

White Paper

# Seeing Both the Forest and the Trees:

## Measuring Progress, Forecasting, and Intervention

by  
Christine Guerrero Burnham, Ph.D.



**Assessment  
Technology  
Incorporated**

**Assessment Technology, Incorporated**  
6700 E. Speedway Boulevard  
Tucson, Arizona 85710  
Phone: 520.323.9033 • Fax: 520.323.9139  
© 2009 Assessment Technology, Incorporated

# Seeing Both the Forest and the Trees: Measuring Progress, Forecasting, and Intervention

By Christine Guerrero Burnham, Ph.D.  
Assessment Technology Inc.

## Table of Contents

Table of Contents .....	i
Acknowledgements .....	ii
Abstract.....	iii
<b>I. Introduction .....</b>	<b>1</b>
<b>II. Formative Assessment and the Cycle of Intervention .....</b>	<b>1</b>
<b>III. Feedback.....</b>	<b>2</b>
A. Feedback Should Be Frequent and Timely .....	2
B. Feedback Should Provide Specific Pointers for the Improvement of Specific Skills .....	3
C. Feedback Should Focus on the Details of the Task and Not on Evaluations of the Competence of the Student .....	4
<b>IV. Interventions as Conversation .....</b>	<b>5</b>
<b>V. No Child Left Behind as an Intervention .....</b>	<b>5</b>
<b>VI. The Intervention System .....</b>	<b>6</b>
A. State and Federal Government .....	8
B. School/District and State.....	8
C. Teacher and School/District .....	9
1. Interim Assessment.....	10
2. Beyond Interim Assessment.....	14
D. Student and Teacher .....	16
<b>VII. Instructional Dialogs .....</b>	<b>16</b>
<b>VIII. How Some Districts Have Implemented Intervention Across Levels in the Intervention System .....</b>	<b>18</b>
<b>IX. Summary and Conclusion .....</b>	<b>19</b>
<b>X. References.....</b>	<b>21</b>

Copyright © 2009 by Assessment Technology Inc.

All rights reserved. No part of this document may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopy, recording, or any information storage and retrieval system, without permission from the publisher.

Assessment Technology Inc., Publishers  
Tucson, Arizona, U.S.A.

Printed in the United States of America.  
V1-022709

Seeing Both the Forest and the Trees  
Assessment Technology Inc.

## **Acknowledgements**

The author is grateful to Jack Bergan and Kathryn Bergan for their insightful comments on an earlier draft of this material.

Christine Guerrera Burnham, Ph.D.  
Senior Research Associate, Assessment Technology Inc.

## Abstract

Interventions in standards-based education take place in a multi-level context typically including classroom, school, district, and state levels. A successful intervention requires a data-driven conversation between parties at each level resulting in improved student mastery of standards as measured by statewide assessment. At all levels, the conversation focuses on the goals to be achieved, the current status with respect to goal attainment, and the instructional steps to be taken to close the gap between the two. As districts strive to improve the education of their students, intervention efforts must involve coordination across all levels of the intervention system, from the individual student to the state and federal governments. However, at some points in the system, the intervention conversation may become difficult due to a lack of continuity in the assessment information used to guide instruction at different levels. The danger is that the formative assessments (which are often used by teachers to guide instruction) may not be adequately linked to the benchmark assessments and that the benchmark assessments (which are often used to guide instruction at the school and district level) may not be adequately linked to the statewide assessment. When linking is inadequate, disconnects in the system can yield instructional guidance that in reality does not adequately further the student's capabilities to show mastery of academic standards as measured by the state assessment. This paper will help district and school administrators to understand the data-driven intervention conversation at different levels of the intervention system. It will identify points where data discontinuities are likely to occur and will recommend steps to ensure a continuous flow of information that will enable system-wide improvement of student learning.

# I. Introduction

Data-driven instruction requires at least two components: the collection of reliable, meaningful data and the use of that data to inform instructional practices. The methodological issues associated with reliability and validity in assessment comprise a vast literature and are beyond the scope of this paper. Rather, it is the use of the data to inform instruction that is under discussion here. This second component often appears to be more difficult to implement than the first, not because of reluctance or diffidence toward data, but because of inherent conflicts in the type of data that is most useful at the district level and that which is most useful in the classroom. Districts must measure progress in terms of global measures: averages of student performance, summaries of performance at the level of grade and subject, and probabilities of success on high-stakes statewide assessments expressed in terms of percentages of students.

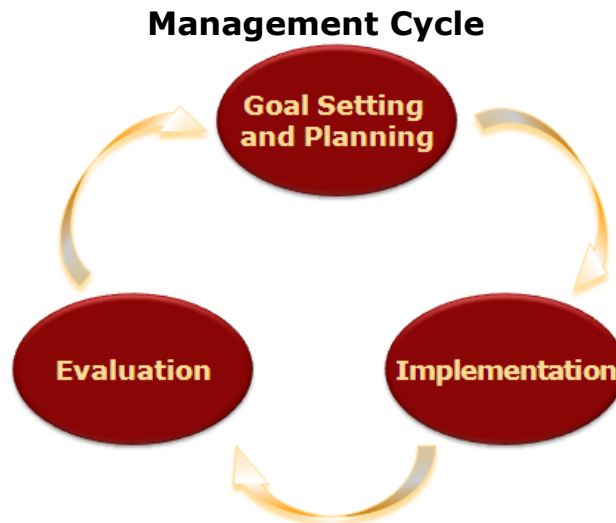
Teachers, on the other hand, must focus on individual skills and individual students. If global success in improving student performance is to be achieved at the district level, much of the work must be done at the detailed level of the interaction between teacher and student. This means that there must be coherence in the goals, implementation, and evaluation of progress across levels, from district to individual student. The language of this coherence is formative data. This paper will explore ways to improve the efficiency of intervention by ensuring that the participants at different levels of district-wide intervention are communicating, and that they are speaking the same language. More specifically, the claim will be made that the principles of the use of feedback from formative assessment data to drive intervention efforts can serve as a common framework, and a common language, at all levels of intervention from student to district.

## II. Formative Assessment and the Management Cycle

Bergan et al. (2008a) described the management cycle, which consists of three main components: goal setting and planning, implementation, and evaluation. Intervention is conceived as a process which cycles through these phases. With each pass through the cycle, the evaluation process leads to the modification of goals and plans. These modifications are implemented, performance is evaluated again, and instructional plans are modified further.

The evaluation phase of the management cycle is implemented by means of formative assessment. It is important at the outset to draw a clear distinction between formative and summative assessment. Summative assessments are administered at the end of instruction to measure achievement and are used to assign grades, certify a status, etc. Formative assessment, on the other hand, is administered during the course of instruction and is intended to provide feedback to modify and improve teaching and learning (e.g. Black & Wiliam, 1998; Shepard, 2006). Formative assessment should help teachers to help students answer the questions: *Where am I going?*, *Where am I now?*, and *How can I close the gap?* (Chappuis & Chappuis, 2007). From its earliest inception, the function of formative assessment has been to "...provide data that permit successive adaptations of a new programme" (Scriven, 1967, as cited in Baroudi, 2007). The phrase, "successive adaptations" highlights the fact that formative assessment is a cyclical function. Formative assessment is part of a dynamic system. Considered in

terms of the management cycle described by Bergan et al. (2008), formative assessment encompasses the *evaluation* and *goal setting and planning* phases of intervention, and is dependent on the third phase, *implementation*.



