



**Testimony of Michael Cohen
President of Achieve
Indiana Interim Study Committee on Education
September 29, 2015**

Senator Kruse, Representative Behning, members of the Committee, thank you for the opportunity to testify before you today. This is the second time this year that I have testified in Indiana on assessment-related issues. I appreciate how complex these issues can be — as they raise an interrelated set of educational, policy, technical, stakeholder, and cost issues, and often touch a range of commercial issues as well. This past week I spent several days at a meeting of some 35–40 state legislative leaders from about half of the states discussing many of these same issues.

You are surely not alone in tackling assessment issues right now. As you well know, in the past several years, testing has become increasingly controversial — and debates about testing increasingly contentious — in many states throughout the country. One concern often voiced is that there is *too much time* spent on testing, at the expense of instructional time. Or that *too much attention* is being paid to tests themselves. Another concern — often heard in legislative halls — is that *too much money*, too much of every education dollar, is spent on testing. Yet another is that tests often *don't provide timely, useful, and actionable information* to students, their parents, and teachers, or to other education leaders. Last, but not least, many are concerned that for too long, state accountability tests, typically dominated by multiple-choice, fill-in-the-bubble tests, promote rote learning and don't measure the critical thinking, problem solving, and application skills that two- and four-year colleges and employers increasingly demand, and that our society needs of all its citizens.

I'm not here to advocate for or against specific solutions, or to sell you a test. I'm here to do what Achieve has done for some 15 years with policymakers in Indiana — offer you our best advice based on our own research and experience, the lessons from the research of leading experts, and the experience of other states facing similar issues.

Achieve has worked with Indiana over a 15-year period to evaluate and, where necessary, recommend improvements to its standards and assessments. Achieve first reviewed Indiana's English language arts and mathematics standards and assessments in 1999 and recommended significant improvements in both, which were largely incorporated in subsequent drafts. In 2003, Achieve conducted an analysis of the newly implemented ISTEP+ to determine the extent to which it was aligned with state standards, as well as whether the proposed scores for passing represented "solid academic performance" and the Pass+ score represented

“exemplary performance.” This study, commissioned by Indiana leaders, demonstrated an unprecedented commitment to transparency and quality with respect to setting cut scores. In 2004, Achieve used the Indiana mathematics standards as a benchmark — a standard of excellence — in our reviews of standards from other states.

Achieve has been a partner and source of expert advice in other ways as well. Indiana was one of five states (along with Kentucky, Massachusetts, Nevada, and Texas) that joined the initial research phase of American Diploma Project. Project researchers from Achieve and our partners, the Thomas B. Fordham Foundation and the Education Trust, worked with college faculty, employers, and high school curriculum experts in each state to identify academic skills in mathematics and English language arts that are essential for success in broad access to postsecondary institutions and careers that pay well and have advancement potential. This was the very first effort by states to anchor academic standards in the best available evidence of the essential demands faced by students preparing for college, work, and citizenship. The resulting American Diploma Project Benchmarks were subsequently used by more than half of all the states, including Indiana, between 2004 and 2009 to develop their own college- and career-ready standards in mathematics and English language arts.

In 2006 Achieve developed a set of recommendations to help Indiana increase its high school graduation rate while also raising its graduation standards, while others thought it would be nearly impossible to simultaneously raise both. In 2009, Achieve featured the importance of the Education Roundtable to develop and sustain a P-16 plan. Just last year, at Governor Pence’s request, we reviewed the draft academic standards in mathematics and in English language arts, and recommended their adoption to the Education Roundtable.

Last Spring I testified to the House Education Committee on testing issues it was considering at the time. I am pleased to be back here now to help continue that conversation.

I know that the cost of testing has been part of the ongoing debate about the future of testing here. Before proceeding to the discussion of specific testing issues, it’s important to put the debate in a broader context. According to the latest report from the Census Bureau, in FY 2012 Indiana spent \$9,719 per pupil in K-12 public schools. Indiana spends approximately \$40 per student for the ISTEP+ English language arts and mathematics tests. *This is less than one half of one percent of per pupil spending.* That may look like a lot of money when it turns up as an item in an appropriations bill, but as a portion of overall education spending, it’s a small amount to spend on something that plays such an important role.

