ADAM GAMORAN: Thanks very much, Ron, Judith. It's a pleasure to be here today. Denise McKeon(?) told me in, I think, 1993 or something like that when she moved over from the AERA, what year was it?

DENISE MCKEON: '99.

ADAM GAMORAN: Okay, 1999, when she moved over from AERA to NEA, that I would have a chance to talk here some day. So, you know, it only. . .

DENISE MCKEON: We made good on that promise.

ADAM GAMORAN: . . . took 11 years. That's all right. But I'm really pleased to have the opportunity to speak here. Over the breakfast meeting before this morning's talk, one of the topics we discussed was the responsibilities of academic researchers to speak with broader audiences about their research. And this is something that is somewhat of a luxury. The incentive structure is not well aligned with creating opportunities for academic research to speak to broader audiences, but it's something that I have had the good fortune to engage in.

And one of the reasons that it's been very important to me is that the topic that I address, research on ability grouping and tracking, is of such close interest to people involved in the practice of education, to teachers and to those who make decisions about how to arrange students in classrooms. And I'm always getting feedback from audiences that helps me think about my research. And so I hope that will occur today as well.

So the title I gave to my talk is Grouping Students to Maximize Learning and Minimize Inequality. Of course, that's our goal, maximizing learning and minimizing inequality, but it's a very challenging goal to meet. And there are some promising sides, but it's going to be very difficult, and so, hence, the subtitle, New Hope or False Promise. By the end of the day, you'll probably still be scratching your head about that, but that, I think, is the challenge that we face.

So why do students assign, why do schools assign students to classes or to groups within classes by their interests or abilities? Well, it seems like a logical and efficient way of organizing students for instruction. After all, students differ in their performance levels, and no one knows this better than teachers. So it seems to make sense to divide students so that we can match instruction more closely to their needs.

A narrower range of student performance makes it easier to organize the curriculum because we can match the curriculum to where students are when they come into the classroom. So why is this problematic? Well, it's problematic for a couple of reasons.

First, due to circumstances outside of the school, separating students by academic performance also tends to separate them by race and social class. We live in a stratified society. And as a result, students from different social backgrounds come to

1

school at different levels of performance. And this is evident when children enter school in kindergarten, and it is manifest all the way through.

And so when we divide students by their academic performance, we also divide them on other grounds. And since one important goal of school is integrating students from all walks of life, then the division of students according to their academic performance level conflicts with that integrative goal.

In addition, homogeneous classes lack the diversity that may foster rich discussions. I know as a college teacher, as the University of Wisconsin has become more and more selective, and, therefore, the classes I teach have become more and more homogenous in terms of student background that we've lost something from our class discussion. And I think the same is true in K-12 education.

Although ability grouping is intended to provide equally effective instruction to all students, that rarely occurs, and there are a number of reasons. And this is an empirical finding. This is not a normative statement but an empirical finding derived from many decades of research.

First of all, when students are divided, teachers are also tracked, in many cases. There are many schools in which teachers with the strongest reputations are the ones assigned to teach the honors tracks or the advanced courses. And this, there are a number of reasons for this. Teaching an honors class is a reward for teachers. Administrators have very few rewards that they can provide. This is one of them.

Also, the parents of high-achieving students are often the loudest voices in speaking up about the education of their children. And in order to maintain a calm relationship with parents, principals will often assign their children to the teachers that they request. So we end up with a system in which not only students, but teachers, are tracked.

Low-track classes have a cycle of low expectations where students have performed poorly, so teachers have low expectations for them, and the children pick up on those low expectations, perform poorly and, thus, reinforcing the low expectations that teachers hold for them.

One researcher described low-level classes as caricatures of regular classes. At a superficial level, they look the same. They have textbooks. They have children sitting in rows and in desks. But if you probe more deeply, you see that the low-level classes are about fragmented instruction and seat work managing to keep order in the classroom but not a lot of in-depth learning. And, indeed, research on instruction in high- and low-level classes has found a greater emphasis in procedure in low-level classes and a greater emphasis on discussion in high-level classes.

As an illustration of this pattern, and I could point to many research studies that have found similar findings. Here are some findings from a study that I was involved with, which appeared in the *American Educational Research Journal* in 2003. And this was a focus on mastery of literature in middle and high school English classes.

And these coefficients are standardized, so they're in, no, these coefficients refer to, well, it depends on which coefficient. So the first row, discussion time, is in minutes per lesson. And we have a very narrow definition of discussion in this study. Discussion was free exchange of comments among students, often without interruption by the teacher and without prompting by the teachers.

