**JOHNS HOPKINS** | **Institute for Education Policy**
SCHOOL *of* EDUCATION

# What Should Researchers Research? Point and Counter-Point from Marc Tucker and Robert E. Slavin

April 23, 2018

*What should education researchers research? Two national leaders in the field, with very different perspectives, debate the issue below. We publish here two new papers, the first from Marc Tucker and the second, in response, from Robert Slavin. Marc Tucker, President and Chief Executive Officer of the National Center on Education and the Economy, is one of the most knowledgeable scholars in the study of best practices in education from around the world. Robert Slavin, Director of the Center for Research and Reform in Education at Johns Hopkins University, led Success for All to become one of the few whole-school reforms with a consistently strong evidentiary base. Slavin is a national leader in researching evidence-based practices. We believe the exchange will prove of great interest to education scholars, policymakers, and practitioners.*

## Paper 1:

### *The Problem with the "What Works" Approach to Education Research and the Case for Focusing on the Determinants of Highly Successful Education Systems at the Scale of a State*
by Marc Tucker
President, National Center on Education and the Economy

Over the last year, I have had the privilege of working closely with the Maryland Commission on Excellence and Innovation in Education ("the Kirwan Commission"). The Kirwan Commission is tasked with recommending education reforms, and funding mechanisms by which they will be executed, to the state legislature. One of the striking aspects of such conversations in the United States is the assumption that by "reform,"

1

one means specific interventions within a generally unchanged system: a new program of wrap-around services here, the expansion of access to AP programs there, more after-school and summer-school programs here or a state-wide tutoring program there.

As I have listened to testimony before the Commission, I have been reminded of the limitations of the What Works Clearinghouse model of education research, the model that has largely guided the U.S. government's approach to the use of research to improve American education. This model is focused on producing lists of discrete interventions that meet stated standards for the conduct of evaluation research. Considerable sums have been invested in such research, and vastly more on implementing the interventions. The government, on the other hand, has invested virtually nothing to improve our understanding of what makes some education systems at the scale of a state or a nation more effective than others. I will argue in this paper that this is a fatal mistake for the United States.

### Solving Maryland's reading problem: A classic application of the "What Works" research paradigm to an urgent problem of student performance

For instance, the Kirwan Commission heard testimony from experts on how to address the fact that Maryland's low-income students have very low proficiency levels in reading. One proposal – from Johns Hopkins University Professor Bob Slavin - called for the creation of a large-scale tutoring program that, if fully implemented, would cost the state more than $1.46 billion per year in new money to employ 17,000 new teachers to tutor struggling students. This sum would clearly take up a major portion of the funds potentially available to the state for reforms needed to propel the performance of Maryland's students from the average of all the countries taking the PISA exams to the top, the goal that was set for the Commission by the legislature.

This particular proposal reflected a framework first developed by researchers in special education called "Response to Intervention." The key to the notion is the idea of "proven programs," that is, carefully specified programs of instruction that have been shown by researchers to produce certain results for students under carefully specified conditions. Now imagine that we divide such programs into three "tiers." In the first tier are proven classroom programs to support regular, core instruction. In the second are proven small-group tutoring programs for targeted groups of students. And, in the third are proven, one-to-one tutoring programs for students who need intensive, individual help. The argument before the Commission was that providing the less intensive forms of help for students in the core program could reduce the need for more expensive and intensive tutoring later, which is why Slavin was advocating for a broader, less-targeted program. Within this suggested intervention strategy, there are a number of programs that now meet the federal government standard for "proven" programs at various levels of strength of evidence for mathematics and reading at both elementary and secondary levels – and a small number of them show particularly good results.

