

Seven Principles for Assessing Effectively Maintained Inequality

American Behavioral Scientist
2017, Vol. 61(1) 132–160
© 2016 SAGE Publications
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0002764216682990
journals.sagepub.com/home/abs



Samuel R. Lucas¹ and Delma Byrne²

Abstract

Effectively maintained inequality (EMI) was proposed as a general theory of inequality, but the theory flows from a decades-long tradition of studying social background effects on educational attainment. After an orienting discussion of several historic challenges of the study of social background effects on educational inequality, proposed and adopted solutions to those challenges, and subsequent critiques of those solutions, we offer and justify seven principles that, if followed, produce a solid assessment of EMI. After conveying the seventh principle, two illustrative ways in which EMI addresses historic challenges with studying inequality are conveyed.

Keywords

qualitative inequality, theoretically focal persons, distractive control variables, EMI bounds, salient standardization

Effectively maintained inequality (EMI; Lucas, 2001) was proposed as a general theory of inequality. Although EMI is resonant with theoretical resources from multiple traditions (Lucas, 2009), it most directly flows from a decades-long tradition of studies of social background effects on educational attainment. That history is composed of many path-breaking innovations and challenging realizations, reflecting a process in which each promising innovation drew critical attention that made visible *its* limitations, spurring further development. Understanding the historic challenges that have hounded efforts to study inequality is essential for efforts to assess EMI and perhaps other theories as well. Indeed, failure to grasp those historic challenges not only may

¹University of California–Berkeley, CA, USA

²The National University of Ireland, Maynooth, Co. Kildare, Ireland

Corresponding Author:

Samuel R. Lucas, Department of Sociology, University of California–Berkeley, 410 Barrows Hall #1980, Berkeley, CA 94720-1980, USA.

Email: Lucas@berkeley.edu

lead one to misunderstand EMI, it can, unfortunately, easily lead one to propose an “advance” that merely reintroduces a dated approach that was rejected years ago by scholars upon realizing the approach’s limitations.

Thus, we open our presentation of seven principles for assessing EMI with an orienting discussion of several challenges that have arisen in the effort to study inequality in education. We then offer and justify each principle. After conveying the seventh principle, we consider how EMI addresses historic challenges with studying inequality. Afterward, we provide concluding remarks.

Historic Methodological Challenges in the Study of Inequality in Education

Because education has long been hypothesized as a powerful, politically palatable means of extending equal opportunity, analysts interested in inequality have turned to the study of education. Much of the interest has been comparative. Analysts have asked whether educational prospects of children of different eras are equally connected to socioeconomic background and have wondered whether some nations are more successful than others at liberating students’ pursuit of knowledge from the accidents of birth. Such questions draw the interest of parents, politicians, policy makers, and the wider public. However, as studies of such comparative questions proceeded it became apparent that answers are often illusive owing to several technical challenges. Those challenges, coupled with statistical developments, technological advance, and social change, set the stage for the development of EMI.

Challenges

Original Challenges With Studying Highest Grade Completed. Post–World War II studies of socioeconomic background effects on education analyzed highest grade completed as an interval-level variable (e.g., B. Duncan, 1967).¹ The approach grew out of path analysis that originated with Wright (1934), was brought into sociology by O. D. Duncan (1966), B. Duncan (1967), Blau and Duncan (1967), and Sewell and colleagues (e.g., Sewell, Haller, & Ohlendorf, 1970; Sewell, Haller, & Portes, 1969), and extended into structural equation models that use unobserved (latent) variables and allow measurement error by Jöreskog (1973) and Hauser and colleagues (e.g., Hauser & Goldberger, 1971). In this approach, years of school completed is one of several interval-level variables that stand between parental status characteristics and children’s ultimate adult occupational and financial attainments (e.g., Hauser, Sewell, & Alwin, 1976; Hauser, Tsai, & Sewell, 1983). Despite the utility of this causal modeling method, and the sociological revelations the research produced, there were limitations.

Notable limitations included that the technology required interval-level variables, such as years of school completed, and that interaction effects were difficult to specify. Only in the late 1970s did structural equation models with categorical variables, such as college entry, become feasible (e.g., Muthén, 1979, 1983), and only in the mid-1980s

did the inclusion of interactions become possible (e.g., Kenny & Judd, 1984). Deeper challenges, however, are posed by the difficulty of comparing across nations or eras.

Substantively, the research found the relation between socioeconomic background and educational attainment was stable across cohorts observed in diverse social contexts of war, peace, poverty, and plenty. The findings sparked critical interest in the method, however, for schools had expanded throughout the period and differed in their level of expansion across countries (e.g., Meyer, Ramirez, Rubinson, & Boli-Bennett, 1977), creating two inferential problems. First, statistically, school expansion changed the variance of the dependent variable, thus problematizing comparison of unstandardized regression coefficients. Yet, standardizing variables mathematically easily creates its own distortions for comparative research (Bollen, 1989). Consequently, no consensus solution was evident.

Second, school expansion altered the meaning of long-standing education milestones (e.g., high school entry, high school graduation, and college entry; Trow, 1961), making changing relations between those milestones and socioeconomic background challenging to interpret. One response to the dilemma was to disaggregate the summary "years of school completed" variable into its component parts, the year-by-year continuation of education.

A Sequential Solution: Boudon's School Continuation Model

Boudon (1974) offered the first sustained analysis of school continuation. From a set of theoretical axioms, Boudon derived several empirical expectations. He concluded that the relationship between years of school completed and socioeconomic background in the United States (and in other societies as well) was nonlinear, such that linear regression models (and, by extension, structural equation models of the period) of highest grade completed could distort the relation.

It was possible that the distortions might underlie the seeming stability of socioeconomic background effects across multiple decades despite changing social conditions. The substantive claim of a nonlinear relation, and the possibility that linear models might distort the impact of social change, partly motivated Boudon's turn to alternative analytic methods that would, presumably, more effectively capture the background/educational attainment relation.

One main Boudon (1974) approach compared class-specific, level-specific attendance ratios. He found that attendance ratios of adjacent classes were not constant throughout the class distribution. Furthermore, as school attendance increased across cohorts, especially for higher levels of education, the cross-class differences in the attendance rates often declined even as the probability that a member of the advantaged class would obtain the level of education at issue increased disproportionately (e.g., Boudon, 1974).

An important criticism of the work, however, is that probabilities are bounded by 0 and 1, and ignoring this fact can distort findings (Hauser, 1976). The higher the baseline, time t probability of making a transition for high versus low socioeconomic status (SES) students, the more constrained high SES students are to make gains with school

expansion between time t to $t + x$. This complicates the consideration of effects of socioeconomic background using Boudon's method.

Mare (1981) showed how Boudon's approach implies a linear probability model that necessarily presumes no such ceiling effects. A troubling result of this assumption is that even if socioeconomic background effects are stable across cohorts, estimated effects can change across cohorts simply because as schools expand the continuation probabilities change.

Boudon's contribution provided a first sustained effort to address the challenges posed by studying years of school completed. While it ultimately failed, its insights suggested what would become a much more successful proposal for resolving the fundamental problems of comparison—the Mare school continuation logistic regression model.

The Mare Model

Like the Boudon model described above, the Mare model treats entry to and completion of each year of schooling as a decision point. Yet, the Mare model uses a more defensible functional form for the analysis. To discern social background effects, the researcher estimates one binary (often logistic) regression equation for each transition. On using this approach, Mare (1980) found that, contrary to claims of stability, socioeconomic background coefficients declined across transitions in the United States.

The Mare model allows one to make up to three possible comparisons: (a) within-cohort comparisons of coefficients across transitions (e.g., Are SES effects on high school graduation and college entry equal for a given cohort?); (b) within-nation comparisons of the same transitions' coefficients across cohorts (e.g., Are SES effects on college entry equal for the high school classes of 1981 and 2011?); and (c) cross-national comparisons of the same transitions' coefficients (e.g., Are SES effects on college entry equal in the United States and the Czech Republic?). By considering one or more of these comparisons, researchers may theorize the pattern of socioeconomic background effects.

Facilitating multiple comparisons, the Mare model allowed deeper theoretical and substantive investigation of the dynamics producing socioeconomic background effects. These gains, however, came at some cost. Despite the deceptive simplicity of the model, it poses several methodological challenges.

The Mare Model: Statistical Challenges and Responses

Over the decades, several problems the Mare model poses have become recognized. For example, De Graaf and Ganzeboom (1993) contend that the later the education transition, the less accurate the estimated coefficients are owing to selection bias induced by sample attrition across transitions. Intriguingly, while the Mare model addressed problems induced by Boudon's treatment of the marginal distribution, the De Graaf and Ganzeboom critique notes a similar problem in that unobserved factors that induce attrition at transition t are not factored into the analysis for transition $t + 1$.

One implication, therefore, is that the marginal distribution is changing owing to unobserved factors that are not appropriately included in analyses of later transitions. Thus, the education transitions model provides an advance on earlier models but still contains pathways through which an earlier challenge reappears. In response, Mare (1993) presented methods for investigating the possible impact of such *unobserved heterogeneity*-induced selection bias. As another example, to this challenge Cameron and Heckman (1998) add that estimated differences between the coefficients for any two transitions are not subject to statistical test owing to an identification failure unless one has time-varying covariates. Taking these critiques together, these observations imply that coefficient patterns are untestable artifacts.

