White Paper

Dynamic Intervention Systems:

Integrating Research and Management to Support Learning in Standards-Based Education

by
John Richard Bergan, Ph.D.
John Robert Bergan, Ph.D.
Christine Guerrera Burnham, Ph.D.
Kathryn S. Bergan, Ph.D.
and Jason K. Feld, Ph.D.

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

Dynamic Intervention Systems: Integrating Research and Management to Support Learning in Standards-Based Education

By John Richard Bergan, John Robert Bergan, Christine Guerrera Burnham, Kathryn S. Bergan, and Jason K. Feld, Assessment Technology Inc.

Table of Contents

Assessment Technology, Inc., Publishers
Tucson, Arizona, U.S.A.
Printed in the United States of America.
V2-071508

Acknowledgements

The authors wish to thank Charles Brainerd for his review and helpful comments, which led to improvements of the manuscript. The authors would also like to thank Jody Jepson for her assistance in the review and preparation of this paper.

John R. Bergan, Ph.D.
President, Assessment Technology Inc.

I. Introduction

We live in a time and place in which education is being called upon to assist students to achieve new levels of learning in order to compete in a rapidly changing global society. Educators are being asked to intervene in the educational process in ways intended to produce new levels of achievement across extended time spans. Educators across the nation have turned to standards-based education as a tool for increasing student learning. Intervention is an essential component of standards-based education. The standards-based approach is rooted in the education reform movement, which is based on the view that the American educational system must increase student learning beyond current levels in order for the nation to remain competitive in our rapidly changing world (e.g., Ravitch, 2001). The need to increase learning beyond currently expected levels calls for interventions that will bring about the desired improvements in educational outcomes. The fundamental assumption underlying this paper is that the increases in learning called for in the educational reform movement will require dynamic intervention systems comprised of interrelated research and management components supporting continuous educational change. Continuous change is demanded not only by the need to increase learning, but also by the need to address the rapid advances in knowledge and technology that are the hallmark of the 21st century.

Research providing credible evidence on intervention effectiveness in promoting learning is needed to insure that selected interventions are likely to produce desired outcomes. In addition, data is required to inform management decisions to ensure that interventions are yielding desired outcomes when they are actually implemented in an educational setting. It is important to note that there are potentially competing reasons for gathering research and management data. The reason for gathering research data is to determine intervention effectiveness. The way in which the intervention is implemented is not affected by the data. The research requirement is that the intervention be faithfully implemented in accordance with predetermined specifications. By contrast, the primary reason for gathering data in intervention management is to obtain information to guide decisions designed to promote the achievement of intervention goals. In some cases, decisions may be made that alter the intervention while it is occurring. Alterations in intervention implementation pose a serious threat to the validity of an experiment because such alterations compromise the fidelity of the treatment.

The task of providing effective interventions in educational settings requires resolution of the potentially competing research and management reasons for gathering data. We argue the resolution can be achieved through the development of intervention systems that integrate research and management components to support the continuous attainment of new levels of learning over an extended time span.

It goes without saying that any initiative to improve student learning requires highly skilled professional educators, and parents, and communities committed to the advancement of learning. These requirements are needed to support the efforts inherent in every child to construct new knowledge from experience. The additional need for dynamic intervention systems integrating research and management components arises from the fact that education in the 21st century requires coordinated initiatives in which groups of individuals work together toward the achievement of shared educational goals in the context of a rapidly changing world. We make no claim that

such a system will produce the desired outcome of continuous advances in achievement. That question must be addressed empirically. What we do claim is that such a system will provide essential information and tools supporting the efforts of schools, parents, and communities to assist children to achieve their potential in the global economy of the 21st century.

II. Intervention Research

Intervention research has been a topic of considerable interest and debate in recent years (Phye, Robinson, & Levin, 2005). Much of the discussion has centered on the need for an evidence-based approach to educational practice. The Federal government has been particularly supportive of an evidence-based approach to the implementation of educational interventions (e.g. Reyna, 2005). Unfortunately, the educational research community has not responded to the need for research aimed at identifying effective interventions. In an address to the American Educational Research Association, Grover J. Whitehurst, Director of the Institute of Education Sciences, lamented the lack of experimental research that could inform educational practice. In addition, Dr. Whitehurst presented evidence that non-experimental research has failed to provide information that could effectively inform instruction (Whitehurst, 2003).

A. The Decline of Experimental Research

Recently a survey of publications in major educational research journals supported Whitehurst's major contentions. This study reported not only a lack of current experimental research, but also a significant decline in experimental studies over recent decades (Hsieh, Acee, Chung, Hsieh, Kim, Thomas, You, and Levin, 2005).

The mandate to increase learning combined with the lack of experimental research to inform instruction leaves local education agencies in a highly vulnerable position. Under the No Child Left Behind Act of 2001, 2002, schools are responsible for increasing learning as measured by the mastery of standards. However, research informing instruction to increase standards mastery is in short supply. Moreover, even in those instances in which an intervention has been shown to be effective in an experimental setting, it may not be effective when implemented in a particular school setting.

