sociological science

Supplement to:

Sotoudeh, Ramina, Ginevra Floridi. 2025. “Pathways to Independence: The Dynamics of Parental Support in the Transition to Adulthood” Sociological Science 12: 833-861.

Online Supplementary Information

Part A: Sequence Analysis

Selecting the best cluster solution

To select the best clustering solution for the sequence analysis results, we evaluated three metrics common in the broader clustering literature for identifying the best solution: point biserial correlation (PBC), average silhouette width (ASW), and the Calinski-Harabasz index (CH). The results from these evaluations can be found in Table A-1. Across all three metrics, the three-class solution performs most optimally.

Table A-1: Evaluating cluster fit

Num. of Classes	PBC	ASW	CH
2	0.25	0.13	162.9
3	0.4	0.15	168.33
4	0.34	0.11	156.32
5	0.38	0.11	145.05
6	0.39	0.11	129.74
7	0.39	0.11	117.92
8	0.39	0.1	109.26
9	0.37	0.09	102

Cluster Imputation Strategy

Because respondents in the TAS portion of the PSID occasionally miss a survey wave, a sizeable portion (58%) of respondents had NAs in their sequence at some point. Traditionally in sequence analysis, researchers will treat NA as its own state and then perform the analysis. This often introduces noise into the clustering process that renders the results unmeaningful. We found far more success–in terms of quantitatively and qualitatively more coherent clusters–when we instead ran clustering only on complete sequences. The downside with this approach is that it leads to a significant reduction in sample size. Our solution was to infer clusters for people with incomplete sequences by examining their average similarity with members of each cluster and then assigning to them the cluster to which their average distance was smallest. Doing so allowed us to balance two objectives of retaining sample size while creating coherent and meaningful clusters.

Part B: Classifying Early-Childhood Behavioral Dispositions

To identify behavioral dispositions, we classified respondents using questions from the Childhood Development Supplement (CDS) of the PSID about difficult behaviors and characteristics as reported by the parents when the child was between 10 and 12 years old. Primary caregivers reported to the following question for each of the different child characteristics: "For the next set of statements, decide whether they are not true, sometimes true, or often true of CHILD's behavior." All characteristics were changed to reflect negative attributes.

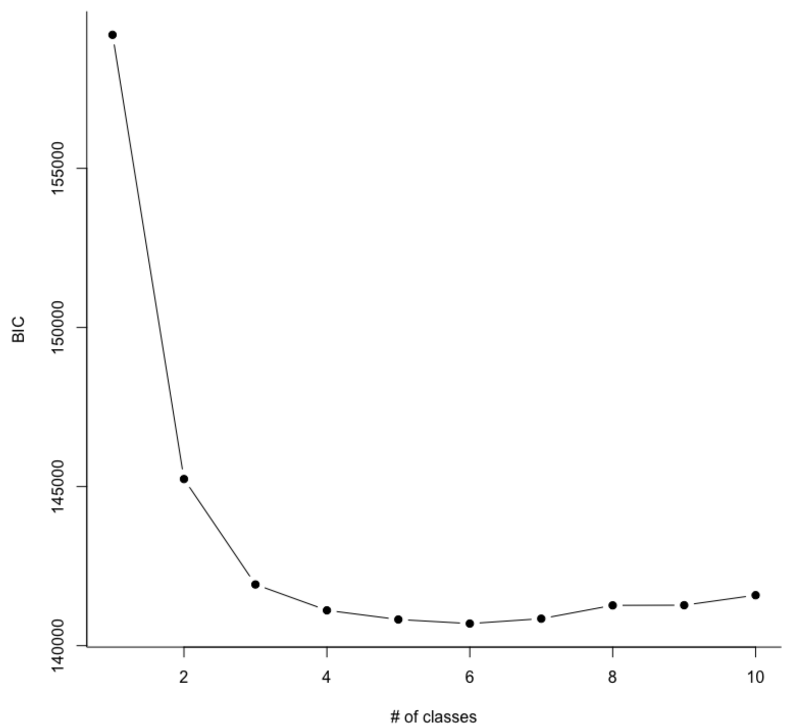
Fearful or anxious	Doesn't get over being upset	Strong tempered	Worries too much
Argues too much	Doesn't solve difficult tasks	Unhappy	Trouble with teachers
Cheats or lies	Disorderly	Withdrawn	Doesn't wait their turn
High strung	Doesn't finish things	Destructive	Does careless work
Feels unloved	Trouble getting along	Cries too much	Not curious
Mood swings	Impulsive	Demands attention	Clings to adults
Difficulty concentrating	Feels worthless	Dependent	
Easily confused	Not liked by other kids	Paranoid	
Disobedient	Has obsessions	Hangs around trouble	
Feels no regret	Stubborn	Secretive	

To classify respondents, we used Latent Class Analysis (LCA). LCA is a statistical method used to identify unobserved subgroups or classes within a population using multivariate categorical data. To do so, it uses a type of finite mixture model to explain the relationships among observed variables by assuming the existence of an underlying latent categorical variable (Sinha et al. 2021). Within each class, observed variables are assumed to be locally independent, such that the latent classes explain the association among the observed variables.

