RESHAPING ABILITY GROUPING THROUGH BIG DATA

Yoni Har Carmel, University of Haifa
Tammy Harel Ben-Shahar, University of Haifa

INTRODUCTION

Ability grouping, otherwise known as tracking, affects millions of students in the country daily. It shapes aspects of schooling that we know to be crucially important for students: the curriculum they study, the resources they receive, the teachers who educate them, and the peers with whom they interact. Critics of ability grouping insist that it reinforces educational inequalities, directing students from racial and ethnic minorities or from poor families to lower tracks in which they receive inferior schooling and limited opportunities. Proponents, on the other hand, argue that homogeneous classes are more effective as they allow teachers to adjust the content and methods of teaching to the needs of the students. All concede, though, that ensuring that the process of grouping is free from biases and does not aggravate racial and class segregation, is imperative.

Despite being one of the most controversial issues in education for almost a century, the educational practice of ability grouping persists, and for the past decade or so, is thriving. This resurgence of ability grouping coincides with another momentous change in education, namely the technological and information revolution. This development, which is already affecting ability grouping practices in many schools around the country, includes the collection of vast amounts of information concerning students, which can then be analyzed by algorithms in order to generate predictions about their educational attainments. One of the most interesting questions concerning this technological development is whether it will alleviate the biases that plague ability grouping, and what role law can play in ensuring it will. These two questions are the focus of this Article.

There is a vast body of legal scholarship concerning equality and segregation in education. Some of it has referred specifically to ability grouping and its relation to segregation. The introduction of technological tools for ability grouping, however, warrants a renewed interest in this topic. There is also significant work pertaining to the ethical and legal ramifications of big data and predictive analytics, but only sparse attention is given within this body to the educational arena.
What little research there is, focuses mostly on issues of privacy, protection of student data, and prevention of monetizing student information. This Article aims to address this gap, and bring together these two distinct areas in law, the integration of which introduces completely novel and complex issues that have import for both areas of law.

The Article argues that information technologies in education can have important benefits: it can improve education by enabling individually tailored teaching, by monitoring student progress and evaluating teachers and schools. Additionally, in light of the biases that persistently affect traditional ability grouping, data driven ability grouping (DDAG) offers significant promise. It can potentially remove prejudice from educational decisions, offsetting implicit biases that teachers may unknowingly hold. In a recent study concerning the use of EVAAS (Education Value-Added Assessment System) for assigning students into different tracks in eighth grade mathematics,1 teachers reported that the algorithm assigned students to a high track that would otherwise not have been identified as suitable for the program, thus increasing shares of children from racial minorities and low socioeconomic status in the program.2 However, DDAG also creates a whole host of unique challenges in terms of equality of opportunity that need to be addressed. Studies concerning data mining and predictive analytics in other domains such as crime prevention, banking, insurance, and commerce suggest that instead of eliminating biases, algorithms recreate social biases.3 In order to identify students likely to succeed in an advanced course, for example, an algorithm examines information concerning past students from which it infers the attributes of successful students. However, if past admission decisions were racially biased, the algorithm’s decisions will simply mirror this bias.

Additionally, algorithms can only use the data they have. Students from privileged background have more access to digital devises outside school meaning that there are likely to be more entries into the system, more academic interaction and task engagement, that will later positively

---

1 According to the Company’s website, EVAAS is widely used to place students in eighth grade algebra. The system evaluates several years of end-of-grade testing to predict a student’s ability to study higher-level subjects, and accordingly, suggests lists of students that would be good candidates for eighth-grade algebra.
2 Shaun M. Dougherty et al., Middle School Math Acceleration and Equitable Access to Eighth-Grade Algebra Evidence from the Wake County Public School System, 37(1) Educ. Evaluation & Pol’y Analysis 80 (2015), http://epa.sagepub.com/content/37/1_suppl/80S. The study also found that the rates of success did not decline subsequently.
affect the predictions concerning them.\textsuperscript{4} Students from privileged background are also considerably more digitally literate, resulting in better functioning in a digital environment. These disparities do not reflect unequal academic ability; however, they can cause the algorithm’s prediction to be biased against children from poor families or racial minorities.

Finally, data driven decision-making may create new classes of children who are disadvantaged. Although law is primarily concerned with biases against students who belong to groups who are historically socially excluded, such as racial minorities or immigrants, algorithmic decision-making may create new groups that are systematically unfairly disadvantaged. If, for some inexplicable reason, children who are color blind, or who engage in after school sports are less able to do well on computerized tasks, and therefore the predictions are less favorable with regard to them, this entails that they will be discriminated against by data driven decision making (DDDM) in general and specifically in processes of ability grouping.

The fact that it is widely believed to be scientific and objective makes biases in algorithmic decision-making worse, at least in one sense, than biases in traditional decision-making. Inequalities that result from algorithmic predictions are treated as inevitable, and may serve to justify existing inequality. Additionally, predictions generated by algorithms are almost impossible to challenge, as algorithms are extremely opaque and complex. This problem is especially challenging in the educational domain. When identifying potential tax evaders, an algorithm-based alert is then validated by an actual audit, so false predictions can be detected. An innocent individual may be inconvenienced as a result of being targeted by the algorithm, but this harm is relatively contained. Conversely, a prediction that leads to the assignment of a student to a high ability track does more than indicate the student’s ability. It constitutes it. By influencing the curriculum she is taught, the skills she develops, the peers she interacts with, the expectations teachers have from her and the expectations she has of herself, the algorithm’s prediction is self-fulfilling. Her subsequent achievement is admittedly the result of her innate ability, but also of all the other benefits she was accrued following the assignment decision, and disentangling the accumulative effects of all these is impossible. This hinders the ability to validate the algorithm’s initial prediction effectively.

What is equally troubling, in terms of protection against biased decision-making in education, is that existing equal protection doctrines, intentional discrimination, disparate impact and rational basis test, cannot appropriately address the challenges of data driven ability grouping. Equal protection doctrine in the US precludes differential treatment according to race and other suspicious classifications. However, the framework of intentional discrimination is unsuited to contend with overrepresentation of minority students in low tracks, because such assignment decisions cannot be characterized as intentionally discriminatory.

Disparate impact doctrine seems more suited for challenging biased decision making in education. However, courts have deferred to school districts’ expertise and accepted that even if such practices have a segregatory effect, it is a justified educational practice. Therefore, it is unlikely to be effective with biased outcomes in DDAG.

The rational basis test, that has been criticized for being too weak in general, is likely to be nothing but a rubber stamp in the case of DDAG. Algorithms, as a whole, are largely inexplicable, as they integrate hundreds of different factors in order to come up with their predictions. As a result, any decision according to algorithms is arguably “rational”, and will therefore withstand judicial review.

Is there, then, no way to ensure equal opportunities in an era of data driven decision making? We argue that the solution lies in the combination of technological solutions and legal regulation, both performed in the stage of the design and use of algorithms.

Information scientists have begun seeking technological solutions to the discriminatory outcomes that algorithms have been showed to have. These include removing from the datasets suspect attributes (such as race or gender) and attributes that correlate with suspect attributes (zip code may correlate with race, for example); and manipulating historical datasets by removing biased decisions. Additionally, algorithms may be able to completely reshape grouping, for example by replacing the traditional criterion of academic performance with other attributes that were previously impossible to ascertain, such as different learning styles and strategies. This kind of ability grouping may promote the goal of designing effective teaching without creating racial and class segregation.

The technological solutions, however, depend on numerous normative decisions, that cannot be divorced from legal doctrine. Law determines which classes are protected (e.g.: is socioeconomic class a suspicious classification?); whether unequal outcomes constitute an actionable
wrongdoing; and whether affirmative action is permissible. These normative decisions must inform the efforts of algorithmic design.

The technological and structured nature of algorithmic decision making, that enables designers to control the attributes taken into consideration and even the desired result, suggests that the involvement of legal and normative considerations at the stage of design can be effective in improving the outcomes in terms of equality. In traditional methods of ability grouping, performed by humans, it is almost impossible to impose rules concerning which data to use (and which to disregard) and to assign specific weight to each piece of information. Teachers used students’ grades, and their own impressions and “hunches” and made a decision. Biases were (hopefully) subconscious and unintended, but could not be avoided. In DDDM, there are many more possibilities to affect the process of decisions and the integration of student data, so it could be possible to correct for some biases.

It should be noted that some of the technological solutions may be, themselves, subject to legal challenges. It could be argued that the process of correcting biases amounts to differential treatment according to race, for example. While these are cogent concerns that would need to be addressed, careful design could probably avoid many such challenges, while still enabling influencing decision making for the better.

I. THE PRACTICE OF ABILITY GROUPING

A. What is ability grouping?

One of the greatest challenges of comprehensive education is that students vary widely in their innate abilities, knowledge and learning styles. Providing instruction that is suitable for all students, sufficiently challenging but not overwhelming, is an excruciating task. Faced with this challenge, many education systems group students based on their academic ability to decrease student diversity and enable teachers to match the content, pace and complexity of their classes to their students, who are all, supposedly, more or less at the same level.5 The primary purpose of this practice is to increase students’ academic growth and achievement by providing instruction

---


- Draft –

Please do not circulate or cite without permission
at their current instructional level. It is also argued that this can enable educators to provide compensatory education and special attention to students who need them.

In its widest interpretation, ability grouping includes also programs for the gifted, on the one hand, and placement in special education, on the other. While we will not refer to these further in the paper, clearly some of the discussion applies to these cases too.

Ability grouping can take on various forms that differ from one another along several dimensions. First, ability grouping can differ with regard to how fixed/flexible it is. Grouping can be ad-hoc, performed by a teacher within a diverse classroom for a specific task, dissolving immediately after the performance thereof. On the other end of the spectrum, students can be assigned to homogeneous groups in separate classes, tracks or schools, from which there is little possibility to move.

A second and related dimension concerns the scope of separation. In some cases, students can be assigned completely different schools or tracks, in which no mixed ability learning or social interaction takes place. In other cases schools are comprehensive, and ability grouping may be used for assigning students to specific courses.

Another difference between types of ability grouping policy concerns the age in which ability grouping is performed. In Germany and Austria, for example, students are tracked at the end of fourth grade into separate schools, constituting a very early grouping system, whereas other educational systems are comprehensive until the higher grades. Ability grouping can sometimes

---

8 For one, assignment decisions in these fields too are likely to become exceedingly data driven. Additionally, traditional decision-making in these fields is also extremely biased against children from disadvantaged backgrounds.
9 Maureen T. Hallinan et. al, Ability Grouping and Student Learning, 6 Brookings Papers on Educ. Pol’y 65, 103 (2003). Assignment to lower track courses can also cause a “locking out” effect, when assignment to high level courses are conditioned on prerequisite course completion.
transcend the classical division into grades, with cross-grade grouping, an option to address high ability students’ need for accelerated teaching in certain topics.\textsuperscript{12}

Ability grouping may differ also with regard to whether the curriculum taught in the different courses, groups or tracks is different. There is no necessary link between ability grouping and curriculum differentiation.\textsuperscript{13} For example, when grouping first grade children according to their ability for reading tutoring, the goal is to promote their reading skills. Although there may be temporary differences in the reading material they are given, the curriculum is ultimately the same, the pedagogical aims are identical, and the main difference concerns the pace of learning. There are, on the other hand, cases of ability grouping in which this is clearly not the case and different tracks are taught completely different content, imparted with different skills, etc.\textsuperscript{14}

Education policy can combine the various dimensions depicted above. Some may be exceedingly separated – tracking students early, separating them completely, and offering them different curriculum; and others can be relatively integrated, with only ad-hoc separation, or separation in older children for specific courses. Notwithstanding the variety described above, there are no education systems in the developed world in which there is no ability grouping at all, making the question of justice in ability grouping universally relevant.