Let’s start by looking at the types of assessments students experience each year. The first two are often referred to as **assessments FOR learning**. They are generally developed and/or selected at the local district level, with minimal

involvement of the state, although I am aware that Indiana has developed its own diagnostic tests and has paid for their use in districts.

Formative assessment is really a process used by teachers and students during instruction that provides feedback to adjust teaching and learning in real-time. The goal is not to measure whether students have mastered a cumulative amount of material; rather, the goal is to improve students' achievement of intended instructional outcomes daily or every few days. Formative assessment is embedded within learning activities. The assessments are small scale (a few seconds, a few minutes, less time than a class period). Formative assessment is done by the classroom teacher for the explicit purpose of diagnosing where students are in their learning, where gaps in knowledge and understanding exist, and determining how to help improve student learning. Providing corrective feedback, modifying instruction to improve the student's understanding, or indicating areas of further instruction are essential aspects of formative assessment. Examples of formative assessments include end-of-unit tests (such as those at the end of a textbook chapter); projects; homework; in-class activities (including lectures and classroom discussions); quizzes; and teacher-developed tests. Note that it does not make sense to aggregate formative assessment information beyond the classroom.

When well designed and implemented, these ways of assessing students provide *periodic, timely, and targeted feedback* to teachers and students. In the context of well-defined learning objectives, these help teachers and students know where they are in their learning, and how to close the gap to where they need to be. This kind of assessment is a crucial component of good teaching. In one form or another, it has been going on as long as teaching has.

Interim or benchmark assessment is the term for the assessments that fall between formative and summative assessment. Interim assessments (1) evaluate students' knowledge and skills relative to a specific set of academic goals, typically within a limited time frame, and (2) are designed to inform decisions at both the classroom level and beyond the classroom level, such as the school or district level¹. Thus, they may be given at the classroom level to provide information for the teacher, but, unlike true formative assessments, the results of interim assessments can be aggregated meaningfully and reported at a broader level. As such, the timing of the administration may be controlled by the school or district rather than by the teacher. This means these assessments are less instructionally relevant than formative assessments. Interim assessments may serve a variety of purposes, including predicting a student's ability to succeed on a large-scale summative assessment, evaluating a particular education program or pedagogy, or diagnosing gaps in a student's skills.

¹ Often these tests are called "formative," though they are not. The terms "formative" or "diagnostic" are often used interchangeably with each other and with the interim or benchmark assessments discussed here, leading to unfortunate confusion.

In many schools and districts, these tests have been added on to the classroom assessments described above, often in response to the pressure experienced by adults by the state end-of-year accountability tests. These tests are usually given quarterly and may be “fixed form” tests, with each student taking the same set of questions, or they may be “adaptive” tests, designed to determine each individual student’s level of performance by giving them a unique item mix based on their responses to an initial set of questions. Either way, these tests, typically used district-wide, are intended to provide information about individual and classroom level strengths and weaknesses that can inform the next phase of classroom instruction. However, in some cases, the results are not just left in the hands of the classroom teacher, but instead are also aggregated and reported up to the school or district level to inform system-wide responses to low performance. Unfortunately, an unintended consequence of this approach is that assessments that are intended to be “formative” or “diagnostic” in nature can frequently become an additional source of test-based pressure, leading some teachers to “prep” students for these tests just as they do for the state summative assessments.

In contrast, state summative assessments are best thought of as assessments OF learning. Summative assessments (also called large-scale or standardized testing), such as ISTEP+, are generally given once at the end of every unit of time, such as the semester or school year, in order to evaluate student performance. Virtually every student enrolled in the tested grades is required to take these tests, with accommodations available for students with disabilities and English language learners. The scores are used in accountability determinations for schools that involve public reporting and consequences. They are also often a factor in educator evaluation. And in some states they are a factor in determining grade-to-grade promotion and/or high school graduation. Summative tests need to be given under standardized conditions to ensure that the results are fair and comparable for all tested students.

