Ifo Institute for Economic Research at the University of Munich

The Economics of Tracking and Non-Tracking

Volker Meier
Gabriela Schütz

Ifo Working Paper No. 50

October 2007

Ifo Working Paper No. 50

The Economics of Tracking and Non-Tracking

Abstract

There exists substantial variation across countries as to whether and how students are grouped in classes according to ability. Economic analyses stress that there is joint production of human capital in schools, where output increases with mean ability in the class. Ability tracking may therefore be particularly helpful for talented students. At the same time, weak students may benefit via tailored and specialised courses. The vast majority of the econometric literature suggests that tracking promotes inequality in academic achievement. By contrast, the empirical literature on the impact of tracking on average student performance is inconclusive. Only few studies find a significant association, including both positive and negative estimates.

JEL Code: I20.
Keywords: Tracking, ability grouping, peer group effects, school systems.

Volker Meier
Ifo Institute for Economic Research
at the University of Munich
Poschingerstr. 5
81679 Munich, Germany
Tel.: +49(0)89/9224-1371
meier@ifo.de

Gabriela Schütz
Ifo Institute for Economic Research
at the University of Munich
Poschingerstr. 5
81679 Munich, Germany
Phone: +49(0)89/9224-1698
schuetz@ifo.de

A. Introduction

In the vast majority of OECD countries, some kind of ability grouping is employed within secondary schooling. In many cases this takes the form of explicit tracking where students somewhere between age 10 and age 16 are sorted into schools of different types. These schools offer a specific degree of academic orientation in their curriculum. The lower track schools often have some vocational orientation, while the higher track schools exhibit a more academic curriculum. In the last decades, several OECD countries moved towards detracking their school systems or decided to defer tracking to a later point in the schooling career. Movements in this direction were enacted in Sweden, the UK, Italy and Norway in the 1960s, in France and the US in the 1980s, and in Spain and Portugal in the 1990s. These measures were usually taken because of a perceived disadvantage of children with less favourable family background at the separation stages in education. Indeed, early tracking seems to induce a higher intergenerational correlation of school careers and wages. Thus, in tracked systems the final school degrees of parents and their rank in the wage distribution will to a higher extent be mirrored in the achievements of their children (Dustmann, 2004; Bauer and Riphahn, 2006; Pekkarinen et al., 2006).

Figure 1 shows the age at which students are first tracked in the different OECD countries. The age of first selection refers to a standard case and may show regional variations in countries in which the organisation of the school system is decentralised. Obviously, Germany and Austria have a boundary position in the comparison with other countries. Selection takes place after four years of schooling at age 10. Most of the OECD countries do not track their students before age 14. The upper boundary is set to age 16 and covers also education systems in which no tracking at all occurs over the course of compulsory schooling, like Sweden, Spain, New Zealand, Iceland, Finland, Denmark and Canada. Thus, among OECD countries there is no consensus on the best way to deal with heterogeneous ability among the student population.

Tracking of students into different types of schools is not the only possible way of grouping students according to ability. Countries with a comprehensive school system often employ streaming within schools as a milder form of ability grouping. In this case, only some courses are taught at different levels of complexity. Streaming entails that students are placed into advanced classes for one subject and into standard classes for another subject. The existing literature largely ignores the distinction between tracking and streaming and treats the notions as synonymous. Since the main research questions

on the impacts of ability grouping also apply to streaming, we also review the empirical evidence from countries in which students are all placed in the same type of school, like in the US. However, the distinction between tracking and streaming is of importance for empirical researchers trying to assess the relative merits of various forms of ability grouping.