So instead of the usual cycle of question, response, evaluation to the response, this is where information is being exchanged freely, and you see that more discussion in high-track than in low-track classes. Envisionment is an activity that teachers and students undertake in which they think prior to writing. They imagine the way the writing will unfold, and it is a conceptual activity that helps children prepare for writing. And this is the standardized coefficient. And so we see, again, much more in high-track classes than in low-track classes.

Revising writing is something that helps young people become better writers. We saw that there was more evidence of that in the high- than in the low-track classes, and time spent on homework more than doubled in the high-track classes compared to the low-track classes.

Most studies in the U.S. that focus on effects of tracking look at differences between high- and low-track classes. This was one where we also had some mixed-ability classes. And it's a little hard to be sure about the representative of the mixed-ability classes, about the schools in which mixed-ability classes are used, even in middle school English, where it's somewhat common. But, you know, it tends to be smaller schools where there are mixed-ability classes.

Nonetheless, you can see that the pattern is that the instruction in the mixed-ability classes tends to be more similar to the middle level of the tracked classes. And, overall, tracking is a system that provides a venue for unequal opportunity, unequal instruction opportunities. Yeah, question?

WOMAN: So this is a very interesting in the mixed classes to see. Because in the mixed classes, all the discussion and the work can be done . . . kids, but and they will be driving the average up. But it will be interesting to see whether the kids that are in the mixed group that are low-achieving are going to actually spend more time discussing in the mixed classes versus in the low classes.

ADAM GAMORAN: Yes, that's a really, yeah . . .

WOMAN: I'm just finding that . . .

ADAM GAMORAN: That's a really good question, and I have a former student who's now a faculty member at Notre Dame University named Shawn Kelly, who's actually pursuing this question. For example, who are the students who participate, and which students actually get the benefits? Is it only those who participate, or does the benefit accrue to everyone in the class?

The general finding is that low-achieving students do better in mixed-ability classes than in low-track classes, and high-achieving students do worse in mixed-ability classes than in high-track classes. That's the general pattern. Of course, it's not true in every case, but that's the general pattern. I think that might be my next slide. It is.

So this is the achievement gap between high and low tracks. So just the observed difference, before taking into account of any other conditions, we had about 1.1 standard deviations difference on our literature achievement test between the high and low tracks. That's the first bar.

Then when we took into account differences in students' writing ability at the beginning of the school year, as well as their social and economic background, that difference was not quite so great. So part of the initial difference was due to, part of the observed difference, rather, was due to initial differences among the students. When we take that into account, though, there's still about .8 of a standard deviation difference.

When we take into account instruction, then that difference shrinks further to about half of the standard deviation. So that suggests that a substantial portion of the achievement difference between students in high and low tracks was a reflection of the differences in instruction that occurred in the high and low tracks. Not all of it. There are other things going, but that's a big part of it.

And you can see that the right-hand bar is only about half the size of the left-hand bar. So we can account for much of the observed track difference in achievement when we understand initial differences among students and the differences in the instruction that students experience in their different classes. So partly as a result of unequal classroom conditions, inequality between students assigned to high and low tracks tends to widen over time. And, therefore, the consequences of ability grouping are, first of all, no effect on achievement productivity.

And by that I mean the production of achievement overall, on average, for the core of students in the school is neither higher nor lower when tracking is used or mixed-ability teaching is used. But there's an increase in achievement inequality. And how do these two statements fit together? They fit together because high achievers benefit, low achievers suffer, and the net is a big zero.

Supporters and critics of tracking often fail to engage in productive dialogue because they focus on different parts of the equation. Supporters focus on productivity, emphasizing that high achievers do best when they're in high-track classes, ignoring the fact that tracking magnifies inequality. Whereas, critics emphasize the link between tracking and inequality correctly, but don't have a response for the statement of advocates who focus on the benefits for high achievers.

So it's a dilemma. And as I say in the research brief, there's no simple solution to this dilemma. If there were, we'd have been doing it a long time ago because tracking is one of the most heavily researched areas in all of education research. But there's, oh, by the way, new international research finds the same pattern as in the U.S. that tracking is linked to increasing inequality.

