I Z A Institute
of Labor Economics
Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17167

What Is Wrong with Ability-Tracking?

Hartmut Esser

JULY 2024

I Z A Institute
of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 17167

What Is Wrong with Ability-Tracking?

Hartmut Esser
University of Mannheim, Mannheim Centre of European Social Research and ZA

JULY 2024

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

What Is Wrong with Ability-Tracking?*

The question of the effects of ability tracking remains unresolved even after decades of research. This is also due to the fact that specific regulations for the stringency of implementing differentiation according to ability and achievement have not been taken into account. The issue is the *conditionalization* of the effects of ability tracking *within* differentiating systems. The question is analysed using the example of the German federal states: There are clear differences in the regulation of differentiation (binding nature of recommendations and/or stronger organizational control of schools) and with the "National Educational Panel Study" (NEPS) a data set is available that contains all the information necessary to identify the effects. Three sequentially related aspects of the effects of stringency are analyzed: Sorting and educational attainment, structuring of school classes and educational achievement in secondary school. For educational attainment, there was an increase in achievement at the end of elementary school and greater equality of achievement at the transition without an increase in social inequality. The structuring of school classes showed a decoupling of cognitive and social segregation. And for achievement in secondary school, was an increase in the effects of cognitive composition without an increase in the effects of social background or social segregation, especially in the lower performance areas. This is practically the opposite of the assumptions known from most international comparative studies, according to which strict differentiation does not improve achievement, but only reinforces the effects of social background.

Keywords: ability tracking, educational achievement, educational inequality, school effects, German Federal States

Corresponding author:
Hartmut Esser
University of Mannheim
Faculty for the Social Sciences
A 5,6, D-68159, Germany
E-mail: hartmut.esser@uni-mannheim.de

* The author would like to thank Hans-Peter Blossfeld, Claudia Diehl, Cornelia Kristen, Clemens Kroneberg, Jeremy Kuhnle, Gianluca Manzo, Stefanie Scherer, Steffen Schindler, Jan Skopek and Werner Raub for their comments and suggestions during the long and varied discussion of the concept and its empirical test, but especially Julian Seuring, without whom the project would not have been possible. I would also like to thank the entire team of the "National Educational Panel Study" (NEPS) in Bamberg, especially Daniel Fuß, who always helped when something needed to be clarified. The theoretical concept and the first results were discussed step by step in a long series of presentations since 2014: in Amsterdam, in Florence, in Luxembourg, in Ascona on Monte Veritá, in Venice several times from 2016, in Lisbon, in Tilburg, in Groningen, in Konstanz, in Cologne, in Bamberg, in Berlin, in Bochum, in Mannheim, in Munich, in Tübingen, in Dortmund, in Jena, at the Leopoldina in Halle and in Potsdam, at the WZB, at the "Sociology of Education" and "Social Inequality" sections of the DGS (German Sociological Association) and at the annual meetings of the GEBF (Gesellschaft für Empirische Bildungsforschung) 2016, 2017, 2019, most recently at the annual congress of the Academy of Sociology 2021 in Leipzig, at the annual meeting of the European Academy of Sociology in October 2021 in Paris, in 2022 at the annual meeting of the GEBF 2022 in Bamberg, in Trento and Venice in 2022 and finally again in Venice in November 2023, as well as, in the meantime, in the field of education policy itself, where there is a process of rethinking whether the previous ideas about the effects of differentiation and integration are really what they have looked like so far.

1. Introduction

The question of the effects of Ability Tracking has always been controversial and, even after a long period of research, cannot really be considered clarified (cf. the overviews in Sörensen 1970, later in Betts 2011, Dollmann 2019, Skopek et al. 2019, Esser 2021, Chapter 1, Terrin and Triventi 2022). There are two *theoretical* approaches. According to the *differentiation position*, the separation into different educational pathways according to cognitive abilities allows curricula to be better adapted to the cognitive differences of children and, due to the greater homogeneity associated with this, easier and more focused teaching, with the result that the achievement level increases for the same effort without reinforcing social educational inequality (Sörensen 1970, Sörensen and Hallinan 1977, following Coleman et al. 1967). According to the *integration position*, the situation would be different and even reversed: peer interactions conducive to learning across the entire range of cognitive abilities and social differences are only possible with cognitive and social heterogeneity, and only in this way can marginalisation and stratification resulting from spatial and social segregation and differences in the quality and social composition of schools and school classes be limited at the same time (Oakes 1985; Zimmer and Toma 2000; Gamoran 2009, Domina et al. 2017, 2019). The *prevailing* view of the *empirical* situation is: Early and strict Ability Tracking reduces access to higher education and does not improve achievement, but only increases social segregation, inequality and injustice in education (see, among others, for *attainment:* Brunello and Checchi 2007, Pfeffer 2008, Bauer and Riphan 2009, Allmendinger 2012, Hadjar and Becker 2016, van de Werfhorst 2019; Traini 2021; for *achievement*: Hanushek and Wößmann 2006, Horn, 2008, Gamoran 2009, Wößmann 2009, Wößmann et al. 2009, Müller and Kogan 2010, van de Werfhorst and Mijs 2010, Chmielewski 2014, Pfeffer 2015, van de Werfhorst 2018, Skopek et al. 2019, Strello et al. 2021, Roller and Steinberg 2022, Terrin and Triventi 2022, Dräger et al. 2023). This could be described as the *standard position*, because it seems to be the largely predominant approach and findings.

However, counterevidence has also been presented from the beginning (Galindo-Rueda and Vignoles 2004, Brunello und Checchi 2007, Waldinger 2007, Betts 2011, Chmielewski 2014, Dunne 2010, Dronkers et al. 2011, 2012, Duflo et al. 2012, Horn 2013, Esser and Relikowski 2015, Sahlgren 2015, Skopek and Dronkers 2015, Korthals and Dronkers 2016, van de Werfhorst 2018, Dollmann 2019, Domina et al. 2019, van de Werfhorst 2022). In particular, it became apparent that empirical studies supporting the standard position often have lacked

centrally important concepts, how above all cognitive ability and cognitive composition of school classes, so that they have may have painted an inaccurate picture of the effects of differentiation. This refers in particular to the overestimation of the influence of social background and of the social composition of school classes (see the overview in Marks 2014 and the ongoing discussions about the effects of social background and cognitive abilities in Betthäuser et al. 2020, Marks, 2019, 2020 or O´Connell and Marks 2021 and genetic anchoring in Klein and Kühhirt 2021, Uchikoshi and Conley 2021, Baier et al. 2022, Mönkediek 2022).

The background to the inconsistencies and ambiguities in the discussion is the particular complexity of the aspects and processes involved. There are three. Firstly, the approaches refer to a set of *common* conditions and effects: social background and cognitive abilities, social and cognitive composition of school classes and school types with effects on teaching and/or peer interactions. However, they derive *different* theoretical predictions for the effects. Secondly, all these aspects have to be theoretically *separated* and *related* to one another. But this has hardly ever been the case. Thirdly, the effects are based on *different phases* in the educational process: Sorting, transition and educational attainment, the resulting social and cognitive structuring of schools and school classes and achievement at secondary school after transition. These are theoretically and empirically different processes, but are often not distinguished, such as attainment and achievement in particular. All this must be observed, brought together theoretically, available in the data and taken into account empirically in the analyses. To date, there is no theoretical modelling that takes these complexities into account, at best, however, for partial aspects, but not in the interplay of the various conditions and relations (as in Brunello and Checchi 2007, Appendix 1; Brunello et al. 2007, Section 4.4 or Brunello et al. 2012, Section 2, Angrist 2014, Section 2, Bowden et al. 2023, Section 2, Dräger et al. 2023, Figure 1). In particular, however, the necessary data are already missing, so that biases and inconsistencies are becoming unavoidable., especially for the international comparative studies such as PISA, TIMSS or PIACC.

In addition, there is another decisive and usually overlooked condition for the effects of ability tracking: The effects of Ability Tracking may depend on whether and to what extent cognitive abilities *actually* have the *primary* influence on the allocation to the various educational pathways and *no* other influences such as social background (Hallinan 1994). This suggests that the alleged positive effects of Ability Tracking may be linked to some *additional* regulations of differentiation *itself*, a particular selectivity according to cognitive abilities, especially

according to *institutionalized* regulations such as the binding nature of school recommendations and the control of organizational processes in schools. In order to correctly identify the effects of Ability Tracking, the effects of such regulations, such as local or regional differences in a country, must therefore be considered in *addition* to those of differentiation alone (as according to the age of the first sorting or the number of options as usual). As far as can be seen, this is not taken into account in any of the country comparisons in the international comparative studies - so that there may be differences in the effects of differentiation *within* the countries that have been masked in the national figures of means and social gradients. And there is indeed some (international) evidence that its more stringent implementation of differentiation leads to more meritocratic sorting, transition and educational *attainment* (Galindo-Rueda and Vignoles 2004, Checchi and Flabbi 2007, Bratti et al. 2012, Esser and Hoenig 2028, Esser 2023, chapter 1 and 3 as to higher *achievement* in secondary school (Dunne 2010, Duflo et al. 2011, Dronkers et al. 2011, 2012, Bol et al. 2014, Cord and Giuliano 2016, Korthals and Dronkers 2016, Esser and Seuring 2020, Esser 2023, chapter 2 and 3) and also in primary school already just *before* transition (Morgan 2012, Jackson 2013, Morgan et al. 2013, Esser 2021, chapter 5).

This could have been the case for the German education system in particular. This system is considered the prototype of a particularly rigid, inefficient and socially unjust tracking system (Skopek et al. 2019, Schindler et al. 2021, Dräger et al. 2023). In the first PISA study in 2000, there was indeed a shocking finding: achievement far below average and an extremely strong influence of social background (OECD 2000). This seemed to confirm the general assumption of the standard position that was already common at the time. But it was a false impression, even then: In Germany, there was and still is a differentiating system across the country, but there are also considerable variations between the 16 federal states according to different regulations: The age at first sorting (10 or 12 years), the number of options (three- or two-tier), but specifically according to the binding nature of the school recommendations and the organizational control for the implementation of differentiation. Accordingly, differences are to be expected in the results of the comparative studies between the federal states and correspondingly to the national averages in international comparison.

Such comparisons were also carried out. For Germany, there was a special analysis for the 16 German federal states for PISA 2000 already (Artelt et al., 2002, Baumert and Schümer 2002). This showed a much more differentiated picture of differentiation: three of the 16 German states (Baden-Württemberg, Bavaria and Saxonia) were equal to or better than the OECD average, in

part not far from Sweden. The other federal states were below this, in some cases quite significantly. One particular result was especially striking: the three best federal states practiced - long before - a particularly strictly regulated form of differentiation: the combination of traditionality in access and regulation in organization, implemented via the two additional implementation regulations mentioned above: the binding nature of school recommendations and stronger organizational control of school processes, including standardization, centralisation and accountability of results (Helbig and Nikolai 2015; Figure 28). This combination is referred to here as the *stringency* of differentiation.

The described pattern of differences in achievement between the German federal states has persisted over time until recently (IQB reports 2009, 2012, 2015, 2018) and has become even stronger in international comparison with the integrated systems: Saxony and Bavaria were better than Sweden and even Finland in 2018, and even after the crises from 2020 onwards in primary and secondary school achievement (IQB 2022, IGLU 2023). In the international comparative studies, all of this remained hidden in the national averages, and the regional differences went largely unnoticed by the public, politicians and educational researchers. The findings on the German federal states thus represent a clear correction of the standard position: there is basically nothing wrong with ability tracking, it would just have to be a differentiation that deserves its designation as *ability* tracking.

The following contribution aims to theoretically systematize and empirically test the Sörensen-Hallinan hypothesis of *conditionalizing* the effects of ability tracking on certain additional regulations to ensure the stringency of implementation. This has two prerequisites: sufficient variations in the organization of differentiation with regulations that can strengthen the stringency of ability tracking and limit interfering external influences, and data that include the constructs and correlations required to identify the effects. Both conditions are fulfilled for the German federal states: There are clearly different regulations for the implementation of differentiation there, and with the 'National Educational Panel Study' (NEPS), a data set has been available (since around 2016) that contains the theoretical constructs and information required according to the Sörensen-Hallinan concept for the empirical identification of the system effects of stringency.

To the best of our knowledge, this is the first time that an analysis has ever included all the information needed to identify the system effects of ability tracking, in particular the additional conditional regulations on stringency in the implementation, but also on the effects of cognitive

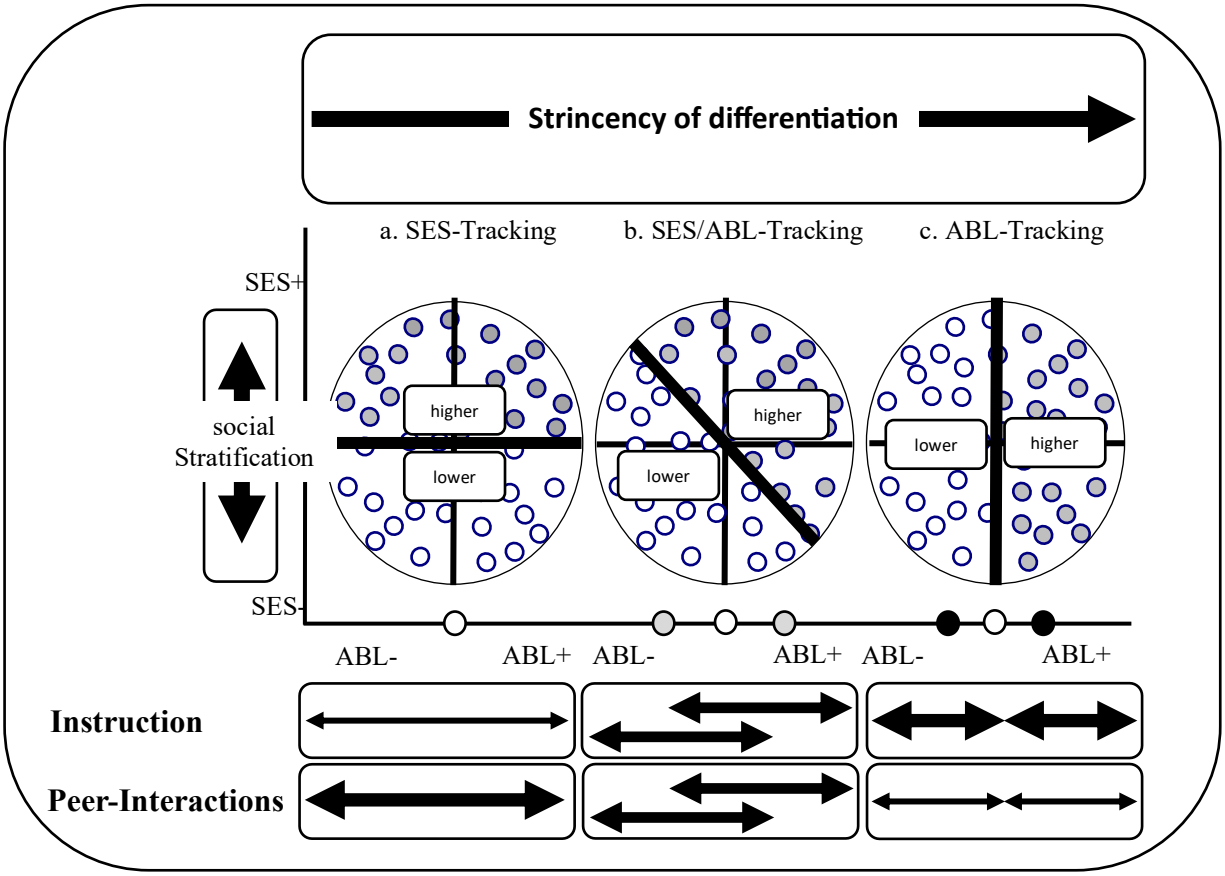
abilities and especially the effects of the social and cognitive composition of school classes, for the niveau as well as for the homogeneity. The effects assumed by the differentiation position were already evident in two, albeit still incomplete, preliminary analyses comparing the German federal states (Esser and Relikowski 2015, Skopek and Dronkers 2015). This is confirmed in the following analyses using the NEPS data, albeit on a not particularly strong basis of case numbers, both overall and for the school classes. The most important limitation, however, is the restriction to the German federal states and its differentiated system. An international comparison to countries without any tracking in lower secondary schools would have been more than desirable. Unfortunately, this is not possible because the necessary data are lacking, and probably will be for the foreseeable future. But the described variance in stringency *within* an otherwise *differentiated* system can also be seen as a variation in the degree of integrative opening and, at least theoretically, extrapolated to full integration, at least as long as a direct comparison is not possible.

