

## **Testing: Point & Counterpoint**

### On Testing: Point & Counterpoint

#### Point: Ted Lingenheld's Presentation, 1998 SAIS Conference

Editor's Note: Ted Lingenheld's consulting firm, Independent School Counsel, based in Atlanta, is the marketing and consulting partner of Educational Value-Added Assessment Services, Inc., VAAS, led by Dr. William Sanders, professor of statistics at the University of Tennessee in Knoxville.

In talking with independent school heads, academic deans, curriculum coordinators, testing coordinators, division heads, and teachers in the last 12 months about value-added assessment, I have seen first-hand that most schools are at best clueless, and at worst hypocritical in how they treat standardized testing information. In ways I will illustrate, school administrators appear to have one philosophy about testing when it comes to promoting their institution and still another when talking about evaluating curriculum and teaching.

I am not on a crusade for emphasizing objective measurements beyond their essential relevance to committed educators. Diverse methods of assessing student growth must be components of every school's measurement repertoire. "Authentic" and other non-objective assessments are critical to understanding how young people develop skills and knowledge. However, I think it a foolish notion that even the most respected schools believe they can exploit assessments for self-promotion, while virtually ignoring their potential value in evaluating the academic program. Furthermore, it is naïve to think that today's parents won't eventually decode these absurd policies and practices that ignore or refute the useful applications of student assessment data. These customs work to the detriment of the school, and ultimately to its students.

The philosophical boundaries of heads' perspectives on assessment are illustrated by these two comments in my recent correspondence with school heads. The first, a head of a school to which I had presented information about the VAAS wrote,

"...an emphasis on this type of assessment either to monitor faculty teaching in any way, discover "holes" in the curriculum or track students runs counter to our recent goals established by the Curriculum Review Committee. The goals set for in the report from this committee refer specifically to forms of assessment other than standardized testing."

The second, from the head of a leading Midwestern school, was uttered only this week in a phone conversation. He said,

"The refusal of so many schools to use data for constant school improvement is irresponsible."

Having thus staked out the philosophical poles on this topic, let me illustrate my point by asking in the manner of Andy Rooney – "Did you ever wonder why schools have such a disingenuous attitude toward objective measurements of student progress? Well, I did."

How is it on the one hand that schools require standardized testing or developmental benchmarks for admission screenings at all grade levels, but on the other often complain that colleges overemphasize SAT/ACT scores in their admission process? How many times have you heard your pre-school director or teacher say, "We can't take little Mikey for PK-4 because his Gesell results show that he is only 3.319 years developmentally in sandbox skills!" Or the admissions director explaining to a parent of an Upper School applicant "Sarah is only in the 23rd percentile in reading comprehension. We are concerned about her ability to handle the academic load at Impressive Country Day." Testing counts at all levels through graduate school. If it weren't useful as a component of screening and prediction, schools and colleges would not use it.

How can school leaders point to their Advanced Placement or IB program as the crown jewel of their curriculum and then treat as blasphemy any suggestion that teachers in lower grades "teach to a test!" Isn't it obvious that AP/IB teachers "teach to a test?" Isn't that one reason the colleges have more confidence in AP/IB results than in a school's own "boutique" of honors courses?

What justification is there for publishing the National Merit list on every billboard in town except the Goodyear Blimp, as putative evidence of a stimulating academic program, and then hesitatingly, reluctantly, anxiously or not at all sharing with parents the results of the annual achievement testing program? I have heard teachers say, "We don't believe in giving test results to parents because they won't understand it" (as if many teachers themselves have been trained to interpret the results). Yet we rush to publicize the outcome of a three-hour PSAT to people we don't even know hoping to convince them that our school is worth taking out an equity loan to pay the tuition!

How do schools justify the expense in lost teaching time and money for the annual administration of the (\_\_\_\_) Test when so little effort is expended to learn from the results of the test and apply that learning to improve instruction and curriculum? In presenting information about the VAAS at the recent conference of the Florida Council of Independent Schools, when I asked the audience to "grade" their schools on how effectively they utilized their testing program results, most indicated a grade of "C" or lower. I think even that mediocre profile reflected grade inflation. I heard an ERB representative in Atlanta last year refer to this as the "Smile and File" approach to test result analysis.

Why do most colleges invest significant time and effort developing statistical models to predict incoming student retention and performance, and yet elementary/secondary school people often treat standardized test results as if they carry infectious disease?

Do accrediting agencies like SACS really believe that requiring schools, as a condition of accreditation, to administer a nationally standardized, norm referenced measurement of academic progress every year has some relevance to how schools should assess their effectiveness, or are they merely sadists? How many schools would actually administer standardized testing if it was not required? Some schools give it in alternating years, rather than annually. They might as well save their money and time and not bother, for all the value this approach yields.

Finally, since when does the word "accountable" translate to "they don't trust my professional judgment," or is this true only in schools when discussing student assessment outcomes as potential indications of teacher effectiveness?

Quoting the one and only Gilbert Gottfried, "Am I missing something here?"