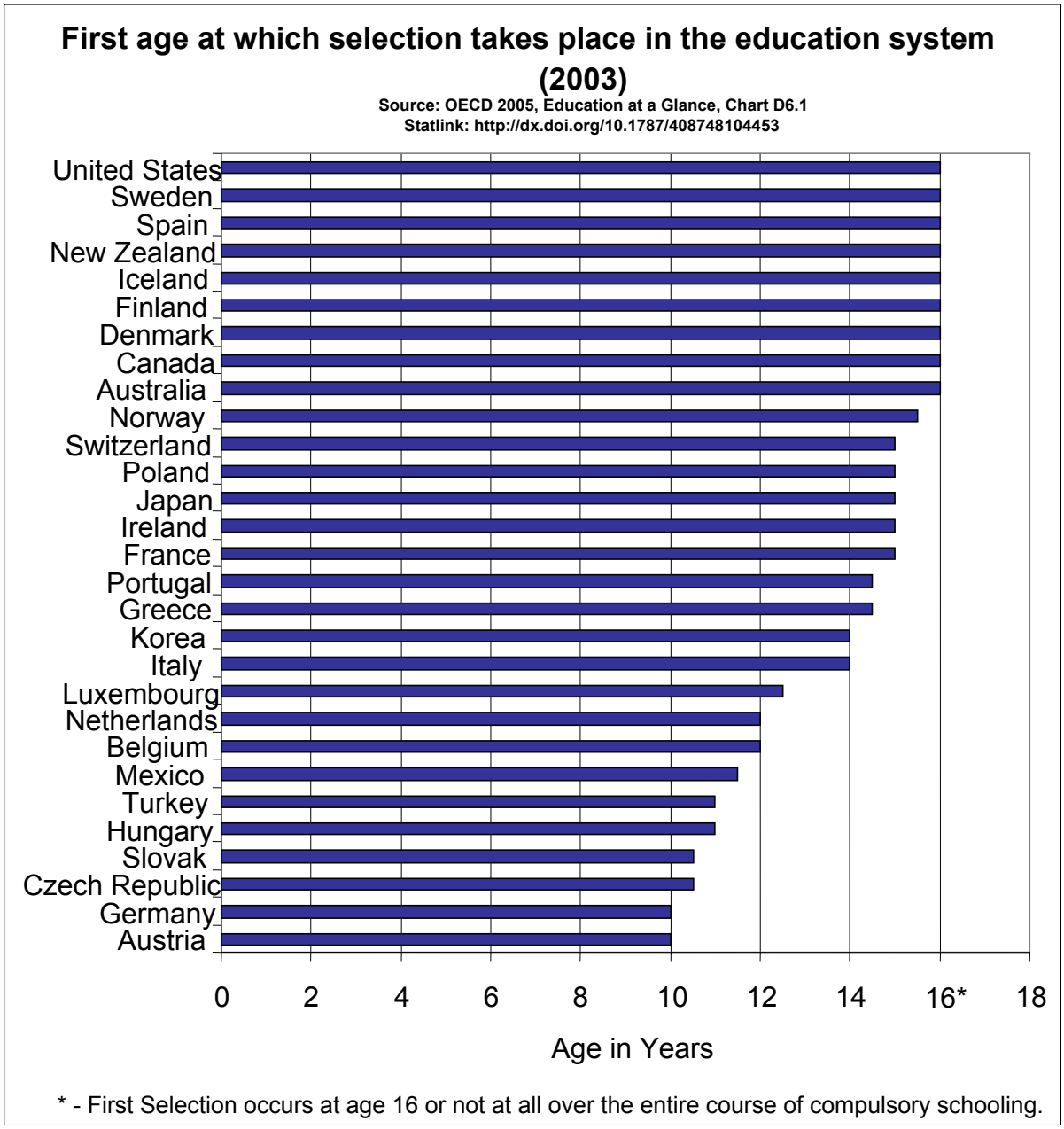

Figure 1: Tracking in OECD countries

Further, it should be noted that ability grouping often also occurs within a school system that officially does not employ any tracking or streaming measures. If the rich live apart from the poor and the children's ability levels are positively correlated with the income of their parents, schools located in rich areas will display a higher mean student ability than schools located in poor areas.

Two main questions are to be answered when considering the economics of ability grouping. The first question relates to a wider notion of *efficiency*. If the choice is between a school system with ability tracking at a given age and no tracking at all, which scheme will yield a higher aggregate sum of individual productivities net of costs of education? In an ex ante sense, a scheme would be called efficient if an individual who does not yet know his type in terms of innate ability would prefer the scheme under a veil of ignorance. In general, it cannot be expected that one scheme dominates the other, such that productivity is higher for any initial ability type. Employing the wider notion of efficiency makes it possible to rank the schemes if, for example, tracking increases productivities of high ability types and reduces productivities of low ability types. The second question deals with the *equality of opportunity* offered by the scheme in a narrower sense. Considering the same choice as before, which scheme leads to higher productivities for disadvantaged students?

Apart from the questions of efficiency and equality of opportunity in tracked or non-tracked systems, the more detailed structure of the tracking scheme is also to be considered. If ability grouping is beneficial, the next questions would be which form of ability grouping should be employed - different school types or streaming within schools -, and at which age tracking should be introduced.

There is clearly some intuitive expectation concerning the answers to the first two questions. The motive behind tracking is to increase output of the school system in terms of aggregate wages. By creating more homogenous classes, teaching becomes easier because the style of the course can be better adapted to the needs of the students. One main channel why this procedure may yield improved outcomes lies in faster accumulation of human capital of the more talented students. If this is indeed the case, a higher aggregate productivity measure may be achieved even if tracking reduces the future productivities of students sent to lower track schools. But it is not even clear whether tracking will reduce the productivity of students in the lower track. One would expect that teachers can help the less able students better when they no longer have to take care of students of high ability. In contrast, the main idea behind non-tracking is to

guarantee equality of opportunity as long as possible. This will be achieved by a policy of equipping all students with the same level of knowledge. While ideally a tracking scheme also implies equality of opportunity ex ante, observations from several countries suggest that early tracking procedures are noisy in the sense that capabilities are imperfectly identified and predicted. Since the influence of parents on school choice is more pronounced when the child is younger, the intergenerational correlation of income and final school degrees is typically stronger the earlier tracking occurs.

The remainder of the paper is organised as follows. Section B reviews the contributions of economic theory to the analysis of tracking. As lots of predictions remain ambiguous, the empirical evidence surveyed in Section C deserves particular attention. We cover the US experience in effectively dealing with different streaming policies across high schools, analyses of regime changes in the UK and Sweden, and international comparisons building on international student achievement tests. Section D concludes and indicates directions for future research.

B. Theory

I. Peer group effects

When considering the question of which school system generates the highest output in terms of aggregate productivities of its students at a given total cost of education, the educational production function has to be taken into account. In an educational production function, an outcome variable describing human capital of a student at the end of school is related to input variables like initial ability of the student, number of schooling hours, class size, or teacher quality. With such a simplified approach, family background variables are often captured by measures of initial ability. If human capital is produced by the student, the teachers and some material resources, tracking would only matter if the material resources are spent in an unequal fashion afterwards or if the characteristics of the teachers would differ between the different school types. The interesting point is that we have joint production, where the abilities of all students in class enter as factors of production. This aspect is usually covered by incorporating mean ability in class in the production function. A higher mean ability increases the individual's output measure for given initial abilities and given resources spent on teaching. This positive impact of mean ability is called the *peer group effect*. It is

usually interpreted as a consequence of joint learning where the students support each other.

Alternatively, the peer group effect may also express that the share of weak students who impose negative externalities on all pupils is reduced. Weak students tend to disrupt the learning process. This problem gives a rationale to sort out the weak students by tracking. The optimal class size in a higher track will exceed the optimal class size in a lower track because there are fewer disturbances. This theory can explain why empirical researchers typically fail to find the expected negative correlation between class size and educational outcomes (Lazear, 2001).

In the following, we consider an educational production function with only three inputs: initial ability of the student, mean ability in class, and resources spent on the student. Using such a simple approach is sufficient to highlight the main issues associated with the tracking decision.

The mere existence of peer group effects suggests that grouping classes according to ability at a given curriculum and given resources spent on each student has an effect on educational outcomes. If the sorting procedure works, individuals in higher tracks would benefit from tracking while individuals in lower tracks are harmed. However, the theoretical prediction on the consequences of peer group effects on the construction of output-maximising class structures is ambiguous. Arnott and Rowse (1987) maximise a welfare function in which welfare increases in each student's final human capital, but may fall with more inequality. They point out that the specific structure of the educational production function has to be considered when constructing an optimal class system. The educational production function they analyse is

$$h = a^{\alpha} m^{\beta} e^{\gamma}, \qquad (1)$$

where h is human capital of the student as the output variable, a is his or her initial ability, m is mean ability in class, e is expenditure spent on the student, and $\alpha, \beta, \gamma \in [0,1)$ are production coefficients. Apart from boundary cases with zero marginal productivities, the production function exhibits positive and diminishing marginal productivities, and the factors of production are complements. Having mixed classes means that mean ability of the whole student population applies for every student. By contrast, tracking assigns students with a high initial ability a to classes with

mean ability m above average. Accordingly, students with low initial ability are taught in classes with mean ability below average.

Increasing the marginal productivity of mean ability in class may both imply a tendency in favour of streamed classes and in favour of mixed classes. For example, starting at β=0, peer group effects are absent. Streaming may be used to exploit that educational resources e are more productive when being spent on more talented students. The predominant effect of introducing peer group effects by increasing β may then consist in the feature that gains from raising mean ability are particularly strong for low levels of mean ability in class. This implies a tendency in favour of mixed classes, as tracking would hurt weak students substantially while losses of able students from detracking are small. Increasing the strength of the peer group effect by further increasing β can then lead to a functional form in which gains from increasing mean ability are similar at all levels of mean ability and given ability of the student, and become very strong for talented individuals. Tracking will then yield the maximum aggregate human capital because the gains of the high ability students from tracking by far outweigh the losses of the low ability individuals.