The contribution is organized as follows. Section 2, immediately following, presents the needed theoretical framework for the interplay of attainment, structuring and achievements in secondary schools and Section 3 summarizes the hypotheses that can be derived from it. Section 4 describes the data, operationalizations and analyses, including the allocation of the German federal states to the additional regulations mentioned above. The findings for the three aspects are presented in Section 5: for sorting, transition and educational attainment section 5.1, for the social and cognitive structuring of school classes section 5.2 and section 5.3 for achievement at secondary school. Section 6 deals with some of the limitations of the studies, and Section 7 provides a brief evaluation, including a look at developments and the current status of the question of the effects of ability tracking compared to integrated systems.

2. The theoretical framework

The starting point for the theoretical systematization is the details of the two theoretical positions and the set of conditions and processes they share, as set out in the introduction: social background and cognitive abilities, social and cognitive composition of school classes and school types with effects on teaching and peer interactions. The relationships between the interplay of conditions and processes can be illustrated in two rather simple sketches (Figures 1 and 2), the specifications of the “Model of Ability Tracking” (abbreviated as MoAbiT; following Esser 2016a, b and Esser 2021, 2023).

Figure 1: The interplay of differentiation rules, classroom structure, learning conditions and achievement at secondary school.



The diagrams show the distribution of children in a school context according to social origin (SES) and cognitive abilities (ABL) in three constellations (1a, 1b and 1c). An even distribution of cognitive abilities according to social origin is assumed, i.e. equality in *opportunities* after controlling for the primary effects of social origin, as is also shown empirically when cognitive abilities are controlled for (Marks 2014, 2019, 2020).

There are two school types (higher track for the *Gymnasium*, lower track for the *Hauptschule/Realschule* in the German system of differentiation). The division is indicated in the three constellations by the bold lines, and those who attend the higher track are marked in dark gray, the others in light gray. The points on the ABL axis indicate the mean values of the cognitive abilities within the respective school types or the respective school classes belonging to them, i.e. the cognitive level to which the instruction should be oriented in each case, highlighted in light, gray and black for the three constellations. They represent the differences

between the school classes. The double arrows then indicate the different variances or ranges of the cognitive abilities within the school types or school classes. These should be covered by the instruction and describe the opportunities for peer interactions across the variance of cognitive abilities within the respective school classes.

Two extreme cases can be distinguished: Sorting exclusively according to cognitive abilities (ABL+ vs. ABL-), i.e. perfect *ability* tracking (Figure 1c), versus a pure status-based distribution exclusively according to social origin (SES+ vs. SES-), i.e. a type of socially exclusive "nobility" tracking (Figure 1a). When sorting only by social origin, cognitive heterogeneity is maximized. Differentiating only by cognitive ability (Figure 1c) changes this structurally: there are now school types or school classes with significantly different cognitive levels and maximum cognitive homogeneity in each case. A purely cognitive differentiation would also maximize the social diversity: If only achievements count, social background fades into the background. In the middle (Figure 1b) is the case in which cognitive ability and social background influence sorting *together*, i.e. as a collider variable, for given primary effects generated via the secondary effects of parents' educational decisions that differ according to social background, but also via the tertiary effects of the different assessments by schools according to social background, particularly in the grades and recommendations.

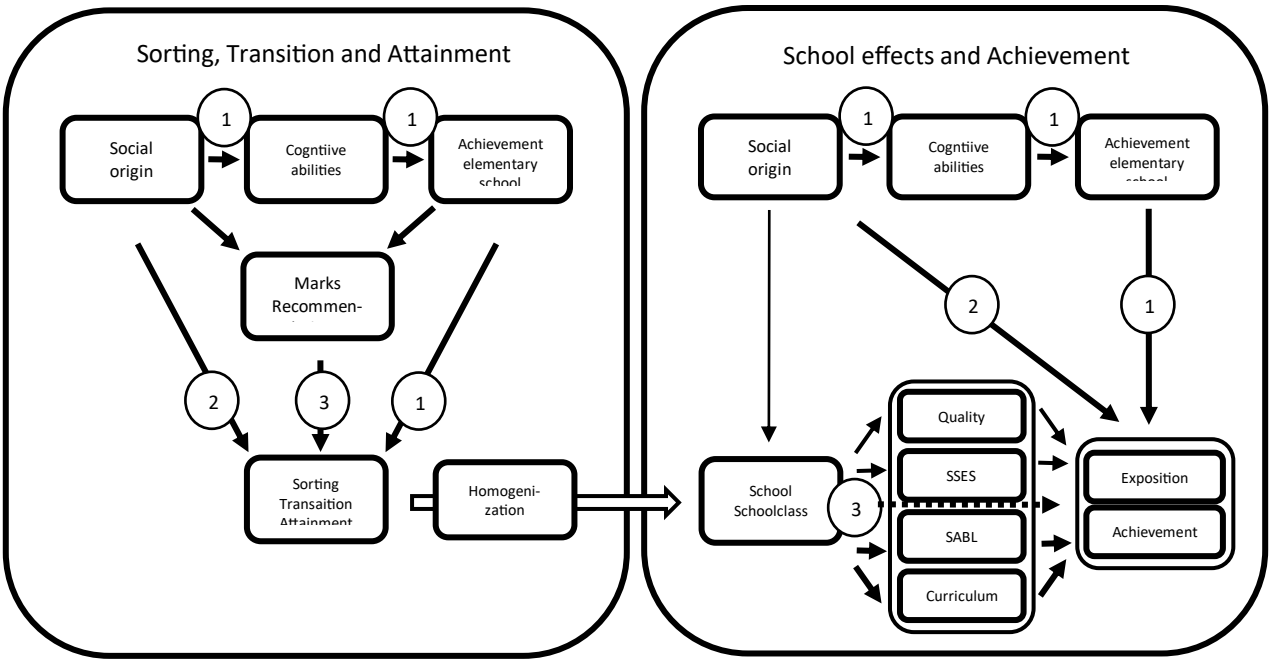
The relationship of learning environments and achievement to system effects can now be simply stated: With stringency, sorting, transition and educational attainment become more meritocratic and this shifts the composition of school classes left to right from social stratification, social homogeneity and cognitive heterogeneity to cognitive differentiation and cognitive homogeneity of school classes with a simultaneous decrease in social stratification and segregation. This has two consequences: *Instruction* becomes more efficient via the now possible adaptation of curricula and the possibility of a stronger focus, while the opportunities for overarching *peer interactions* decrease (cf. the arrows of varying strength below the distributions for the school classes). And vice versa. Because both effects can exist empirically simultaneously, a test of the MoAbiT would therefore depend on the *gross* effect. A positive effect of stringency (after controlling for other relevant influences) would support the differentiation position, while a negative effect would support the integration position.

Behind the processes shown in Figure 1 lies the empirical sequence of the different processes of the educational trajectory in sorting, transition and educational attainment, structuring of

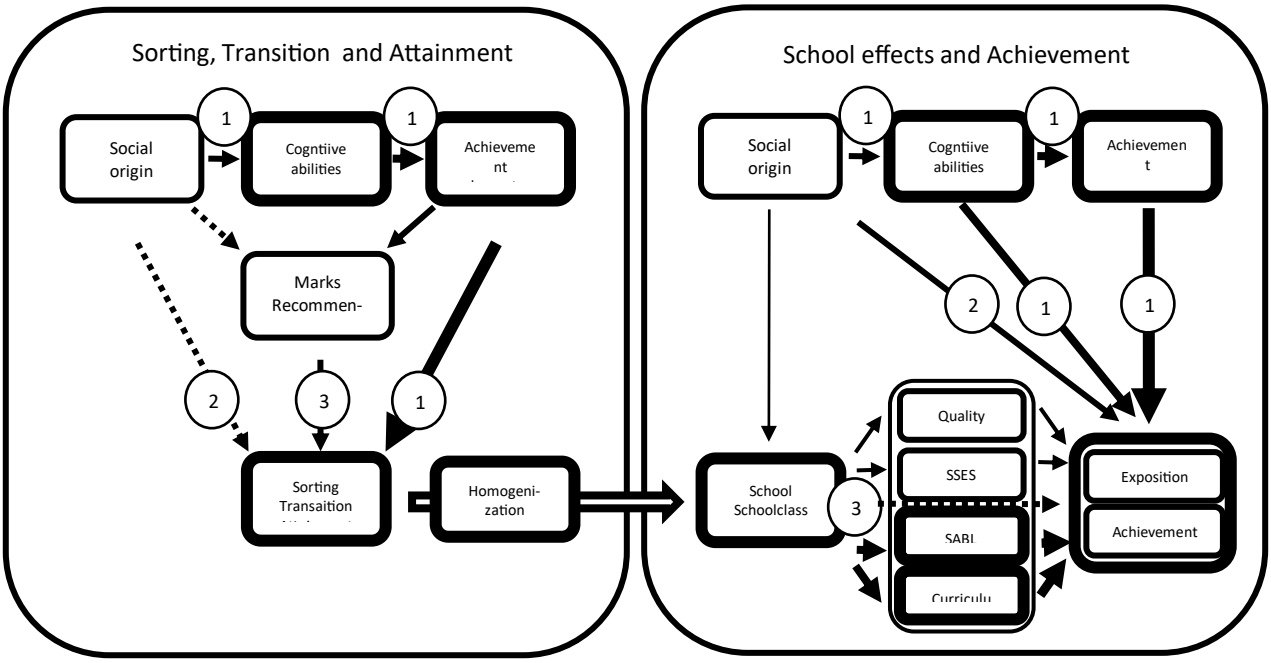
school classes and achievement at secondary school. This is outlined in Figure 2. Part 2a. describes the *general* relationships as they exist in principle in all systems, part 2b describes the *system* effects, i.e. the changes in the general relationships with the characteristics of the respective education system, in this case: for stringency. The numbers 1, 2 and 3 indicate the allocation of the effects to the three basic processes in the interaction of learning by children, activities of parents and of schools, respective teachers (see Esser 2021, chapter 2 following and extending the wellknown distinctions by Boudon): primary effects (social conditions and cognitive development), the secondary effects (effects of educational decisions and family support) and the tertiary effects (expectations, assessments and activities of teaching staff).

Figure 2: General relations and system effects in explaining achievement in secondary education according to the Model of Ability-Tracking.

2a. General relations



2b. System effects



In *general*, sorting, transition and *educational attainment* in differentiation are primarily determined by achievement in elementary school, which in turn depends on cognitive abilities and cognitive development in the family and thus on social background. Secondary effects, if there is a choice, are educational decisions and the family's choice of school, as well as tertiary influences such as the assessment of achievements via grades and recommendations. The *structuring* of school classes is largely an analytical consequence of patterns in educational attainment: Varying degrees of social and cognitive homogenization depending on the accuracy and objectivity of cognitive placement as a consequence of the primary, secondary and tertiary effects before (see the block arrow “homogenization” between educational attainment and school/school class). The social composition of school classes is always also an indicator of vertical differences in school classes in terms of facilities, resources and quality, and possibly also in terms of stigmatization and devaluation. Finally, *achievement at secondary school* depends primarily on prior achievement and cognitive ability, secondarily on further family support and thirdly on school effects, i.e. teaching and peer effects. in each case determined by the structures of the school learning environments: the social and cognitive composition of school classes according to niveau and homogeneity, and thus also according to quality and equipment on the one hand and the change in curricula and the focus of teaching and peer interactions on the other.

The *system effects* are marked in part 2b by highlighting the conditions (edges) and relationships (arrows) that have changed as a result of the stringency regulations. According to the theoretical justifications of differentiation, stringency enhances early cognitive development in the family, as well as achievement in elementary school and thus the primary effects of sorting and transition and achievement in secondary school. At the same time, the secondary effects are weakened (cf. the broken arrows from social background to the assessments in grades and recommendations and to educational attainment). This may lead to a reduction in misplacements and thus to greater cognitive homogenization without increasing social segregation (cf. the block arrow again). The effects of cognitive and social composition can thus be decoupled via stringency. For the school effects themselves, it is ultimately assumed that the effects of the social composition of school classes do not change, but that the effects of cognitive composition, both for niveau and for homogeneity, increase and that this then leads to better achievements via better adapted and more focused instruction.

3. Hypotheses and empirical approach

Table 1 summarizes the hypotheses for the three central processes of sorting, transition and educational attainment (columns 1 to 3), the structuring of school classes (columns 4 and 5) and achievement in secondary school (column 6). They follow the assumptions of MoAbiT as outlined above. The upper part of the table shows the non-conditional *general* effects, while the lower part shows the *system* effects. The system effects are particularly important for the test of the model (highlighted in grey). Columns 1 to 3 describe the hypotheses for the processes before the transition: cognitive development and achievement in primary school (ACE), school recommendations (REC) and the transition itself (GYM), columns 4 and 5 those for the social and cognitive structuring of the school classes (NSES and HSES or NABL and HABL) and finally column 6 for achievement in secondary school (ACS). Listed in each case are the expected direct effects after controlling for preconditions; possible indirect effects are indicated via parentheses. Some of the possible influences are not considered further because they would lead to over-controls (see Elwert and Winship 2014: 35f., Rohrer 2018: 38, Cinelli et al. 2001: 6): At recommendations (REC, column 2) and transition (GYM, column 3), the grades because they are already formally included in the recommendations in some federal states (Esser 2023: section 5.1), for achievements in secondary school (ACS, column 6) the achievements before, the grades, recommendations and aspirations, because the changes in the *efficiency* of learning that are decisive for the system effects relate solely to this, regardless of how they are transmitted. (cf. Esser 2023, section 7.1).

Table 1: Hypotheses for the general conditions and processes on sorting, transition and educational attainment, structuring of school classes and achievement in secondary school and the system effect of stringency on these in the implementation of differentiation according to the MoAbiT.

		Sorting, transition and attainment			Structuration schoolclasses		Achieve-ment
		1	2	3	4	5	6
	Abbrev iations	ACE	REC	GYM	NSES HSES	NABL HABL	ACS
general relations							
individual effects							
socal background	SES	**(+)**	**(+)**	**(+)**			**(+)**
cognitive abilities	ABL	+	**(+)**	**(+)**			+
Achmnt. elem. school	ACE		**(+)**	**(+)**			**...**
Marks	MRK		**...**	**...**			**...**
Recommendation	REC			**(+)**			**...**
Aspirations	ASP			+			**...**
School effects							
Niveau SES	NSES						+
Homogeneity SES	HSES						+
NSES*HSES							≥0
Niveau ABL	NABL						+
Homogeneity ABL	HABL						+
NABL*HABL							≥0
school type	GYM						≥0
System effects							
Stringency (uncond.)	(T)	≥0	≤0	≤0	0	0	≥0
T*SES		≤0	≤0	≤0			≥0
T*NSES						-	0
T*HSES						-	0
T*NSES*HSES							0
T*ACE/T*ABL		+	+	+			≤0
T*NABL					-		+
T*HABL					-		+
T*NABL*HABL							0
T*GYM							≥0

For the aspect of sorting, transition and educational attainment, the *general* conditions for cognitive development and achievements in elementary school (ACE; column 1),

recommendations (REC; column 2) and transition (GYM; column 3) all have a *positive* effect, both directly and indirectly. For the *system effects*, there is a non-conditional effect equal to or greater than zero (T≤0 in column 1) and an interaction effect of T with the social background of less than or equal to zero (T*SES≤0) as well as a positive system effect with the achievements before (T*ACE): with stringency, the achievements increase even before the transition. For the *recommendations* (REC; column 2), a non-conditional effect of zero or less is expected with stringency (T≤0): Recommendations decrease with stringency. The interaction effect of stringency with social background on recommendations would tend to be weaker (T*SES≤0) and that with prior achievement (T*ACE) would tend to be stronger. The recommendations would therefore be fairer according to social background and prior achievement. The pattern is repeated for the *transition* to the Gymnasium (GYM, column 3): a reduction in the transition as a non-conditional effect of stringency, a rather decreasing effect of social origin (T*SES) and a stronger link between the transition and previous achievement (T*ACE).

Only the system effects are relevant for the *structuring* of the school classes (columns 4 and 5): How do the respective covariations of the cognitive and social composition of the school classes change with stringency? The hypotheses according to the MoAbiT are simple: there is a *de*coupling of the cognitive from the social structuring with stringency and vice versa, both for the niveau and for the homogeneity (NSES on NABL, HSES on HABL, NABL on NSES, HABL on HSES).