Two lines of response developed from recognition of problems with the model. One approach maintains the framework of school continuation but modifies the model. Interest in continuing to study educational attainment as composed of a sequence of binary branching points exists in part because binary variables provide perhaps the greatest chance for cross-national, cross-cohort, and cross-life course comparative analyses. Thus, Tam (2011) offers a latent-class estimator that, while nonparametric, presumes equal socioeconomic background effects across transitions. Although Tam's (2011) approach assumes an answer to a substantively and theoretically intriguing question the education transitions model is otherwise able to answer, models exist to test the assumption (e.g., Hauser & Andrew, 2006) under some conditions. In contrast, Lucas, Fucella, and Berends (2001) and Holm and Jæger (2011) offer bivariate probit models with selection to account for selection bias in an earlier transition, thereby proposing a parametric response that may preserve analysts' ability to answer all three of the key questions of education transitions research but forces analysts to find theoretically appropriate exclusion restrictions to identify parametric selection models, a usually difficult task. Lucas, Fucella, and Berends (2011) identified eight methodological challenges with the classical education transitions model and proposed a neoclassical education transitions approach that addresses all methodological challenges. The neoclassical education transitions model encompasses both nonparametric and parametric modifications of the Mare model.

The second response to emergent challenges with the Mare model has been to treat at least some education transitions as involving more than two options—we refer to this as the polytomous outcomes framework. Three noteworthy factors underlie analyses that elaborate the positions within schooling levels: (a) statistical developments, (b) computer technology improvements, and (c) social change. Statistical developments have made it possible to address much more complex questions with models much more flexibly tuned to problems, allowing analysts to relax more and more assumptions deemed implausible. Computer technology improvements, including increases in data storage capacity (Walter, 2005), working memory, heat dissipation, and chip speed (Schaller, 1997), coupled with high-quality open source (e.g., R) and quasi-open source (e.g., Stata) software for statistical analysis have made many advanced statistical techniques feasible. All this would be of little value, however, in the absence of a need for different kinds of analyses. This need is manifest by social change in the form of increasing college attendance, possibly partly driven by the loss

of manufacturing (i.e., low-education/high-pay jobs), which has been accompanied by the elaboration of degrees, certificates, and multiple types of precollege education. Such educational and economic change makes it necessary and potentially revealing to study multiple categories of education within levels of schooling.

Thus, Breen and Jonsson (2000) propose a multinomial model to analyze polytomous outcomes; on using it, they find path dependence in Swedish students' educational attainment. Breen and Jonsson's model can apply whenever one has qualitatively distinct categories, whether the categories are ordered or not.

Lucas (2001) proposes a more parsimonious ordered probit model with censoring to investigate persons' navigation of the stratified curriculum. To use this model, outcomes must be ordered. Using this stratified outcomes model, Lucas (2001) investigated whether the pattern of results was consistent with effectively maintained inequality.

EMI in the Context of Historic Challenges to the Study of Inequality

Estimation, calibration, and comparison of effects constitute the foundational challenge of the effort to assess inequality. EMI addresses these challenges by highlighting consequential inequality, thus distancing its assessment from the standard rote calculation of coefficients' statistical significance. Assessing EMI requires careful reflection in the selection of the dependent variable, the modeling framework adopted, the population for study, the grounding of the comparison in theoretical relevance, the omission of distractive variables, the use of an extensive set of socioeconomic background indicators, and the reflective assessment of model results. Notably, a fully successful analysis of EMI will necessarily address the problems that have hounded the assessment of inequality in both the summary years of school completed and sequential accumulation of years of schooling approaches. To produce such an analysis, seven principles are essential.

Principle I: To Assess EMI's Qualitative Hypothesis, Select an Outcome With Meaningful Categorical Variation

EMI's Qualitative Hypothesis

Lucas (2001, 2009) conveyed the tenets of EMI; Lucas (2001) summarizes EMI's claims while elaborating their implications and connections to other theories. A distinctive aspect of EMI is a hypothesis for the pattern of allocation to important, qualitatively different outcome categories owing to socioeconomic background.

Lucas (2001) translated the theoretical claim concerning the qualitative dimension into an expectation for statistical analyses. Under EMI, statistical significance—that is, the difference between the statistical coefficient and zero—is not the focus. Instead, under EMI, we should predict different category outcomes for theoretically focal persons simply on the basis of socioeconomic background.

Because quantitative outcomes (e.g., years of education) vary smoothly over many ordered values, point-predicted values differ for those of disparate origins if the estimated regression coefficient is discernibly different from zero. However, qualitative outcomes are “lumpy,” such that it is possible for the socioeconomic background coefficient to be statistically significant yet still imply theoretically focal persons of disparate socioeconomic backgrounds would have the same modal categorical outcome.

Lucas (2009) demonstrated that EMI implies bounds on the socioeconomic background coefficient, for only some coefficients make the predicted outcome category for those of high and low socioeconomic background differ. Indeed, most positive coefficients (i.e., almost all positive coefficients) fail to produce an EMI pattern. Thus, it is possible for the association between the outcome and socioeconomic background to be positive, statistically significant, yet inconsistent with EMI (Lucas, 2009). Consequently, EMI is falsifiable even amid ubiquitous findings showing a positive association between socioeconomic background and education outcomes.

Selecting a Plausibly Relevant Outcome

Because of this atypical need to calculate predicted values to assess a more focused hypothesis, one must carefully select a plausibly relevant outcome variable for study. The univariate distribution of the dependent variable may make it more or less difficult and more or less sensible to study EMI’s qualitative hypothesis. Two aspects of the dependent variable matter: (a) the number of categories and (b) the distribution of cases across the categories.

Taking the distribution of cases across categories first, no matter how many categories there are, the higher the relative probability of the modal category compared with other categories, the more difficult it may be for the background-specific predicted values to display the diverging trajectories pattern EMI posits. Certainly, the proportions of persons of different socioeconomic statuses may have different chances of falling into an advantaged category. But differences in magnitude, while important, are not the focus of the qualitative question of EMI. This observation recognizes that EMI is a specific theory focused on specific patterns of inequality. One need not deny the existence of other possible inequality patterns, but one must recognize that EMI highlights particular ones and does not highlight others. Thus, EMI may not be an accurate characterization for every pattern of inequality.

One theoretical implication of the above is that EMI is a relevant hypothesis whether an outcome is universal or not (Lucas, 2009). Notably, an EMI pattern becomes *more* difficult to discern when a category of an outcome is highly dominant. Thus, a key factor in whether one can discern EMI on a qualitative outcome is the peakedness of the distribution of cases across the qualitative categories. The more peaked the distribution, the more difficult it will be to discern EMI patterns.

That the distribution of cases across values of a variable matter for our ability to assess a theory appears true of every theory and every statistical means of studying inequality. For example, if one’s class category scheme contains capitalists, petite bourgeoisie, and proletarians, but every adult in the society is a proletarian, one will

not be able to identify factors that determine being a capitalist *in* that society (though one could study the structures and extractive relations that enable a society composed only of proletarians). The extreme case makes the point, but the point applies in general. Thus, all research methods require sufficient variation to parse the factors that produce the outcome. Without the needed variation, the theory cannot be explored via empirical means.

With respect to the number of categories, it is clear that one cannot distinguish variance in the qualitative dimension from variance in the quantitative dimension of an outcome if the outcome has only two categories. At the same time, if there are several categories, and many with only slight differences from others (e.g., law school vs. medical school) from the perspective of stratification, one may find it easy to produce the appearance of diverging trajectories.

A missing literature on the impact of number could greatly aid EMI-focused analyses. Kanter (1977) provided one beginning of such a literature, but subsequent work focused on the substantive claims and the implications for gender and racial disadvantage in organizations (e.g., Jackson, Thoits, & Taylor, 1995). Such work is important, but Kanter's (1977) meta-methodological point that analysts should attend to numbers both relative and absolute *independent* of the content to which those numbers were attached (e.g., numbers of people from Group X, numbers of categories available for diagnoses), has not spurred a visible literature. Yet, it is just such work that is needed to aid analysts' efforts to discern whether the outcome is likely to be one worthy of investigating the qualitative question posed by EMI. The implication echoes the earlier observation that systems need certain characteristics—scope conditions—for EMI patterns to be discernible and/or relevant.

Because that literature appears to be missing, we can only provisionally note that the sweet spot for testing the qualitative question of EMI may be three or four categories. Social outcome categories associated with training may be imagined as less constrained, but even there, certain dynamics should reduce the realized number of truly distinct such locations. For example, families play a role in steering children into training and in allocating needed resources from other domains as well (e.g., health insurance and care), and evidence suggests contestation among families, managed by officials of the state and system involved, can affect the number of options available and their content (e.g., Dougherty, 2001). Yet, human cognitive constraints can reduce the desirable or feasible number of truly distinct strata for, as research indicates, as options increase so does decision-making difficulty (Iyengar & Lepper 2000; Tversky & Shafir, 1992). Thus, political pressures may motivate the multiplication of positions, but many new positions may be equivalent from the perspective of stratification.

