

# Issue Paper

Santa Clara University

Institutional Research

## Measuring Socio-Economic Status Among Santa Clara Undergraduates (SES-SCU)

### Issue

The Ignatian model of social justice refers to the idea of a just society which gives individuals and groups fair treatment and a just share of the benefits of society. Operationally, social justice requires analytical tools to identify, investigate, and respond to institutional and structural injustices found in economic, political, cultural, or ecclesial systems. One such tool is a socio-economic status (SES) indicator. Under APA best practices is the recommendation that “an important determinant of the approach [used] to measure SES is the level at which [the effects are to be accessed]”. This research describes a composite indicator of SES specific to the Santa Clara undergraduate population derived from self-reported items collected by the HERI census survey administered to all first-time students at regular intervals between 1985 and 2017.

### SES Indicators

The standard indicators of SES, per the Department of Health and Human Services (HHS), include measures of level of education, income, occupation and industry, and family size and household composition. This same 2012 report (Carr, 2012) concludes that “the measurement of SES is a critical but very challenging area, and it is too early at this time to provide more detailed recommendations regarding the adoption of specific standards for the measurement of SES”. Nonetheless, the HHS does define education, income, and occupation as essential components.

The challenge in constructing an SES score is reflected among recent analyses linking SES to student achievement. The SES composite of Sackett et al (2009) is derived as an “equally weighted composite of the three variables” self-reported by SAT test-takers: father’s education, mother’s education, and [natural log of] family income. This derivation does not meet the HHS standard. Hurtado et al (2005) abandons any attempt to construct an SES score, instead relying on individual indicators: education, family income, white collar/blue collar occupations, even though this is an analysis of the same HERI data employed in this paper.

The SES score for the Santa Clara first-time student population presented here is a composite of these three components. SES-SCU is derived from and coincident with parent/guardian attributes (i.e. a student ‘inherits’ their SES from their parents). This paper defines general socioeconomic status (SES) as the overall standing of a person in the stratification system of a society, based on education, income, and occupational prestige.

# Issue Paper

**Santa Clara University**

**Institutional Research**

## **Data Collection**

First administered in 1967, the HERI freshman survey is regarded as the most comprehensive source of information on U.S. college students. Since 1985, the survey has posed items to collect self-reported data on the three key components of SES-SCU. The survey has evolved over time and as questions have been added, dropped, or modified, the three components are represented by more than 20 distinct items. Even questions that have remained constant (e.g. family income) have had changing response options over time which must be integrated to create a longitudinal series.

These survey data were coalesced into five items: occupation of parent one, occupation of parent two, education of parent one, education of parent two, and parent total income per year. A complete set was derived for every year of administration between 1985 and 2017. Responses to all five items are available from 85 percent of all respondents. Due to concerns of data not missing at random (NMAR), no imputation of missing data was performed. Since the SES-SCU composite is a linear combination of indicators which are not sample dependent, the constructed scores are unaffected by missing data.

Data coverage is affected by three factors: survey administration cycle, unit non-response and item non-response. Starting in 2010, Santa Clara adopted a biennial cycle for administration of the Freshman Survey, the survey was also not administered in 1986. The SCU-SES for these students is therefore unknown. During years of administration, item non-response constitutes less than one percent of missing data; instead students with unknown SCU-SES result from unit non response. This survey response rate averages 85% over time, ranging between 71 and 97 percent.

Though responses differ by sex, Santa Clara's experience with these data is that non-response bias is minimal and rarely benefits from remediation. In the case of SCU-SES, there is the potential that non-response is associated with SES status. To examine for this, a CHAID analysis of response-status and HEOA financial aid status was conducted. HEOA financial aid status is part of student records and independent of survey responses. It classifies students into mutually exclusive categories: Pell recipients, Title IV aid recipients (exclusive of Pell), and students not receiving Title IV aid. As such, HEOA financial aid status is highly correlated with SCU-SES. The analysis was conducted for all survey years between 1999 and 2017 (Santa Clara financial aid data is available only starting in 1999). Even with an aggressive p-value threshold of 0.10, of the 15 years under study, only two show any difference; in one (1999) Pell grant recipients responded at a higher rate than the other groups, and in the other (2005), the effect size is negligible. No significant non-response bias for SCU-SES is observed in this analysis.