On *achievement* in secondary school (column 6), social origin, cognitive ability, and prior achievement have as *general* conditions a significant positive effect for the general conditions and processes. For the school effects positive effects can be expected for social niveau and social homogeneity insofar as empirically the equipment of schools and the quality of teaching covary with social segregation. No effects on achievement are expected for the interaction of social niveau and social homogeneity because, according to the theoretical model, this should have no relation to learning efficiency via the general effects when controlling for cognitive ability. Arguably, however, there should be additional positive effects for the interaction of cognitive niveau and cognitive homogeneity: As heterogeneity in school classes is reduced, the efficiency of learning becomes even stronger under favorable conditions, as hypothesized by the aptitude-treatment interaction concept of learning (see Esser 20121: Section 4.1). This is also true for the effect of school type (GYM). It is hypothesized that beyond the clustering of favorable cognitive conditions, there is an additional increase in efficiency related to particular

motivations and a more stable learning environment - both above and below in the niveau of cognitive composition as a consequence of certain specialization gains for all (see the different "Dimensions of Organizational Differentiation in Schools" of Tracking as curriculum, exclusiveness, stability and „scope" of fields by Domina et al. 2019).

For the *system* effects of stringency, the (unconditional) main effect of stricter differentiation (T) is expected to be an improvement in achievement levels in general. Values of the effect at least greater than zero are expected for T after *mediation* of the other influences, but no significantly negative ones. For the interaction effects of the system properties with the general conditions in *moderation*, there would be no changes in the effects of the origin-related conditions with stringency T as a conditional effect. In contrast, the influences of the cognitive composition of the school classes should become larger, those of the niveau as well as those of the homogeneity. A positive system effect of its own is also expected for the Gymnasium as a school type: The effects of motivation, stability, exclusiveness and adaptation of the curricula as further "Dimensions of Organizational Differentiation in Schools" should be even higher with stringency than they already are in general. However, this effect should become smaller with the control of cognitive niveau and homogeneity: The cognitive *structures* bear a significant part of the effect of school type, and when they are controlled, the system effect of school type should become smaller.

There are thus *two* crucial empirical estimands for the test of the theoretical model in the decisive criterion of achievement in secondary education. 1. For the *non-conditional* effect of stringency in mediation, effects equal to or greater than zero are expected. 2. For the respective *conditional* system effects, a *three*-way interaction between cognitive niveau, homogeneity, and stringency (T*NABL*HABL) achievement is expected to *improve* with cognitive homogeneity in combination with educational system stringency, but possibly in different ways in interaction with cognitive niveau. However, the direction of changes conditional on niveau is left open. Consistent with the theoretical model are scissor effects such that differences *increase* upward, because that is an implication of the assumed increase in learning efficiency with cognitive homogeneity and adapted curriculum and instruction. But with the gains in specialization possible with stringent differentiation in the presence of cognitive homogeneity in school classes with low niveau, the reverse may also be true: Stringency may especially help children at *low* levels of achievement.

4. Data, Operationalizations and Analyses

The theoretical considerations apply in a general sense, i.e., also to fully integrated systems. As mentioned, the international comparative studies did not allow for the theoretical model to be empirically tested because the necessary data were missing. Rather coincidentally, data from the "National Educational Panel Study" (NEPS*;* Blossfeld et al. 2011) were available from 2016 to 2018, covering early developments in family and primary school, the transition to secondary school and achievement in secondary school after two years for the starting cohort in 2011/12. In addition, for this cohort, the 16 German states differed considerably in the regulations that determined the actual implementation of differentiation: Binding and control according to the concept of stringency in the theoretical model. This, in turn, allowed the model to be empirically tested, albeit only conditionally on the differentiation otherwise generally applied in Germany. Here is a description of the most important details concerning the data basis, the operationalizations and the analysis strategy.

4.1 Data Basis and Analysis Sample

The cohort in the analyses considered here follows students who attended fifth grade in Germany in the 2010/2011 school year in annual surveys and competency tests.[2] The variables used are from the first to third waves of a written survey and ability test of the students and a telephone survey of one guardian each. The three late-sorting states of Berlin, Brandenburg, and Mecklenburg-Western Pomerania were not included in the analyses because the necessary data for transition and achievement were not yet available in the third wave. From the remaining 13 states, 4719 students participated in the third wave of the survey. After excluding cases without available results from the achievement tests in grade 7 (N=769) and those with incomplete data for the model variables (N=1288), the analysis sample included 2636 students from 181 schools and in 313 grades.[3]

---

[2] Start Cohort 3, Grades 5-7, doi:10.5157/NEPS:SC3:8.0.1. NEPS data were collected from 2008 to 2013 as part of the Framework Program for the Promotion of Empirical Educational Research, which was funded by the German Federal Ministry of Education and Research (BMBF). Since 2014, NEPS has been continued by the Leibniz Institute for Educational Progress (LIfBi) at the Otto Friedrich University of Bamberg in cooperation with a Germany-wide network.

[3] The high proportion of missing data due to incomplete information is a result of the relatively low willingness to participate on the part of the parents. For about 30% of the students, the parent survey was not completed, which is why some key information is missing for these cases, such as information on social and ethnic background, which was recorded exclusively via the parents in grade 5. Additional control analyses, in which missing values from later waves were replaced or taken into account by means of multiple imputation, provided essentially

4.2 Variables and distributions

An overview of the variables used and key metrics for them can be found in Table 2. The listing follows the order of the constructs in the model of Ability-Tracking shown in Table 1 in Section 3 above.

Achievement in secondary education (ACS), the dependent variable of interest, was measured in seventh grade via competencies in reading comprehension and mathematics. Both competency domains were assessed using test instruments developed and tested specifically in NEPS (see Gehrer et al. 2013 for reading comprehension and von Neumann et al. 2013 for mathematics). The WLE estimators of the two competency tests form the basis for the index of achievement, which was calculated as the individual mean of both competency domains. The results of the administration tests were z-standardized in advance (M=0, SD=1) to account for the difference in the test scales.[4]

comparable results, as did earlier analyses with a similar survey design (Esser and Relikowski 2015). Even accounting for sample or panel missing values by means of appropriate weights did not change the core findings, which is why the results of the unweighted analyses are reported below.

[4] Separate analyses for reading and mathematics yielded mostly similar patterns of findings, but were somewhat weaker for reading. In particular, the effects of cognitive homogeneity were more pronounced for mathematics than for reading.

Table 2: Range, means values and standard deviations of the constructs.

Constructs	Min./Max.	Means	Standard Deviations
ACS	-3,5/3,8	0,00	1,00
SES	0/1	0,53	0,22
ABL	0/1	0,60	0,21
NSES	0/1	0,55	0,13
HSES	0/1	0,61	0,11
NABL	0/1	0,65	0,17
HABL	0/1	0,49	0,12
GYM (in %)	0/1	58,4	
T: Stringency (%)	1/3		
T1		41,9	
T2		20,3	
T3		37,9	
FEM (in %)	0/1	48,4	
MBG (in %)	0/1	29,8	
PSA	0/83	39,32	10,84
N_Students	2662		
N_Classes	349		

ACH=Achievement grade 7, SES=social background, NSES=Niveau SES, HSES=Homogeneity SES, ABL=cognitive abilities, NABL=Niveau ABL, HABL=Homogeneity ABL, GYM=school type, T=Stringency of Tracking, FEM=female, MBG=Migration Background, PSA=preschool attendance (in months).

Social background (SES) is recorded as the highest status of the parents, measured by the ISEI-88 index (after Ganzeboom et al. 1992).[5] To measure cognitive ability (ABL), the NEPS-MAT developed for NEPS, a matrix test of reasoning (Lang et al., 2014), was administered in fifth grade. The mean values of social background and cognitive ability represent the social and cognitive niveau of school classes (NSES and NABL, respectively) in secondary school, and the corresponding standard deviations represent the respective homogeneity (HSES and HABL, respectively), with higher expressions corresponding to increasing homogeneity. The school type (GYM) is operationalized as a dummy of the type of school attended at the beginning of fifth grade (Gymnasium with 1 vs. no Gymnasium with 0), with children attending the

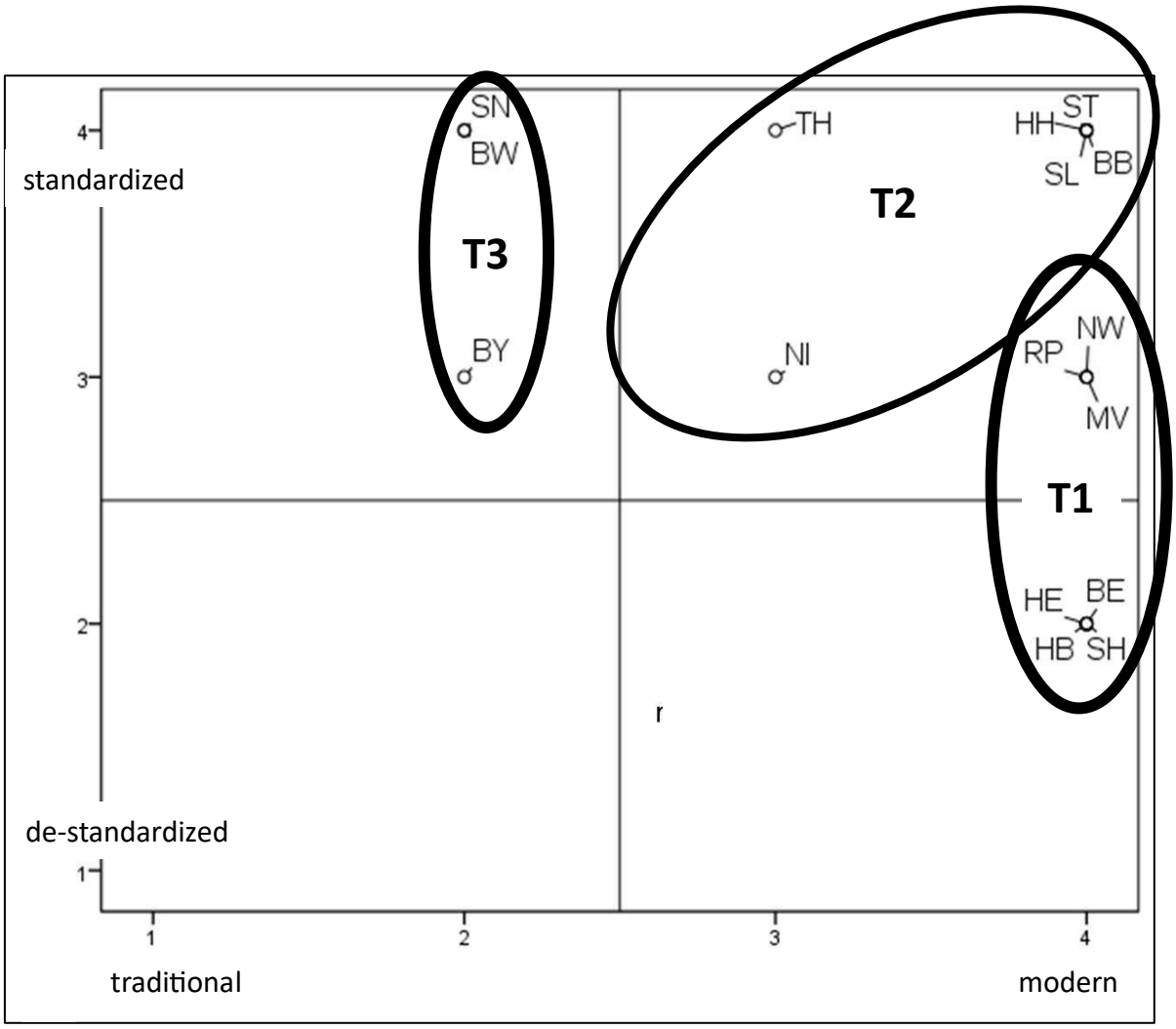
[5] For the purpose of comparison, analyses were also performed with the highest level of education in the family according to the CASMIN index. In the main structures, there were no differences regardless of whether parental education was considered separately from or with status. To avoid overcontrol, only the HISEI index was used in the analyses.

Gymnasium branch of a cooperative comprehensive school also classified as Gymnasium students.

4.3 Stringency regulations in the German federal states

Similar to the Netherlands, Austria and Switzerland, Germany is one of the countries with a differentiated education system with spatial and organizational separation of educational pathways after elementary school. The specific organization is a matter for the 16 federal states. There are clear differences in the specific regulations: The age at first division, the number of options, offers of integrated schools and also in the regulations of stringency on bindingness and control, which are the focus of the analyses here. The core of the analyses here is a classification of stringency (T) consisting of three values, formed according to a typology in von Below (2011, section 4.3) and a further expanded classification in Helbig and Nikolai (2015; Figure 28: 286). This distinguishes between two fundamental dimensions: modernization vs. traditionalism and de-standardization vs. standardization, each with 4 values. Figure 3 shows the localization of the 16 federal states in the space of these dimensions for the year 2010/1, the period to which the analyses in this article also refer.

Figure 3: Stringency as combination of (de)traditionalization and (de)standardization in the German federal states for the year 2010/11; own illustration; after Helbig and Nikolai 2015: Figure 28: 286.



BB=Brandenburg, BE=Berlin, BW=Baden-Württemberg, BY=Bavaria, HB=Bremen, HE=Hesse, HH=Hamburg, MV=Mecklenburg-Western Pomeraia, NI=Lower Saxony, NW=Northrhine-Westfalia, RP=Rhineland-Palatinate, SH=Schleswig-Holstein, SL=Saarland, SN=Saxony, ST=Saxony-Anhalt, TH=Thüringia.

In 2010/2011, Baden-Württemberg, Bavaria and Saxonia are the three states with the highest stringency (index value T3), Berlin, Bremen, Hesse, Mecklenburg-Western Pomerania, North Rhine-Westphalia and Rhineland-Palatinate are the most open and liberal (T1), with all other states in between (T2). Three federal states are not included in the analyses: Brandenburg, Berlin and Mecklenburg-Western Pomerania. Here, the transition to secondary school does not take place until the 6th grade (at the age of 12) and no corresponding data was available at the time of the analyses.

Table 3 shows the distributions for the 13 states included in the analyses, with stringency for the index according to Helbig and Nikolai 2015 with five expressions between 4 and 8 (with 4 as the lowest value and 8 as the highest value). Based on these five values, the states were divided into three categories. They correspond to combinations of the two specific regulations mentioned in the model: The binding nature of the recommendations to limit secondary effects and the control of school procedures via standardization and regular evaluations to limit tertiary effects. The two types of empirical conditions can thus be assigned to the three types theoretically delineated in Section 3 via specific institutional rules: T1 has neither bindingness nor control, T3 has both bindingness and control, and T2 has only one of each of these conditions (see Table 3).

Table 3: Assignment of the federal states to the values of the index index (from the list in Helbig and Nikolai 2015, Figure 26: 286) and the three types of stringency.

Index Stringency	4	5	6	7	8
	liberal				stringent
	BE HB HE SH	*MV* NW RP	*BB* HH NI SL ST TH	BY	BW SN
n5	352	762	539	354	655
Type Stringency	**T1**		**T2**	**T3**	
n3	1114		539	1009	
N	2636				

BB=Brandenburg, BE=Berlin, BW=Baden-Württemberg, BY=Bayern, HB=Bremen, HE=Hessen, HH=Hamburg, MV=Mecklenburg-Vorpommern, NI=Niedersachsen, NW=Nordrhein-Westfalen, RP=Rheinland-Pfalz, SH=Schleswig-Holstein, SL=Saarland, SN=Sachsen, ST=Sachsen-Anhalt, TH=Thüringen; BE, BB, and MV excluded from analyses as late-sorting countries (each indicated in italics; case counts exclude these three countries); n5=case counts for the five categories between 4 and 8; n3=case counts for the three types/categories included in the analyses here).

The reason for *not* operationalizing stringency as an index and continuous variable but as classification of “types” of different rules for the implementation of *Ability* Tracking is a theoretical peculiarity of institutional regulations. Following Lazarsfeld's well-known distinction, these are "global" characteristics that affect *all* individuals within the respective context – equally and independently from other traits. Therefore. Such types do *not* involve a further level of aggregation, the decomposition of within-group and between-group variations and usual "multi-level" analysis. This is different with school classes: niveau and homogeneity for social background or cognitive abilities are "analytically" formed from the characteristics of the *individuals* as context characteristics, so that variance decompositions make sense for them. The number of units at the (federal) state level is therefore also irrelevant for identifying

the effects of certain regulations. It would become significant if the analyses were conducted at the macro level of aggregate data and there were no controls for the distributions of individual conditions and school classes as in some analyses of the standard approach. Such analyses assume that there are no changes in the individual relations with the system conditions, but this is precisely what is envisaged for the system effects (according to the covariance theorem).