**In my judgment, however, what the "proven program" research paradigm actually does is identify programs that produce marginal results in a dysfunctional system, when the real issue is how to fix the system, a problem that cannot be addressed with this paradigm.**

The underlying logic is simple. Start with the problem - say, a large proportion of students leave elementary school two or more years behind in reading. Come up with a theory about the cause of the problem and, to test the theory, use the theory to develop a "treatment" (in medical terms) or a program (in educationists' terms). Administer the treatment with statistical controls that will enable the researcher to establish the size of the effect of the treatment on student performance under the specific conditions under which the research was conducted. Then, put all the programs whose effect size crosses a certain threshold and meet certain criteria for research quality on a list of proven programs. Then stand back and watch the policymakers implement them in great numbers, replicating everywhere the results the researchers observed.

Except, of course, they don't. They never have, and when they do, we don't see much improvement at scale. What is going on here? Are the educators simply misguided, or stuck in their ways? Or is there something fundamentally wrong with this model? I am in the latter camp and I will explain why.

My organization has been studying the countries with the most effective education systems for close to 30 years. This has been very instructive. Consider that, when a demonstratively powerful and effective policy or practice has been identified in other countries and brought back to the United States, it is almost sure to be abandoned before long. Why? Because it does not work in our country. And why is that? Because the effectiveness of the policy or practice depends for its success on other elements of policy or practice that are absent or very weak in our country, or are counteracted by policies or practices in the United States that are absent in other countries. Other countries, for example, might limit the right to offer teacher education programs to their research universities, thus denying the right to pursue teaching as a career to high school graduates who cannot get into the research universities. In our country, such a policy would quickly dry up the supply of applicants and, pretty soon, there would be no teachers. But that does not happen in these other countries, because their teacher salaries are significantly higher and they offer better working conditions to their teachers.

What I am saying is that these other countries are out to design effective *systems*, while we are looking to identify effective *programs*. It is actually worse than that. The What Works research paradigm is designed to measure the unique effects of the program while using all kinds of very sophisticated methods of holding all other variables except the dependent variable constant. In other words, the whole architecture of the method is intended to screen out system effects, to ignore the ways in which the system is affecting the outcome of interest. In our effort to identify the independent effect of the intervention, we go to ingenious lengths to ignore the myriad ways in which the system

itself is working to minimize the effect of any single intervention. Many researchers do not, in practice, ignore these factors, but describe them as important features of the context, acknowledging that they can defeat effective implementation of the specified program.

Again and again, we celebrate what are, in fact, very weak effects of our interventions on student performance. Why? Because that is the best we can get, given the context in which they will be implemented. That context is highly dysfunctional. How do I know it is dysfunctional? Because 15-year-old students in the typical American state are performing two to three grade-levels below their counterparts in the top-performing countries, many of which are the size of the typical American state. The top performers are usually getting much better results on equity, and there is good reason to believe that the top performers are spending no more than the average American state. Why this stunning difference in results? It is at least in part because those countries have powerful, comprehensive, highly aligned, systemic strategies for improving their education systems, and our states typically do not. The reality in the United States is that our state departments of education are much weaker than the ministries of education in countries the same size as our states, and the school districts are much stronger, each pursuing their own agendas. Each district is led by superintendents whose time in the job gets shorter and shorter, which creates strong incentives to search not for systemic solutions that would take years to implement, but rather for short-term fixes that will have little or no impact on the systemic issues that stand in the way of enabling our states to match the performers of the countries that lead the global league tables.

How else do you explain the fact that some of our richest states get such mediocre results compared to much less wealthy countries? If the key levers of top performance in the countries with the best education systems don't work in this country because they are not supported in this country by the other elements of policy that make them work in the top-performing countries, why would we expect that good ideas *developed* here would work any better? Only Massachusetts has succeeded in entering the ranks of the global top performers and that is because only Massachusetts, in its landmark Education Reform Act (1993), created the kind of systemic approach embraced by all the other top performers, and, through thick and thin, and with considerable political skill, stuck to it long enough to produce impressive results.