There are a few exceptions. There's a study from Israel that I was involved with, another study from Taiwan that shows that when there are incentives for performance for students in low levels of academic classes that those students respond positively. So in the case of Israel, there's a national examination that young people take which determines their opportunities for post-secondary study. And in the Israeli education system, those classes are structured so that the higher-achieving students studied for more points towards the college admissions program, and the lower-achieving academic students studied for fewer points.

So those are like ability group classes within the same subject, and it turns out that the more differentiated the instructional opportunities, the better low-achieving students did, not the worse. So that's an example of how incentives can provide a framework in which low-level classes are meaningful.

You know, in the U.S., there is no incentive for low-achieving students to do anything better than get a C in our high school classes. Yeah, to get into a community college or even a nonselective, four-year college, you need a high school diploma. That's it. So if you're not shooting for the stars in terms of your post-secondary opportunities, there's not an incentive to do anything more than graduate. And that's part of the reason that we see inequality associated with tracking.

But if you provide a meaningful incentive, such as a performance standard that even lower-level students need to maintain in order to have post-secondary opportunities, or, in some cases, labor market opportunities, then that may be an inducement to low-achieving students to thrive. But, overall, the general pattern, internationally, as well as in the U.S., is that more differentiation of instruction is associated with more inequality.

My favorite example of this is a study that I've cited. I've given you the citation in the research brief. It's a study from Great Britain from the mid-1980s, which had a fantastic dataset. It includes every child born in England, Scotland, and Wales in the first week of March, 1958. So talk about representative. This is a representative sample.

And the British system at that time, so now we're talking the early '70s, the British system at that time had both separate schools for students aiming towards different destinations and comprehensive students. They had undergone a comprehensive reorganization, so the schools were not entirely, not all schools were selective, but there were also still the grammar schools and the secondary, modern schools, which were the selective and nonselective schools.

And the research shows the same pattern as in the U.S. The more selective schools promoted higher achievement for high achievers. The secondary, modern schools pulled down students' performance compared to the comprehensive schools. Similarly, within the comprehensive schools, there was grouping or what they call streaming in setting.

So students assigned to high-level classes did better. Students assigned to low-level classes did worse compared to other comprehensive schools that had mixed-ability teaching. So that is a particularly striking and comprehensive analysis of this challenge. And it puts together many of the pieces that we see across a lot of the other studies. There's' another question from the back?

WOMAN: There was. So is this . . . literature . . . and it seems that this is confirming that the . . . effects, and the . . . effects . . . instructions are extremely important, but if you ignore them, then I would assume that no matter which class you are, a good instruction will get you to good achievement.

ADAM GAMORAN: Yeah, it's very difficult to disentangle teacher effects from peer effects when you're only looking at one teacher with one class at a time. So what I'm calling teacher effects, probably, are a mixture of teacher and peer effects. And this research literature, at least, doesn't try to disentangle that.

So what are we going to do? Well, there are, basically, two solutions. On the one hand, we can reduce the use of ability grouping, but in doing so we need to find new mechanisms of providing challenging instruction to high-achieving students. Or we

can maintain the ability grouping, but if we do, we need to do a much better job of providing effective instruction to low achievers, to students in low tracks.

Now there is some new research, and that, I think, is the main focus. And the main reason I'm here today is to bring your attention to some new research that provides promising new directions for both of these two responses, either get rid of tracking but challenge high achievers, or maintain tracking, but do a better job on the lower tracks.

And so I want to report to you today on some conditions that support successful mixed-ability teaching and also some conditions that support effective instruction in low groups. So the first case I want to bring to your attention is some research by a woman named Carol Burris, who is a Rutgers Professor of Education, and she's done a case study of detracking in a New York school district.

Now you may know there's been a lot of research on detracking. And most of the research on detracking shows that the schools are unsuccessful at detracking, that tracking is maintained. And Jeannie Oakes and Amy Stuart Wells have provided some very nice insights about why that is. There are three main barriers to detracking, political, normative, and technical.

The political barrier is that people have vested interests in tracking, and whether teachers or parents of high achieving, teachers who like to teach the honors classes or parents of high-achieving children who want their children to have separate opportunities, politics intervenes.

Normative barriers, there is a belief that people are different. And since we believe they're different, we believe we should serve them differently, and that's sometimes hard to overcome. But perhaps the most important barrier, in my judgment it's the most important barrier, is the technical barrier. Teachers are not well prepared to instruct students with widely varying levels of performance all in the same classroom.

So this is a case study of where the school found a way around the technical barrier in, and I think this is the kind of evidence that is necessary if detracking is to become more widespread. So what happened in this school district that Burris studies was that tracking was replaced with mixed-ability teaching in middle and high school mathematics.

