

The Optimal Reference Guide:

Why Eva Baker Doesn't Seem to Understand Accountability **The Politimetrics of Accountability**

Extraordinary insight™ into today's education information topics

By Glynn D. Ligon, Ph.D, ESP Solutions Group



ESP Solutions Group

Table of Contents

Origins of Politometrics.....	5
Growth Models	7
Contrasting Accountability and Formative Assessments	8
A Formative Assessment System	9
Standards-Based Report Cards.....	10
School Improvement System.....	11
The NAEP Identity.....	12
Help from the Federal Government	13
NAEP Lessons	14
Classroom Data Needs – The Real Formative Data	15
Attachment A – An Open Letter to Eva Baker.....	18
Attachment B – Case Study of the Formative Data Needs of a Teacher.....	21
An Eighth-Grade Mathematics Teacher’s First Day of School.....	21



Formative assessment has been under-valued and under-funded in education. Formative assessment is what really helps teachers focus their instruction on students' immediate needs.

A major complaint educators have about statewide accountability assessments is that they make poor formative assessments for teachers. True, but the real problem is we can't seem to let accountability assessments simply do their job without faulting them for not being formative assessments as well. We should all be demanding separate assessments—one designed to be an excellent accountability measure, and many designed to be excellent formative assessments. But no, educators who disagree with the money and time invested in accountability measures have lobbied politicians to stretch the use of those assessments beyond the capability of a well-designed accountability test.

There is plenty of money to have two separate assessment programs—one to rate schools and one to diagnose and prescribe instruction. Plenty of money if we automate test administration, scoring, and reporting. Plenty of money if we apply extreme security and confidentiality standards only to the accountability assessments, not to the formative assessments.

po-lit-i-met-rics \p•-li-t•-me-triks\ *n pl but sing or pl in constr* (ca. 1972)


1 : the quantitative study of political groups, institutions, nations, and international systems **2** : statistics and indicators that are determined by a combination of scientific, mathematical, and political processes **3** : the art or science of determining high-stakes measures and criteria for accountability, esp. in the field of education

What do the words decimated, income tax rate, and proficiency level have in common? These are all terms derived through a combination of political and psychometric decision making. Politometrics are used to determine each.

Decimated refers to drawing lots to select one in ten soldiers to be executed. While the measurement of one in ten is rather precise, the setting of 1/10th as the cut point was rather political—enough to make a point, but not too many to wipe out a useful unit. The income tax rate is set mathematically to generate a target revenue, but the rate is also politically determined by a vote of Congress and a signature from the President—to curry favor or avoid retaliation by the voters. The determination of proficiency levels on an assessment is informed by a projection of how many students will perform within each level, but ultimately a political body adopts the official cut scores.

Separating psychometrics, accountability, and annual objectives for adequate yearly progress from the political context within which education lives is impossible. Some significant politometrics of our time are:

- 100% of students proficient by 2014
- The National Assessment of Educational Progress' (NAEP) standard for being proficient rather than basic
- Criterion scores for eligibility for Title 1 services

 **ESP Insight**
Formative Assessment:
*Tests' results used to
diagnose and prescribe
instruction for students.*

- Formula for calculating a dropout rate
- Average daily attendance (rules for excused absences, tardies, etc.)
- Persistently dangerous school
- Highly-qualified teacher
- Percent of students by race/ethnicity
- Age requirement to enter kindergarten
- Percent expenditures on instruction
- Income guidelines for National School Lunch Program eligibility


The governing body exercising its politimetric responsibilities may be a local school board, the Office for Civil Rights, a state legislature, a school parent advisory committee, a state school board, or Congress. The result is that the comparability, validity, and reliability of our education statistics are susceptible to politics. Many of us have worked hard to raise the level of data quality within the education statistics arena. However, a major component of quality is definition—especially setting a standard and the process for measuring that standard. Policy and politics play a significant role in data quality and our perception of data quality in education’s metrics.

Let’s revisit the list of education politimetrics and rate each by our level of confidence in them.

Figure 1: Confidence Levels in Education’s Politimetrics

Politimetric	Perceived Confidence by Educators
100% of students proficient by 2014	0%
Formula for calculating a dropout rate	5%
Persistently dangerous school	20%
Percent expenditures on instruction	50%
Average daily attendance (rules for excused absences, tardies, etc.)	60%
Percent of students by race/ethnicity	65%
Highly-qualified teacher	70%
Criterion scores for eligibility for Title 1 services	75%
Income guidelines for National School Lunch Program eligibility	80%
NAEP’s standard for being proficient rather than basic	90%
Age requirement to enter kindergarten	95%

I made up these ratings for the sake of discussion. Agree or disagree with these ratings, the fact is, politimetrics are simpler for some measures and certainly some

 **ESP Insight**
Politimetrics didn't work well when Congress decided 100% of students must be proficient by 2014.

politimetric decisions have more face validity than others. One could argue that anything rated above 50% in Figure 1 may be over-rated.

Origins of Politimetrics

At one time, a brief time, I thought I had created the term politimetrics. However, credit goes to Thomas Gurr, *Politimetrics: an introduction to quantitative macropolitics*, *Prentice-Hall*, 1972. He thought of politimetrics more as statistics about political entities. My notion of politimetrics is more as the artful combination of psychometrics and policy—how we arrive at tolerable criteria for accountability.

In this paper, we'll expand the term even more to encompass the whole arena of assessments, the standards they measure, the rigor they impose, and the uses to which the scores are applied—appropriately and inappropriately.

In the process, I will challenge the establishment and those people who have become established in the assessment and accountability world. So why not start with someone who has become one of the most respected authorities in the assessment and evaluation field for education. Dr. Eva Baker, past President of the American Educational Research Association and Director of the Center for Research on Evaluation, Standards, and Student Testing (CRESST). So significant is she that her email address is simply eva@ucla.org.