The United States education system can be characterized, roughly, as a system in which ability occurs late and does not regularly create complete separation between tracks by maintaining comprehensive schools. In Elementary schools, within class ability grouping is typically used for reading and math instruction, and instruction is targeted to the proficiency level of each group. Most Americans are familiar with assignment of elementary school students to reading groups, often with inoffensive names designed to obscure ability level, such as the Blue Birds and Red Birds, the Robins and the Sparrows.\textsuperscript{15}

\textsuperscript{12} James A. Kulik & Chen-Lin C. Kulik, \textit{Meta-analytic Findings on Grouping Programs}, 36(2) Gifted Child Q. 73, 75 (1992). The most popular between-class grouping plan is the Joplin Plan. The earliest version of this plan included the cross-grade grouping of elementary students in reading. During the time reserved for reading, students in grades 4, 5, and 6 would proceed to different classrooms to receive instruction and use materials geared to their readiness levels.


\textsuperscript{14} \textit{Id}.

\textsuperscript{15} Kulik & Kulik, \textit{supra} note 12, at 75.
Another common ability grouping option is tracking. Scholars have defined tracking in various, and sometimes conflicting, ways. Tracking as we use the term, is a form of ability grouping in which students are separated not only for ad-hoc, specific tasks, but rather have a relatively fixed, separated educational program, with differentiated curriculum. In this paper we refer to ability grouping as the general phenomenon that encompasses all the different forms described above. The process of grouping is done according to the students’ purported capacities for learning; assignment to math tracks is based on math proficiency, English tracks on reading proficiency, etc. However, the dimensions of ability grouping vary relatively independently between schools and within schools over time.

B. Ability Grouping and Educational Equality

Education researchers fiercely debate the effectiveness of ability grouping. Despite being hotly contested for over three decades, the jury is still out on whether teaching students together with peers of the same ability level enhances educational attainment. While some studies have found

---

16 Oakes states, “Tracking is the process whereby students are divided into categories so that they can be assigned in groups to various kinds of classes”. Jeannie Oakes, Keeping Track: How Schools Structure Inequality 3 (1985). Gamoran et al., refer to “tracking” as: “… the practice of dividing students for instruction according to their purported capacities for learning.” Adam Gamoran et al., An Organizational Analysis of the Effects of Ability Grouping, 32(4) Am. Educ. Res. J. 687, 688 (1995). Akos et al. define it in the following way: policy involves a school organization structure that increases the homogeneity of instructional groups by stratifying students by curriculum standards, educational career goals, or ability. See Patrcik Akos et al., Early Adolescents’ Aspirations and Academic Tracking: An Exploratory Investigation, 11(1) Prof. Sch. Couns. 57, 58 (2007).

17 The names of high school courses signal curricular differences. Advanced math students in tenth grade, for example, may take Algebra II while others take Geometry, Algebra I, or Pre-Algebra. Advanced tenth graders in English language arts (ELA) may attend a class called “Honors English” while other students attend “English 10” or “Reading10.” Excellent science students may take “AP Chemistry” while others take a course simply called “Chemistry” or “General Science.” History may also be tracked, as when Advanced Placement courses are offered in U.S. or European history that not all students take.

positive effects for students studying in homogeneous classes,\textsuperscript{19} others have found little or no such effects.\textsuperscript{20} The effects may also differ between the high tracks and the low tracks, with various studies suggesting that students in high tracks enjoy education advantage as a result of the separation and students in the lower tracks have no comparable gains,\textsuperscript{21} and are even disadvantaged by the separation into ability based classes.\textsuperscript{22}


\textsuperscript{21} In a meta-analysis of findings from 52 studies about ability grouping, studies in which high-ability students received enriched instruction in honors classes produced especially clear effects, while studies of average and below average students produced near-zero effects (Chen-Lin C. Kulik & James A. Kulik, Effects of Ability Grouping on Secondary School Students: A Meta-analysis of Evaluation Findings, 19(3) Am. Educ. Res. J. 415 (1982)); Ireson & Hallam (supra note 5) found high self-esteem in high ability groups; Shields showed that some form of homogeneous grouping benefits the most able and gifted students in terms of their academic achievement, as well as their attitudes concerning themselves as learners and regarding their school experiences; Carolyn M. Shields, A Comparison Study of Student Attitudes and Perceptions in Homogeneous and Heterogeneous Classrooms, 24(3) Roeper Rev. 115 (2002); Gamoran et al. found that growth in student achievement is significantly lower in general-track classes than in college-preparatory classes (Adam Gamoran et al., Upgrading High School Mathematics Instruction: Improving Learning Opportunities for Low-Achieving, Low-Income Youth, 19(4) Educ. Evaluation & Pol’y Analysis 325 (1997).

\textsuperscript{22} In a study of data from the Early Childhood Longitudinal Study, Lleras & Rangel found that students placed in low achieving reading groups actually learn less every year as they progress through school, thus increasing the achievement gap. Christy Lleras & Claudia Rangel, Ability Grouping Practices in Elementary School and African American /Hispanic Achievement, 115(2) Am. J. Educ. 279 (2009).
Overall, most writers on grouping have concluded that grouping students by academic performance typically contributes to widening the achievement gap between high-level and low-level classes over time, even after accounting for initial differences in ability.\textsuperscript{23} Clearly, the success of students in ability-based groups depends on other factors too, such as adapting instruction to the needs of the group, and being allocated sufficient resources.\textsuperscript{24} Unfortunately, a sufficient body of research shows that ability grouping results also in inferior treatment of students in the lower tracks, in the following ways:

\textit{Resources} students in lower tracks are allocated less resources than students in the higher tracks.\textsuperscript{25}

\textit{Teachers} students in higher tracks are systematically assigned better and more experienced teachers than the students in lower tracks.\textsuperscript{26}

\textit{Peer effects} some studies show that grouping students by ability results in a reduction of peer effects in general.\textsuperscript{27} Others, however, show that there are differences between the low and high tracks in peer effects: while grouping creates a resource-rich environment for high-level students,


\textsuperscript{24} Noting the importance of adapting instruction in order for ability grouping to succeed see: Lou et al. (supra note 19, at 448). See also Slavin (supra note 10, at 311): “regrouping for reading and/or mathematics can be effective if instructional pace and materials are adapted to students’ needs, whereas simply regrouping without extensively adapting materials or regrouping in all academic subjects is ineffective.”

\textsuperscript{25} Karl L. Alexander et al., \textit{Curriculum Tracking and Educational Stratification: Some Further Evidence}, 43(1) Am. Soc. Rev. 47, 64 (1978);


it deprives students in the low tracks of an important classroom resource, namely the positive input of high ability peers.\textsuperscript{28}

Skills and Curriculum Some research suggests that learning opportunities are influenced by the educational structure of ability grouping. Those channeled into lower classes are frequently provided a substantially different curriculum and set of learning experiences.\textsuperscript{29} Oakes found that the students who were placed in the more advanced tracks learned skills essential to critical thinking, thereby improving their problem solving skill set.\textsuperscript{30} Others too have found that teaching in the higher tracks is more rigorous.\textsuperscript{31} In contrast, instruction in low-ability classes tends to be more fragmented, dwelling on isolated bits of information, using workbooks and low-level pedagogy, and it progresses at a slower pace.\textsuperscript{32}

Discipline and Dropout Students placed in lower academic tracks were found to have higher dropout rates.\textsuperscript{33} This does not mean that the grouping, in itself, is to blame for dropout rates, however, it may aggravate the tendency.\textsuperscript{34} Moreover, student misbehavior is often treated differently in high and low levels and disciplined more severely in the lower tracks.\textsuperscript{35}

Labeling and Self Esteem Ability grouping labels students, thus creating long-term effects such as lower teacher expectations and self-expectations in terms of their ability for academic success.\textsuperscript{36} Expectations have a surprisingly significant influence on educational ability and


\textsuperscript{29} Gamoran et al., \textit{supra} note 16, at 692.

\textsuperscript{30} Oakes, \textit{supra} note 16.


\textsuperscript{32} Oakes, \textit{supra} note 16, at 93-112. James Rosenbaum’s Making Inequality: The Hidden Curriculum of High School Tracking (1976) described working class youth at a New England high school who were channeled into vocational and remedial tracks that were nothing more than boring, academic dead ends.


\textsuperscript{34} Jacob Werblow et al., \textit{On the Wrong Track: How Tracking is Associated with Dropping Out of High School}, 46(2) Equity & Excellence Educ. 270, 272 (2013).

\textsuperscript{35} Mary H. Metz, \textit{Classrooms and Corridors: The Crisis of Authority in Desegregated Secondary Schools} 106 (1978).

\textsuperscript{36} Alexander et al., \textit{supra} note 25, at 60; Harker & Tymms, \textit{supra} note 26.

- Draft –

Please do not circulate or cite without permission
achievement, so the fact that ability grouping diminishes expectations for certain students is problematic. A related problem concerns the influence on children’s self-esteem that is harmful regardless of whether it affects educational achievement. In most cases, once students are placed in a lower academic track in the early grades, they remain there through high school, where the differences between tracks become more pronounced. Additionally, students assigned to lower track courses often find themselves “locked out” of higher-level courses in the upper grades due to school policies that require prerequisite course completion into higher-tracked courses.

The evidence presented raises grave concerns that instead of promoting the educational position of students with lower abilities, investing in them, and pushing them to perform better, ability grouping in fact causes them further disadvantage. The findings are all the more troubling since considerable research supports the assertion that racial minorities and students from poor background are severely overrepresented in low tracks whereas students from privileged backgrounds tend to be assigned in higher proportions to higher tracks. Thus, Latinos/as, African Americans, and students from lower socioeconomic backgrounds are often less likely to enroll in Advanced Placement (AP) or higher-level courses. African American, Native American, and Latino/a students are also significantly underrepresented in programs for the gifted, while Asian Americans are overrepresented in relative to their representation in the school

37 Pallas et al. found that students in high-ability classes typically are exposed to a more positive learning environment, in regards to attitude, aspirations, and self-esteem, than those in low-ability classes. Aaron M. Palls et al., *Ability-Group Effects: Instructional, Social, or Institutional?*, 67(1) Soc. Educ. 27 (1994). See also: Losen, *supra* note 33, at 522.


population. International studies draw similar conclusions: countries that track their students into ability groups have greater educational gaps than countries that do not track. As a result, ability grouping seems to be aggravating racial and social segregation, causing injustice far beyond schooling and denies students meaningful opportunities in life. It undermines the ability of education to allow upward social and economic mobility for poor and minority students.

These conclusions cast doubt not only as to the desirability of ability grouping as an educational practice but also to its moral permissibility. While conceding that ability grouping is a contested educational practice, this Article does not address these issues in detail. As was noted previously, ability grouping is performed, routinely, in all education systems, and this is unlikely to change. Therefore, we focus on what we believe may be a more valuable contribution to the promotion of educational justice – on legal methods of optimizing data driven assignment decisions so that biases may be minimized.

The fact that children from disadvantaged backgrounds are overrepresented in low tracks can be attributed to one of the following causes: The first explanation consists of “pre-grouping causes”. Ability grouping, according to this explanation, merely reflects social inequality that makes children from marginalized groups less well equipped for school. Individuals from disadvantaged groups tend to have less nurturing environments, resulting in diminished abilities when they enter school. As presented above, there is disagreement whether ability grouping is able to remediate these past disadvantages, or whether they actually aggravate them.