The successive adaptations that are guided by formative assessment data are adaptations to teaching and learning with the goal of bringing the learner closer to the desired goal of mastery. Black and Wiliam (1998a) further clarify the process:

The core of the activity of formative assessment lies in the sequence of two actions. The first is the perception by the learner of a gap between the desired goal and his or her present state (of knowledge, and/or understanding, and/or skill). The second is the action taken by the learner to close that gap in order to attain the desired goal. (Black & Wiliam, 1998).

### III. Feedback

Just as formative assessment is a driving force in intervention, feedback is a driving force of formative assessment. Feedback from formative assessment highlights the gap between a present state and the desired goal. Ideally, it also provides strong clues with regard to modifications in instruction or the student's studying behavior that will serve to close the gap. Much has been written about what makes feedback effective. Some common themes about feedback from formative assessment that have emerged in the literature are that formative assessment should:

- be frequent and timely, such that feedback is provided during the course of instruction, when the student can still work to modify her skills and knowledge (Martinez & Martinez, 1992; Spitzer, 1939),
- provide feedback with specific pointers for improvement of specific skills or kernels of understanding (Elawar & Corno, 1985; National Mathematics Advisory Panel, 2008), and
- provide feedback that focuses on ways to master the details of the task, and not on making judgments about the student's competence (Butler & Neuman, 1995; Kluger & DeNisi, 1996).

## **A. Feedback Should Be Frequent and Timely**

An early source of evidence supporting the benefits of assessment that is administered immediately after instruction comes from a classic experiment in cognitive psychology (Spitzer, 1939). The experiment involved 3,605 sixth-grade students who read an age-appropriate article that presented information that was expected to be new for the students. The students were divided into 10 groups that were tested on the material according to various schedules. For the purposes of this discussion, the final outcome measure was student performance on a 25 item multiple-choice test on the material that was administered 21 days after the students had read the article. Group I had been tested on the material the day the article was first read and again the following day before being tested a final time on day 21. Group VI was tested for the first time on day 21. On average, the students in Group I answered 12.2 items correctly while those in Group VI answered 6.5 items correctly. It was concluded that being tested on the material immediately after its presentation substantially reduced the degree of forgetting for the Group I students relative to the Group VI students, who had not been tested.

It is important to note that the experiment conducted by Spitzer specifically precluded the use of feedback from the initial tests. Teachers were instructed not to discuss the reading material or the tests with their students. It seems, therefore, that there is some benefit to memory that is derived simply from the task of being tested. Similar results were obtained more recently by Martinez & Martinez (1992) who divided college students taking an introductory algebra course into two groups, one of which was tested three times on each chapter, and the other only once per chapter. The students who had been tested more frequently performed better on a posttest than those who had been tested less frequently. Again, this experiment addresses frequency of testing without any consideration of whether the results of the tests were used formatively, either by the teacher or by the students themselves, to guide further learning. Of course, in intervention we are concerned with the use of formative assessment data to improve instruction, and there is evidence that such use of feedback increases student achievement.

## **B. Feedback Should Provide Specific Pointers for the Improvement of Specific Skills**

In an investigation of the benefits of constructive feedback, Elawar and Corno (1985) trained an experimental group of nine teachers of sixth-grade students in providing specific written feedback on homework assignments. The teachers' feedback was guided by the following questions: (1) What is the key error? (2) What is the probable reason the student made the error? (3) How can I guide the student to avoid the error in the future? and (4) What did the student do well that can be noted? This feedback procedure was applied to three homework assignments in mathematics per week for a period of 10 weeks. The students of the nine teachers who comprised the control group completed the same number of homework assignments, but their feedback consisted simply of the number correct. The performance of the students in the experimental group on a mathematics posttest was significantly higher than that of the students in the control group.

More recently, the National Mathematics Advisory Panel (2008) conducted a meta-analysis of research on the benefits of the use of formative assessment in mathematics instruction. The first set of papers that they reviewed compared regular use of formative assessment to a control condition. This meta-analysis yielded an overall effect size that was marginally significant, suggesting that students who had had the benefit of regular formative assessment performed better on an outcome measure such as the Math Computation Test-Revised than students in a control condition. The use of formative assessment leaves open the question of whether, or how effectively, teachers were using the data from the assessments to provide feedback to students and guide further instruction. The math panel therefore conducted a second analysis of experiments that investigated the use of formative assessment with enhancements. The enhancements in the experiments were intended to guide teachers in the use of formative data. Enhancements included such things as a formative assessment system that provided teachers with a detailed analysis of student strengths and weaknesses in each skill, software that draws on formative assessment data to provide teachers with specific instructional suggestions, and self monitoring, in which teachers responded to questions such as “What skill(s) should be targeted for this student in the next two weeks?”. The math panel meta-analysis yielded a statistically significant effect size for experiments that contrasted formative assessment plus enhancements to a control condition, and a marginally significant effect size for experiments that contrasted formative assessment with enhancements to formative assessment without. These results led the panel to recommend the use of formative assessment in mathematics education, and to “cautiously call [the use of enhancements] promising” (National Mathematics Advisory Panel, 2008, p. 6-182.)

### **C. Feedback Should Focus on the Details of the Task and Not on Evaluations of the Competence of the Student**

A further point about the use of feedback from formative assessment to improve learning that is emerging from research concerns the question of whether feedback should be directed specifically to the task at hand or to student characteristics, such as the use of grades or other forms of assessment of student ability. The consensus seems to be that feedback should be directed to the details of the task and leave evaluative judgments out of the picture. Butler & Neuman (1995) conducted an experiment with second- and sixth-grade students who were asked to solve a set of puzzles. The students were divided into two groups: a Task Focus group and an Ego Focus group. The Task Focus group was told that the goal of the puzzle task was to learn how to do puzzles and how to improve in solving puzzles. The Ego Focus group was told that children who could solve the puzzles are known to be very smart. They found that, for both age groups, children in the Task Focus group were significantly more likely to ask for help than those in the Ego Focus group. Student achievement was measured in terms of the number of puzzles children were able to solve on their own at the end of the experiment (taking into account any differences in the number they were able to solve alone at the beginning). They found that students who had asked for more help were able to solve more puzzles alone by the end of the task. They had achieved greater mastery.

Kluger & DeNisi (1996) conducted a meta-analysis of 131 controlled studies that looked at the effect of the use of feedback in formative assessment on student performance. Their meta-analysis revealed an average effect size of 0.41 which they



interpreted as suggesting that, on average, the use of feedback has a moderate positive effect on performance. However, they also noted that over 38 percent of the effects were negative, suggesting that feedback can actually sometimes *detract* from performance. They therefore conducted a detailed analysis pulling out specific components of the use of feedback, such as whether praise is given, whether the correct solution is provided, the degree of complexity of the task, and so on. One of their major conclusions after this detailed analysis was that feedback that directs attention to the self (e.g. classroom grades or even praise) tends to attenuate the effect of the feedback on performance, whereas feedback that directs attention to task motivation or task-specific processes augments the effects of feedback on performance. Their conclusion is that feedback directed to the student's characteristics detracts attention away from the task at hand.

Having reviewed some of the principles regarding effective formative assessment in the service of intervention in the classroom, we are now able to apply these principles to all levels of intervention.