What researchers in the United States are doing is identifying programs that are at least making a little difference in a highly dysfunctional system. They tell you nothing whatsoever about how to build a highly effective system. They are a prescription for assembling a house of Band-Aids, when we could be building a great house. Not only do they not tell you what a highly effective education system looks like, but they do not even tell you what effective programs would look like if we had an effective system. They will only tell you how to get marginal improvements in a dysfunctional system that is very similar to the dysfunctional system in which the research was done.

But that is not the only problem with the prevailing paradigm. The paradigm only works if the researched interventions are faithfully implemented in just the form in which they were researched, under the same conditions. This is an unworkable model.

## The crippling limitations of a model of education research based on the idea of "replication"

Much of my work is with states at the policy level and with the governments of other countries. I have yet to meet a governor, minister, commissioner or governor general who is interested in copying any other country, state or province. Many are intensely interested in learning as much as possible from their colleagues elsewhere, but no one needs to tell them that those colleagues were addressing different constellations of problems and a very different political context, had very different opportunities and faced very different obstacles. They know that their colleagues did some of the things they did as workarounds for particular problems they faced and would never have done them if they had not been forced to do so. They also know that some of their colleagues were able to do some of the things they did because they were just plain lucky. Every situation is different. Under those circumstances, the idea of a research-into-practice model that is based on faithful implementation of someone's research design or reported intervention is simply laughable.

This does not mean that nothing is generalizable or that everything is context-specific. There are common principles that explain the effectiveness of highly effective systems. It is the search for those principles that underlies our work and the work of others who want to help education system designers around the world.

## Why systems and the way they function are so important

And that bring us to the main point, which is that effective schools, districts and states are not compilations of effective programs. They are effective systems. You may have a great way to teach reading, but, if you have lousy teachers, it won't produce great reading results. You may have great teachers, but, if the school leader is a petty tyrant and does not support good teaching, the good teachers will either leave or give up while going through the motions of teaching. And student performance will lag. You may have good leadership, but the district might insist that the school use a strategy for teaching reading that the assistant superintendent for instruction fell in love with at a workshop a couple of years ago and the teachers think is just junk. You might have a great reading program, but student mobility is very high as families in your community face a great deal of housing instability, resulting in frequent moves for many low-income students.

The typical What Works Clearinghouse evaluation design for new programs does not control for any of this. It might control for student poverty, social class, race and ethnicity, age, parent's education background and on and on, but is less likely to control for the kinds of system features I just mentioned. This is true also of the Best Evidence

Encyclopedia – a resource provided by Slavin and his colleagues that enables policymakers to review the evidence supporting different educational interventions.

And that is precisely the problem. Many of the problems faced by our schools are the result of having a system in which many different decisions about our schools are made by different bodies and levels of government that operate independently of one another and often in conflict with one another. Nothing fits together. We typically don't see this because we are so used to it. The only way to see it is to compare it with systems outside the United States. When we do that, we see in the top-performing countries systems that work much better, in which the parts and pieces work in harmony with one another.

**The What Works Clearinghouse approach to research intended to improve outcomes for students could not have been better designed to be manifestly unsuited to research on effective systems**

The What Works Clearinghouse approach to research could not have been better designed to be manifestly unsuited to the real problem we face, which is how to build more effective education systems. When researchers in the United States tell teachers to use findings that are based on statistical methods for holding everything constant except for the program being evaluated, they are really telling teachers, administrators and education policymakers to assume the perpetual existence of the very factors that make our system dysfunctional, excessively expensive and unusually ineffective. They are prescribing fixes that are doomed to produce very modest effects, because they are, by design, ignoring the very factors, which, if changed in the right ways, could produce very large improvements in the performance of virtually all of our students at very little increase in cost.

**The beauty of the comparative method**

How do I know this? Because close to two dozen countries are doing this right now and have been doing it for years, many of them routinely graduating high school students who are two to three years ahead of ours on average. A higher proportion of their students than ours place in the top PISA performance band. The gaps between their lowest-performing students and their top-performing students is smaller than ours. The proportion of their disadvantaged students who end up in high school doing very well is higher than ours.