The second possible cause for overrepresentation lies within the process of ability grouping itself. Biases in the process of determining the assignment of each student results in the

---


44 Mickelson & Greene found that during middle school and high school gaps in student achievement levels become increasingly larger as a consequence of not only the differentiated early instruction and curriculum exposure, but also because of the vast differences in learning opportunities associated with participation in the honors and college preparatory programs in the middle grades and high schools, respectively. Roslyn A. Mickelson & Anthony D. Greene, _Connecting Pieces of the Puzzle: Gender Differences in Black Middle School Student’s Achievement_, 75(1) J. Negro Educ. 34 (2006).

45 Although it could be argued that creating and improving methods for ability grouping may have the effect of embedding this practice
overrepresentation of disadvantaged students in lower tracks, including students that could have been successful in higher tracks.

Clearly, the two causes are not mutually exclusive. Longstanding social inequality is surely to blame for inequalities in educational capabilities for children from different social groups, however there is evidence that educational decision making is deeply afflicted with racial and class biases. Though the educational inequality that results in pre-existing social inequality is an important challenge, the rest of the paper focuses only on the second cause, namely biases in decision-making, and examines whether data driven decision-making, coupled with appropriate legal regulation, is likely to overcome these biases.

C. The Resurgence of Ability Grouping in the Digital Age

Ability grouping has been a part of education for well over a century. While a full history of ability grouping is beyond the scope of this paper, it is nevertheless important to recap a few key historical markers. This discussion is all the more important as much of this history has followed a perplexing pattern of use and disuse that has tended to maneuver around social class and racial/ethnic compositions of the schools and communities.

In the early days of comprehensive schooling, ability grouping was motivated largely to separate children of lower classes and immigrant children who were largely uneducated from those of educated gentry. After Brown ability grouping expanded dramatically, coming to represent a means of circumventing desegregation by substituting within-school segregation for what had previously existed between schools.

In the 1985 groundbreaking Keeping Track: How Schools Structure Inequality, Jeannie Oakes argued that although tracking was typically justified by educators as a strategic response to student heterogeneity, the practice was undergirded by normative beliefs regarding race and class, and politically defended by white, middle-class parents to protect privilege. Ability grouping provided a convenient excuse to separate students within schools, ostensibly by ability. Racial, ethnic and socioeconomic segregation continued to exist because minority and poor

47 Losen, supra note 33, at 521.
children were disproportionately assigned to, and remained in, low levels, and white, wealthier children to high levels. Ability grouping caused racial and ethnic separation both because students of disadvantaged groups were objectively academically inferior, but also because racial and socio-economic biases entered into grouping decisions.\footnote{Id.}

Ability grouping did not just mirror the inequalities of the broader society, as Oakes and other critics charged, it produced and perpetuated inequality.\footnote{Oakes, \textit{supra} note 16; Jeannie Oakes & Amy Stuart Well, Beyond the Technicalities of School Reform: Policy Lessons from Detracking School (1996).} This ongoing de-facto segregation attracted public criticism, and was even legally challenged. As a result, the practice of ability grouping in elementary schools dropped significantly from the beginning of the 1990s.\footnote{Tom Loveless, \textit{The Resurgence of Ability Grouping and Persistence of Tracking}, 3(2) The Brown Center Report on American Education 12 (2013).} But ability grouping has made a strong comeback in the 2000s in classrooms all over the country. According to a 2013 study, over 70 percent of fourth-grade teachers who participated in a survey said they had grouped students by reading ability in 2009, up from 28 percent in 1998. In math, over 60 percent of fourth-grade teachers grouped students by ability in 2011, a 40 percent increase since 1996. As for tracking, it has remained commonplace in eighth-grade mathematics for the past two decades, with about three-quarters of students enrolled in separate ability-level math classes. The reemergence of ability grouping has surprised education experts who believed the outcry had all but ended its use.\footnote{Id.}

The concerns over the effect of ability grouping on the achievement gap between white students and their minority peers have not eased in the new resurgence in the last decade\footnote{Richard R. Verdugo, \textit{The Heavens May Fall: School Dropouts, the Achievement Gap, and Statistical Bias}, 43(2) Educ. & Urb. Soc. 184 (2011).} as empirical research on the formation of ability groups and curricular tracks consistently indicate that ability group assignment is still correlated with socioeconomic status, race, and ethnicity, whereas racial-ethnic minority students are disproportionately placed into lower levels.\footnote{Werblow et al., \textit{supra} note 34.}

One central explanation for the resurgence of ability grouping in the 2000s concerns the accountability measures adopted in the bipartisan 2002 reauthorization of the federal Elementary and Secondary Education Act (ESEA), commonly known as the No Child Left Behind Act.
(NCLBA). The NCLBA maintained that high standards and measurable goals for individual students would improve student outcomes in education. The Act calls for each state to develop an accountability system that describes how the state will take responsibility for the academic achievement of all students, including subgroups of students considered most vulnerable to failure. Subgroups include groups of students defined by race or ethnicity, those eligible for subsided lunch, English language learners, and students with disabilities. By forcing teachers to focus on students who fell just below the threshold for “proficiency” on state tests, the law may have encouraged teachers to group struggling students together to prepare them for standardized tests. We will discuss this legislation further in part III.

There is a further development in this final phase, which is at the center of this Article, namely the rise of information and communication technologies. This development will be detailed further in the second section of the Article, but for now suffice to say that it has had effect not only on the methods with which grouping is performed but also on their prevalence. Information technologies offer easy access to data that can facilitate ability grouping. This access enables and encourages ability grouping. Before moving to discuss this issue further, we describe the biases in the traditional method of ability grouping.

D. Biases in the process of grouping

Well documented evidence points to bias in traditional educational decision making against racial minorities, children from low social class, and girls. Teachers are unaware of their biases and believe they are judging students equally, however it is proved that they tend to judge

56 NCLBA § 1001(1) (Statement of Purpose).
57 § 1001(2).
60 George Ansalone, Keeping on Track: A Reassessment of Tracking in the Schools, 7(3) Race, Gender & Class in Educ. 108 (2000).
equally qualified students from racial minorities as less academically and socially competent than students from non-minority students, and to underestimate the actual academic abilities of students belonging to racial minorities. These biases pervade all spheres of schooling. Black children, for example, are more likely to be disciplined for misconducts that white children could get away with, and to suffer more severe punishments for the same behavior. Biases have also been found in decisions concerning placement in special education, with research showing that children from racial and ethnic minorities are three times more likely to be found in need of special education in cases of disabilities that involve subjective teacher evaluations, a bias that is not found in cases of “objective” disabilities. And while the legal treatment of discrimination and the attitudes in society toward racial equality have developed significantly since these topics were first studied, these findings have not changed in recent years. Implicit biases still pervade subconscious attitudes that affect decision making.

Moreover, studies show that even when schools employ a set of criteria in placement decisions (most often grades, test scores, teacher and counselor recommendations, parental preference, and student choice) nonacademic factors play a significant role in determining the ability group level to which a student is assigned. Things like students’ social skills, physical attractiveness, style of dress, whether both parents are present in the home, and even their first names, affect teachers’ evaluations of student ability, leaving space for biases to infiltrate decision-making.

---

62 Linda van den Bergh et al., The Implicit Prejudiced Attitudes of Teachers: Relations to Teacher Expectations and the Ethnic Achievement Gap, 47(2) Am. Educ. Res. J. 497 (2010); Regina Cecelia McCombs & Judith Gay, Effects of Race, Class, and IQ Information on Judgments of Parochial Grade School Teachers, 128(5) J. Soc. Psychol. 647 (1988); La Yonne I. Neal et al., The Effects of African American Movement Styles on Teachers’ Perceptions and Reactions, 37(1) J. Special Educ. 49 (2003); Felicia R. Parks & Janice H. Kennedy, The Impact of Race, Physical Attractiveness, and Gender on Education Majors’ and Teachers’ Perceptions of Student Competence, 37(6) J. Black Stud. 936 (2007); Glock et al. showed that, in Luxembourg, minority students more frequently received recommendations for lower school tracks and that they were expected to have a less successful educational career than majority students. Sabine Glock et. al, Beyond Judgment Bias: How Students’ Ethnicity and Academic Profile Consistency Influence Teachers’ Tracking Judgments, 16(4) Soc. Psychol. Educ. 555 (2013).


64 Steve Knotek, Bias in Problem Solving and the Social Process of Student Study Teams: A Qualitative Investigation, 37(1) J. Special Educ. 2 (2003). Findings show that these self-fulfilling prophesies were especially harmful for black students, girls and for students from low socio-economic status.


66 Ansalone, supra note 60, at 127.
Parental influence is also an important factor in determining placement. Not only are poor minority parents less likely to complain about ability grouping decisions than white middle-class parents, but white middle-class parents are also more likely to be assertive in gaining access to the upper tracks' and gifted and talented programs than are minority or poor parents.\(^67\)

These findings are not surprising. Extensive research has been conducted examining the impact of implicit biases against members of minority groups in contexts other than education such as law enforcement, employment, etc., that shows the pervasiveness of implicit bias in every field.\(^68\)

Teachers and education administrators are, like all humans, subject to biases, and there are no immediate solutions available. Can algorithms offer a solution to this problem by doing a better job and making unbiased decisions? This will be examined shortly, but first we examine legal challenges to inequality in ability grouping.

II. LEGAL CHALLENGES TO ABILITY GROUPING

Like many other educational controversies over the past half a century, the issue of student grouping has been almost as likely to be tested in the courtroom as in the classroom. Ability grouping has been challenged through the Fourteenth Amendment Equal Protection Clause and Title VI of the 1964 Civil Rights Act.

A. Equal Protection

The first and most publicly known case to deal with the segregatory and discriminatory effect of ability grouping was *Hobson v. Hansen*\(^69\). The case challenged the practice in schools in the District of Columbia, in which students were assigned to one of several tracks that ranged from "basic" for the slow students to "honors" for gifted students and completed virtually all their course work within such a differentiated curriculum. The children were assigned to the different tracks based on intelligence, achievement and aptitude test scores. The policy resulted in blatant segregation in schools, with the higher tracks serving an overwhelming majority of white students, and black students assigned to mostly lower tracks. The District Court ruled that ability grouping as it was practiced in the D.C. Schools violated the Due Process Clause of the Fifth Amendment.

---

\(^{67}\) Losen, *supra* note 33, at 525.


Amendment. In thus deciding, the court stressed several important points: first, the segregatory effect of grouping in a school district that was under a desegregation order; second, that education in the lower tracks was so watered down that it could more aptly be described as warehousing; and that the program did not involve review of the initial tracking assignments. Another important finding was that the process determining the grouping misclassified a large number of students. Therefore, the court found that the tests used in this case do not give an accurate estimation of the ability of the children from racial minorities and that “rather than being classified according to ability to learn, these students are in reality being classified ... according to environmental and psychological factors which have nothing to do with innate ability”.

The applicability of the Hobson ruling, however, is critically limited. In Hobson, the plaintiffs had, themselves, been the victims of racial segregation throughout their prior education. The Hobson court was clear that ability grouping is not unlawful per se, but objected to the fact that the tests reflected the inequality caused by this segregation, and that ability grouping did nothing to aid students overcome it. However, if it is reasonably related to a legitimate educational objective, and implemented in a non-arbitrary, capricious, or discriminatory way, ability grouping was deemed a legitimate policy.

In another case, Moses v. Washington Parish School Board the court was faced with ability grouping in a recently desegregated school district. Here, the previously white school absorbed all students, and continued a grouping system it practiced prior to desegregation, that was comprised of eleven homogeneous levels rather than following the traditional elementary school structure of six grades. Tracking was not held illegal per se in this case either, but rather not allowed in recently desegregated schools due to the fact that they suffered from inferior prior education.