B. Reasons for the Decline

Many reasons may be offered to explain the observed shortage of experimental studies that could inform educational interventions. Hsieh and her colleagues suggest the possibility that given the increased popularity of qualitative research and the post-modern relativist view, some researchers may have rejected the underlying assumptions of experimental research. Hsieh and her associates also cite the rigorous methodological standards, practical constraints, and resource requirements of experimental research as potential sources for the observed decline in experimental studies (Hsieh et al., 2005). For example, the need for random assignment to experimental conditions may contribute to the reduction in the incidence of experimental studies. Given the mandate to leave no child behind, it may be difficult to make a convincing argument that students should be assigned to a control condition in which they will be prohibited from reaping whatever instructional benefits may accrue from an

experimental treatment. Resource requirements almost certainly affect the incidence of experimental studies, particularly when a study involves a large number of students and occurs over an extended time span. These conditions are especially likely to limit the experimental contributions of scholars early in their careers because at that time scholars lack the necessary track record to attract the large amounts of funding required for the conduct of large-scale experimental research.

C. The Role of Short Experimental Studies in Standards-Based Education

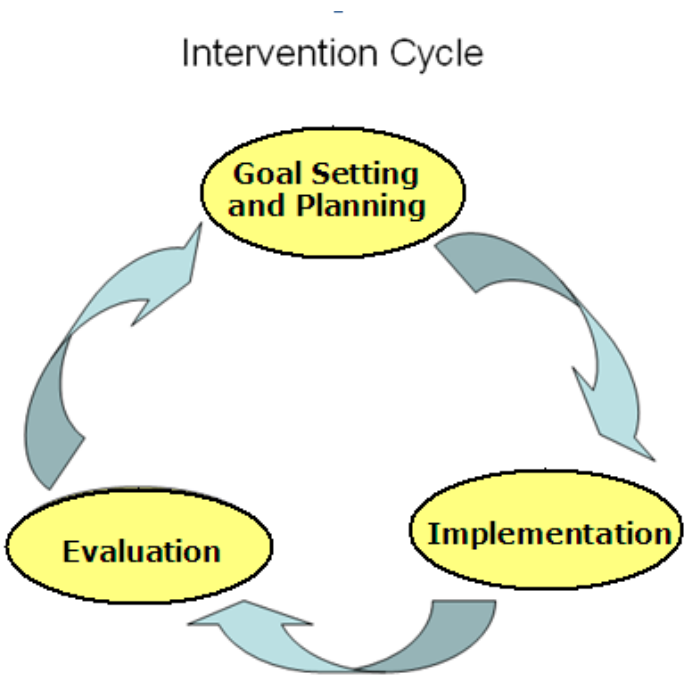
Not surprisingly, Hsieh and her associates point out that much of the experimental research published in educational journals is conducted on small samples of students over time spans lasting less than one day (Hsieh et al., 2005). Although Hsieh and her associates see this finding as a limitation negatively affecting research quality, it is important to point out that to a large extent the great wealth of scientific information that is available to guide instruction has come from short experimental studies conducted with small samples of subjects. Moreover, many of the most important findings in fields such as learning, development, cognition, and memory, which provide the foundation for the science of instructional design, have come directly from studies of this kind. Finally, there is every reason to believe that short studies will continue to make significant contributions to instruction. For example, recently a compelling case has been as to the likelihood that experimental research in the field of neuroscience can be useful in informing instruction (Varma, McCandliss, & Schwartz, 2008). Studies in this field often occur in a short time period.

The short experimental study has a number of advantages that make it ideally suited for use in the development and implementation of evidence-based interventions in the context of standards-based education. Short studies are inexpensive to conduct. As a consequence, they can be implemented by experienced scholars and by young scholars who do not have access to large amounts of funding. They are easy to conduct under highly controlled conditions. The problem of insuring the fidelity of treatment implementation is greatly simplified. Because they require little time to implement, the necessary peer review process can be implemented within a shorter time frame than is the case for an extended experiment. As a result, findings can be made available more quickly than is the case with a large-scale study. Replications and adaptations to accommodate potential contextual effects can be accomplished quickly. Thus, in many instances information involving support for the generality of findings is easier to obtain than would be the case in a long study. Typically, short experiments are implemented by an experimenter working with one subject (student) at a time. This approach eliminates dependencies among subjects that may exist when experiments are conducted in groups within the classroom (see, for example, Raudenbush, 2008). The control conditions in short experiments are generally more precisely designed than is the case in extended instructional experiments. The control condition in extended instructional experiments is often what is described as standard instruction. Standard instruction is a moving target. It may vary from one situation to the next. Consequently, in those cases in which there is a significant finding favoring an experimental treatment, it may be difficult to evaluate the benefit of the treatment. Finally, short experiments can be conducted in the context of a standards-based educational initiative without disrupting necessary management activities needed to maximize the likelihood of intervention success.