#### **IV. Intervention as a Conversation**

It is clear from the review of research on the use of feedback to improve learning that effective intervention is a conversation of sorts. The conversation should begin with the teacher communicating to the students the goal associated with a particular lesson. The student's performance on an assignment or quiz serves as a communication to the teacher about his or her current state of knowledge or skill. With the formative assessment data in hand, the teacher communicates to the student any gaps between the student's current state and the desired goal and, crucially, provides information about *how* the student can close that gap, whether it be through formally re-teaching the concept, through written comments on a homework assignment, or by some other means. The conversation continues via further formative assessment, by which the student communicates his or her (hopefully improved) state, the formative data are analyzed relative to the goal and the conversation continues as necessary.

#### **V. No Child Left Behind as an Intervention Initiative**

Thus far we have considered intervention and formative assessment at the level of student and teacher. This is the pinpoint where the improvement of student learning is necessarily focused. But it is a mistake to assume, particularly in the current educational climate, that the intervention conversation between student and teacher occurs in isolation. Rather, it is most often the functional tip of a much larger intervention initiative. Intervention initiatives frequently and appropriately are enacted at the school or district level. The spirit of the No Child Left Behind Act (NCLB) of 2001 is that of an intervention effort at a national level. This is clear in the introduction of the Act into public law:

An Act

To close the achievement gap with accountability, flexibility, and choice so that no child is left behind (NCLB, 2001, p.1).

By phrasing NCLB in terms of closing a gap, the act certainly takes on the appearance of an intervention effort and presumably it was intended as such by its authors. But notice that the proposed remedy for closing the gap is accountability. Accountability translates as summative assessment, not formative. The high-stakes, statewide assessments by which districts fulfill their NCLB accountability obligations are absolutely summative assessments. Accountability and summative assessment do not serve the intervention conversation in the way formative assessment must. It could be argued that high-stakes summative assessment provides some feedback regarding the current state (the present percent of students demonstrating proficiency on the assessment) and the desired goal of 100 percent proficiency or, more realistically, the current Annual Measurable Objective (AMO). But the feedback provides no guidance about how to close the gap, nor is it timely with regard to instruction. To be fair, high-stakes, statewide assessments were not designed to play a formative role.

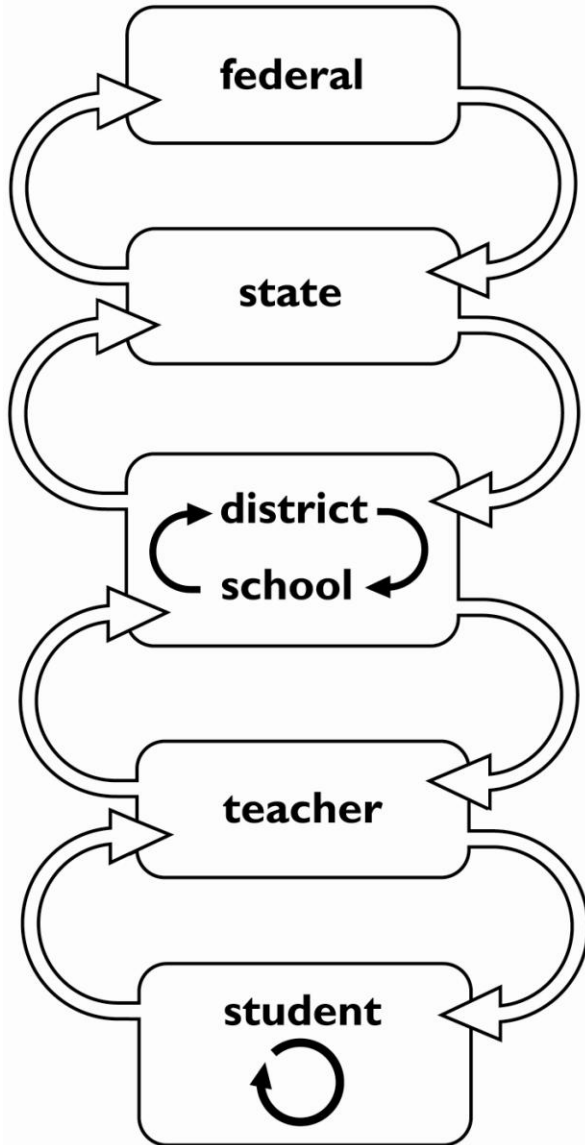
High-stakes statewide assessments can identify gaps in very broad terms. They can identify for districts the grade levels and subject areas in which students are attaining the desired goals and those that are problematic, and they can identify subgroups of students who are experiencing difficulties. But it is left to districts, schools, and teachers to identify the specific student skills that must be addressed and *how* to improve student understanding in those areas. High-stakes statewide summative assessment could not fulfill this function and it should not be expected to do so. But it is important to recognize that NCLB only provides a portion of the intervention conversation. Districts and schools must work to provide the rest, so that the management cycle can function appropriately and student achievement can be improved.

## VI. The Intervention System

Within each state, intervention in the context of NCLB accountability is best conceived as a system that spans the levels of student, teacher, school, district, state, and the federal government. The system will only function properly if the intervention conversation functions appropriately at all levels. This means that, between any two levels, there must be relevant information flowing up from the lower level to the higher level (bottom-up information) and information flowing back down from the higher level to the lower level (top-down information). A breakdown in the intervention conversation between any two levels means that the bottom-up or top-down flow of information is disrupted, and the performance of the intervention system as a whole is compromised. Figure 1 illustrates how the intervention system is conceived<sup>1</sup>.

---

<sup>1</sup> This conception of the intervention system is inspired, with apologies, by interactive activation models in the study of human cognition (e.g., McClelland & Rumelhart, 1981).  
Seeing Both the Forest and the Trees  
Assessment Technology Inc.



**Figure 1**  
**The Intervention System**

In the intervention system, the bottom-up and top-down flow of information at any given point in the spectrum is carried by the intervention conversation between two levels. The character of the conversation differs at different levels, but if the system is to work as a whole, each level must be able to communicate with the levels above and below it. Some conversations are fairly straightforward. Others are not. It is worth considering the flow of information at each level closely, so that likely breaks in communication can be identified and prevented. In the diagram, the district and school levels are treated as a single unit because these two levels tend to speak the same language (e.g. aggregated district-wide assessment data), and so this conversation is relatively straightforward. This does not mean that conversation between school and district always flows smoothly, but for the purposes of this discussion, we are concerned with places where the conversation is inherently difficult because of differences in the kind of data that are most relevant at the different levels. The circular arrow within the

student level is meant to represent the student's use of feedback in self-assessment and in monitoring her own progress toward goals (e.g., Fontana & Fernandes, 1994). In the following sections we will consider the characteristics of the intervention conversation at each level of the intervention system.

### **A. State and Federal Government**

Under NCLB, the highest level of conversation in the interaction system is between the states and the federal government. At this level, the conversation is primarily in the form of the Adequate Yearly Progress (AYP) workbook submitted by each state to the U.S. Dept. of Education on an annual basis. The workbook serves as the bottom-up information in the intervention conversation at this level, and it specifies the steps the state is taking to ensure accountability and comply with NCLB. The top-down side of this conversation comes primarily in the form of mandates from the federal government, and decisions regarding the details of compliance, such as whether or not a growth model is to be permitted in accountability. The only student performance data that is invoked in this conversation comes in the form of biennial National Assessment of Educational Progress (NAEP) assessments. Under NCLB, all states who wish to receive Title I money must participate in the biennial NAEP assessments in reading and math in the fourth and eighth grades. No rewards or sanctions are delivered in response to NAEP data. Data are not reported for individual students or schools. The relative infrequency and limited coverage of the required NAEP assessments make them poor candidates for providing feedback that is useful in the intervention conversation, nor were they designed to be used in that manner. The conversation between states and the federal government essentially boils down to mandates from the federal government telling states that they must monitor student progress and take action with schools that are failing to make progress, and with states documenting that they have procedures in place to comply with that mandate. The conversation provides no pointers with regard to how to improve student performance.