In 1976, the Supreme Court draw a firm distinction between de jure discrimination which results from intentional state action, and that violates the Equal Protection Clause, and de facto discrimination, that refers to racially disproportionate impacts of state action. As a result, subsequent legal challenges of tracking have met with inconsistent results. Rather than address

---

70 Id. at 514.
72 Id. at 852.
the constitutionality of academic tracking generally, courts have decided its constitutionality on a district-by-district basis.

In the mid1970s the Fifth Circuit in the McNeal case,74 ruled that ability grouping that has a racially disparate impact may be permitted if it satisfies two conditions: first, the school district achieved “unitary” status, and the second is that ability grouping can be justified by demonstrating either that the assignments do not result from past segregation, or that they will remedy such results through better educational opportunities.75 Where there was a historic educational inequality, the court assumed that classroom groupings would “predictably cause students from the inferior system to immediately be re-segregated within the lower classroom sections.” Therefore, the school districts must not employ grouping until they have operated unitary systems for a sufficient period of time needed to ensure that students do not underachieve because of past injustices. This period will end, according to the court, when the district can show that “steps taken to bring disadvantaged students to peer status have ended the educational disadvantages caused by prior segregation.” However, this statement was not intended as a hard-and-fast rule, and the court did leave space for cases in which evidence might show that a given system of grouping is in the best interest of students.

The time lapsed since formal racial segregation was abolished following Brown, has made the McNeal standard, that applies only to students who were themselves educated in segregated schools, practically obsolete. Courts do not view current racial segregation in ability grouping as a result of past educational segregation, and therefore do not regard it as discrimination.

In NAACP v. Georgia in which no student had attended a segregated school, the court found that segregation could not be blamed for inequality, even if the district was under a desegregation decree and had not achieved unitary status.76 The fact that the students’ parents attended segregated schools, and that educational disadvantage is largely inherited, and therefore

74 McNeal v. Tate County School District 508 F.2d 1017 (1975).
75 Id. at 1020.
76 775 F.2d 1403 (11th Cir. 1985). The court not only found that the ability grouping practice that tracked students as early as kindergarten was justifiable, it also stated that the practice corrected the effects of past segregation by providing remediation to black students. It is unlikely that a court would decide in the same way today, given that the view that all students benefit from high expectations is, in large measure, a shared conclusion among researchers and politicians alike. See also Montgomery v. Starkville 854 F.2d 127 (1988) in which the fifth circuit court ruled that because the district had been under desegregation order for 20 years, past segregation could not be blamed for ability grouping’s disparate impact.
continues to afflict students, was deemed irrelevant to the current grouping system. The court not only found that the ability grouping practice that tracked students as early as kindergarten was justifiable, it also stated that the practice corrected the effects of past segregation by providing remediation to African American children. Since NAACP, challenges to practices of tracking that were based on racial imbalance have been a tough battle to win. Courts have upheld racial imbalance in cases of student choice and racial achievement gaps. As a result, in order to address racial disparity in ability grouping, opponents must prove that racial disparities in academic placement are the result of intentional racial discrimination.

To conclude, the judiciary is increasingly reluctant to exercise of local decision makers regarding ability grouping. And while school districts that were under desegregation orders in the past are considerably more vulnerable to equal protection arguments than those that who were not, the

77 Id. at 1412.
78 The NAACP case was followed by decision in Montgomery v. Starkville 854 F.2d 127 (5th Cir.1988) where the fifth circuit court ruled that because the district had been under desegregation order for 20 years, past segregation could not be blamed for ability grouping’s disparate impact. The Court in this case concluded that the “minimal segregative effect” caused by the achievement grouping was “outweighed by better educational opportunities afforded the students” by the achievement grouping program. A similarly detrimental decision was made in
79 Quarles v. Oxford Mun. Separate Sch. Dist., 868 F.2d 750 (5th Cir. 1989). In this case the court could not deny that there was “a high concentration of White students in the upper level groups and” a high concentration “of Black students in the lower level groups.” However, the Quarles court determined that the present unequal impact that the current grouping system had on White and Black children was not a result of the school's former segregated school system, which met the requirements previously stated by the Supreme Court.
80 In People Who Care v. Rockford Board of Education the district court found that the acts and omissions of the school district in its tracking system not only had a discriminatory effect, but also were strong evidence of discriminatory intent. Based on the conclusions of the expert, the court ordered that “ability grouping and/or tracking will no longer be allowed in the Rockford schools”. On appeal, however, the court found that the provisions of the decree generally forbidding grouping of students by ability and establishing racial quotas for the permitted gifted and talented programs were inequitable. The court further stated that “[w]ere abolition of tracking the only means of preventing the school district from manipulating the tracking system to separate the races, it might be a permissible remedy.” The People Who Care court concluded that the school should be enjoined from “twisting the criteria to achieve greater segregation than objective tracking alone would have done ... [but] not, on this record, enjoined from tracking”. 111 F.3d 528, 536 (1997).
81 See: Simmons v. Hooks 843 F Supp. 1296 (E.D. Ark. 1994) which required the school to alter its tracking system, at least to the extent that the ability grouping practices were vestiges of previous intentionally discriminatory practices; Price v. Austin Indep. Sch. Dist., 945 F.2d 1307 (5th Cir. 1991), ruling that once the school system had been held "unitary," the burden shifts to the plaintiff to show that a newly adopted student assignment plan with a disparate impact on minorities is intentionally discriminatory; In the Yonkers case the court said that since ability grouping in the district was based on teacher's attitudes and expectations that could be traced to prior segregation, the ability groups themselves were a form of segregation. United States v. Yonkers Bd. of Educ., 123 F. Supp. 2d 694, 695 (2000).
willingness to intervene even in those cases in quite minimal. At the same time, there is a growing number of racial and ethnic minority children whose ancestors did not attend segregated schools, either because they were not in the South or because they immigrated to the US after Brown, but who are nonetheless being disadvantaged by ability grouping. These students, as well as students in previously segregated districts that achieved unitary status cannot, generally, successfully challenge ability grouping practices that result in racial segregation. The remaining path to ground an Equal Protection claim, therefore, is to show that ability grouping consisted in intentional discrimination. It is, however, extremely difficult to prove the requisite element of intent, therefore most constitutional challenges are unlikely to succeed.

B. Rational Basis Test

The rational basis test is a less stringent standard of review. It is applied in cases in which the differential treatment is based on a classification that is not regarded as suspicious (or when the rights that are curtailed are not “fundamental”). Socioeconomic class, for example, which is especially pertinent classification with regard to ability grouping is one such classification. In order to challenge a policy concerning ability grouping that entails treating poor individuals differently than their privileged peers, plaintiffs are required to prove that the policy bears no rational relation to a legitimate governmental interest.

Given that the Supreme Court declared that education is not a fundamental right, and as persons other than those in “suspect” categories may be involved, it seems reasonable to utilize the lesser standard suited to the problems arising within the special and unique confines of the public education system.

The rational basis test is, even in theory, very permissive. The courts application of it through the years has made it practically unusable. In one case the court explained that “if there is any

---

82 Losen, supra note 33, at 532.
84 Angelia Dickens suggests that the Court should adopt the “belief in the fundamentality of education” adopted by Justice Marshall in his dissent in San Antonio Indep. Sch. Dist. v. Rodriguez, 411 U.S. 1, 55 (1973) and further argues that the practice constitutes a classification based on race that should he subject to strict scrutiny. Thus, under Dickens's formulation, a school district would be required to show that ability grouping is “narrowly tailored to serve a compelling state interest.” In the author's view, a district will likely not be able to establish a compelling interest for tracking; and therefore, an Equal Protection challenge to ability grouping under her framework for strict scrutiny analysis would likely succeed. Note, Revisiting Brown v. Board of Education.: How Tracking Has Resegregated America's Public Schools, 29 COLUM. J.L. & SOC. PROBS. 469, 473-74 (1996).
conceivable state of facts that could provide a rational basis” for a challenged law, it will survive rational basis review. Moreover, the court stated that it was irrelevant whether the rationale given for challenged distinction actually motivated the legislature, suggesting that any plausible reason can suffice whether or not it was the true reason for legislation. Additionally, the standard of proof required of plaintiffs is extremely high, creating a “virtually irrebuttable presumption of constitutionality under the rational basis test”. The plaintiff must negate “every conceivable justification” for the challenged regulation. As a result, critics argue that the rational basis test, as it stands, is too weak and provides little protection for rights.

It should be noted that there have been several cases over the years in which courts apply the rational basis test differently, sometimes called “rational basis with a bite”. In these cases the court did not give automatic credence to the rationales offered by the government, and instead examined each reason to examine whether the ordinance or regulation was reasonably related to the governmental interest, however these have not been in the educational domain.

To conclude, it is unlikely that the rational basis test can be an effective means for judicial intervention. The cases described above show that courts generally defer to education

89 As Jackson (supra note 86, at 493) says: “The Court has essentially made the rational basis test the equivalent to no test at all”.
90 The term “rational basis with bite” originated with Gerald Gunther. See Gerald Gunther, Foreword: In Search of Evolving Doctrine on a Changing Court: A Model for a Newer Equal Protection, 86 HARV. L. REV. 1, 21 (1972); See also Gayle Lynn Pettinga, Note, Rational Basis with Bite: Intermediate Scrutiny by Any Other Name, 62 Ind. L. J. 779 (1987); David O. Stewart, Supreme Court Report: A Growing Equal Protection Clause, 71 A.B.A.J. 108, 112-114 (1985). Some commentators (and indeed some dissenting justices) argue that in these cases the court is, in effect, applying an intermediate level of scrutiny, without explicitly saying so. Others however point to the fact that the test is a means-ends test, that doesn’t follow the intermediate scrutiny tests (Jackson, supra note 86, at 540). Robert C. Farrell, Successful Rational Basis Claims in the Supreme Court from the 1971 Terms Through Romer v. Evans, 32 Ind. L. Rev. 357 (1999) counts ten cases in 25 years in which this rational basis with a bite has been applied, compared to 100 cases in which it has been rejected.
administrators even in equal protection cases, when the standard of review is higher, therefore there is virtually no chance that ability grouping could fail the rational relations test.

C. Title VI of the 1964 Civil Rights Act

An additional normative source for protection against discrimination is title VI of the Civil Rights Act. While ability grouping litigation has most often involved contentions of racial segregation and discrimination, grouping practices on the basis of national origin or language, in addition to race, may also be challenged under Title VI of the 1964 Civil Rights Act, a general antidiscrimination law that bars discrimination on the basis of race and national origin in programs and services operated by recipients of federal financial assistance.\textsuperscript{92}

Under Title VI, where ability grouping results in significant levels of classroom segregation, even when no intention to discriminate exists, the district may find itself in noncompliance. The difficulties in proving intentional discrimination (as well as the fact that most cases of inequality in ability grouping is, luckily, unintended) makes doctrines of indirect discrimination extremely important for promoting equality. However, such policies can be redeemed if they are justified from an educational perspective, and if the policy is the least segregative instructional approach from among equally effective educational alternatives. As was previously noted, Courts have deferred to professional expertise concerning the question whether ability grouping is overall better for students. The practical meaning of this deference is that in school districts in which ability grouping is employed, courts have viewed this as educational necessity and therefore upheld the practice.\textsuperscript{93} The idea that disparate racial impact derives from nonracial factors has been a foundation for judicial deference to the expertise of school officials under both the Equal Protection Clause and Title VI. Courts have refrained from seriously considering the possibility that even good faith efforts at grouping could be biased. Instead, they have view ability grouping as a scientific or technical issue.\textsuperscript{94}

\textsuperscript{92} Ability grouping policies or processes that operate to discriminate on the basis of student gender are also prohibited by Title IX of the Education Amendments of 1972, 20 U.S.C. §§ 1681-86.