The implication of these observations is that EMI likely offers little illumination for the analyst seeking to assess the correlates of, for example, students' allocation to any one or two of, say, the 119 different majors at the University of California–Berkeley. Such an analysis will parse, for example, students' chances of majoring in anthropology as opposed to sociology as opposed to philosophy as opposed to history. From the perspective of the student majoring in a field, these differences may be massive; from the perspective of the reproduction of inequality, they may be nonexistent. Indeed, for

any given level of evenness of univariate distribution, the more categories there are, the more theoretically trivial any finding of EMI is likely to be. Thus, faced with such data, the analyst must use theory to coherently construct a small number of truly distinct strata out of the many possible categories.

In sum, analysts must either select outcomes with a modest set of positions that differ qualitatively in stratification-related ways (e.g., zero vs. a great deal of future job autonomy), or they must use theory and substantive knowledge to recode a dependent variable with many categories into a smaller set that differ qualitatively in stratification-related ways. Ideally, such outcomes should not have highly peaked distributions. Otherwise the thesis of EMI is entertained only in a mechanistic and likely uninformative way.

Principle 2: Select and Test an Appropriate Modeling Framework

There are two main modeling frameworks from which one might choose—ordered categorical or unordered categorical models. Within the former, there are two dominant models; within the latter, there is one. Each framework has advantages and disadvantages.

Ordered Categorical Regression Models

One dominant model is the ordered probit model. With a four-category outcome as an example, an ordered probit model can be specified as

$$\begin{aligned}
 P(y_{it}=0) &= \Phi(\mu_1 - \beta_t' \mathbf{X}_{it}) \\
 P(y_{it}=1) &= \Phi(\mu_2 - \beta_t' \mathbf{X}_{it}) - \Phi(\mu_1 - \beta_t' \mathbf{X}_{it}) \\
 P(y_{it}=2) &= \Phi(\mu_3 - \beta_t' \mathbf{X}_{it}) - \Phi(\mu_2 - \beta_t' \mathbf{X}_{it}) \\
 P(y_{it}=3) &= 1 - \Phi(\mu_3 - \beta_t' \mathbf{X}_{it})
 \end{aligned} \tag{1}$$

for I individuals across t transitions, where Φ signifies the normal probability density, μ_s represent thresholds on that density, \mathbf{X}_{it} represents a vector of explanatory variables for person i at time t , and β_t represents a vector of estimated parameters linking variables to the outcome at time t . An analogous ordered logit model can be specified instead. Note that the specification equates thresholds for all transitions, but the effects of characteristics as well as some of the characteristics themselves are allowed to vary over transitions.

The ordered logit model is the alternative dominant model. Because little of consequence for assessing EMI differs across these models, we discuss them together.

One advantage of ordered models is parsimony. If there are K independent variables, the model estimates $K + 1$ coefficients. Parsimony facilitates interpretation and,

by concentrating social background effects in a small set of coefficients, allows a more focused test.

However, there are costs to parsimony. One cost is that if a variable has effects on one category but not on another, the model will misestimate the effect on at least one outcome. For example, measured achievement might have a positive effect on students' likelihood of entering top-tier colleges but might have no effect on entry to colleges outside the top tier. In such a case, the ordered model will either estimate a globally positive coefficient—erring for colleges outside the top tier—or will estimate a statistically nonsignificant coefficient—erring for top-tier colleges.²

Even if such is not the situation, the framework is based on certain assumptions. One assumption is that the outcome variable categories are a priori ordered, though distances between the categories need not be known. One could implement an empirical strategy for assessing or establishing the validity of the order assumption (e.g., Goodman RC II model; Goodman, 1979), but empirical methods are rarely employed if one can make a persuasive theoretical or substantive argument. If one cannot make or defend an assumption about the order of the outcome variable categories, one could switch to an unordered model (see the next section).

The ordered probit/ordered logit models also require a parallel regression (proportional odds) assumption; formally, the model assumes the first derivatives of all categories' predicted cumulative distribution functions are equal. If the first derivative equality (FDE) assumption is not satisfied, results will distort reality, undermining proper inference.

Consequently, analysts must test the FDE assumption prior to using an ordered model. In order to test the assumption, the researcher should have already established or defended the specified order of the categories (e.g., the academic track is above the general track, which is above the vocational track). Assuming the a priori order is justified, one can test the FDE assumption using the likelihood ratio test. This test tests all coefficients simultaneously. If it rejects the parallel regression assumption, one does not know if the rejection occurs because most or all of the relationships between independent variables and the outcome violate the assumption or, instead, whether one or two of the independent variables are the culprits. To assess this possibility, one can use a Wald test, as described by Brant (1990; which is why it is sometimes called a Brant test). A statistically significant test result means the FDE assumption is violated.

If the FDE assumption is not satisfied, one could switch to the multinomial logit model. Yet, if one wants to use a model for ordered outcomes, the generalized ordered logit model can be estimated (Williams, 2006).

Unordered Models

The dominant unordered outcome model for assessing EMI is the multinomial logit model (Nerlove & Press, 1973), which can be specified as

$$Pr(Y_{it}=j) = \frac{1}{1 + e^{\sum_{k=0}^K \beta_{jk} X_{ikt}}} \tag{2}$$

for I individuals across t transitions, where \mathbf{X}_{ikt} represents each of K explanatory variables for person i for transition t , β_{jkt} represents parameters linking variables to each category J for each transition (t) studied, and $X_{i0t} \equiv 1$ to specify the time-specific (or transition-specific) constant.

One advantage of the multinomial logit model is that it does not require ordered outcome categories. Although EMI requires ordered outcomes, if partially ordered outcomes are specified in an EMI study they could be studied with the multinomial logit model. In using a multinomial logit model, the researcher should note the unordered nature of the categories. Using the multinomial logit model when an ordered model could have been used will produce estimates with larger standard errors than could have been obtained with an ordered model, which can lead to erroneous inference. This occurs because using the multinomial logit estimates $(J - 1)(K + 1)$ coefficients while an ordered outcome model estimates $K + 1$ coefficients.

If outcome categories are ordered, one could switch to a model that assumes ordinality, such as those discussed above. However, as those models have other assumptions, one may still prefer to use the multinomial logit, even given its reduced efficiency.

For example, an advantage of the multinomial logit model is that if a variable has effects on one category but not on another, the model will correctly estimate the association for both outcomes. Indeed, the multinomial logit coefficient will capture the effect (relative to the baseline), *even* if the sign of that effect differs across outcome categories. To concretize the issue, consider that, relative to lacking a PhD, a sociology PhD might be highly positively associated with obtaining a faculty position in the social sciences, moderately positively associated with obtaining a faculty position in the humanities, but *negatively* associated with obtaining a faculty position in the physical sciences. A multinomial logit model could discern such a difference in sign because the model is very flexible. This flexibility means that an independent variable's relation with an outcome category has no connection with that independent variable's relation with some *other* outcome category. In contrast, in the fully ordered models, a global constraint exists that destroys the ability of each coefficient to adjust for specific categories.

Alas, this very flexibility is a disadvantage in some ways because it means the model is not parsimonious. As noted above, the model produces $(J - 1)(K + 1)$ coefficients, making the model very difficult to interpret. Not only does the profusion of coefficients hinder interpretation but also all interpretations of coefficient "effects" are relative to a baseline. Consider the following verbalization of illustrative mock statistical model results: Children of higher SES are more likely to enter 4-year college than are children of low SES, compared with not attending college; however, children of higher SES are *less* likely to enter 2-year college than are children of low SES, compared with not attending college. Even if an audience understands the multinomial logit model, such sentences may easily produce confusion, owing largely to the inherent "double comparison" aspect of the model.

The multinomial logit model assumes that the effects of variable X on outcome j have nothing to do with the effects of variable X on outcome $j - 1$. This is known as the independence of irrelevant alternatives (IIA) assumption.

Researchers should test the IIA assumption. One can test the IIA assumption using a Hausman test or a Small–Hsiao test. These tests are said to sometimes conflict, and some scholars (e.g., Long & Freese, 2006) advise against their use. However, if both are consistent in supporting the IIA assumption, it is at least some warrant for using the multinomial logit model.

If the IIA assumption is not satisfied, yet one wants to use a model for unordered outcomes, there are alternatives. One promising alternative is the stereotype logistic regression model (Hilbe, 2009). This model relaxes the IIA assumption while still allowing unordered categories.

Principle 3: Specify a Delimited Population

At first blush, it would seem that the population for study is obvious: It is the population at risk of the outcome of interest. Yet, this obvious claim does not fully identify the study population, for some in the at-risk population are such that including them can bias results. Hence, one must carefully consider the various subpopulations to construct a delimited population for study. This counsel is, in a sense, a theory-focused manifestation of the general idea that one must study comparable (e.g., matched) cases to make inferential progress.

To concretize the issue, consider, for example, some of the challenges cross-national migration might pose for study of EMI. It is no secret that people are on the move around the world (e.g., Byrne, McGinnity, Smyth, & Darmody, 2010). The rise of immigration makes many nation-specific data sets contain many people who are not native to the nation under study or whose parents are not native to the nation under study. Immigrants are participants in stratification processes in the nation under study, whether in their own generation or intergenerationally. They or their children may be at risk for experiencing the outcome of interest, and their experiences may be directly relevant to the maintenance of inequality (e.g., Darmody, Byrne, & McGinnity, 2014). How should one proceed analytically in such cases?