In all analyses, we control for child gender (FEM), migrant background (MHG, with 1 vs. 0), and length of preschool attendance before enrollment (PSA), because the main competing general socio-demographic influences on achievement should be found in these, but also important aspects of state and residential district socio-structural differences. However, no further macro-conditions are controlled for the country classifications according to stringency, such as the degree of urbanisation, social segregation or the respective GDP: all such possible effects are more than adequately controlled for in the NEPS via the controls for individual conditions (such as social origin, migration status and pre-school attendance) and the characteristics of the immediate school learning environments (social composition of school classes according to niveau and homogeneity). Further controls would only unnecessarily weaken the stability of the estimates.

No other characteristics of the federal states are taken into account, such as the two- versus three-tier system or the age at the first sorting. The latter is excluded anyway because the analyses refer to the seventh grade and there is no data on this in the NEPS (see above). The combination of stringency and number of options would fall well below the limits of the number of cases. Stringency is simply used to analyze another criterion of differentiation that can be compared with the effects of other classifications within the given framework of the number of systems and cases compared, but not in combination.

4.4 Estimations

To determine the effects, OLS regressions were calculated for metric variables (achievement in elementary and secondary school, social und cognitive composition of school classes) and logistic regressions for categorical variables (recommendations, transition). In the empirical analyses of system effects on achievement at secondary school, the focus is on the achievement of a given educational standard for verbal and mathematical skills in a comparison between systems. To test the theoretical assumptions for school effects, hierarchical linear regression

models were estimated with students at level 1 and school classes at level.[6] To quantify the effect sizes, the dependent variable was z-standardized. Thus, the regression coefficients can be interpreted as achievement differences in standard deviations. Furthermore, all continuous independent variables, both individual and context characteristics, were transformed so that 0 corresponds to the lowest empirical value and 1 to the highest empirical value, respectively. The regression coefficients thus show the maximum effect of the respective indicator. This procedure also facilitates the interpretation of the interaction terms: The conditional main effects each represent the effect of x1 at minimum expression of x2, and vice versa. The interaction effects then represent the change to the conditional main effects when x1 or x2 assume the highest empirical expression.

5. Results

The overarching goal of the empirical analyses is to identify the *system* effects of the stringency of ability tracking on sorting, transition and attainment, on structuring of school classes and on achievement in secondary school (see also section 4 above on the hypotheses). The results for all three aspects are presented in two steps: 1. the *non*-conditional system effects of stringency under control of the general conditions according to the theoretical model (mediation) and 2. the *conditional* system effects of the change in the general contexts in their *interaction* with the stringency (moderation). The analyses follow the order of the constructs in the hypotheses in Table 1 above: Effects of stringency on cognitive development and achievement in elementary school, recommendations and sorting, transition and educational attainment (Section 5.1), on the social and cognitive structuring of school classes (Section 5.2) and on achievement in secondary school (Section 5.3).

[6] For the identification and causal interpretation of school effects, there is a discussion that has gone on for quite some time, the core of which is that context effects can be overestimated if the context characteristics are formed from the respective dependent individual variables and/or the respective individual variables exhibit measurement errors (Sacerdote 2011, Angrist 2014, Pokropek 2014, Marks 2015, Feld and Zölitz 2017). Here, this risk is to be assessed as rather low, at least in comparison to the studies to which the replication of the standard approach applies here: to characterize the achievement-related composition of school classes, the dependent variable of achievement is not used, as is often the case, but that of cognitive abilities. Therefore, the so-called leave-it-out rule is not applied either, which would tend to generate biases in the analyses of school effects here (cf. Esser and Seuring 2023). With the NEPS data, comparatively well validated tests are available, so that the measurement errors should also be smaller than usual. Moreover, the findings prove robust to various other techniques of controlling for other influences and measurement errors: In a recent study using more specialized techniques such as multilevel structural equation modeling and propensity score matching (Lauermann et al. 2019), the same relations are found as in conventionally designed analyses of school effects in differentiated systems.

5.1 Sorting, transition and educational attainment.

The analyses for the first step of the sequence concern the cognitive development of the children in the skills acquired by the end of elementary school (in reading and mathematics), the recommendations at the end of elementary school and the actual transition to different educational pathways. The hypotheses according to the theoretical model (MoAbiT) are shown in columns 1 to 3 in Table 1 in Section 3 above. The corresponding results are summarized in the three columns 1 to 3 from left to right in Tables 4a and 4b.

Table 4a: Achievement at the end of primary school (ACE, OLS), recommendations (REC, logistic regression) and transition to secondary school (GYM, logistic regression), System effects 1: Non-conditional effects of stringency (Mediation, bold: $p<0.05$).
ABOUT HERE

Table 4b: Achievement at the end of primary school (ACE), recommendations (REC) and transition to secondary school (GYM) System effects 2: Conditional effects of stringency (Moderation, bold: $p<0.05$).
ABOUT HERE

5.1.1 Achievements at the end of primary school

The results for achievement at the end of elementary school can be found in column 1 on the left of Table 4a. Social background in general (at 0.14) and cognitive abilities in particular (at 0.33) have a significant influence on achievement at the end of primary school. In addition, there is a significant *non-conditional* system effect of stringency (with 0.03 at T3). This means that children in the respective federal states start at a higher niveau of achievement right from the beginning in secondary school. It is a *separate* system effect to everything else, even before there are any changes to the secondary school later. Behind this are apparently considerations of anticipating the gains/losses if the achievements are sufficient for a positive recommendation or not and if there is bindingness and *not* the option of choosing a school against the recommendations (cf. Bach and Fischer 2021, Grewenig 2021, Osikominu et al. 2021). Further analyses also show that this effect is stronger in *three*-tier-systems than in two-tier systems (Esser 2023, Table 8.2: 471). This would also indicate that the estimated opportunity costs drive efforts at the end of primary school, because the impending losses are even higher with three-tier-systems systems (and stringency).

The influences of social background and primary school achievement (resp. cognitive ability) do not change further in the moderation with stringency (Table 4b, column 1) – with the exception for the semi-liberal systems at T2 (with -0.07 in relation to the main conditional effect at T2 with 0.05).

The result for the system effects of stringency on achievement before primary school is thus easily summarized: It generally raises achievement levels already at the transition with stringency, but then nothing more is added for this in secondary school. This would relate to other processes, especially those that have to do with school effects after the transition (see section 5.3 below).

5.1.2 Recommendations

The assessment of children's achievements in primary school forms the basis of the influence that the *school* has on their subsequent educational trajectory. The schools' recommendations are of particular interest here, as they directly determine the transition. The grades are indirectly in the background and are included in the recommendations; therefore, they are not included further due to the risk of over-control. For the general context, it is known and widely documented that latent cognitive abilities and achievements control the ratings and thus also the recommendations to a very considerable extent and that these are also largely "fair", both socially and in terms of achievement, but that there are also significant inaccuracies and distortions in which children from the lower classes or with a migration background are particularly disadvantaged. But what effect do the rules of stringency as a combination of bindingness and control have on the accuracy and biases in the assessments, especially according to social background? This is described in Tables 4a and 4b in the centre column 2 for the system effects on the *recommendations*, first for the non-conditional system effects after mediation, then for the conditional system effects during moderation.

The general influences show clear effects of social background as tertiary effects via expectations and stereotypes of teaching staff and as primary effects on cognitive abilities in the background rather and specifically on achievement at the end of primary school. This confirms the known results: They exist, the tertiary effects of school assessments, but the much stronger part is accounted for by the meritocratic part, specifically school achievement.

There are still *no* non-conditional system effects in mediation, nor are there any significant changes or conditional system effects in moderation. However, this would not actually be

expected after the MoAbiT: the stringency should ensure that the assessments are socially *fairer* and that recommendations are *more closely* linked to prior achievement (or cognitive abilities) with all the consequences for the cognitive structuring of schools and school classes. However, this is *not* the case with a high degree of stringency: the value of the corresponding coefficient, T3*ACE, is not significant, but at -1.98 it is significant and *negative*. It would appear that the otherwise strict federal states in particular have a lot of catching up to do in terms of accuracy and objectivity of assessments. They could therefore easily become *even* better than they already are.

5.1.3 The transition to secondary school

The right-hand column 3 in each of Tables 4a and 4.b describes the results on transition and educational attainment. Again, the results are summarized briefly. In general, social background, cognitive ability and - in particular - prior achievement in primary school have strong direct (as well as indirect) effects, but so do recommendations once again and aspirations. This is in line with the expectations of the theoretical model and, as far as it has been tested, also with what is known about the general conditions of educational attainment. There are no conditional system effects of stringency. There is also hardly any conditional change with stringency. And there is also *no* strengthening of the effect of origin with stringency. However, there is a *closer* link between achievement in primary school and the transition and educational attainment, although this is (just) not significant (at -1.98 for T3*ACE). This means that the transition becomes more performance-oriented with stringency, i.e. more *meritocratic*.

Overall, however, and in combination with the findings on cognitive development in primary school achievement, this is a truly remarkable result: there is *no* further social stratification in educational attainment with stringency of differentiation, there are *gains* in achievement level right from the start of secondary school and *greater* equity in the distribution across the various educational tracks. In other words, the results are completely different to those assumed by the integration position and the findings of the standard approach.

5.2 The structuring of school classes

Transition and educational attainment generate the school structures at secondary level *analytically*: the cognitive and social composition of school classes are *statistical* consequences of the allocations at transition. The central aspect is the question of

whether and how stringency changes the respective *correlative* relationship for the structures of school classes via educational attainment, on which causal follow-up processes may then be based. In particular, the question is whether the social and cognitive structures of the school classes are more strongly connected with stringency, as assumed by the integration position, or whether decoupling occurs, as postulated by MoAbiT. The effects of these structural changes on achievement at secondary school are then discussed in the following section 5.3.

This concerns the hypotheses in columns 4 and 5 in Table 1 in section 3 above. According to the theoretical model, stringency should lead to a *reduction in* any covariations of cognitive and social structuring. This would be reflected in a *negative* system effect of the covariation of cognitive and social structuring. In the respective tables (5 and 6), therefore, the control of the correlation via some other structural variables of the school classes comes first, the mediation (models 1 to 3), then the moderation of the effect via the interaction with the stringency (model 4). Three variants are specified for mediation in each case: The *overall* effect as it empirically results when nothing else is controlled (Models 1 in each case of tables 5 and 6), then the effect controlling for the *other* structure, NSES on NABL and NABL on NSES, HSES on HABL and HABL on HSES, (Models 2 in each case), and then controlling for class compositions by socio-demographic composition for gender, migration background and pre-school attendance (composition GMV; Models 3 in each case). Model 4 contains the theoretically decisive empirical estimand in each case: the interaction effect of stringency with the other structural condition, which describes the change in the covariation of the social and cognitive structures. It is highlighted in gray for T3, i.e. the strongest stringency. Class size is controlled for in all models.[7]

Table 5 shows the results for structuring at the cognitive and social *niveau* (NABL and NSES), i.e. cognitive differentiation and social stratification of the school classes, viewed once from the cognitive to the social structuring (5a) and then from the social to the cognitive (5b). Tables

[7] In the NEPS data, the sizes of the school classes are smaller than they normally are empirically. Among other things, this is a consequence of dropouts over the three waves of the study. The different sizes could have distorted the effects because the weights of the influences would have been distributed differently across the classes. However, this would only have had systematic effects if there had been selective dropouts according to class size. However, this was not the case (see section 5.3 below). Statistical replacement procedures were generally not used due to the often even more distorting assumptions. This produces rather conservative estimates, but no sign reversal. Since we are not dealing here with causal relationships, the various controls could also be omitted. Variations of the controls showed no evidence of systematic changes, especially not with regard to the sign in the system effects, which are particularly relevant in this case (cf. Esser and Seuring 2023).

6a and 6b show the findings for structuring in cognitive and social *homogeneity* and segregation, respectively.

Table 5a: Cognitive differentiation (NABL) and social stratification (NSES) of school classes according to stringency; NEPS; OLS; control class size; bold: p<0.05; abbreviations in the text.
ABOUT HERE

Table 5b: Social stratification (NSES) and cognitive differentiation (NABL) of school classes by stringency; NEPS; OLS; control class size; bold: p<0.05.
ABOUT HERE

Table 6a: Cognitive homogeneity (HABL) and social segregation (HSES) of school classes by stringency; NEPS; OLS; control class size; bold: p<0.05; abbreviations in the text.
ABOUT HERE

Table 6b: Social segregation (HSES) and cognitive homogeneity (HABL) of school classes by stringency; NEPS; OLS; control class size; bold: p<0.05.
ABOUT HERE

The results for all four constellations in tables 5 and 6 can be summarized briefly. In the case of mediation (models 1 to 3 respectively), there are clear covariations in *niveau*, cognitive differentiation and social stratification (with values of 0.71 and 0.57 non-conditional in models 2 and 3, and 0.64 conditional in model 4). The situation is different for *homogeneity* and segregation. There are no covariations of cognitive homogeneity and social segregation, nor vice versa for social segregation and cognitive homogeneity. This corresponds to the current findings of Brinkmann et al. (2024), according to which the 'selectivity' of individual schools in differentiation alone does not change the cognitive and social structuring as '*homogenization*'.

Irrespective of this, however, all four comparisons show clear *negative* interaction effects of cognitive and social structuring with stringency in the moderation (Models 4 in each case: T*NSES or T*NABL and T*HSES or T*HABL). This means that cognitive differentiation is *de*coupled from social stratification by stringency, in *both* directions and for niveau *and* homogeneity, regardless of the value of the coupling beforehand. This is what the theoretical model says and was specified as a model in Figure 1. In other words, the opposite of what the integration position assumes with the hypothesis that cognitive differentiation and social stratification are as it were already *logically* inseparable also applies here.

5.3 Achievements in secondary school

The analyses for the achievement in secondary school follow again the order of the constructs in the hypotheses in Table 1 in Section 3, column 6: the effects of individual social background on achievement in secondary school first, then those of social composition of school classes, then those of individual cognitive ability and cognitive composition of school classes, finally also those of school type in secondary school. All constellations of school effects are examined in the analyses: niveau, homogeneity, individually and together and also in their interaction. According to the theoretical model (cf. Sections 2 and 3), it is expected that there will be a gradual approximation to the differentiation position with the completion of the model from the standard position to the MoAbiT: a stable, non-conditional system effect of stringency across the general controls of mediation and conditional effects of stringency of the amplification of the effects of the cognitive composition of the school classes in the moderation, but no increase in the influence of social origin or that of the social composition of the school classes. Irrespective of this, positive effects beyond the structural influences of the social and cognitive composition of the school classes are expected for the school type.

5.1 System effects 1: Non-conditional (Mediation)

Table 7 displays the findings on the mediation of the system effects of differentiation. First the analyses on social origin, then on social niveau and homogeneity of the school classes, then analogously on the cognitive influences and finally on the effect of the school type.

Table 7: System effects 1: Non-conditional effects of stringency (Mediation, OLS, bold: $p<0.05$)
ABOUT HERE

There are nine special observations. 1. there is a stable effect of stringency (T3) across all controls. 2. there is also a significant effect of social origin (SES), which is also stable across all controls. 3. the effect of individual social origin decreases (by one third) with the control of the social niveau of the school classes (NSES in model 3). 4. there are no effects of social homogeneity of school classes (HSES in model 3), nor of the interaction of social niveau and homogeneity (NHSES in models 4 and 5). 5. the effect of social niveau is maintained across all further controls, including that of cognitive composition, and only disappears when the school type (GYM) is controlled for (model 11). 6. Individual cognitive abilities have a clear and

additional influence on achievement, but they hardly change the influence of social background (ABL in model 6). 7. the cognitive niveau in the school classes (NABL in model 7) also has a strong influence, the effect of the social niveau is thus strongly reduced. 8. as before with the social composition, the control of cognitive homogeneity or the interaction of cognitive niveau and homogeneity changes virtually nothing (models 8 to 10). The only remaining non-conditional influences are social background, cognitive abilities and the social and cognitive niveau of the school classes (model 10). Social and cognitive homogeneity generally have no influence, and the interaction of niveau and homogeneity does not change this. 9. With the inclusion of school type, this changes again in two places (GYM in model 11): The effect of social niveau disappears completely, but that of cognitive niveau remains. However, this is also greatly reduced (from 2.20 in model 10 to 1.40 in model 11). The school type itself has its own strong positive influence. This points to additional, pareto-optimal advantages of stringent differentiation, quite similar to the findings by Domina et al. (2019: 309ff.) on exclusivity, curriculum, stability and scope.