# Issue Paper

**Santa Clara University**

**Institutional Research**

## **Data Transformation**

Composite indicators such as SES-SCU are a linear combination of suitably transformed individual indicators. It is useful therefore to transform all individual indicators to a common domain. Since all individual indicators in this measure of SES are ordinal, a bounded interval is required, and the domain of [0,...,100] is conventional.

Education items present eight ordinal response options from “less than junior high” to “graduate degree”. These values were mapped to the [0,...,100] interval via min-max normalization. Min-max normalization, unlike z-scores or percentile transformation, is sample-independent, i.e. the scale is derived solely from the construction of the items posed to respondents. This is a valuable property for longitudinal items that may have changed over time, and such scales can be transformed to be commensurate. Items “education of parent1” and “education of parent2” were further mapped to “maximum education of parents” and “minimum education of parents”, creating scores independent of parent sex (earlier versions of the question items posed “father’s education” and “mother’s education”).

Occupation items are comprised of 93 distinct, nominal occupations, though due to idiosyncratic variations in occupation titles over time, the complete set comprises 155 distinct values. These 93 occupations were mapped to occupational prestige scores developed by NORC for the 2010 Census Standard Occupational Classification codes (SOC) based on the 2012 General Social Survey (GSS) (Holt, 2015). GSS occupational prestige scores are standardized as a continuous value on [0,...,100]. These two occupation category schemas overlap but are not coincident, requiring case-by-case evaluation of a many-to-many mapping. Of particular challenge are HERI occupations mapping to multiple GSS occupations with significantly different prestige scores (e.g. HERI: Lawyer/Judge to GSS: Judges and GSS: Lawyers). The prestige score for a HERI occupation was estimated as a weighted average of multiple corresponding GSS scores. Items ‘occupation of parent1’ and ‘occupation of parent2’ were mapped to a maximum prestige score and a minimum prestige scores, again creating scores independent of parent sex (as before, earlier versions were posed as “father’s occupation” and “mother’s occupation”). The HERI response options are subset of GSS occupations, and yield a range of prestige scores compressed to [20,...,80]. To be commensurate with income and education, a min-max normalization maps these data to the [0,...,100] interval.

Income items pose the same question over time but with varying ordinal response options. Response options are presented as income bands (e.g. ‘\$60,000-\$74,999’, ‘more than \$100,000’). To devise a continuous measure of income, closed interval bands were mapped to the midpoint value, open intervals were mapped to the boundary value plus/minus an amount equal to the difference to the nearest midpoint. As would be expected, income bands increase over time, representing both nominal and real growth in family income. To create a longitudinally comparable income indicator, independent min-max normalizations were

# Issue Paper

Santa Clara University

Institutional Research

applied for each survey year, essentially representing family income as a percent of the maximum reportable income.

## Construction of SES-SCU Composite Score

As the *OECD Handbook on Constructing Composite Indicators* emphasizes, “the relative importance of the indicators is a source of contention”. This is an understatement, and this contention applies in equal measure to SES indicators. While there should be a theoretical framework to define the concept represented by the composite indicator, rarely does this framework provide guidance on the mathematical construction itself and the researcher is left with a variety of ad hoc techniques, including the geometric mean, principle component analysis (PCA), and factor analysis (FA). As the SES-SCU is essentially a rank, the geometric mean is inappropriate; PCA/FA are sample dependent and, in the case of quite distinct individual indicators will identify multiple “composite” scores, or select the individual score with the highest load. Instead, we require that SES-SCU be constructed with a transparent, sample-independent function. Indeed, in the large national Education Longitudinal Study conducted by the National Center for Education Statistics (NCES), SES is “based on five equally weighted, standardized components” (NCES 2004–405, p.H-5). We modify this approach in the construction of the SES-SCU.

With social assortative mating, we see a high correlation between parent one education and parent two education; this is also the case for parent one occupational prestige and parent two occupational prestige. Under a model that assigns equal weights, this correlation will expand scores at the high and low ends of the scale and compress those in the middle. This of course is also true among the three key components of education, income, and occupation, though the correlation among them is typically smaller than between parents’ education or parents’ occupation. To compensate for this effect, the two education indicators are replaced with their L-2 norm and normalized to [0,...,100] via min-max. The two occupational prestige indicators are likewise replaced by their L-2 norm. This results in three indicators, one each for family income, family occupational prestige, and family education. The SES-SCU is constructed as the mean of these equally weighted indicators.