How would the following situation be allowed to continue? For two, three, or more consecutive years no student taking an AP course with the same teacher ever scored higher

than a 2? Or a 3? How long would a responsible department or division head accept these results before sensing a problem? Wouldn't they conclude that the teacher is primarily responsible for these recurring indicators of modest student learning in the AP course? Wouldn't they take appropriate action to protect the interests of the students, even if parents had not already stood in line to complain?

Now the same school that has this AP problem is faced with the challenge of selecting a standardized assessment instrument for the lower/middle grades. It finally decides to use \_\_\_\_ Test after weeks or even months of thorough teacher and departmental analysis, comparisons, and review. In making this choice, are not experienced educators concluding that this particular assessment instrument will most accurately measure student progress toward the school's published curriculum objectives?

If we accept this last premise, that the instrument chosen comes closest to matching the school's curriculum goals, then assume that the students of these same teachers will be taking this carefully selected test at the end of each year from second to eighth or ninth grade. BUT THEN WHY isn't student performance on these standardized tests (see the AP example above) over time a reflection of the teacher's effectiveness in having their students reach curriculum objectives using the chosen yardstick? The test may not touch every topic of the year's curriculum, but it presumably includes the basic and core elements of skills and content in the various sub-tests.

A footnote: Dr. Sanders' research on standardized assessments in Tennessee and elsewhere has yielded convincing evidence of the teacher's primacy in K-8 student assessment outcomes, just as with the AP example above.

Independent schools have been the model for significant reforms in public schools, witness the exploding charter school movement - public schools that are being run like private schools. Yet in the area of studying standardized measures to study how schools "add value" to students, the public schools are now setting the agenda for private schools. That is not to imply that all states are using their testing results in ways that illuminate this issue - many, including my own NC, publish annual achievement testing results as illustrations of school "quality" and allow the media to "rank" the "best" schools on the basis of average test scores. How ridiculous! Such approaches are better indicators of gene pools, economic strata, and/or neighborhood than school effectiveness.

The two key words in public education today are "standards" and "accountability." When you really get down to it, what are the benchmarks/standards of your schools? Are they internal or external? How do you measure whether your students are keeping pace with those benchmarks, or if your school is adding more or less "value" to students at all achievement levels relative to your reference points?

The search for benchmarks for student progress explains why the work of Bill Sanders and others on "value-added assessment" is attracting significant national attention (Education Week, USA Today, Washington Post in recent months). These researchers believe that the truest measure of a school's academic influences on its students can be detected by looking longitudinally at its scale score gain patterns, not its average scores. That is, in by using a statistically proven approach to study student gains from year to year on the accreditation-

mandated, but individual-school-chosen tests of academic achievement, the factors that have influenced group and individual outcomes can be identified.

Their research in Tennessee and now other states, using a unique, sophisticated statistical methodology, has shown irrefutably that student achievement is most highly influenced by the teacher – not by race, economic level, prior academic achievement record, urban/rural setting, etc. Dr. Sanders found that "high effectiveness" teachers were 85% likely to produce high student gains the next year regardless of the background, income, or ethnicity of their students. Conversely, "low effectiveness" teachers produced lower test gains or negative student progress, likewise on a predictable, consistent basis.

The VAAS was first offered to independent schools in the spring of 1998. A Pilot Study was done with five schools – Lovett, BGA, St. Andrew's in Jacksonville, McCallie, and Park Tudor in Indianapolis. Each submitted its standardized testing data for the past three years or more (Four use the ERB, and one the Stanford), and three of them also submitted high school assessment data – PSAT/SAT/ACT, courses taken, grades received. EVAAS presented each school with a comprehensive report that traced student gain patterns and identified factors (including teacher influences in two schools) that appeared to have affect student achievement indicators. Many leading independent schools are now reviewing their policies towards, and use of standardized assessment information. Other schools will join the VAAS program during this school year.

The use of the VAAS or any analytical methodology adds an empirical element to the traditional methods of assessing student progress by measures unique to each individual teacher; that is to say, based largely on experience and anecdotal evidence. Only by analyzing standardized assessment data as an essential component of measuring total student growth can schools understand how curriculum, course placement, and teaching quality and methods influence advances in student academic achievement. With that insight, even the finest schools can add a margin of improvement in meeting the needs of all their students.

Counterpoint: Opinion from the ISED-L Listserv by Jamie Neilson, Shipley School

I have little faith in the capacity of standardized tests to measure much of anything useful. Many able students score well on them, but I've seen plenty of others who don't. Some studies have shown that these tests predict little beyond success in the first year of college, and if that's true, then their function would seem to be pretty narrow.

But I also see the effect that test results can have on students. The whole discourse around these scores at college preparatory schools the likes of mine reminds me of the Puritans and their obsession with the Elect. The handful of colleges and universities that have chosen to do without the scores in their admissions process seem to be doing fine, so it's possible that they serve no important purpose at all--apart from providing the top echelon of colleges and universities with safe admissions bets and a tool for obtaining high ratings on such travesties as the US News & World Report annual rankings.