Eva Baker doesn't seem to understand accountability because of politimetrics. Politimetrics has worked to influence her phrasing if not her thinking. Being a leader of CRESST and President of AERA, she must be ever vigilant to the politics of psychometrics as presented by the pending decisions of Congress, the rambling priorities of a major education research association, and the leanings of her in-crowd. She does not have the luxury of seeing accountability from a simple, clear perspective. She appears to be obligated to couch every thought she issues publicly inside a complete context of political propriety (aka correctness).

This paper sets out to do one simple thing—separate accountability from all of the politically proper, politically expedient, politically encrusted context that assessments and accountability have accumulated around them. This paper intends to call for accountability assessments to be singular in design, purpose, and use. When we say accountability assessment, we should think of only one thing—a standardized test that provides a measure solely for the evaluation of student learning progress. In the process, a compelling case will be presented in support of true formative data and assessments.

As I listened to Dr. Baker's Presidential Address at the 2007 AERA Annual Meeting, I made notes on the five "accountability fixes" she proposed. I found it interesting to read her "expanded version" entitled "The End(s) of Testing" available on CRESST's website. The fixes are now six, and they are called "mitigations" and "tactics." The tone of the expanded version is much less critical of accountability than was the address. Attachment A is an open letter to Dr. Baker written after the address. Do I really think Eva Baker doesn't understand accountability? No. That's why the title



Politimetrics for education is more the artful combination of psychometrics and policy—how we arrive at tolerable criteria for accountability.



Education needs to separate accountability from all of the politically proper, politically expedient, politically encrusted context that assessments and accountability have accumulated around them.



cleverly says “Why Eva Baker Doesn’t Seem to Understand Accountability.” She understands accountability, but like so many others in education, she uses the term rather expansively. She’s managing the politimetrics. She’s trying to squeeze formative dollars out of the accountability budget. I chose her so this letter could argue the opinions and conclusions rather than the agreed-upon facts or Dr. Baker’s qualifications. I also know she has a bigger platform if she chooses to respond.

When I attended the AERA Annual Meeting (not a convention as less collaborative associations call their annual gatherings) in Chicago in 2007, I was struck by how correct that association is. Just read the list in the program of almost 200 officially recognized divisions and special interest groups (SIGs, of which I have been a member of several over the years). NCLB was the piñata hanging from every meeting room chandelier. Unfortunately, few AERA members seemed to expect any candy out of it. I recall from past annual meetings when NCLB was new that the majority of educators thought NCLB was already broken from the start. I was simply overwhelmed by the “glass half empty” conclusions that predominated the professional papers and presentations. Bias? Yes. Because many educators have not yet separated accountability from school improvement, or differentiated accountability assessments from formative assessments. Even NCLB mandates they be reported back to teachers in a useful manner.

AERA members, in general, seem to want the NCLB assessments to be formative. Accountability appears to have joined norms, grade equivalents, multiple-choice tests, and randomized trials as targets of the researchers who focus on formative, constructive, narrative, performance measures, and ethnographic approaches. Somewhere, behind the scenes must be the researchers who appreciate the efficacy of all of these approaches when each is properly applied.

The main point of the open letter is simply to say that accountability assessment is different from formative assessment. They have different purposes, different psychometric requirements, and different policy foundations.

The distinction between politimetrics and political correctness is important. A politimetric decision may or may not be politically correct. The decision is politimetric because it is a compromise between the statistical or psychometric factors and the political ones. In this context, political means policy in general more than government in particular. Politimetric decisions can be correct without being politically correct.

In 2000, when I was consulting with the Governor’s Office in Colorado during the creation of their school accountability reports, the question arose of where to set the dividing lines for CSAP (Colorado’s statewide assessment) performance between school ratings. Initially, Governor Owens wanted to use A, B, C, D, and F, but eventually agreed to descriptions (excellent, high, average, low, and unsatisfactory). As the Governor’s policy advisors and members of the Legislature debated the relative merits of various methodologies, I asked “How many schools can Colorado tolerate being unsatisfactory? How many schools will the public accept as being excellent? The answers were 8% excellent, 25% high, 40% average, 25% low, and 2% unsatisfactory. With the policy determined, the psychometrics, statistics, and mathematics of establishing the rules were straightforward. Yes, in year one the cut points were arbitrary (based upon actual performance of all schools). Critics

complained that the system was normative, a pejorative word used to discredit accountability systems deemed as dooming a set percentage of schools to failure.

The reality is that after year one, any number of schools could be rated excellent or unsatisfactory as they changed their performance. That accountability system has been in use for seven years without substantive modifications. A major reason for its persistence is the face validity of the published schools' ratings. The creative blending of psychometrics, statistics, and policy resulted in an accountability system that worked. The new Governor and Legislature will have their opportunity to apply their own politimetrics to the next generation of accountability rules.

Is the infusion of politics into accountability anathema to valid ratings? Not at all. In fact, without the balance of policy makers in the design, accountability systems would be inflexible, statistical theories. D3M (data-driven decision making) isn't just about the numbers. When policy decisions are made, the facts are balanced with the politics. National politics may be over-sensitive to the political dynamics with all the polling that goes on before Congress or presidents and candidates claim their policy ground.

Growth Models

Growth models represent a new generation of politimetrics in education. The idea is very simple—recognize schools making gains on assessments. The implementation has become very complex. Hierarchical linear models (HLM), which few educators understand and most statisticians I have met trust too blindly, are being touted as the most sophisticated way to tease out gains. With the error measurement of assessments, the mobility of students, the small cell sizes for subgroups, and the resistance of student performance to rapid/sustainable improvement, HLM frequently splits hairs as it combs through data to find statistically significant differences that translate into tiny practical advantages.

The preceding statement was a blatant generalization that does not recognize the existence of clear academic gains within effective schools. The admonition in the statement is for us not to get our hopes too high for what growth models will show. Many schools full of low-performing students are really ineffective with academically disadvantaged students. Many schools full of high-performing students are restricted in how high they can perform because of assessment ceilings. Many—not all. The pursuit of those exceptions is both noble and necessary. Even for those low-performing schools that achieve miracles with their students, the ultimate goal doesn't change. Maybe they still need assistance to reach that goal. The prime objective of No Child Left Behind and most state accountability systems is to establish a goal line that is the same for all students regardless of how unlevel their playing fields are.