One main effect of tracking is to make public schools more attractive relative to private schools. Epple et al. (2002) consider an environment with school competition, where private schools can set tuition fees and individuals are differentiated according to initial ability and household income. Private schools will always be stratified according to tuition, where the most expensive schools have the best peer quality among its students, and its student population exhibits the highest average household income. While private schools may in principle also employ ability tracking, a similar effect is achieved by prices that can be differentiated according to ability and parental income. Without tracking, mean ability in the student population in any public school will lie below mean ability in the weakest private school. This is a consequence of the fact that even the weakest private school charges a positive fee while all public schools are free. Introducing tracking in public schools with two tracks will attract students with intermediate ability from families with high income to public schools. Returning to the public system is attractive when these students are assigned to the higher track. The public lower track schools are still the weakest, but the public higher track schools will typically show a higher quality in terms of average ability in class than some private schools. Tracking then tends to harm students of low ability with low household income and particularly helps more able students with poor parents.

The use of tracking allows for a differentiation of the academic level across different types of schools. Employing a higher academic level in schools attended by stronger students may simply reinforce the ordinary peer group effect. Effinger and Polborn (1999) study a hierarchical school system with two school types in which raising the academic level yields a stronger increase in a higher school student's productivity if mean ability in class is higher. Households are free to choose to send children to the lower or the higher track school. The weakest students are always best served at the lower track school. Sending a child from a threshold household to the higher track school rather than to a lower track school is associated with negative externalities, because mean ability both in the higher track class and in the lower track class deteriorates. As these negative externalities are not taken into account by parents, too many students attend higher track schools under free school choice. This argument justifies binding recommendations. Further, schools choose the academic level to maximise the educational output of their students (Costrell, 1994). Higher track schools may not take into account that increasing the academic level deters weaker students from entering and imposing negative externalities. This provides an argument for centralised setting of academic standards, which yields higher standards than under a decentralised organisation.

A drawback of this argument is that higher academic levels also imply a risk of failure at school, In the US, the dropout rate in high school has been around 20 per cent in the nineties (Arum and Hout, 1998), and the loss of income arising from dropout rates has been substantial (Domazlicky et al., 1996; Thompson, 1998). If output losses are higher for more talented students, tracking procedures associated with adapting the academic standard implies a higher risk of human capital losses of talented students. In such a situation, a comprehensive system without tracking may well lead to a higher aggregate income (Meier, 2004).

II. Other arguments

Economic theory has also stressed the signalling value of education. According to this view, education is not necessarily productive per se. One main function of the school system is to provide prospective employers with some information about the productivity of the applicant. Given that ability screening in a tracking scheme works reasonably well, a selective system tends to be more informative than a comprehensive school system (Brunello and Giannini, 2004a). This is particularly true with respect to the assessment of students at the end of the ability distribution. If final grades are not

comparable across students due to individualised courses and exams, the only reliable pieces of information are completion of school and school type, where only the selective scheme provides the latter signal. It can be argued, however, that this problem may be mitigated by introducing uniform examination procedures in order to make grades more informative.

Another key idea behind tracking is to exploit gains from specialisation. Rather than teaching similar curricula at different levels, tracking is often used to teach a more practical and vocational curriculum in the lower track and a theory-orientated and more academic curriculum in the higher track. A comprehensive system would incorporate elements of both tracks, which allows for more flexibility of the students. As expressed by Brunello and Giannini (2004b), human capital is kept more versatile. On the other hand, accumulation of human capital of a given type – academic or vocational – would be faster if the student specialises early. As the jobs ultimately require predominantly academic or predominantly vocational skills, a stratified school system with tracking producing the desired shares of skills is superior to a comprehensive scheme without tracking under perfect foresight. This argument is similar to the idea that segregated clubs are more efficient because the provision of public goods can be tailored to the preferences of the club members. The comprehensive scheme without tracking displays its strength in environments with a substantial amount of uncertainty.

If tracking is used for introducing specialised curricula, an interesting problem is to determine the optimum age at which students are separated. If separation starts too early, there is some risk of mismatch unemployment. As the separation of students determines the future shares of workers of different types, the resulting proportions may not fit well with the needs of the employers after completion of schooling. In addition, if there are frictions in matching in the job market, mismatches will occur where positions are filled with workers displaying specialisations that are not used in the job. Such mismatches are associated with higher output losses when specialisation at school starts earlier. Therefore, increasing frictions in the labour market imply that the optimal tracking age goes up (Ariga et al., 2006).

Moreover, the earlier tracking sets in, the noisier the test signal is when measuring true ability. Apart from the problem that ability is revealed only gradually, younger students show stronger differences in maturity. In fact, the younger students in class more often achieve an upgrading as a correction of a previous tracking decision (Puhani and Weber, 2007). Hence, losses from sending children to the wrong schools will be higher with

earlier tracking. By contrast, if specialised curricula are introduced late, the gains from specialisation tend to be lower (Judson, 1998, Ariga et al., 2005).

Finally, Brunello et al. (2004) argue that technological progress in firms leads to a higher depreciation rate of vocational training. This reduces the gain from earlier specialisation as the acquired knowledge may already be obsolete when entering the job market. More academic tracks are not hurt by skill depreciation in a similar fashion. The appropriate policy in view of such a skill-biased technological change is to increase the share of students in the academic track and to separate the students later.

C. Empirical Analyses

I. Overview

For empirical researchers, the most obvious approach to analyse the consequences of tracking policies consists in monitoring major reforms or local reform experiments. However, because whether to employ tracking or not is a central structural element in education systems, reforms or experiments are rare. Therefore, empirical analyses of the effects of tracking often struggle with credible statistical methods to identify causal effects. The missing availability of policy experiments also explains why empirical research has taken different alternative approaches.

One main option is to exploit variation of ability grouping policies within and across schools. This route is pursued in many studies from the US. While tracking in the shape of sorting children into different types of secondary schools is uncommon, the US high schools display a substantial amount of heterogeneity with respect to their streaming policies for different subjects. As a consequence, the notion of tracking is not well-defined in many datasets. Additional problems arise if the measures of student performance are correlated to unobserved student characteristics that determine both school choice and track placement.

Another strand of the literature analyses institutional variations across countries using international achievement tests. This approach can avoid problems of explaining selection of students into tracking and non-tracking schools by omitting countries in which both types of schools exist. The major disadvantage lies in the fact that it is almost impossible to control for all relevant institutional differences between countries.