Composite indicators can create an impression of greater precision than may be warranted, given the types of ad hoc techniques employed in their construction. In the case of individual indicators that increase over time (e.g. income), nominal growth can distort the SES composite; this effect is exacerbated by the use of unequal income bands that change over time. In practice then, it is convention to group the scores into percentiles, typically deciles, quintiles, or quartiles. This research adopts a quintile grouping, calculated independently for each survey year. This allows comparison by quintile over time. In this report and associated visualizations, the SES-SCU quintiles are labeled and ordered as [5: upper, 4: upper-middle, 3: middle, 2: lower-middle, 1: lower].

# Issue Paper

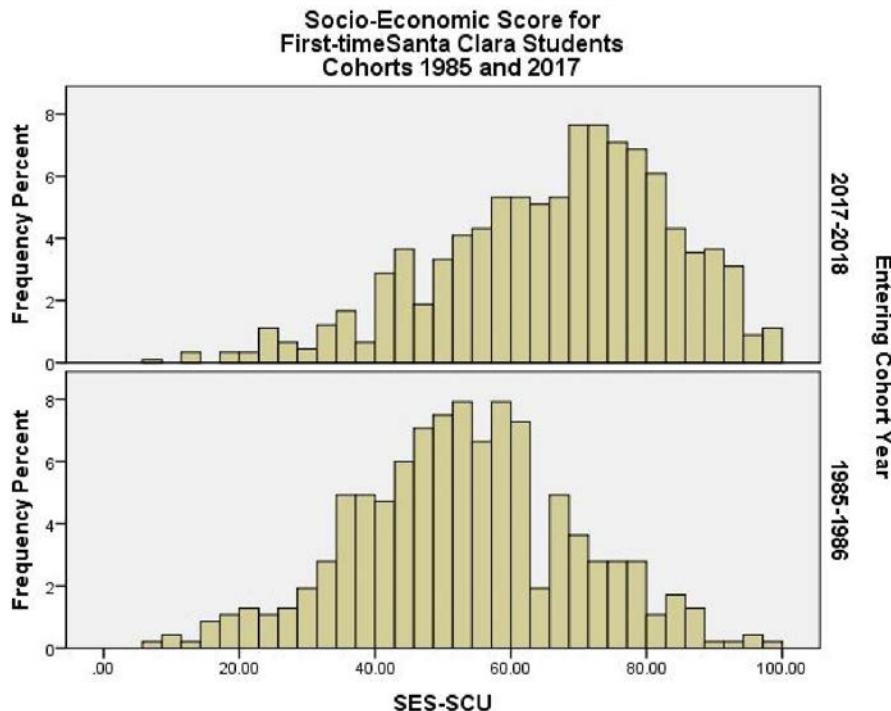
Santa Clara University

Institutional Research

Results: SCU-SES over time and by individual indicator

## Chart one

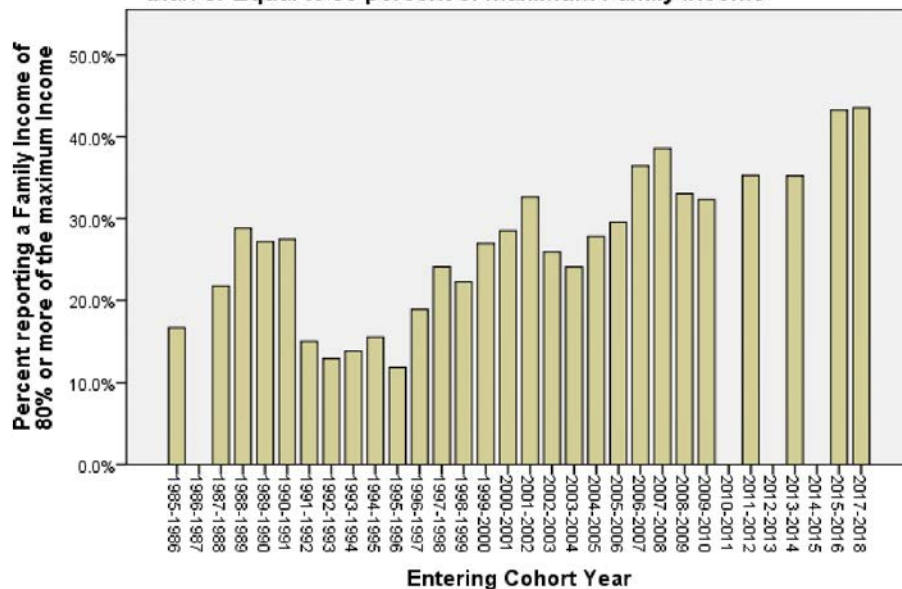
The distribution of SES scores changes over time. Additional research is needed to discern the role of exogenous changes (e.g. increasing income disparity, nominal income growth) from the role of institutional decision making, ranking, and prominence.