\textsuperscript{93} In the NAACP case referred to above, the court referred to both equal protection claims and claims according to title VI. The court ruled that a racially disparate grouping system did not violate Title VI because grouping was necessary to meet the needs of the student population and was an "accepted pedagogical practice". 775 F.2d at 1418 (quoting the district court record).

\textsuperscript{94} Note, supra note 84, at 1326.
Therefore, unless it can be demonstrated that the choice of ability grouping was unreasonable in the circumstances of the case, Title VI cases are likely to be just as ineffective in challenging ability grouping as strict scrutiny or the rational basis test. To conclude, there has been limited success in challenging ability grouping in courts, and the framework enabling a successful challenge is largely obsolete as time lapses from formal de-jure segregation. Current doctrine does not seem promising in terms of challenging the unequal effects that ability grouping creates for members of disadvantaged groups. As will be argued later on, these traditional tools are even less likely to be effective in challenging data driven ability grouping, and therefore different means of regulation are required. We move on to examine whether data driven decision making is likely to be less biased than traditional decision making, and after that put forward an alternative framework for regulating DDAG, consisting of a combination between technological and legal solutions, that will be better equipped to contend with the challenge of ensuring equal educational opportunities.

III. BIG DATA-DRIVEN ABILITY GROUPING

Two related challenges have spurred educational reform in the US Over the past thirty years: how to make schools more effective and equitable in their outcomes, and how to make them publicly and politically accountable for delivering those outcomes. Increasingly, there has been a strategic convergence of these two approaches by using more stringent accountability as the prime driver to improve educational performance. There has also been a growing commitment to a particular approach for achieving accountability. This has involved collecting, analyzing, and reporting performance data of various kinds at the student, teacher and school levels. More recently, “a fundamental philosophical shift has occurred from data for compliance to the principles of data for continuous improvement”. Data are no longer just to hold people and

---


schools accountable. Instead, data are to be used to stimulate continuous improvement and inform educational decisions.97

A. Educational Data-Driven Decision Making

The literature often characterizes DDDM in the educational context as a practice where teachers are active participants who systematically collect data, interpret these data, formulate action plans, and continuously evaluate and adjust their plans based on further data.98 In the practice of data-driven decision making, educators are supposed to examine data, make sense of it, and then translate their understanding of data into action, such as instruction.99

Although law-created expectations for informed decision making in education is not new,100 this process received a boost with the NCLB Act, that has played a major role in further propelling education administrators towards educational data based teaching. Specifically, the NCLB Act imposed financial and administrative sanctions based on student test scores, thereby making test scores a primary concern to all educators. Part of the NCLB legislation required a focus on closing the achievement gap in each school based on its demographics and achievement scores.101 To achieve this goal, the law requires that states measure achievement annually and evaluate that achievement in light of interim achievement goals established by the state.102

The broad implementation of standards-based accountability under the NCLB Act has presented new opportunities and incentives for data use in education by providing schools and districts


100 Hargreaves & Braun, supra note 95, at 1. The concept of DDDM in education is can be traced to the debates about measurement-driven instruction in the 1980s; state requirements to use outcome data in school improvement planning and site-based decision making processes dating back to 1970s and 1980s; and school system efforts to engage in strategic planning in the 1980s and 1990s.

101 § 1001(3).

102 § 1011. The state determines annually whether each district and school has made “Adequate Yearly Progress (AYP)”. For a school and district to meet AYP each subgroup must reach an identical minimal level of proficiency for each school year. Failing to meet the AYP entails sanctions for the school and district’s operation and autonomy.
with additional data for analysis, as well as increasing the pressure on them to improve student test scores in mathematics, language arts, and science. As mentioned earlier, one of the strategies that educators have used to meet these targeted needs and to increase achievement based on standardized assessments is through the practice of ability grouping. These different motivations coincide encouraging both ability grouping and using information technology to do so.

The Race to the Top (RTT) reform was meant to contend with the widening of the achievement gap that occurred during the period of the NCLBA. It turned the focus from student achievement to student growth. Many states have passed legislation that link teacher evaluation with student growth measures rather than student achievement levels. Another change that RTTT created was to move away from sanctions for schools and districts, and instead to offer states a heavy financial incentive to implement data-use policies and invest in data-use infrastructure. Almost every aspect of the RTTT competition guidelines required states and educators to use data to make instructional decisions. From teacher evaluations to the creation of longitudinal state-wide data bases to evaluating colleges of education, RTTT pushed states and local districts to incorporate DDDM.

The federal emphasis on accountability and measurement coincides the development of data related technology and fuels it. States and local districts were encouraged to implement data-use policies that are intended to foster data use. A recent study indicated that 30 states and DC now have legislation or policies that require student performance data be used to “significantly” inform the criteria for the evaluation of teacher effectiveness and subsequent decision-making

104 The RTTT program is a $4.35 billion fund created under the American Recovery and Reinvestment Act of 2009. It is the largest competitive education grant program in U.S. history, warranting unprecedented transparency and participation to ensure the best possible results. RTTT competition is designed to provide incentives to states to implement large-scale, system-changing reforms that improve student achievement, close achievement gaps, and increase graduation and college enrollment rates. See U.S. Department of Education, Race to the Top-Game-Changing Reforms (n.d.), available at https://www.ed.gov/open/plan/race-top-game-changing-reforms.
A common theme across states and districts data-use policies is the requirement of teachers to regularly analyze student performance data in order to make and justify instructional decisions. The increased demand to base decisions on data consequently is bolstered by an industry of assessment systems, longitudinal statewide databases, data warehouses, and other technological tools for educators. These systems and annual standardized assessments fostered an influx of data into schools, which encouraged educators to access, analyze, and use these data to inform their decisions. Consequently, today’s educators are not only exposed to more data than ever before but also expected to use it, particularly in low-performing schools that are under great pressure to meet their accountability targets.

In December 2015, a new Federal legislation was enacted, Every Student Succeeds Act (ESSA). Under the new law states are embarking on a new era of school accountability. However, the law is remarkably consistent with NCLB in terms of helping states enhance their efforts to use more accurate and transparent data about the performance of all students. While law based educational reforms played a role in encouraging DDDM, there has been little scholarly discussion concerning law’s role in regulating its effects. The research that has been conducted in the field of educational DDDM, has focused on issues of privacy, data protection and the commercialization of student information. The possible ramifications of DDDM in terms of equality, which is the topic of this Article, have yet to receive serious scholarly attention.

A. Using Educational Data Mining for Ability Grouping

As education becomes more data-driven, today’s core educational activities increasingly rely on technology. Learning management systems (LMSs), digital whiteboards, digital textbooks, educational applications, mobile devices, online assessments, and others, enable optimizing teaching techniques, communication between teachers and students and other advantages.

---

110 Id.
use of these technologies creates vast amounts of granular learning related data including students’ everyday interaction with digital environments.

In the current learning environments, users learn in online communities like discussion forums, online chats, instant messaging platforms and various LMSs like Moodle. Moodle is a popular open-source Learning Management System (LMS) that can be used for, inter alia, task assignments, quizzes, content delivery, etc. The system contains detailed logs of user actions available through teacher or administrator user interface. Moodle will log user information, view and upload commands, start and end time, order of questions, question time, and correctness. Students have started using smart phones to access learning content access their courses anywhere and indulge in learning activities. Every keystroke of every student is incidentally recorded in log files and are potentially open to analysis. Although not yet used in most operative systems, technology already enables the systems to follow bodily movements and conditions such as heart rate and eye movement.

The information collected as a side effect of the use of education technologies opens up a new world of evidence and insight on students: how they learn, what their abilities are, and what are their educational needs.

The amount of data made available in the above scenarios is so enormous that traditional processing techniques simply cannot be used to process them. Indeed, the limitations of conventional applications in utilizing the enormous quantities of data have caused schools and reformers to increasingly explore “big data” technologies to process the educational data. Big data is not easily defined, but, in general refers to “large and complex datasets collected from digital and conventional sources that are not easily managed by traditional applications or processes”. By integrating data gleaned from all applications, systems and sources into a central technology foundation, institutions can more easily analyze and convert data into

---

113 Moodle is the acronym for Modular Object-Oriented Dynamic Learning Environment.


actionable insights. In order to do so, it is not enough to collect and record activity and interactions and technologies of Educational Data Mining (EDM) and Learning Analytics are required.

EDM is described as “the area of scientific inquiry centered around the development of methods for making discoveries within the unique kinds of data that come from educational settings, and using those methods to better understand students and the settings which they learn in”. The term refers to techniques, tools, and research designed to automatically extract meaning from large repositories of data generated by or related to people’s learning activities in educational environments. EDM can be used to determine patterns in large and noisy datasets, such as incidentally recorded data (e.g. log files, keystrokes), unstructured data (e.g., text files, discussion threads), and complex and varied, but complementary data sources (e.g., different environments, technologies and data models).

The goals of EDM include both descriptive findings concerning learning and learners and predictive findings pertaining to students’ future learning behavior. A descriptive model identifies relationships or patterns in data - the objective is to uncover and summarize patterns or features that exist in data sets. Predictive data analytics is designed to characterize specific cases, generating a predicted value or classification of each case without regard to the utility of the model for understanding the underlying structure of the data. For example, using predictive

---

121 EDM researchers view the following as the goals for their research: (1) predicting students’ future learning behavior by creating student models that incorporate such detailed information as students’ knowledge, motivation, metacognition, and attitudes; (2) discovering or improving domain models that characterize the content to be learned and optimal instructional sequences; (3) studying the effects of different kinds of pedagogical support that can be provided by learning software; and (4) advancing scientific knowledge about learning and learners through building computational models that incorporate models of the student, the domain, and the software’s pedagogy. See B R Prakash et al., Big Data in Educational Data Mining and Learning Analytics, 2(12) Int’l J. Innovative Res. in Computer & Comm. Engineering 7515, 7516 (2014).
models, educational institutions developed systems to identify students at risk of dropping out to consequently take needed measures to retain those students in advance.

Grouping students is one of the most common uses of EDM. The objective is to create groups of students according to their customized features, personal characteristics, etc. Grouping students using DM techniques can be performed in two distinct methods: classification or clustering.

Clustering is a descriptive data mining model which consists in dividing the different students into clusters that consists of objects that are similar to one another and dissimilar to objects of another group. Clustering students would result in grouping students according to their similarities. These would typically include their grades, knowledge in a particular field, capabilities, skillsets, etc., but could also include learning styles and habits, or family background, hobbies and the like. Importantly, as this is a descriptive model, there may well be surprising results to clustering, allowing the algorithm to discover unexpected similarities and dissimilarities that traditional education decision makers would have no way of knowing.

---


123 Milan Vukicevic et al., *Grouping Higher Education Students with RapidMiner*, in RapidMiner: Data Mining Use Cases and Business Analytics Applications 185, 185 (Markus Hofmann & Ralf Klinkenberg eds., 2013).


125 The clusters that are formed need to satisfy the following two principles: (1) homogeneity: elements of the same cluster are maximally close to each other; (2) separation: data elements in separate clusters are maximally far apart from each other. Id., at 7.