The discussion of benefits associated with short experiments does not imply that there may not also be important benefits associated with large-scale experiments. For example, Raudenbush (2008) has argued for the value of large-scale experiments in informing educational policy. As more data from large-scale experimental studies becomes available, the utility of large-scale experiments for informing policy will become clearer than is now the case. What is indisputably clear is that short experiments have made an enormous contribution to the scientific underpinnings of instruction in the past, and there is no reason to believe that they will not continue to do so in the future.

III. Managing Standards-Based Educational Interventions

Intervention management in standards-based education may be defined as a cyclical process. The first phase of the process is goal setting and planning. The second is implementing the intervention, and the third is evaluating the intervention. The figure below illustrates the management cycle. Each cycle begins with goal setting and planning. The goal setting and planning phase is followed by implementation and evaluation. Evaluation provides information to inform planning in the next cycle.



The process is goal directed meaning that interventions are designed to achieve specific objectives. For example, an intervention may be designed to enable at-risk students to master a given set of standards. The process is also data driven meaning that data analyzed during the evaluation phase of the cycle informs the next iteration of the planning process.

The discussion that follows outlines the activities associated with the three phases of the intervention cycle and describes the ways in which intervention research can be integrated into the management process.

A. Intervention Goal Setting and Planning

Goal setting and planning begins with the identification of students participating in the intervention. In some cases, an intervention may involve all students in a district or in a particular grade in a district. In other instances, the intervention will be targeted toward a subset of students. In cases in which the intervention is designed for a subset of students, the selection of students may be based on prior test performance. For example, an intervention might be planned for students identified through an assessment as being at risk for not meeting standards.

Intervention planning requires the specification of procedures for tracking student participation as well as for identifying which students are participating in the intervention. Tracking is particularly important for identifying intervention attrition and for identifying students not included at the start of the intervention who subsequently become intervention participants.

When students have been identified and procedures for tracking specified, the objectives to be achieved through the intervention are addressed. Intervention goals are generally a subset of a larger set of goals reflecting the major objectives to be achieved through instruction. The larger set is often set forth in a district pacing calendar, which indicates the objectives targeted for instruction during a series of specified time periods. For example, district pacing calendars often divide the school year into quarters and outline the objectives to be covered during each quarter. Intervention goals are selected from the larger set included in the pacing calendar.

After goals have been selected, instruction aimed at promoting goal attainment is planned. Planning involves the specification of lessons, assignments, tutorials, and other forms of instruction designed to promote standards mastery. Online technology can play an important role in the development of instructional plans. For example, online resources can be used to target instruction at the mastery of specific standards. In some cases the planning process itself can be implemented online. For instance, users can develop online lessons or assignments to address specific objectives targeted for instruction in an intervention (Bergan, Bergan, & Guerrera, 2007).

Effective instructional planning can be facilitated by the inclusion of information on the quality of instructional procedures to be implemented to achieve intervention goals. Several kinds of ratings may be useful in this respect (Bergan, Bergan, & Guerrera, 2007): The first is a teacher rating of the instruction. The second is a student rating of the instruction. The third is the percentage of students receiving instruction who displayed mastery of the standards that the instructional was designed to address. The fourth indicates the number of teachers who implemented the instruction during a given time period, and the fifth is experimental evidence supporting the effectiveness of the instruction in promoting standards mastery.

B. Intervention Implementation

The major management problem to be addressed in implementing an intervention is to determine the extent to which the intervention is actually implemented as planned. This problem is directly analogous to the problem of ensuring treatment fidelity in intervention research. As Hsieh and her colleagues have noted, evidence

regarding the fidelity of treatment implementation is rarely available in experimental research (Hsieh et al., 2005). It is safe to assume that information on the fidelity of intervention implementation is even rarer in practice than it is in research. One obvious reason for this is that schools lack the necessary technology to monitor intervention implementation effectively.

Electronic educational management systems provide a number of options for estimating the fidelity of intervention implementation. One widely used approach involves curriculum maps that indicate the number of lesson plans and/or assignments targeted at intervention objectives. Indicators obtained from curriculum maps can be used to estimate the relative emphasis given to intervention objectives in instruction. The shortcoming of this approach is that it is difficult to know the extent to which these indicators accurately reflect what has actually occurred in instruction. This is the case because the maps do not detail the instructional experiences of students. When instruction occurs online, it is possible to produce a permanent record of teacher and student interactions (Bergan, Bergan, & Guerrera, 2007). This information can be used to estimate the fidelity of intervention implementation.

Short formative quizzes may play a role establishing the fidelity of intervention implementation. For example, suppose that a teacher assigns students to go to a particular website that contains math problems of a particular type. Typically, it will not be easy to produce a permanent record of student instructional experiences at the website. However, it is possible to administer a short formative assessment covering the standards addressed on the website. The fact that students have participated in the assessment provides an indication that there has been exposure to instructional content covered on the website.

