Linking Assessment and Instruction: Teacher Preparation and Professional Development

March 2010

John L. Hosp
University of Iowa

Copyright © 2010 National Comprehensive Center for Teacher Quality, sponsored under government cooperative agreement number S283B050051. All rights reserved.

This work was originally produced in whole or in part by the National Comprehensive Center for Teacher Quality with funds from the U.S. Department of Education under cooperative agreement number S283B050051. The content does not necessarily reflect the position or policy of the Department of Education, nor does mention or visual representation of trade names, commercial products, or organizations imply endorsement by the federal government.

The National Comprehensive Center for Teacher Quality is a collaborative effort of ETS, Learning Point Associates, and Vanderbilt University.
## Contents

<table>
<thead>
<tr>
<th>Overview</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Importance of Linking Assessment and Instruction</td>
<td>2</td>
</tr>
<tr>
<td>Innovation Configuration for Linking Assessment and Instruction</td>
<td>4</td>
</tr>
<tr>
<td>Teacher Preparation and Professional Development to Promote Linking Assessment and Instruction</td>
<td>5</td>
</tr>
<tr>
<td>Components of the Linking Assessment and Instruction Innovation Configuration</td>
<td>5</td>
</tr>
<tr>
<td>Fundamentals of Assessment</td>
<td>5</td>
</tr>
<tr>
<td>Standards for Comparison of Performance</td>
<td>5</td>
</tr>
<tr>
<td>Considerations for Decision Making</td>
<td>6</td>
</tr>
<tr>
<td>Assessment Procedures</td>
<td>7</td>
</tr>
<tr>
<td>Identification of Content to Teach</td>
<td>8</td>
</tr>
<tr>
<td>Identification of Student Response</td>
<td>12</td>
</tr>
<tr>
<td>Recommendations</td>
<td>14</td>
</tr>
<tr>
<td>Course Structure</td>
<td>14</td>
</tr>
<tr>
<td>Use a Variety of Practice Activities</td>
<td>16</td>
</tr>
<tr>
<td>Develop/Use a Checklist to Complete When Selecting or Administering an Instrument</td>
<td>17</td>
</tr>
<tr>
<td>Conclusion</td>
<td>18</td>
</tr>
<tr>
<td>References</td>
<td>19</td>
</tr>
<tr>
<td>Appendix. Linking Assessment and Instruction Innovation Configuration</td>
<td>23</td>
</tr>
</tbody>
</table>
Overview

Making decisions about instruction is as core a component to teaching as providing the instruction itself. When providing services to students at-risk for poor educational outcomes or students with disabilities, it is especially salient to ensure that the decisions we make have the highest likelihood of accuracy as possible and lead to improving those outcomes. The students with the greatest needs require the most accurate and effective decisions. In addition, recent increases in the need for accountability have put additional pressure on teachers to document their decisions and decision-making processes. Now more than ever, effective use of assessment data to plan and judge instruction is a fundamental competency for good teaching.

The purpose of this paper is to provide a framework and justification for effective ways that teachers can collect and use assessment data to make instructional decisions. This framework is provided as an indication of what effective linking of assessment data to instruction decisions ought to look like rather than a summary or survey of current practices. The framework and Innovation Configuration are primarily designed to provide a blueprint for preservice teacher preparation; however, they may also be used as an evaluation rubric or development guide for inservice professional development. Although many schools and districts may not currently have the practices discussed in this issue paper in place, these practices are strongly endorsed by both the No Child Left Behind Act (NCLB)(2002) requirements as well as the Race to the Top Fund’s (2009) competitive grants to states.