Because LCA is an inductive method, it is useful for uncovering heterogenous groups in categorical data. Researchers, however, must tell LCA how many latent classes it should look for in the data; and therefore, much like the sequence analysis detailed in Appendix A, researchers often use goodness-of-fit statistics to inform their choice between multiple, competing models. The most common goodness-of-fit statistic for this purpose is Bayesian Information Criterion (BIC). BIC is derived from the log likelihood statistic but crucially adjusts it by including a penalty term for the number of classes in its formulation to reward parsimony. In practice, as the number of classes grow, the penalty term offsets gains in fit from adding additional terms (e.g., latent classes) to the model. Researchers often plot BIC against the number of terms and look for an “elbow” in the plot, e.g., when the returns to BIC from adding terms level off and become negligible.

We followed this protocol to select the number of classes for our LCA model. The results are visualized in Figure B-1. The elbow is at three classes, which is the number we use for the model in the main paper.

Figure B-1: Returns to BIC from adding classes to LCA model



Part C: Descriptive Statistics

Table C-1: Average Characteristics by Cluster and Tests of Differences from Other Clusters

Variable	In School & Supported	Residentially Supported Workers	Married & Independent
LCA category: “Diligents”	**0.529*** **	*0.353****	**0.5**
LCA category: “Middles”	*0.391**	**0.462** **	*0.385*
LCA category: “Strugglers”	*0.08****	**0.185*** **	*0.115*
Age	*25.967***	**26.132 * **	**26.095**
Family income: Quintile 1	*0.125****	**0.346*** **	*0.157**
Family income: Quintile 2	*0.136****	**0.228*** **	*0.179*
Family income: Quintile 3	*0.181**	**0.222**	**0.234**
Family income: Quintile 4	**0.238*** **	*0.137****	**0.241**
Family income: Quintile 5	**0.321*** **	*0.067****	*0.19*
Female	**0.561** **	*0.469****	**0.588* **
First Born	**0.439* **	*0.368***	**0.431**
Year joined TAS	**2008.538* **	*2008.341*	*2008.052**
Respondent feels close to parent	*3.261**	**3.335**	**3.332**
Metro	**0.868*** **	*0.78**	*0.684****
Mother’s years of education	**13.645*** **	*11.803****	*12.621*
Number of children born to mother	*2.602****	**3.094*** **	*2.785*
Experienced Pandemic Year	**0.198*** **	*0.127****	*0.144*
Parents currently married	**0.642*** **	*0.501****	*0.543*
Asian	**0.029*** **	*0.004****	*0.01*
Black	*0.274****	**0.516*** **	*0.181****
Hispanic	*0.139*	**0.154**	*0.105*
Other Race	**0.027**	*0.02*	*0.007*
White	**0.532*** **	*0.306****	**0.697*** **
Experienced Recession Year	*0.464*	**0.487**	**0.536**
Standardized and imputed GPA	**0.826*** **	*0.798 ****	**0.824**
Family wealth: Quintile 1	*0.136****	**0.262*** **	*0.139**
Family wealth: Quintile 2	*0.125****	**0.278*** **	*0.153*
Family wealth: Quintile 3	*0.15***	**0.218*** **	*0.168*
Family wealth: Quintile 4	**0.219**	*0.17***	**0.266** **
Family wealth: Quintile 5	**0.37*** **	*0.072****	**0.274**

*Note: Bolded values indicate that the cluster average or proportion is higher than the other two. Red and italicized values indicate that it is lower. Proportions are reported for categorical variables and means are reported for numerical variables. Asterisks are based on t-tests that test whether the difference between the cluster average for a given variable is significantly different than the members of the other two clusters. Significance Levels: * p<0.05, ** p<0.01, *** p<0.001*

Part D: Multinomial Logistic Regressions

We estimate two sets of multinomial logistic regression models to examine the predictors of sequence cluster membership. In both models, cluster 1 ("school, single, and financially supported") serves as the reference category. As in the main models, the between-family model includes demographic characteristics (gender, race), family structure (maternal education, parents' marital status, number of children born to mother), socioeconomic indicators (family income quintiles, wealth quintiles), and whether they live in a metropolitan area (metro). The within-family model includes individual characteristics (gender, standardized GPA, health status), birth order, closeness to family (helped siblings, closeness to parents), behavioral profiles (from the latent class analysis), and sibling group fixed effects to account for unobserved family characteristics. Both models include year controls. The coefficients are presented as log odds ratios, indicating the relative likelihood of membership in each cluster compared to the reference cluster.

