

How should we measure the effect of ability grouping on student performance?

D.I. Rees^{a,*}, D.J. Brewer^b, L.M. Argys^a

^a Department of Economics, University of Colorado at Denver, Denver, Campus Box 181, P.O. Box 173364, Denver, CO 80217-3364, USA

^b RAND, 1700 Main Street, Santa Monica, CA 90401, USA

Received 1 February 1998; accepted 15 July 1998

Abstract

In this volume, Betts and Shkolnik [Betts, J. R., & Shkolnik, J. L. (1999) The effects of ability grouping on student math achievement and resource allocation in secondary schools. *Economics of Education Review*, 19, 1–15] argue that studies that compare students in tracked versus untracked classes overestimate the impact of tracking on student achievement by not adequately controlling for student ability and motivation. In this paper we discuss the shortcomings of their analysis and reinterpret their results. The data used by Betts and Shkolnik do not allow one to accurately classify tracked and untracked classroom, since identification of heterogeneous classes is impossible. They compare ability-grouped students in schools that report formally engaging in tracking to ability-grouped students in schools that track only informally. Our interpretation of their results suggests that there is little difference in student performance and resource allocation between schools that formally and informally group students by ability. [JEL I20] © 1999 Published by Elsevier Science Ltd. All rights reserved.

Keywords: Resource allocation; Educational economics; Efficiency

1. Introduction

There is a great deal of controversy surrounding the practice of ability grouping or tracking in America's schools. On the one hand, the education establishment appears to consider tracking as unambiguously detrimental to student learning. For instance, the president of Teachers College at Columbia University recently told the New York Times that research showed that putting students of varying abilities in the same class does not hurt fast learners, but can help slow learners (Sengupta, 1997). In keeping with this view, there has been considerable movement towards eliminating the practice of grouping students according to ability.¹

On the other hand, a number of recently published articles have called into question this conventional wisdom. Work by Argys, Rees and Brewer (1996) and Hoffer (1992), suggests that tracking creates winners and losers. Students placed in upper-track classes benefit in terms of achievement with higher scores on standardized tests, but students placed in lower-track classes are clearly hurt.

The latest contribution to the debate over tracking by Betts and Shkolnik (1999) [henceforth "B and S"] is another important addition to the literature which challenges the conventional wisdom. They find that tracking helps some students but harms others. However, they argue that prior work failed to adequately control for unmeasured factors such as motivation, and therefore

* Corresponding author. E-mail: drees@carbon.cudenver.edu

¹ For example, data from the *Schools and Staffing Survey* suggest that 20% of school with programs for gifted children in

1990 had eliminated the programs by 1993 (Figlio & Page, 1998).

overestimated the magnitude of these differential effects. In contrast, we believe that B and S have interpreted their results incorrectly.

2. Student comparisons and the effect of tracking

In their empirical work, B and S compared students in high, middle and low ability classes at schools that tracked, with students in high, middle and low ability classes at schools that supposedly did not track. While this is an interesting approach, it does not strike us as the correct comparison; indeed it is rather like comparing apples to apples and finding little difference.

B and S determined whether a school practiced tracking based on an assessment made by the school principal. The level of the class (i.e. whether the class was composed of high, middle or low ability students) was based on answers to a teacher questionnaire. B and S investigated the effects of tracking by comparing the achievement of students in classes of similar composition across the two types of schools.

Previous researchers were interested in a completely different comparison. If a school principal claims that his or her school does not track and yet teachers in that same school can divide their classes based on the average abilities of their students, then clearly there is informal tracking taking place. Thus, the achievement differential estimated by B and S should be interpreted as the effect of formal versus informal tracking, and is *conceptually distinct* from the effect of placement in a class composed of high ability students versus placement in a heterogeneous class. Therefore, statements such as, “although the paper does find some evidence that grouping has differential effects across students of differing ability levels ... the sizes of these effects are shown to be far smaller than previous estimates” (p. 28) are misleading. The fact that formal and informal tracking have similar effects has no bearing on previous results. In order to make this study comparable to previous studies the authors would have to have been able to observe students in a heterogeneous environment. Yes, B and S identify students in schools that, according to their principals, are not tracked, but teachers in these same schools are distinguishing between five categories of classes, ranging from classes composed of students of “much higher than average” ability to those composed of students of “much lower than average” ability. Call it informal, but when a teacher is in front of a class composed of students whose academic ability is much lower than average, that is tracking in the traditional sense of the word.