While summative tests provide useful information for educators about the overall performance and progress of each of their students, and more detailed analyses can provide information about particular strengths and weaknesses, the primary audiences for the results of large-scale assessments are *external* to the school, though their consequences are certainly experienced within the school. This starts with parents, who need to know whether their students have mastered grade-level knowledge and skills (are “proficient”) and have demonstrated academic growth over the previous year, are adequately prepared for the next grade, and are on track to be prepared for postsecondary success when they graduate from high school. Policymakers — including local and state school boards, legislators, chief state school officers and governors — as well as the public and taxpayers, rely on data from annual summative tests to determine, for example:

- Which schools, and which subgroups of students, have demonstrated adequate performance and/or growth this year

- How the performance of schools and districts compare with one another throughout the state
- How many students have demonstrated proficiency this year, and how that compares with last year's performance
- Whether more students are graduating high school each year with the academic skills needed for postsecondary success
- Whether policies or programs adopted by the state contributed to improved performance; for example, are charter schools or schools that accept vouchers outperforming traditional public schools? Have investments in technology or new curriculum contributed to improvements in student performance?

Under both state and federal law, annual summative testing in at least grades 3–8 and once in high school is required and is the responsibility of each state. Those requirements do not appear likely to change even if the Congress enacts legislation to replace No Child Left Behind. Indiana is at a pivotal position with respect to you summative testing program. As a result of the adoption of new Indiana Content Standards for College and Career Readiness in mathematics and English language arts, and the decision to withdraw from PARCC, the Department of Education is now in the process of developing and implementing a new assessment. My understanding is that you will first see preliminary results from these assessments by December. As this new program is being rolled out, some are calling for a different approach to summative assessments; one that relies on commercially-developed interim assessments described above.

As you consider the right direction for Indiana, I'd like to suggest some things you should keep in mind:

Don't use interim tests for accountability purposes. In light of widespread concerns about overtesting, it is tempting to try to use interim or benchmark tests for multiple purposes. Don't. Most commercially-available benchmark assessments were not developed to reflect the Indiana standards, nor the Common Core State Standards. Many are computer-based adaptive tests, well designed to pinpoint with precision the level at which each student can perform. This is extraordinarily useful diagnostic information in the hands of teachers. However, the moment it is also used for high-stakes accountability, its diagnostic value will be diminished, as genuine instruction is too often replaced by test-prep in order to improve the scores. It's like going on a crash diet for a couple of weeks before you go for your annual check-up; you will weigh slightly less when you get on the scale in the doctor's office, but you won't give the doctor an accurate picture of your physical condition.

Insist that any test used for accountability purposes be well aligned with state standards, as determined by an independent third-party review. In practical terms, this means making sure that the test does a good job of carefully measuring the knowledge and skills included in the state standards, as this is what you have

determined is necessary for students to be well prepared for the postsecondary opportunities they wish to pursue. These standards call on students to demonstrate a deep understanding of mathematics and literature, as well as strong reasoning and critical thinking skills. In mathematics, they expect students to be able to both get the right answer, and in some cases, when dealing with real-world problems, to be able to explain their reasoning. In English language arts, they expect students to be able to read and synthesize progressively more complex texts, and draw on evidence in them to make an evidence-based coherent argument in writing. This demands assessments that require students to read one or more texts and write an essay that draws on the information and evidence in what they have read. This is a foundation of college and career readiness — ask any faculty member or employer

You can't measure these kinds of writing and reasoning skills with multiple-choice questions alone. That is why a growing number of states are incorporating performance assessments into their summative tests. These are often multi-step, open-ended tasks that require students to think their way through a problem and explain their approach and reasoning. This is not just a measurement issue. It's an instructional issue. You can't "teach to" tests that require students to explain their reasoning, create mathematical models of real world problems, and make evidence-based arguments. You have to actually help students develop those skills.

Experience has shown that you also should not rely on the assurances of the test publisher alone to determine if the tests actually do measure the full range of skills built into the standards. There is simply too much self-interest involved. The Indiana DOE should instead commission an independent study to provide evidence of alignment. In addition, the U.S. Department of Education released guidance last week requiring such evidence in order to comply with NCLB requirements.