### **B. School/District and State**

The next level in the intervention system is the conversation between the school district and the state. The obvious channel for this conversation is the high-stakes statewide assessment. Although these high-stakes assessments are a cornerstone of the NCLB program for improvement of student performance, they are summative in nature and are not intended to provide detailed pointers to improve learning. They do, however, provide some information that is relevant to intervention. For example, they identify the broad subject areas in which students in general are struggling, and those in which they are demonstrating mastery. They identify which schools, grades, and subgroups are performing adequately, or even excelling, and which are falling behind. This is useful information and districts should glean from it what they can. However, there is no getting around the fact that these are summative assessments and their usefulness for intervention for the students who took the assessment is severely limited. They are generally administered toward the end of the school year, when instruction is coming to a close. The results are often not available for several months by which time the students are in the next grade pursuing a new curriculum. Student performance within a subject area is often reported with reference to some of the broader strands within a set of state standards, but is rarely if ever reported for individual learning standards within each strand. For example, the data may indicate that, on average, the

students in a given district or school demonstrated mastery in the 'number sense' strand, but the aggregated nature of the data may disguise the fact that many students are excelling on items in which fractions are represented by physical representations such as shaded squares but failing on items that require more abstract representations of fractions.

In spite of the limitations of high-stakes summative data as an instrument in intervention, it is the data that is available at this level of the intervention system, and schools and districts should use it to the best of their ability. Using high-stakes summative data, districts can and do identify schools and subgroups that require intense remediation efforts. The data can also be used to identify which portions of the curriculum are and are not working. This analysis can go beyond the general (e.g., "Fifth-grade math seems to be fine, but sixth-grade math is not attaining the desired goals."). It can identify sections within the curriculum that need to be re-worked. For example, it may turn out that the data analysis portion of the sixth grade math curriculum is attaining the desired goals in terms of student performance on the statewide assessment, but that the number sense portion is not. Although the individual students who took the statewide assessment have moved on to the next grade, the school and district curriculum for each grade and subject continues into the next year. So while the feedback comes too late for the individual students, it is still useful at the school and district level, for which the feedback is actually intended.

The conversation between the state and the school/district is certainly unbalanced, but it is not without value. The bottom-up portion of the conversation is the students' demonstration of their skills and knowledge on high-stakes statewide assessments. The NCLB mandate ensures that this portion of the conversation comes through loud and clear. The top-down portion of the conversation, the feedback that can guide improvement of student performance, is sometimes faint and its timing is not optimal. But it is there. This is one juncture of the intervention system that requires determined and well-organized effort on the part of district and school administrators if it is to function efficiently.

### **C. Teacher and School/District**

The interaction conversation between a teacher and school or district can easily be dominated by top-down information. Much of this information can be very helpful. For example, teachers can receive detailed information about the performance of their students on district-wide assessments known as benchmark or interim assessments. Within the context of standards-based education, interim assessments can usually provide detailed information about specific learning standards that students have not yet mastered. But it is not clear that the bottom-up information in this exchange is as efficient. Schools and districts may not have a consistent or clear picture of the work that is being done in the classroom.

It is at the level of teacher and school/district that the language barrier in the intervention conversation is most severe. The levels above the teacher all speak in terms of averages, aggregated scores, probabilities of success, and broad subject matters. To that extent they can communicate with each other although, as we have seen, those conversations can have their difficulties for other reasons. But for the

classroom teacher, the intervention conversation must be in terms of specific students and specific skills.

Intervention conversations are primarily in the form of data. Data at the school, district, and state level are most often based on formal, high-stakes statewide assessments. We have seen that the time frame and aggregate nature of these assessments serve to make them less than optimal for intervention efforts in the classroom. However, these are generally good, solid assessments that are constructed under strict guidelines for quality (i.e. AERA, APA, & NCME, 1999) and which are backed by research establishing their validity and reliability. The data derived from these assessments can be trusted as objective measures of student ability. In contrast, classroom formative assessments are immediate and specific, which serves the task of intervention well. However these assessments are much less formal and are not aligned to any criteria that would establish reliability or validity. Even Black & William (1998), staunch advocates of a movement for more formative assessment in the classroom, acknowledge that teacher classroom assessments often measure and encourage superficial learning. It is not clear that the demonstration of mastery on such assessments would necessarily imply that students are ready to demonstrate mastery on high-stakes statewide assessments.

#### 1. Interim Assessment

Perie, Marion, Gong, & Wurtzel (2007) have studied a level of assessment, which they term *Interim Assessment*, that can bridge the gap between informal classroom formative assessment and high-stakes, statewide summative assessment. Interim assessment refers to a level of assessment that is often called “benchmark,” “diagnostic,” or a variety of other names. Interim assessments, like classroom formative assessments, are designed to evaluate student performance relative to a very specific set of goals. However, like high-stakes statewide assessments, they are designed according to high standards, have established reliability and validity, can be aggregated meaningfully for consideration at the school or district level, and are usually aligned to the high-stakes assessments so that they can be used to forecast student performance on those assessments. Interim assessments, such as the benchmark assessments prepared by ATI for school districts, serve the intervention conversation between districts and teachers by being somewhat bilingual: they can provide information to participants at each level in the language that they require.

One of the goals of interim assessments is to forecast student performance on high-stakes statewide assessments. If the interim assessments are properly aligned, such forecasting can be quite accurate. For example, a review of 648 Galileo K-12 Online benchmark assessments administered during the 2006-07 school year along with the students’ subsequent performance on the spring, 2007 high-stakes statewide assessment indicated an average correlation coefficient of 0.76. In addition, based on those same assessments, on average 96 percent of the students who passed all three interim assessments went on to pass the statewide assessment, while 86 percent of those who failed all three went on to fall below the required cut score for proficiency on the statewide assessment (Bergan et al., 2008b).