Hoffer (1992) clearly considered informal and formal tracking to be equivalent. He wrote, “... past experience has shown that some schools that in fact have ability grouping do not have a formal policy of grouping: students are informally encouraged to take or discouraged

from taking certain classes by guidance counselors, teachers, and peers, even while admission to all classes are formally open to all” (p. 212). In his analysis, Hoffer made sure to classify students in these schools in the same fashion as he did students in schools with formal tracking policies. In order to accomplish this classification, he turned to information collected from, “teachers, school documents, and—when the teacher reports and documents proved ambiguous or contradictory—phone calls to key informants” (p. 212). Hoffer, in fact, found that only 15% of the seventh grade math students in his sample were in schools that did not track, whereas B and S, using the same data source (*The Longitudinal Study of American Youth or LSAY*) found that fully 27 percent of their sample were in schools that did not track.² Obviously, relying simply on the assessments of principals understates the amount of tracking in the data. This problem alone would tend to bias any estimates of the effect of tracking toward zero. In addition, because teachers were not given the option of categorizing their class as heterogeneous, B and S are unable to utilize a more suitable measure of tracking.³

In our 1996 analysis we relied on teacher assessments as to the composition of a particular class. This method too has its drawbacks.⁴ First, teachers will not all use the same criteria in defining classes of different levels. Second, as pointed out by Loveless (1997), classes that seem heterogeneous to a teacher may lack students from the extremes of the ability distribution.⁵ The answer to these issues is better data. Ideally, researchers would be able to access test scores from all students in each class in a school, and in a large number of schools. Tracking information would be gathered from multiple survey sources—district and school administrators, teachers, and students—and perhaps supplemented by school administrative records, and on-site observation. Only then can we be sure that we are really measuring the effect of tracking or detracking on student outcomes.

² It should be noted that the B and S sample included students from grades 7 through 12. Because tracking has been shown to be more prevalent in higher grades the discrepancy between these figures is especially worrisome.

³ B and S state that teachers would naturally classify heterogeneous classes as “average”. But it is unclear to us how a teacher with a heterogeneous class would respond to the question, “how would you rate the average academic ability of the students in this class ...? Much higher than average, somewhat higher, about average, somewhat lower, much lower than average?”

⁴ These are discussed in more detail in Brewer, Rees and Argy (1995).

⁵ A priori, it is not clear to us what the effect might be of including these students in the heterogeneous group.

3. Resources and tracking

B and S devote a considerable portion of their paper to showing the effect of ability grouping on resource allocation. They find that neither class size nor teacher characteristics vary much according to whether a student is in a tracked or detracked environment. This finding is counter to claims made by Oakes (1990, 1992) and others that tracking leads to the unequal structuring of resources across classes such that upper-level classes receive the better teachers, smaller classes, computers, etc.

However, the B and S analysis again utilizes a tracking measure based on the principal's assessment as to whether the school sorts its math students into classes according to ability. In effect, they simply found that schools with and without formal tracking policies have similar patterns of resource allocation.

Using a sample of tenth graders drawn from the *National Educational Longitudinal Study* (NELS) (National Center for Education Statistics, 1992), we re-examined the relationship between tracking and resources. Table 1 shows the mean value of selected resources for four subjects by class track.⁶ The first row of the table suggests that average class size is actually lower in classes composed of below average students in all subjects; e.g. for English, the mean class size is 24.5 in above average and average classes, but is 20.0 in below average classes. This result, then, does not support the common contention of detracking proponents that less able students are consigned to larger classes. However, there is some evidence to support their position with respect to other resources.

For example, while teachers appear to be about equally experienced across tracks in English and history, there is a sizeable gap (0.7 years in each case) in mean years of total experience between above and below average mathematics and science classes. Similarly, for these subjects in particular, teachers of high-ability classes are much more likely to have undergraduate and graduate majors in mathematics and science than those teachers assigned to low ability classes: 77.8% (82.3%) of teachers of above average classes have undergraduate degree majors in mathematics (science) compared to just 67.0% (65.2%) in below average classes. The differences are again less pronounced in English and history. There is far less variation across tracks in terms of whether the teacher is certified or whether the teacher has a Masters

⁶ Class track was determined from teacher responses to the question: "[w]hich of the following best describes the achievement level of the 10th graders in this class compared with the average 10th grade student in the school? Higher achievement levels, average achievement levels, lower achievement levels, or widely differing achievement levels".