Now the focus on mathematics is particularly interesting because mathematics is a subject area that most teachers believe has a rigid sequence, and unless you proceed along the lines of those sequence, you can't master the content. There are cognitive science who challenge that assumption, but that's the prevailing assumption in schools, at least those that I've encountered and that others have written about. So math makes a particularly good example to challenge assumptions about detracking.

What she found in this case was improved outcomes for low achievers without losses by high achievers. And, again, that flies in the face of the typical evidence, which suggests that high achievers do best when they are assigned to, segregated to high-level classes. The middle school reform involved an accelerated curriculum for all students and an extra support workshop for struggling students. It also involved common planning time for teachers and increased use of calculators by students.

The high school reform meant that all students were assigned to regents classes. As you may know, in the state of New York there's an examination at the end to see if you're a college, if you can satisfy college entrance requirements, the regents exams.

But in the past and, typically, across the state, in the past, only the high-achieving students went into the regents classes. And if you were low achieving, you didn't get into the regents classes, you couldn't go to one of the, or it's much more challenging to go to one of the four-year New York State colleges.

Now this reform also involved a supplementary class for students who struggled with the more advanced materials. This class met three times each week. The research design was what we call an interrupted time series. And by that I mean that there, if you can imagine a school system chugging along, there's some sort of interruption. And what we're looking to see is whether the achievement trajectory, after the interruption, is different than the achievement trajectory before the interruption.

Now one problem with this kind of design is that this interruption may be the reform we're looking at, or it may be something else that was going on in the same time. So the way to deal with that is to look at some other nearby and comparable school systems that may have experienced the same overall environment, the same overall education milieu but without this particular reform. And so that that's what Burris and her colleagues have done. They've looked at, they've compared success of cohorts of students in the same school and to other schools that did not undergo the reform.

Now this is not a failsafe methodology because there could still be other changes within this school district where the reform was that are accountable for any changes in the achievement trajectory, if those same changes did not occur in the other district as well.

So as I say, this is not a failsafe methodology, but it's a reasonable quasi-experimental approach to testing the effects of a reform. And what they found, as I mentioned, was that high achievers did similarly well, and low achievers did much better. This is one of their findings. It's from the high school paper. And it looks at questions about whether students took the regents exam or not and what percent passed the regents exam or not.

So there are two cohorts that examine. One is before the reform, and one is after the reform. And the two pairs of bars on the left are the comparison schools, and the two pairs of bars on the right are the study schools. And you can see there are two things to pick up from this.

First, the main thing, the thing that the authors want you to see is that in the comparison schools, there were similar levels of taking the test and passing the test before and after this time of reform. By contrast, in the study schools, there were a significantly higher level of students both taking the test and passing the test after the reform compared to before the reform. So that's the main finding. That's the takeaway point.

It's also the case, however, that there was more taking and passing of the exams in the study schools before the reform compared to the comparison schools. So one raises the question how good a comparison are the comparison schools? And that is just going to have to be an open question. But, you know, I think it's a reasonable, if not ironclad conclusion that the reform is responsible for this increase.

And by the end of the period, the study schools actually had 105% of the students sitting for the regents exams. How could that be? Well, there were some students who had failed the exams who had not sat for the exams the previous year who then did it with a younger cohort. So, and I think that's a very positive result. And

then they had 96% of the students passing the exam, so that's' quite an extraordinary rate. So something going on in this study and, I think, bears close attention.

What were the conditions that supported this successful case of mixed-ability teaching? Well, cutting across both the high school and the middle school findings is a substantial supplementary instruction for low-performing students. In the high schools, the low-achieving students had 50% more instructional time. Notice, by the way, that this is a form of grouping. The low-achieving students had a low-achieving group, which they went to and received their supplementary instruction.

So I think this shows us that even in a case of mixed-ability teaching, some differentiation is necessary. And detracking efforts that simply try to wave a magic wand and pretend all students are the same are unlikely to be successful. This finding about supplementary instruction resonates with research that I and some of my colleagues did earlier in a study of highly restructured schools.

We looked for cases that had eliminated tracking and ability grouping. And over the eight high schools, there were three that had done so. And then within these three, we looked for cases where achievement among high achievers was no lower than elsewhere, and achievement among low achievers was brought up.

And we found one such case, same condition, supplementary instruction for low achievers. That school, which was, oh, I think this is in the next slide. Hold on. I'll get to that in a second. I'll tell you about that school. Anyway, so supplementary instruction for low achievers and some manner of differentiation.