When considering growth models, the measure of growth must be as objective, numerical, reliable, valid, and comparable as possible. Again, we get back to needing a true accountability measure for the task. Formative assessments have a single shortcoming related to growth. They are most useful when they focus on a




A major reason for the persistence of Colorado's school accountability report cards is the face validity of the published schools' ratings.



HLM frequently splits hairs as it combs through data to find statistically significant differences.

limited number of specific skills and objectives about to be taught. This characteristic makes them poor measures across grade levels and school years. An accountability assessment should measure skill and objectives across multiple years to avoid floor and ceiling effects—and to fit the assumptions of emerging growth models.

Contrasting Accountability and Formative Assessments

 **ESP Insight**
Educators always want to squeeze formative data out of the accountability turnip.

Reading through Figure 2 makes one wonder why anyone ever tried to make accountability and formative assessments the same. The reason is simple actually. Educators always want to squeeze every ounce of utility out of their efforts. Extracting formative data out of the accountability turnip is understandable. Unfortunately, accountability assessments are not up to the formative task.

Politemetrics has loosened the focus of accountability assessments by pandering to the proponents of formative assessments. What makes a good formative assessment does not make a good accountability assessment. Does your state’s assessment go on the long list of those that would be more precise measures for accountability if they had not been developed to also provide objective-level proficiency scores for individual students?

About 1990, Darvin Winick, now Chair of the National Assessment Governing Board, led a study group in Texas to recommend the next generation of assessments for accountability. As a member of that group, I recall agreeing with Dr. Winick that a nationally standardized test best fit the requirements for a single measure for evaluating the achievement of Texas students. In addition, we agreed that this test would never satisfy the need for formative, diagnostic data for teachers, so a separate diagnostic test should be developed aligned with Texas’ curriculum standards. Those discussions and the insights about separate measures have remained valid through today. Unremarkably, Texas’ TAKS assessments are another generation of state assessments that try to be both accountability and formative measures at the same time.

Figure 2: Contrasting Formative and Accountability Assessments

Accountability Assessment	Formative Assessment
Representative items from across all knowledge and skills	Selected objectives representing knowledge and skills to be taught now
About 50% items correct by average student provides maximum measurement precision	About 75% correct by a proficient student provides expectation of success on the accountability assessment in the future
More total items on assessment for reliability of the overall proficiency level of the student	More items per individual objective provides confidence in diagnosis of areas in need of instruction

Accountability Assessment	Formative Assessment
High security and confidentiality to protect the integrity of the test items and the results for individual students	No need for security because the whole idea is for teachers to use the items on demand
Scheduled administration times or windows for comparability	On demand administration to coincide with instructional planning
Timely scoring and reporting for decision making	Immediate scoring and reporting for diagnosis and prescription
Fresh items with only a few reused for alignment and equating	Reusable items for next groups of students as long as alignment with standards is maintained; released items from accountability assessment used
Major concern about cheating	No concern about cheating; no incentive for teachers or students to cheat
Content measured is the same for all students	Content measured is what each student needs at the moment
On-line administration supports security and lowers costs	On-line administration supports the on-demand nature of formative assessment and lowers costs
Vertical scaling desired for measurement of growth	Measurement of current status on objectives for diagnosis
Politemetric establishment of cut points for proficiency	Teacher decision of cut points for prescription of interventions

A Formative Assessment System

Testing has some distressing drawbacks.

- Loss of instructional time
- Cost to build, administer, and score
- Complexity of the psychometrics to ensure valid, reliable, aligned, and performance-level appropriate forms
- Anxiety, both student and teacher
- Misinterpretation and misuse of the results
- Security and confidentiality procedures
- Motivation of students to perform their best
- Standardization of administration and scoring to ensure accuracy
- Cheating
- Differences of opinion about all of the above



No need to worry about security and cheating on a formative assessment. That alone saves dollars and time.

Ultimately the discussion of how to balance these issues with the benefits from formative assessments comes around to computer administration and web-based applications. There are dozens of legitimate software applications available today for performing the basic functions required.

- Creating tests from item banks
- Composing a form
- Authenticating and authorizing the student to be tested
- Presenting the form to the student
- Recording the responses (not just multiple choice)
- Deriving scores (e.g., raw, scale, performance levels, percent correct)
- Producing reports
- Linking to curricular resources (e.g., media, lessons)
- Tracking mastery of objectives

Optimistically, the costs of these applications will be driven down by competition and the accumulation of items available for use.

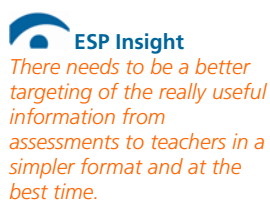
The key to successful delivery of a formative assessment system will be on-demand access by teachers. Teachers need to be capable of assigning an individual student to a formative assessment at the time an instructional decision is to be made. Maybe even better, the system itself will assign the next formative assessment based upon completion of the prior instructional module. Some already do.

Standards-Based Report Cards

Enhanced formative assessments are basic to the movement toward standards-based report cards for parents. In a previous Optimal Reference Guide (Using Assessment Results to Get Performance Results, 2006), Dr. Evangelina Mangino reported that parent reports from accountability assessments are too hard to understand, and that scale scores are meaningless to teachers because they do not provide a context for interpretation. "Teachers do not use (accountability) assessment reports as much as they might because they are overwhelmed by the quantity and complexity of the reports. Training on the vast scope of these reports is not realistic. There needs to be a better targeting of the really useful information to teachers in a simpler format and at the best time."

That is a difficult challenge for accountability assessments. That is the prime objective for formative assessments. In fact one of the final recommendations of Dr. Mangino's study was to support on-line diagnostic testing.