Table 1
Resources by track for sample of 10th graders from NELS

History	Above (628)	Average (1066)	Below (388)	Heterogenous (427)
Mean class size	25.9	25.4	22.4	25.1
Mean experience	5.8	5.6	5.8	5.3
% Masters degree	45.6	50.8	50.4	51.1
% Undergraduate history major	84.3	77.2	76.9	82.8
% Graduate history major	44.8	32.9	30.6	28.7
% Certified in history	99.7	99.4	98.2	99.8
% Part-time	0.1	1.0	0.0	0.2
Science	Above (1053)	Average (1659)	Below (625)	Heterogenous (396)
Mean class size	23.7	24.2	21.5	24.6
Mean experience	5.7	5.3	5.0	4.7
% Masters degree	55.8	55.0	54.9	44.8
% Undergraduate science major	82.3	75.9	65.2	65.1
% Graduate science major	36.6	35.0	31.8	25.4
% Certified in science	100	99.7	97.9	99.7
% Part-time	0.8	1.5	0.2	2.5
Math	Above (1203)	Average (1785)	Below (1040)	Heterogenous (518)
Mean class size	23.5	24.2	21.6	24.0
Mean experience	6.0	5.4	5.3	5.2
% Masters degree	57.7	47.2	48.5	35.5
% Undergraduate math major	77.8	70.6	67.0	69.8
% Graduate math major	28.3	25.6	22.7	23.2
% Certified in math	99.9	99.0	98.9	98.4
% Part-time	1.1	1.5	1.7	1.4
English	Above (1429)	Average (2256)	Below (903)	Heterogenous (766)
Mean class size	24.5	24.6	20.0	23.8
Mean experience	5.7	5.6	5.6	5.3
% Masters degree	56.0	49.8	56.1	47.8
% Undergraduate English major	78.4	74.7	70.8	73.1
% Graduate English major	31.0	31.1	31.9	22.6
% Certified in English	98.7	98.2	99.2	98.9
% Part-time	0.8	1.9	2.1	0.4

Notes: Sample sizes in parentheses. Population weights were used to correct for the oversampling of certain groups. See National Center for Education Statistics (1992) (p. 45) for a discussion of the weights in NELS.

degree in any subject. Overall, these data suggest a mixed picture in terms of the disproportionate allocation of resources across tracks.⁷

4. Concluding thoughts

It is encouraging to see high quality econometric research being done on institutional schooling issues like ability grouping. We believe that economists have a useful contribution to make to this and other educational policy debates.

B and S argue that past studies comparing students in tracked vs. untracked classes overestimated the impact of tracking on student achievement by not adequately controlling for student ability. However, we believe that their attempt to control for student ability by comparing the achievement of students in classes of similar composition is not the correct approach and that their estimates should be interpreted as the effect of formal versus informal tracking. B and S show that, controlling for class composition, neither class size nor teacher characteristics vary much according to whether a student is in a tracked or detacked environment. This suggests to us that there are few observable differences between schools with and without formal tracking policies. Given this similar pattern of resource allocation, it is not surprising that B and S find that students in “tracked” and “untracked” schools have similar achievement gains.

References

- Argys, L.M., Rees, D.I., & Brewer, D.J. (1996). Detracking America's schools: equity at zero cost? *Journal of Policy Analysis and Management*, 15(4), 623–645.
- Betts, J.R., & Shkolnik, J.L. (1999). The effects of ability grouping on student math achievement and resource allocation in secondary schools. *Economics of Education Review*, 19, 1–15.
- Brewer, D.J., Rees, D.I., & Argys, L.M. (1995). Detracking America's schools: the reform without cost? *Phi Delta Kappan*, 77(3), 210–215.
- Figlio, D. N., Page, M. E. (1998). *Ability tracking, school choice and social mobility: Does separation increase equality?* Institute for Research on Poverty Discussion Paper. University of Wisconsin.
- Hanushek, E. (1989). The impact of differential expenditures on school performance. *Educational Researcher*, 18(4), 45–62.
- Hedges, L., Laine, R., & Greenwald, R. (1994). A meta-analysis of the effects of differential school inputs on student outcomes. *Educational Researcher*, 23(3), 5–14.
- Hoffer, T.B. (1992). Middle school ability grouping and student achievement in science and mathematics. *Educational Evaluation and Policy Analysis*, 14(3), 205–227.
- Loveless, T. (1997). *The uses and misuses of research in educational reform*. Presented at the State of Student Performance in American Schools, Brookings Institution, Washington, DC.
- McDonnell, L. (1995). Opportunity to learn as a research concept and a policy instrument. *Educational Evaluation and Policy Analysis*, 17(3), 305–322.
- National Center for Education Statistics (1992). *NELS first follow-up: student component data file user's manual*. National Center for Education Statistics, Washington, DC.
- Oakes, J. (1990). *Multiplying inequalities: the effects of race, social class, and tracking on opportunities to learn math and science*. RAND, Santa Monica, CA.
- Oakes, J. (1992). Can tracking research inform practice? Technical, normative, and political considerations. *Educational Researcher*, 21(4), 12–21.
- Sengupta, S. (1997). Are teachers of teachers out of touch? *The New York Times*, 22 October, B8.

⁷ Several points are worth making in connection with our consideration of classroom resources. First, we are only able to document differences across tracks in *measured* classroom characteristics. It is quite possible, for example, that more able and motivated teachers are assigned to upper-level classes but if these traits are independent of years of experience and formal qualifications we will not have adequately captured this inequitable distribution across tracks. Second, our analysis concerns only classroom *resources* rather than “opportunity to learn” (OTL) more broadly (McDonnell, 1995). Third, it is not clear what effect, if any, that a differential resource distribution across tracks has on student achievement. Many researchers argue that there is little evidence of a systematic relationship between schooling characteristics such as class size and teacher experience and standardized test scores (Hanushek, 1989), although this view is not unchallenged (Hedges, Laine & Greenwald, 1994).