The problem is that the SES of immigrants—parents of the children of immigrants—are unlikely to be sufficiently well-measured by their characteristics in the destination country. Poor data on immigrants' SES could bias estimation of socioeconomic background effects on both the children of immigrants and the children of the natives in a way that undermines the validity of the assessment of EMI.

For example, immigrants (to the United States) are more highly educated than their nonimmigrant peers back home (Feliciano, 2005). Yet, many immigrants who worked in the primary labor market in their home country, or who were on track to do so, may work in the secondary labor market or in an enclave economy in the destination country (Bailey & Waldinger, 1991). Thus, many first-generation children will appear to have low socioeconomic background even as they benefit from resources their parents could not capitalize on occupationally in the destination country but *could* utilize in helping their children, such as socialization in line with a history of sole proprietorship (e.g., Kohn, 1963; Lareau, 2002) or access to an enclave labor market that provides some of the benefits traditionally associated with the primary labor market (e.g., Bailey & Waldinger, 1991). Using parental occupation in the destination country for immigrants

could bias results in complex ways. The bias in socioeconomic background coefficients could be downward because if some immigrant parents have destination occupations lower than their home country occupations, and if their home country occupations facilitate their efforts to socialize their children for success, then the coefficients for parents' occupation will be biased downward because the children of some persons with parents of low destination-country occupations will actually be receiving advantages typically available only to children of parents with higher occupations. This will reduce outcome differences between those of measured low and high parental occupation, attenuating the coefficient for that variable.

Notably, the bias could be in the opposite direction. For example, if some immigrants face discrimination, and discrimination is unmeasured, its operation may reduce the advantages they can transmit to their children below those typical for their socioeconomic position. If immigrants' receiving country occupations are concentrated at the lower ends of the occupational distribution, the suppression produced by discrimination will increase the difference between those at the top and bottom, consequently raising the slope on socioeconomic background.

Whichever way the bias goes, for assessing EMI, the result is the same: Under such conditions, predicted values may be less likely to show a diverging trajectories pattern, as the estimated coefficient may be pushed beyond the EMI bounds (Lucas, 2017). Similar reasoning applies to other socioeconomic background indicators (e.g., parents' education).

One easy response is to add a dummy variable for first-generation status. Alas, this "fix" implies that all the coefficients are the same for first-generation children and others, which is likely insufficient, as the previous paragraph noted the likelihood of major deficiencies in the measures of socioeconomic background for immigrants. Yet, if one interacts all the socioeconomic background variables (and the other variables as well) with first generation, the implication is that unless the aim is to study the difference between first-generation immigrants and others, one would do as well to simply estimate models on separate data.

If one has data on parents' occupations, educations, earnings, and more in the origin country, one might be able to address these problems. Alas, few if any nationally representative data sets contain the array of information necessary to assess EMI while also containing measures of socioeconomic location in the origin country for immigrant parents. Thus, it is likely that the best option is to either estimate separate models for natives and children of immigrants, or remove the children of immigrants from the analysis altogether.

The specific example concerns immigration. But the general principle is that analysts must proceed carefully to assure they have selected for analysis an appropriate population.

Principle 4: Specify Theoretically Focal Persons

The theory calls on analysts to compare two people who are the same on everything else except socioeconomic background. Ideally, at least some of those comparisons will be specified prior to the estimation process, to reduce the chance of analysts'

searching for a set of covariate values that will confirm or disconfirm the theory. The need to specify at least some theoretically focal persons prior to model estimation motivates the placement of this issue as the fourth principle.

Of course, once we say we want people who are “the same on everything except socioeconomic background,” we are committed to identifying specific values for the covariates in the model. There are a few clear options for how to select values for the comparison. Even so, the most important counsel is that the selection should have theoretical coherence, not simply be a series of mechanistically imposed averages.

One tactic is to set all continuous variables at their mean or median. An advantage of this approach is that these values are such that there are likely multiple cases in and around that location, strengthening inference about such “theoretical” persons.

A second tactic is to set one or more continuous or count variables at theoretically interesting values. For example, one could set number of siblings, a count variable, at zero because one has theoretical interest in only children. Such an analysis might be of special interest for data on China for some cohorts, for example. Or, as another example, one could set a standardized test score two standard deviations above the mean because one wants to consider the outcomes of students with high levels of prior achievement.

A third tactic is to set categorical variables at some value of interest. For example, one might set the value of male at 0, which means the calculations will concern females. Leaving categorical variables at zero can simplify calculations; thus, one might want to think ahead of time about how to code the categorical variables to make the zero values of categorical variables specify a case of theoretical interest.

A fourth possible tactic is to use the mean of any dichotomous variables. This tactic is strongly discouraged, because the calculations then refer to no one. What, after all, would be the point of plugging in .4 for Male and multiplying it by the coefficient for Male? The result conveys something about the aggregate, but as there is no one scored .4 on female in most datasets, it means *all* of your calculations refer to no one. As testing EMI requires identifying theoretically focal *persons* for comparison, using this tactic would make one immediately violate the logic of the theory and its testing.

Principle 5: Avoid Specifying Distractive Controls

A challenge posed by the necessity of identifying theoretically focal persons is that the only factors that can be used to identify those persons are the variables in the data set. Although one could select subsets of the data for analysis, the typical approach is to specify covariates for inclusion. Thus, asking what groups form the basis upon which comparisons will be made implies asking what covariates should be included in the model so that one can set covariate values so as to identify the theoretically focal.

One way to proceed is to place variables into two categories: (a) those variables that specify socioeconomic background and (b) those variables that help identify the persons who are viable candidates for the transition or outcome of interest. The second set draws our attention here.

One important counsel is to include only those variables that help identify the viable candidates for categories of the outcome of interest. For example, if one of the

outcome categories for the variable “enters college” is “top-tier college,” one would want to include covariates that play a role in allocating persons to such colleges. Clearly, measured achievement would be one such variable.

However, some variables available in the data set might not be appropriate variables to include despite their possible relation to the outcome of interest. For example, a data set might contain information on whether the student received financial aid. If the aim was simply to include all correlates of the outcome, this variable might belong in the model. However, if the aim is to assess the veracity of EMI, inclusion of this variable is likely inappropriate. Given an interest in socioeconomic background effects, immediately we must wonder what the financial aid variable means. Is it an indicator of poverty and thus itself a socioeconomic background variable? Is it another measure of achievement, perhaps reflecting a dimension not tapped by standardized tests such as musical virtuosity or athletic prowess? Is it another manifestation of the dependent variable; for example, are the students judged most promising given financial aid, those judged least promising denied admission, and the students in the middle admitted without aid? Alas, matters may be even more complex, as it may be that one of the above interpretations applies to some students in the data set while a different interpretation applies to other students in the data set? Furthermore, the above does not exhaust the possibilities, as financial aid could indicate many other phenomena (e.g., ethnicity, religion, or other categorization). The variable’s ambiguity makes its inclusion in models estimated to assess EMI inappropriate. Such a variable is, from the perspective of the assessment of EMI, a *distractive control*.

Because the institutionalization of data collection often has depended on multiple research communities working together to encourage states to support data collection, it is likely that most nationally representative data sets contain many variables that from the point of view of assessing EMI are distractive controls.

Once a variable is available in a data set, it can be difficult to resist the urge to include the variable in one’s model. To help with the decision, one could add the removal of distractive controls to an analysis of a directed acyclic graph (DAG). Through DAGs one may assess which variables need be included in or excluded from a model to secure proper causal inference (Elwert, 2013; Morgan & Winship, 2007; Pearl, 2010). Thus, DAGs already justify removal of *methodologically* distractive controls from statistical modeling. We advise analysts to construct a causal graph, determine and remove any variables that are (owing to measurement or other reasons) *theoretically* distractive controls for study of EMI, and then analyze the DAG to determine which coefficients must be estimated to identify the causal effect of socioeconomic background. It is likely that efforts to assess EMI will improve greatly if analysts proceed in such a manner.

Principle 6: Use and Test a Broad Specification of Socioeconomic Background

Many analysts use multiple measures of socioeconomic status, such as mother’s education, father’s education, mother’s occupation, father’s occupation, family income, farm family status, and number of siblings, in the same model. This approach more fully

specifies socioeconomic background. However, it means each single regression coefficient reflects the association between one socioeconomic background indicator and the outcome while *holding the other socioeconomic background factors constant*.

Note, however, that EMI makes no claim about the impact of one socioeconomic variable (e.g., mother's education), and certainly makes no claim about the effect of one socioeconomic variable holding other socioeconomic variables constant. EMI asks whether socioeconomically advantaged and socioeconomically disadvantaged persons follow a diverging pathways pattern. Thus, the best test of the theory simultaneously sets all socioeconomic background variables at all high and then all low values after setting all covariates at the value needed to specify theoretically focal persons.³

Because socioeconomic background is a larger conceptual factor, only partly captured by any one indicator, it may be that analyses that use only one or two indicators of socioeconomic background may be more likely to reject the theory than are those with several indicators. Thus, analysts are encouraged to use multiple indicators and to conduct tests that use all socioeconomic indicators simultaneously, for rejecting EMI under those conditions would be illuminating.

