

Research Contradicts  
Heterogeneous Grouping  
Claims and  
Supports Ability Grouping

Presented May 26, 2009

### Why we compiled this research

There is much to like about the Middle School Reform efforts. Movement toward flexible grouping (e.g., being in different math and reading groups) and toward more frequent shifting between levels, both have widespread appeal, as does ongoing staff development efforts, etc.

However, the narrowing of ability groups down to just two or three levels, as part of an overall movement toward greater heterogeneous grouping, is deeply concerning and strongly opposed by many parents (see [petition](#) and [signatures/comments](#)) and, as it turns out, by the research community as well (see notebook “Research Contradicts Heterogeneous Grouping Claims and Supports Ability Grouping”).

The “research” presented to the Middle School Reform Committee (MSRP) and to the School Board paints a picture that moving towards heterogeneous grouping is a “no-lose” proposition: that it helps lower performers with no cost to higher performers. This seemed so at odds with our own experiences on grouping from within the school system that a group of concerned parents set out to review the research on the subject directly.

### Summary of Findings

- Even a cursory review of the literature makes it apparent that the research in the MSRP and the Feb 2009 presentation is a one-sided and slanted representation of the academic research on the topic of grouping. This is an extremely controversial topic in the literature, yet only studies in favor of heterogeneous grouping have been cited, including several studies that have been disproven, and other studies that don’t apply to our demographics. Even content within the studies cited have in some cases been cherry-picked to tell a story, without revealing the dark side of heterogeneous grouping.
- In fact, we consulted a nationally renowned expert on this topic, Tom Loveless, former professor at Harvard and currently at Brookings, and he reviewed the list of research that has been presented in Stamford. He said parts of the research have been “called into question” (and referred us to research disproving these), and noted that the research being cited “uses very small samples” of few students and single districts. He referred us to a number of studies that provide greater clarity on the subject.
- We found that the claims of their being “no cost” to heterogeneous grouping and “minimal benefits” to homogeneous grouping are simply inaccurate: Studies, such as Slavin (cited in MSRP), showing “no cost to higher performing students” were based on data mostly going back to the 70s; and have subsequently been disproven such as by [Brewer, Rees & Argys]; [Page & Keith]; and [Kulik] among others.
- These studies were based on a straw man – the so-called “XYZ grouping” plans first implemented in Detroit in the 1920s, which group children by ability but then makes no changes to content based on ability. These types of systems indeed have little benefit [Slavin] [Kulik]. But they shouldn’t be used! When curriculum is appropriately varied and pace of learning is matched to the group, students outperform other students “of the same age and IQ by almost one full year on achievement tests.” [Kulik]

- Kulik (University of Michigan) performed a massive meta-study of hundreds of other studies and found gains to all levels in homogeneous grouping, provided instruction is varied and not simply “XYZ.” [Kulik]
- Furthermore high-performing minorities are the most harmed by heterogeneous grouping. [Page & Keith]. They found “Schooling in a homogeneous group of students appears to have a positive effect on high-ability students’ achievements, and even stronger effects on the achievements of high-ability minority youth. Grouping does not seem to affect negatively the achievements of low-ability youth. Indeed, ability grouping seems to have no consistent negative effects on any group or any outcome we studied” [Page & Keith]
- The case studies presented as evidence in MSRP (e.g., Railside and Rockville Centre) are from districts bearing little resemblance to Stamford – in one case a highly homogeneous small town, and in another case an urban, nearly entirely minority district. Neither were balanced districts trying to serve many varied needs as in Stamford. Railside (Boaler) has been refuted by colleagues at Stanford and Cal State as well [enclosed].
- Some of the teaching methods in the Railside study are in fact quite disturbing when read in its entirety. The vast majority of the “benefits” can be gained within a homogeneous system– high expectations, praising effort, etc. The remaining “benefits” come from what amounts to enlisting top students in “group work” in which they become the tutors of the lower performers, and are held accountable for other students’ performance for instance by being graded on their partners’ test results. [Boaler]
- Flight risk in a balanced system like ours also needs to be considered, not as a rhetorical threat of some sort, but as an important academic matter for those left behind. Studies that include controlling for the reality that top students have choices, find that homogeneous grouping benefits all levels, even when they analyze the same dataset that had previously been used by proponents of heterogeneous grouping, [Figlio & Page].
- The Petition attests to flight risks in Stamford: within the period of one week, several hundred parents signed on, many leaving important comments, opposing the move to heterogeneous grouping in fewer than four groups.