This is the beauty of the comparative method. If your system—for making steel, providing health care or educating children—is not working as well as others, go and take a look at the ones that are working well and figure out how they do it. As you will see in a moment, this is not a search for models to copy, the model on which the What Works Clearinghouse approach is based. It is not about replicating anything. It is a search for the principles underlying the design and implementation of highly successful systems,

principles that can typically lead to different designs differently implemented. The designers of the most successful systems use the principles, examples and narratives of this kind of research to design their own systems in ways that are carefully adapted to their own aims and circumstances, informed by, but not constrained by, the kind of research I am describing.

## Applying the critique to the problem with which we began

Thus far, I have been making a rather abstract argument. Suppose now we apply it to the proposition on tutoring placed before the Kirwan Commission: You can solve or at least greatly ameliorate the reading problem in Maryland by hiring 17,000 additional teachers to do tutoring at a cost of $1.46 billion per year, over and above what Maryland is currently spending.

You will say, "Marc, if you do not like that program, what program do you like?" And I will say, "No program." And you will be more than a little impatient. If I am not offering a better program, why am I wasting your time?

This no place for a full description of the relevant policies and practices of the top-performing nations. I will pick out just a few to make my point.

First, the top performers do far more than the U.S. to make sure that young children get the support they need before they enter compulsory education.

Reading with comprehension involves far more than decoding the words on a page. Students who hear words spoken and used all the time are far more likely to recognize and use those words in their own speaking, reading and writing. Students who can connect a word to the cultural content of their everyday experience are in a much better position to recognize and use that word than those whose only meaning for that word comes from a dictionary. So, students who come from culturally impoverished homes or no homes at all are deeply disadvantaged from a beginning reading point of view before they ever arrive at the school house door. The top-performing countries typically spend much more than does the United States on child allowances, child bonuses, child care, child nutrition and early childhood education than the U.S. does. And they do so in an environment in which the disparity between rich and poor is much smaller than in the United States and the proportion of students who are poor is smaller. The result of this combination of policies is that children come to school ready to learn to read in much higher proportions in the top-performing countries than in the U.S. American researchers have known all this for a long time, but most proposals in this country do nothing about any of this.

The U.S. recruits its teachers mainly from the middle of the distribution of high-school graduates going to college; the top performers recruit from the middle to the top. They can do this because they offer compensation comparable to that offered to students going

into the high-status professions, as well as comparable working conditions. In a growing number of countries, only research universities can offer teacher education programs, so only students with academic achievement high enough to get into a research university can become teachers. Once in, they have to meet research university standards of accomplishment to get their degree. Future elementary-school teachers are given a sophisticated understanding of the reading research worldwide and enter practice with a deep understanding of that literature and the capacity to apply it to a wide range of situations. But discrete, programmatic interventions like the tutoring proposal presented to the Commission do nothing to raise the quality of our teachers.

When they enter the profession, new teachers are apprenticed to master teachers, much like residents in a teaching hospital. The schools are organized very differently than ours. Teachers have a real career in teaching, usually culminating in the position of master teacher, with salary and status comparable to that of a principal. Teachers spend much less time teaching. They spend much more time working together in teams to systematically improve every aspect of instruction and support for students. They typically meet every week by grade and subject.

In this system, when the teachers meet by grade, they are expected to identify students who are in danger of falling behind. All the teachers of a student who is having trouble will work as a team to figure out whether the problem is the result of a home eviction or shooting, or of a specific issue in deciphering the code of the language in reading. They will agree on a course of action, which could range from a home visit to extensive tutoring by the student's regular teacher, all of whom have had extensive training in specific tutoring techniques. Of course, it might be the case that there is a whole collection of issues, and the teachers might have to form a team to deal with that student until she is back on track. In such systems, the teachers not only have the time to figure out what the problem is, they have the time to do something about it.