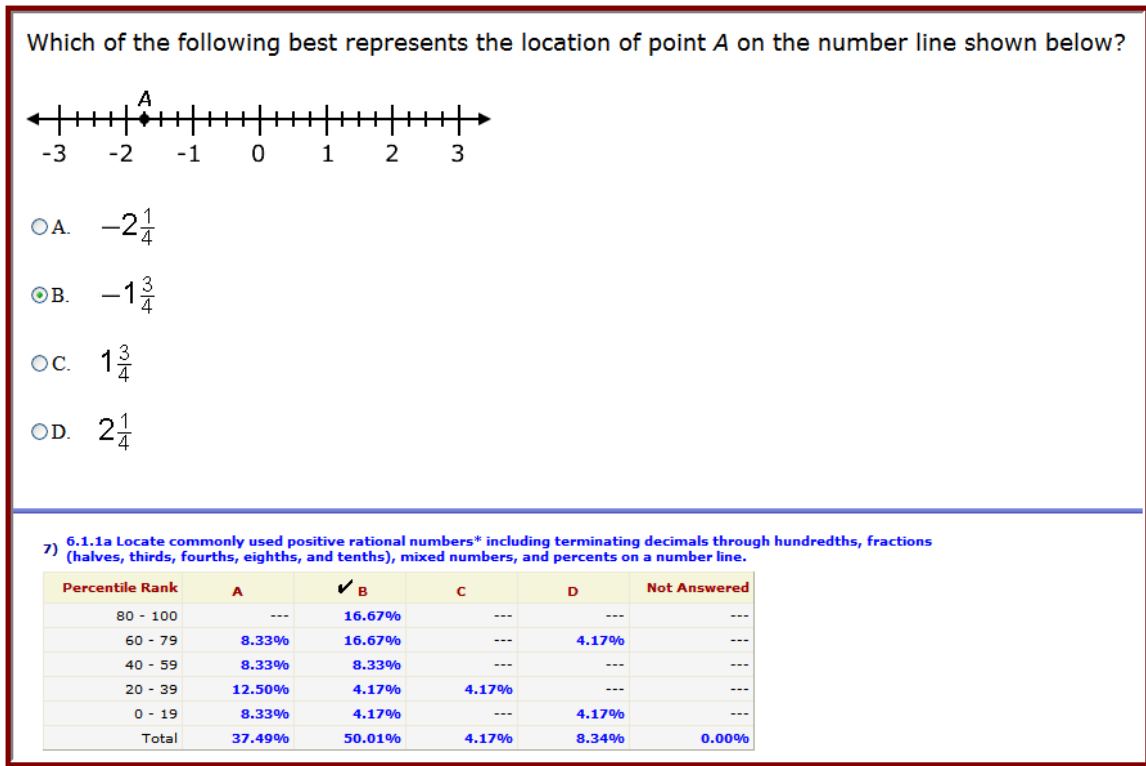
Accuracy in forecasting student performance on statewide assessments means that interim assessments can play the formative role of identifying gaps in the percent of students demonstrating proficiency in a given grade and subject relative to a stated goal. The goal is most frequently expressed in terms of a state-mandated AMO. For example, if the state-mandated AMO for demonstrating proficiency in sixth grade math is 67 percent of students and the interim assessment forecasts that only 57 percent of students are likely to demonstrate mastery, then the gap, or current position relative to the goal, has been clearly identified. Feedback expressed in these terms is directly relevant for intervention at both the district and school levels, the only difference being the level of aggregation, which is why it is claimed that districts and schools speak the same language in the intervention conversation. Turning to the intervention conversation between school/district and teacher, it is important that schools/districts make sure that each teacher (1) knows the current AMOs for his or her grade/subject and (2) has access to the benchmark assessment data for his or her students including information regarding the percent of students forecast to demonstrate proficiency on the statewide assessment. In addition, the teacher should know which students in her class are identified as being likely to succeed on the high-stakes statewide assessment and which are not. In other words, the gap (if any) identified by forecasts of student performance on high-stakes assessments should be made salient to the teacher. The school/district may have to take specific steps to make sure that the teacher has this information.

A well designed interim assessment with a good reporting system can also provide feedback with specific pointers for improvement. One advantage of standards-based education is that this step becomes relatively simple. Reports generated from such assessments should be able to identify for the teacher the standard(s) with which his or her students are having difficulty, as well as which students are experiencing the difficulty.

An interim assessment can provide the teacher with feedback that is much more specific than identifying problematic learning standards. For example, a well-written multiple-choice assessment question includes distracters that represent typical errors of students (Haladyna, 2004). Therefore, a review of the pattern of student responses to benchmark assessment items should locate precisely for the teacher the gap in understanding with regard to a particular skill. Figure 2 presents the *Galileo Detailed Item Analysis* report for a sixth grade math question that was presented on a district benchmark assessment in the fall of 2008. The item analysis is presented in terms of student percentile ranks, so that different patterns from high-achieving students and low-achieving students can be identified. This report was generated at the level of an individual class. A number of cells in the grid are blank because of the relatively small number of students. Even with the small sample size, the pattern of responses becomes clear.

Although 50 percent of the students in this class responded to the item correctly, 37 percent selected distracter A instead. A review of the item makes the gap in understanding clear. Although the students understood that the minus sign indicated that they should look to the left of the 0 on the number line (i.e. they did not select distracter D), they failed to apply this knowledge to the fractional portion of the target number. A further drill-down from this report would provide the teacher with a list of

students who selected distracter A, and a quick re-teaching exercise could be implemented to close this particular gap.



**Figure 2**  
**A Detailed Item Analysis**

A district-wide interim assessment, then, can play a formative role by identifying gaps in learning to a teacher and by providing her with feedback with specific pointers for improvement. It is important for districts to make sure that teachers have access to this information. As we have seen, feedback should also be provided during the course of instruction so that it can be useful in guiding further instruction. In this regard, district-wide interim assessments may not be as well-suited to the teacher's task as less formal classroom formative assessment. If interim assessments are to serve the function of forecasting student performance on statewide assessments, they must live up to a set of psychometric standards that will ensure their validity and reliability (AERA, APA, & NCME, 1999). In order to fulfill these requirements, the assessment must be of a minimal length and must be administered to a relatively large sample of students. Interim assessment items must be constructed with a great deal of care and must comply with professional standards ranging from consistency of format to assurances against bias. The administration of interim assessments must be as standardized as possible. For example, it is not acceptable for some students to be permitted to use calculators while others are not. To the extent possible the assessment should be administered to all students at the same point in time in the instructional calendar. And it should measure a reasonably broad set of learning standards to assure a more complete estimate of student ability. In addition, there are practical concerns, such as the time required to print up test booklets for an entire district if the assessment is to be taken via paper and pencil or scheduling arrangements for computer labs if it is to be taken online. All of these factors combine to dictate that district-wide interim assessment



must take place at appointed times in the school year, usually no more often than three or four times during the year.

District-level interim assessments must be valid and reliable instruments, and so they must meet all of these requirements. The tradeoff is that, by the time the assessment is administered, at least some of the content that is assessed will have been covered weeks earlier in the school year. This state of affairs appears to violate the suggestion that formative assessment should be administered during the course of instruction. This is indeed a point in the intervention system where the conversation between levels is not seamless. It is important to recognize the slightly awkward nature of this juncture and supplement it with other means.

The most obvious solution to the timeliness problem is that less formal classroom formative assessment should be administered frequently, as discussed throughout the earlier sections of this paper. These more frequent, less formal formative assessments will serve to supplement the benchmark assessment on a more finely grained time schedule. With feedback from classroom formative assessments, adjustments to instruction can be made on the fly to improve student learning. The primary work of improving a student's skills and knowledge takes place in the classroom, and routine formative assessment should bring students as close as possible to the target skills and knowledge during the course of instruction, in anticipation of district interim assessments. The interim assessments should, in turn, act as a further check on where students stand relative to the goal. It is possible for classroom formative assessments to fail to plumb the depth of knowledge or level of difficulty that will be expected on high-stakes statewide assessment. A well-aligned interim assessment may reveal that students who appeared to have attained the goal in response to classroom formative assessment are still short of the goal when assessed by more objective methods. It is also possible for students to forget. A district interim assessment can reveal gaps in student skills and knowledge that can be addressed via re-teaching. They can also reveal to the teacher gaps between the rigor of her classroom formative assessment and the rigor expected on high-stakes assessments, and teachers can adjust their formative assessment practices accordingly.