## Conclusion

- The claim that the research shows gains for the bottom at no cost to the top are clearly disputed, and controversial at best. Our read is that the research shows far more benefits of ongoing ability grouping for all levels.
- The case studies presented as evidence are of districts bearing little resemblance to Stamford – in one case a highly homogeneous small town, and in another case an urban nearly entirely minority district. Other studies presented have been disproven.
- The fact that only one side of a highly controversial body of research was presented raises serious questions about the process and should cause significant pause before making such momentous decisions.
- While many aspects of the MSRP are excellent, the unintended consequences of this move to minimize groups may create a very negative and irrecoverable outcome for the school system, for all levels of students, and for the city as a whole.

Following are 1 page summaries of several studies on this topic.

Links to full studies are provided next to citations.

## Tom Loveless (Brookings Institution) Finds Holes in Stamford's Research

Stamford parent Elise Kipness contacted the renowned expert on tracking Tom Loveless, former professor at Harvard and currently at Brookings, to evaluate the research presented in both the February presentation and the April "Middle School Reform Plan" document.

He said:

- "The Boaler study has been called into question" and he referred us to a detailed study refuting Boaler's claims.

*See Bishop (Cal State), Clopton (VAMC San Diego), Milgram (Stanford), "A Close Examination of Jo Boaler's Railside Report." [Link](#)*

What we found: In a devastating rebuke to the findings cited in the Middle School Reform Plan document from Prof. Jo Boaler, colleagues at Stanford and Cal State say that Boaler's attempts to hide the identity of the other two schools that Railside was compared to failed, and that when they uncovered the identity they found the other schools actually outperformed Railside in virtually every respect, and that "Railside students in fact do not perform well on state tests... or on AP or SAT exams."

- "A recent study out of Chicago is making quite a splash. It examined detracked 9th grade courses in math and English and found several negative effects, including negative effects for high achievers."
- "Brewer and Argys [show] a loss of achievement by high achievers placed in heterogeneous groups"

*See this study later in this document as well.*

- "I would also point out that the research being cited in support of detracking involves very small samples (a few students or a single district)"

## The move to end ability grouping can harm all groups

David Figlio (University of Florida, now at Northwestern) & Marianne Page (University of California-Davis) "*School Choice and the Distributional Effects of Ability Tracking: Does Separation Increase Equality?*" (National Bureau of Economic Research, 2000) [Link](#)

**Summary: This study finds that previous research has not considered the fact that higher performing students tend to leave school systems that abandon homogeneous grouping. It finds that when this form of "choice" is taken into account, homogeneous grouping creates gains for the bottom third.**

### Relevant excerpts

- Note: this study uses the term "tracking" to mean "homogeneous grouping," - it is not making a claim in favor of remaining in rigid, inflexible groups; it is supporting ability groups.
- "Previous studies have been based on the assumption that students' enrollment decisions are unrelated to whether or not the school tracks. When we take school choice into account, we find evidence that low-ability children may be helped by tracking programs." (abstract)
- "We find no evidence that low ability students are harmed by being grouped together, and conclude that the trend away from tracking has been misguided. In fact, we find that programs targeted to specific parts of the test score distribution have a substantive effect on a school's ability to attract high-income students, which may benefit low-ability students in a number of ways." (p3)
- Including "attracting better teachers" and "positive school-level peer group effects" (p3)
- "[When we] address the possibility that school choice is partly determined by tracking status, we find that tracking programs are associated with test score *gains* for students in the bottom third of the initial test score distribution. We conclude that the move to end tracking may harm the very students that it is intended to help." (p3)