When performance levels ("cut scores") are set for the new ISTEP+, make sure that "proficient" really means prepared to succeed at the next grade level and, ultimately, in postsecondary education or training programs required for middle-skills jobs. This has not been the case in recent years with ISTEP+ program. Several months ago, Achieve looked at the results of state tests in all 50 states, comparing them with the results on the state administration of the National Assessment of Education Progress (NAEP), also known as the Nation's Report Card. NAEP is widely recognized as the "gold standard" for U.S. assessment, with respect to the technical quality of the tests, the rigor of their items, and the rigor of their performance levels. We looked in particular at the results in 4th and 8th grade reading and math, as those are the grades and subjects in which NAEP is administered every other year. More specifically, we compared the percent of students who were deemed "proficient" on the state's test with the percent proficient on NAEP.

As you can see in the table below, in three out of four comparisons, there is a difference in the percent proficient of 40+ percentage points. In the 4th grade math, the difference is slightly less — 31 points. Depending on the grade level and subject,

these gaps are *larger in Indiana than in at least 40 other states*. This means that many Indiana students and their parents have been told that their students are proficient when there is every reason to doubt that they are academically well prepared.

	Percent Proficient on 2013-14 ISTEP	Percent Proficient on 2013 NAEP	Gap
4th Grade Reading	86	38	48
4th Grade Math	83	52	31
8th Grade Reading	76	35	41
8th Grade Math	81	38	43

The first requirement for summative test is that they provide honest and accurate information about the performance of each student. Students and their parents certainly deserve it. So do Indiana’s educators, citizens, and taxpayers. A student who scores “proficient” should have a strong understanding of the grade-level expectations. They should be very likely to succeed at the work of the next grade level and, ultimately, in postsecondary education.

The Indiana Department of Education has the responsibility for overseeing that process with the test publisher. As legislators, you shouldn’t involve yourselves in the details of this process, but you can play an important oversight function. At the end of the day, you should ask two basic questions regarding the performance standards:

- Have we substantially closed the “honesty gap” between Indiana’s definition of proficient and NAEP’s?
- Will postsecondary institutions in Indiana accept the results of the most advanced high school assessment that is administered to all students as an indicator of whether a student can enter and succeed in credit-bearing courses, without having to take an additional placement test when they enroll as a freshman?

Make sure the summative assessments measure both status and growth.

“Status” or “proficiency” indicates whether or not a student has met a single bar. Growth measures the change in a student’s achievement over two or more points in time. Growth does not measure whether a student has met a particular bar. Both are important to know in order to evaluate the performance of individual students, schools, school districts, or the state as a whole. Achieve recommends that states balance student growth and proficiency (or status) when measuring progress. Simply relying on status does not allow schools and districts to demonstrate student achievement growth, which may be substantial, even if it falls short of proficiency. Conversely, simply relying on growth fails to recognize that low-performing

students or schools may be making substantial progress, but still fall short of the mark. Measuring growth well will likely require somewhat more test items than a test that simply measures status in order to reliably assess performance both well above and well below grade level.

Finally, as Indiana develops high-quality and aligned summative assessments, this is the right time to address the growing concerns about the amount of time and attention to devoted to testing each year. Those concerns are real and legitimate.

The best estimates nationally are that annual statewide summative tests take up approximately 1-2% of instructional time each year. Most of the assessments students take are locally determined, in the form of various formative, diagnostic, periodic benchmark and other tests. Many of these tests provide teachers with valuable information about student strengths and weakness that can inform instruction over the course of the year. But it is likely that in many districts, the number of such tests can be reduced, as some are duplicative of others or were adopted long ago and have outlived their usefulness.

I understand that a study is underway to gather information about the amount and types of assessments administered in Indiana's school districts. When those findings are available, I encourage you to work with local school boards, superintendents, and educators in every district to promote a deep analysis of the assessments given in their own districts and schools. Encourage them to make sure that they keep the ones that are providing useful information, eliminate or replace the ones that don't, and explain to parents the purposes and value of those they continue to administer.

Senator Kruse, Representative Behning, members of the Committee, I thank you again for the opportunity to speak with you today. I will be happy to answer any questions you may have.