For readers interested in automating the formative assessment process, the paper provides a crosswalk between its recommended D3M Assessment Report Standard and the SIF (Schools Interoperability Framework) standard for exchanging data about assessment results among software applications. With horizontal interoperability, the formative data can flow hands-free and timely among an education agency's student information system, directory management system, data



warehouse (or collection of data stores), special education system, media services application, food services application, transportation/GIS application, finance system, human resources system, instructional management system(s), etc. The D3M Framework and D3M Education Portal represent the architecture that supports this interoperability.

School Improvement System

Accountability and formative evaluation should be components of an overall school improvement system, not systems unto themselves.

The overall school improvement system includes these components as illustrated in Figure 3.

- Standards—academic expectations including specific objectives for student performance
- Formative Data—the full complement of information a teacher uses to diagnose and prescribe interventions for individual students
- Resources—printed and electronic media aligned with the standards
- Teaching-Learning—the activities managed by the teacher to effect learning by the students
- Accountability—the measures of outcomes that evaluate the effectiveness of the teaching-learning activities in meeting the standards

Significantly, formative data, including formative assessments, and accountability, including accountability assessments, are represented as different components, not the same activities with dual purposes.


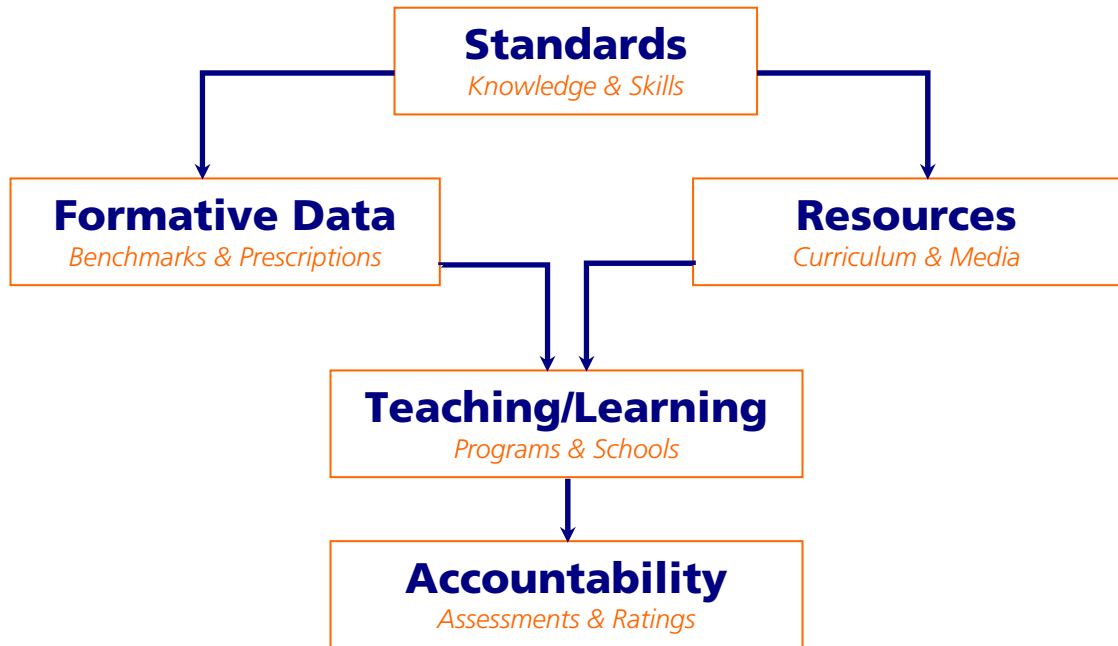
 **ESP Insight**
Accountability and formative evaluation should be components of an overall school improvement system, not systems unto themselves.

Figure 3: The School Improvement System

School Improvement System



The NAEP Identity

ESP Insight
NAEP does not attempt to be useful to teachers, to give diagnostic information on individual students.

NAEP is excellent example of an accountability assessment (evaluation in their terms). NAEP does not attempt to be useful to teachers, to give diagnostic information on individual students. NAEP is relatively unchallenged as the nation’s accountability assessment. No real argument here. However, another person whom I respect in the assessment field has used NAEP as the ultimate benchmark to show that state accountability assessments are too easy. Dr. Robert Linn, University of Colorado at Boulder and another CRESST/AERA luminary published the following information to illustrate his conclusion. See Figure 4.

Figure 4: Achievement Level Differences: Four States’ NAEP vs. State Performance (2003 Reading/Language Arts 4th Grade – percent proficient and above)

	Alabama	Alaska	Arizona	Arkansas
NAEP	22	28	23	28
State	77	74	77	62

Dr. Linn and everyone else I've heard discuss this relationship concluded that the states' tests are too easy. We could also conclude that NAEP is too hard. The latter conclusion is especially compelling if you trust the politimetrics within each state more than the politimetrics at the national level. Think of the conclusion this way. Less than half of the nation's students are proficient. Really? The parents of students in Lake Wobegon would be particularly upset at this statement.

Maybe the problem is one of semantics. Proficient. Basic. Which one should we require of ALL students?

The purpose of bringing NAEP into this discussion is that NAEP is an excellent example of an assessment from which educators are not trying to use individual student scores and objective-level scores to plan the next week's instructional activities. That is how educators should approach their own state's test results. Let the policy makers use the results, but turn to your own formative assessments for what you need.

If this perspective were to be translated to a state's assessment program, not only would there be separate tests for accountability and formative purposes, the nature of both would improve for the better.

Help from the Federal Government

I trust your expectations for a section titled "Help from the Federal Government" are not too high. What follows is an analogy (as illustrated in Figure 5) to the three branches of government and the three uses of data within an education agency.

- Legislative Branch = Needs Assessment (Congress, legislatures, school boards)

What needs to be done? What funds are required? What hurdles need to be removed to allow improvement to be made?

- Executive Branch = Formative Assessment (Local education agency administrations, school leadership teams, teachers in the classroom)

How should improvement be implemented? How do we use resources to do it?