Now this was an affluent district with few high-needs students, so it's not clear how far these results will generalize. But there were similar results from a study that I was involved with, which I was just describing, which was in an urban environment where just over half the students were on free and reduced-price lunch. However, in that case, again, additional resources allowed a Saturday tutoring program.

And I have a great quote from a teacher in my article about this where the teacher says, if you can't add fractions, we don't slow down for you. We say you have to come in Saturday to learn how to add fractions. So they maintained a rigorous level of instruction in the main class, but then there was supplementary instruction for low achievers. Also, admission required an interview for students so they could ensure a level of commitment that would allow low-achieving students to keep pace with their high-achieving counterparts.

So successful instruction, to the best of our knowledge today, successful, mixed-ability instruction, particularly with a broad range of students, requires some form of differentiation, a way of reaching students where they are. And that means supplementary instruction so that the main class can be taught at an accelerated level without slowing down and, yet, more struggling students who are struggling more have opportunities to keep pace.

Again, how far this will generalize to all environments, I think, is an important question. Clearly, it requires resources. It's not, you know, this supplementary instruction is not free. It means extra investment. Yeah?

MAN: I have to ask . . . what happens to the students who don't pass . . .

ADAM GAMORAN: They go to their neighborhood school. This is a choice school. Yeah?

WOMAN: . . . Burris study . . . how many . . . schools . . .

ADAM GAMORAN: I believe two study schools and six comparison schools. That's my recollection. The citation is in the research brief, and you can check it. There may, I can't remember if there were. Maybe there were more middle schools in the study. I can't, I don't remember.

Okay. So evidence is accumulating that successful mixed-ability teaching is possible, that an enabling condition is extra resources to support low-achieving students. Now what about the alternative, maintaining grouping, but doing so in a way that is successful with low-achieving students? Whoops.

Well, there is some new research on grouping systems that close gaps instead of magnifying gaps. And here I'm focusing on the research of Carol Connor, a faculty member at Florida State University in their reading research center. And she and her colleagues have conducted a series of studies on grouping students for early reading instruction.

Now the thing about early reading instruction, there's no other area where we have more fine-grained research on effective practices for helping students achieve. So that's one great advantage that this line of research has is it can build on very well-specified information about effective instructional practices.

And my term for this approach, for the broader approach into which this is imbedded, is diagnosis and instructional response. Because that, in my judgment, that's the core of it, figuring out where students are and then providing an instructional response to students' performance.

So the steps involved are, first, an assessment of students' reading performance. Second, the assessment results are input into a computer algorithm called assessment to instruction, which was developed by this research team. And assessment to instruction, or A2i diagnoses students' performance, and then it recommends an instructional response. And then the software takes one further step, which is it recommends within class groupings to facilitate instructional responses.

Now perhaps an optimal solution would be instead of recommending groupings would be to recommend an instructional regime for each student individually. But that's not practical. We don't have time to spend one-on-one interaction with each student. That's why schools are organized into classrooms, and that's why instruction in classrooms is either done as a whole class or organized into groups. It's a resource limitation.

But what A2i does is it organizes, it identifies an instructional arrangement that optimizes the matching of students' instructional needs with their classmates so that they can be taught in small groups with instruction effectively targeted. And one key feature of the A2i software is that it is programmed to identify a steeper trajectory for low achievers so that low achievers have a chance of catching up to their high-achieving peers.

The research design for this study is a randomized evaluation . Now I hesitate to say, there were many years of trial and error, of naturalistic observation, of what we

might call design experiments, where the researchers try something out, see how it's going, and then improve it before getting to the randomized evaluation. So, you know, it took years of work to design the system.

But once that was done, it's been evaluated in a randomized evaluation where teachers were randomly assigned to either a group that would receive the A2i software and receive training on how to use the software, and then their use of the software was monitored, and their instruction was monitored, and their children's achievement was monitored, in comparison to a comparison group of teachers who did not get the A2i access, but their instruction and student performance was also monitored.

So the results showed that students whose teachers were assigned to the A2i group outperformed those in control conditions. Not only that, low-achieving students received the largest benefits, and the benefits were greatest for students who made the most use of A2i.

And so, well, I guess, all right. So this is a slide from an article that appeared in the *Journal of Science*, which I strongly recommend to you. As you may know, science is a general audience science journal published by the American Association for Advancement of Science, and so it's a very accessible article.