Principle 7: Use Appropriate Means to Assess EMI

For theoretical reasons noted above and elsewhere, one must use predicted probabilities, not regression coefficients, to assess EMI. In calculating predicted values, high SES should be one-half or one standard deviation above the mean or median, and low SES should be one-half or one standard deviation below the mean or median. Using more extreme values of the socioeconomic variables stacks the deck in favor of finding support for EMI, while at the same time greatly increasing the risk of drawing inferences off the support. Neither of those implications will lead to deepened knowledge of the stratification system.

The preference for predicted probabilities has a methodological basis as well as a theoretical one, even as it requires analysts to proceed with reflection rather than rote practice. Reflection indicates that analysts should calculate point estimates of predicted values but *not* calculate confidence intervals for those predictions.

Using Predicted Values

One methodological argument for using predicted values draws on Long's (1997) observation that to identify regression model coefficients for categorical outcomes, one must make assumptions about the mean and variance of either the error or a latent outcome variable. Yet, there is rarely a theoretical reason for any particular identifying assumption. The assumptions are usually invoked for convenience, only. Long (1997) observes that

if we change the assumption regarding $\text{Var}(\varepsilon|x)$, the β 's also change. *Accordingly, the β 's cannot be interpreted directly since they reflect both: (1) the relationship between the x 's and [the latent variable] y^* ; and (2) the identifying assumptions.* While the identifying

assumptions affect the β 's they do not affect $\Pr(y = 1|\mathbf{x})$. More technically, $\Pr(y = 1|\mathbf{x})$ is an *estimable function*. An estimable function is a function of the parameters that is invariant to the identifying assumptions. (p. 49, italics in original)

Furthermore, categorical model coefficients do not reveal how the estimated probabilities change as a regressor increases. For middle categories in ordered models, and all categories in multinomial models, the probability can rise *and* fall as regressor values rise (Long, 1997). As one needs predicted categories to test EMI, and coefficients do not convey this information, one cannot test EMI using coefficients. Indeed, marginal effects estimates are also insufficient, failing to convey the needed information. Instead, predicted probabilities—the estimable (i.e., invariant) function of categorical variables (Long, 1997)—are needed. These methodological observations culminate in our recognizing that we should prefer predicted probabilities in assessing EMI because the probabilities (and thus the findings) are robust to the identifying assumptions and clearly indicate the category-specific role of a variable.

Do Not Calculate Predicted Probability Confidence Intervals

The next step would seem to be to calculate confidence intervals for the predicted values. This, however, is incorrect for assessing EMI. To show the error, let us return to the bedrock reasoning behind tests of coefficients, using ordinary least squares regression as an example. Consider the following model:

$$Y = b_0X_0 + b_1X_1 + b_2X_2 + b_3X_3 + b_4X_4 + e \quad (3)$$

where $X_0 \equiv 1$ for all cases. All coefficients are estimates with a nonzero standard error, s_{bk} . One can use the standard error to calculate a confidence interval to test whether 0 is a nontrivial possibility for the value of b_1 . If to test a theory one needed to test whether $b_1 - b_2 = 0$, the standard error for the difference, calculated as $(s^2_{b_1} + s^2_{b_2} - 2s_{b_1,b_2})^{1/2}$, would be used. The principle in both tests is that estimates' uncertainty must be factored in if that uncertainty could affect the result.

Yet, if to test a theory one needed to see whether the predicted value of Y for $X_1 = 0$ (i.e., $Y_{X_1=0}$) is different than the predicted value for Y if $X_1 = 1$ ($Y_{X_1=1}$), one would need to fix values for all variables in the model except X_1 . Thus, any *focused* test of whether $Y_{X_1=0} = Y_{X_1=1}$, would consider only the standard error for b_1 . But, if the test for b_1 was conducted earlier, there is no need to conduct the test again.

One might rightly contend, however, that the coefficients contribute to the predicted value and, as estimates, contain uncertainty. Table 1 reflects that contention by showing a range for the values of each coefficient. However, note that there is no way to set the coefficient at, say, r , for the calculation of $Y_{X_1=0}$ while setting the coefficient at *not- r* for the calculation of $Y_{X_1=1}$. Thus, the coefficients are fixed across the two calculations, and thus the variance of their estimates should not contribute uncertainty to the calculation. Consequently, the appropriate focused test of the equality of the two predicted values would still use only the standard error for b_1 , a test that has already been conducted.

Table 1. Predicted Value Calculation Spreadsheet.

Coefficient	Variable	Fixed value	Differentiating value for comparison	
			$Y_{X1=0}$	$Y_{X1=1}$
Lo - b_0 - Hi	X_0	1	—	—
Lo - b_1 - Hi	X_1	—	0	1
Lo - b_2 - Hi	X_2	7	—	—
Lo - b_3 - Hi	X_3	10	—	—
Lo - b_4 - Hi	X_4	22	—	—

The point does not change when one moves to the categorical dependent variable case. Consider the following logistic regression model and four possible outcome categories:

$$\Pr(Y=1) = 1 / \left(1 + e^{\lambda S + \eta X - \delta_1} \right) \tag{4.1}$$

$$\Pr(Y=2) = \left[1 / \left(1 + e^{\lambda S + \eta X - \delta_2} \right) \right] - \left[1 / \left(1 + e^{\lambda S + \eta X - \delta_1} \right) \right] \tag{4.2}$$

$$\Pr(Y=3) = \left[1 / \left(1 + e^{\lambda S + \eta X - \delta_3} \right) \right] - \left[1 / \left(1 + e^{\lambda S + \eta X - \delta_2} \right) \right] \tag{4.3}$$

$$\Pr(Y=4) = 1 - \left(1 / \left(1 + e^{\lambda S + \eta X - \delta_3} \right) \right) \tag{4.4}$$

where λ is a coefficient on socioeconomic origins, η is a coefficient on some non-socioeconomic background factor, and δ_j is a threshold dividing category j from category $j + 1$ in a latent, logistically distributed variable underlying the observed categorical variable Y (Long, 1997). If we set high SES to equal 1 on S , and low SES to equal 0 on S , then the predicted values for each category are as follows for low SES:

$$\Pr(Y=1|S=0) = 1 / \left(1 + e^{\eta X - \delta_1} \right) \tag{5.1}$$

$$\Pr(Y=2|S=0) = \left[1 / \left(1 + e^{\eta X - \delta_2} \right) \right] - \left[1 / \left(1 + e^{\eta X - \delta_1} \right) \right] \tag{5.2}$$

$$\Pr(Y=3|S=0) = \left[1 / \left(1 + e^{\eta X - \delta_3} \right) \right] - \left[1 / \left(1 + e^{\eta X - \delta_2} \right) \right] \tag{5.3}$$

$$\Pr(Y=4|S=0) = 1 - \left(1 / \left(1 + e^{\eta X - \delta_3} \right) \right) \tag{5.4}$$

and as follows for high SES:

$$\Pr(Y=1|S=1) = 1 / \left(1 + e^{\lambda + \eta X - \delta_1}\right) \quad (6.1)$$

$$\Pr(Y=2|S=1) = \left[1 / \left(1 + e^{\lambda + \eta X - \delta_2}\right)\right] - \left[1 / \left(1 + e^{\lambda + \eta X - \delta_1}\right)\right] \quad (6.2)$$

$$\Pr(Y=3|S=1) = \left[1 / \left(1 + e^{\lambda + \eta X - \delta_3}\right)\right] - \left[1 / \left(1 + e^{\lambda + \eta X - \delta_2}\right)\right] \quad (6.3)$$

$$\Pr(Y=4|S=1) = 1 - \left(1 / \left(1 + e^{\lambda + \eta X - \delta_3}\right)\right) \quad (6.4)$$

The only difference between Equations 5.c and 6.c is the coefficient λ , where c is the category for which the predicted value is calculated. We should have already tested whether λ is discernibly different from zero, for λ being discernibly different from zero is a necessary precondition for usefully exploring whether EMI's qualitative pattern exists. If λ is not discernibly different from zero, we should not be entertaining the qualitative aspects of EMI.

The remaining elements of Equations 5.c and 6.c— η , δ_1 , δ_2 , δ_3 , and X —are equal across all the calculations, so they contribute nothing to distinguishing 5.3 and 6.3. Two implications follow. First, assessing whether $A - B = 0$, where $A = \max \text{category} [\Pr(Y_c|S = 0)]$ and $B = \max \text{category} [\Pr(Y_c|S = 1)]$ is already an incredibly stringent test, because, except for the social background variable, all of the regressor values are exactly the same for each predicted value pair, and all of the coefficients are exactly the same for each predicted value pair. With almost everything in both calculations the same, it is very difficult for a categorical difference to emerge.

Second, the thresholds (δ s) are also the same, making the size of the categories the same for both high and low SES students. In reality, some students might have much tougher-to-meet thresholds than other students (e.g., Lucas, 1999, pp. 101-114). However, setting thresholds equal for all students concentrates all differences in the coefficients of interest. With such a strategy, making the test include imprecision in the thresholds is not only unnecessary, it is detrimental.

