CHAPTER ONE

Introduction

How Is Inequality of Educational Opportunity Generated? The Case for Primary and Secondary Effects

Michelle Jackson

When sociologists write about inequalities in educational attainment, they frequently get under way by emphasizing the extraordinary transformations of educational institutions over the course of the 20th century. And indeed, it is hard to imagine how to open a volume on inequalities in educational attainment without acknowledging the significant educational expansion and reform in all Western societies over the last century. The basic features of expansion and reform are well known and may be summarized as comprising three fundamental steps: the establishment of near-universal primary education at the beginning of the 20th century, the rise of near-universal secondary education toward the middle of the century, and the (as yet unfinished) development of a system of mass higher education toward the end of the century. An important consequence of this expansion has been an increase in the average level of educational attainment, such that most of the students who enter secondary education today can expect to obtain a tertiary-level qualification by the end of their educational career. Alongside increasing average levels of attainment, we also observe increasing differentiation in educational systems, so that students may choose from a range of academic and vocational courses in many diverse specialist fields.

The development of educational systems can be understood principally as a response to the demands of changing economic and occupational structures but also as an attempt to create a greater equality of educational opportunity. Yet a great deal of research has demonstrated that significant inequalities in educational attainment between members of different social groups remain. One important area of research focuses on social-class inequalities in educational attainment; children of professional or managerial

background generally achieve higher levels of educational performance and make more ambitious educational choices than do children from working-class backgrounds (e.g., Shavit and Blossfeld 1993; Breen et al. 2009).

Arguably of more interest than the current state of class inequalities in educational attainment is the question of how far these inequalities have changed over time. *Persistent Inequality* (Shavit and Blossfeld 1993) argues that there has been a relatively high degree of temporal stability in the association between class origin and educational attainment. More recent work suggests that a trend toward a weakening association between class origin and educational attainment is present in many European countries, particularly if changes over a relatively long period are considered (Jonsson, Mills, and Müller 1996; Vallet 2004; Breen and Jonsson 2005; Breen et al. 2009, 2010). But this observation should not lead us to lose sight of the following: even if weaker now than in the past, class inequalities in educational attainment remain as a feature of modern societies and this feature is likely to linger for some time. The durability of these inequalities is particularly striking when compared to the far more substantial changes in gender, ethnic, and racial inequalities observed in many countries.

In this volume we aim to understand why social-background inequalities in educational attainment, or inequality in educational opportunity (IEO),[1] should exist and persist in eight Western countries. Should IEO be understood as a consequence of differences in academic ability and performance between members of different social classes? Or should it be understood as a consequence of differences in the educational decisions made by members of different social classes, such that students from advantaged backgrounds choose higher levels of education more frequently than students from disadvantaged backgrounds, regardless of their academic performance? These basic questions outline extreme positions on how IEO is created. In this volume we consider social-background inequalities in educational attainment to be a consequence of *both* social-background differences in academic performance and social-background differences in the choices that students make, holding performance constant. Our main aim is to determine the relative importance of these two features in creating IEO. Insofar as changes in IEO are observed, we ask whether they can be attributed to changes in the relationship between class and performance or to changes in the class-biased choices that are made, conditional on perfor-

mance. If we observe declines in one or both effects, leading to a decline in IEO, this provides us with important evidence about which policies and institutional innovations hold most promise for further reducing class effects.

IEO is a term that carries some ambiguity in that it can refer either to a summary measure of *all* inequalities related to social background generated by an educational system or to only the social-background inequalities generated *at a given transition*. For example, Boudon defines IEO as meaning "differences in level of educational attainment according to social background" (1974, xi) but also states that "IEO rates are subject to variations as a function of national context, point in time, and school level. . . . [A] certain amount of IEO is present . . . at each school level" (1974, 41). In this volume we are concerned with IEO at given educational transitions, and we therefore take the latter understanding of IEO to be our own. In each chapter IEO is discussed in relation to the transition under consideration and to the risk set of students eligible for that transition. On the whole, we do not address the question of how far IEO assessed for different transitions and risk sets accumulates to a summary measure of all IEO generated by an educational system.[2]

DEFINING PRIMARY AND SECONDARY EFFECTS

Our understanding of IEO has at its heart an individual-level model, in which a student achieves a certain level of academic performance and then makes a decision about how to proceed in the educational system. The decision that students make when faced with an educational transition is shaped by their previous academic performance, which provides information about the likelihood of successful completion of higher levels of education. But the decision is also influenced by factors other than previous academic performance, because a student takes into account the costs and benefits of the different choices that might be made in relation to the transition.