It might be that the teachers' analysis reveals a problem that is not specific to one student, but instead is being experienced by many. In that case, the teachers might form a team to build new lessons for all the students that do a better job of engaging the students and conveying the content.

In systems of this kind, very high-quality tutoring is built right in to the way the school does its work. The tutoring is done by the regular classroom teachers, so it is closely aligned with the regular classroom instruction. It is not an add-on program, poorly aligned with the regular classroom work of the student and teacher, requiring the hiring of thousands of additional teachers of unknown quality.

It is, of course, true that at least some aspects of what I have described can be found in the practice of some of our richest school districts. What distinguishes our states from the countries I am describing here—many the size of American states—is that these

practices are systemic and universal in those countries, whereas they are rare and spotty in the U.S.

Trying to raise the academic performance of America's school children through the implementation of a series of discrete programs will, in my judgment, do nothing at all to improve the regular, core instruction the students get or the likelihood that regular classroom teachers will have the time or skill to identify and correct reading problems.

Here is the bottom line. When our team visits the top-performing countries, we often ask middle-school principals what they do for students who enter middle school two or three years behind in reading, a common occurrence in the U.S. Invariably, they just look at us, puzzled. It does not happen.

And why does it not exist? It is not because they have better programs than we do. It is not because they have implemented programs of the sort that What Works Clearinghouse endorses and we failed to do so. I have described no programs. What I have described are system features. It is those features of their systems that largely account for their success in teaching reading and everything else. If we fail to adopt similar design features, our performance will continue to be sub-par to mediocre.

The Kirwan Commission understands this. In its effort to find ways to enable the state to match the performance of the top-performing countries, it has worked hard to understand the key strategies used by the top performers to get to the top and to adapt those strategies to fit Maryland. It is determined to help the state build a more coherent, more powerful system not by copying any country's system, but by learning from the experience of all the top performers.

## An indictment

The program evaluation model – the What Works Clearinghouse model - ignores all the system design features that explain the success of the top-performing countries. It takes as given the very system which, if changed in the right ways, could and in fact does address all the issues the program being studied is addressing, but with much more success. The program evaluation model is doomed to failure, because at best it can only triage a very ineffective system, and even the triage is often so expensive that a state must choose between implementing the triage or changing the system.

But, as I pointed out above, that is not all that is wrong with this model. It guarantees the advertised results only with faithful replication of the conditions under which it was researched, even though, for very good reasons, educators are rarely interested in replicating anything.

This last point is especially important to consider when asking whether there is any value in this kind of research. The answer is an emphatic "yes," if it is used appropriately. As

you saw in my description of how new teachers are trained and how they work in teams to systematically improve instruction, they are constantly reading the global research literature, looking for analysis and findings that can help them, as any high-status professional would. They are looking for solid research from multiple sources that they can use not to copy anything but to craft their own solutions to the problems their own students face, just as high-status professionals do in many other fields.

But that does not make this model of research benign. To the extent that this model of research largely defines the U.S. government's approach to school improvement in the United States, it has become not merely useless, but a major roadblock standing in the way of the only strategy that can get our schools to parity with the top-performing nations.

## The way forward

So what is needed instead? The answer should be obvious by now. The question facing Maryland is not what programs Maryland should adopt but how it should spend the money it has available to change its *system* to produce the biggest gains for its students. That question—how do we build highly effective education systems at the scale of a state? —should be leading the agenda of the American education research community.

But that question is nowhere to be seen on the nation's education research agenda. It is not on the map. What is so strange is that we know that education systems the size of a state can perform much better than the vast majority of them do now, because we have a growing number of education systems about the size of a typical state that do so right now. Our researchers should be all over those countries, finding out how they did it, but they aren't. The Massachusetts Education Reform Act was passed in 1993. The state went on to become the only state in the U.S. to rival the global top performers. That success was fully visible more than 20 years ago. There has yet to be a single solid analysis of how Massachusetts did it. That, in my view, is a case of criminal neglect.