These findings confirm the hypotheses according to the MoAbiT for a generally and unconditionally positive effect of differentiation and thus correct the standard- as well as the integration position, according to which differentiation does not improve achievement in secondary school, but only reinforces the effects of social background.

The question is then, on what already this positive *non*-conditional system effect is based. According to the theoretical model, it would be reasonable to explain the effect with the improvement of achievement already in *elementary* school (see the analyses above in sections 5.1 on cognitive development and achievement in elementary school): If there is bindingness of recommendations and strong organizational control, the transition depends much more on prior achievement, and this would be a strong incentive to make a special effort beforehand. This is marked in Figure in Figure 2b on the right as the *higher* starting point for achievement just *after* the transition. This means that with stringency from the beginning, even probably long before elementary school, there is a certain *advantage* in achievement even *before* there may be further effects after the transition. There is strong further evidence for such an effect, specifically of the binding rule (Bach and Fischer 2021, Grewenig 2021, Osikominu 2021, Esser 2023, Chapter 5).

5.2 System effects 2: Conditional (Moderation)

In addition to the non-conditional system effects, there may still be changes in the general relationships across the system characteristics. Table 8 reports the corresponding findings. In it, the system effects of stringency T under control of the general conditions and correlations are listed, in each case also for the change in the effect of school type. The order of the analyses is the same as for mediation: the interaction effects of stringency (T) with social background, then with social school effects, with cognitive ability and cognitive school effects, and finally with school type.

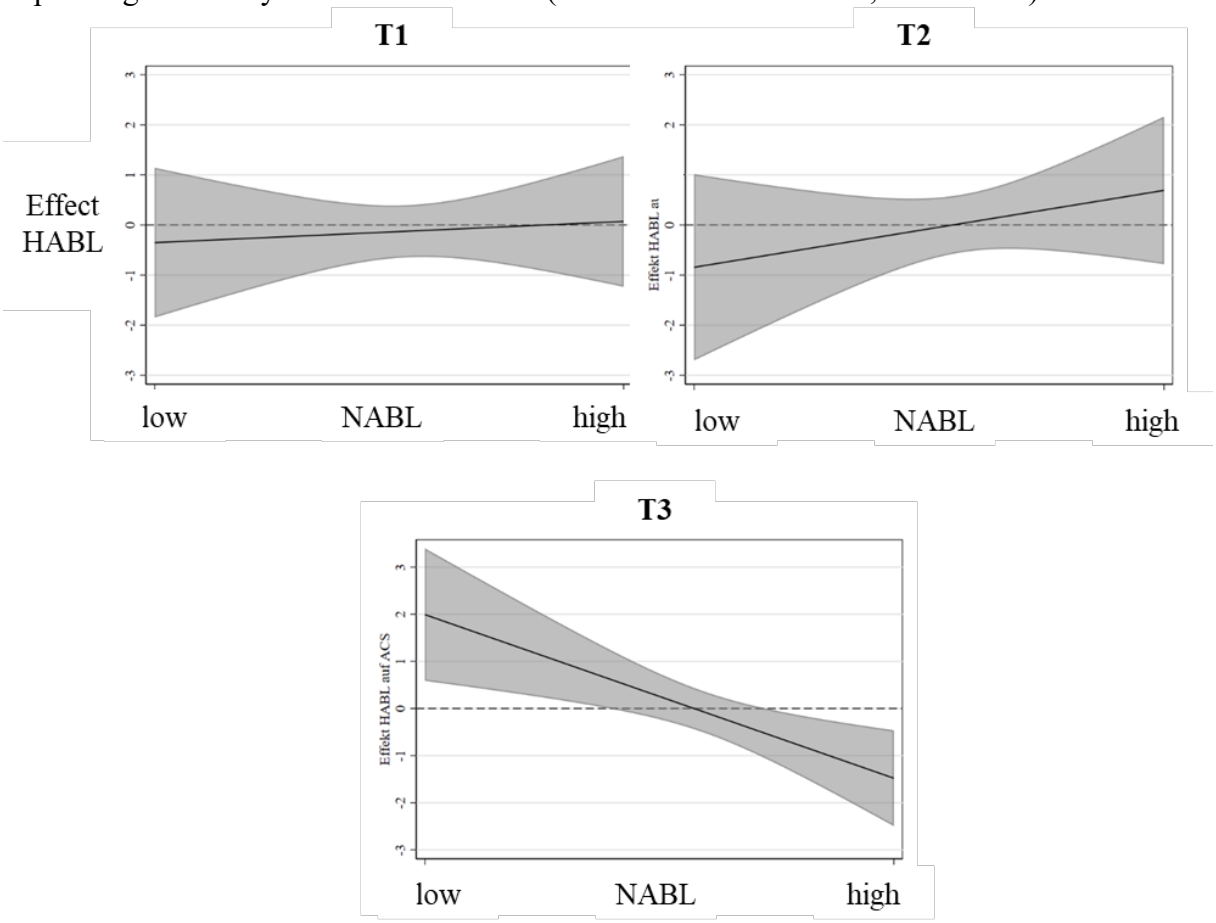
Table 8: Conditional System effects of stringency 1 (Moderation, Interactions, OLS, bold: $p<0.05$)
ABOUT HERE

The results can be summarized in six points. 1. there are no effects of stringency on the effects of social background across all models of moderation, neither individually nor in the school effects 2. the influence of cognitive abilities does not change with stringency, it even tends to decrease (for T2). 3. the cognitive niveau of the school classes, on the other hand, improves achievement consistently and under all other conditions (models 8 to 12). 4. on its own, cognitive homogeneity of school classes has no influence (models 9 and 10). 5 However, this changes with the interaction of stringency, cognitive niveau and cognitive homogeneity: there are clear positive effects of niveau and homogeneity with stringency and an equally clear negative effect of the (triple) interaction (model 10). This means that the positive effects of stringency on niveau and homogeneity are (relatively speaking) higher in the lower performance areas than at the top. 6 If the school type is added (model 11), nothing changes. Compared to the effect of school type only for the social influences, the effect decreases slightly (from 0.29 in model 6 to 0.22 n.s. in model 11). This is an indication that part of the conditional effects of stringency are carried by the special features of the Gymnasium. These are gains that are added and not lost elsewhere, as can be seen in the findings of Model 10 on the triple interaction of stringency, cognitive niveau and cognitive homogeneity.

Taken together, this would mean a further correction of both the integration- and the standard position, especially models 11 and 12 with the theoretically suggested triple interaction. However, this combination of conditional effects of cognitive niveau, cognitive homogeneity, and stringency points to a rather complex pattern of interrelationships. Figures 4 and 5

graphically describe the findings for the crucial three-way interaction of stringency (T) with the interaction of cognitive Niveau (NABL) and cognitive homogeneity (HABL) according to the findings in Model 7 of Table 8. In figure 4, the interaction of NABL*HABL is shown separately according to the three categories of stringency T1, T2, and T3, and in figure 5, as well as according to the low and high cognitive niveau for the combinations of stringency and homogeneity or heterogeneity. This allows us to look at the structures from all sides. We first present figure 4.

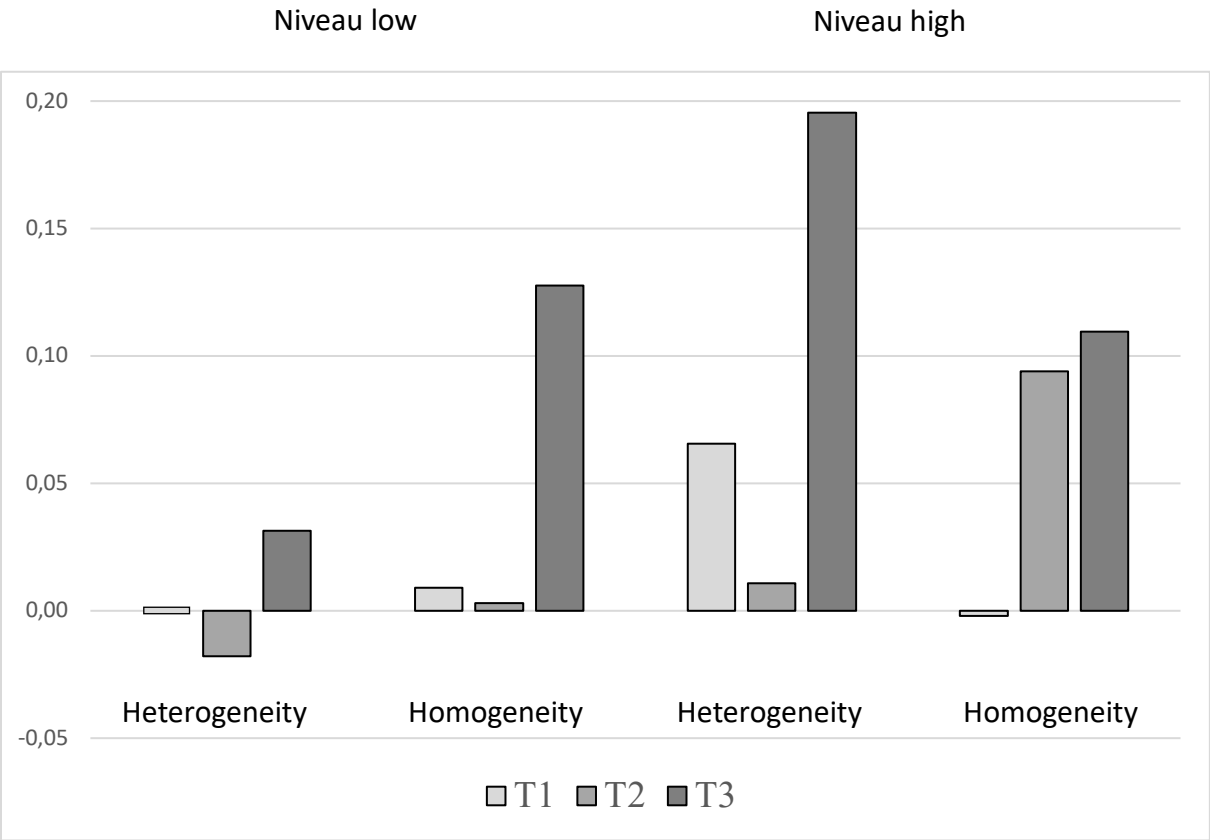
Figure 4: The interaction of cognitive niveau, cognitive homogeneity, and stringency in explaining secondary school achievement (from Model 7 in Table 8, all controls).



The result is clear: under the liberal rules of T1 and T2, the effects of cognitive homogeneity are independent of the cognitive niveau (and vice versa); under stringency in T3, they become larger for children in *low*-niveau school classes. This is an effect of *compensation*, explainable by the specialization gains just at the bottom of the niveau.

This is confirmed in the other plot of the interaction of cognitive homogeneity and stringency with cognitive niveau in figure 5.

Figure 5. The interaction of cognitive niveau, cognitive homogeneity, and stringency in explaining secondary school achievement (from Model 7 in Table 8, all controls).



The graph shows the effects for the combinations of high and low niveau of cognitive ability in the school class, lower or higher cognitive homogeneity, and the three types of stringency (T1, T2, T3). The estimation of achievement by heterogeneity versus homogeneity and low vs. high niveau is based on the values for the respective 10% vs. 90% quantile of the corresponding distribution, i.e., not for the extreme values. In combination with the three systems, this results in the 12 different constellations in figure 5. Accordingly, the values shown characterize the achievement differences of the respective combination for the quantiles compared to the reference category (constellation 1: liberal regime T1, low cognitive niveau and heterogeneous cognitive composition). The main result is that T3 has the strongest upward effect in all constellations of niveau, homogeneity and stringency. However, the asymmetry of the effects of homogeneity and heterogeneity according to niveau is also recognisable: at the *bottom* there are positive effects of *homogeneity*, at the *top* there are positive effects of *heterogeneity*. In other words, where there are no particular problems with achievement, heterogeneity is actually

an advantage. This may explain why openness and integration is a (political) preference of the social classes close to education.

For less distant comparisons, the differences become smaller, but they do not disappear. Nevertheless, the results and the graphical representations should only be evaluated with a certain degree of caution. The obvious problem is the increasing number and complexity of the estimates, also in view of the rather narrow data basis of the NEPS. The constellations are still quite simple and stable for social origin only, as analyzed by the standard position. It becomes more difficult with the social school effects alone, but is still manageable (cf. the contributions by Dunne 2010, Dronkers et al. 2011, Dronkers and Skopek 2015). With the inclusion of cognitive effects, the uncertainties continue to grow, but they are still manageable: Analyses with only the cognitive school effects (not further documented here) yielded the same findings up to that point. For the analyses of the triple interaction and its quite surprising findings, however, the limit may have been reached.

Table 9 shows the findings for analyses separated according to stringency. They allow simpler and therefore more stable comparisons between the systems with, at most, two-way interactions. To further simplify the estimates, the two less stringent system types, T1 and T2, were combined. Controls with all three types yielded the same result. The comparisons refer to two models each: one without and one with the two-way interaction of NABL*HABL (models 1 and 3 vs. 2 and 4), which may be so sensitive for the estimates.

Table 9: Conditional System effects 2 (Moderation, separated analyses, OLS, bold: $p<0.05$)
ABOUT HERE

The result is rather easy to summarize. In all comparisons, the effects of social origin are *not* higher for stringency (T3) than for openness (T1/T2) and the influence of the social niveau in the school classes is in almost every constellation *weakened* in the case of stringency. As before, there are virtually no differences in the effect of individual cognitive abilities. In *all* analyses, the effect of cognitive niveau of school classes is stronger for stringency than for openness. Effects of cognitive homogeneity are only evident in the two-way interaction with cognitive niveau. Here, the additional finding of a stronger effect in the lower achievement levels from table 7 is also recognizable. With stringency, there is also a stronger effect of school type, i.e. the (pareto-optimal) scissor effect of the accumulation of favorable conditions across

curriculum, stability, exclusivity and scope in the upper tracks, as was already visible in the non-conditional findings for the school type (Table 7 above).

Taken together, this means a clear correction of the standard results and the integration position, even if the stability of the estimates had to weaken with the increasing complexity of the analyses: There are *no* increases in the effects of social origin via the stringency of differentiation, not even in the effects of social segregation of school classes, but *positive* effects of cognitive niveau and school type. And all of this is always already without the theoretically recommended triple interaction of stringency, cognitive niveau and homogeneity. and all the doubts that can certainly be attached to the stability of the statistical estimates in view relatively low case numbers of the NEPS.

6. Limitations

Compared to the familiar standard studies, the NEPS offers in all its limitations a comparatively sound basis, especially with regard to the completeness of the necessary theoretical constructs, and also in part with the possibility of empirical analyses that could and have not been conducted in this field before. As already mentioned in the introduction it is the first (and sofar only) study that contains all the constructs necessary in the theoretical reasoning for and against Ability-Tracking. In this respect, the aspects in which the NEPS goes *beyond* the *previous* limitations in practically all studies before should be named first: no accounting for cognitive abilities, the social and cognitive composition of the school classes and the conditionalization of the effects on additional rules: the stringency of implementation.

Two limitations of this study, however, must be mentioned: The scope of the evidence and the range and representativeness of the data, especially regarding the structural conditions of school classes and school effects. The first point on the generalizability of the findings to international comparisons undoubtedly touches on an important point that calls for caution. It needs certainly to be clarified further, but this cannot be done in the foreseeable future and has not yet been attempted. The findings indeed apply *conditionally* to a differentiating system, and any generalization depends largely on speculation. Nevertheless, as already mentioned in the introduction, the findings have also their *own* significance: It seems that it really depends on the *concrete* implementation of Ability-Tracking, and this, it seems, is dependent on the

additional regulations of bindingness and control. Three aspects add to the relevance of the findings for the discussion of the system effects of Ability-Tracking. *First*, with the comparison of the regional states in *one* country, it would be a particularly conservative test of the theoretical model precisely because of the smaller variation in the national context, and when restricted to one national context, the problem of confounding across other macro variables, economic, political, social, and cultural characteristics, for example, is considerably smaller than in international comparisons. *Second*, the variation toward integrative systems in international comparison is not as great as is often inaugurated. This was already shown by Jackson and Jonsson (2013: 319f.), for example, who found to their own surprise that the countries with different education systems hardly differed in the primary effects in an international comparison and that it was rather the secondary effects that mattered, i.e., regulations restricting free choice. *Third*, the discernible internal differences among the German federal states indicate that the national averages commonly used in the PISA reports, for example, may have painted a *misleading* picture of the effects of *strict* Ability-Tracking: Because there have been only a few German federal states with stringent differentiation (and there are fewer and fewer), a wrong impression could be created that it was the stringency that created the negative picture of low achievement and high social stratification in Germany. However, it was (and is) quite different.