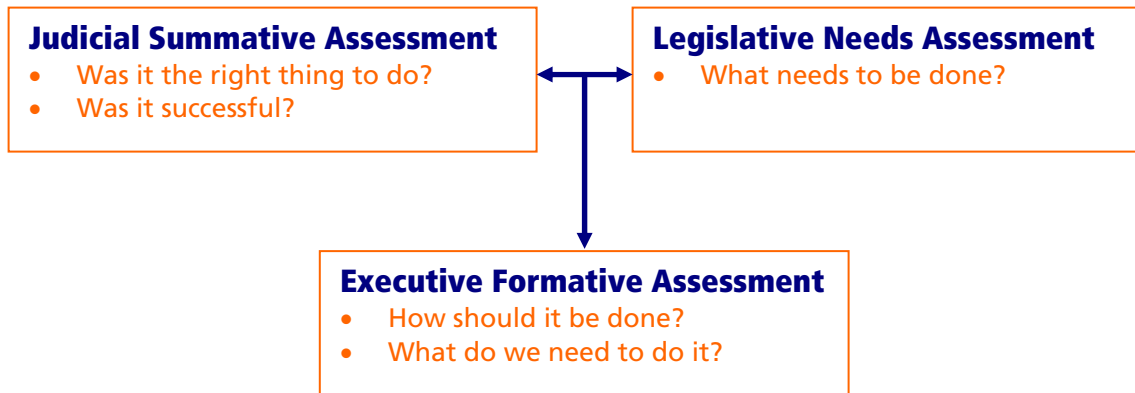
- Judicial Branch = Summative Assessment (Evaluators, the public, parents, stakeholder groups)

Was it the right approach? Was the intervention successful?

Just like in the Federal government, the separation of powers gets blurred at times. The point, however, again, is that we need to separate the functions of accountability from those of formative processes. In this comparison, the weakest alignment may be to the judicial branch. Who really sits at this level for the

schools? The legal courts get involved at times, but the court of public opinion may rule.

Figure 5: Branches of Government



NAEP Lessons

The National Assessment Governing Board in 2005 declared:

The National Assessment of Educational Progress (NAEP) should provide each state with results on the achievement of 12th graders. All states should participate in reading and mathematics assessments in grade 12 as they do now in grades 4 and 8.

Not only are they supporting continued testing, but expanded grade levels. What's significant about this recommendation is that NAEP is remaining true to its charter—evaluating the achievement of the nation's students. Instead of straying into formative assessments, the Governing Board recognizes that 12th grade is the time to hold schools accountable for the full education experience.

In Education Week, May 17, 2006, Kay and Houlihan made this observation about 21st Century Assessments.

A balance of assessments, including high-quality standardized testing along with effective classroom assessments, offers students a powerful way to master the content and skills central to success.

This balance of assessments can best be achieved by having two assessment components to our school improvement system than by continuing to attempt a balance within a single assessment.

Classroom Data Needs – The Real Formative Data

Teachers in classrooms use a different set of data than do other education professionals and the public. Teachers need to know very personal and immediate details about their students. These details are about not only instruction and learning, but also the students' individual lives. Much of this information is not shared with other levels of the education enterprise. Much is anecdotal, momentary, and unrecorded in any formal way. Much would be inappropriate to record permanently or share widely.

Classroom data can be categorized into two major sets:

1. **Leading Indicators:** Information useful for diagnosing, planning, organizing, and delivering the instruction for individual students on a daily basis.

These data are very personal for the student and would not typically be shared beyond the classroom or school.

2. **Trailing Indicators:** Information that documents activities and is reported outside the classroom for compliance and accountability purposes

These data are less confidential, more descriptive and are typically collected for purposes beyond the classroom and school.

Teachers need assistance in gathering, organizing, and using these classroom-level data. Automated systems with standards for defining and describing these data are (or would be) helpful. The efficiency gained from this automation and organization can contribute to maximum time and focus on instruction and learning during classroom time.

Attachment B describes a composite case study that was created for a study of decision support systems for the U.S. Department of Education. The perspective is that of an eighth-grade mathematics teacher. The narrative that follows has been expanded in three tables.


Table 1: Information Needed on the First Day of School

Table 2: Information Needed on a Daily Basis

Table 3: Classroom Records Needed

In each table, the information required is linked to the most likely *data source*. These include:

- Student Information System (software that schedules classes, reports grades, maintains attendance records, and performs various other record keeping functions for schools)
- Student Profile (compilation of information about an individual student into a permanent record folder or a printed report)

 **ESP Insight**
Teachers need leading indicators to plan instruction. Trailing indicators come too late to be useful in the classroom.

- Instructional Management System (software that links a student's performance with instructional activities and resources)
- Assessment DSS Reports (decision support system for assessment records that allows for grouping of students by performance on specific objectives or totals)
- Special Education Records (representative of any special program that diagnoses student needs and creates an individual education plan to monitor progress)
- Personal Student Report (informal information gathered by the teacher about a student)
- District Curriculum (curriculum adopted by the LEA)
- State Standards (objectives adopted by the SEA as a basis for state assessments, accountability, or accreditation)
- School References (student handbook, school directory, and other documents describing school rules and procedures)
- District References (policies and regulations, including state rules)
- Staff Profile (record of a staff member's degrees, preservice and inservice training, annual evaluations, and other performance instruments)

Each table also includes the levels at which the information is most likely to be shared. These levels include:

- Classroom
- School
- District
- State
- USED (U.S. Department of Education or other Federal agency)

Conclusion

The point of this Optimal Reference Guide: Formative is great, but everything doesn't have to be formative to be valuable. Yes, we have struggled forever to capture and provide formative data for teachers. However, being formative is not the sole criterion for valuable data. We have to make our case and seek formative data without diminishing the value of data for other purposes—like accountability. Even more important, we must not try to turn all data into formative data. Some data will not meet our high standard for being formative just like most formative data do not meet the high standard for accountability.

The exasperating problem in front of us is that education has gotten off track. Let me change that. Education has built only one track—formaccountability. Through the misguided application of politometrics, we have tried, and failed, to build a single state assessment system that satisfies the need for both accountability and formative data for teachers. We need two tracks that ensure schools get value from true formative assessments while the public gets real accountability measures.