### All levels benefit from grouping that includes varying the curriculum

Kulik, James A. (University of Michigan). *"An analysis of the research on ability grouping: Historical and contemporary perspectives"* (1992 and 1993). [Link to full text](#)

**Summary:** This study that "painstakingly catalogued the features and results of hundreds of studies" found that there are some old approaches to grouping that create few benefits – so-called "XYZ grouping" where kids are ability grouped but then the curriculum remains the same. These are the studies that have been used to supposedly show "grouping doesn't create benefits." However, proper grouping, in which curriculum is different for different groups, creates clear benefits for all levels. The study concludes that "Bright, average, and slow youngsters profit from grouping programs that adjust the curriculum to the aptitude levels of the groups."

#### Relevant excerpts

- Note that this article is based on "meta-analysis of... hundreds of studies" for "drawing a composite picture of the studies and findings on grouping"
- "Research literature on ability grouping used to be like the Bible. You could quote from it to support almost any view.... Now, reviewers are using new statistical methods to organize and summarize the literature on grouping, and its message has become clearer." (ix)
- "The reviewers have painstakingly catalogued the features and results of hundreds of studies... Their reviews have shown that certain approaches to grouping consistently produce positive effects on children, while other programs seldom produce measurable effects."
- The idea that there's no benefit to grouping comes from systems dating back to the 1920s, where groups were created based on ability but where the course content remains the same for all groups - so-called "XYZ grouping." (xi, xii) – both Slavin and Kulik did extensive analysis of this model and both found little benefit to grouping XYZ style. "Programs that entail only minor adjustment of course content for ability groups usually have little or no effect on student achievement." (vii)
- But why throw the baby of grouping, which does work, out with the bathwater of XYZ grouping which does not? [See pages ix to xv for a fascinating history of grouping.]
- "Grouping programs that entail more substantial adjustment of curriculum to ability have clear positive effects on children." (vii)
- When top groups use an accelerated curriculum they "outperform nonaccelerates of the same age and IQ by almost one full year on achievement tests." (vii)
- Conclusion: "Bright, average, and slow youngsters profit from grouping programs that adjust the curriculum to the aptitude levels of the groups. Schools should try to use ability grouping in this way."
- Of Jeanie Oakes (one of the "experts" cited who favors elimination of grouping), Kulik says "meta-analytic evidence suggests that [her approach] could greatly damage American education." (xv).

## Correcting the record on Boaler – “Railside students in fact do not perform well...”

**Summary: In a devastating rebuke to the findings cited in the Middle School Reform Plan document from Prof. Jo Boaler, colleagues at Stanford and Cal State say that Boaler’s attempts to hide the identity of the other two schools that Railside was compared to failed, and that when they uncovered the identity they found the other schools actually outperformed Railside in virtually every respect, and that “Railside students in fact do not perform well on state tests... or on AP or SAT exams.” [Link to full study](#)**

### Quote from Middle School Reform Plan Document

Boaler, J. & Staples, M. (2008). Creating Mathematical Futures through an Equitable Teaching Approach: The Case of Railside School. *Teachers College Record*, 110, 604-645.

- This longitudinal study was conducted in three high schools. Railside, one of the schools included in the study, is a diverse high school located in an urban area. Unlike the other two high schools in the study, Railside’s teachers used a reform-oriented approach to instruction and grouped all students in heterogeneous algebra classes....