The second general objection to the scope of the findings is directed against the data situation in the NEPS (cf. Heisig and Matthewes, 2022): The sample is said to be too small and biased overall, and the number of cases in the school cohorts in particular is lower than the average. These objections are also not unjustified, nor can they be fundamentally eliminated for the present data set. Even with the greatest care, these problems affect panel studies in particular and thus, as is certainly the case, NEPS as well. Large-scale cross-sectional surveys, such as PISA or IQB, have an easier time of it: one does not notice the inconsistencies and omissions, and the high number of cases can only suggest the validity of the findings as a mere statistical certainty.

However, effects on the central findings with the NEPS data could only have been expected if the realized exhaustion at class level, and thus the reliability of the composition characteristics, varied systematically according to the characteristics of the education systems - which are ultimately at the core of the question about system effects. This, however, was not the case. The core findings of the analyses hardly changed as a function of the samples; in fact, they tended

to emerge more clearly with increasing restriction to classes with high exhaustion. This suggests that compositional characteristics are indeed less accurately represented in classes with low case counts, which may be associated with an underestimation of the compositional effect overall in the analyses. However, no evidence of systematic bias on the part of the (differential) system effects could be found. There was also a total of 36 classes coded with a sample size of one, making them inappropriate for determining within-class variances. Re-analysis adjusting for this and others then revealed a reduction in the total number of cases by 26 (from a total of 2662 to now 2636), including via some further adjustments in the data resulting from the widespread use and careful supervision of the NEPS.

.

None of this had any further impact on the substantive findings; the core findings were confirmed after further cleanups (see Esser 2023, chapter 9). This at least suggests that it is not the number of cases and the dropouts that may have caused biases in *favor* of the theoretical hypotheses, but rather the opposite. Nevertheless, the question remains open, of course, and the conclusion would be to proceed to investigations that contain everything on a broader basis and then in international, as well as in regional comparison with sufficient case numbers, especially then for the school classes.

7. Summary and evaluation

The paper focused on the question of the effects of ability tracking, which remains unresolved even after decades of empirical educational research. Three fundamental shortcomings of previous studies are assumed to be the reason for the inconsistencies in the findings and the ongoing debates: 1. The approaches refer to a set of *common* conditions and effects, but derive *different* theoretical predictions for the effects. 2. *All* relevant aspects have to be theoretically separated and related to one another. But this has hardly ever been the case. 3. The effects are based on *different phases* in the educational process: Sorting, transition and educational attainment, the resulting social and cognitive structuring of schools and school classes and achievement at secondary school after transition. But these are often not distinguished, such as attainment and achievement in particular. The aim of the paper was to theoretically systematize and complete the explanation of the system effects for the three topics and to empirically correct it, especially for the achievements in secondary school.

The results on the effects of stringency can be summarized in ten points. 1.stringency increases *achievement* right from the end of primary school, apparently because there are greater incentives to prepare for possible downgrading at an early stage. 2. in the stringent sorting federal states, the *recommendations* for the higher educational track tend to be more generous, so that misplacements are more likely to occur there, where achievement is otherwise more important. The background to this could be an avoidance of conflicts with parents from the upper classes, who are not simply prepared to accept this, in view of the higher cognitive requirements. 3. the rates of *transition* to the higher trajectories and educational attainment are lower in the stringent systems. There is no particular increase in the effects of social background, nor is there a decrease, but there are indications that the link between the transition and achievement is strengthened, i.e. that educational attainment becomes more performance-oriented and meritocratic. 4. the *structuring of school* classes with educational attainment shows a *decoupling of the* cognitive and social composition of school classes with stringency in all variants as a result of the higher performance equity in the transition. 5. the *achievements in secondary school* show, even after controlling for all relevant general conditions, a significant positive non-conditional system effect of the stringency of (early) differentiation on achievement in secondary education in the German federal states. 6. controlling for cognitive ability and school structures up to school type, the effects of social origin *shift* to those of cognitive ability and the corresponding school structures, respectively. 7. after all controls, a clear *un*conditional general effect of individual social origin remains, i.e., in both open and stringent education systems. 8. in the system effects, there is *no* reinforcement of the effects of social origin or social composition of school classes with the stringency of differentiation. 9. in all constellations of the interaction of stringency with the *cognitive niveau* is obtained with stringency. 10. cognitive homogeneity supports achievement in the *lower* achievement ranges, if at all, but that in each case only when there is stringency in the arrangements for sorting, transition, and educational attainment, i.e., a combination of binding and recommendations.

So, all in all, it does *not* look as the integration position and the findings of the standard approach had suggested: there are *increases* in achievement and meritocracy with the stringency of differentiation, but *no* reinforcements of social stratification in education - to say the least.

However, the analyses corresponding to the theoretical model were only possible in one specific respect: the investigation of the effects of different regulations of Ability Tracking in the *German federal states*. The international comparative studies lack key constructs and

analyses, especially the cognitive abilities and the cognitive composition of the school classes, but also the stringency with which the implementation of Ability Tracking should *actually* take place: the binding nature of the school recommendations at the transition and the monitoring of organizational processes in the schools. For the German federal states, on the other hand, the National Educational Panel Study (NEPS) provides a data set that contains everything that is theoretically necessary. The disadvantage is obvious: a direct transfer of the findings to fully integrated systems in an international comparison is not possible. It is therefore a question of *conditionalising* the effects of an otherwise comprehensively differentiated system. However, the open systems in the German federal states already come quite close to such full integration: there is a free choice of educational tracks and the relaxation and flexibilisation of many requirements and thus the core points of an educational system without sorting in the compulsory school period and a clear flexibilization of teaching and interactions.

Even if there is no doubt that the NEPS data used as a basis did not have the necessary numerical breadth for (even) more stable estimates, especially in the theoretically suggested estimands for the interaction effects, and some technical doubts cannot be refuted, the results do not correspond with the idea that opening up and relaxing the differentiation would improve anything, even in the simple models where only the social origin and social composition of the school classes are involved. Quite the opposite. Education policy, at least in Germany, has largely followed the recommendations that the standard position so emphatically suggested from the outset: Abolition of bindingness, relaxation of standards and requirements the transition to a two-tier system and de-differentiation in curricula. All of this may also explain why, roughly after gradually catching up in international comparisons after PISA 2000 to 2012, there has been a gradual decline again in achievement levels in Germany after 2012 (cf. Wößmann 2023) and that after 2020 the two most stringent federal states, Bavaria and Saxony, have survived the various turbulences in the last years almost unimpaired (IQB, 2022). They, unlike all others, did *not* drop in primary school achievement and later in secondary school. It is primarily the children in those federal states that have always pushed ahead with openings and relaxations, or have done so in the meantime, who have suffered.

These are all clear indications that the standard approach to identifying the systems effects of differentiation has pointed in the wrong direction via its theoretical and empirical incompleteness: It is not further openness that would have been the right path for the goals of efficiency and equity in educational success, but rather the retention or (re-)introduction of the

additional regulations that can ensure that Ability-Tracking is indeed *Ability*-Tracking: bindingness and control.

Tables 4 - 9

Table 4a: Achievement at the end of primary school (ACE, OLS), recommendations (REC, logistic regression) and transition to secondary school (GYM, logistic regression), System effects 1: Non-conditional effects of stringency (Mediation, bold: p<0.05).
ABOUT HERE

	Achievement elementary school (ACE)	Recommendations (REC)	Transition (GYM)
SES	**0,14**	**1.89**	**1.82**
	(0.01)	(0.25)	(0.37)
ABL	**0.33**	**1.16**	**0.81**
	(0.01)	(0.27)	(0.37)
ACE		**8.11**	**5.37**
		(0.48)	(0.60)
MRK		**...**	**...**
REC			**2.69**
			(0.15)
ASP			**2.36**
			(0.17)
T2	0.01	-0.18	0.52
	(0.01)	(0.22)	(0.44)
T3	**0.03**	-0.18	0.50
	(0.01)	(0.19)	(0.30)
FEM	0.00	**0.36**	0.01
	(0.00)	(0.10)	(0.13)
MHG	**-0,02**	**0.03**	**0.38**
	(0.01)	(0.11)	(0.16)
VSB	0.00	0.00	0.00
	(0.00)	(0.00)	(0.01)
c	**0.18**	**-5.35**	**-7.28**
(Pseudo-)R^2	0.33	0.25	(0.55)
N	3042	3042	3042

Table 4b: Achievement at the end of primary school (ACE, OLS), recommendations (REC, logistic regression) and transition to secondary school (GYM, logistic regression), System effects 2: Conditional effects of stringency (Moderation, bold: p<0.05).
ABOUT HERE

	1	2	3
	Achievement elementary school (ACE)	Recommendations (REC)	Transition (GYM)
SES	**0.14**	**1.72**	**1.81**
	(0.02)	(0.38)	(0.84)
ABL	**0.35**	**1.16**	**0.79**
	(0.02)	(0.27)	(0.37)
ACE		**9.03**	**5.79**
		(0.75)	(0.40)
MRK		**...**	**...**
REC			**2.74**
			(0.16)
ASP			**2.49**
			(0.17)
T2	**0.05**	0.11	0.68
	(0.03)	(0.68)	(0.53)
T3	**0.04**	0.44	0.85
	(0.02)	(0.57)	(0.54)
T2*SES	0.00	-0.22	-0.79
	(0.03)	(0.70)	(1.37)
T3*SES	0.00	0.61	0.74
	(0.03)	(0.56)	(1.45)
T2*ABL/ACE	**-0.07**	-0.37	-2.98
	(0.03)	(1.13)	(2.42)
T3*ABL/ACE	-0.01	-1.98	2.65
	(0.02)	(1.03)	(2.53)
FEM	0.00	**0.36**	0.01
	(0.00)	(0.10)	(0.13)
MHG	**-0.02**	**0.03**	**0.38**
	(0.01)	(0.11)	(0.16)
VSB	0.00	0.00	0.00
	(0.00)	(0.00)	(0.01)
c	**0.16**	**-5.60**	**-7.78**
(Pseudo-)R^2	0.34	0.25	(0.63)
N	3042	3042	3042

Table 5a: Cognitive differentiation (NABL) and social stratification (NSES) of school classes according to stringency; NEPS; OLS; control class size; bold: p<0.05; abbreviations in the text.

	Mediation			Moderation
	Systemeffect on NABL	Control NSES	Control Composition GMV	Interaction System*NSES
	1	2	3	4
T2	0.01	0.00	-0.02	-0.05
	(0.03)	(0.02)	(0.02)	(0.09)
T3	0.01	-0.03	0.02	**0.17**
	(0.02)	(0.02)	(0.02)	(0.08)
NSES		**0.71**	**0.55**	**0.64**
		(0.19)	(0.06)	(0.08)
T2*NSES				0.06
				(0.02)
T3*NSES				**-0.31**
				(0.02)
c	**0.61**	**0.19**	**0.16**	**0.13**
R^2	0.00	0.32	0.39	0.40
N	339	339	339	339

Table 5b: Social stratification (NSES) and cognitive differentiation (NABL) of school classes by stringency; NEPS; OLS; control class size; bold: p<0.05.

	Mediation			Moderation
	Systemeffect on NSES	Control NABL	Control Composition GMV	Interaction System*NABL
	1	2	3	4
T2	0.02	0.02	0.00	0.05
	(0.02)	(0.02)	(0.02)	(0.06)
T3	**0.06**	**0.05**	**0.05**	**0.23**
	(0.02)	(0.02)	(0.02)	(0.05)
NABL		**0.46**	**0.36**	**0.48**
		(0.04)	(0.04)	(0.02)
T2*NABL				-0.07
				(0.09)
T3*NABL				**-0.29**
				(0.08)
c	**0.59**	**0.31**	**0.25**	**0.20**
R^2	0.02	0.34	0.38	0.39
N	339	339	339	339

Table 6a: Cognitive homogeneity (HABL) and social segregation (HSES) of school classes by stringency; NEPS; OLS; control class size; bold: p<0.05; abbreviations in the text.

	Mediation			Moderation
	Systemeffect on HABL	Control HSES	Control Composition GMV	Interaction System*HSES
	1	2	3	4
T2	-0.02	-0.02	-0.03	-0.05
	(0.02)	(0.03)	(0.02)	(0.08)
T3	0.03	0.03	0.02	**0.21**
	(0.02)	(0.06)	(0.02)	(0.18)
HSES		0.03	0.04	0.16
		(0.03)	(0.06)	(0.09)
T2*HSES				0.03
				(0.18)
T3*HSES				**-0.28**
				(0.08)
c	**0.66**	**0.64**	**0.55**	**0.46**
R^2	0.02	0.02	0.05	0.07
N	339	339	339	339

Table 6b: Social segregation (HSES) and cognitive homogeneity (HABL) of school classes by stringency; NEPS; OLS; control class size; bold: p<0.05.

	Mediation			Moderation
	Systemeffect on HSES	Control HABL	Control Composition GMV	Interaction System*HABL
	1	2	3	4
T2	-0.00	-0.00	-0.00	0.02
	(0.02)	(0.02)	(0.02)	(0.09)
T3	0.00	0.00	0.01	**0.23**
	(0.01)	(0.02)	(0.02)	(0.09)
HABL		0.03	0.03	0.15
		(0.03)	(0.05)	(0.02)
T2*HABL				-0.03
				(0.14)
T3*HABL				**-0.33**
				(0.07)
c	**0.68**	**0.66**	**0.76**	**0.68**
R^2	0.00	0.00	0.02	0.04
N	339	339	339	339

Table 7: System effects 1: Non-conditional effects of stringency (Mediation, OLS, bold: p<0.05)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
SES	**0,63**	**0,41**	**0,63**	**0,41**	**0,41**	**0,39**	**0,39**	**0,39**	**0,39**	**0,39**	**0,39**
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)
NSES		**2,39**		**2,40**	**3,78**	**2,85**	**1,73**	**2,81**	**1,74**	**1,89**	0.79
		(0.18)		(0.18)	(0.95)	(0.81)	(0.76)	(0.81)	(0.76)	(0.77)	(0.71)
HSES			-0.39	-0.48	0.51	0.43	0.30	0.41	0.31	0.38	0.09
			(0.33)	(0.27)	(0.72)	(0.62)	(0.58)	(0.62)	(0.58)	(0.58)	(0.53)
NSES*HSES					-1.95	-1.60	-1.13	-1.56	-1.15	-1.35	-0.73
					(1.33)	(1.13)	(1.05)	(1.13)	(1.05)	(1.06)	(0.96)
ABL						**1,60**	**1,41**	**1,60**	**1,41**	**1,41**	**1,41**
						(0.07)	(0.08)	(0.07)	(0.08)	(0.08)	(0.08)
NABL							**1,42**		**1,42**	**2,20**	**1,41**
							(0.17)		(0.17)	(0.53)	(0.49)
HABL								0.13	-0.06	0.66	0.40
								(0.18)	(0.17)	(0.50)	(0.45)
NABL*HABL										-1.29	-0.88
										(0.85)	(0.77)
GYM											**0,56**
											(0.06)
T2	0.19	0.08	0.18	0.07	0.07	0.07	0.06	0.07	0.06	0.05	0.03
	(0.11)	(0.09)	(0.11)	(0.09)	(0.09)	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)	(0.06)
T3	**0,26**	0.13	**0,26**	0.14	**0,14**	**0,16**	**0,18**	**0,16**	**0,18**	**0,17**	**0,19**
	(0.09)	(0.07)	(0.09)	(0.07)	(0.07)	(0.06)	(0.05)	(0.06)	(0.05)	(0.05)	(0.05)
FEM	-0.06	**-0,06**	-0.06	**-0,07**	**-0,06**	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
MHG	**-0,10**	**-0,09**	**-0,10**	**-0,10**	**-0,10**	**-0,08**	**-0,07**	**-0,08**	**-0,07**	**-0,07**	**-0,09**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
VSB	0.00	0.00	0.00	0.00	0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
c	**-0,58**	**-1,70**	-0.32	**-1,38**	**-2,08**	**-2,57**	**-2,80**	**-2,62**	**-2,78**	**-3,25**	**-2,48**
	(0.10)	(0.12)	(0.24)	(0.22)	(0.52)	(0.45)	(0.42)	(0.45)	(0.42)	(0.52)	(0.48)
R^2: level 1	0.01	0.01	0.01	0.01	0.01	0.14	0.14	0.14	0.14	0,14	0,14
R^2: level 2	0.19	0.54	0.20	0.55	0.55	0.72	0.79	0.72	0.79	0,79	0,86
N: students	2636	2636	2636	2636	2636	2636	2636	2636	2636	2636	2636
N: classrooms	313	313	313	313	313	313	313	313	313	313	313