An important guideline regarding the feedback function of formative assessment is that it should focus on skills and not on an assessment of competence. When applied to the feedback function of district-wide interim assessment, this directive suggests that for interim assessments to play a valuable formative role in the classroom, it is crucial that the focus, both of administrators at the district and school levels and of the teachers in the classroom, be on the presence or absence of particular skills as determined by student performance on the interim assessments. The value of the interim assessment is negatively impacted if the student performance on interim assessments is used as a measure of teacher ability. For teachers to maximize the value of student information from interim assessment, the teachers must have the confidence to accept the formative data that an interim assessment provides them without feeling defensive, threatened, or in other ways at risk. This can be a very tricky issue. A survey conducted in 1995 found that:

Most of the teachers in this study were caught in conflicts among belief systems, and institutional structures, agendas, and values. The point of friction among these conflicts was assessment, which was associated with very powerful feelings of being overwhelmed, and of insecurity, guilt, frustration, and anger. (Johnston et al., 1995, p. 359)

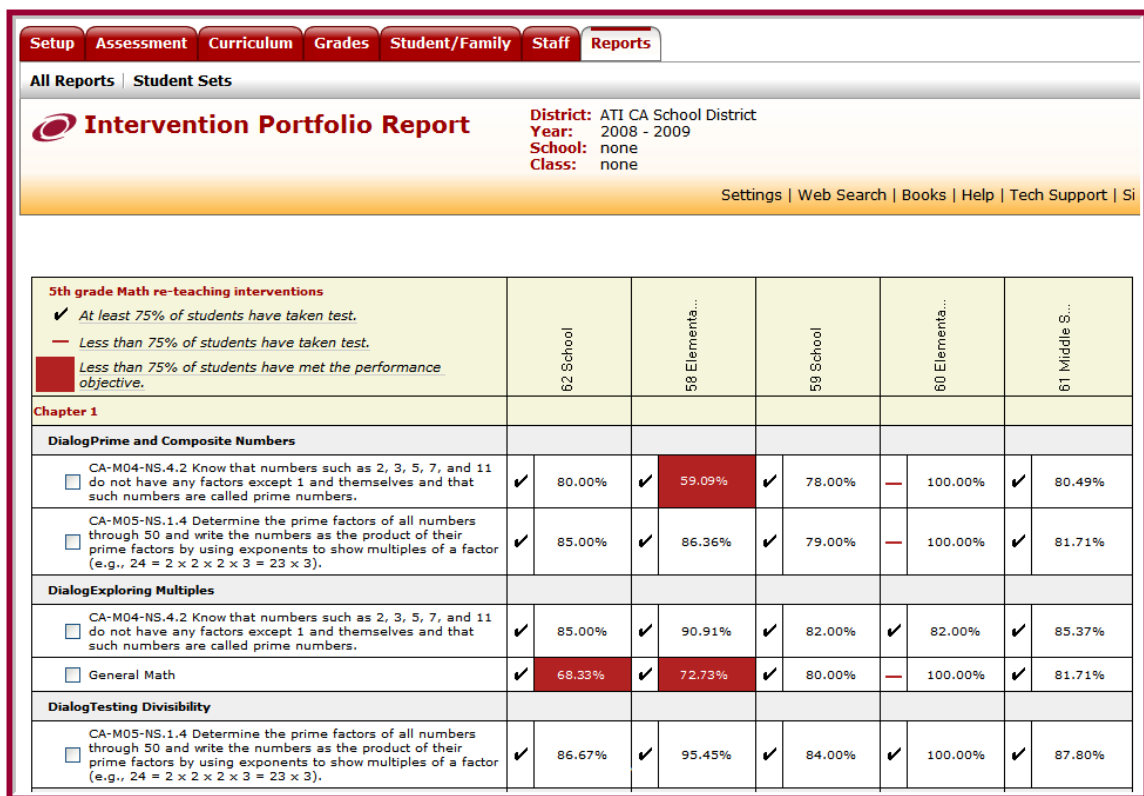
It is difficult to imagine a teacher using data from interim assessments effectively to guide adjustments in instruction if she resents the assessment program. It is likely that feedback from assessment that causes a teacher to focus on her own confidence and self-esteem will distract her from focusing on the details of the task at hand, finding ways to close the gap between her students' current level of understanding and the goal, just as Kluger & DeNisi (1996) conclude that it does for individual students. A further problem in this vein is the growing trend of using the results of interim assessments to determine which teachers will receive bonuses. If interim assessments are to guide intervention, they must identify gaps in student learning. That is their entire purpose. If teachers feel they are being judged by the assessment data, they will be frustrated and angry if the majority of their students have not earned scores of 85 percent or 95 percent correct. An assessment on which most students respond to most items correctly serves no formative purpose and cannot benefit intervention. An interim assessment generally cannot be used simultaneously to inform instruction and reward teachers. The two objectives are often mutually exclusive.

## 2. Beyond Interim Assessment

Thus far the discussion of the intervention conversation between the school/district and the teacher has focused on the top-down portion of the conversation, on the ability of a district-wide interim assessment to provide feedback to guide instruction in the classroom. The bottom-up portion of this conversation is just as important, and is often overlooked. Interim assessment provides the district with data about student progress that is timely relative to that of high-stakes statewide assessments and which can provide important periodic updates on student progress while there is still time to take action in preparation for high-stakes statewide assessments. But for the reasons considered earlier, interim assessment is generally limited to three or four times a year. Schools and districts should have more frequent updates about student progress in the classroom. They must rely on aggregate data, of course. It is impossible for school or district administrators to monitor the progress of every student. But they should be able to monitor whether teachers are keeping up with the pacing guide and it would be beneficial for them to have some idea of how the students are performing on the content leading up to the interim assessments. Ideally, teachers should be able to provide schools and districts with a summary of progress based on their classroom formative assessment data. However, this requirement could potentially be so time consuming that it is prohibitive. An integrated, computer-based system that automatically records formative assessment data in the classroom and provides school and district administrators with a summary of progress, both in terms of content covered and student mastery of the content, would greatly enhance the intervention conversation between teacher and school/district.

Galileo K-12 Online *Instructional Dialogs* can improve the bottom-up flow of information from individual classrooms to school and district administrators. *Instructional Dialogs* are instructional materials that are integrated with the Galileo K-12 Online benchmark assessment system, the details of which will be described further in the

following section. For the current purposes, the structure of *Dialog Books* serves as an outline of the district curriculum. Each *Instructional Dialog* within a book is an individual, interactive lesson aligned to a specific learning standard in the state or district curriculum, and each contains several (usually five) formative assessment items with feedback on errors designed to guide the student to a better understanding. Each Dialog may have an optional assessment at the end, without feedback to the students, the results of which are recorded in Galileo. The *Intervention Portfolio* provides teachers and administrators alike with an overview of each dialog in the book that includes information about whether the dialog has been administered and which reports the average performance of the students on the optional assessment. Teachers can review the report for individual students. School and district administrators can view the report aggregated at the class, school, or district level so that progress through the curriculum is visible at a glance. Figure 3 provides an illustration of the *Intervention Portfolio* aggregated at the school level.