This paper begins with a discussion of why assessment and instruction should be linked. It continues with an overview of the Innovation Configuration describing essential components in preservice and inservice teacher training to better facilitate data-based decision making. Next, each of the major points within the Innovation Configuration are provided with a rationale for their importance and elaboration of some of their core characteristics. Last, recommendations are provided regarding how these components of the Innovation Configuration might be included into teacher preparation and professional development practices.
The Importance of Linking Assessment and Instruction

There are different arguments for why assessment and instruction should be closely linked or aligned, some legal, some ethical, and some practical. The legal reasons include that both federal law and state regulations have shown an increase in the requirements of collecting assessment data and use of those data for accountability purposes at the state, district, school, teacher, and student levels (Salvia, Ysseldyke, & Bolt, 2010). The No Child Left Behind Act (2002) codified that assessment is to be used to evaluate schools, districts, and states. Accountability in teacher performance or quality is also being advanced through the Higher Education Opportunity Act (HEOA; 2008). This is often at the teacher level. And at the individual level, the Individuals with Disabilities Education Improvement Act (IDEIA; 2004) mandates different types of assessment to document effectiveness for individual students as well as programs. While these are clearly important influences in the assessment practices of teachers, they are not forces that generally drive day-to-day instructional decisions; nor are many of the assessment methods required by federal or state law or regulation useful in making decisions about what to teach or how well students are learning the presented material.

The ethical basis for linking assessment and instruction is that most every professional organization includes assessment and the use of assessment data to make decisions in their guidelines for ethical and/or best practice and training. As examples, organizations for reading teachers (International Reading Association and National Council of Teachers of English, 2010), math teachers (National Council of Teachers of Mathematics, 1995), and special education teachers (Council for Exceptional Children, 2003) all provide standards for training and practice in the use of assessment. While the necessity and role of high stakes testing is addressed in each of these, the primary focus is on the use of assessment data to make decisions about teaching and learning, and this is the embodiment of the practical reason for linking assessment and instruction.

Teachers need to make instructional decisions frequently. Estimates have put the number of decisions teachers make each day at 1,300 (Jackson, 1968) or about 10 interactive decisions per hour (McKay, 1977), but empirical work has also identified that teachers make 9.6 to 13.9 instructional decisions per lesson (Morine-Dershimer & Vallance, 1975). However, Peterson and Clark (1978) reported that decisions were only made when instruction was not effective and changes were only made in half of the situations in which students were not learning sufficiently. Much of the research on the frequency of teacher decision making was conducted in the 1970s and 80s (see Shavelson & Stern, 1981 or Clark & Peterson, 1986 for reviews). Since that time the focus of research has changed.

Research on teacher decision making since the early 1980s has often focused on the outcomes of those decisions. The most common outcome is that when teachers use assessment data to make their instructional decisions, student performance increases (L. Fuchs & Fuchs, 1986; Black & William, 1998). In addition, teachers who collect systematic progress monitoring data (and use it to make decisions) make instruction changes more frequently for their students who are experiencing difficulty (L. Fuchs, Fuchs, Hamlett, & Stecker, 1991). Given the current focus on accountability and outcomes in education, training preservice and inservice teachers to more effectively and efficiently collect and use assessment data to make instructional decisions for their students and classes should be a core component of any professional development.
Innovation Configuration for Linking Assessment and Instruction

This Issue Paper presents an Innovation Configuration that can be used to evaluate general or special education preservice teacher preparation or inservice professional development in terms of their content relevant for linking assessment and instruction. This matrix is included as an Appendix.

Innovation configurations typically identify and describe the critical components of a practice that is important to training within a field. They are across two dimensions: essential components and degree of implementation (Hall & Hord, 1987; Roy & Hord, 2004). The essential components are typically listed as the row heads of the matrix within the leftmost column. Additional descriptors or subcomponents are also included for clarification and use with more specific evaluations. The degree of implementation is typically presented as column heads in the topmost row with multiple levels of implementation specified ranging from zero (no mention) through progressively higher scores to a maximum that is used to represent exemplary inclusion and implementation of the component.