The decomposition of IEO into a part determined by differences in previous performance across social groups and a part determined by the choices made by members of those groups is well established in the literature, in which performance effects are labeled primary effects and choice effects are labeled secondary effects[3] (Girard and Bastide 1963; Boudon 1974). As Breen and Goldthorpe write,

> Primary effects are all those that are expressed in the association that exists between children's class origins and their average levels of demonstrated academic ability. Children of more advantaged backgrounds . . . perform better, on average, than children of less advantaged backgrounds in standard tests, examinations, and so on. . . . [S]econdary effects . . . are effects that are expressed in the actual choices that children, together perhaps with their parents, make in the course of their careers within the educational system—including the choice of exit. (1997, 277)

The history of the conceptual distinction between primary and secondary effects is discussed in more detail below.

Before proceeding any further, an apology on the matter of terminology is in order. While the concepts of primary and secondary effects can be defined with precision, the labels attached to the concepts are unfortunately rather ambiguous. Although one can understand the rationale behind the labels once the concepts have been defined, in that social background *primarily* affects performance and then *secondarily* affects choices (conditional on performance), there are few clues to the meanings of these labels for the uninitiated. It is also curious that these terms should be used in a field concerned with education systems, in which "primary" and "secondary" are understood first and foremost to refer to different phases of the school career or school system, not to the decomposition of IEO presented here. Despite the problems with the terminology of primary and secondary effects, the terms are now fully established in the literature on IEO, and in this volume we continue to use these labels. We also refer to primary and secondary effects as "performance" and "choice" effects, so as to further emphasize their meanings. To avoid confusion, we largely avoid using "primary" and "secondary" in isolation to refer to the school career.

THE UTILITY OF PRIMARY AND SECONDARY EFFECTS

There are clear advantages to treating IEO as the overall consequence of the operation of primary and secondary effects: the concepts allow sociologists to gain greater precision in identifying the determinants of IEO, and they also have obvious implications for social policy.

As discussed below, primary and secondary effects are likely to be generated by different processes, and as a consequence, the explanatory tools needed to explain the primary effects of differences in performance

between social groups differ from the tools needed to explain the secondary effects of the differences in choices, conditional on performance (e.g., Erikson and Jonsson 1996a). Therefore, by determining to what extent IEO should be understood as resulting from primary or secondary effects, sociologists obtain valuable information about where explanatory effort should be directed. If the distinction between primary and secondary effects is ignored, sociologists are in danger of attempting to explain IEO as an unwieldy whole, thus mistaking a dual phenomenon requiring distinct and separate explanations for a single composite requiring a single explanation.

Aside from the utility of primary and secondary effects for explaining IEO, the concepts are highly relevant to policy. Just as the explanatory tools differ depending on whether the focus is on primary or on secondary effects, the appropriate policy interventions will also differ depending on which type of effect is to be addressed. If differences in *performance* are the main drivers of educational inequality, policies aimed at reducing those differences will have a very large impact in reducing overall inequality. Early interventions, such as intensive preschool education or economic support for families with young children, seem to have great potential for reducing primary effects (see Cameron and Heckman 1999; Carneiro and Heckman 2003; Heckman 2006). On the other hand, if differences in the *choices* made by students at the same level of performance have a significant impact, policies aimed at changing constraints and incentives hold more promise (Jackson et al. 2007).

PRIMARY AND SECONDARY EFFECTS

History of the Concepts

The distinction between primary and secondary effects has a long and colorful history. It is a history marked principally by unrealized promise, a state of affairs that can perhaps be attributed to the lack of a rigorous operationalization of the concepts of primary and secondary effects but also to other idiosyncratic features that had a further suppressive effect.