For example, social policies may be crucial for achieving mixed classes in a comprehensive school system. Peer group effects are among the forces creating segregation of communities by internal migration, with schools in richer areas displaying a higher peer quality (see, for example, de Bartolome (1990)). In such a situation, moving toward more heterogeneous classes requires measures like augmenting the space for free school choice and subsidise transportation for pupils from poor areas to schools in rich areas. Other institutional features that affect the performance of schools in a country are the degree of school autonomy, centralised exams, and competition between schools (Woessmann, 2003).

II. Older literature on peer group effects

The effects of tracking have first been discussed in the United States in the context of peer group effects. Since it is a widely held belief that peer groups exert substantial influence on the educational production process, it is interesting to analyse the effect of a change in the composition in the peer group that may be caused by tracking.

Summers and Wolfe (1977) analyse the effect of peer group composition on students' achievements in the Philadelphia school district, using the change in the composite achievement score between the third and sixth grade as outcome variable. They find that a higher share of high achievers in a grade has a positive impact on low achievers, while the impact on high achievers is negligible. At the same time, a higher share of low achievers reduces the score of both low and high achievers. Hence, peer group effects are present for everybody and work strongest for weak students. However, the peer group measure used refers to school grades, but not classes. Yet, the presence of tracking implies that the composition of classes within the same grade differs with respect to their peer group quality.

Henderson et al. (1978) estimate peer group effects in elementary schools using achievement scores in French and mathematics between the first and the third grade of French speaking students in Montreal. The authors control for innate ability by including the first recorded IQ score of the student and at the same time also control for achievement in the previous grade. Peer group quality is measured by the mean IQ in the class or by the relative position of this mean IQ among all classes. In each case the coefficients of the peer group variables do not change with the student's ability, indicating similar peer group effects for all ability groups. Since the peer group effect is positive and concave, test score gains from increasing the class mean IQ by one point is

diminishing with increasing mean ability. Henderson et al. (1978) therefore conclude that maximising the sum of scores would be achieved by mixed classes. Assuming that tracking separates students on the basis of IQ, and that peer group effects work similarly across grades and subjects, these results indicate a negative effect of tracking on overall achievement.

III. US studies

Since then, several contributions have aimed at determining the sizes of peer group effects in tracked and non-tracked US high schools. Hoffer (1992), Argys, Rees and Brewer (1996), Betts and Shkolnik (2000a), and Figlio and Page (2002) use rather similar datasets but slightly different approaches to address the effects of tracking. Hoffer (1992) and Betts and Shkolnik (2000a) both use the Longitudinal Study of American Youth (LSAY), while Argys et al. (1996) and Figlio and Page (2002) use the National Education Longitudinal Study of 1988 (NELS). Their results are similar in that tracking does not seem to be related to substantial efficiency gains. However, with respect to differential effects for students of different ability levels, the authors come to widely different conclusions.

Table 1: Evidence from the US

Paper	Data	Country	Tracking /Streaming information	Outcome variable	Results
Hoffer (1992)	Longitudinal Study of American Youth (LSAY)	USA	a) Dummy variable if school uses ability grouping in math classes b) Teacher's evaluation of average ability level of class compared to other classes in same school	Test scores in mathematics and science in ninth grade	No effect of ability grouping on overall achievement. Placement in high group has positive effect, placement in low group has negative effect
Betts and Shkolnik (2000a)	Longitudinal Study of American Youth (LSAY)	USA	a) Dummy variable if school uses ability grouping in math classes. b) Teacher's evaluation of average ability level of class compared to other classes in same school.	Test scores in mathematics and science in ninth grade	No effect of ability grouping on overall achievement. Low ability children not affected by ability grouping, middle ability students harmed, high ability students gain
Argys et al. (1996)	National Education Longitudinal Study, 1988	USA	Teacher's evaluation: a) achievement level in mathematic class considered above average, average, below average, widely differing b) class is honours, academic, general or vocational track	Test scores in mathematics tenth grade	Students in lower tracks gain from detracking, students in higher tracks lose. Detracking reduces average test scores by two percent
Figlio and Page (2002)	National Education Longitudinal Survey, 1988	USA	Tracking dummy a) from principals' reports: whether school uses ability grouping or tracking in math classes. b) combining answers from teachers' reports	Gain in math test scores between 8th and 10th grade	Overall, no effect of tracking. Using two-stage least squares: Tracking helps students in bottom third of test distribution

Hoffer (1992) examines the effects of ability grouping between seventh and ninth grade and performs different analyses using mathematics and science scores. Information on the schools' use of ability grouping and on the level of the ability-grouped class that the students attended were gathered from teachers and school documents. The analysis starts with simple ordinary least squares regressions of later achievement on previous achievement, measures of social background and two sets of grouping indicators. First using a dummy indicating whether or not the school uses ability grouping, no evidence is found for a significant effect of ability grouping on average achievement. The second set of grouping indicators consists of dummies indicating ability group level placement in high, middle and low group with non-grouped as the reference category. Thus, students in high, middle and low ability classes are compared to students in schools that supposedly do not use ability grouping. The results across different specifications indicate that placement in the high group has a weak positive effect, and that placement in the low group has a stronger negative effect. The results for the overall effects of grouping and the effects of group placement are not greatly changed by employing a propensity score method to encounter the criticism that grouped and non-grouped students are not comparable. This criticism relates to the fact that selection into grouped or ungrouped classes might be based on unobservable factors that are correlated with achievement, such as motivation, which would lead to biased estimates.

Betts and Shkolnik (2000a) prefer to use information on the ability level of the class provided by the teachers to compare grouped and non-grouped students at given class ability levels. Using this information might even better control for unobserved heterogeneity than a propensity score method. They control for average class ability both at grouping and at supposedly non-grouping schools using two different measures of class ability. Their first measure is based on teachers' reports on the average ability level of the class as compared to other classes in the same school. Their second measure is obtained by subtracting the mean achievement of the grade from the initial achievement and subsequently grouping the students into quartiles based on their normalized scores. Betts and Shkolnik regress student achievement on previous achievement, control variables and a set of dummy variables indicating membership in different ability groups and whether or not the school employs grouping according to the principal. The difference in the coefficients on class ability between grouped and non-grouped schools are used to identify the effect of being placed in a class of given ability in a school that officially groups. The overall effect of formal ability grouping on the average student is insignificant. This finding is in line with Hoffer (1992), but stands in contrast to Argys, Rees and Brewer (1996) who find a positive significant overall

effect. Further, Betts and Shkolnik (2000a) find that children at the bottom of the ability distribution are not affected by ability grouping, that middle students are harmed, and that high ability students gain. The authors discuss and present several robustness checks which lead to insignificant estimates for the effects of grouping in the case of a propensity score method and to significant effects of grouping only for high ability children in the case of an instrumental variable approach. Their conclusion is that previous research by Hoffer (1992) and Argys et al. (1996) has overstated the differential effects of ability grouping across students of differing ability levels due to inadequate controls for class ability levels at supposedly ungrouped schools. On the other hand it can be argued that the approach used by Betts and Shkolnik (2000a) does not allow the correct identification of grouped and ungrouped classes and that the results should be reinterpreted as the effect of formal versus informal tracking (Rees et al., 2000).