This calls for our information systems within education agencies to open up. Now is the time to begin including formative, unofficial, data in our data stores. Great progress was required to move education agencies to the point of organizing their data for access by decision makers. More great progress is required to now incorporate those softer, less compliance-mandated data elements that can help schools improve.

Attachment A – An Open Letter to Eva Baker

Dr. Baker,

I enjoyed your 2007 AERA presidential address--except for the part where you suggested "accountability fixes."

The real world, Congress, state legislatures, and the public are serious when they criticize education for being reluctant to be accountable. We must be cautious when suggesting moving from tests to softer, subjective "accountability" measures.

In your keynote address, you laid out five "accountability fixes" for the No Child Left Behind Act. Unfortunately, you could not have been more wrong about what needs to be done for accountability. Simply put, most of the fixes do not belong in accountability. They belong in a school improvement system. The distinction between the two escapes most educators. Of course, the purpose of accountability is to verify that the resources being invested in education are delivering the expected benefits--successful schools. This is very different from telling schools where individual students need to improve.

As I listened to your address, I reacted to each of your proposed accountability fixes.

- Fix 1: More Indicators

We get very confused by having more and more indicators to interpret. The fact is NCLB already mandates multiple indicators rolled into a single adequate yearly progress rating. As long as additional indicators get combined into a single rating rather than present a confusing and conflicting array of separate indicators, this is a great idea. Otherwise, when it comes to indicators, the more the murkier.

- Fix 2: Opportunity to Learn

Opportunity to learn is a process indicator, not an outcome indicator. We will not be satisfied knowing whether or not students were taught, we want to know if they learned. States should definitely monitor opportunity to learn as part of their overall implementation of NCLB.

- Fix 3: Performance Assessment

Have we already forgotten that performance assessments withered as accountability measures because they are too costly, unreliable, and rater-biased to be practical? Beyond limited constructed-response items, writing samples are the signature survivor of performance measures within statewide assessment systems.

- Fix 4: Formative Assessment

Wait a minute. Aren't these accountability recommendations?

- Fix 5: Prioritized Standards

Great idea—for formative assessments like James Popham advocates. However, we should be expanding the scope of the content of our accountability assessments.

The message in your fixes is that accountability should be more like formative evaluation. You continued with a call for an accountability system that leads to instructional decisions. You criticized current accountability systems as having an absence of feedback for teachers. This is formative not accountability assessment. This just isn't a reasonable expectation for an accountability system. We should define accountability appropriately and narrowly. We must accept the expense and burden of accountability. We can then construct accountability assessments that measure a broad range of knowledge and skills. We can have more affordable, shorter forms that are tightly aligned with the full core academic standards.

In the process of creating a truer accountability solution, we should keep a focus on the need for formative data.

First and foremost, we need two different assessment programs—one for accountability and one for formative decisions. Your fixes perpetuate the same mistake NCLB codified in 2001. They call for accountability results to be useful to teachers. This is not likely to happen and sets up accountability assessments to disappoint teachers. I'm also critical of the policy makers for buying into the notion that if the tests are not useful for teachers to plan instruction, then they are failures.

I see that we are trying to satisfy everyone with a single assessment and accountability system. What we need is to satisfy the accountability requirement. Then we need to have a separate, differently designed and crafted system for formative evaluation.

To say the NCLB's accountability can be fixed by making it into formative evaluation is just wrong. Accountability can be fixed by separating it cleanly from the formative evaluation process. Then we can set about to build the infrastructure and processes to do formative evaluation and assessment right. The scope of NCLB is far beyond accountability. Formative goals fit, but formative goals are not accountability fixes.

Accountability assessments are like stock prices for a corporation. There is an incredible array of components that can be analyzed to discover what went right or wrong with a corporation and how to improve, but the accountability function is not tasked with that diagnosis and prescription. The stock price is not very helpful to management and workers to design improvements, but it is the essential way to value the worth of a company. Shareholders are not satisfied to know that there was an "opportunity to earn," or that performance evaluations were high for all employees, or that the corporation focused on a smaller set of standards for the year. They want a higher stock price.

Want a sports analogy instead of a business one? A professional baseball team is ultimately judged by its won/lost record or by championships won. A .333 won/lost record doesn't tell anyone what needs to be done to improve, but it is a clear accountability measure. Separately, management (or the fans) must analyze RBIs, ERAs, LOBs, BAs, and HRs. If you have no idea what those are, that's fine, because you know .333 is bad. Let the fans argue the statistics and management rebuild the team.

OK, so here's the education example. Parents see that their school has a 33.3% proficiency rate. Bad. They won't be satisfied knowing that their children had an opportunity to learn 90% of the standards, or students averaged over 85% on formative assessments, or the teachers reported student performances to be acceptable on report cards. Parents know there's something wrong. Policy makers know that the school must improve. That's accountability.

What to do is the next step after accountability. If you want to roll all assessment together into a complex system of "formaccountability," that is wrong. Instead, we need to separate them even more. Formative assessment—accountability assessment. Two different types of tests.

How can that be simpler?

I am ready to support an increase in formative information for teachers. Doing that requires information systems changes far beyond over-analyzing accountability test results.

Sincerely,

Glynn D. Ligon, Ph.D

Attachment B – Case Study of the Formative Data Needs of a Teacher

An Eighth-Grade Mathematics Teacher's First Day of School

On the way to school this morning, I was thinking about those 150 strangers whom I'll be meeting in a few minutes. I could have gone down to the office and leafed my way through 150 permanent record folders to try to get to know them. Instead I decided to wait for the class reports that will find their way to me over the next few weeks—or months. I'll probably get a chart of the students by "State Assessment" objective mastery from last year's testing. For the special education students, I should get notice of the accommodations each needs. Most other reports will just list who's in each period with their student number and gender. Those lists will be out-of-date before they get to me, so the office, the students, and I will pass paper notes around for changes and late entries. Our middle school does have an automated scheduling, attendance, and grade reporting system; and I do have a computer right there in my classroom; however, I still send notes to the office for absences, and my grades go in on bubble sheets.