Many elements of what is needed are at hand. I am hardly alone in this critique of the standard What Works model. Many researchers in the United States and abroad have been developing techniques and methods that can be used as is or further elaborated in combination with others to investigate the drivers of success in large-scale systems. Some have been pursuing research intended to facilitate effective systems design. Others have been exploring approaches that focus on finding common drivers in very different system contexts. Still others have been working on the kind of multi-level analyses required to understand complex dynamic systems.

Tony Bryk deserves a special shout-out here. Throughout his whole career, Bryk has worked hard to figure out how to produce education research that would be useful to practitioners, and has understood that would require a fundamental reorganization of the way the research is done. I am a big fan of his current work, focused on the creation

of networks in which researchers join practitioners and stakeholders in defining the goals of the research and the way in which the research is conducted and used.

But, even here, to my knowledge, the focus of Bryk's work is the local scene. The larger policy system sits above it, unchanged, outside the orbit of the local network. So, in the sense in which I have used that word here, the larger system remains unresearched and unchanged, accepted as an unchanging context for the local network's program of research and action. It is precisely that larger system that must be redesigned and rebuilt to produce the kind, extent and scale of change that I have been describing.

To be clear, what I am advocating for is a large program of research on the most successful education systems in the world, organized to help American states understand what combination of features of their systems account for their success, or, put another way, what the common principles are that underlie the different approaches they have taken. What is needed is a design orientation, which is to say that the purpose of this research should be to facilitate the redesign of our current state systems of education for high performance.

If I had my way, the National Academy of Education, the National Academy of Science and the American Education Research Association would each convene their own high-level panel of researchers to develop a research agenda around the goals I just put forward. Those panels would be charged not just with developing a research agenda, but also with examining the research tools and methods now available to see whether the tools and methods are adequate to the task ahead or whether others will be needed. Over time, I would hope that the U.S. government's image of research-driven school improvement would change from replication of a heap of programs to the design and implementation of highly effective education systems.

## ACKNOWLEDGEMENTS

# Paper 2:

## *Using What Works is Maryland's Best Way Forward*
by Robert E. Slavin
Director, Center for Research and Reform in Education
at Johns Hopkins University

I have known and respected Marc Tucker for many years, since we both led major school reform projects funded by New American Schools starting in the early 1990s. Even back then, we represented two different approaches to school reform. He has always believed that school reform depends primarily on educational policies and less on practices or programs. I have always believed that to make a substantial difference for students, schools must implement programs and practices that rigorous research has proven effective. I also believe that policy and systems are also crucial and have fought throughout my career for policies designed to put into place proven programs. Medicine and public health inform my model. These fields have made extraordinary progress from a combination of rigorous research on the effectiveness of drugs and procedures and public health policies that put proven therapies into widespread and effective use. In contrast, Marc has become our most articulate advocate of the idea that the solution to America's educational problems is to borrow from "top performers," by which he means other countries, such as Singapore, Finland, and Hong Kong, that perform very well on international tests.

In all this time, Marc and I have never had a chance to address each other's viewpoints directly. I'm delighted to have an opportunity to do so, occasioned by the Kirwan Commission's consideration of new funding formulas for Maryland's schools. In that context, Marc has been consulting with the Commission for more than a year. My involvement has been much less: only an hour or so in October when I presented a paper for the Commission entitled, "Achieving Proficiency for All: Maryland's Opportunity."

Marc prepared a paper, published with mine at the Johns Hopkins Institute for Education Policy, in which he rejects the entire concept of "what works." He argues that because "what works" research seeks to hold constant all factors except the experimental treatment, programs validated in rigorous research can only make marginal differences in the United States, because our systems rather than individual practices constitute the underlying problem. Why create, evaluate, and disseminate better reading, math,

science, or whole-school reform approaches, when we should be working on becoming Singapore instead?