### Boaler gets skewered in peer review

Bishop, Wayne (Cal State), Clopton, Paul (VAMC San Diego), Milgram, James (Stanford), *A Close Examination of Jo Boaler’s Railside Report*. [Link to full study](#)

- The authors took the time to write this rebuttal because “This study makes extremely strong claims for discovery style instruction in mathematics, and consequently has the potential to affect instruction and curriculum throughout the country.”
- “Prof. Boaler has refused to divulge the identities of the schools to qualified researchers. Consequently, it would normally be impossible to independently check her work.”
- However, in this case, the names of the schools were determined and a close examination of the actual outcomes in these schools shows that Prof. Boaler’s claims are grossly exaggerated and do not translate into success for her treatment students.”
- They found that the results Boaler got were in part from creating “un-validated tests tailored to favor their program and assessing low-level skills.”
- They analyzed the tests Boaler used for the study and found them to be “on average, roughly 3 years below the grade level where they were being administered,” and “Thus, the content validity (see Appendix 3) of these tests was shown to be extremely small.”
- They said “Whatever was being tested by these exams was unlikely to be relevant to student outcomes in the sense of their being able to actually use school mathematics either in their continuing educations or in their daily lives.”
- “We also found evidence that Dr. Boaler obtained her results by focusing on essentially different populations of students at the three schools. At Railside, her population appeared to consist primarily of the upper two quartiles, while at the other two schools the treatment group was almost entirely contained in the two middle quartiles.”
- When they checked data on SAT’s, AP exams and standardized state tests they found the other schools actually did much better. The “data does not support her conclusions” and “Indeed, there is only one year in the last five where any of these various measures for any cohort of students gives any advantage to the Railside students.”

## Boaler's Frightening Approach: sacrifices student growth while helping no one

**Summary of concern:** In addition to the fact that Boaler's outcomes have been called into question (see [Bishop, Clopton, Milgram](#)), the study itself raises grave concerns about the direction the Stamford school system might head if it were to embrace her approach – with students being required to act as what amounts to in-class tutors for slower students and being held accountable for other student's results, at severe cost to their own growth.

### **An Inapplicable and frightening approach**

- Not only did Boaler fail to show gains for Railside [Bishop et al], but this study also serves as a case study in what's wrong with the implementation of heterogeneous grouping. The greatest fear for top students is that the intent is for faster kids to have to constantly help the slower kids, and that they will be turned into in-class tutors when they finish their own work, instead of being challenged and continue to grow.
- That approach could theoretically improve scores for those receiving the "tutoring" (though even the claim of helping the bottom is called into question by Bishop et al). And it may not cause top students to fail state tests, since they are ahead of standards anyway. So it *could* serve the claim that is being made that you can do "no harm" to top kids while helping the bottom (through the free tutoring of "group work").
- Of course the issue/concern is that "no harm" is being very poorly defined. Does it mean that you didn't cause a child to literally fail? Or does it mean that you caused them not to reach anything near their full potential. On the first measure, yes maybe that's achievable (i.e., they'll still pass), but that's a very poor definition of doing no harm. On the second measure, it is doubtful that tying a slow runner to a fast runner is going to help the fast runner much.
- Is that too cynical a view? You would say "of course the schools are going to try to help everyone, and there's no intent of having the kids sacrifice their own learning time while having to help the other kids keep up..."
- However, that's just what has happened in Boaler. [Boaler's own article](#) and [slide deck](#) start with the big quote on the first page from a student: "What makes the class good is that everybody's at different levels so everybody's constantly teaching each other and helping each other out." (Zane, Railside school, presumably a slower student)
- Boaler states it as a positive that "teachers expected students to be responsible for each other's learning" and "ensure that students took their responsibility to each other very seriously." One method was that "the teachers occasionally gave group tests" and they "graded only one of the individual papers and that grade stood as the grade for all the students in the group."
- Another quote from Boaler of a lower performing student is revealing: "*Most of them, they just like know what to do and everything. First you're like "why you put this?" and then like if I do my work and compare it to theirs. Theirs is like super different 'cos they know, like what to do. I will be like – let me copy, I will be like "why you did this? And then I'd be like: "I don't get it why you got that." And then like, sometimes the answer's just like, they be like "yeah, he's right and you're wrong" But like – why?"* (Juan, Y2) – Should that really be the

responsibility of a student to tutor him on this? Wouldn't it be better for him to be in a smaller class of kids at his level with a good teacher?