So, just out of curiosity, I thought I'd list out what information I could really use on the first day of school.

Here's the list.

1. What is my own class schedule and which students are assigned to each class period?
 - Student last name, first name, middle name, generation code
 - Student identifier (local, state)
 - Grade level
 - School year, term, starting/ending dates
 - School name, district name
 - Course name, course number, credits
 - Room location
 - Period, start/end times
2. What language do the parents speak? (When I need to call a student's home, knowing ahead of time if they can speak English would be very helpful. By the way, this will change for some students during the year as they move from one relative's house to another, or parents come and go from their home.)
 - Preferred home language spoken by each parent and household member
 - Preferred language for written communications
 - Student dominant language

3. Who are the parents/guardians and what are their relationships to the student? (When I call home, I'd like to know whom I'm talking to and be sure they are responsible for the student.)
4. Which students mastered each of the "State Assessment" objectives on the last administration? (This initial chart would help me plan for where my students are starting. I could see the range of achievement levels. I would be able to plan the administration of my own diagnostic tests to group for instruction.)
5. What special accommodations do students with an IEP require? (Last year I found out about accommodations after the first report cards went out. I would regret failing a student because I didn't know about special considerations or modifications I should have been making.)
6. What final grades did each student make in math from kindergarten through grade 7? (Some students will be consistent, others will be going up or down, and others may have had one really bad year along the way. If I know how they have been doing, I can look for gaps that might need to be addressed. Even though report card grades are notoriously subjective, I can interpret them to some extent based upon which teacher each student had.)
7. What final grades did each student make last year in every course? (Math success requires students to know how to read and have a broad vocabulary. If a student is struggling to read, I'll need to know that separately from their specific math skills.)
8. What crucial events have occurred in each student's life? (Last year, a student responded to an exercise to write three facts I needed to know about him by stating that he had no idea who his father is, his mother is in prison, and he has an arrest record already. I needed to be aware of these when sending notes home or following up on missed assignments.)
9. What honors and awards has each student earned?

So that gets the year started. On a daily basis, I need other types of information to help me plan and deliver instruction.

Here's that list:

1. Important events happening in each student's life on a daily basis (On Algebra "State Assessment" testing day last year, a girl came to school and worried that she might not pass because her mother had thrown her out of the house the night before and she had spent the night on the streets. I let her sleep in the back of the room for an hour before starting the test. Because the test is untimed, she was able to finish and pass. Good events happen too, but they are easier to discover.)

2. What my students do from 4 – 6 each day (which ones play sports, take lessons, care for younger siblings, work, watch TV, roam the streets, or study?)
3. Logs and lists of standard school activities beyond my classroom, like attendance in other classes, major discipline actions, events that excuse a student from my class, excuses that make a student eligible to get a grade on make-up work, etc.
4. Extracurricular activities that students participate in and require them to have passing grades.
5. The district's required curriculum for each course I'm teaching.
6. The "State Assessment" objectives for the tests they will take this year—or next.
7. The curriculum resources that have been aligned with both the curriculum and the "State Assessment" objectives.
8. My prior lesson plans and any others that I might refer to when planning this year.
9. A set of assessments I can administer as needed to know which students have mastered what was taught in class and which students need more help.
10. Classroom management resources I can refer to if student discipline or motivation get to be issues.
11. The student handbook and all other materials handed out to students and parents describing the school's expectations for them.
12. The district policy book for those times when I might not be sure what I can and can't do.
13. Some way to locate a master teacher who has been successful teaching a skill my students are having trouble mastering.
14. My past evaluations, a self-assessment, and other resources that help me choose training to attend. Also, a list and schedule for required training everyone must attend.
15. A list of support services available from the school, district, and community for students who encounter problems.
16. Logs and charts showing which students have completed assignments; a grade book of assignments completed and grades awarded; some way to

summarize these grades so I can monitor how each class is doing compared to past classes.

17. A list of phone numbers and names for all the offices I'll have to contact.
18. Honors and awards.
19. A list of names and numbers for each student for contacting their parents and guardians.
20. The school year calendar with major events shown so I can plan assignments to avoid conflicts.
21. My own class schedule and student rosters whenever a change occurs.

Then as my classes progress, I need to have classroom records on...

1. Performance on classroom diagnostic tests tied to the curriculum
2. Performance on daily work, assignments, and projects
3. Performance on classroom tests
4. Discipline incidents
5. Attendance
6. Participation in tutoring, conferences
7. Parent contacts

So, any help I can get organizing and using all this information would save me a lot of time—not to mention improving the instructional services I can offer my students.

Table 1: Information Needed on the First Day of School

Information Needed on the First Day of School	Data Source	Data Sharing
1. What is my own class schedule and which students are assigned to each class period?	Student Information System	Classroom School
2. What language do the parents speak? (When I need to call a student's home, knowing ahead of time if they can speak English would be very helpful. By the way, this will change for some students during the year as they move from one relative's house to another, or parents come and go from their home.)	Student Information System Student Profile	Classroom School District State USED
3. Who are the parents/guardians and what are their relationships to the student? (When I call home, I'd like to know whom I'm talking to and be sure they are responsible for the student.)	Student Information System Student Profile	Classroom School
4. Which students mastered each of the "State Assessment" objectives on the last administration? (This initial chart would help me plan for where my students are starting. I could see the range of achievement levels. I would be able to plan the administration of my own diagnostic tests to group for instruction.)	Assessment DSS Reports	Classroom School District State
5. What special accommodations do students with an IEP require? (Last year I found out about accommodations after the first report cards went out. I would regret failing a student because I didn't know about special considerations or modifications I should have been making.)	Special Education Records Student Profile	Classroom School Special Education Program