## Chart Two

Despite the downturn observed during the recession of the early 1990's, there is a positive trend in the proportion of students arriving from families earning 80 percent or more of the maximum income presented by the survey item. Though provocative, recasting these ordinal measures of income into constant dollars is required to examine for changes in growth of real income.

**Distribution of Percent of First-Time Students Reporting a family Income Greater than or Equal to 80 percent of maximum Family Income**



note: family income is derived from ordinal response options defined by unequal income bands. Of necessity, income bands are compressed in earlier years. However the maximum income is censored at a value commensurate with the range of income offered in each year.

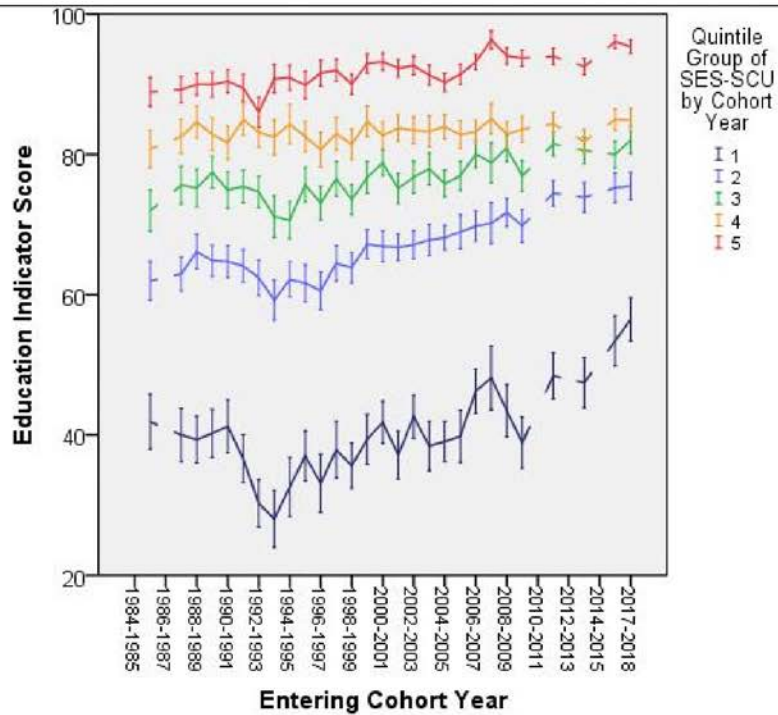
# Issue Paper

Santa Clara University

Institutional Research

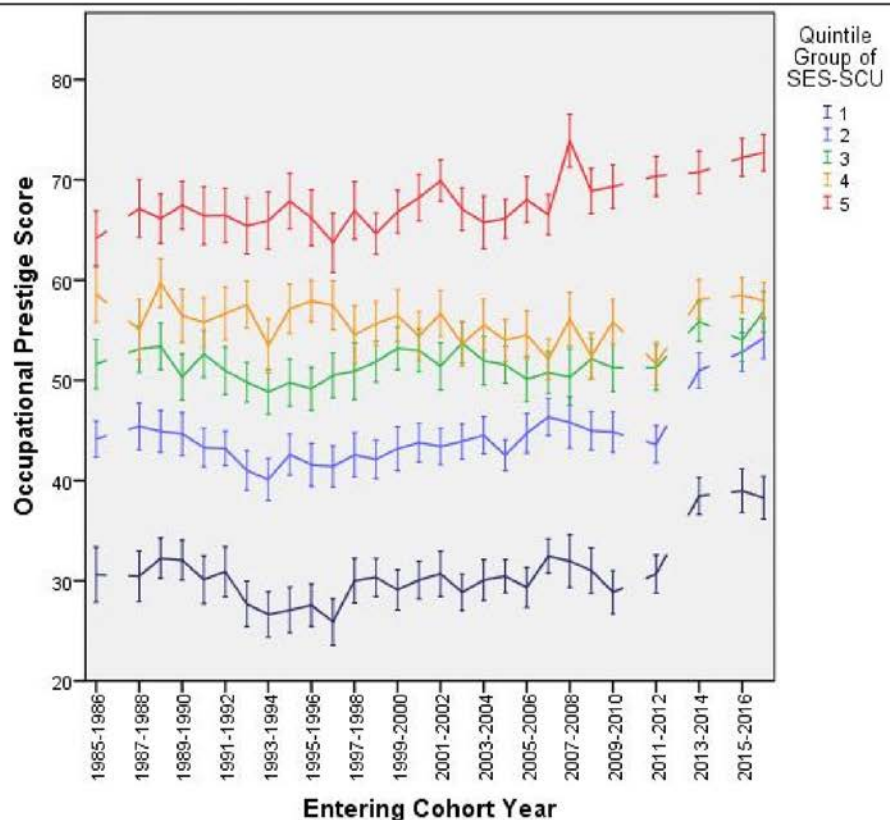
## Chart Three

This chart shows the level of education that defines each quintile. The gap between the fifth and fourth quintiles is persistent over a long period. There appears to be convergence among the middle three quintiles though this may be due in part to a ceiling effect (education is measured only to graduate degree). The level of education has been increasing among the lower three quintiles.