We have listed some of the ways in which the fidelity of intervention implementation can be monitored. There are, of course, many other possibilities. A number of factors are important to consider in establishing an effective monitoring approach. Monitoring the fidelity of intervention implementation is essential to effective intervention management. In the absence of information regarding whether or not an intervention has actually been implemented, it is not possible to make informed decisions to promote the achievement of intervention goals.

Because monitoring is essential to effective management, it must be a continuous component of the management process. Yet, at present it is safe to say that monitoring is rarely implemented. Monitoring typically requires a significant level of effort, and in many cases resources are not readily available to support monitoring. One of the goals of management technology is to minimize the effort required for monitoring. In the ideal case, monitoring data should be provided automatically as part of the intervention process. The automatic provision of monitoring data will require new technology linked to instructional resources used in interventions. For example, when a school adopts a particular set of curricular materials, those materials will need to include technology to provide data on the use of the materials. Technology for monitoring online instruction mentioned earlier provides one example of how monitoring data can be linked to the provision of instructional resources.

C. Intervention Evaluation

The most important question to be answered in evaluating an intervention is whether or not intervention goals have been attained. At the end of the school year, this is easy to do, but the only corrective action possible in the event that goals have not been attained is action that will take effect the following year. End-of-year evaluations are available based on performance on statewide assessments. While statewide test performance provides useful summative information, it is not useful for guiding instructional decisions (Perie, Marion, Gong, & Wurtzel, 2007). Determining goal attainment during the course of the year provides opportunities for immediate corrective action, which is obviously important for effective intervention management. In standards-based initiatives, interim evaluations are typically conducted using formative and benchmark assessments.

D. Formative Assessment in Standards-Based Interventions

Formative assessments are generally short assessments given by teachers for the purpose of guiding instruction related to a limited set of objectives. For example, a teacher might administer a five- or ten-item quiz to determine the extent to which students have mastered a particular performance objective. Formative assessments are generally much too short to be reliable or valid, and the reliability and validity of these assessments is typically not assessed. It is generally assumed that the teacher will have many opportunities to informally assess student achievement related to a particular set of objectives. When this assumption is justified, the inconsistencies associated with unreliability assume less importance than they otherwise might.

Formative assessments can play an important role in standards-based interventions. For example, they can be used to assess student mastery of standards to be included on an upcoming benchmark assessment. The results then can be used to plan instruction to promote mastery of those standards. Formative assessments are also useful following a benchmark assessment. Formative assessments can be used following a benchmark test to assess the results of enrichment and/or re-teaching initiatives designed to provide targeted instruction for students at risk of not meeting standards on the statewide test.

In some cases formative tests can be combined with other formative or benchmark assessments. The new combined assessment may be used along with other evidence to assess student progress and forecast student risk of not meeting standards.

E. Benchmark Assessment in Standards-Based Interventions

Benchmark tests are assessments designed to measure the achievement of standards that have been targeted for instruction by a district. They are generally district-wide measures. However, in some cases they may be administered to selected groups or schools within a district. Benchmark assessments are interim measures of standards mastery, designed to be related to subsequent statewide assessments, which ultimately determine overall standards mastery. Insofar as benchmark tests are intended to reflect what is being taught in the district, they are generally customized to reflect the district curriculum. Since benchmarks are used as estimates of probable

standards mastery on statewide tests, they are subject to reliability and validity requirements not associated with short formative assessments.

Benchmark assessments serve four major functions in the evaluation process. First, they can provide information on the mastery of standards targeted for instruction during specific time periods in the school year. Second, they can provide guidance as to which standards should be targeted next to promote further learning. Third, they can be used to estimate the probability of standards mastery on statewide assessments. Fourth, they can be used to measure progress toward standards mastery. This information can be used to adjust intervention instruction in cases in which students are not making adequate progress in meeting standards.

F. Mastery of Standards Targeted for Instruction

Determining the mastery of standards on a benchmark assessment is typically accomplished by constructing a test containing multiple items for each standard assessed. For example, a benchmark test could be constructed containing four items for each assessed standard. The next step is to establish cutoff points reflecting varying levels of mastery. These cutoff points are arbitrary. For example, a district may define the cut point for standards mastery as three correct responses out of four. This approach is attractive because it is easy to understand. However, in general a better approach is to use Item Response Theory (IRT) to estimate the true score for the subset of items (see, for example, Lord, 1980). The cut off points are then applied to the estimated true score rather than to the observed score. It is also possible to use an augmented scoring approach, which takes advantage of information from the entire test in estimating the score for the subset of items of interest (Thissen & Wainer, 2001). This approach has the advantage of increasing the reliability of estimated scores.

G. Determining Which Standards to Target in an Intervention

Benchmark tests can be particularly useful for determining what to teach next. Benchmark results can be used to determine the probability that students of a given ability level will be able to master a given set of standards. This information can be used to guide instruction. The estimated probability of standards mastery for students of a given ability level may be computed using IRT. The IRT estimate is based on ability level and characteristics of the items used to assess the standards.