Innovation configurations have been used for over 30 years as tools to develop, implement, and evaluate education innovations (Hall, Loucks, Rutherford, & Newton, 1975). The innovation configuration described in this Issue Paper is designed to provide educators who are involved with designing or evaluating preservice teacher preparation or inservice teacher professional development activities a tool to evaluate the degree to which their preparation or professional development activities incorporates current best practices for linking assessment and instruction. It is designed for use with general education teachers, instructional specialists or coaches, special education teachers, paraprofessionals, itinerant specialists or related service providers (e.g., school counselors, school psychologists, speech-language pathologists), or educational administrators. Some components may be important to elaborate upon and adapt for some specialties, but all components are important considerations for all educators.
Teacher Preparation and Professional Development to Promote Linking Assessment and Instruction

The essential components of the Linking Assessment and Instruction Innovation Configuration are as follows:

- Fundamentals of assessment
- Standards for comparison of performance
- Considerations for decision making
- Assessment procedures
- Identification of content to teach
- Identification of student response

These six components are based on the research and best practice literature detailing how assessment and instruction can be linked as well as important considerations in assessment and instruction. The following sections briefly describe the essential components of the Linking Assessment and Instruction Innovation Configuration. As stated previously, training for specific roles may warrant additional elaboration of some of the components and some details may vary by the grade level of students the educators are being trained to work with, but these six components should be addressed in any system of training for educators. Preparation in these components establishes a fundamental competency that is critical for teaching—particularly when working with students with special needs.

**Fundamentals of Assessment**

This is the fundamental information about assessment and measurement—topics such as reliability and validity, types of scores that might be produced through assessment and their interpretation, legal provisions regarding assessment, issues of cultural and linguistic diversity, statistical bias and fairness, and accommodations and modifications for use with students with disabilities and English learners. It is also an area that should cover the types of decisions that teachers and other educators routinely make. These topics are ones that are generally covered in any introductory assessment text (e.g., Miller, Linn, & Gronlund, 2008; Popham, 2010; Salvia, Ysseldyke, & Bolt, 2010).

**Standards for Comparison of Performance**

Once a student’s performance has been measured, a key component to being able to make decisions about her/his performance and plan instruction is to be able to make comparisons to a standard for performance. There are three ways of determining standards typically used in education: normative, criterion, and ipsative. *Normative* standards involve comparing a student’s performance on the assessment to that of other students in a comparable peer group. This might be other students in the same grade (e.g., 3rd grade), other students taking similar course work (e.g., high school biology), other students of the same age (e.g., 3 year olds), or other students with similar demographic characteristics (e.g., students with disabilities).

*Criterion* standards involve comparing a student’s performance to an empirically-derived level of proficiency, i.e., a cut score that is used to determine whether or not a student has sufficiently mastered
the material. For example, high stakes accountability tests have cut scores to determine whether or not a student has reached proficiency in that area. Typically these are criterion referenced. Another example would be if the core curriculum includes that students in first grade should be able to compute basic subtraction facts with 90% accuracy. This provides a criterion when giving students a sheet of basic subtraction facts and having them work the problems to determine how many they get correct.

Ipsative standards are when a student’s prior performance is the basis for comparison of their current performance. Ipsative standards are often used for goal setting and motivation. For example, if a child completed a task such as completing a sheet of independent-level work (i.e., work the child can perform accurately without support or guidance) in 20 minutes, the teacher could ask him/her to try it again, but see if he/she could do it more quickly (i.e., completing it in less than 20 minutes or with a specific goal of 18 minutes). Ipsative standards are also often considered in monitoring student progress as the student’s current performance can be compared to prior performance (yesterday or last week) as well as future performance (tomorrow or next week).

Considerations for Decision Making

The term assessment is often used to mean different things. It can be used to mean a specific task or test, the process of assigning numbers to characteristics of people or objects, or to the process of making decisions. One way to keep these multiple usages distinct is to use other terms such as instrument to refer to a specific task or test, measurement to the process of assigning numbers, and evaluation to the process of making decisions. In this framework, the term assessment can be used to mean the process of collecting information through measurement (conducting using instruments) for the purpose of evaluation (J. Hosp, 2008).