### Correcting the record on Slavin – findings are “unwarranted”

Summary: Findings cited in the Middle School Reform Plan document from Slavin are clearly contradicted by later research, including [Kulik](#) (see section on Kulik) and Brewer, Rees and Argys below, among others. They call Slavin’s findings “unwarranted” and criticize it for being based on “small samples” and “with only one exception all the examined were done prior to 1978.” They explain “conventional wisdom that tracking does not have beneficial effects on student achievement has been undermined in more recent nonexperimental research”

### Quote from Middle School Reform Plan Document

Slavin, R.E. (May, 1993). Ability grouping in the middle grades: Achievement effects and alternatives. *The Elementary School Journal*, 93, 535-552.

- Through a review of the literature, the author concludes that there is no effect of ability grouping on high, middle, or low achieving students in the middle grades.
- Since there are no benefits to grouping, the author recommends discontinuing this practice, particularly because of the detrimental effects of ability grouping on students in the low groups.
- Districts moving away from grouping must also make improvements to curriculum and instruction to accelerate student learning and contribute to improved student achievement.

### Later research correcting this point of view

Brewer, Dominic (USC, Rossier School of Education); Rees, Daniel; Argys, Laura, “*Detracking America’s Schools: The reform without cost?*” (1995). [Link](#), [link2](#)

- “Perhaps the most influential review of tracking research was done by Robert Slavin. After summarizing 29 separate studies at the secondary level, Slavin concluded that the effect of tracking on students of any ability was ‘essentially zero.’
- “After reexamining this work, we have come to believe that such strong conclusions are unwarranted, for the following reasons. First, of the studies examined by Slavin, many were unpublished dissertations, which, of course, were not subjected to independent peer review. Second, of the experimental studies, most used small samples, often taken from a single school. Third, of the nonexperimental studies, none used nationally representative data. Finally, with only one exception, all the studies examined were done prior to 1978.”
- “Furthermore, the conventional wisdom that tracking does not have beneficial effects on student achievement has been undermined in more recent nonexperimental research that was based on large-scale data sets and that used more sophisticated statistical models.”
- Also see research by Adam Gamoran and Robert Mare, as well as Thomas Hoffer on this.

## Conclusion

- There is much to like about the Middle School Reform efforts. Movement toward flexible grouping and toward more frequent shifting between levels both have widespread appeal, as does ongoing staff development efforts, etc. There is no dispute about the vast majority of the reform effort, and Dr. Starr, the committee, and all those involved should be praised for their efforts on this.
- However, the movement toward heterogeneous grouping is concerning and highly controversial, both locally and within the research.
- The claim that “the research on heterogeneous grouping shows gains for the bottom at no cost to the top” is clearly disputed, both by direct experience within the school system as well as by the research. Our read is that the research shows far more benefits *for all levels* from ongoing ability grouping.
- Furthermore, the case studies presented as evidence are of districts bearing little resemblance to Stamford – in one case a highly homogeneous small town, and in another case an urban nearly entirely minority district. Other studies presented have been subsequently disproven by later research.
- While the stated intent is not to move to heterogeneous grouping, if the outcome turns out to be two groups it would still be highly heterogeneous with a wide range of students in the classroom. Also the fact that the research presented in support of whatever is about to happen was drawn exclusively from the heterogeneous camp causes concern about the underlying intent and direction. Hopefully this concern is unwarranted.
- More broadly though, the fact that only one side of a highly controversial body of research was presented raises questions about the process in this particular area of the reform, and should cause pause before making such a significant decision on drastically reducing the number of groups
- While many aspects of the reform plan are excellent, the unintended consequences of the move to minimize groups may create a very negative and irrecoverable outcome for the school system, for all levels of students, and for the city as a whole.
- We respectfully request that you consider enacting the vast majority of the reforms, but consider maintaining more groups, or at least taking a more deliberate approach to dramatically reducing the number of groups, such as running a pilot to see how it actually works *within our own system* instead of doing it city-wide in one giant step.