Information Needed on the First Day of School	Data Source	Data Sharing
<p>6. What final grades did each student make in math from kindergarten through grade 7? (Some students will be consistent, others will be going up or down, and others may have had one really bad year along the way. If I know how they have been doing, I can look for gaps that might need to be addressed. Even though report card grades are notoriously subjective, I can interpret them to some extent based upon which teacher each student had.)</p>	<p>Student Information System Student Profile</p>	<p>Classroom</p>
<p>7. What final grades did each student make last year in every course? (Math success requires students to know how to read and have a broad vocabulary. If a student is struggling to read, I'll need to know that separately from their specific math skills.)</p>	<p>Student Information System Student Profile</p>	<p>Classroom</p>
<p>8. What crucial events have occurred in each student's life? (Last year, a student responded to an exercise to write three facts I needed to know about him by stating that he had no idea who his father is, his mother is in prison, and he has an arrest record already. I needed to be aware of these when sending notes home or following up on missed assignments.)</p>	<p>Student Profile</p>	<p>Classroom</p>
<p>9. What honors and awards has each student earned?</p>	<p>Student Profile</p>	<p>Classroom</p>

Table 2: Information Needed on a Daily Basis

Information Needed on a Daily Basis	Data Source	Data Sharing
<p>1. Important events happening in each student's life on a daily basis (On Algebra "State Assessment" testing day last year, a girl came to school and worried that she might not pass because her mother had thrown her out of the house the night before and she had spent the night on the streets. I let her sleep in the back of the room for an hour before starting the test. Because the test in untimed, she was able to finish and pass. Good events happen too, but they are easier to discover.)</p>	<p>Personal Student Report</p>	<p>Classroom</p>
<p>2. What my students do from 4 – 6 each day (Which ones play sports, take lessons, care for younger siblings, work, watch TV, roam the streets, or study?)</p>	<p>Personal Student Report</p>	<p>Classroom</p>
<p>3. Logs and lists of standard school activities beyond my classroom, like attendance in other classes, major discipline actions, events that excuse a student from my class, excuses that make a student eligible to get a grade on make-up work, etc.</p>	<p>Student Profile</p>	<p>Classroom</p>
<p>4. Extracurricular activities that students participate in and require them to have passing grades.</p>	<p>Student Information System Student Profile</p>	<p>Classroom School</p>
<p>5. The district's required curriculum for each course I'm teaching.</p>	<p>District Curriculum</p>	<p>Classroom School</p>
<p>6. The "State Assessment" objectives for the tests they will take this year—or next.</p>	<p>State Standards</p>	<p>Classroom School District State</p>

Information Needed on a Daily Basis	Data Source	Data Sharing
7. The curriculum resources that have been aligned with both the curriculum and the "State Assessment" objectives.	District Curriculum	Classroom School District
8. My prior lesson plans and any others that I might refer to when planning this year.	Instructional Management System	Classroom
9. A set of assessments I can administer as needed to know which students have mastered what was taught in class and which students need more help.	Instructional Management System	Classroom School District
10. Classroom management resources I can refer to if student discipline or motivation get to be issues.	Instructional Management System	Classroom School District
11. The student handbook and all other materials handed out to students and parents describing the school's expectations for them.	School References	Classroom School
12. The district policy book for those times when I might not be sure what I can and can't do.	District References	Classroom School District
13. Some way to locate a master teacher who has been successful teaching a skill my students are having trouble mastering.	District References	Classroom School District
14. My past evaluations, a self-assessment, and other resources that help me choose training to attend. Also, a list and schedule for required training everyone must attend.	Staff Profile	Classroom School

Information Needed on a Daily Basis	Data Source	Data Sharing
15. A list of support services available from the school, district, and community for students who encounter problems.	District References	Classroom School District
16. Logs and charts showing which students have completed assignments; a grade book of assignments completed and grades awarded; some way to summarize these grades so I can monitor how each class is doing compared to past classes.	Instructional Management System	Classroom
17. A list of phone numbers and names for all the offices I'll have to contact.	District References	Classroom School District
18. Honors and awards.	Student Information System Student Profile	Classroom School
19. A list of names and numbers for each student for contacting their parents and guardians.	Student Information System Student Profile	Classroom School

Table 3: Classroom Records Needed

Classroom Records Needed	Data Source	Data Sharing
1. Performance on classroom diagnostic tests tied to the curriculum	Instructional Management System Assessment DSS Reports	Classroom
2. Performance on daily work, assignments, and projects	Instructional Management System	Classroom
3. Performance on classroom tests	Instructional Management System	Classroom
4. Discipline incidents	Student Information System Student Profile	Classroom School District State
5. Attendance	Student Information System	Classroom School District State USED
6. Participation in tutoring, conferences	Student Information System Student Profile	Classroom
7. Parent contacts	Student Information System Student Profile	Classroom



About ESP Solutions Group

ESP Solutions Group provides its clients with *Extraordinary Insight*™ into PK-12 education data systems and psychometrics. Our team is comprised of industry experts who pioneered the concept of “data driven decision making” and now help optimize the management of our clients’ state and local education agencies.

ESP personnel have advised school districts, all 52 state education agencies, and the U.S. Department of Education on the practice of K-12 school data management. We are regarded as leading experts in understanding the data and technology implications of the **No Child Left Behind Act (NCLB)**, **Education Data Exchange Network (EDEN)**, and the **Schools Interoperability Framework (SIF)**.

Dozens of education agencies have hired ESP to design and build their student record collection systems, federal reporting systems, student identifier systems, data dictionaries, evaluation/assessment programs, and data management/analysis systems.

To learn how ESP can give your agency *Extraordinary Insight* into your PK-12 education data, email info@espsg.com.

This document is part of *The Optimal Reference Guide Series*, designed to help education data decision makers analyze, manage, and share data in the 21st Century.

Why Eva Baker Doesn't Seem to Understand Accountability – The Politimetrics of Accountability. Copyright © 2007 by ESP Solutions Group. All rights reserved. No part of this paper shall be reproduced, stored in a retrieval system, or transmitted by any means, electronic, mechanical, photocopying, recording, or otherwise, without written permission from the publisher.



ESP Solutions Group

(512) 879-5300

www.espsolutionsgroup.com