Table 8: Conditional System effects of stringency 1 (Moderation, Interactions, OLS, bold: $p<0.05$)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
SES	**0,41**	**0,44**	**0,41**	**0,44**	**0,44**	**0,44**	**0,43**	**0,43**	**0,43**	**0,43**	**0,43**	**0,43**
	(0.11)	(0.12)	(0.11)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
NSES	0.80	0.70	0.65	0.56	0.79	1.05	0.74	1.03	0.76	1.03	0.93	1.05
	(0.71)	(0.72)	(0.71)	(0.73)	(0.98)	(0.98)	(0.98)	(0.99)	(0.98)	(0.99)	(0.98)	(0.98)
HSES	0.10	0.08	0.15	0.13	0.29	0.35	0.29	0.30	0.31	0.31	0.31	0.36
	(0.53)	(0.53)	(0.55)	(0.55)	(0.71)	(0.71)	(0.71)	(0.71)	(0.71)	(0.71)	(0.70)	(0.70)
NSES*HSES	-0.75	-0.72	-0.50	-0.49	-0.81	-0.93	-0.80	-0.89	-0.82	-0.90	-0.79	-0.85
	(0.96)	(0.96)	(0.97)	(0.98)	(1.35)	(1.34)	(1.35)	(1.34)	(1.34)	(1.34)	(1.34)	(1.33)
ABL	**1,41**	**1,41**	**1,41**	**1,41**	**1,41**	**1,41**	**1,50**	**1,56**	**1,51**	**1,56**	**1,56**	**1,56**
	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.08)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)	(0.12)
NABL	**1,41**	**1,43**	**1,32**	**1,35**	**1,33**	**1,35**	**1,30**	**1,04**	**1,34**	**1,08**	0.29	0.38
	(0.49)	(0.49)	(0.49)	(0.49)	(0.50)	(0.49)	(0.50)	(0.51)	(0.50)	(0.52)	(0.72)	(0.72)
HABL	0.41	0.42	0.33	0.35	0.32	0.38	0.30	0.44	0.22	0.37	-0.38	-0.39
	(0.45)	(0.46)	(0.45)	(0.46)	(0.46)	(0.46)	(0.46)	(0.47)	(0.48)	(0.49)	(0.69)	(0.68)
NABL*HABL	-0.88	-0.91	-0.77	-0.80	-0.76	-0.83	-0.72	-0.90	-0.78	-0.92	0.46	0.48
	(0.77)	(0.77)	(0.77)	(0.77)	(0.78)	(0.77)	(0.78)	(0.78)	(0.78)	(0.79)	(1.18)	(1.18)
GYM	**0,56**	**0,56**	**0,56**	**0,56**	**0,56**	**0,45**	**0,56**	**0,56**	**0,56**	**0,56**	**0,56**	**0,47**
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)	(0.08)
T2*SES	-0.13	-0.18	-0.14	-0.18	-0.18	-0.18	-0.16	-0.16	-0.16	-0.16	-0.16	-0.16
	(0.18)	(0.20)	(0.18)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)	(0.20)
T2*NSES		0.22		0.19	0.21	0.08	0.27	-0.08	0.08	-0.23	-0.01	-0.17
		(0.37)		(0.37)	(2.51)	(2.52)	(2.50)	(2.51)	(2.55)	(2.55)	(2.53)	(2.54)
T2*HSES			0.57	0.56	0.58	0.50	0.56	0.48	0.42	0.35	0.54	0.49
			(0.69)	(0.69)	(1.95)	(1.93)	(1.94)	(1.94)	(1.97)	(1.97)	(1.96)	(1.95)
T2*NSES*HSES					-0.04	0.11	0.14	0.44	0.37	0.66	0.41	0.51
					(3.62)	(3.59)	(3.61)	(3.61)	(3.67)	(3.67)	(3.64)	(3.63)
T2*ABL							**-0,39**	**-0,42**	**-0,39**	**-0,42**	**-0,42**	**-0,42**
							(0.19)	(0.20)	(0.19)	(0.20)	(0.20)	(0.20)
T2*NABL								0.21		0.18	-0.36	-0.50
								(0.43)		(0.44)	(1.35)	(1.40)
T2*HABL									0.22	0.17	-0.28	-0.30
									(0.41)	(0.41)	(1.25)	(1.26)
T2*NABL*HABL											0.76	0.78
											(2.12)	(2.14)
T2*GYM						0.04						0.11
						(0.15)						(0.17)
T3*SES	0.01	-0.02	0.01	-0.02	-0.02	-0.02	-0.02	-0.02	-0.01	-0.02	-0.02	-0.02
	(0.16)	(0.17)	(0.16)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)	(0.17)
T3*NSES		0.11		0.13	-0.46	-1.17	-0.48	-0.99	-0.50	-0.99	-0.17	-0.54
		(0.29)		(0.29)	(1.47)	(1.49)	(1.47)	(1.49)	(1.47)	(1.49)	(1.51)	(1.52)
T3*HSES			-0.53	-0.53	-0.98	-1.04	-0.99	-1.03	-1.00	-1.04	-0.51	-0.56
			(0.41)	(0.41)	(1.15)	(1.15)	(1.15)	(1.15)	(1.15)	(1.15)	(1.16)	(1.15)
T3*NSES*HSES					0.84	1.10	0.86	1.09	0.88	1.10	0.02	0.16
					(2.04)	(2.03)	(2.04)	(2.04)	(2.04)	(2.04)	(2.06)	(2.05)
T3*ABL							-0.00	-0.14	-0.00	-0.14	-0.14	-0.14

							(0.17)	(0.18)	(0.17)	(0.18)	(0.18)	(0.18)
T3*NABL								**0,72**		**0,70**	**3,18**	**2,93**
								(0.34)		(0.34)	(1.09)	(1.10)
T3*HABL									0.19	0.12	2.49*	2.46*
									(0.35)	(0.35)	(1.05)	(1.04)
T3*NABL*HABL											**-4,14**	**-4,09**
											(1.74)	(1.74)
T3*GYM						**0,29**						0.22
						(0.12)						(0.13)
T2	0.10	0.00	-0.27	-0.35	-0.36	-0.31	-0.22	-0.20	-0.23	-0.18	-0.01	0.11
	(0.11)	(0.20)	(0.46)	(0.48)	(1.34)	(1.34)	(1.34)	(1.34)	(1.35)	(1.35)	(1.47)	(1.49)
T3	0.18	0.13	0.54	0.48	0.80	0.99	0.80	0.66	0.71	0.61	-1.22	-0.99
	(0.10)	(0.16)	(0.29)	(0.32)	(0.83)	(0.82)	(0.83)	(0.83)	(0.85)	(0.85)	(1.11)	(1.12)
FEM	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02	-0.02
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
MHG	**-0,09**	**-0,09**	**-0,09**	**-0,09**	**-0,09**	**-0,09**	**-0,09**	**-0,09**	**-0,09**	**-0,09**	**-0,09**	**-0,09**
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)
VSB	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
c	**-2,50**	**-2,47**	**-2,48**	**-2,45**	**-2,55**	**-2,66**	**-2,57**	**-2,59**	**-2,55**	**-2,57**	**-2,14**	**-2,23**
	(0.49)	(0.49)	(0.49)	(0.49)	(0.58)	(0.58)	(0.58)	(0.58)	(0.58)	(0.58)	(0.63)	(0.63)
R^2: level 1	0.86	0.86	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0.87	0,87	0,88
R^2: level 2	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0,14	0,14
N: students	2636	2636	2636	2636	2636	2636	2636	2636	2636	2636	2636	2636
N: classrooms	313	313	313	313	313	313	313	313	313	313	313	313

Table 9: Conditional System effects 2 (Moderation, separated analyses, OLS, bold: p<0.05)

	T1/T2		T3	
	(1)	(2)	(3)	(4)
	b/se	*b/se*	*b/se*	*b/se*
SES	**0,37**	**0,37**	**0,41**	**0,41**
	(0.10)	(0.10)	(0.13)	(0.13)
NSES	1.16	1.14	-0.43	0.49
	(0.94)	(0.94)	(1.02)	(1.05)
HSES	0.44	0.45	-0.91	-0.24
	(0.68)	(0.68)	(0.82)	(0.84)
NSES*HSES	-0.88	-0.85	0.58	-0.64
	(1.29)	(1.29)	(1.39)	(1.42)
ABL	**1,40**	**1,40**	**1,42**	**1,42**
	(0.10)	(0.10)	(0.13)	(0.13)
NABL	**0,69**	0.38	**1,05**	**3,20**
	(0.23)	(0.64)	(0.23)	(0.75)
HABL	-0.09	-0.38	-0.06	**1,99**
	(0.20)	(0.59)	(0.23)	(0.72)
NABL*HABL		0.53		**-3,47**
		(1.02)		(1.15)
GYM	**0,49**	**0,50**	**0,69**	**0,69**
	(0.08)	(0.08)	(0.09)	(0.09)
FEM	-0.03	-0.03	-0.01	-0.02
	(0.04)	(0.04)	(0.05)	(0.05)
MHG	**-0,10**	**-0,10**	-0.08	-0.08
	(0.04)	(0.04)	(0.05)	(0.05)
VSB	0.00	0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
c	**-2,45**	**-2,28**	**-1,24**	**-3,03**
	(0.50)	(0.59)	(0.61)	(0.85)
R^2: level 1	0.13	0.13	0,16	0,16
R^2: level 2	0.84	0.84	0,92	0,93
N: students	1632	1632	1004	1004
N: classrooms	198	198	115	115

Allmendinger, J. 1989. Educational Systems and Labor Market Outcomes. In: European Sociological Review, 5: 231–250.

Angrist, J. D. 2014. The perils of peer effects. Labour Economics, 30, issue C: 98–108.

Artelt, C., Schneider, W., and Schiefele, U. .2002. Ländervergleich zur Lesekompetenz. In: Deutsches PISA-Konsortium, Eds., PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich. Opladen 2002: Leske und Budrich: 56–94.

Bach, M., and M. Fischer. 2021. Understanding the Response to High Stakes Incentives in Primary Education. IZA-Discussion Paper Nr. 13845. Bonn: Forschungsinstitut Zukunft der Arbeit. Institute for the Study of Labor.

Baier, T, V. Lang, M. Grätz, J. Kieron, K. J. Barclay, C. Dalton, D. C. Conley, Ch. T.Dawes, Th. Laidley and T. H. Lyngstad .2022.. Genetic Influences on Educational Achievement in Cross-National Perspective. In: European Sociological Review, 38: 1–16. https://doi.org/10.1093/esr/jcac014.

Bauer, Ph., and R. T. Riphan. 2009. Age at School Entry and Intergenerational Educational Mobility. In: Economic Letters, 103: 87–90.

Baumert, J., and G. Schümer. 2002. Familiäre Lebensverhältnisse, Bildungsbeteiligung und Kompetenzerwerb im nationalen Vergleich. In: Deutsches PISA-Konsortium, Ed., PISA 2000 – Die Länder der Bundesrepublik Deutschland im Vergleich. Opladen 2002: Leske und Budrich: 159–201.

Betts, J.R. 2011. The economics of tracking in education. In: E.A. Hanushek, S. Machin and L. Woessmann, Eds., Handbook of the Economics of Education, Band 3. Amsterdam: North Holland: 341–381..

Blossfeld, H. P., H.G. Roßbach and J. von Maurice. 2011. Education as a Lifelong Process – The German National Educational Panel Study (NEPS). Wiesbaden: Springer VS.

Bol, Th., J. Witschge, H. G. van de Werfhorst and J. Dronkers. 2014. Curricular tracking and central examinations: Counterbalancing the impact of social background on student achievement in 36 countries. In: Social Forces 92: 1545–1572.

Bowden, A. B., V. Rodriguez and Z. Weingarten. 2023. The Unintended Consequences of Academic Leniency. EdWorking Paper: 23-836. Annenberg Institute at Brown University: https://doi.org/10.26300/m6sa-m733

Bratti, L., L. Cappellari, O. Groh-Samberg and H. Lohmann. 2012. School Tracking and Intergenerational Transmission of Education. In: J. Ermisch, M. Jäntti and T. M. Smeeding, Eds., From Parents to Children: The Intergenerational Transmission of Advantage. New York: Russell Sage Foundation: 311–344.

Brinkmann, M., N. Huth-Stöckle, R. Schunck and Jana Teltemann. 2024. Achievement or Social Background? The Impact of Tracking on the Composition of Schools in an International Comparison. In: Zeitschrift für Soziologie 53, https://doi.org/10.1515/zfsoz-2024-2014.

Brunello, G., and D. Checchi. 2007. Does school tracking affect equality of opportunity? New international evidence. Economic Policy, 22: 782–861.

Brunello, G., M. Giannini and K. Ariga. 2007. The Optimal Timing of School Tracking. A General Model with Calibration for Germany. In: L. Wößmann and P.E. Peterson, Eds., Schools and the Equal Opportunity Problem. Cesifo Seminars Series. Cambrigde, Mass.: MIT: Press: 229-169.

Brunello, G., L. Rocco and K. Ariga. 2012. In: Education Economics, 20: 117-138.

Checchi, D., & L. Flabbi (2007). Intergenerational Mobility and Schooling Decisions in Germany and Italy: The Impact of Secondary School Tracks. IZA-Discussion Paper Nr. 2876. Bonn: Forschungsinstitut Zukunft der Arbeit. Institute for the Study of Labor.

Coleman, J. S., E. Q. Campbell, et al. 1967. Equality of Educational Opportunity. Washington, D.C. Government Printing Office.

Chmielewski, A. K., 2014. An International Comparison of Achievement Inequality in Within- and Between-School Tracking Systems. In: American Journal of Education, 120: 293–324; http:// www.jstor.org/stable/10.1086/675529.

Cinelli, C., A. Forney und J. Pearl. 2021. A Crash Course in Good and Bad Controls. Technical Report R-493.

Cord, D. & L. Giuliano (2016). Can Tracking Raise the Test Scores of High-Ability Minority Students? In: American Economic Review, 106: 2783-2816.

Dollmann, J. 2019. Educational institutions and inequalities in educational opportunities. In: R. Becker, Ed., Research Handbook on the Sociology of Education, Cheltenham: Edward Elgar Publishing Limited: 268–283.

Domina, T., A. Penner und E. Penner. 2017. Categorical Inequality: Schools as Sorting Machines. In: Annual Review of Sociology, 43: 311–330.

Domina, T., A. McEachin, P. Hanselman, P. Agarwal, N. Hwang and R. W. Lewis. 2019. Beyond Tracking and Detracking: Dimensions of Organizational Differentiation in Schools. In: Sociology of Education, 92: 293–322.

Dräger, J., Th Schneider, M. Olczyk› A.Solaz, A.Sheridan, E. Washbrook, V. Perinetti Casoni, S. J. Kwon & J. Waldfogel. 2003. The relevance of tracking and social school composition for growing achievement gaps by parental education in lower secondary school: a longitudinal analysis in France, Germany, the United States, and England. In: European Sociological Review, 8, 1–17 https://doi.org/10.1093/esr/jcad076.

Dronkers, J., R. van der Velden and A. Dunne. 2011. The effects of educational systems, school- composition, track-level, parental background and immigrants‹ origins on the achievement of 15-years old native and immigrant students. A reanalysis of PISA 2006. Maastricht: ROA, Research Centre for Education and the Labour Market.

Duflo, E., P. Dupas, and M. Kremer. 2011. Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya. In: American Economic Review, 101: 1739–1774.