Where to begin with this history? The distinction between primary and secondary effects is most commonly associated with Boudon's 1974 book *Education, Opportunity, and Social Inequality*. However, while Boudon's

work is clearly the most extensive treatment of the roles of primary and secondary effects in creating educational inequality, the roots of these concepts are elsewhere. In the 1940s Boalt carried out an empirical analysis of social-class inequalities in educational attainment in Sweden, in which he considered the roles of previous performance and class background in the transition to secondary school (Boalt 1947; summarized in English in Boalt and Janson 1953). Calculating partial correlations between class, school performance, and school selection, on data describing all 5,000 students in primary school in Stockholm in 1936, Boalt concludes that

> "social class" measured by father's occupation gave a significant correlation with marks in primary schools (0.32) but a high correlation with selection to secondary schools (0.56). This latter correlation remained high even when partial correlations for marks, income, and the factor "known to social welfare authorities," were worked out. (Boalt and Janson 1953, 323)

Boalt clearly distinguishes between the effects of social-class background on performance and its effects on the transition to secondary school, conditional on performance. Here we find, therefore, a distinction between primary and secondary effects in all but name.

It is not apparent that Boalt's distinction was widely applied in the educational research of the time, and indeed the 1947 book and 1953 paper have been cited only 59 times.[4] However, Boalt's work has much in common with that of his contemporaries, in particular, with research in the 1950s and 1960s on the "reserve of talent." The talent reserve referred to "how many young people . . . with sufficient ability did for various reasons *not* proceed to upper secondary and university education" (Husén 1974, ix). In this literature, a concern about class inequalities in educational participation rates was expressed as concern about wastage of talent, in that there must be working-class children who were not realizing their full potential. Several studies aimed to assess the size of the talent reserve in relation to class origin, and these studies considered the gap between more advantaged and less advantaged students to be a consequence of differences in ability (or performance) between classes and differences in transition propensities conditional on ability (or performance) (see, e.g., Anderberg 1948; Ekman 1951; Wolfe 1954; Härnqvist 1958; Härnqvist 2003 provides an excellent summary of some of this research). Again, in all but name, these studies distinguished between primary and secondary effects in determining class

inequalities in educational attainment. Indeed, the expectancy method, applied by Gösta Ekman (1951), in which the proportion of students from one class making a transition is compared with the proportion of students from another class making a transition for given ability levels, is a precursor of the methods applied in this volume.

In the 1960s the distinction that Boalt had recognized was once again applied in sociological research, although this application was seemingly independent of Boalt's work. In France, Girard and Bastide published two papers in the journal *Population* that discussed the effect of social background on performance and its effect on transitions conditional on performance (Girard and Bastide 1963; Girard, Bastide, and Pourcher 1963). In an analysis of a nationally representative sample of French schoolchildren facing the transition to secondary education, Girard and his colleagues demonstrated that students' grades were associated with their social origins, such that around 55 percent of students originating in the most advantaged social class achieved "good" or "excellent" grades while only 29 percent of students originating in the least advantaged class achieved equivalent grades. Social-class origins were also found to affect the chances of making the transition to secondary education, even after taking into account these differences in performance. So among the students with good or excellent grades, almost all of those from the most advantaged class made the transition to secondary education, while only 80 percent of those from the least advantaged class made the transition.

The work of Girard and his colleagues offered evidence that social-class inequalities in educational attainment were driven by both class differences in performance and class differences in transition propensities, given performance. But this research is probably more important for a theoretical than an empirical contribution: although not generally acknowledged, Girard and Bastide (1963) were the first to introduce the language of primary and secondary effects to the analysis of educational inequality.[5] They write, "C'est là la première cause de la non-démocratisation: l'influence du milieu familial sur le développement de l'enfant et, par suite, sa réussite scolaire" (437), and "C'est là la seconde cause de la non-démocratisation: même à égalité de notes, la chance pour l'enfant d'entrer en sixième est en relation avec sa condition sociale" (439).[6] In these two passages, Girard and Bastide describe a decomposition of overall educational inequality into a part determined by performance differences and a part determined by differences

in choices, conditional on performance, and also identify these features as primary and secondary causes of the inequality.