The estimated probabilities of standards mastery for a set of standards can be used to guide instruction for an individual student. For example, in providing individualized instruction, a teacher might initially choose to target the standard that the student would be most likely to be able to master. Instruction in more difficult standards would follow. Estimated probabilities of standards mastery can also be used in planning instruction for groups of students. For example, a teacher might choose to plan instruction for a group of students at risk for not meeting state standards. In this case the selection of standards to be targeted for instruction would be based on the average ability of the group and characteristics of the items used to assess standards mastery.

H. Forecasting Standards Mastery on Statewide Tests

One of the most important functions of benchmark tests is to forecast standards mastery reflected in statewide test performance. For example, suppose that a group of students have met the standard on three benchmark math tests. It would be useful to know the likelihood that those students would go on to meet the standard on the statewide test. Forecasts based on benchmark test performance can provide information on the probability of standards mastery on statewide tests. Without forecasting information, school personnel are left to guess whether or not their teaching is likely to lead to successful statewide test performance.

In order for forecasting information to be most useful, it must be based on what is currently being taught. This point can be illustrated by considering cases in which forecasts are not based on assessments of standards currently targeted for instruction. For example, previous statewide test performance could be used to forecast subsequent statewide test performance. However, the value of the forecasting would be limited because the results could not be used to determine what to teach to promote standards mastery on the subsequent statewide test. By contrast, benchmark tests used in forecasting do provide information about what to teach next.

In order for benchmark assessments to be effective in forecasting standards mastery, they must be reliable and valid. Reliability is essential to insure consistency in measuring student achievement. When reliability is low, measurement error is high. When this is the case, test scores can yield misleading information.

Establishing the validity of a benchmark assessment involves assessing the relationship between the assessment and a criterion measure such as performance on a statewide test. If the benchmark test is to be useful in forecasting the mastery of standards, it needs to be related to the statewide test ultimately used to determine standards mastery.

Although benchmark tests should generally be related to the statewide test, it is not expected that they be equated to the statewide test. The purposes of benchmark assessment and statewide assessment are different. Benchmark assessments are administered periodically during the school year to guide instruction. The standards assessed on a benchmark test are typically selected to match standards targeted for instruction during a particular time period. By contrast, statewide tests are summative assessments generally administered toward the end of the school year for accountability purposes. The content of a statewide test is not intended to reflect instructional goals targeted for instruction at a particular time.

The fundamental forecasting question of interest in forecasting statewide test performance is that of identifying a student's risk of not achieving standards mastery on the statewide test given his or her performance on benchmark assessments. We have repeatedly found that risk forecasting is sufficiently accurate to provide useful information for guiding instruction. This is illustrated in the table below, which shows data collected by Assessment Technology Incorporated during the 2006-2007 school year.

TABLE 1
Risk of not Meeting State Standards Given Varying Patterns of Benchmark Standards Mastery

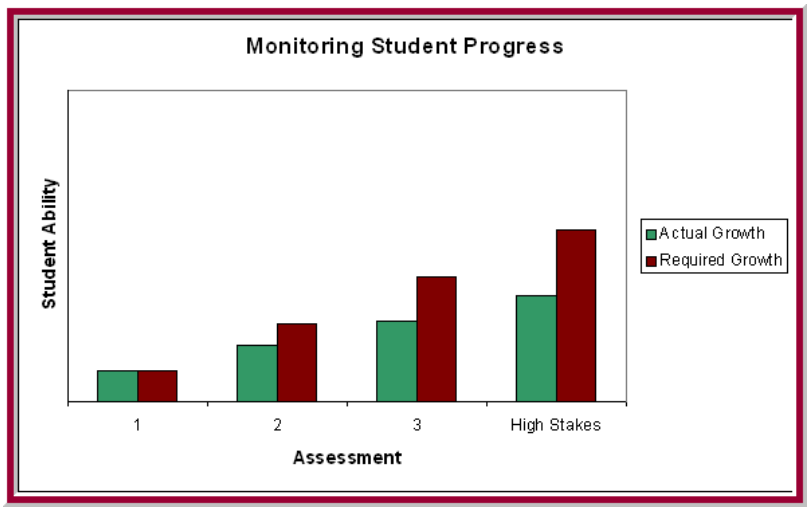
Sample District 2006-07 06th Grade Math						
Time Frame:	**Benchmark Assessments**			**Number of**	**AIMS**	
After…	**1**	**2**	**3**	**Students**	**Met**	**Not Met**
Benchmark 1	Met	?	?	732	0.89	0.11
1378	Not Met	?	?	646	0.27	0.73
Benchmark 2	Met	Met	?	509	0.95	0.05
1283	Met	Not Met	?	161	0.65	0.35
	Not Met	Met	?	148	0.66	0.34
	Not Met	Not Met	?	465	0.16	0.84
Benchmark 3	Met	Met	Met	450	0.98	0.02
1239	Met	Met	Not Met	45	0.73	0.27
	Met	Not Met	Met	77	0.84	0.16
	Met	Not Met	Not Met	77	0.47	0.53
	Not Met	Met	Met	78	0.87	0.13
	Not Met	Met	Not Met	65	0.42	0.58
	Not Met	Not Met	Met	51	0.57	0.43
	Not Met	Not Met	Not Met	396	0.11	0.89
	Correctly Classified Overall:			**0.86**		

I. Measuring Progress toward Standards Mastery

Benchmark assessments can play an important role in measuring progress toward standards mastery. Measuring progress can be achieved by placing scores from benchmark assessments administered over time on a common scale. This can be accomplished using scaling procedures based on IRT. The measurement of progress is useful for assisting districts to determine the amount of progress students are making during the school year. This information is obviously useful in instructional planning.