## Chart Four

Occupational prestige is clearly delineated for the top and bottom quintiles, while moderate convergence is observed among the three middle quintiles.



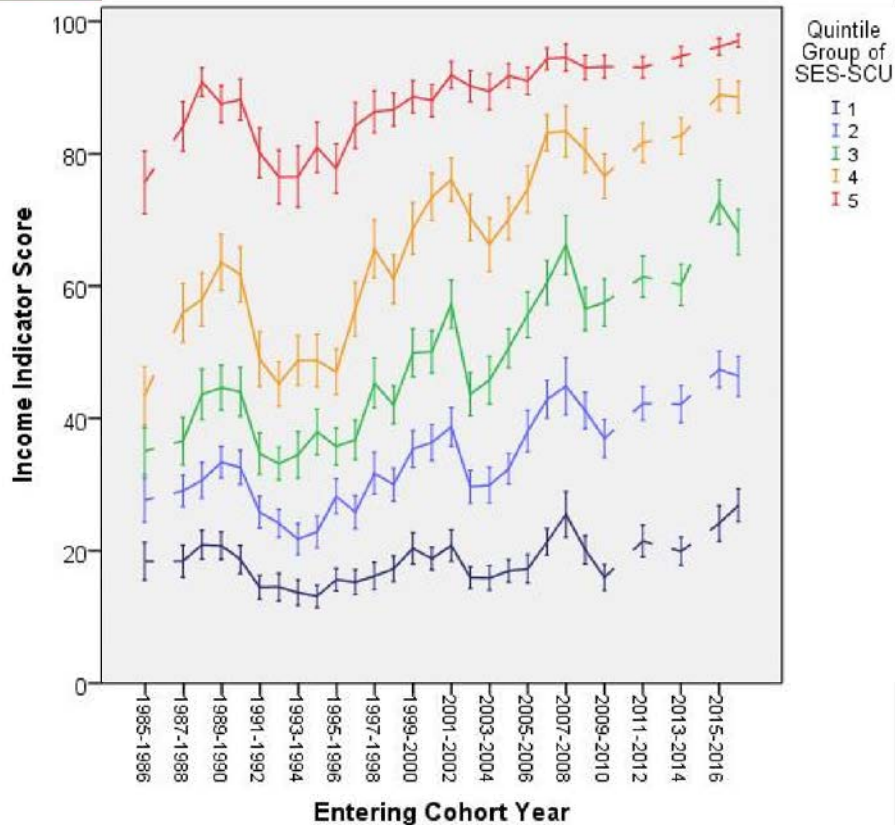
# Issue Paper

Santa Clara University

Institutional Research

## Chart five

As would be expected by construction, the income score of the quintiles are quite distinct. Even measured in nominal dollars, the recessions of the early 1990's and 2008 are visible. The gaps show little evidence of convergence. Indeed the gaps between adjacent quintiles may be increasing over time,-- however the effects of item-income compression must be addressed.



## Discussion

This is the initial construction of an SES indicator specific to Santa Clara first-time degree-seeking undergrads. Constructed from self-reported data collected just prior to matriculation, this measure conforms to the requirements of both the theoretical construct of socio-economic status and minimum data elements. The use of an SES composite score reflects the multi-dimensional aspect of social stratification and is an improvement over the more readily available, but crude (binary) proxy of Pell-grant recipient status. With students mapped to specific quintiles, the university can begin to explore the complementarity and interaction among socio-economic status and other demographic and social characteristics.