Argys et al. (1996) infer from their analysis that abolishing tracking would lead to losses in efficiency. They consider the impact of tracking on test scores in mathematics at the end of the 10^{th} grade. Two measures of grouping of students that were derived from survey questions to teachers are used separately in the analysis: the achievement level of the students in the class as compared to the average student in the same grade (above average, average, below average, heterogenous), and the track that the class could best be described as (honors, academic, general or vocational). For each track or ability group, achievement equations are estimated separately. Averaging predicted achievement in each track across all students in the sample gives an estimate of what the achievement of the average student would be if placed in a certain track or ability group. Comparing mean predicted achievement between tracks or ability groups, the authors find uneven effects of detracking schools. Students in below average classes or tracks would gain and students in above average classes or tracks would lose. The overall net effect of placing all students in heterogeneous classes would be a 2 percent drop in mathematics test scores.

Betts and Shkolnik (2000b) claim that the sizes of the effects reported by Argys et al. (1996) are overstated. They argue that group or track placement may be correlated with the error when students' initial achievement is only imperfectly measured, and that therefore the estimates will be systematically biased. Moreover, the selectivity correction used by Argys et al. (1996) appears not to correct for the correlation between track placement and initial achievement, but rather to increase the omitted ability bias. Betts and Shkolnik (2000b) also criticise that a ´heterogeneous` ability level of a class

should not be used to compare grouped classes of differing ability levels to ungrouped classes, since it may mean different things to different teachers in different schools.

Using the same data as in Argys et al. (1996), Figlio and Page (2002) compare students' achievement growth between 8th and 10th grade only across schools, with some schools employing ability grouping measures while others do not. Figlio and Page divide the students according to their 8th grade math achievement into top, middle and bottom thirds of the test score distribution. Subsequently, they estimate separate regression equations for each of these subsets and include a dummy for whether the principal reports that the school tracks. The effect of tracking is statistically insignificant within each third of the student achievement distribution. Figlio and Page interpret this as evidence that lower test score gains observed among students in low ability tracks does not stem from tracking placement, but from unobserved factors that are correlated with track placement. Using alternative approaches to define the school as tracking or non-tracking, the authors again do not find evidence that tracking harms low-ability children. The authors also argue that a school's tracking status affects school choice. Therefore, a two-stage least squares analysis is performed to address the potential endogeneity of tracking status. The results from this exercise indicate that low-ability students may actually gain from being placed in schools that employ tracking.

The major problem of these US studies is the nonexistent definition of tracking and ability grouping in the survey questionnaires. Therefore, grouping measures may not be comparable across schools or even teachers. Moreover, two possible sources of endogeneity make it difficult to estimate causal effects of different tracking policies within any survey data. First, track placement may be correlated with unobserved factors such as motivation that both determine achievement and track placement. Second, the selection of students into schools is presumably endogenous. If parents tend to send talented children to schools employing tracking programmes, the results will be biased towards finding a positive effect of tracking on achievement.

IV. Impacts of reducing the variance in mean ability

In the last few years, some attempts have been undertaken to identify the pure peer group effect generated by tracking. By creating homogenous classes, tracking not only changes the mean ability in classes for students assigned to different tracks, but it also reduces the variance in ability. The impact of tracking on the level of the peer group effect is seen by measuring the effect of reducing the variance on the level of the

student's test scores at given mean ability. In addition, tracking also affects the strength of the peer effect, which would be expressed by the change in test scores associated with varying mean ability. If individuals benefit from tailored courses in homogenous classes, the sum of these two effects is expected to be positive. As the results presented by Zimmer and Toma (2000) and Zimmer (2003) show, this is not necessarily the case.

Estimating peer group effects in a sample of students from five countries, Zimmer and Toma (2000) surprisingly find that a higher standard deviation of test scores within the class leads to higher individual student achievement. The strength of this effect, however, also depends on the mean ability of the class. Increasing the variance in ability within a class generates smaller effects when the mean of the classroom is higher; and higher effects when the mean is lower. At the same time, low-ability students capture higher gains from a better peer set than high-ability students. Thus, low-ability students are the ones most affected by classroom composition. Zimmer (2003) uses only the US portion of the same dataset and shows that for low-ability and average-ability students tracking reduces the positive peer effect, whereas for high-ability students the peer effect is unaffected by the tracking procedure. His results also suggest that the institutional practice of tracking has a positive effect on low- and high ability students at low levels of peer quality; a situation which is more likely to occur in the presence of tracking. Thus, the positive effect of tracking may compensate weak students for losses associated with a reduction in peer group quality upon separation from stronger students. However, if both effects are combined, the loss of more able peers seems to outweigh the benefits of tracking for low- and average ability students.

V. Policy reforms

Other contributions to the empirical literature have tried to exploit institutional change or reforms over time. Since determining causal effects is very difficult in the context of student performance due to unobserved ability and family background and due to the difficulties in establishing the treatment effects on the non-treated, reforms in education systems offer a much more reliable basis for establishing causal effects. This approach is followed by Galindo-Rueda and Vignoles (2004) and Manning and Pischke (2006) for the United Kingdom and by Meghir and Palme (2005) for Sweden.