Table D-1: Multinomial Logistic Regression: between family models (OR with 95% CI)

	$\frac{P(\textbf{Coresident worker})}{P(\textbf{In school} \,\&\, \textbf{supported})}$	$\frac{P(\textbf{Married} \,\&\, \textbf{Ind.})}{P(\textbf{In school} \,\&\, \textbf{supported})}$	$\frac{P(\textbf{Married} \,\&\, \textbf{Ind.})}{P(\textbf{Coresident worker})}$
Female	0.588***	1.014	1.724***
	(0.475, 0.729)	(0.764, 1.346)	(1.29, 2.305)
Asian	0.129**	0.206*	1.601
	(0.03, 0.548)	(0.043, 0.986)	(0.249, 10.275)
Black	1.154	0.284***	0.246***
	(0.881, 1.511)	(0.193, 0.419)	(0.168, 0.361)
Hispanic	0.767	0.36***	0.469**
	(0.521, 1.128)	(0.21, 0.616)	(0.273, 0.806)
Other Race	0.756	0.158	0.209
	(0.277, 2.061)	(0.02, 1.271)	(0.026, 1.694)
Mother’s Education	0.914***	0.925**	1.012
	(0.876, 0.954)	(0.875, 0.977)	(0.96, 1.066)
Parents Currently Married	0.797*	0.842	1.057
	(0.636, 0.999)	(0.621, 1.142)	(0.777, 1.437)
Number of siblings	1.229***	1.162*	0.945
	(1.121, 1.348)	(1.023, 1.319)	(0.835, 1.07)
Family Income Quintile 2	0.804	1.057	1.314
	(0.585, 1.106)	(0.65, 1.719)	(0.834, 2.072)
	0.551***	0.928	1.686*
Family Income Quintile 3	(0.392, 0.773)	(0.565, 1.527)	(1.044, 2.722)
	0.473***	0.697	1.473
Family Income Quintile 4	(0.323, 0.694)	(0.404, 1.201)	(0.858, 2.527)
	0.153***	0.697	4.565***
Family Income Quintile 5	(0.09, 0.258)	(0.383, 1.268)	(2.316, 8.998)

Family Wealth Quintile 2	1.083 (0.779, 1.505)	1.288 (0.787, 2.108)	1.19 (0.746, 1.897)
Family Wealth Quintile 3	0.726 (0.521, 1.012)	1.03 (0.638, 1.662)	1.418 (0.887, 2.267)
Family Wealth Quintile 4	0.577** (0.402, 0.83)	0.723 (0.437, 1.196)	1.252 (0.753, 2.082)
Family Wealth Quintile 5	0.252*** (0.157, 0.405)	0.397** (0.224, 0.703)	1.572 (0.828, 2.983)
Metro	0.815 (0.637, 1.045)	0.56*** (0.415, 0.756)	0.687* (0.504, 0.936)
Year entered TAS	1.005 (0.946, 1.069)	0.948 (0.871, 1.031)	0.943 (0.865, 1.027)
Recession Year	1.349 (0.914, 1.993)	0.965 (0.564, 1.654)	0.715 (0.414, 1.236)
Number of Obs.	2040	2040	2040
AIC	3538.26	3538.26	3538.26
Log Likelihood	-1729.13	-1729.13	-1729.13

*Significance Levels: * p<0.05, ** p<0.01, *** p<0.001*

Table D-2: Multinomial Logistic Regression: Within-Sibling (OR with 95% CI)

	$\frac{P(\textbf{Coresident worker})}{P(\textbf{In school \& supported})}$	$\frac{P(\textbf{Married \& Ind.})}{P(\textbf{In school \& supported})}$	$\frac{P(\textbf{Married \& Ind.})}{P(\textbf{Coresident worker})}$
Female	0.457*** (0.325, 0.643)	0.775 (0.493, 1.218)	1.694* (1.075, 2.668)
Standardized GPA	0.799 (0.564, 1.132)	0.592* (0.376, 0.93)	0.74 (0.483, 1.135)
First Born	0.731 (0.514, 1.039)	1.507 (0.928, 2.45)	2.063** (1.283, 3.315)
Feels close to parents	2.622*** (1.884, 3.649)	1.732* (1.059, 2.833)	0.661 (0.405, 1.079)
Diligent (vs Strugglers)	0.202*** (0.106, 0.388)	3.155** (1.503, 6.62)	15.583*** (10.587, 22.937)
Middle (vs Strugglers)	0.384** (0.203, 0.726)	3.062** (1.457, 6.436)	7.981*** (5.375, 11.849)
Year entered TAS	0.503*** (0.503, 0.503)	0.318*** (0.318, 0.318)	0.633*** (0.633, 0.633)
Recession Year	0.135*** (0.064, 0.283)	0.039*** (0.014, 0.107)	0.29* (0.107, 0.787)
Number of Obs.	1177	1177	1177
Number of Sib. Groups	853	853	853
AIC	3752.52	3752.52	3752.52
Log Likelihood	-154.26	-154.26	-154.26

*Significance Levels: * <0.05, ** p<0.01, *** p<0.001*

Reference:

Sinha, P., Calfee, C. S., & Delucchi, K. L. (2021). Practitioner's guide to latent class analysis: methodological considerations and common pitfalls. *Critical care medicine*, *49*(1), e63-e79.