A question of special importance in standards-based education is whether or not the amount of progress is sufficient to lead to an increase in the number of students meeting Adequate Yearly Progress (AYP) targets. The answer to this question requires a determination of how much progress is required to increase the number of students meeting AYP measurable objectives. For example, suppose that a district has the goal of increasing the number of students meeting AYP measurable objectives by 10%. For the sake of simplicity, assume that growth is linear. Consider the hypothetical progress data shown in the figure below. The red bars show linear growth required to produce a 10% increase in the number of students meeting AYP measurable objectives based on

statewide test performance. The green bars indicate actual growth. Both indices show growth. The problem is that the actual growth is not sufficient to increase the number of students meeting AYP objectives by 10%. Unless the district knows in advance how much growth is needed to achieve the goal, district staff can easily be misled into thinking that an intervention is going well when in fact the amount of growth is inadequate to meet district goals.



The amount of required growth can be effectively estimated when prior growth data are available, when cut points for statewide assessments are stable, and when classification errors related to statewide test performance are within acceptable limits. It is important to recognize that cut-point stability and classification accuracy related to statewide assessments may not always be adequate to support forecasts of changes in the attainment of AYP measurable objectives (Betebenner, Shang, Xiang, Zhao, and Yue, 2008).

When required conditions are met, categorical data analysis procedures can be implemented that provide easily interpretable forecasting results. The following table displays a simple approach for achieving the desired information. The table shows changes in standards mastery and non-mastery for two benchmark assessments. The cells in the table indicate the number and percentage of students who:

- maintained mastery,
- failed to maintain mastery,
- remained at a non-mastery level,
- moved from non-mastery to standards mastery.

TABLE 2
Mastery Classification Changes for Two Benchmark Assessments

		Benchmark 2 Same	Benchmark 2 Different	
Benchmark 1	**Pass (87)**	83	4	5% *did not maintain* mastery
	Fail (59)	17	42	71% *acquired* mastery

The sum of the percentage of students who maintained mastery and the percentage of students who moved from non-mastery to mastery provides an estimate of the percentage of students likely to meet standards on the statewide test. If the goal is to exceed performance from the previous year, that sum should be higher than the percentage of students who met standards on the statewide test during the previous year

The assumption that standards mastery is increasing under a given intervention can be assessed by comparing the number of students who failed to maintain mastery to the number of students who moved from non-mastery to mastery. The hypothesis that standards mastery is increasing is supported if significantly more students move from non-mastery to mastery than those who fail to maintain mastery. This hypothesis can be tested using the chi-square test of equiprobability for the two cells in question. Note that this test is not intended to address the question of whether or not the intervention causes an increase in standards mastery. However, a positive finding would be consistent with the assumption that the intervention caused the change. Nonetheless, other factors occurring at the same time as the intervention might also be responsible for the desired increase in standards mastery.

J. Evaluation and Intervention Adjustment

The primary purpose of the evaluation phase of the management cycle is to determine whether or not there is a need to make intervention adjustments to promote the achievement of intervention goals. Formative and benchmark assessment results provide the information required to determine the need for adjustments. Formative assessments indicate the extent to which specific standards targeted for instruction have been learned. When learning has not occurred as planned, adjustments are warranted. Adjustments based on formative assessment results are typically instituted by the teacher in the classroom setting. For example, assignments may be given to address specific objectives not yet mastered.

Benchmark results provide information not available through short formative assessments. Results not obtainable through short formative tests include information regarding what to teach next, Information on students at risk for not meeting standards on statewide tests, and information on student progress. Benchmark results may signal the need for adjustments initiated by teachers. For example, a teacher might individualize instruction based on benchmark information indicating what to teach next for a student of a given ability. However, they may also call for adjustments initiated at administrative levels. For example, a benchmark result identifying students in multiple

classes at risk for not meeting standards might call for a re-teaching intervention initiated for students in those classes.

IV. Integrating Intervention Research and Intervention Management

The utility of intervention research in standards-based education can be enhanced by procedures that integrate short experimental studies into online systems for the management of standards-based educational initiatives. Integration provides a mechanism for directly linking research findings to their application in standards-based initiatives. The well-documented lengthy time lag between publication of findings and their application in educational settings is virtually eliminated by integration. Integration also supports evaluations of the extent to which an intervention supported by research is faithfully implemented in the classroom. Information of this kind is critical to evaluating the scalability of an intervention.