But had American medicine taken the same view, we would not have experienced the dramatic improvement in life expectancy and the ability to prevent and treat a broad range of diseases. Better health policies surely helped, but so did research that identified effective treatments for diseases. The virtual eradication of polio, smallpox, measles, chicken pox, and many other diseases did not happen due to structural changes alone. They required research, development, and dissemination, using exactly the research standards increasingly required in educational research.

In a recent blog, Marc wrote about his experience of spending two weeks in a Florida hospital after a bicycle accident. I'm delighted to report that he came through fine. In his blog, he praises the doctors and nurses who are exquisitely trained to know which medicines to give, for which symptoms, in what doses, and for how long. I share his admiration, but I have several questions: Where does he think those medicines came from? How did we learn which medicines work, how much to give, and how long? Did the exquisitely trained doctors and nurses make this up by trial and error? Of course not. Marc's life was saved not just by skilled doctors and nurses, but by a huge enterprise, devoted to creating and rigorously evaluating promising treatments for every disease. Once research validates a medicine or treatment, it is disseminated worldwide, and doctors and nurses are trained in its proper application. It becomes part of "standard of care," which means that hospitals and doctors <u>must</u> use proven treatments in well-defined situations, or face possible legal penalties. Marc's blog contrasts medicine to education but focuses entirely on the training and mentoring that doctors and nurses, but not teachers, receive (by his reckoning). However, he leaves out the process by which new methods and materials for education (and medicine) are created, evaluated, and disseminated, and it is this very "what works" process that his paper rejects.

### The Critique of Response to Intervention (RTI) Done Right

Marc's central argument relates to my proposal that Maryland implement a substantial tutoring program to advance the state's goal for reform: To ensure that <u>every</u> child achieves proficiency on the annual PARCC assessments. My presentation attempted to demonstrate powerfully that this goal requires substantial intervention. Maryland is the equivalent of merely 10 PARCC points[1] behind Massachusetts on reading and math tests. It is, in fact, possible to readily improve by 10 points - which would make Maryland the highest-performing state - by helping all schools select and effectively implement proven programs, i.e., ones known from rigorous, medical-quality research to improve outcomes by 10 PARCC points or more.[2] Such a policy would cost a tiny fraction of the costs I

---

[1] The standard deviation of PARCC is 50 points, so increasing PARCC by 10 points is 20% of a standard deviation, or an effect size of +0.20. In this way, any effect size can be transformed into a PARCC-equivalent.

[2] The Center for Education Research and Reform, which I lead, created and maintains a website, www.evidenceforessa.org, that lists numerous reading and math programs meeting this standard.

proposed and would raise the average PARCC score from 240 to 250 (proficient is 250). But this would still leave fully half of all Maryland children below proficient. Is that all we want?

No. I tried to make the case to the Kirwan Commission that we could enable far more than half of Maryland students to reach math and reading proficiency, also using proven programs – of a different kind.[3] The only programs known from research that routinely add the equivalent of 20 or more PARCC points involve tutoring. This is particularly true when tutoring exists in a response-to-intervention format, in which students receive only the services they need. Tutoring is expensive. However, its costs can be greatly reduced by hiring high-quality paraprofessionals (teacher assistants), such as ones who have a B.A. Also, effective tutoring is likely to reduce special education costs in the long term. The Center for Research and Reform in Education (CRRE), which I lead, recently completed a research review and found that tutoring from high-quality paraprofessionals exercised substantially positive outcomes on student achievement, averaging the equivalent of 26 PARCC points for one-to-one tutoring in reading or math, and 14 points for one-to-small-group tutoring. If continued with integrity and care across multiple years, a growing number of students would reach "proficient" each year, Maryland's students eventually could advance far beyond those in Massachusetts and "top-performing" countries. And there would be additional benefits: the apprenticeship model of hiring and training high-quality tutors could bring talented, eager, recent college graduates into the teaching profession.