And this particular slide, or this graph, which I cut and pasted from the article, shows the steeper trajectory of students with initially low-achieving scores compared to those with initially higher-achieving scores. And this is not an accident, and this did not happen on their own. This happened because they programmed the software to identify an instructional pattern that would elevate the performance standards for low-achieving students over time. And as you can see, it was successful.

So this is showing you across the bottom axis, the X axis, this is how much time teachers used the A2i software, and then it's, then the Y axis is the spring reading performance. And the blue versus the red lines are differences between students who started with high vocabularies versus students who started, high vocabulary scores versus students who started with low vocabulary stores. So you can see the impact of the program is greatest for students who started at the lowest levels.

Now so the experimental result is the first and second bullets. The first bullet is that students in the A2i group performed better than those in the other group at the end and that the biggest increases were among the lowest achievers. This, the last bullet, is quasi-experimental. That is to say, it's not random which teachers use the program more and which teachers use it less.

Teachers who are more motivated, more interested, struggling more, many reasons why one teacher might use the program. You know, and that's the same goes for any instructional resource, right, or any resource at any job, not just teaching. Some people use it more. Other people don't.

So you can't say for sure that it was the use of the program based on this kind of finding because maybe it was the more motivated teachers. Maybe teachers who are just better instructors used this program, and the student performance reflects their better instruction.

But when you put this finding in combination with the experimental findings, we feel pretty confident about the interpretation that the, not only the availability of the A2i software, but actually the use of the A2i software was associated with student

performance. And, again, what was the structure of the classroom? What was the arrangement? Grouping, right? So students were differentiated.

Once again, we see that some form of differentiation of instruction is necessary to maximize students' outcomes. The key here is very careful assessment and a lot of knowledge about what instruction responds to students with particular assessment levels.

These findings, the results of this new research echo the conclusions of Robert Slavin, a very well-known writer whose early work concerned grouping. You may know him as the leader of the Success for All program. Success for All also involves a lot of grouping, and so his practice also is based on findings about this.

But his conclusions long ago were that ability grouping can be effective if three conditions are met. If students are assigned to groups based on the specific skill that is to be taught, and that might seem obvious. But through the history of grouping, there's many cases of where one test is given, and students are assigned, for the whole school day, based on their performance on one test. So in the A2i, we can see an example of where grouping is based on a very fine-grained assessment of student skills that are very specifically linked to the skill that is the instructional objective.

Second, instruction is targeted to the specific skill, and, third, that grouping arrangements are flexible, that we don't assume that a student who does well with one kind of skill is at the top of the distribution on another type of skill. So we have frequent monitoring of student progress, and we can reshape the groups as necessary in response to students' different needs in different areas of performance.

So this seems like a promising new direction. It's very demanding of knowledge about the link between student performance and instruction. It is very demanding of high-quality assessments. If you don't have high, if you have crude assessments, then you can't do this very effectively. Early reading is one area where we have fine-grade information about student performance, we can assess student performance well, and we can match instruction to it.

Finally, one other approach to maximizing achievement through the differentiation of students, through grouping students, but doing so in a way that allows us to elevate student performance at all levels instead of letting low achievers fall further and further behind, and this I use a label for this. My label for this is the optimal matching of teachers and students.

Don't have to speak to this audience about the annual testing of students and whether or not it can provide evidence of teachers' contribution to students' achievement. But this optimal matching idea rests on the fact that there's so much testing going on, and that in many cases we can link students from one year to the next, and we can link students to their teachers.

If we can do that, then we can address the question as to whether some teachers are more effective with one type of student than others. That's not an easy thing to do. Yeah?

WOMAN: I have a question about that because you hear from teachers that principals take certain students in their classroom year after year after year. Are the, is that going along with identifying them as effective with a specific type of student?

ADAM GAMORAN: If it were, that would be a good thing. I suspect that it's not. I suspect that, based on prior literature, it's, the prior literature has identified two mechanisms that go into those decisions. One is rewarding teachers. So if a principal wants to reward a teacher, he or she will give that teacher either the students the teacher wants or students the teacher thinks will be, the teacher, will make it a comfortable year for the teacher.

And then second is, the second thing that is addressed in the research literature is assigning students to teachers to minimize complaints among parents. So if there's a teacher you know are going to get you a lot of complaints, then you put in the kids whose parents are not going to complain. So I'm not speaking from cynicism. I'm speaking from, you know, this is what the research literature says.