**Figure 3**  
**Intervention Portfolio**

The intervention conversation between teacher and school/district is a complex and crucial one. A discussion of how some districts have tried to make this conversation as effective as possible will follow a review of the intervention conversation between student and teacher and a further description of *Instructional Dialogs*.

#### D. Student and Teacher

Finally, we return to the intervention conversation between the student and teacher. Most of the relevant issues at this level have been discussed in earlier sections of this paper. Here, the student communicates with the teacher by means of his or her performance on formative assessment items, including classroom assignments, homework, quizzes, and so on. The teacher in turn provides feedback that is intended to highlight any gaps between the student's current state and the desired goal. Such feedback should be frequent and should be provided during the course of instruction, so that the student can modify her skills and knowledge (Martinez & Martinez, 1992; Spitzer, 1939). Feedback from formative assessment should provide specific pointers for improvement, and should be directed toward the task at hand rather than at judgments about the student's competence (Butler & Neuman, 1995; Elawar & Corno, 1985; Kluger & DeNisi, 1996).

We have also seen that there are some shortcomings with classroom formative assessment as it is generally practiced. Teachers may assess at a shallow level that focuses on recall rather than greater depth of knowledge (Black & William, 1998). Since classroom formative assessments have no established validity and reliability and are not necessarily carefully aligned to state learning standards, there is typically no way of verifying whether the demonstration of mastery on such assessments indicates a likelihood of success on the statewide assessment. District-wide interim assessments play an important role in checking to make sure the mastery demonstrated in the classroom is aligned to the expectations of statewide assessments. But in the ideal situation, such alignment would be checked more frequently, in the course of ordinary instruction.

An integrated system that includes classroom instruction, classroom formative assessment and district-wide interim assessment can bring classroom formative assessment in line with higher level expectations and intervention efforts. Galileo K-12 Online *Instructional Dialogs* were designed to serve this purpose.

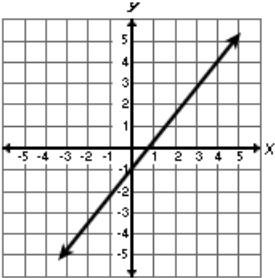
### VII. Instructional Dialogs

*Instructional Dialogs* are online, interactive lessons that are designed to integrate instruction and progress monitoring in the classroom. Each Dialog begins with a clearly stated goal for the lesson, such as, "In today's lesson, we will be identifying the slope of a line." Each successive slide in the Dialog takes the students through another step in the process or further explains the target concept. Every *Instructional Dialog* contains at least one, usually several, questions embedded within it that test the students' understanding of the concept that has been introduced. When students submit their answers, they are immediately provided with feedback. If the student selects an incorrect response, the feedback points out the error to the student and provides information that is intended to guide him or her toward the correct solution. Figure 4 provides a sample of feedback in response to the selection of an incorrect answer. In this example, the question was presented after a series of slides that explain how to identify the slope of a line, and the student selected Graph C instead of the correct answer, Graph B.

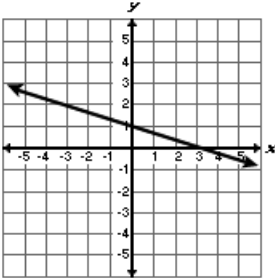
Dialog Completion  
11 of 15

Slide 11

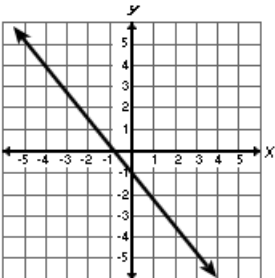
Which of these graphs shows a line with a slope of -1?



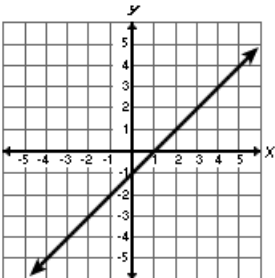
Graph A



Graph C



Graph B



Graph D

A. Graph A
  C. Graph C

B. Graph B
  D. Graph D

Save My Answer

This is a negative slope, but it goes over 3 spaces not 1. What should the denominator be?

**Figure 4**  
**Example of Feedback to an incorrect student response in an Instructional Dialog**

As this example shows, *Instructional Dialogs* provide immediate feedback that provides specific pointers to the student regarding how to close the gap between his or her current level of understanding and the goal of complete understanding. The student may change his answer as many times as he would like until the correct answer is selected at which point, of course, he is told that he was correct. Since the feedback comes from a computer, it is clearly directed toward the skills involved in the task at hand rather than a judgment about the student's ability. The benefit of computer-based feedback has been noted before. In their meta-analysis of research on feedback in intervention, Kluger & DeNisi (1996) noted that the effect of feedback on student performance was greater when it came from a computer than when it came from a person.

In an *Instructional Dialog*, the questions embedded within it, along with the automatic feedback to students, are only the first level of formative assessment. The next level comes in the form of an optional formative assessment at the end of the Dialog. The optional assessment generally contains about five items that address the content of the Dialog, and the student responds to these in the same way she responded to the items embedded within the Dialog. The difference is that these items do not contain automatic feedback, and the student's responses are recorded. The teacher can then review the student responses and use them to guide further instruction as she would with any other formative assessment. The student scores on the optional assessment provide a record that the Dialog was administered, and also an indication of the degree of student mastery of the concepts that were addressed by the Dialog. These student scores form the basis of automatic record keeping that can facilitate the bottom-up communication between teacher and school/district. Administrators can pull up reports for the Dialogs, see which Dialogs have been administered in which classes, and can see aggregated data that indicates, on average, how well the students learned the material. *Dialog Books* can be designed that reflect the district's curriculum. The data that are collected from Dialogs can provide school and district administrators with a means for continuously monitoring progress in the district curriculum at the district, school, and class levels of aggregation.

In addition to facilitating record-keeping and communication between teacher and school/district, *Instructional Dialogs* can ensure continuity in the degree of rigor between classroom formative assessments, district-wide interim assessments, and high-stakes statewide assessments. As was previously discussed, interim assessments are already carefully designed to align to the high-stakes statewide assessments in terms of both content and rigor. *Instructional Dialogs* are, in turn, designed to align to interim assessments. This means that, if a student demonstrates mastery on the optional formative assessment at the end of an *Instructional Dialog*, there is good reason to expect that she will demonstrate mastery with regard to that particular concept on the district-wide interim assessment and, consequently, on the high-stakes statewide assessment.

## **VIII. How Some Districts Have Implemented Intervention Across Levels in the Intervention System**

This final section will review some solutions that school districts using Galileo K-12 Online have devised to overcome the difficulties of coordinating intervention efforts between the school/district level and the level of individual teachers.

One solution that has been very effective involves using the data from interim assessments to guide very intense, school-wide re-teaching efforts. One key to the success of this initiative is that time for the re-teaching was set aside as the school year was planned. Before the first interim assessment was administered, everyone knew that time was to be devoted to re-teaching in response to the results of the interim assessment. Teachers collaborated on the creation of formative assessments to be administered at the end of each instructional unit, and students needed to get 80 percent correct in order to demonstrate mastery. If a student fell below this mark, she received further instruction. Quarterly district-wide interim assessments were then administered and, again, students who fell below 80 percent correct on any given concept received

further instruction. The post-interim assessment re-teaching was a school-wide event, with different teachers addressing different topics, and groups of students going to the different teachers according to their needs. The initiative involved a great deal of professional development, and was piloted at one school within the district. By implementing this intervention program, the school moved from having 45 percent of their fifth-grade students demonstrating mastery on the statewide assessment in math in 2002 to 61 percent in 2003, 76 percent in 2004, and 97 percent in 2005 (Sassone, 2007). The district has expanded the pilot program and continues to achieve an extremely high level of success.