Of course, there are many different purposes of decision making. A useful framework is to consider them as inside the classroom or outside (J. Hosp, in press). Inside classroom decisions are those that are directly relevant for instructional planning or the day-to-day operations of a classroom. Examples of inside decisions are grouping of students for small group instruction, determining whether or not a student or group of students is making adequate progress, or which method to use to teach a concept or skill. Outside decisions are others that do not directly impact daily instructional planning. This should not imply that they are not important, but only that they do not have a direct or immediate impact on the teaching within a classroom. As such, they are also often ones that are not made by teachers or are made by groups of which teachers may be members. Examples of outside decisions are eligibility for specialized programs or services, Adequate Yearly Progress (AYP) of classrooms or schools, or which core program to adopt school- or district-wide.

One of the distinctions occasionally made about types of decisions is the summative/formative dichotomy. These are sometimes listed as summative and formative assessments (e.g., Black & William, 1998; Shepard, Hammerness, Darling-Hammond, Rust, Snowden, Gordon, Gutierrez, & Pacheco, 2005) and sometimes as summative and formative evaluation (Araisian & Madaus, 1972; L. Fuchs & Fuchs, 1986; Howell, J. Hosp, & Kurns, 2008). One problem with consideration as assessments is that there are purposes which do not fit the dichotomy, requiring the addition of another term—interim assessments, to bridge the gap (Perie, Marion, & Gong, 2007). Interim assessments are given less frequently than formative assessments, but with more relevance for teaching decisions than summative assessments. As
such, they might encompass periodic benchmark or screening assessments. Within this summative/formative framework, summative can be conceived of as assessment/decisions of learning, whereas formative is assessment/decisions for learning (Torgesen & Miller, 2009).

Summative decisions are made at a single point in time to summarize the learning or performance of a student or group of students. For example, high stakes tests administered at the end of a school year are for the purpose of summative outcome decisions—determination of whether or not each student met the criterion for mastery of that year’s curriculum standards and determination of AYP of the school or district. Formative decisions are those to help teachers provide the most effective instruction to their students. For example, a curriculum-based measure (CBM) of oral reading fluency (ORF) can be administered once per week to those students experiencing difficulty in order to determine the effectiveness of instruction. When the progress monitoring data indicate that a student is not learning at a sufficient rate to be proficient by the end of the school year, the educator can alter the instruction to better meet the student’s needs. Within the evaluation framework, there is no need to consider ‘interim’ decisions as these would fall under formative or summative depending on their frequency and purpose. Within the framework of assessments, interim assessments would address a little bit of both formative and summative characteristics. They are administered at periodic intervals to gain snapshots of student performance, but can also provide some feedback useful to instructional planning. For example, benchmark screening measures administered to all students in the Fall, Winter, and Spring can be used for summative decisions about student learning and the effectiveness of instruction, but also may provide feedback on which students need additional support or which areas of the content needs more instruction.

In the context of decision making, there are many different needs for teachers to make decisions. Some classroom decisions are quick and made immediately (e.g., to praise a child or not, which student to call on for response, whether or not to repeat directions). Others require being more planful in the collection of data. When a decision has high stakes associated with being wrong (i.e., making an incorrect decision), the need to have enough information to make a good decisions increase (J. Hosp, 2008). In this case, use of a structured set of procedures for collecting information and making decisions can be useful. Two structured approaches are Curriculum-Based Evaluation (CBE; Howell, J. Hosp, & Kurns, 2008) and standard treatment protocol approaches of Response to Intervention (RtI; Jimerson, Burns, & VanDerHeyden, 2007). These provide explicit guidelines and decision rules for types of information to collect, determining why to collect it, and how to make decisions—all with explicit links to providing instruction.