Table 2: Evidence from policy reforms

Paper	Data + Country	Reform information	Outcome variable	Results
Galinda-Rueda and Vignoles (2004)	National Child Development Study, England and Wales	Change from selective to comprehensive school system	Test scores in mathematics at age 16	Tracking helps high ability students, insignificant effects on others
Manning and Pischke (2006)	National Child Development Study, England and Wales	Change from selective to comprehensive school system	Test scores in mathematics at age 16 and 11	Selection bias cannot be ruled out. Applies also to Galinda-Rueda and Vignoles (2004)
Meghir and Palme (2005)	1948 and 1953 cohort survey, Sweden	Abolishment of tracking, extension of compulsory schooling, introduction of national curriculum.	Final educational attainment and annual earnings	Educational attainment and earnings increased for individuals with low-skilled fathers

Galindo-Rueda and Vignoles (2004) use the National Child Development Study consisting of a cohort of individuals born in England and Wales in March 1958. The data contains the test scores of the individuals at age 7, 11, and 16, together with lots of individual and family characteristics. As the movement from the selective to the comprehensive school system was delayed or accelerated by local education authorities, both systems coexisted in the years under consideration. In their study, Galindo-Rueda and Vignoles take the test score in mathematics at age 16 as the dependent variable. Using a matching approach in which individuals with similar characteristics are compared to each other, their results indicate that a selective school system favours high ability students, whereas the impacts on middle and low ability students are insignificant. In contrast, Manning and Pischke (2006) argue that nothing can be learned from the education policy reform in England and Wales. They analyse the same data and perform similar analysis, but also test additional specifications. Using ability measures at age 11 as the outcome variable, that is before even entering secondary education, the authors are still able to show a negative effect of attending a comprehensive school. Since it is not possible that later attendance of comprehensive schools influences test scores during primary school years, these results indicate the presence of selection bias. Hence, results for specifications using age 16 ability measures are most likely biased as well and should not be interpreted casually. The

selection into comprehensive and selective schools during the transition period seems to be based on unobservable characteristics that cannot be controlled for.

The consequences of the Swedish education reform in the 1950s give strong evidence for an equalising effect of detracking, with gains arising for individuals of a less favourable educational background and losses for children of skilled parents. The reform not only abolished tracking after the sixth grade, but also entailed the extension of compulsory schooling and the introduction of a national curriculum. Meghir and Palme (2005) use the 1948 and 1953 cohort survey that collected information on individuals when they were in sixth grade. Data from these two cohort surveys was supplemented by information on final educational attainment from the 1990 Swedish education register and by information on annual earnings and employment status between 1985 and 1996 from the Swedish tax registers. Meghir and Palme find that the reform increased the educational attainment of individuals with unskilled fathers, with a particularly strong effect on the more able students. The increase in educational attainment beyond the new compulsory schooling level was stronger for women. Earnings of individuals with unskilled fathers increased significantly, with a more pronounced effect for women. The difference in the earnings increase between individuals of high or low ability was significant only for women. However, earnings of individuals with skilled fathers were negatively affected by the reform. One reason for this phenomenon can be traced to the fact that this group used to attend the academic track prior to the reform. The reform appears to have reduced both the quality of education for these individuals and their subsequent earnings. Since the Swedish reform included an increase in compulsory schooling and the establishment of a nationwide unified curriculum, it is not completely clear which part of the effects of the reform can be attributed to detracking.

VI. Interregional and international comparisons

Other papers have even more strongly focussed on family background, thereby not only considering the effects of tracking versus non-tracking, but more frequently also on the effects of the timing of tracking and of the number of distinct tracks available to secondary students. This applies both to studies exploiting regional variations across countries and to analyses of international student achievement tests.

Bauer and Riphahn (2006) estimate the effect of the timing of tracking on the educational mobility using within country variation in the timing of tracking between

the 26 cantons in Switzerland. If tracking takes place at an early stage in the educational career, students' innate ability will be measured with a lot of noise. In this case, track placement will be dominated by the parental background of the students. Bauer and Riphahn analyse the difference in the predicted probabilities of attending the college-bound track for children of high and low educated parents and compare those differences between early and late tracking cantons. The results show that the impact of parental education on track placement varies depending on the timing of tracking. Early tracking greatly increases the relative advantage of children with highly educated parents. Applying a similar methodology to data referring to the German Federal States, Woessmann (2007) confirms these results.

Table 3: Evidence from international studies

Paper	Data	Country	Tracking information	Outcome variable	Results
Ammermüller (2005)	PIRLS 2001 and PISA 2000	12 countries	Number of distinct school types (tracks) in secondary education.	Change in impact of family background on test achievement between primary and secondary schooling	Impact of family background higher with more school types (tracks) in secondary schooling
Hanushek and Woessmann (2006)	Several waves of PIRLS, PISA and TIMSS	18-26 countries	Dummy representing early tracking.	Change in dispersion of student test scores between primary and secondary schooling	Early tracking increases inequality, tendency for tracking to reduce mean achievement
Schütz et al. (2005)	TIMSS 1995 and 1999	53 countries	Timing of tracking	Impact of family background on test achievement	Impact of family background higher when tracking occurs earlier
Brunello and Checchi (2006)	European Community Household Panel (ECHP), International Adult Literacy Survey (IALS), International Social Survey Programme (ISSP)	12-25 countries	Weighted indicator capturing share of population tracked into vocational stream and time spent in tracks	Educational attainment, literacy, employment, earnings	Effect of family background stronger with tracking, tracking reinforces dispersion of earnings
Waldinger (2006)	PIRLS 2001, PISA 2000 and 2003 and TIMSS 1995 and 1999	8-15 countries	Number of years spent in untracked system	Student test scores in reading and mathematics	Family background more important in countries that track early, inequality already present before tracking occurs

Ammermüller (2005) employs data from two international student achievement studies PIRLS (Progress in International Reading Literacy Survey) and PISA (Programme for International Student Assessment) to analyse the effect of changes in the institutional structure of the education systems between primary and secondary schooling on the educational opportunities of children with differing family backgrounds. His dataset consists of student micro level, school and country level information for 12 of the 14 countries that participated both in PIRLS in 2001 and in PISA 2000. Using a difference-in-differences estimation approach, Ammermüller investigates how changes in institutional variables that occur between the primary and the secondary schooling level influence the strength of the influence of family background on the test score of students. One of the institutional changes that are considered in this analysis is the number of distinct school types or tracks that are available to students in secondary education. The results indicate that the number of school types is positively linked to the impact of parents' education and origin. Thus, the influence of parental background on student achievement is higher and - thus equality of educational opportunities is lower - in education systems that select their students into many different types of secondary schools.

Hanushek and Woessmann (2006) employ a differences-in-differences approach to analyse the question of whether the timing of educational tracking affects performance and inequality. Their dataset consists of different waves of the international student achievement tests PIRLS, PISA and TIMSS (Trends in International Mathematics and Science Study, formerly Third International Mathematics and Science Study). By matching tests measuring performance at the end of primary school (PIRLS 2000, TIMSS 1995 and 2003) with tests taken in secondary schools (PISA 2000 and 2003, TIMSS 1995, 1999 and 2003) they are able to compare inequality in student outcomes before selection takes place with inequality in outcomes after selection has taken place in some countries, but not in others. The matching of tests produces datasets containing between 18 and 26 countries depending on the subjects tested. The econometric analysis shows that early tracking increases inequality. With respect to the effects of tracking on average country performance, the picture is not consistent across countries and subjects but overall shows a tendency for early tracking to reduce average performance.