A second approach to improving communication between the classroom and school/district level is a bit simpler, and was implemented by a district in 2007-08. This district requested that ATI provide a five-item formative assessment for each learning standard in addition to the periodic interim assessments. The formative assessments were made available to all teachers and they were encouraged, though not required, to use them to assess student learning at the end of each unit. Since the items were drawn from an item bank that mirrors the interim item bank in terms of item characteristics, quality, and so on, the results from the formative assessments can be expected to indicate whether students would demonstrate mastery on the interim assessments. District-wide interim assessments were also administered three times during the year. When the high-stakes statewide assessment was administered, all grades except for one showed an increase in the percent of students demonstrating mastery in math and in reading over the previous year. In many cases the increase was by three to five percent.

A third approach comes from a client that has just begun using Galileo K-12 Online this year. Their approach to assessment is unusual in that, instead of administering three interim assessments across the year, they are administering three sets of pre- and post-interim assessments each year. A district-wide pre-assessment is administered at the beginning of the trimester, and is followed by a district-wide post-assessment at the end of the trimester that assesses the same material. The idea is that the pre-assessment should serve as a guide for instruction during the trimester, and the results should be evident in student scores on the post-assessment at the end of the trimester. The effect of this creative approach on student performance on the statewide assessment remains to be seen, but the students are certainly improving from the pre-assessment to the post-assessment. If nothing else, the pre-assessment serves as a clear communication from the district to the teachers regarding what material is expected to be covered in the coming trimester and to what degree of depth.

## **IX. Summary and Conclusion**

The mandate of accountability under NCLB is an important step in the improvement of student learning in the United States. Accountability, however, is only one half of the intervention conversation. If intervention efforts are going to succeed, the conversation must be an exchange of both bottom-up and top-down information at all levels of the intervention system. At some points of the intervention system the conversation flows smoothly, but at other points schools and districts must take determined steps to make sure all parties are getting the information they need. One such juncture is the top-down information from the state to the district. Data are

available from statewide assessments that can inform intervention at the district level, but the limitations of this source of information must be recognized and steps must be taken to make sure that it is disseminated to teachers. The intervention conversation between individual teachers and the school/district can be particularly difficult, because their needs, and the level of data that is relevant to each, differ. District-wide interim assessment can act as a bridge at this level, providing teachers with detailed feedback and pointers regarding specific schools and students, and districts with valid, reliable, aggregated data that can forecast student performance on high-stakes statewide assessments. In addition, technology can be used to improve the bottom-up communication between teachers and schools/districts by allowing them to share continuously the status of the progress through the curriculum within their classroom.

A common complaint about accountability initiatives under NCLB is that too much emphasis is being placed on testing. The improvement of student learning cannot take place without some form of formative assessment data that identifies gaps in student understanding and provides pointers for improvement. This is something teachers have always relied on, long before the mandates associated with NCLB. The task for global improvement of student learning is to expand the management cycle to all levels of aggregation. Most of the mechanisms to make this happen are already in place, especially for districts that engage in interim assessment. What is suggested here is not more testing, but a more effective coordination of assessment and intervention efforts. It is hoped that the ideas expressed in this paper will serve to identify gaps between the current state of district-wide intervention efforts and the goal of improved student learning, and that it will provide detailed pointers toward an efficient and effective intervention system.



## X. References

- AERA, APA, & NCME (1999). *Standards for educational and psychological testing*. Washington, DC. American Educational Research Association.
- Baroudi, Z.M. (2007). Formative assessment: Definition, elements, and role in instructional practice. *Post-Script: Postgraduate Journal of Education Research*, 8, 37-48. Retrieved January 16, 2009 from [http://www.edfac.unimelb.edu.au/research/resources/student\\_res/pscript\\_past.html](http://www.edfac.unimelb.edu.au/research/resources/student_res/pscript_past.html)
- Bergan, J.R., Bergan, J.R., Burnham, C.G., Bergan, K.S., and Feld, J.K. (2008a) *Dynamic intervention systems: integrating research and management to support learning in standards-based education*. (Available from Assessment Technology Inc., <http://www.ati-online.org>)
- Bergan, J.R., Burnham, C.G., Bergan, J.R., and Bergan, K.C. (2008b). *Benchmark assessment development in the Galileo educational management system*. (Available from Assessment Technology Inc., 6700 E. Speedway Blvd., Tucson, AZ 85711)
- Black, P. and Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: principles, policy and practice*, 5, 7-74.
- Butler, R. & Neuman, O. (1995). Effects of task and ego-achievement goals on help-seeking behaviours and attitudes. *Journal of Educational Psychology*, 87, 261-271.
- Chappuis, S. & Chappuis, J. (2007). The best value in formative assessment. *Educational Leadership*, 65, 14-19. Retrieved January 15, 2009 from [http://www.ascd.org/publications/educational\\_leadership/archived\\_issues.aspx](http://www.ascd.org/publications/educational_leadership/archived_issues.aspx)
- Elawar, M.C. & Corno, L. (1985). A factorial experiment in teachers' written feedback on student homework: changing teacher behaviour a little rather than a lot. *Journal of Educational Psychology*, 77, 162-173.
- Fontana, D. & Fernandes, M. (1994). Improvements in mathematics performance as a consequence of self-assessment in Portuguese primary school pupils. *British Journal of Educational Psychology*, 64, 407-417.
- Haladyna, T.M. (2004). *Developing and validating multiple-choice test items*. Mahwah, NJ. Lawrence Erlbaum Associates, Inc.
- Johnston, P., Guice, S., Baker, K., Malone, J. & Michelson, N. (1995). Assessment of teaching and learning in literature-based classrooms. *Teaching and Teacher Education*, 11, 359-371.
- Kluger, A.N. & DeNisi, A. (1996). The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, 119, 254-284.

- Martinez, J.G.R. & Martinez, N.C. (1992). Re-examining repeated testing and teacher effects in a remedial mathematics course. *British Journal of Educational Psychology*, 62, 356-363.
- McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: Part 1. An account of basic findings. *Psychological Review*, 88, 375-407.
- National Mathematics Advisory Panel (2008). *Reports of the task groups and subcommittees*. Washington, DC: U.S. Department of Education. Retrieved January 16, 2009, from <http://www.ed.gov/about/bdscomm/list/mathpanel/reports.html>
- No Child Left Behind Act of 2001, 20 U.S.C. 6311. Retrieved Jan. 22, 2009, from <http://www.ed.gov/policy/elsec/leg/esea02/index.html>
- Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007, November). The role of interim assessment in a comprehensive assessment system: A policy brief. Retrieved January 12, 2009, from <http://www.nciea.org/index.php>
- Sassone, J. (2007, February). *Success with standards: Standards-based strategies for improving student performance*. Paper presented at the Innovative Technology Advancing Standards-Based Education workshop, Tucson, AZ.
- Shepard, L. (2006). Classroom assessment. In R. L. Brennan (Ed.), *Educational measurement*. (4<sup>th</sup> ed.). Westport, CT: Praeger Publishers.
- Spitzer, H.F. (1939). Studies in retention. *The Journal of Educational Psychology*, 30, 641-656.