126 Different clustering algorithms have been used in k-12 and higher education to form homogenous groups of students, such as: hierarchical agglomerative clustering, K-means and model-based clustering to identify groups of students with similar skill profiles; a clustering algorithm based on large generalized sequences to find groups of students with similar learning characteristics based on their traversal path patterns and the content of each page they have visited; a hierarchical clustering algorithm for user modeling (learning styles) in intelligent e-learning systems in order to group students according to their individual learning style preferences; a fuzzy clustering algorithm to find interested groups of learners according to their personality and learning strategy data collected from an online course; hybrid method of clustering and Bayesian networks to group students according to their skills; a K-means clustering algorithm for effectively grouping students who demonstrate similar learning portfolios (students assignment scores, exam scores and online learning records); and more. See: Vukicevic et al., *supra* note 123, at 189; Ashish Dutt et al., *Clustering Algorithms Applied in Educational Data Mining*, 5(2) Int’l J. Info. & Electronics Engineering 112, 113 (2015); Li Li et al., *Clustering Students for Group-Based Learning in Foreign Language Learning*, 9(2) Int’l J. of Cognitive Informatics & Natural Intelligence 55, 57 (2015).

---

- Draft –
Please do not circulate or cite without permission
The second method is classification. This is a predictive data mining task in which algorithms are on existing datasets in order to establish connections between different attributes of datasets.\textsuperscript{127} Thus, an attribute is chosen as the one to be predicted – success in a certain course for example. The algorithm then is required to classify students into the different categories. In order to make this prediction the algorithm uses all the information it is fed, generating high predictability rates, and finding correlations with attributes that cannot be learnt any other way.\textsuperscript{128}

Schools are increasingly utilizing these technologies for optimizing decision making in education.\textsuperscript{129} Of the different systems that are available today, SAS® EVAAS® (Education Value-Added Assessment System, developed by SAS), for example, provides analytical services, including value-added modeling and projection analyses, for the assessment of schooling effectiveness at the district, school and, when requested, at the classroom level. The system allows users to access reports that show everything from teacher effectiveness, to student probability of success in a future course, to reports that provide all students who are at-risk of not being proficient on state-wide assessments. EVAAS analyses make use of scores on standardized tests such as those provided by major educational testing companies and those used by states to fulfill their NCLBA obligations.\textsuperscript{130}

According to the Company’s website, EVAAS is widely used to assign students to eighth grade algebra. The system evaluates several years of end-of-grade testing to predict a student’s ability

\textsuperscript{127} Pedro G. Espejo et al., \textit{A Survey on the Application of Genetic Programming to Classification}, 40(2) IEEE Transactions on Sys., Man, and Cybernetics-Part C: Applications and Reviews 121 (2010).

\textsuperscript{128} Several classification algorithms have been applied in order to group students, such as: discriminant analysis, neural networks, random forests and decision trees for classifying university students into three groups (low-risk, medium-risk and high-risk of failing); classification and regression tree, chi-squared automatic interaction detection and C4.5 algorithm for the automatic identification of the students’ cognitive styles; decision trees for classifying students according to their accumulated knowledge in e-learning systems; genetic algorithms for grouping students according to their profiles in a peer review content; C4.5 decision tree algorithm for discovering potential student groups with similar characteristics who will react to a particular strategy; etc. See Romero & Ventura, \textit{supra} note 122, at 9.

\textsuperscript{129} Some adaptive learning systems, use Bayesian network to assign students into groups based on similarities in their learner models, either similarities in their current knowledge state, or by similarities in their cognitive and non-cognitive characteristics. Ioannis Magnisalis et al., \textit{Adaptive and Intelligent Systems for Collaborative Learning Support: A Review of the Field}, 4(1) IEEE Transactions on Learning Tech. 5, 8 (2011).


- Draft –
Please do not circulate or cite without permission
to study higher-level subjects, and accordingly, suggests lists of students that would be good candidates for eighth-grade algebra.\textsuperscript{131}

Although systems such as EVAAS have not been operational for long, research is already beginning to emerge. One study, aimed at examining teachers’ use and perceptions of EVAAS, found that 19 percent of teachers who used EVAAS data, stated they used it for ability grouping, to differentiate instruction, to and provide remedial education.\textsuperscript{132} EVAAS’s ‘probability of success’ reports have also become a determinant factor in math placement policy at district levels. For example, Wake County North Carolina decided that achieving a certain level of ‘probability of success’ would be the criterion for assigning student to an accelerated track in Math.\textsuperscript{133}

\textbf{A. Will EDM Mitigate or Exacerbate Biases in Ability Grouping Decisions?}

In light of the persistent biases that plague traditional methods of educational decision-making, DDDM with its purported scientific and objective nature, may well be a welcome step in the right direction. Data, it is argued, “doesn’t lie”,\textsuperscript{134} therefore, decisions based on data mining can provide an objective, accurate indicator of students’ learning and abilities. Although the extent to which data drives or supports educators’ decisions is debated, overall, proponents of DDDM believe that the decision-making process is more accurate and effective when decisions are based on data, as opposed to educators’ judgment or professional wisdom.\textsuperscript{135} In this part of the paper we show that while this method of decision-making may potentially alleviate some of the problems with traditional decision-making, it raises an independent set of concerns in terms of equality of opportunity. Considering the high stakes involved in ability grouping decisions and

\begin{itemize}
  \item \textsuperscript{132} Clarin Collins, Houston, \textit{We Have a Problem: Teachers Find No Value in the SAS Education Value-Added Assessment System (EVAAS®)}, 22(98) Educ. Pol’y Analysis Archives 1, 14 (2014) However, even among the teachers who indicated that they used EVAAS for ability grouping, differentiating instruction, and, almost no one actually articulated how the data were specifically used.
  \item \textsuperscript{134} Duncan, \textit{supra} note 97.
  \item \textsuperscript{135} Mandinach, \textit{supra} note 97; Jeffrey R. Henig, \textit{The Politics of Data Use}, 114(11) Tchr. C. Rec. 1 (2012).
\end{itemize}
the pervasive biases that are currently demonstrated in ability grouping, addressing these concerns is particularly important.

One underlying assumption driving the use of data within educational systems is that it introduces objectivity and lessens discrimination in an evaluative process that has traditionally been heavily influenced by subjectivity. Students from racial minorities or those living in poverty stricken environments may view DDAG as an opportunity to have objective decisions made about them that are derived from data, as opposed to stereotypes or prejudice. Indeed, the U.S. Department of Education promotes the collection and analysis of information generated by and about students as a means to help close achievement gaps, increase educational opportunities and college access, and reduce discrimination against underserved students. Yet, As Coburn and Turner assert, “one of the central lessons from research on data use in schools and school districts is that assessments, students’ tests, and other forms of data are only as good as how they are used”. In other words data, in itself, is not a panacea able to solve all the ailments of educational inequality, and a more critical reflection is required in order to ascertain whether DDDM holds real promise, and what are the pitfalls we should be wary of.

Since the use of big data in education is in its early days, the evidence is still not conclusive as to its ability to solve biases in decision-making. It is often assumed that big data techniques are unbiased because of the scale of the data and because they are implemented through algorithmic systems. However, the evidence that already exists together with insights on data mining and predictive analytics from other fields can inform this discussion.

In one study, a North Carolina school district sought to increase the disproportionately low rates of enrollment in advanced math classes among female, black, Hispanic and low-income students. To achieve this goal, in the 2010–2011 school year, the district decided to use masked performance data, analyzed by EVAAS rather than teacher recommendations to determine mathematics course assignment. By assigning students based on this data, rather than on intuitive decisions, the district substantially improved overall rates of math acceleration among black, Hispanic, and low-income students, and raised female math acceleration to reflect

- Draft –
Please do not circulate or cite without permission

138 Dougherty et al., supra note 2.
their proportional enrollment in the district – all without impacting the successful completion rates.\textsuperscript{139}

Another interesting finding concerns the reaction teachers had to the assignment recommendation EVAAS generated: teachers were surprised when the model identified many students who had not been considered promising candidates for advancement.\textsuperscript{140} Presumably, these students might have been overlooked for the recommendation to take accelerated-level courses as a result of variation in course-grading practices and subjective beliefs about which students are capable of success in these courses. Naturally, further research is required to demonstrate the scope of variance between traditional decision making and big data-driven decision making, however these initial findings are encouraging, suggesting that big data can solve some of the implicit biases involved in traditional educational decision making.

Despite the justified optimism, the belief that large data sets “offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy”\textsuperscript{141} is largely exaggerated. The use of big data creates a unique set of challenges, including that it, too is biased, and therefore recreates inequality. These concerns must be acknowledged and regulation should be designed in order to minimize their effect.

The claim that algorithms will classify more ‘objectively’ (thus solving previous inadequacies or injustices in classification) cannot simply be taken at face value given the degree of human judgment still involved in designing the algorithms. This human involvement includes defining features, pre-classifying training data, and adjusting thresholds and parameters. In developing models data miners link their own meanings, values, and assumptions to similar ones taken from the problem and the intended intervention.

A White House report from 2014, titled “Big Data: Seizing Opportunities, Preserving Values” (also known as the Podesta Report), similarly suggests that there may be unintended discriminatory effects from data mining, stating that “The increasing use of algorithms to make eligibility decisions must be carefully monitored for potential discriminatory outcomes for disadvantaged groups, even absent discriminatory intent”.\textsuperscript{142}

\begin{itemize}
\item[\textsuperscript{139}] Id.
\item[\textsuperscript{140}] SAS\textsuperscript{®}, supra note 131.
\item[\textsuperscript{141}] Danah Boyd & Kate Crawford, Critical Questions for Big Data, 15(5) Info., Comm. & Soc’y 662, 663 (2012).
\item[\textsuperscript{142}] White House, Big Data: Seizing Opportunities, Preserving Values 47 (2014), available at http://perma.cc/8VPX-DB8E.
\end{itemize}
A further report issued in May 2016, titled “Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights”, divided the possible discriminatory effects of big data into two categories: “(1) Challenges relating to data used as inputs to an algorithm”; and “(2) Challenges related to the inner workings of the algorithm itself”. We will follow this framework and discuss technical issues which potentially hold discriminatory implications for data driven ability grouping process.

i. Data used as inputs

The decision to use certain inputs for algorithms to process rather than others can result in discriminatory outputs. Suspicious classifications theorem expression of this involves allowing attributes such as race ethnicity or gender to feature in the datasets. Moreover, as algorithm scientists working in this field have recognized, merely removing the suspicious attribute may not completely eliminate the discriminatory effects because other attributes that are not themselves suspicious often strongly correlate with those that are. An additional problem is that removing suspicious classifications prevents education administrators and researchers from using algorithms to identify cases of inequality, to evaluate their persistence over time and to experiment with ways to alleviate inequality. Therefore, although ‘color-blind’ algorithms seem “bias-safe”, this comes with a price in terms of the ability to use the technology in the service of promoting equality.

The control over inputs – which to use and which to preclude – can also be used to mask and justify discrimination. As several writers have persuasively argued, the use of correlations uncovered by data science can potentially give rise to “inequality on an unprecedented scale triggered by … ‘smart discrimination’”. A programmer can embed bias in a complex algorithm such that discrimination will be very difficult to detect. The machine can find strong correlations, which result in discriminatory outcomes that are based on neutral factors. It is wrong to discriminate based on race; yet it will be exceedingly difficult to detect such

---

discrimination if it is based on a dozen factors that through big data analytics are found to be positively correlated to race.146

Attributes that are correlated with suspicious classifications while data mining certainly introduces novel ways to discriminate intentionally and to conceal those intentions, concerns with this potential should not crowd out careful consideration of the unintentional discrimination that is likely to be more common than the kinds of discrimination that commenters fear could be pursued intentionally.147 Situations of this sort are possible when the criteria that are genuinely relevant in making rational and well-informed decisions also happen to serve as reliable proxies for class membership.148

Even when algorithms are not supplied with attributes such as race, gender, nationality, there are often other attributes that correlate with suspicious classifications. Thus, due to residential segregation, the seemingly neutral attribute of zip code can denote race. In other cases, a combination of different attributes may indicate membership in a racial or ethnic group and removing all traces of suspicious attributes may prove virtually impossible.