Intervention research differs from intervention management in a number of important respects: The goal of intervention research is to provide evidence regarding treatment effects. By contrast, the goal of intervention management is to insure that students master standards targeted for instruction. Data collected in intervention research provides evidence as to treatment effects. Data collected in intervention management indicates whether or not intervention goals are being achieved. Intervention research requires that the treatment of interest be faithfully implemented during the time the experiment is in effect. Intervention management is cyclical. Successive cycles may lead to intervention modifications. Intervention research typically occurs in a short time period (Hsieh et al., 2005). Intervention management typically covers an extended time span. Intervention research calls for random assignment of experimental units (e.g. students, or classes) to experimental conditions. Intervention management does not require random assignment. Indeed, the goals driving the management process often militate against randomization. The many differences between research and management underscore the need for an organizational structure that accommodates both and makes it possible for management initiatives to support research and for research to be carried out in ways that enhance the effectiveness of management.

The integration process described here is not intended to supplant the standard approach by which experimental research has informed educational practice. In the past the anointed road from research to practice involved the following steps: Research studies were carried out and published in peer-reviewed journals. Interventions were developed on the basis of research findings. Applied experiments were conducted to evaluate the interventions. These studies were also published. Organizations such as the federally funded educational laboratories disseminated the findings to schools. The traditional road from research to practice will, no doubt, continue to be followed. In some cases, the extent to which research can be integrated directly into intervention management will be minimal. For example, as indicated earlier, a convincing case can be made that research in neuroscience has significant implications for educational practice. However, it is generally impractical to carry out research of this kind in the context of an educational setting. For example, the data in a neuroscience experiment often comes from magnetic resonance imaging. The technology required to produce this type of data is not likely to found in a school setting. In those instances in which direct

integration of research and practice is impractical, the traditional approach for informing educational practice based on research will continue to play an important role in serving the needs of the nation's schools.

The integration of intervention research and management can be facilitated at each of the three stages in the management process. The discussion that follows illustrates integration during the planning, implementation, and evaluation phases of intervention management.

A. Integration and Intervention Goal Setting and Planning

The initial phases of the planning process are marked by the specification of goals to be achieved through instruction. In standards-based initiatives, these goals are typically expressed in terms of the achievement of state standards. Perhaps the most basic way to integrate intervention research and management is to design short experimental studies that have direct implications with respect to the achievement of specific standards targeted for instruction. Studies of this kind ensure that research will have direct relevance for instruction.

The link of research to specific standards is not intended to imply that the implications of any particular experiment are limited to the specific standard(s) to which the research has been related. Clearly, research on topics such as learning, memory, cognition, or motivation may be associated with a broad range of standards. The purpose of relating specific experiments to specific standards is to facilitate, where possible, the direct application of the research in practice.

The planning process includes the specification of lessons and assignments to be implemented to achieve instructional goals. Instructional procedures assessed through experiments should generally be short enough to support instruction occurring during a short time period. For example, it should be possible to implement a new instructional period within a single lesson or assignment. The major benefit of the criterion is that it supports rapid application of experimental findings in instructional initiatives.

B. Integration and Intervention Implementation

Integration can be supported during plan implementation in two ways: First, experiments can be designed to emulate the manner in which instruction occurs in the standards-based educational environment. Instruction is typically rendered through lessons implemented during class periods and through assignments that can be carried out either in class or through homework. In re-teaching interventions, instruction may be rendered to special groups of students or through individual tutorials. In some cases, it may be appropriate to design experimental studies to reflect the various approaches to instructional implementation. When this is feasible, instructional procedures addressed in an experiment may be transferred directly into instructional practice.

The second approach to integration during implementation involves implementation monitoring. As indicated earlier, monitoring is essential to ensure the fidelity of implementing an experimental treatment. Monitoring is essential in intervention management to ensure that the planned intervention is actually being implemented. Inconsistent intervention implementation is among the greatest obstacles to intervention scalability. Consistency is difficult if not impossible to achieve in the

absence of monitoring information that can be used identify instances of failure to implement the intervention.

Despite its importance, there are many obstacles to the integration of research and management through monitoring. As indicated earlier, monitoring treatment fidelity is often not implemented in experimental research. The potential reasons for this are not hard to understand. Experimental research is typically carried out by well-trained graduate students who are thoroughly familiar with the experimental design. Moreover, the treatment is typically specified in great detail. Consequently, experimenters have confidence that the treatment will be implemented as designed. In addition, in cases in which monitoring occurs, the research procedures used may be impractical to implement in managing an intervention. For example, monitoring may take place through observations of video recordings of instruction.

It is safe to assume that monitoring is rarely used as a tool in the management of interventions. As already indicated, schools often lack the resources to implement effective monitoring procedures. Even in cases in which monitoring techniques are relatively straightforward, the level of effort required to implement them may be excessive. For example, a rudimentary level of monitoring can be achieved by asking teachers to fill out a questionnaire indicating what they have taught. Filling out a questionnaire does require teacher time. Entering and analyzing questionnaire data takes additional time.