What we'd prefer to see, however, is a principal assignment pattern where principals are taking into account each teacher's strengths and weaknesses in trying to assign students to teachers in a way that optimizes the strengths and minimizes the weaknesses. And that is exactly what this approach is about, except it takes a more scientific approach to doing so.

So instead of relying on the principal's instincts and intuition about which students do well with which teachers, it says, well, let's look at the evidence. Let's look at each teacher's past history with certain types of students, and see which students have done the best with this teacher, and then let's increase the assignment. Let's over assign those kinds of students to this teacher and under assign them to the other teacher who does better with some other kinds of students.

So I guess I want to pull back a little bit from what I just said about the research literature and principal's assignment of teachers, students to teachers. The two examples that I gave are found in the literature, but that's not to say that many principals don't do what you said.

So it may be that a lot of principals are trying to optimize the match between teachers and students. That just hasn't been the focus of very much research. So I probably overstated it, and I want to pull back a little bit from that. But if that's true, if principals really are trying to provide good matches between teachers and students, then the approach that I'm about to describe would be a new technology that would help principals do so more effectively.

Another comment, another side comment I would make about this, my own stance about the annual testing of students, which is that although it's an accountability system that has imposed this testing system on schools and teachers and students, we can do a lot better with this than accountability. We can get information from this that can help us teach more effectively, that can help identify strengths and weaknesses among teachers that can identify which students are more successful and in which circumstances, that can identify needs for professional learning among teachers.

There's a lot we can do with this evidence. It's a shame that it's almost exclusively being used for accountability purposes and not for growth and development purposes. This, in my judgment, would be an example of making better use of the testing data that we're gathering all across the country. So, again, yeah?

WOMAN: . . . you just mentioned said that they take a look at a teacher's past history. Does that include . . . preparation program . . .

ADAM GAMORAN: Not in this system. Yeah, this system is only focusing on performance of students who had this teacher for instruction in past years.

WOMAN: Okay.

ADAM GAMORAN: In principle, one could bring other factors into account. There are very few places where one could do that because there are very few places with the data systems that would allow you to do that. The state of Florida is the one place that I'm aware of that can monitor teacher performance and link them back to their teacher preparation programs. Probably, we're going to see more of that in the future.

So as I said, to do this, we would need annual achievement data. We need student linked across years and to their teachers. We need to test for differential effects. And this is the hard, so what we're asking here is do teachers produce the same effects with all students or not? And this is difficult to do.

As you know, when you have only a few students per teacher in a data system, it's hard to get reliable estimates of the teacher's contributions to student performance. It's even, it's much more difficult to obtain reliable estimates if you subdivide the students and say, well, there's some high achievers and some low achievers. There's some boys and some girls, and you try to look for differential effectiveness.

So this is really only feasible if we have multiple years of student performance linked to teachers and better if we have multiple classes in the same year linked to teachers. So, probably, this is only viable for middle school, for grades five through eight, let's say, where teachers have multiple classes, you know, where one math teacher might have five classes.

Now you've got 125 kids in a year, and then let's say you have 3 years of data. Now you have 375 kids per teacher. And that starts to be a large enough sample that you can obtain reliable estimates of teacher effects, not only on the whole but for subgroups of students. So that's one big challenge.

The second is that it's not easy to derive these estimates of differential effects because the same information that you're using to see how well students are doing, their annual test scores, is also the information that you're using to see which students the teacher is more effective with, the high achievers or the low achievers. So that's cutting the information very thin, and it's difficult to do that.

There are two teams of researchers that I know that are working on this, one at Rand and one at my research center, the Wisconsin Center for Education Research, in our value-added research center. Both have demonstrated some success. Neither are using it in a widespread fashion. But I think this is an area that, if we continue the annual testing of students, and we see the increasing linkages to teachers, which is what we're seeing so far, that this may be fruitful in the future.

So what I'm especially interested in is whether some teachers are more effective with high achievers and others with low achievers. I'm especially interested in that for two reasons. First of all, my research is on tracking. This is a tracking question. This is a question about whether, if we assign teachers to students in a way that maximizes the teachers' past success, then we could use tracking in a way that doesn't hurt low achievers.

The second reason I'm interested in this is because I can imagine taking student prior achievement into account when assigning students to classes. We could also do differential effects on gender, free lunch, race ethnicity. But I wouldn't be interested in assigning students to classes by free lunch or race ethnicity because that would counteract other goals that I have for schools, which is, as I said in the beginning, that they should bring students together from different backgrounds.