Evidence for the negative effects of streaming on the equality of educational opportunities is also provided by Schütz et al. (2005). Their analysis tries to provide an answer to the question of why the degree to which family background determines student achievement differs between countries. The authors use data from two international student achievement tests in mathematics (TIMSS 1995 and TIMSS 1999)

to obtain a sample consisting of 54 countries. The degree of equality of educational opportunities that a country achieves is measured as the extent to which the number of books in the students' home influences test achievement. Several organisational features of the education systems are found to determine educational opportunity. The equality measure increases with a longer duration of the pre-primary educational cycle, a higher level of enrolment in pre-primary education and a higher age of first streaming. Later tracking is shown to reduce the strength of the influence of family background on student achievement. The authors also analyse whether the timing of streaming is connected to the overall level of educational achievement in the participating countries. Since country mean test achievement is independent of the age of first streaming, delaying the timing of tracking does not appear to have negative effects on average student performance.

Brunello and Checchi (2006) and Waldinger (2006) also investigate whether tracking contributes to increasing inequality of educational outcomes by strengthening the impact of family background variables. Brunello and Checchi use several datasets like the European Community Household Panel, the International Social Survey Programme, and the International Adult Literacy Survey, that all measure outcome variables of young adults in terms of wages, final educational attainment, or literacy. The stronger effect of family background with tracking is persistent when considering educational attainment, and tracking reinforces the dispersion of earnings. By contrast, the effect of family background on reading skills of young adults is weaker in countries that track earlier. This finding indicates that the opposite result that arises with the PISA 2003 literacy score data has to be taken with caution. Waldinger considers international student test scores of primary school students (PIRLS) and secondary school students (TIMSS and PISA). He confirms the findings that family background effects on student achievement are stronger in countries that track earlier. However, using a difference-in-differences approach, he shows that differences in the importance of family background between countries are already present before tracking takes place, and that actual tracking does not increase the importance of the family background characteristics. Hence, the differences across countries cannot directly be traced back to the tracking procedures. Yet, one drawback of this study is the relatively small sample of countries which is why it cannot be ruled out that some important variation is missing.

D. Conclusions

Although the literature on the consequences of ability grouping has spread out substantially during the last two decades, clear-cut messages have not emerged. Neither theory nor the empirical evidence gives a definitive answer to the question of which school system yields the maximum of total productivities, or the maximum aggregate test scores at some given grade in school. Theoretical considerations suggest that ability grouping is useful if either talented students can achieve a strong increase in performance or if substantial gains from specialisation can be expected. In contrast, keeping students of different ability levels in the same class makes sense when weak students benefit a lot from the presence of strong students. Delaying the age when tracking starts lowers the risk of mistakenly sending children to the wrong school type, but also reduces gains of appropriately placed students.

The empirical literature has until now failed to provide clear answers to the question of the impact of tracking on average achievements of students. While some studies have found negative effects of tracking, others have found evidence for positive effects. Since lots of studies fail to find a significant overall effect, it seems most likely that on average the tracking regime has no major effect on average achievement. However, the relative advantages of tracking or detracking may still depend on timing, the composition of the student population and institutional features. Further, insignificant or even negative effects of tracking on average test scores do not exclude positive impacts on productivities, because test results may not be perfect indicators of subsequent labour market performance. In particular, the tests may not capture possible gains from specialisation.

Empirical research has also investigated the impacts of tracking on the equality of educational opportunities and on the dispersion of achievements. There is a clear tendency that ability grouping increases the variance of school achievements in terms of test scores. Moreover, children of disadvantaged family background seem to benefit from delaying tracking and from a reduced availability of distinct types of secondary school tracks. Yet, speculations that ability grouping helps the talented individuals and harms the weak students are still hotly debated.

References

Argys, L. M., D. I. Rees and D. J. Brewer (1996). Detracking America's schools: equity at zero cost? Journal of Policy Analysis and Management 15, 623-645.

Ariga, K., G. Brunello, R. Iwahashi and L. Rocco (2005). Why is the timing of school tracking so heterogeneous? IZA Discussion Paper No. 1854, Bonn.

Ariga, K., G. Brunello, R. Iwahashi and L. Rocco (2006). On the efficiency costs of de-tracking secondary schools. IZA Discussion Paper No. 2534, Bonn.

Arnott, R., and J. Rowse (1987). Peer group effects and educational attainment. Journal of Public Economics 32, 287-305.

Arum, R., and M. Hout (1998). The early returns - the transition from school to work in the United States. In: W. Müller and Y. Shavit (eds.), From School to Work, Oxford, 471-510.

Bauer, P., and R. T. Riphahn (2006). Timing of school tracking as determinant of intergenerational transmission of education. Economics Letters 91, 90-97.

Betts, J. R., and J. L. Shkolnik (2000a). The effects of ability grouping on student achievement and resource allocation in secondary schools. Economics of Education Review 19, 1-15.

Betts, J. R., and J. L. Shkolnik (2000b). Key difficulties in identifying the effects of ability grouping on student achievement. Economics of Education Review 19, 21-26.

Brunello, G., and D. Checchi (2006). Does school tracking affect equality of opportunity? New international evidence. IZA Discussion Paper No. 2348, Bonn.

Brunello, G., and M. Giannini (2004a). Selective schools. Bulletin of Economic Research 56, 207-225.

Brunello, G., and M. Giannini (2004b). Stratified or comprehensive? The economic efficiency of school design. Scottish Journal of Political Economy 51, 173-193.

Brunello, G., M. Giannini and K. Ariga (2004). The optimal timing of school tracking: a general model with calibration for Germany. In: L. Woessmann and P. E. Peterson (eds.), Schools and the Equal Opportunity Problem, Cambridge, MA: MIT Press, 129-156.

Costrell, R. (1994). A simple model of educational standards. American Economic Review 84, 956-971.

de Bartolome, C. (1990). Equilibrium and inefficiency in a community model with peer group effects. Journal of Political Economy 98, 110-133.

Domazlicky, B. R. et al. (1996). Measuring the cost of high school noncompletion in Southeast Missouri. The Journal of Economics 22, 81-86.

Dustmann, C. (2004). Parental background, secondary school track choice, and wages, Oxford Economic Papers 56, 209-230.

Effinger, M. R., and M. K. Polborn (1999). A model of vertically differentiated education. Journal of Economics 69, 53-69.

Epple, D., E. Newlon and R. Romano (2002). Ability tracking, school competition, and the distribution of educational benefits. Journal of Public Economics 83, 1-48.

Figlio, D. N., and M. E. Page (2002). School choice and the distributional effects of ability tracking: does separation increase inequality? Journal of Urban Economics 51, 497-514.