Despite the relatively long history of primary and secondary effects in sociology, it was not until Boudon's 1974 book, *Education, Opportunity, and Social Inequality*, that these concepts received an extensive sociological treatment. Boudon's book put forward a general theory designed to explain why IEO and inequality of social opportunity (ISO, or social-class immobility) should exist and persist in modern, industrialized societies. The distinction between primary and secondary effects is part of the general theory of IEO.

Boudon argues that a puzzling feature of modern societies with well-developed educational systems is that children of less advantaged social background choose to acquire less education, on average, than children of more advantaged social background, even though these children must know that their chances in the labor market would be substantially improved if they obtained more education rather than less. He argues that this is a consequence of two different processes, which he labels as primary and secondary effects. His treatment of these effects is very similar to Girard and Bastide's, in that primary effects are seen to be the effects of social background on performance and secondary effects are seen to be the effects of social background on educational choices, conditional on performance. The existence of secondary effects, for Boudon, results from the differential costs and benefits attached to educational decisions for students from different social classes. In sum, he writes that

> IEO is generated by a two-component process. One component is related mainly to the cultural effects of the stratification system. The other introduces the assumption that even with other factors being equal, people will make different choices according to their position in the stratification system. In other words, it is assumed (1) that people behave rationally in the economic sense of this concept . . . but that (2) they also behave within decisional fields whose parameters are a function of their position in the stratification system. (1974, 36)

After the publication of Boudon's book, the distinction between primary and secondary effects was firmly established in the literature on IEO, although as Jackson and colleagues note, these concepts were "surprisingly neglected" by sociologists of education (2007, 212; see also Nash 2006). While many researchers discussed the primary-versus-secondary-effects distinction, no rigorous method was available for operationalizing the

concepts and few shaped their empirical analyses around the distinction. But there are also two historically idiosyncratic reasons why the concepts were not more influential.

First, as Stephen Morgan, Michael Spiller, and Jennifer Todd note in this volume (Chapter 10), a crushing review of Boudon's book by Hauser (1976) in the *American Journal of Sociology* is unlikely to have encouraged scholars to take the distinction to heart.[7] Hauser's criticisms focused on Boudon's overarching theories of IEO and ISO and the tests of these theories (although because Hauser's review dealt with the whole of Boudon's book, many criticisms do not in fact relate to the conceptual distinction between primary and secondary effects).

Second, to some extent the primary-versus-secondary-effects distinction has been subsumed in the wider literature of rational choice theories in education. Many of the influential rational choice models of education, in which educational decisions are related to the costs, benefits, and expected probabilities of success of different courses of action, make the distinction between primary and secondary effects, either implicitly or explicitly. In the Breen-Goldthorpe model (1997), the distinction is acknowledged explicitly:

> We assume, to begin with, that class differentials in educational attainment come about through the operation of two different kinds of effect which, following Boudon (1974), we label as "primary" and "secondary." . . . Some educational choices may of course be precluded to some children through the operation of primary effects. . . . But, typically, a set of other choices remains, and it is further known that the overall patterns of choice that are made are in themselves—over and above primary effects—an important source of class differentials in attainment. (277)

Other rational choice models of educational decisions might not be linked to the primary-versus-secondary-effects distinction as explicitly as the Breen-Goldthorpe model is, but performance and choice effects are understood to have separate roles in determining the main components of many of these models (e.g., Erikson and Jonsson 1996a).

Aside from these two relatively particularistic explanations for why the distinction between primary and secondary effects has been underapplied in the sociology of education, another, more straightforward reason explains why the distinction did not achieve instant or wholesale adoption. The distinction between primary and secondary effects is most useful when applied to educational transitions: given a transition point, we wish

to understand the separate roles of performance and choice in determining whether a student makes the transition. While western European systems are transition-based systems, in other school systems transitions are less easily identified, which limits the applicability of the distinction. For example, in the United States, the first major educational branching point is not a transition but instead relates to whether students complete the high school degree. In such a system, where years of schooling or completion of educational levels are treated as the appropriate components of educational systems, a transition-based analysis might not be seen as the most natural analytic approach.