In sum, calculating a confidence interval for the $\max[\Pr(Y_c|S = 0)] - \max[\Pr(Y_c|S = 1)] = 0$ test makes imprecision in all other parameters *that are the same for* $\max[\Pr(Y_c|S = 0)]$ and $\max[\Pr(Y_c|S = 1)]$ factor into the assessment of our estimates of $\max[\Pr(Y_c|S = 0)]$ and $\max[\Pr(Y_c|S = 1)]$. Using confidence intervals in this manner is to overcorrect the test. Thus, we advise against calculating standard errors for the predicted values in testing EMI.

Assessing EMI as a Response to Historic Challenges

As noted earlier, several historic challenges bedevil efforts to study inequality. A solid study of EMI necessarily addresses many of those challenges. We discuss two quite stubborn challenges—unobserved heterogeneity and the challenge of comparability—to illustrate how study of EMI addresses historic impediments to the investigation of inequality.

Unobserved Heterogeneity and Assessments of EMI

Education transitions analyses are discrete-time event history analyses. For any discrete-time event history process unobserved heterogeneity becomes a complex threat to proper inference through its relation to either time-constant or time-varying variables (Vermunt, 1997), which we can regard as the first and second pathways of harm. We can decompose \mathbf{X} in Equation 1, above, into time-constant (C) and time-varying (V) variables, making Equation 7:

$$\begin{aligned}
 P(y_{it}=0) &= \Phi(\mu_1 - \lambda_t' \mathbf{V}_{it} - \gamma_t' \mathbf{C}_i) \\
 P(y_{it}=1) &= \Phi(\mu_2 - \lambda_t' \mathbf{V}_{it} - \gamma_t' \mathbf{C}_i) - \Phi(\mu_1 - \lambda_t' \mathbf{V}_{it} - \gamma_t' \mathbf{C}_i) \\
 P(y_{it}=2) &= \Phi(\mu_3 - \lambda_t' \mathbf{V}_{it} - \gamma_t' \mathbf{C}_i) - \Phi(\mu_2 - \lambda_t' \mathbf{V}_{it} - \gamma_t' \mathbf{C}_i) \\
 P(y_{it}=3) &= 1 - \Phi(\mu_3 - \lambda_t' \mathbf{V}_{it} - \gamma_t' \mathbf{C}_i) \tag{7}
 \end{aligned}$$

Tam (2011, p. 288, Note 1) contends that Lucas (2001) “does not adjust for stable unobserved heterogeneity other than adding [seven] test scores as covariates.” This claim is mistaken in two ways. First, it forgets that in models focused on studying effects of socioeconomic factors (e.g., family income, parents’ education) analysts have often viewed most of the unobserved heterogeneity as owing to cognitive achievement (e.g., Mare, 1980, on ability [which cannot be measured independent of achievement]; Cameron & Heckman, 1998, on ability and motivation [which arguably interact to produce achievement]). Realization of this historical tendency suggests that adding seven achievement tests to the model (as well as time-varying grades in relevant courses) is an important advance. Second, Tam (2011) ignores the way that the comparisons required for assessing EMI further address the threat unobserved heterogeneity can pose. Explicitly considering the pathways through which unobserved heterogeneity can have deleterious effects may assuage such concerns.

Using Equation 7, if unobserved time-constant determinants (C) are correlated with time, then spurious interactions between time and C will be produced that will distort the comparison between the effect of any C variable at time t and time $t + 1$. Note, however, that not only does the analysis of EMI specify multiple variables that are regarded as accounting for much of the unobserved heterogeneity but also study of EMI does not entail comparing coefficients across time or space. Thus, the impact of this pathway by which unobserved heterogeneity can bias the assessment of the theory is muted.⁴

Considering the second pathway through which unobserved heterogeneity can matter, time-varying variables pose a problem owing to unobserved heterogeneity, for unobserved risk factors may create changes in the values of time-varying independent variables. If this happens, then the effects observed for those covariates will be at least partially spurious. Yet, studies of EMI highlight independent variables, such as mother’s education, that are largely invariant during children’s education trajectory.

Unobserved heterogeneity cannot cause changes in the values of variables whose values do not change. In other words, the focal independent variables for assessing EMI are factors such as parents' occupations, parents' educations, income, number of siblings, and farm background. Farm background is a stable factor, and number of siblings of middle and high school students changes little. Permanent income rather than transitory income could be studied, and parents' education also changes little after their children are middle and high school age. Finally, occupation is stable enough to be proposed as an indicator of permanent income (e.g., Hauser & Warren, 1997). In sum, research design and the focus of the theory undermines the power of unobserved heterogeneity to bias results.

These observations indicate that the stratified outcomes model, when used to evaluate EMI, has several features that respond to the challenges posed by unobserved heterogeneity specifically, and many questions posed to the education transitions model in general. Although unobserved heterogeneity always remains a possibility, the focus on EMI, and specification of a model tuned to that aim, can greatly reduce the threat unobserved heterogeneity usually poses.

Cross-National Comparative Analysis

Cross-national comparative analysis raises challenges to all research. Even though EMI comes out of a tradition of cross-national comparative research, special challenges confront those interested in assessing EMI cross-nationally.

Cross-national comparative researchers agree on the importance of assuring that phenomena must be comparable enough to make comparison illuminating. By far the most common view of what is required for cross-national research is reflected in a strategy we term *institutional standardization*.⁵ To invoke institutional standardization one studies the same institutional location in each nation. So, for example, an analyst might study the determinants of college entry in several different countries, as in Figure 1, focusing on differences in the coefficients for X while Z is controlled. College entry, X , and Z all mean the same and are measured the same across all J countries. The analyst directly compares the magnitudes of β_1 across countries.

This approach is useful, perhaps even ideal. But, the diversity of nations means that it can be difficult to truly standardize beyond a few nations. For example, not only do education systems vary cross-nationally, but data collection programs, survey procedures, and even the variables collected and the answers available for respondents to select vary as well. For example, some nations often bar collection of racial classification (e.g., France) or religious identification (e.g., the United States) data. Historical trajectories and idiosyncratic features may explain such differences in data collection, but the impact is to make it impossible to standardize study variables of potential importance in every nation one might desire, *even* for some variables that may be important in nations for which the data is not collected.

Such particularities led Shavit and Blossfeld (1993) to standardize by asking each country analysis team to use the same model, muting most country-specific issues (e.g., race was omitted from models for the United States). Shavit and Blossfeld

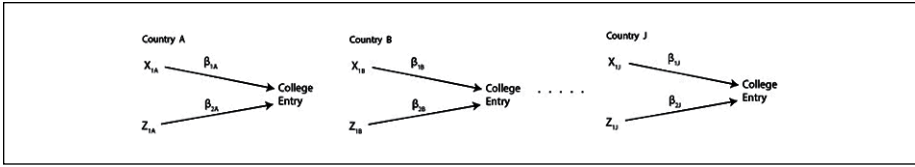


Figure 1. Institutional standardization.

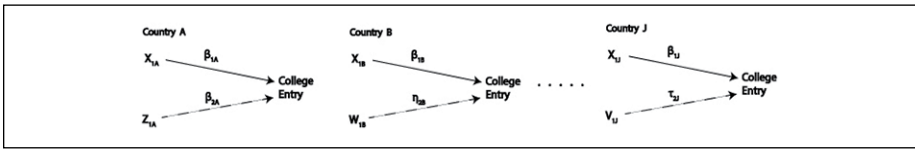


Figure 2. Partial standardization.

termed their approach *partial standardization*, which is reflected in Figure 2. The outcome is measured the same in each country, but potentially important covariates for any particular country are absent from the model (as reflected by the dashed lines for Z for Country A, W for Country B, and V for Country J in Figure 2). Partial standardization does not allow analysts to directly compare the magnitudes of β_1 ; instead, analysts devise a narrative for each country, using the relevant coefficients. After the narratives are constructed, they may be compared and contrasted qualitatively. Partial standardization offers a tractable response to many of the challenges of cross-national research. Both classic analyses of causality (Lieberson, 1985) and recent research on the structure of causal evidence (Elwert, 2013; Morgan & Winship, 2007; Pearl, 2010) suggest the approach may be able to estimate causal effects under many conditions.

However, for a theory like EMI, partial standardization will fail because, while in every education system students eventually end their schooling, the number, content, and meaning of positions persons navigate as they pursue education varies across countries. Thus, the dependent variable for a polytomous outcomes model varies across countries, making partial standardization both much more challenging and likely of limited value.

The challenge is even more acute for assessing EMI. EMI focuses on *substantive* significance, and thus points attention to the point(s) in the system where consequential effects are likely to be generated. Certainly, study of all polytomous positions throughout the entire series of education transitions is desirable. Yet, study of all such positions over all levels of schooling is often infeasible. Study of the entire series is rendered unnecessary; however, if analysts can identify stage(s) either most likely or least likely to contain EMI processes. Finding EMI in either case confirms an EMI dynamic, while rejecting it in the former place suggests EMI may *not* characterize inequality dynamics in the country studied.