## Future Research on SES development

Among the initial considerations in devising a Santa Clara specific SES indicator was whether to source data from self-reported data or student record data (i.e. admission applications). Despite the substantial challenges and compromises inherent in using self-reported data, the deficits of student record data are not easily remedied: though parent occupation and education are reported on the Common App, they are not loaded in PeopleSoft--even if so, such data would extend only to 2007, and while accurate and detailed family income is available from the financial aid application process, nearly half of Santa Clara students do not apply for aid. If SES proves its value in understanding the college experience, academic performance, and degree attainment, and even perhaps advancing the social mobility and

# Issue Paper

development of cultural capital among Santa Clara undergraduates, then development of alternative (and improved) indicators of SES should be considered, to wit:

## Topics for research:

- i. Income is measured in nominal amounts; for longitudinal analysis it should be adjusted to constant dollars.
- ii. Is there evidence of “education inflation”, i.e. does it take higher levels of education to enter the same SES-SCU quintile over time?
- iii. Derive weights used in scale construction analytically (n.b. this is an area of active research with no consensus in the literature).
- iv. Create an analogous SES score with the university as the unit of analysis, supporting normative institutional comparison.
- v. Define a reference SES period and test for changes in relative SES composition over time: is SCU becoming more or less SES diverse?
- vi. Imputation of missing data post score construction, for increased coverage of first-time students
- vii. Are changes in the distribution of SES-SCU over time driven by exogenous factors (e.g. rising incomes and/or increasing income disparity) or by institutional factors (e.g. is Santa Clara’s pattern of enrolling students driving changes in the SES distribution)?
- viii. Re-access occupational prestige directly based on HERI occupational career categories using a best-worst scaling (BWS) model.

# Issue Paper

Santa Clara University

Institutional Research

## References

- American Psychological Association, "Measuring Socioeconomic Status and Subjective Social Status" Accessed May 4, 2017. <http://www.apa.org/pi/ses/resources/class/measuringstatus.aspx>.
- Carr, Justine M. *Development of Standards for the Collection of Socioeconomic Status in Health Surveys Conducted by the Department of Health and Human Services*, National Committee on Vital and Health Statistics, 22 June 2012, [www.ncvhs.hhs.gov/wpcontent/uploads/2014/05/120622lt.pdf](http://www.ncvhs.hhs.gov/wpcontent/uploads/2014/05/120622lt.pdf). Accessed 22 Aug. 2017
- Hout, Michael, Tom W. Smith, and Peter V. Marsden. "Prestige and Socioeconomic Scores for the 2010 Census Codes." *Methodological Report MR124*, Chicago, NORC. Gss. NORC. Org/Get-Documentation/Methodological-Reports, 2015. <http://gss.norc.org/Documents/reports/methodological-reports/MR124.pdf>.
- Hurtado, Sylvia, et al. *Black Undergraduates From Bakke to Grutter*. Los Angeles: Higher Education Research Institute, 2005. Print. ISBN: 1-878477-37-4
- OECD, European Union, and Joint Research Centre-European Commission. *Handbook on Constructing Composite Indicators: Methodology and User Guide*. OECD Publishing, 2008.
- Sackett, P. et al. *Socioeconomic Status and the Relationship Between the SAT® and Freshman GPA: An Analysis of Data from 41 Colleges and Universities*. New York: The College Board, 2009 Print Report No. 2009-1
- Walpole, MaryBeth. "Socioeconomic Status and College: How SES Affects College Experiences and Outcomes." *The Review of Higher Education* 27, no. 1 (September 11, 2003): 45–73. <https://doi.org/10/czrcm4>.
- U.S. Department of Education, National Center for Education Statistics. *Education Longitudinal Study of 2002: Base Year Data File User's Manual*, NCES 2004–405, by Steven J. Ingels, Daniel J. Pratt, James E. Rogers, Peter H. Siegel, and Ellen S. Stutts. Project Officer: Jeffrey A. Owings. Washington, DC: 2004.
- Zhou, Xueguang. "The Institutional Logic of Occupational Prestige Ranking: Reconceptualization and Reanalyses." *American Journal of Sociology* 111, no. 1 (July 2005): 90–140. <https://doi.org/10.1086/428687>.