So that's why, for me, it's differential effects by prior achievement that's the most important and interesting, but it's also the most difficult to do because, again, you're looking for differential effects on the same measures that you're using to see whether the outcome is effective overall or not.

There's another reason why this is difficult to do. With something like gender, boy versus girl, that's a fixed characteristic. So you can look at the whole student trajectory and see whether it differs for boys and girls and whether it differs for certain teachers across boys and girls.

But achievement is not a fixed characteristic. It's a dynamic or changing or time-varying characteristic. So to do this really well, you have to check whether teachers affects change according to students changing achievement trajectories. So it gets very complicated because it's not like you can take a one-time characteristic of a student and see whether the trajectory goes this way or this way. We have to take into account the dynamics of student performance over a year. So it's very complicated. But some brilliant statisticians are coming, and economists, are coming up with ways of addressing this problem.

So if there are differential teacher effects, then we may assign students to teachers who are particularly effective with students who have those qualities. And students would get teachers who, based on their past performance, are expected to bring out the best in those students. Teachers would get students who are like those with whom they've had success in the past.

Unfortunately, there's some challenges to optimal matching. First, we don't know if there are differential effects. It may be that good teachers are effective with everyone, and weak teachers are less effective with everyone, and it doesn't vary. The past research, which doesn't test this question in a very rigorous way, but the past research is mixed.

There's some research suggesting the differential effects do exist and others are questioning that. So we don't know if there are differential effects. Obviously, if teachers do not vary in their effectiveness, then this is not going to provide any solution to the tracking problem.

A second issue is what if there are many teachers who are especially effective with high achievers, but very few who are especially effective with low achievers? This is certainly plausible. If that's the case, how are we going to optimally assign low-achieving students? There wouldn't be enough teachers to serve them optimally, and so this approach is not going to work.

WOMAN: . . .

ADAM GAMORAN: Another problem where limitation of the optimal matching approach is it's not clear that the assessments are good enough to be meaningful. As I

was arguing before, in the case of early reading, we have assessments that can be very closely linked to the skills that students need. But most of the assessment that we're using for our, in our accountability systems are pretty crude measures of student performance, and it's not that straightforward to link them to instructional differences among teachers. So it's not clear that the assessments are good enough.

Of course, we're making huge federal investments right now in new assessment systems. It's possible that the assessments that are being developed now will be better, will be richer, and will be more meaningful for making instructional allocation decisions. And so this may be more feasible in the future than it is at present.

Finally, no study has examined this optimal matching idea in practice. There's no case where principals have, at least no case that I'm aware of, where principals have looked at teachers' past history of performance with students at a particular achievement level and then tried to assign students to that teacher on that basis and then tested whether, over time, that those, that teacher's students did better than similar teachers whose students were not assigned through this optimal matching process. So we don't know if this would work. Even if all these conditions that I've identified are there, we still don't know if this would work, so more research is needed.

So did you know, do you know who said this? Well, if you've read the research brief, you know. But if you don't, you might wonder. Neither tracking nor heterogeneous grouping is necessarily good or bad. The effectiveness of grouping depends on the specific situation and the needs within a school. So this was the conclusion of a National Education Association taskforce in 1990, 20 years ago. And I believe the conclusions remain true today that there are strengths and limitations to grouping and strengths and limitations to mixed-ability grouping. There's no one-size-fits-all solution.

There are some conclusions that are clear. One is to eliminate dead-end courses. These are courses such as general math or business English where they don't lead to anywhere in the student's future, and they have a hard constraint on the level of content to which students are exposed.

We can demonstrate in studies that these kinds of courses do not serve students well because students do better if they're assigned to academic courses, even if their initial performance is low. For example, a study that I did of algebra in ninth and tenth grade showed that even low achievers benefited from being assigned to algebra compared to being assigned to general math. So eliminate dead-end courses. That's a clear conclusion.

Beyond that though, we have, there's no clear solution. We have tradeoffs. And so where ability grouping is maintained, it's essential to implement high standards for low-achieving students. Where ability grouping is eliminated, we need to see that standards for high-achieving students are not lowered. I think the latest research is pointing at some directions and conditions that may support this.

I think we need to avoid saying there's a single solution that's going to work in all contexts, but rather we need to challenge principals and teachers to reflect on the issues in whatever approach they take and to try to be, to confront those issues head on, whatever direction they choose to go. Thank you.