Importantly, some of the correlating attributes can be inherently relevant for educational decision making so that removing them is likely to impede the accuracy of the prediction. For example, the classification of students as ELL (English Language Learners) correlates with immigrant status. This might be deemed a suspect classification and we might not want the algorithm to have access to that information. On the other hand, this seems like relevant input concerning the students, that would be important for optimal educational decision-making, not only in terms of assignment decisions but also in order to offer them appropriate educational support.

Other examples of data that might be relevant for educational decision making but may also correlate with a suspect classification include data concerning discipline or attendance records,

146 Omer Tene & Jules Polonetsky, Judged by the Tin Man: Individual Rights in the Age of Big Data, 11(2) J. Telecomm. & High Tech. L. 351, 358 (2013). In 2015, This American Life detailed the story of a school district in Normandy, Missouri, that demonstrates how data can be used to justify excluding disadvantaged students. In Normandy school district, a mostly black and poor school district, with abysmal graduation rates and test scores, students had the option to attend a higher performing and mostly white school in their district. Using crime statistics and test scores, parents at the higher performing school advocated blocking disadvantaged students from entering their school. While this story is not one about big data, it does demonstrate how neutral data that correlates race can be used in this way, and this ability will be reinforced by big data analytics.

147 Barocas & Selbst, supra note 3, at 23.

148 Id. at 21.
which all can correlate with racial or socioeconomic background. If these correlations become implemented into the grouping algorithms, big data would allow for subliminal forms of discrimination of members of protected groups.

Poorly selected data often, members of a protected class are disadvantaged because the factors that better account and describe them are not well represented in the set of selected features. Members of protected classes may find that they are subject to systematically less accurate classifications or predictions because the details necessary to achieve equally accurate determinations reside at a level of granularity and coverage that the features fail to achieve.149 This problem stems from the fact that data are by necessity reductive representations of an infinitely more specific real-world object or phenomenon. These representations may fail to capture enough detail to allow for the discovery of crucial points of contrast. Increasing the resolution and range of the analysis may still fail to capture the mechanisms that account for different outcomes because such mechanisms may not lend themselves to exhaustive or effective representation in the data, if such representations even exist.150 Although this is a general characteristic of data analytics, it has disparate effect for minorities, both because they often have less exposure to technology rich environment and because the programmers may be biased. Obtaining information that is sufficiently rich to permit precise distinctions can be expensive, because some of it may not be readily available as a side product of existing activities. Even marginal improvements in accuracy may come at significant practical costs, and may justify a less granular and encompassing analysis. To take an obvious example, some student grouping models tend to assign a considerable weight to the amount of time a student spends doing homework on an LMS. Algorithm developers may favor this type of data because it communicates pertinent information at no cost, despite the fact that it may communicate very little about a student’s academic abilities, and this disadvantages students from poor background that tend to use LMS after school hours much less than their privileged peers. A NAEP research from 2012, suggests that students with less familiarity with computers devote more cognitive resources and time to entering information on keyboards and navigating digital menus with less

149 White House, supra note 143, at 7.
150 Toon Calders & Indrė Žliobaitė, Why Unbiased Computational Processes Can Lead to Discriminative Decision Procedures, in Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases 43, 47 (Bart Custers et al. eds., 2013).
time and focus to organize and communicate ideas.\textsuperscript{151} This may disproportionately affect minorities, poor students, or English language learners who generally are less likely to have home Internet access.\textsuperscript{152}

Also, relying on information obtained from technological applications have discriminatory effects caused by the gap in technological proficiency between students from privileged background, who have high quality access to internet at home, and those who did not.

\textit{Perpetuating past inequalities} algorithms make their predictions based on past data. If we wish to predict student success in a course, for example, algorithms analyze the data of past students (called “training dataset”), and find which attributes (or the complicated combination thereof) best correlates success. Thus, if successful participants in honors classes were, historically, mostly white and privileged, the algorithm will try to locate candidates who are similar to successful past graduates and a perpetuation of inequality is to be expected. Social inequality and biased decision making in the past, is therefore likely to resurface through the training dataset. Data mining can “inherit the prejudices of prior decision-makers, or simply reflect the widespread biases that persist in society”.\textsuperscript{153} This problem can be contended with by manipulating the training dataset, although this raises separate issues that will be discussed shortly.

\textit{ii. The Algorithm}

Two main issues are relevant in this category, the infiltration of human biases into the design of algorithms, and the opaqueness of algorithms. \textit{Human biases in algorithm design} despite the fact that algorithms operate “independently”, discovering the connections that are simply there “in” the data, they are still, ultimately, designed and programmed by humans. Human biases are therefore able to seep into the process of big data driven ability grouping. Human involvement in algorithm design occurs in various stages.

\textsuperscript{153} Barocas & Selbst, \textit{supra} note 3, at 671.
Programmers define the features and attributes the algorithm analyzes; they classify and organize training data that algorithms base their predictions upon; they adjust certain thresholds, thus determining acceptable outcomes and predictions; they also determine (in predictive models, at least) the “question” the algorithm aims to answer. This framing function is far from neutral. An algorithm used to assign students to a course, for example, can be programmed to predict which students are most likely to succeed in a course, but it could also be designed to predict which students are likely to benefit the most from the course (in terms of added value). The difference between these is extremely important, and programmers must make these value-laden decisions when designing algorithms.

Opaque algorithms even for the best informed programmer, an algorithmic predictive model is something of a “black box”—an opaque machine that takes inputs, carries out some inscrutable process, and delivers unexplained outputs based on that process. Their complexity may often hinder attempts to unpack their rationale. Algorithms are often difficult to diagnose and their outcomes problematic to challenge. Moreover, the technical processes involved in algorithmic systems are often zealously guarded trade secrets treated as confidential or proprietary to the entities that use them making outside agents, even the individual to whom the decision applies, unable to access anything but the most general explanation as to the decision process and the information it was based upon. Therefore, the affected individuals, such as those students who are being grouped, have limited ability to learn the reasons why such decisions were made, and cannot challenge the results.

iii. Why are These Problems Especially Troubling in Big Data-Driven Ability Grouping?

Big data driven decision-making processes are, therefore, also subject to biases. To add to this concern, is the fact that at least in one aspect data driven decision-making is actually worse than traditional decision-making. The purported objectivity of big data masks discrimination and prevents meaningful discussion pertaining to it. Discriminatory outcomes are excused and explained by data’s scientific nature, and therefore made to feel benign. The algorithms’

---

154 Citron & Pasquale, supra note 144, at 6.
155 Id., at 5.
unrestricted ability to identify patterns in endless piles of data facilitates the masking of intentional or unintentional discrimination behind layers upon layers of mirrors and proxies. In education this is especially problematic. In other fields it is possible to examine the predictions after the fact: when an algorithm points to a potential suspect, a policeman can search the individual for drugs or illegal weapons and thus to disprove the algorithms prediction. However, when an algorithm assigns students to different ability groups, there is no parallel way to test the validity of the algorithm and refute its prediction. Examining the final achievements of students in the courses they are assigned to is unhelpful because educational predictions have an influence upon students’ abilities. Teachers made aware of students’ ability levels are affected by this information, and unintentionally have different interactions with students according to what the teachers perceive as their level of ability. This differential treatment reinforces and advantages those with higher perceived ability. Additionally, as ability grouping most often involves curriculum differentiation, students perceived as having higher ability are also granted better resources and taught higher skills, thus further affecting their abilities.

Moreover, because ability grouping creates a status hierarchy, the assignment of students to groups constitutes a status allocation in which some are elevated above others. When ability-group assignment is correlated with sociodemographic factors, it reinforces status distinctions that originate outside the school. This labeling effect has a long lasting detrimental effect in terms of students’ self-respect and sense of worth. Being diagnosed by the algorithm, therefore, has constitutive power. It not only assesses one’s ability, it also influences it. Disentangling the effects of prior ability, expectations, and instruction in the educational outcome is almost impossible. What results is a method of prediction that is virtually unsusceptible to criticism.

iv. Technological Solutions

Scientists have attempted to offer technological solutions to some of these problems, and have used various methods to do so. While these attempts are commendable, a normative discussion of these solutions is indispensable. Algorithms, as a technological tool, are able to reach different results, according to the way they are programmed, but it is up to their designers to decide which...

156 Tene & Polonetsky, supra note 146, at 358.
157 Gamoran et al., supra note 16, at 689.
of the results are desirable. For example, algorithms can be required to assign students to a course without taking gender into consideration, in a (seemingly) gender neutral way. Conversely, they can be programmed to ensure equal gender representation, thus changing the other parameters for decision (instating differential criteria for girls and boys). The choice between these two outcomes is not technological, but rather normative, and therefore one that should be discussed by lawyers and educators.

We distinguished previously between unequal outcomes that are caused by social inequality (one that exists before the grouping decision and is unrelated to it) and those that are caused by biases in the process of decision-making. This distinction resurfaces now, as we move to examine regulation of DDAG.

It should be stressed from the outset that using algorithms to correct pre-existing social inequality compromises the predictability of the algorithm. The unequal outcomes of the algorithm in this case indicates that the algorithm “got it right”.158 There may still be reasons to design algorithms in a way that does not reflect this reality, that have been well argued in contexts of affirmative action. Most pertinent to our discussion is that some of the educational disadvantage that minority students suffer is relatively easy to overcome with appropriate instruction. As social inequality is to blame for the lack of nurturing of natural talents, perhaps the algorithm is able to overlook the neglect and deprivation and detect the innate abilities. If so, although at the given moment the algorithm was not wrong, these predictions can be overcome with reasonable effort.

There are, however, reasons for unequal outcomes in algorithms for which social inequality cannot be blamed and that are caused, rather, by biases in the process. These are cases in which the algorithm “got it wrong”, and therefore correcting the algorithm would not compromise the predictive accuracy of the algorithm, but rather would enhance it. The challenge, of course, is finding a way to discern the two.

Some of the technological solutions have been already referred to above. Scientists have examined the effect of removing certain attributes from the datasets. However, for reasons described above, this is unlikely to be helpful, and more active interventions are probably imperative. One central method involves manipulating training data sets so that more equal results are achieved. A training data set can be manipulated by choosing those cases that are

---

158 As Barocas & Selbst (supra note 3) put it.
borderline and changing their classification. Thus, members of minority groups that were classified as not entering a high track class but only slightly so, are reclassified as suitable for the high track, and members of majority group that the algorithm placed within the high track but only just, are reclassified and assigned to the lower track. Clearly, it can be argued that this is a form of affirmative action, and indeed, in the next part we examine whether these methods would be upheld according to current doctrine. However, it is also possible to argue that this action merely corrects the inaccuracy of the algorithm. It would have to be shown that the scope of correction is small enough to be a correction of bias rather than full-scale affirmative action.

Additionally, as stated previously, it is possible to think of completely novel ways to group students in ways that induce learning, such as clustering them according to learning styles and preferences instead of the traditional ability. While this method would need to be empirically tested for pedagogical effectiveness, it offers a novel approach to grouping that is likely to have a positive effect in terms of social integration.

Adopting any of these methods, and others, requires addressing the normative issues embedded in them. Questions such as what actions constitute discrimination, which classes are protected classes, and whether differential treatment aimed at promoting equal outcomes are permissible, must be determined in order to inform technological solutions.

