	0.076*** (0.018)	0.069*** (0.020)	0.069*** (0.020)
Medium * % high-SES	0.025 (0.031)	0.036 (0.032)	0.065 (0.034)	0.063 (0.034)
High * % high-SES	0.621*** (0.031)	0.634*** (0.032)	0.592*** (0.033)	0.590*** (0.033)
Medium * % non-western	-0.192*** (0.029)	-0.183*** (0.029)	-0.229*** (0.032)	-0.230*** (0.032)
High * % non-western	-0.008 (0.031)	-0.001 (0.031)	-0.145*** (0.035)	-0.146*** (0.035)
SD(% actual peers)	0.630*** (0.015)	0.631*** (0.015)		
SD(% deviation actual-predicted peers)			1.544*** (0.031)	1.543*** (0.031)
Loglikelihood	-78108.08	-78076.69	-86002.53	-85995.28

Source: Netherlands Cohort Study on Education (NCO).

Notes: $N = 500,044$, $n = 62,620$. All continuous variables are mean-centered and standardized ($SD = 1$), except for distance (in km). Cluster-robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.7. Conditional logit models, secondary school choice, similar-track peer groups.

	% actual peers		Dev. actual-predicted peers	
	(1a)	(2a)	(1b)	(2b)
Main effects				
% actual similar-track peers	1.011*** (0.008)	1.080*** (0.020)		
Dev. similar-track peers			0.482*** (0.006)	0.515*** (0.015)
Interactions parental education (ref. = low)				
Medium * % actual similar-track peers		-0.082*** (0.022)		
High * % actual similar-track peers		-0.079*** (0.022)		
Medium * Dev. similar-track peers				-0.036* (0.017)
High * Dev. similar-track peers				-0.040* (0.017)
Controls				
	Yes	Yes	Yes	Yes
Loglikelihood	-68695.6	-68683.68	-85489.46	-85483.33

Source: Netherlands Cohort Study on Education (NCO).

Notes: $N = 498607$, $n = 62468$. The models control for distance, distance * parental education, denomination match, school quality, school quality * parental education, % high-SES students, % high-SES students * parental education, % non-western students, % non-western students * parental education. All continuous variables are mean-centered and standardized ($SD = 1$), except for distance (in km). Cluster-robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.8. Conditional logit models, secondary school choice, subsamples by gender.

	% actual peers				Dev. actual-predicted peers			
	Boys		Girls		Boys		Girls	
	(1a)	(2a)	(1a)	(2a)	(1b)	(2b)	(1b)	(2b)
Main effects								
% actual peers	0.820*** (0.011)	0.944*** (0.032)	0.834*** (0.012)	0.960*** (0.032)				
Dev. actual-pred. peers					0.321*** (0.007)	0.377*** (0.019)	0.303*** (0.007)	0.335*** (0.018)
Interactions parental education								
<i>(ref. = low)</i>								
Medium * % peers		-0.118*** (0.035)		-0.147*** (0.035)				
High * % peers		-0.153*** (0.033)		-0.139*** (0.033)				
Medium * Dev. peers						-0.070** (0.022)		-0.016 (0.021)
High * Dev. peers						-0.062** (0.021)		-0.049* (0.020)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Loglikelihood	-39771.89	-39752.52	-39274.23	-39257.18	-45450.21	-45441.27	-45292.72	-45285.52
Students	31453	31453	31453	31453	31453	31453	31453	31453
Observations	250881	250881	250881	250881	250881	250881	250881	250881

Source: Netherlands Cohort Study on Education (NCO).

Notes: The models control for distance, distance * parental education, denomination match, school quality, school quality * parental education, % high-SES students, % high-SES students * parental education, % non-western students, % non-western students * parental education. All continuous variables are mean-centered and standardized ($SD = 1$), except for distance (in km). Cluster-robust standard errors in parentheses. Separate models including (tree-way) interactions with gender (not displayed) indicate that gender differences in the effects of the peer variables, and their interactions with parental education, do not reach statistical significance.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table B.9. Conditional logit models, secondary school choice, same-gender peer groups.

	% actual peers		Dev. actual-predicted peers	
	(1a)	(2a)	(1b)	(2b)
Main effects				
% actual same-gender peers	0.871*** (0.009)	0.993*** (0.023)		
Dev. same-gender peers			0.386*** (0.005)	0.443*** (0.015)
Interactions parental education (ref. = low)				
Medium * % actual same-gender peers		-0.119*** (0.025)		
High * % actual same-gender peers		-0.149*** (0.024)		
Medium * Dev. same-gender peers				-0.053** (0.017)
High * Dev. same-gender peers				-0.073*** (0.016)
Controls	Yes	Yes	Yes	Yes
Loglikelihood	-76336.89	-76300.25	-88755.23	-88735.47

Source: Netherlands Cohort Study on Education (NCO).

Notes: $N = 500,044$, $n = 62,620$. The models control for distance, distance * parental education, denomination match, school quality, school quality * parental education, % high-SES students, % high-SES students * parental education, % non-western students, % non-western students * parental education. All continuous variables are mean-centered and standardized ($SD = 1$), except for distance (in km). Cluster-robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.