Dunne, A., 2010. Dividing lines. Examining the relative importance of between- and within-school differentiation during lower secondary education. PhD-Thesis, Department of Political and Social Sciences. Florence: European University Institute.

Elwert, F. und Ch. Winship. 2014. Endogenous Selection Bias: The Problem of Conditioning on a Collider Variable. In: Annual Review of Sociology, 40: 31–53.

Esser, H. 2016a. Bildungssysteme und ethnische Bildungsungleichheit. In: Diehl, C., Ch. Hunkler and C. Kristen, Eds, Ethnische Ungleichheiten im Bildungsverlauf. Mechanismen, Befunde, Debatten. Wiesbaden: Springer VS: 331–396.

Esser, H. 2016b. Educational Systems and Educational Inequality. The Model Ability Tracking and Empirical Findings. In: H. P. Blossfeld, S. Buchholz, J. Skopek, and M. Triventi, Eds., Models of Secondary Education and Social Inequality – An International Comparison. eduLIFE Lifelong Learning Series, Band 3. Cheltenham, UK and Northampton, MA, USA: Edward Elgar Publishing: 25–44.

Esser, H. 2021. „Wie kaum in einem anderen Land?“ Die Differenzierung der Bildungswege und ihre Wirkung auf Bildungserfolg, -ungleichheit und -gerechtigkeit. Band1: Theoretische Grundlagen. Frankfurt/M. und New York: Campus.

Esser, H. 2023. „Wie kaum in einem anderen Land?“ Die Differenzierung der Bildungswege und ihre Wirkung auf Bildungserfolg, -ungleichheit und -gerechtigkeit. Band 2: Empirische Zusammenhänge. Frankfurt/M. und New York: Campus.

Esser, H. und I. Relikowski. 2015. Is Ability Tracking (Really) Responsible for Educational Inequalities in Achievement? A Comparison between the Country States Bavaria and Hesse in Germany. IZA-Discussion Paper Nr. 9082. Bonn: Forschungsinstitut Zukunft der Arbeit. Institute for the Study of Labor.

Esser, H., and K. Hönig. 2018. Leistungsgerechtigkeit und Bildungsungleichheit. Effekte der Verbindlichkeit der Grundschulempfehlungen beim Übergang auf das Gymnasium. Ein Vergleich der deutschen Bundesländer mit den Daten der „National Educational Panel Study“ (NEPS). 2018. In: Kölner Zeitschrift für Soziologie und Sozialpsychologie, 70, 419-447.

Esser, H. and J. Seuring. 2020. Kognitive Homogenisierung, schulische Leistungen und soziale Bildungsungleichheit. Theoretische Modellierung und empirische Analyse der Effekte einer strikten Differenzierung nach den kognitiven Fähigkeiten auf die Leistungen in der Sekundarstufe und den Einfluss der sozialen Herkunft in den deutschen Bundesländern mit den Daten der „National Educational Panel Study“ (NEPS). In: Zeitschrift für Soziologie, 49: 277–301.

Esser, H., and J. Seuring. 2023. Was ist Dein Replicandum? Eine Antwort auf die Replik von Heisig und Matthewes (2022) zum Beitrag von Esser und Seuring (2020) über „Kognitive Homogenisierung, schulische Leistungen und soziale Bildungsungleichheit“. https://doi.org/10.1515/zfsoz-2023-2021.

Feld, J., and U. Zölitz. 2017. Understanding Peer Effects: On the Nature, Estimation, and Channels of Peer Effects. Journal of Labor Economics, 35: 387–428.

Galindo-Rueda, F., and A. Vignoles. 2007. The heterogeneous effect of selection in UK secondary schools. In: L. Wößmann and P.E. Peterson, Eds., Schools and the equal opportunity problem. Massachusetts: MIT Press: 103–128.

Gamoran, A. 2009. Tracking and Inequality: New Directions for Research and Practice. WCER Working Paper Nr. 2009–6. Madison: University of Wisconsin-Madison: Wisconsin Centre for Education Research.

Ganzeboom, H. B., P. M. de Graaf and D. J. Treiman. 1992. A standard international socio-eco- nomic index of occupational status. In: Social Science Research, 21: 1–56.

Gehrer, K., S. Zimmermann, C. Artelt and S. Weinert. 2013. NEPS Framework for Assessing Reading Competence and Results From an Adult Pilot Study. In: C. Artelt, S. Weinert and C. H. Carstensen, Eds., Competence Assessment within the NEPS. In: Journal for Educational Research Online: 50–79.

Grewenig, E. 2021. School Track Decisions and Teacher Recommendations. In: E. Grewenig. Human Capital and Education Policy: Evidence from Survey Data. Ifo Beiträge zur Wirtschaftsforschung, 96, Kapitel 4: 113–161.

Hadjar, A., and R. Becker. 2016. Education systems and meritocracy: Social origin, educational and status attainment. In: A. Hadjar und Ch. Gross, Eds., Education Systems and Inequalities: International Comparisons. Bristol: Policy Press: 231–258.

Hallinan, M.T. 1994. Tracking: From Theory to Practice. Sociology of Education, 67: 79–84.

Hanushek, E.A., and L. Wößmann. 2006. Does educational tracking affect performance and inequality? Differences-in-differences evidence across countries. In: The Economic Journal, 116: C63-C76.

Heisig, J. P., and S. H. Matthewes. 2022. Keine Belege für leistungsfördernde Effekte von strikter Leistungsdifferenzierung und kognitiver Homogenisierung: Eine kritische Reanalyse von Esser und Seuring (2022). In: Zeitschrift für Soziologie, 51: 99-111.

Helbig, M., and R. Nikolai. 2015. Die Unvergleichbaren. Der Wandel der Schulsysteme in den deutschen Bundesländern seit 1949. Bad Heilbrunn: Klinkhardt.

Horn, D. 2008. Age of selection counts: a cross-country comparison of educational institutions. Mannheim: Arbeitspapiere – Mannheimer Zentrum für Europäische Sozialforschung; No. 107.

Horn, D. 2013. Diverging performances: the detrimental effects of early educational selection on equality of opportunity in Hungary, In: Research in Social Stratification and Mobility, 32: 25–43.

IGLU 2023. IGLU 2021. Lesekompetenz von Grundschulkindern im internationalen Vergleich und im Trend über 20 Jahre. Münster und New York: Waxmann.

IQB-Bericht 2009. 2010. Sprachliche Kompetenzen im Ländervergleich. Münster/New York/ München/Berlin: Waxmann.

IQB-Bericht 2012. 2013. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I. Münster/New York/München/Berlin: Waxmann.

IQB-Bericht 2015. 2016. Sprachliche Kompetenzen am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich. Münster/New York/München/Berlin: Waxmann

IQB-Bericht 2016. 2017. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangstufe im zweiten Ländervergleich. Münster/New York/München/Berlin: Waxmann.

IQB-Bericht 2018. 2019. Mathematische und naturwissenschaftliche Kompetenzen am Ende der Sekundarstufe I im zweiten Ländervergleich. Münster/New York/München/Berlin: Waxmann.

IQB-Bericht 2021. 2022. Bildungstrend 2021. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich. Münster/New York/München/Berlin: Waxmann.

Jackson, M. 2013. How ist Inequality of Opportunity Gernerated? The Case of Primary and Secondary Effects. In: M. Jackson, Ed., Determined to Succeed? Performance versus Choice in Educational Attainment, Stanford, CA: Stanford University Press: 1-46.

Jackson, M., and J.O. Jonsson. 2013. Why does inequality of educational opportunity vary across countries? Primary and secondary effects in comparative context. In: M. Jackson, Ed., Determined to Succeed? Performance versus Choice in Educational Attainment, Stanford, CA: Stanford University Press: 306–337.

Klein, M. und M. Kühhirt (2021). Direct and Indirect Effects of Grandparent Education on Grandchildren's Cognitive Development: The Role of Parental Cognitive Ability. In: Sociological Science. DOI: 10.15195/v8.a13.

Korthals, R. A., and J. Dronkers. 2016. Selection on performance and tracking. Applied Economics. DOI: 10.1080/00036846.2015.1130789.

Lang, F.R., S. Kamin, M. Rohr, C. Stünkel & B. Williger, 2014: Erfassung der fluiden kognitiven Leistungsfähigkeit über die Lebensspanne im Rahmen des Nationalen Bildungspanels: Abschlussbericht zu einer NEPS-Ergänzungsstudie (NEPS Working Paper No. 43). Bamberg: Leibniz-Institut für Bildungsverläufe: Nationales Bildungspanel.

Lauermann, F., A. Meißner and R. Steinmayr. 2019. Relative Importance of Intelligence and Ability Self-Concept in Predicting Test Performance and School Grades in the Math and Language Arts Domains. In: Journal of Educational Psychology 112: 364–383.

Lundberg, I., R. Johnson und B. M. Stewart. 2019. What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory. American Sociological Review. Online-First. https://doi.org/10.1177/00031224211004187.

Marks, G. N. 2014. Education, social background and cognitive ability. The decline of the social. London: Routledge.

Marks, G. N. 2015. Are school-SES effects statistical artefacts? Evidence from longitudinal population data. In: Oxford Review of Education, 41: 122–144.

Marks, G. N. 2019. Socioeconomic background, education, cognitive ability and genetics: A commentary on Bourne, Betthäuser, Bourne and Bukodi. In: Research in Social Stratification and Mobility, 61: 20–23.

Marks, G. N. 2020. Occupational mobility and cognitive ability: A commentary on Betthäuser, Bourne and Bukodi. In: British Journal of Sociology, 71: 898–901.

Mönkediek, B. 2022. How variants of tracking affect the role of genes and environment in explaining child attendance at upper secondary school. In: *Research in Social Stratification and Mobility,* 81. 100714. https://doi.org/10.1016/j.rssm.2022.100714.

Morgan, St. L. 2012. Models of College Entry in the United States and the Challenges of Estimating Primary and Secondary Effects. In: Sociological Methods & Research, 41: 17–56.

Morgan, St. L., M. W. Spiller and J. J. Todd. 2013. Class Origins and Education in the United States. In: M. Jackson, Ed.., Determined to succeed? Performance versus choice in educational attainment, Stanford, CA: Stanford University Press, 279–305.

Müller, W., and Kogan, I. 2010. Education. In: St. Immerfall and G. Therborn, Eds., Handbook of European Societies: Social Transformations in the 21st Century, New York: Springer: 217–289.

Oakes, J. 1985. How Schools Structure Inequality, New Haven und London: Yale University Press.

OECD.2000. PISA database; http://www.oecd. org/pisa/data/.

O´Connell, M., and G. S. Marks. 2021. Are the effects of intelligence on student achievement and well-being largely functions of family income and social class? Evidence from a longitudinal study of Irish adolescents. In: Intelligence, 84: https://doi.org/10.1016/j.intell.2020.101511.

Pfeffer, F. 2008. Persistent Inequality in Educational Attainment and its Institutional Context. In: European Sociological Review, 24, 543–565.

Pfeffer, F. 2015. Equality and quality in education. A comparative study of 19 countries. In: Social Science Research, 51: 350–368

Pokropek, A. 2014. Phantom Effects in Multilevel Compositional Analysis: Problems and Solutions. Sociological Methods and Research, 44:1-29.

Roller, M., and D. Steinberg. .2020. The distributional effects of early school stratification - non-parametric evidence from Germany. In: European Economic Review, 125. https://doi.org/10.1016/j.euroecorev.2020.103422.

Rohrer, J. M. 2018. Thinking Clearly About Correlations and Causation: Graphical Causal Models for Observational Data. In: Advances in Methods and Practices in Psychological Science, 1: 27–42.

Sacerdote, B. 2011. Peer Effects in Education: How Might They Work, How Big Are They and How Much Do We Know Thus Far? In: E.A. Hanushek, S. Machin and L. Wößmann, Eds., Handbook of the Economics of Education, Volume 3. Amsterdam: North-Holland: 249-277.

Sahlgren, G. H. 2015. Real Finnish Lessons. The true story of an educational superpower. Centre for Policy Studies: London.

Schindler, St., et al. 2021. Educational tracking and long-term outcomes by social origin: Seven countries in comparison. DIAL Working Paper Series 02/2021.

Skopek, J., and J Dronkers. (2015). Performance in Secondary School in German Country States – A longitudinal Approach, Working Paper. European University Institute und Maastricht University (ROA).

Skopek, J., M. Triventi and S. Buchholz. 2019. How do educational systems affect social inequality of educational opportunities? The role of tracking in comparative perspective. In: R. Becker, Ed., Research Handbook on the Sociology of Education, Cheltenham: Edward Elgar Publishing Limited: Cheltenham: 214–232.

Sörensen, A. B. 1970. Organizational Differentiation of Students and Educational Opportunity? In: Sociology of Education, 43: 355–376.

Sörensen, A.B., and M. T. Hallinan. 1977. A Reconceptualization of School Effects. In: Sociology of Education, 50: 273–289.

Strello, R. Strietholt, I. Steinmann and Ch. Siepmann. 2021. Early tracking and different types of inequalities in achievement: difference-in-differences evidence from 20 years of large-scale assessments. In: Educational Assessment, Evaluation and Accountability. https://doi.org/10.1007/s11092-020-09346-4.

Terrin, E., and M. Triventi. 2022. The Effect of School Tracking on Student Achievement and Inequality: A Meta-Analysis. In: Review of Educational Research, 20: 1–39.

Uchikoshi, F. & D. Conley (2022). Gene-environment interactions and school tracking during secondary education: Evidence from the U.S.. In: Research on Social Stratification and Mobility. 2021. doi: 10.1016/j.rssm.2021.100628

van de Werfhorst, H. G. 2018. Early tracking and socioeconomic inequality in academic achievement: Studying reforms in nine countries. In: Research in Social Stratification and Mobility, 58: 22–32.

van de Werfhorst, H. G. 2019. Early Tracking and Social Inequality in Educational Attainment: Educational Reforms in 21 European Countries. In: American Journal of Education, 126: 65–99.

van de Werfhorst, H. G. 2022. Sorting or mixing? Multi-track and single-track schools and social inequalities in a differentiated educational system. In: British Educational Research Journal. https://doi.org/10.1002/berj.3722

van de Werfhorst, H.G., and J.J.B. Mijs. 2010. Achievement inequality and the institutional structure of educational systems: A comparative perspective. In: Annual Review of Sociology, 36: 407–428.

von Below, S. 2011. Bildungssysteme im historischen und internationalen Vergleich. In: R. Becker, Ed., Lehrbuch der Bildungssoziologie, 2. Aufl. Wiesbaden: Springer VS: 139–162.

von Neumann, I., Ch. Duchhardt, M. Grüßing, A. Heinze, E. Knopp and T. Ehmke. 2013. Modeling and assessing mathematical competence over the lifespan.In: Journal for educational research online 5: 80-109 - URN: urn:nbn:de:0111-opus-84267 - doi: 10.25656/01:8426

Waldinger, F. 2007. Does tracking affect the importance of family background on students test scores? London: LSE.

Wößmann, L. 2009. International Evidence on School Tracking: A review. In: Cesifo DICE Reports, 7: 26-34.

Wößmann, L., E. Lüdemann, G. Schütz, und M.R. West. 2009. School Accountability, Autonomy and Choice around the World. Celtenham: Edward Elgar Publishing.

Wößmann, L. 2016. The Importance of School Systems: Evidence from International Differences in Student Achievement. In: Journal of Economic Perspectives, 30: 3–32.

Wößmann, L., F. Schoner, V. Freundl and F. Pfaehler. 2023. Wie ungerecht sind die Bildungschancen von Kindern aus verschiedenen Familien in Deutschland verteilt? ifo-Schnelldienst 4/2023. München: ifo-Institut.

Wößmann, L. 2024. Erkenntnisse aus aktuellen Schulleistungsstudien zur Evaluation des Bildungssystemes. Eine bildungsökonomische Perspektive. In: N. McElvany, M. Becker, H. Gaspard, F. Lauermann und A. Ohle-Peters, Eds, Evakuation des Bildungssysteme. Welche Erkenntnisse liefern die Schulleistungsstudien? Münster und New York: Waxmann: 9-32.

Wößmann, L., F. Schoner und F. Pfaehler. 2024. Ungleiche Bildungschancen: Ein Blick in die Bundesländer. Ifo Schnelldienst 5/2024. München: ifo-Institut.

Zimmer, R. W., and E. F. Toma. 2000. Peer effects and educational vouchers: evidence across countries. In: Journal of Policy Analysis and Management, 19: 75–79.