IV. **LEGAL REGULATION OF DATA DRIVEN ABILITY GROUPING**

After understanding the promises and pitfalls of big data for ability grouping, it is time to examine how law can promote equality in data driven ability grouping. It is helpful to distinguish between two roles that law can play in affecting education policy – challenging education policy and guiding it. The first consists in the legal tools with which individuals or groups whose rights have allegedly been infringed upon can challenge the educational practice. Thus, if any given education policy has an adverse effect on a racial minority, for example, they may turn to the law to challenge that policy. However, as will be demonstrated next, the existing legal tools demonstrates that these tools are likely to be of limited effectiveness in the case of data driven ability grouping. The second possible role law can play is to regulate, in advance, the design and implementation of data driven ability grouping. Legal regulation of this sort, we argue, is a more
promising direction for effectively regulating DDAG and removing biases that have historically afflicted ability grouping.

A. Challenging Data Driven Ability Grouping

i. The Inadequacy of Judicial Review

As was described above, judicial review of ability grouping policies and decisions has been largely ineffective in safeguarding equality of opportunity for disadvantaged groups. As we draw away from the painful history of de-jure segregation, the possibility of courts applying a stricter standard of review decreases. Ability grouping performed by school districts are quite unlikely to be struck down, except in the most exceptional cases. Despite the substantial disagreement between experts concerning the efficacy of ability grouping, some of which was described above, courts have been willing to adhere to policy makers on this point, and accept that ability grouping is a legitimate pedagogical measure. This decision causes challenges based on disparate impact doctrine and rational basis test to fail. The shift to data driven ability grouping, which is viewed as more accurate and less biased than traditional ability grouping will make courts even less inclined to intervene. Given that courts have repeatedly approved of ability grouping, and that DDAG improves the process in terms of equality, it is also undesirable to strike it down.

What follows from this, however, is that despite the challenges described above, individuals have little possibility to challenge big data driven ability grouping. How then, can law provide useful and effective tools to prevent bias in DDAG? We argue that a more viable possibility is by using law to regulate the design and use of algorithms in advance. Before doing so, however, a comment is required concerning the rational basis test and data mining.

ii. The Rational Basis Test: data mining is always rational

Naturally, not any method of ability grouping would be deemed constitutional, and the rational basis test, although extremely lenient, discerns between those that are legitimate and the few that are not. Notice however, that EDM creates a unique problem regarding the use of the rational basis test. At first glance, EDM seems to be extremely successful in terms of the rational basis test. It is a better predictor of educational success, and it overcomes some of the biases associated with traditional educational decision-making. There is, however, something special algorithmic
decision making that raises doubt as to the appropriateness of the rational basis test to it, as a matter of principle.

The point of data mining, according to Barocas and Selbst, is to provide “a rational basis upon which to distinguish between individuals and to reliably confer to the individual the qualities possessed by those who seem statistically similar.” Data mining attempts to locate statistical relationships and patterns between attributes (or between an integration of several of attributes) in a dataset. These statistical correlations are always rational in the sense that they are statistically valid. If so, then any finding of an algorithm is rational, and passes the legal test. Algorithms are able to find surprising and unexpected relations and patterns in large databases. It is these connections, between multiple different inputs and attributes that give data mining its predictive power. This also means, however, that the predictions incorporate connections between too many attributes and are too complicated to be understood by the human mind. The algorithm’s ‘reasons’ for finding one child suitable for a certain track, and the other unsuitable, is a combination of tens, maybe hundreds of different attributes put together, in certain doses. We are unable to fully grasp these intricate relations making the algorithm’s prediction wholly inexplicable.

In terms of the rational basis test, this raises a concern, because this test requires being able to demonstrate a reasonable relation between the measure and the outcome. However, when asked why a certain student was placed in a certain track, and others in another, we are unable to explain the algorithm’s finding. The only answer we can possibly give is “because the algorithm said so”. And although the algorithm is indeed relatively successful at predicting, definitely better than teachers have been for decades, its inexplicable, “black box” nature raises doubt as to whether it can be argued that decisions generated from it can satisfy the rational basis test. For a mechanism to be rational, it must offer some substantive explanation for its decisions. Another problem is that in each and every prediction offered by the algorithm, the explanation would always be “because the algorithm said so”, and therefore these predictions supposedly satisfy the rational basis test always. Absent a possibility to sometimes fail the test, the rational basis test seems to have no meaning at all – when everything is rational, nothing is rational.

159 Barocas & Selbst, supra note 3, at 677.
The argument does not suggest that any use of DDAG must always fail the rational basis test, and that the use of technology for DDAG should be struck down in all cases. It aims to show, rather, that the traditional tests are too weak to locate those cases in which intervention is required, and that other means are needed in order to ensure that DDAG is put for good use, and does not aggravate inequality. These means, we argue, are means to regulate the design and use of DDAG.

**B. Legal Regulation of the Design and Implementation of Big Data Driven Ability Grouping**

We mentioned above that technological solutions must be developed within a certain legal framework. We now move on to offer some general guidelines for such solutions. One of the first attempts information scientists perform in order to prevent discrimination is to remove suspicious classifications from the dataset. We argue that this strategy is undesirable. Although removing these classifications can safeguard against the danger of intentional discrimination, this risk is relatively small, and the gains from having ongoing access to information about race in education can be extremely instrumental in recognizing inequality, learning its reasons and promoting equality. Additionally, as stated above, removing the suspicious attributes hinders the ability to design the algorithm so that it will be able to correct some of the biases that are embedded in the process of choosing. As a result, removing this data is counter-productive. Finally, removing these attributes seems quite senseless as this information can be recreated at will, using other correlating attributes or a combination of characteristics.

We argue, therefore that algorithms should be able to use classifications such as race and ethnicity in order to promote educational equality in general and in grouping decisions specifically. Algorithms can be programmed to produce equal outcomes. By instructing an algorithm to give more weight to certain factors in certain cases, ability groups can be completely reflective of the population in terms of race, gender or class. Assuming that at least some of the inequality is caused by social inequality this would entail that students from racial minorities would have to be evaluated differently than non-minority students. Additionally, as was also explained before, training datasets can be manipulated to ensure a more equal outcome.
These actions, however, are problematic both from a technological perspective and from a legal perspective, and should, therefore, be performed with caution. From a technological perspective, treating members of racial minorities (or other disadvantaged background) differently would decrease the algorithm’s predictive accuracy. If a relatively small decrease in accuracy results in a significantly more equal outcome, this might be desirable, whereas if accuracy was substantially curtailed, this would have counter-effective outcomes – assigning students to tracks that do not provide for their educational needs and are not suited to their abilities. It is also reasonable to assume, however, that at least to some extent, differential treatment of disadvantaged students actually constitutes a correction of biases that exist against them, rather than unjustified preferential treatment and therefore would not impede upon the predictive accuracy, but rather improve it. Discerning the inequality that results from biases within the process as opposed to social inequality is not an easy task, and therefore it is also difficult to know how much of the correction will improve the accuracy of the algorithms predictions, and when further correction would entail sacrificing accuracy.

The legal challenge concerned with such practices is, of course, that differential treatment of certain groups and communities constitutes affirmative action, which is, in the current legal atmosphere, a non-starter. Courts have struck down policies that treat members of different racial groups differently when this differential treatment was designed in order to facilitate integration and promote equal opportunity.\(^\text{161}\) In order to withstand strict scrutiny, educational policy that gives preferential treatment to racial minorities must promote a compelling state interest, and be sufficiently narrowly tailored. In the seminal case of *Parents Involved*, the Supreme Court struck down assignment policies in two school districts that considered students’ race in assigning students to schools, when this was performed in order to promote racial diversity. Striking down the policy, the court stated that it was not sufficiently narrowly tailored,\(^\text{162}\) and that while race may be considered, it could only constitute one consideration among many and that students must be evaluated holistically rather than merely according to their race. Following *Parents Involved*, the US Department of Education Office for Civil Rights and the US Department of Justice Civil Rights Division issued joint Diversity Guidelines for

---


\(^{162}\) Four of the five majority justices went further to state that racial diversity was not a compelling state interest. The fifth majority Justice, however, joined the dissent in ascertaining that integration is a compelling state interest. This case has been subject to wide scholarly critique.
school districts, in which they detail the measures that school districts may adopt in order to promote diversity in a constitutional manner. The guidelines advise school districts to first examine race neutral measures, and then use generalized race-based approaches that do not refer to any specific student. Individualized racial examination should be used as a last resort, and be narrowly tailored to the specific goals of the district. In these cases race may be considered alongside other considerations in assessing a student’s assignment.\(^\text{163}\) Although these guidelines refer to school assignment, its rationale may apply to the case of ability grouping too, and suggest that as long as race is merely one consideration, it may be legitimately taken into account in order to realize the compelling state interest of racial integration.

If algorithms are programmed to treat racial minorities differently in order to promote reflective assignment of students to different tracks, this can reasonably be viewed as fulfilling the requirements put forward in *Parents Involved*. Algorithmic decision making is based upon an integration of multiple attributes, and each student is examined holistically. Race is merely one factor among many others that are taken into consideration in the decision. As a result, such policy could be defended in court and might well withstand judicial review.

Additionally, while the focus on race is understandable, we should keep in mind that racial disparities are not the only inequalities that ability grouping recreates. Children from lower socioeconomic classes are also overrepresented in lower tracks, as are immigrants, and in science and math courses, gender inequality is also often observed. These classifications are treated differently by courts, requiring only intermediate scrutiny (in the case of gender and nationality) or the lenient rational basis test in the case of class.\(^\text{164}\) Noting this difference between categories of race and class, several writers suggest promoting equality and diversity by using socioeconomic class instead of race.\(^\text{165}\) Empirical studies stress the educational disadvantage that


\(^{164}\) *San Antonio Independent School District*, 411 U.S. 1 (deciding that class was not a suspicious classification that triggers strict or intermediate scrutiny).

comes from poverty, meaning that this kind of differential treatment can have enormous effect in increasing access to high track courses for children from disadvantaged backgrounds. Because of the correlation between poverty and race, this would have the side-effect of increasing access to higher tracks also to students of racial minorities and immigrants.

The conclusion is that while direct challenges to programs of DDAG are likely to be unsuccessful in courts, regulation requiring algorithms to take into consideration race in a way that corrects biases in decision making and in training datasets can be justified and defended in court. This strategy also seems effective in preventing social segregation and the perpetuation of disadvantage. If DDAG proved especially effective, and became easily available, it might even become a required standard, so that if a district failed to adopt it, and their ability grouping created severe inequality, it would be possible to challenge the policy on the grounds that they failed to perform their duty to pursue the policy that least infringes on individual’s rights.

CONCLUSION

Brown marked the beginning of the end of de jure apartheid in America, but segregation in education underwent modification rather than ended. Attending the same school is hardly a remedy for school segregation if blacks and whites are separated upon entering the schoolhouse doors. Regardless of the alleged neutrality of the policy, where the assignment of students creates separate and racially identifiable classrooms, minorities are disadvantaged as a result of tracking, receiving fewer educational benefits and resources, and having inferior opportunities in life.

Technological developments, and specifically data-driven decision making, have the potential to improve the processes of ability grouping, and to begin to deliver the long promised educational justice to all children. Whether DDAG will indeed succeed in doing so depends on multiple factors, of which legal regulation is only one. Educators and regulators alike must watch the implementation of DDAG closely, and to adjust its design and use according to the outcomes. If, after all, DDAG is unable to promote equality of opportunity and decrease segregation, there may be no choice but to revisit the struggles to eliminate it altogether.