Lucas (2001) studied high school transitions and college entry, but did not explicitly justify selection of these levels, possibly because their importance for the period

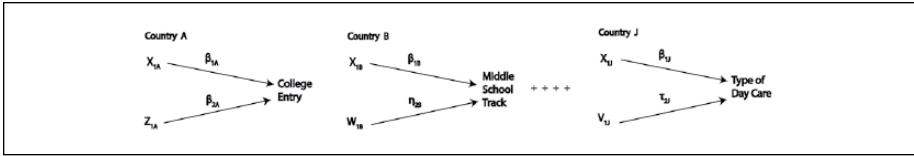


Figure 3. Salient standardization.

and nation studied was well-known. For the era studied, the college track/noncollege track measured achievement gap exceeded the size of the achievement gap between vocational students and high school dropouts (Gamoran, 1987). College track students generally obtain more education than their noncollege track peers (Hauser et al., 1976), and evidence indicates that students taking college preparatory courses outearn their peers 10 years after graduation (Rose & Betts, 2004) while college entrants outearn high school graduates (e.g., Card, 1999).

Yet, EMI is a general theory, not specific to only high school track location and college entry. Thus, in order to assess the qualitative claims of EMI, the analysis must focus on the place(s) in the institutional structure where consequential qualitative distinctions likely occur. Hence, a cross-national comparative analysis of EMI would use what we term *salient standardization*. Under salient standardization, one selects focal points for analysis that may differ across countries, but with each regarded as consequential in the country for which it is analyzed. In one country the salient location for qualitative distinctions might be entry to colleges of different types, whereas for another country the salient location might be assignment to different types of tracks in middle school; and for a third country the salient question may be whether the child attends day care and, if so, what kind. Each institutional location is salient both for the theory under investigation and for the specific nation observed. Figure 3 alters Figure 2 to illustrate the strategy of salient standardization.

The underlying reason for the necessity of salient standardization is that EMI is a general theory that has been applied to education systems. General theories do not require institutional standardization. For example, if a theory claimed that leftist parties are more likely to address issues of gender equality, cross-national study of the theory would not need to standardize on the *same* issue of gender equality across countries (just as the theory would not need to standardize on the *same* leftist party across countries). In some countries, the key issue of gender equality could be access to abortion, and the analyst would ask whether the leftist parties there do or do not address abortion. In other nations, the key issue of gender equality could be access to elementary school, access to the courts for partner violence, access to the labor market, access to finance capital, or access to professional jobs. In such a case, analyzing all nations with the same issue would distort the assessment of the theory. The distortion would occur because by studying a nonsalient dimension (e.g., Do girls in the United States have access to elementary schools?) one would risk missing evidence germane to the theory by focusing on an issue irrelevant, in that national context, to the theory. Institutional standardization in the context of studying diverse nations denies the possibility that the salient outcomes differ

depending on national context. Thus, not only will an informative cross-national analysis of EMI standardize on salience, not institutional location, but, indeed, *any* solid study of EMI must highlight salient outcomes, not necessarily the outcomes other researchers have studied in other temporal or national contexts.

Of course, if one has a narrow theory, then the scope is smaller and one must use institutional or partial standardization. If one's theory were "leftist parties are more likely to support abortion," then one would have to study each party's position on abortion. It would be wrong to analyze other issues (Do leftist parties support girls' entry to elementary school?) while ignoring the theoretically focal issue of abortion. In essence, with a narrow theory institutional standardization and salient standardization become one—abortion, salient or not in the specific context, *is* salient for the narrow theory, and thus must be the object of focus. However, with a general theory institutional standardization and salient standardization may differ. As EMI is a general theory, salient standardization is indicated.

Concluding Remarks

Historically, investigation of the relation between socioeconomic background and education outcomes has contributed multiple landmark studies. Work continues for many reasons, including new developments in statistical modeling, advances in technology that facilitate such research, and continuing social change. A result of the continued work is the development of theories of multiple dimensions of inequality. Effectively maintained inequality, as a 21st-century contribution, is dependent on each of those sources of continued work.

EMI differs from many sociological theories of inequality in an important and potentially illuminating way. Many sociological theories are tested by considering the sign of one or more regression or regression-like coefficients. However, because EMI posits the existence of a specific pattern connecting socioeconomic status and allocation to distinct, qualitatively different categorical outcomes, key aspects of EMI cannot be assessed with the typical test of the sign of a regression coefficient. Instead, one must search for the patterns EMI hypothesizes. This need implies other features analyses require in order to allow critical, appropriate test of the theory.

Those features address multiple problems that hound analyses of inequality. An ideal study of EMI's qualitative contention will select outcomes with meaningful variation across qualitatively distinct categories, will use an appropriate modeling framework on data for a population selected for its ability to allow clear assessment of socioeconomic factors, and will measure socioeconomic position in an expansive way. In specifying theoretically focal persons for comparison, the analysis will avoid the inclusion of methodologically and theoretically distractive control variables. Finally, analysts will calculate the appropriate predicted values to allow direct assessment of EMI's qualitative claim. In the modeling effort, and owing to EMI's focus and assessment procedure (i.e., calculation of predicted values for categorical positions), an ideal analysis of EMI will simultaneously address the challenges posed by selection bias, unobserved heterogeneity, and theoretical salience.

At its core, EMI posits that those possessing socioeconomic advantage will use that advantage to secure additional advantages for themselves and for their children, including in nonsocioeconomic domains. In a sense, EMI as a theory posits a Liebersonian (Lieberson, 1985) basic cause—advantaged actors' power to secure advantages. EMI encourages analysts to assess whether advantaged actors are successful in this endeavor even given the absence or diminishment of quantitative inequality. Nothing in EMI denies the potential importance of quantitative inequality. But EMI points analysts' attention to qualitative inequality simultaneously.

EMI's qualitative contention may not prove operative in every nation. It would be a sad world were EMI to exist everywhere. Inequality appears ubiquitous, but EMI may not be. Thus, EMI is not the only theory analysts need consider. Yet, if policies to further equality and opportunity are more likely to be successfully designed and effective upon implementation when design and implementation are ground in deep understanding, we submit that EMI is a general theory of inequality worthy of analyst, policy maker, and general public consideration.

Acknowledgments

We thank our EMI project participants—Megan Andrew, Hannah Ayalon, Paolo Barbieri, Aaron Benavot, Soo-yong Byun, Vikki Boliver, Sabrina Carrossa, Martina Dieckhoff, Yariv Feniger, Yvette Grelet, Martin Hällsten, Cristina Iannelli, Tomáš Katrňák, Irena Kogan, Aleksander Kucel, Noémie Le Donné, Gary N. Marks, Selina McCoy, Oded Mcdossi, Matthew McKeever, Clemens Noelke, Hyunjoon Park, David Reimer, Peter Robert, Stefan Schindler, Stefanie Scherer, Natalie Simonová, Emer Smyth, Sue Thompson, Florencia Torche, Louis-André Vallet, Rolf Van Der Velden, Montserrat Vilalta-Bufi, Christopher Whelan, and Felix Weiss—for several illuminating conversations; the EQUALSOC International Network of Excellence of the European Union 7th Framework Programme for project funding; Herman G. van de Werfhorst for useful conversations and support; and Jan Jacobs, Susan Schacht (posthumously), Wai Kit Choi, Aimée Dechter, H. Sorayya Carr, Olivia Garcia, and many Madison, Wisconsin analysts for helpful comments and consistent support.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: EQUALSOC International Network of Excellence of the European Union 7th Framework Programme.

Notes

1. We use *social* background to refer to all background factors (e.g., geographic location), reserving *socioeconomic* background for the socioeconomic subset (e.g., education, occupation).

2. The estimated thresholds could adjust instead, but only to the extent that the adjustment makes sense given the other variables in the model. Thus, even so, the fundamental issue—that, in reality, two coefficients are needed to capture the disparate associations—would remain.
3. Lucas (2001) calculated predicted values using each separate socioeconomic background indicator holding all others constant, but noted this approach treated indicators as statistically independent. The last predicted probability calculations (Lucas, 2001, pp. 1677-1678) treated socioeconomic background indicators as correlated. Because socioeconomic background indicators are correlated, this last comparison is the ideal comparison to make to test EMI.
4. If one estimates a stratified outcomes model and compares coefficients or predicted values across transitions, the analyst must assure those coefficients and predicted values are not contaminated or are minimally contaminated by unobserved heterogeneity.
5. We prefer the term *institutional standardization* to avoid confusion with the demographic techniques of *direct standardization* (Neison, 1845).