**Assessment Procedures**

When presented with the word assessment, many educators often equate it with testing. However, in meeting the demands of collection and use of information to make decisions about instruction, it is useful for teachers to think more broadly about what constitutes assessment. In preservice teacher preparation and inservice teacher professional development, there are a lot of variations in the specific instruments used to collect information. Different procedures are required for measuring reading at the elementary level than mathematics in high school or behavior in early childhood for example. All methods of assessment can be considered within one of four different
categories: review of information, interview, observation, and testing (which fits into the handy rubric, RIOT). **Review** of information includes collecting and systematically organizing information that has been collected previously about a student—records from her/his cumulative folder, prior test results, work samples, etc. **Interview** involves talking to others who have knowledge of the student and his/her performance. This might be other teachers, related service personnel, the student’s parents or siblings, and the student him- or herself. **Observation** is watching the student perform a task, typically in the learning environment (such as the classroom). There are methods a classroom teacher can use to collect observation data on students while teaching and there are other methods that are more appropriate for an external observer to come into the classroom to collect data (Shapiro & Kratochwill, 2002). **Testing** is the most common understanding of assessment. It includes methods ranging from informal inventories to individually administered norm-referenced tests.

**Identification of Content to Teach**

Within the confines of the general classroom and the general curriculum, there are certain externally predetermined standards for what every child is expected to learn within a grade level or at a certain age level. This is the state’s core curriculum or standards for grade-level learning. The majority of students will be held to these standards and most likely progress through the expectations at a fairly typical rate. However, for those students who are not progressing through the curriculum, it is important to identify those areas in which they are having difficulty and need extra instruction. The first step is to compare the student’s performance in each broad content area. In the elementary grades, the state or district probably has expectations within areas such as reading/language arts, mathematics, science, and social studies/history for example. The student’s performance should be compared to two different standards—how his/her performance compares to the cutoff for proficiency or mastery (criterion) and to the performance of other students in the classroom (normative). If the student’s performance is below the criterion for acceptable performance, he/she needs additional instruction in that area. If the student’s performance is similar to the peers’ performance (and below the criterion), then changes to instruction should involve the entire class and the general, or Tier 1, instruction. If his/her performance is below that of his/her peers (as well as below the criterion), the instruction should be supplementary and it is important to conduct additional assessment to determine more specifically where the breakdown in learning is occurring.

When a student’s performance is significantly below the criterion for acceptable performance and his/her peers’ performance, it is necessary to identify more specifically what difficulty the student is experiencing. This area of decision making can encompass three types of skills to examine: prerequisites, related skills, and subskills. **Prerequisites** are things the student must be able to do in order to perform the task at hand, but are not necessarily skills that would be taught previously. It includes such things as visual acuity (i.e., being able to read the materials), language proficiency, and other personological characteristics that may impact the student’s ability to access the learning materials. They are characteristics that are distal to the learning action and might need to be accommodated in order to allow the student access. For example, a student with poor vision might need to wear corrective lenses, sit closer to the board, and/or have larger print materials. These would accommodate the prerequisite of being able to see the materials.
Related skills are skills the student must be able to perform or knowledge the student must have mastered that are related to the content area of interest, but are not included within it. They are often skills that should have been taught/learned previously, but in a different content area. For example, many math instructional materials require the student to read the problems in order to derive the information for computation or application. Reading is not a component of math, per se, but is important when needing to solve story problems, geometry theorems, or other mathematical applications within sciences such as biology or physics. As such, being able to decode the text in order to comprehend the information contained therein and associate it with one’s vocabulary and prior knowledge covers a series of related skills and subskills.

Subskills are skills that are actually components of the content area of focus that must be learned before being able to master that content. They are sometimes derived through a task analysis of a skill (i.e., explicit identification of the subskills necessary to complete it) or through an explicit scope and sequence of a curriculum. For example, the student having difficulty in math, may be having specific difficulty with computation—particularly with double-digit addition with regrouping. This is a relatively specific skill within the curriculum; however, there are other subskills that are critical to being able to add two double-digit numbers with regrouping. The student must understand the concepts of regrouping, conservation of quantity, and place value. The student must know procedures for regrouping and column addition. The student must have number sense and know basic addition facts as well as understand the concept behind and procedures for adding two numbers.