Galinda-Rueda, F. and A. Vignoles (2007). The heterogeneous effect of selection in UK secondary schools. In: L. Woessmann and P. E. Peterson (eds.), Schools and the Equal Opportunity Problem, Cambridge, MA: MIT Press, 103-128.

Hanushek, E. A., and L. Woessmann (2006). Does educational tracking affect performance and inequality? Difference-in-differences evidence across countries. Economic Journal 116, C63-C76.

Henderson, V., P. Mieszkowski and Y. Sauvageau (1978). Peer group effects and educational production functions. Journal of Public Economics 10, 97-106.

Hoffer, T. B. (1992). Middle school ability grouping and student achievement in science and mathematics. Educational Evaluation and Policy Analysis 14, 205-227.

Judson, R. (1998). Economic growth and investment in education: how allocation matters. Journal of Economic Growth 3, 337-359.

Lazear, E. P. (2001). Educational production. Quarterly Journal of Economics 116, 777-803.

Manning, A., and J.-S. Pischke (2006). Comprehensive versus selective schooling in England and Wales: what do we know? IZA Discussion Paper No. 2072, Bonn.

Meghir, C., and M. Palme (2005). Educational reform, ability and family background. American Economic Review 95, 414-424.

Meier, V. (2004). Choosing between school systems: the risk of failure. Finanzarchiv 60, 83-93.

OECD (2005). Education at a glance, OECD: Paris.

Pekkarinen, T., R. Uusitalo and S. Pekkala (2006). Education policy and intergenerational income mobility: evidence from the Finnish comprehensive school reform. IZA Discussion Paper No. 2204, Bonn.

Puhani, P. A., and A. M. Weber (2007). How persistent is the age of school entry effect in a system of flexible tracking? Mimeo, University of Hannover.

Rees, D. I., D. J. Brewer and L. M. Argys (2000). How should we measure the effect of ability grouping on student performance? Economics of Education Review 19, 17-22.

Schütz, G., H. W. Ursprung and L. Woessmann (2005). Education policy and equality of opportunity. IZA Discussion Paper No. 1906, Bonn.

Summers, A. A., and B. L. Wolfe (1977). Do schools make a difference? American Economic Review 67, 639-652.

Thompson, M. A. (1998). Assessing the economic costs of high school noncompletion Journal of Economics and Finance 22, 109-117.

Waldinger, F. (2006). Does tracking affect the importance of family background on students' test scores? Mimeo, London School of Economics.

Woessmann, L. (2003). Schooling resources, educational institutions and student performance: the international evidence. Oxford Bulletin of Economics and Statistics 65, 117-170.

Woessmann, L. (2007). Fundamental determinants of school efficiency and equity: German states as a microcosm for OECD countries, CESifo Working Paper 1981, München.

Zimmer, R. (2003). A new twist in the educational tracking debate, Economics of Education Review 22, 307-315.

Zimmer, R. W., and E. F. Toma (2000). Peer effects in private and public schools across countries. Journal of Policy Analysis and Management 19(1), 75-92.

Ifo Working Papers

No. 49 Buettner, T. and A. Ebertz, Quality of Life in the Regions – Results for German Counties, September 2007.

No. 48 Mayr, J. and D. Ulbricht, VAR Model Averaging for Multi-Step Forecasting, August 2007.

No. 47 Becker, S.O. and K. Wohlrabe, Micro Data at the Ifo Institute for Economic Research – The "Ifo Business Survey", Usage and Access, August 2007.

No. 46 Hülsewig, O., J. Mayr and S. Sorbe, Assessing the Forecast Properties of the CESifo World Economic Climate Indicator: Evidence for the Euro Area, May 2007.

No. 45 Buettner, T., Reform der Gemeindefinanzen, April 2007.

No. 44 Abberger, K., S.O. Becker, B. Hofmann und K. Wohlrabe, Mikrodaten im ifo Institut – Bestand, Verwendung und Zugang, März 2007.

No. 43 Jäckle, R., Health and Wages. Panel data estimates considering selection and endogeneity, March 2007.

No. 42 Mayr, J. and D. Ulbricht, Log versus Level in VAR Forecasting: 16 Million Empirical Answers – Expect the Unexpected, February 2007.

No. 41 Oberndorfer, U., D. Ulbricht and J. Ketterer, Lost in Transmission? Stock Market Impacts of the 2006 European Gas Crisis, February 2007.

No. 40 Abberger, K., Forecasting Quarter-on-Quarter Changes of German GDP with Monthly Business Tendency Survey Results, January 2007.

No. 39 Batchelor, R., Forecaster Behaviour and Bias in Macroeconomic Forecasts, January 2007.

No. 38 Sülzle, K., Innovation and Adoption of Electronic Business Technologies, December 2006.

No. 37 Overesch, M. and G. Wamser, German Inbound Investment, Corporate Tax Planning, and Thin-Capitalization Rules – A Difference-in-Differences Approach, December 2006.

No. 36 Kempkes, G. and C. Pohl, The Efficiency of German Universities – Some Evidence from Non-Parametric and Parametric Methods, October 2006.

No. 35 Kuhlmann, A., German Productivity – A Reassessment via the New Ifo Productivity Database, October 2006.

No. 34 Kuhlmann, A., What is the X-Factor in the German Electricity Industry?, September 2006.

No. 33 Temple, J. and L. Wößmann, Dualism and cross-country growth regressions, August 2006.

No. 32 Baumann, F., V. Meier and M. Werding, Transferable Provisions in Individual Health Insurance Contracts, July 2006.

No. 31 Abberger, K., Qualitative Business Surveys in Manufacturing and Industrial Production – What can be Learned from Industry Branch Results?, May 2006.

No. 30 Ruschinski, M., Investigating the Cyclical Properties of World Trade, May 2006.

No. 29 Holzner, Chr., V. Meier and M. Werding, Time Limits in a Two-tier Unemployment Benefit Scheme under Involuntary Unemployment, April 2006.

No. 28 Eggert, W. and A. Haufler, Company Tax Coordination cum Tax Rate Competition in the European Union, April 2006.

No. 27 Lachenmaier, S. and H. Rottmann, Employment Effects of Innovation at the Firm Level, April 2006.

No. 26 Radulescu, D.M. and M. Stimmelmayr, Does Incorporation Matter? Quantifying the Welfare Loss of Non-Uniform Taxation across Sectors, March 2006.

No. 25 Lessmann, Chr., Fiscal Decentralization and Regional Disparity: A Panel Data Approach for OECD Countries, March 2006.

No. 24 Fuchs, Th., Industry Structure and Productivity Growth: Panel Data Evidence for Germany from 1971–2000, December 2005.

No. 23 Holzner, Chr. and A. Launov, Search Equilibrium, Production Parameters and Social Returns to Education: Theory and Estimation, December 2005.