In general I see a number of ill effects from the SAT II and its standardized ilk (such as the AP) for independent schools:

- 1.) They provide a blow to a great number of able young people who deserve better at a time when they can ill afford it. These unfortunates then spend inordinate amounts of time on prep programs and general desperation trying to bring up their scores.
- 2.) They ossify the "disciplines" in such a way that the traditional means of organizing knowledge become unquestioned and unassailable. Interdisciplinary options are thereby relegated to secondary or elective status and generally accorded considerably less respect.
- 3.) They influence otherwise sound institutions to farm out their standards for what constitutes an honors curriculum. Instead of establishing an independent, clearly articulated institutional pattern for honors curricula, these institutions either separate the "smarter" students from the others and drive them as hard as they'll go, or alternatively, they slavishly teach to the AP.

I will take issue first with Ted Lingenheld's apparent characterization of my post (among others) as containing "hints" of "a lamentably strong anti-testing bias." I had hoped that my views would be read as nothing so gentle as this. As I mentioned, I was feeling cantankerous; I intended not to hint at a bias, but to state one overtly.

But notwithstanding my biases concerning standardized testing, I do think it's important for those of us who are skeptical of such assessment to hear and be heard by those who have a fuller faith in its usefulness. I would gently suggest that the assertion that "most school administrators and teachers have a shallow understanding at best of what can be learned from studying student assessment outcomes over time" is unnecessarily condescending and provocative. There may indeed be deficiencies in the way teachers and administrators understand and use standardized testing. But it is in the nature of teaching that direct personal experience (inherently anecdotal, sometimes, alas, irrelevant) tends to loom larger than the theoretical spreads, schemata and paradigms developed by professors of education and consultants.

And it is certainly true in retrospect that claims for early standardized assessments made by the testing establishment, going back at least to the emergence of what would one day be the Educational Testing Service in the 1940's, were grander than experience would justify. (See Nicholas Lehmann's very fine two-part series in the Atlantic Monthly for more on this.) Theories of multiple intelligences that are now gaining currency call into question the usefulness and predictive validity of these tests. So given where we are with education and assessment in these times, it makes sense at least to be skeptical of "the man behind the curtain," the great and powerful Wizard of Educational Testing Service (ETS), the purveyors of the SATs.

To address a few of the points that Mr. Lingenheld makes...

The requirement that schools administer and review standardized testing as part of their accreditation process is, I should think, a good argument for ensuring that these tests measure and predict what they claim to do. If the results of the tests are at variance with what line teachers and administrators see in the classroom, I would think that education professionals

would want to know this. Moreover, I would expect to see more interest on the part of organizations like ETS, as well as on the part of individual educational theorists, in measuring interdisciplinary competency and achievement (to use one example). Just because I'm not aware of it, doesn't mean it isn't happening. But the complications for measuring and predicting this sort of thing may simply be too many for the appropriate bodies or individuals to contemplate. If so, where does that leave us in terms of real curricular innovation?

Mr. Lingenheld notes that the AP curriculum is similar in its one-test-fits-all aspect to the English system. This is so, but I wonder what we're to make of the point. Do we wish to be more like the English system? I don't know what this fact means for our own system. As far as teaching to the Advanced Placement tests is concerned, I've made my views clear on that above, but I would add one additional set of concerns. For well-trained students of middling ability or motivation, AP's can boil down to a crapshoot, the odds for which are better if one has the money to spend on the testing. Many independent schools have curricular standards in the upper levels that are inherently AP. Since the tests are not required (but are desirable when one's scores are 4's and 5's) in the college admissions process, it's worth the while of those who can afford to do so to take any test on which they might possibly score a 4. (Human nature being what it is, this "possibly" is elastic, to say the least.) Long shots among this set are reasonably common because bad scores can simply be withheld. I would have a lot more faith in the AP test if the results were withheld entirely by the ETS until a student's date of graduation or, alternatively, if colleges and universities simply refused to look at them until after admission. Why is neither option seriously considered? One reason may be that to adopt either policy would curtail the marketability of the AP. Simply put, far fewer students would be interested in taking it.

Mr. Lingenheld indicates by the use of scare-quotes around certain phrases ("teach to a test," "selling one's soul to the devil," "Evil Empire") that there is a lot of hyperbole surrounding this issue of standardized tests, most of it not useful. I think here we can agree. However, I think he may be overlooking the fundamental reality of teaching--that it is an uneasy mix of art and science, blending people and policies in ways that are frequently uncomfortable and--when things are out of balance--destructive of one in the name of the other. I'm suggesting that some of us in the teaching life and culture see just such a disequilibrium in the standardized testing process as it is used in this country. I may be wrong, but I don't think I'm the only person whose experience has generated these impressions. I am still an agnostic on the question of whether standardized tests can ever be brought into line with their stated purposes and capacities. But with each monolithic defense of these tests, I come closer to giving up on them for good and all. And it is very hard indeed to coolly defend their predictive virtues and power to measure to some of the hardworking (and in some cases, high-achieving) young people who are insulted and demoralized by their own scores, year after year.

-----  
Editor: Patrick F. Bassett, 4/22/99. Published as insert in Spring '99 ISACS Update newsletter.