References

- Bailey, T., & Waldinger, R. (1991). Primary, secondary, and enclave labor markets: A training systems approach. *American Sociological Review*, *56*, 432-445.
- Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure*. New York, NY: Free Press.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.
- Boudon, R. (1974). *Education, opportunity, and social inequality: Changing prospects in western society*. New York, NY: Wiley.
- Brant, R. (1990). Assessing proportionality in the proportional odds model for ordinal logistic regression. *Biometrics*, *46*, 1171-1178.
- Breen, R., & Jonsson, J. O. (2000). Analyzing educational careers: A multinomial transition model. *American Sociological Review*, *65*, 754-772.
- Byrne, D., McGinnity, F., Smyth, E., & Darmody, M. (2010). Immigration and school composition in Ireland. *Irish Educational Studies*, *29*, 271-288.
- Cameron, S. V., & Heckman, J. J. (1998). Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males. *Journal of Political Economy*, *106*, 262-333.
- Card, D. (1999). The causal effect of education on earnings. In O. C. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 3, pp. 1801-1863). Amsterdam, Netherlands: Elsevier.
- Darmody, M., Byrne, D., & McGinnity, F. (2014). Cumulative disadvantage? Educational careers of migrant students in Irish secondary schools. *Race, Ethnicity and Education*, *17*, 129-151.
- De Graaf, P. M., & Ganzeboom, H. B. G. (1993). Family background and educational attainment in the Netherlands for the 1891-1960 birth cohorts. In Y. Shavit & H.-P. Blossfeld (Eds.), *Persistent inequality: Changing educational attainment in thirteen countries* (pp. 75-99). Boulder, CO: Westview Press.
- Dougherty, K. J. (2001). *The contradictory college: The conflicting origins, impacts, and futures of the community college*. Albany: State University of New York Press.
- Duncan, B. (1967). Education and social background. *American Journal of Sociology*, *72*, 362-373.
- Duncan, O. D. (1966). Path analysis: Sociological examples. *American Journal of Sociology*, *72*, 1-16.

- Elwert, F. (2013). Graphical causal models. In S. L. Morgan (Ed.), *Handbook of causal analysis for social research* (pp. 245-273). New York, NY: Springer.
- Feliciano, C. (2005). Educational selectivity in U.S. immigration: How do immigrants compare to those left behind? *Demography*, 42, 131-152.
- Gamoran, A. (1987). The stratification of high school learning opportunities. *Sociology of Education*, 60, 135-155.
- Goodman, L. A. (1979). Simple Models for the analysis of association in cross-classifications having ordered categories. *Journal of the American Statistical Association*, 74, 537-552.
- Hauser, R. M. (1976). Review: On Boudon's model of social mobility. *American Journal of Sociology*, 81, 911-928.
- Hauser, R. M., & Andrew, M. (2006). Another look at the stratification of educational transitions: the logistic response model with partial proportionality constraints. *Sociological Methodology*, 36, 1-26.
- Hauser, R. M., & Goldberger, A. S. (1971). The treatment of unobservable variables in path analysis. *Sociological Methodology*, 3, 81-117.
- Hauser, R. M., Sewell, W., & Alwin, D. (1976). High school effects on achievement. In W. Sewell, R. M. Hauser, & D. Featherman (Eds.), *Schooling and achievement in American society* (pp. 309-341). New York, NY: Academic Press.
- Hauser, R. M., Tsai, S.-L., & Sewell, W. (1983). A model of stratification with response error in social and psychological variables. *Sociology of Education*, 56, 20-46.
- Hauser, R. M., & Warren, J. R. (1997). Socioeconomic indexes for occupations: A review, update, and critique. *Sociological Methodology*, 27, 177-298.
- Hilbe, J. M. (2009). *Logistic regression models*. New York, NY: Chapman & Hall/CRC.
- Holm, A., & Jæger, M. M. (2011). Dealing with selection bias in educational transition models: The bivariate probit selection model. *Research in Social Stratification and Mobility*, 29, 311-322.
- Iyengar, S. S., & Lepper, M. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 76, 995-1006.
- Jackson, P. B., Thoits, P. A., & Taylor, H. F. (1995). Composition of the workplace and psychological well-being: The effects of tokenism on America's Black elite. *Social Forces*, 74, 543-557.
- Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85-112). New York, NY: Seminar Press.
- Kanter, R. M. (1977). Some effects of proportions on group life: Skewed sex ratios and responses to token women. *American Journal of Sociology*, 82, 965-990.
- Kenny, D. A., & Judd, C. M. (1984). Estimating the nonlinear and interactive effects of latent variables. *Psychological Bulletin*, 96, 201-210.
- Kohn, M. L. (1963). Social class and parent-child relationships: An interpretation. *American Journal of Sociology*, 68, 471-480.
- Lareau, A. (2002). Invisible inequality: Social class and childrearing in Black families and White families. *American Sociological Review*, 67, 747-776.
- Lieberson, S. (1985). *Making it count: The improvement of social research and theory*. Berkeley: University of California Press.
- Long, J. S. (1997). *Regression models for categorical and limited dependent variables*. Thousand Oaks, CA: Sage.
- Long, J. S., & Freese, J. (2006). *Regression models for categorical variables using Stata* (2nd ed.). College Station, TX: Stata Press.

- Lucas, S. R. (1999). *Tracking inequality: Stratification and mobility in American high schools*. New York, NY: Teachers College Press.
- Lucas, S. R. (2001). Effectively maintained inequality: Education transitions, track mobility, and social background effects. *American Journal of Sociology*, *106*, 1642-1690.
- Lucas, S. R. (2009). Stratification theory, socioeconomic background, and educational attainment: A formal analysis. *Rationality & Society*, *21*, 459-511.
- Lucas, S. R. (2017). An Archaeology of Effectively Maintained Inequality Theory. *American Behavioral Scientist*, *61*(1), 8-29.
- Lucas, S. R., Fucella, P., & Berends, M. (2001, August). *Neo-classical education transitions of boomers and post-boomers in the United States: A corrected tale for three cohorts*. Presentation at the International Sociological Association Research Committee No. 28 Meeting, Berkeley, CA.
- Lucas, S. R., Fucella, P. N., & Berends, M. (2011). A neo-classical education transitions approach: A corrected tale for three cohorts. *Research in Social Stratification and Mobility*, *29*, 263-285.
- Mare, R. D. (1980). Socioeconomic background and school continuation decisions. *Journal of the American Statistical Association*, *75*, 295-305.
- Mare, R. D. (1981). Change and stability in educational stratification. *American Sociological Review*, *46*, 72-87.
- Mare, R. D. (1993). Educational stratification on observed and unobserved components of family background. In Y. Shavit & H.-P. Blossfeld (Eds.), *Persistent inequality: Changing Educational attainment in thirteen countries* (pp. 351-376). Boulder, CO: Westview Press.
- Meyer, J. W., Ramirez, F. O., Rubinson, R., & Boli-Bennett, J. (1977). The world educational revolution, 1950-1970. *Sociology of Education*, *50*, 242-258.
- Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: Methods and principles for social research*. New York, NY: Cambridge University Press.
- Muthén, B. O. (1979). A structural probit model with latent variables. *Journal of the American Statistical Association*, *74*, 807-811.
- Muthén, B. O. (1983). Latent variable structural equation modeling with categorical data. *Journal of Econometrics*, *22*, 48-65.
- Neison, F. G. P. (1845). Contributions to vital statistics, especially designed to elucidate the rate of mortality, the laws of sickness, and the influences of trade and locality on health, derived from an extensive collection of original data, supplied by friendly societies, and proving their too frequent instability. *Journal of the Statistical Society of London*, *8*, 290-350.
- Nerlove, M., & Press, S. J. (1973). *Univariate and multivariate log-linear and logistic models* (R-1306-EDA/NIH). Santa Monica, CA: RAND.
- Pearl, J. (2010). The foundations of causal inference. *Sociological Methodology*, *40*, 75-149.
- Rose, H., & Betts, J. R. (2004). The effect of high school courses on earnings. *Review of Economics and Statistics*, *86*, 497-513.
- Schaller, R. P. (1997). Moore's law: past, present and future. *IEEE Spectrum*, *34*(June), 52-59.
- Sewell, W. H., Haller, A. O., & Ohlendorf, G. W. (1970). The educational and early occupational status attainment process: Replication and revision. *American Sociological Review*, *35*, 1014-1027.
- Sewell, W. H., Haller, A. O., & Portes, A. (1969). The educational and early occupational attainment process. *American Sociological Review*, *34*, 82-92.
- Shavit, Y., & Blossfeld, H.-P. (1993). *Persistent inequality: Changing educational attainment in thirteen countries*. Boulder, CO: Westview Press.

- Tam, T. (2011). Accounting for dynamic selection bias in educational transitions: The Cameron–Heckman latent class estimator and its generalization. *Research in Social Stratification and Mobility*, 29, 287-310.
- Trow, M. (1961). The second transformation of American secondary education. *International Journal of Comparative Sociology*, 2, 144-166.
- Tversky, A., & Shafir, E. (1992). Choice under conflict. *Psychological Science*, 3, 358-362.
- Vermunt, J. K. (1997). *Log-linear models for event histories*. Thousand Oaks, CA: Sage.
- Walter, C. (2005). Kryder's Law. *Scientific American*, 293, 32-33.
- Williams, R. (2006). Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal*, 6, 58-82.
- Wright, S. (1934). The method of path co-efficients. *Annals of Mathematical Statistics*, 5, 161-215.

Author Biographies

Samuel R. Lucas is Professor of Sociology at the University of California-Berkeley, with research interests in stratification, research methods, sociology of education, and statistics. His recent works include "Where the Rubber Meets the Road," an analysis of the epistemological justifications for case selection strategies across multiple data collection methods. He is currently a deputy editor of the *American Sociological Review*.

Delma Byrne is a Lecturer in Sociology of Education and Social Inequality at Maynooth University Departments of Sociology and Education. Her main research interests focus on social stratification in education and the labour market, with a focus on social class, migration, race/ethnicity, gender and disability.