The above-mentioned studies of tutoring and other response-to-intervention approaches engage with all sorts of teachers and tutors, i.e., a cross-section of the types of individuals likely to be available in Maryland or other states. My plan does not depend on waiting for who-knows-who to arrive from who-knows-where. Instead, it depends on giving better tools and better supports to the outstanding teachers who are already in our schools and to para-professionals in Maryland and nearby who already have B.A.'s.

Marc's paper does not lay out a clear plan or clear costs, but instead asks for support for "a large program of research on the most successful education systems in the world." In Maryland, this is not the moment to ask for more research. It is not the moment to try out solutions that seem to be working in far-away countries but have never been transferred to the U.S. Maryland has a once-in-a-decade chance to make a substantial difference with its students. We should start to implement programs that we already know will work, and then figure out how to create sustainable structures and policies to improve the outcomes we strive for in our commitment to all students' success.

The solutions I am proposing are not theoretical. They do not ask us to wait for more research before we act. The federal government has defined the types of research needed to certify a program's effectiveness, and ties its funding to the use of such programs. Every program I proposed for Maryland already meets these federal effectiveness standards.

---

[3] Of course, the farther students are below 250, the more expensive it is to get them to proficient.

Research will of course continue, and more solutions (structural as well as classroom-focused) will surely emerge. Some of these will derive from foreign examples. But we owe it to kids and taxpayers to use what works, now, while we continue to learn and improve over time.

The most important problem in America's schools is not our middling PISA scores. It is the persistent gaps in achievement according to social class and ethnicity. Middle-class, White, and Asian students do not present major achievement challenges for our country. It is African American, Hispanic, and Native American students, and disadvantaged students of all ethnicities, whose learning demands our full attention. Massachusetts, where 24% of the population are African American or Hispanic, may be able to reach world-class status without having a substantial impact on minority achievement, but Maryland, with 45% minorities, cannot.

My proposal goes to the heart of this problem. There is nothing wrong with struggling learners that tutoring and other proven programs cannot substantially improve. As one example of this, CRRE evaluated a high-octane version of our Success for All program in some of Baltimore's most disadvantaged schools. Success for All provides proven programs for reading, parent involvement, social-emotional skills, and much more, but it also provides one-to-one or one-to-small-group tutoring for students who need it. In some schools, we had as many as six tutors using proven strategies.

The outcomes, compared to other Baltimore schools, were extraordinary: the average school-wide effect compared to control schools was a PARCC equivalent of 25 points (see Madden et al., 1993; Borman & Hewes, 2003; and Borman et al., 2007). For the students who started in the lowest quarter of their classes, and were therefore most likely to receive tutoring, the difference was a PARCC equivalent of 37 points. An independent follow-up to eighth grade found that former Success for All students were still far ahead on standardized tests and half as likely to have been retained or assigned to special education.

In Steubenville, Ohio, a hardscrabble, post-industrial city a bit like Cumberland, Success for All has been used for 18 years. Its elementary schools have long performed at the top of the state, ahead of suburban districts.

Using Success for All or other proven approaches that include adequate amounts of tutoring statewide could move Maryland forward and make it the envy of Singapore. More importantly, such a strategy would apply attention and resources to the identified problem as it actually exists, rather than to improving our ranking on a list somewhere.

My vision for Maryland ten years from now is a state that has figured out how to ensure that all students, regardless of background, succeed in school, and that constantly finds better, more cost-effective ways to build upon what has been proven to work.

What I hope *not* to see are thousands of once-promising young men and women condemned to lives of poverty and hopelessness, and leaders who are ashamed for having squandered their chance to do something about it in 2018, because they were chasing Singapore dreams.

The children of Maryland have waited long enough. While we debate about structural changes and learn what other countries do, let's make a bold commitment to ensure success for the students in our system today, and those yet to come, using strategies already proven to work in American schools.