There is no simple solution to the monitoring problem. Yet, it must be solved if intervention management is to be effective. It is clearly not possible to ensure effective intervention management in the absence of credible evidence regarding the extent to which the intervention is implemented. Technology for monitoring online instruction discussed previously offers one possible approach to address the monitoring problem because it provides a permanent record of instruction.

C. Integration and Intervention Evaluation

The evaluation phase of the management process includes the implementation of formative and benchmark assessment to provide information to guide instruction. Both of these forms of assessment could be used in intervention research as well as intervention management. Short formative assessments can be used to measure performance on specific standards targeted for investigation in an experiment. Benchmark assessments can be used to assess experimental effects on the mastery of specific standards, on the probability of mastering standards on statewide assessments, and on progress toward standards mastery. The benefit of integration at the evaluation phase is that it links experimental research findings directly to the measurable achievement of intervention goals.

V. Intervention Essentials in the 21st Century

This paper outlined a new vision for the integration of intervention research and intervention management. This vision assumes a dynamic system capable of supporting continuous research to inform instruction. In addition, the vision assumes the integration of research with a management process supporting continuous course corrections designed to promote the achievement of intervention goals. The effective

implementation of such a system requires that certain conditions be present. These essential conditions include:

- An intervention comprised of well-defined instructional procedures designed to promote the mastery of standards;
- An experimental research program integrated into the intervention and comprised of a series of short experiments designed to inform intervention instruction aimed at promoting standards mastery;
- A cyclical intervention management process that supports planning, implementation, and intervention evaluation;
- A set of formative and benchmark assessments designed to assess standards mastery, to forecast mastery of standards on statewide assessments, and to measure progress toward standards mastery;
- A procedure for gathering continuous credible information on the extent to which the intervention is being implemented as planned.

Achievement of these essentials will require partnerships including at a minimum schools committed to implementing interventions to improve student learning, members of the research community, and technology providers. If these partnerships are to flourish, they must be designed to thrive in the rapidly changing world of the 21st century. They must operate as dynamic systems capable of adapting to changing societal needs, changing educational standards, changing environments for delivering instruction, rapid advances in knowledge and knowledge access, and advances in technology. Schools leading the way toward a new era of instructional excellence will be at the heart of these partnerships. Members of the research community will contribute to the partnerships by designing and implementing ongoing programs of research providing the science of education for a rapidly changing society. Technology providers will supply the technological innovations needed to support a new era of education.

VI. References

Bergan, J.R., Bergan, J,R, & Guerrera Burnham, C. (2007). Curriculum Innovations: Instructional Dialogs. Tucson, AZ: Assessment Technology Incorporated.

Betebenner, D. W., Shang, Y., Xiang Y., Zhao Y., & Yue, X. (2008). The Impact of performance Level Misclassification on the accuracy and precision of percent at performance level measures. *Journal of Educational Measurement, 45*, 119-137.

Hsieh, P., Acee, T., Chung, W., Hsieh, Y., Kim, H., Thomas, G.D., You, J., Levin, J.R., & Robinson, D.H. (2005). Is Education intervention research on the decline? *The Journal of Educational Psychology, 97*, 523-529.

Lord, F.M. (1980). Applications of item response theory to practical testing problems. Hillsdale: Lawrence Erlbaum Associates.

National Center for the Improvement of Educational Assessment. (2007). A Framework for considering interim assessments. M. Perie, S. Marion, & B. Gong.

Phye, G.D., Robinson, D.H., & Levin, J. (2005). *Empirical Methods for Evaluating Educational Interventions.* San Diego: Elsevier Academic Press

Raudenbush, S. (2008). Advancing educational policy by advancing research on instruction. *American Educational Research Journal, 45*, 206-230

Ravitch, D. (Ed.). (2001) Brookings papers on education policy. Washington D.C.: Brookings Institution Press.

Reyna, V.F. (2005) The No child left behind act, scientific research and federal education policy: A View from Washington, D.C. In G. D. Phye, D. H. Robinson, & J. Levin (Eds)., Empirical Methods for Evaluating Educational Interventions (pp 29 -52). San Diego: Elsevier Academic Press

Thissen, D. & Wainer, H. (2001). *Test Scoring.* Mahwah: Lawrence Erlbaum.

Varma, S., McCandliss, B.D., & Schwartz, D.L. (2008). Diversity, group identity, and citizenship education in a global age. *Educational Researcher*, *32*, 129-139.

Whitehurst, G. (2003). A Presentation at the American Educational Research Association Conference: New wine, new bottles. Retrieved June 20, 2008 from http://www.ed.gov/searchResults.jhtml?oq=Whitehurst&odq=Whitehurst&rq=1&tx=New+wine+new+bottles&GO+-+Submit+Search.x=19&GO+-+Submit+Search.y=10