RECENT DEVELOPMENTS

For all the foregoing reasons, the distinction between primary and secondary effects was to a certain extent neglected in the literature on IEO, yet recent years have seen a resurgence of interest in these concepts. In part this resurgence has been driven by methodological developments that allow researchers to assess the relative importance of primary and secondary effects in creating inequalities in educational attainment.

In the empirical literature on the talent reserve in the 1950s, researchers were concerned with estimating the size of the reserve of talent by comparing ability distributions across classes and comparing the transition rates for those same classes. In Ekman's expectancy method, as described in Härnqvist (1958, 2003), class differences in transition probabilities were calculated for different levels of performance, enabling the researchers to show that students from the highest class were more likely to make the transition to upper secondary school than students from the lower classes at all levels of academic performance (Härnqvist 2003, 487–89; see Duru-Bellat, Jarousse, and Mingat 1993 for a similar approach using French data). In a somewhat similar fashion, Erikson and Jonsson (1996b) observed that the primary and secondary effects of class origin could be represented graphically; Figure 1.1 shows how grade point average (GPA) distributions and transition propensities for two different social classes would be presented under this approach. First, to represent primary effects, they presented class-specific GPA distributions, to demonstrate that students from the highest class (EGP I) had higher average GPAs than students from the lowest class (EGP VII). Second, to represent secondary effects, they plotted

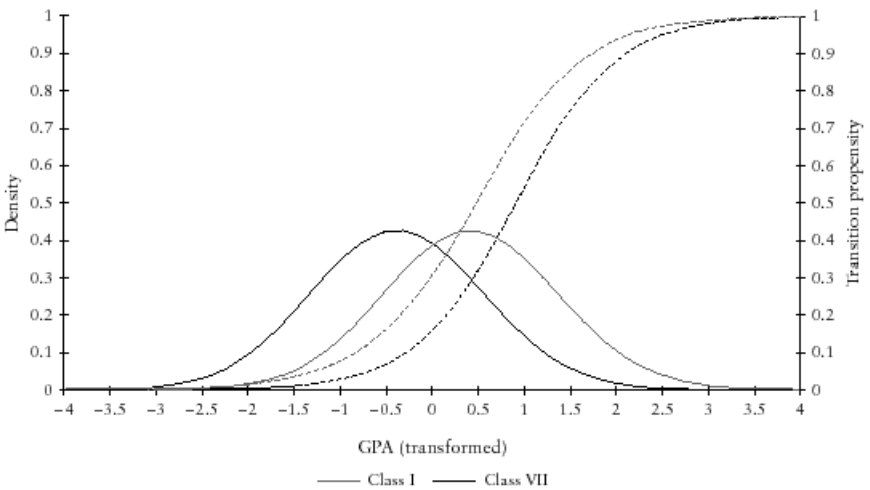

Figure 1.1. Grade point average (GPA) and transition propensities for two social classes

NOTE: Adapted from Erikson and Jonsson (1996b). The curves shown here are for illustrative purposes only and do not exactly correspond to the results reported in Erikson and Jonsson (1996b).

separate curves for each class showing the probability of making the transition (to upper secondary school) at every level of GPA (1996b, 76–77). Through integration, the two sets of curves could be used to estimate the total proportion of students making the transition for each class, how these proportions would be affected if arbitrary grade limits were defined, and the subsequent effect on IEO.

The Erikson-Jonsson approach was further developed in work by Erikson et al. (2005) and Jackson et al. (2007), in which class-specific performance distributions and transition probabilities were manipulated to obtain a quantitative estimate of the contributions of primary and secondary effects to IEO. This method will be applied throughout the volume, and it is described in detail and extended in Chapter 2. In essence, the method decomposes the odds ratios describing inequalities between social groups into a part due to primary effects and a part due to secondary effects; a percentage estimate of the relative importance of the two effects in creating IEO can thus be derived.