When considering which procedures to use to collect information about content areas, prerequisites, related skills, or subskills, it is important to ensure that the assessment procedures used are aligned with the form of knowledge that is expected: fact, concept, or strategy (Howell & Nolet, 2000). Facts (also called rote or declarative knowledge; Marzano et al., 1988) are types of information that are discrete and stand alone. For example, knowing that the capital of the United States is Washington DC does not give any information about the capital cities of states within the U. S., capital cities of other countries, or details about Washington DC such as where it is, how to get there, or how many residents it has. Concepts are groups of objects, events, or actions that share a set of distinguishing characteristics. These characteristics are generally defined through rules for differentiating examples and nonexamples of the concept. For example, the concept of ‘squares’ would be defined by the rules: two-dimensional figure, four sides of equal length, and four right angles where the sides meet. Nonexamples would include near distracters (i.e., those that are similar in that they share one or two rule-traits, but not all—such as a rectangle) and far distracters (i.e., those that share few or no rule-traits—such as a sphere). Strategies are often defined as processes of work rather than products (Marzano et al., 1988). As such, they can be considered knowledge of how to do something, or procedures for its demonstration. They involve applying or generating other forms of knowledge (i.e., facts and concepts). For example, in math there are strategies for conducting numeric operations; in reading there are strategies for decoding a word the reader does not recognize. They are procedures for conducting an action or solving a “problem” of sorts.

To put all these together, consider the case of trying to determine the area of a circle. Concepts involved include what a circle is and that mathematical equations can be used to represent physical attributes. Facts involved would be the equation for determining the area of circle, multiplication facts, and the value of pi. Strategies involved would be to find the radius of the circle and substitute that for ‘r’
in the equation as well as the process of solving the equation (which involves application of facts such as when to multiple pi by \( r^2 \) and to square \( r \), i.e., multiple it times itself). So the smooth performance of this seemingly simple activity requires the learner to combine different forms of knowledge in rule-governed ways, but also to know when and how to apply them.

When a student does not perform a task or subskill to proficiency (i.e., above the criterion for acceptable performance), it is important to determine whether the student cannot perform the task or will not perform the task because remediation of each requires different instructional methods (Noell et al., 1998). This is identifying if the student’s difficulty is the result of a skill deficit or a performance deficit (Gresham, 1981; J. Hosp & Ardoin, 2008). A skill deficit occurs when the student is not able to perform the task at the level of proficiency required for successful performance. A performance deficit occurs when the student does not have sufficient motivation to perform the task at a proficient level or to sustain performance enough to complete a task. When exhibiting a performance deficit, the student is capable of performing the task when there is sufficient motivation, but the difficulty lies within generating the motivation. Note that although it is possible that some students actively decide to not perform a task, more often there are other reasons that negatively impact the student’s motivation.

Identification of a performance deficit should not be used to automatically indicate that a student is willfully not performing. It is also possible that a student exhibits a combined skill/performance deficit wherein the student cannot quite perform the task to proficiency but also has difficulty sustaining motivation to perform the task. The type of performance deficit can be distinguished through the use of a ‘can’t do/won’t do’ assessment (VanDerHeyden & Witt, 2008). This approach uses repetition of the task (using parallel materials) combined with implementation of reward conditions in order to determine whether or not the student cannot or will not perform the skill to proficiency.

In addition to determining whether a student cannot or will not perform the task to proficiency, the teacher or educator should consider the stage of learning at which the student can perform the task (Idol, 1989). This is also sometimes referred to as the instructional hierarchy (Haring & Eaton, 1978), and is related to the work of Benjamin Bloom (1971). There are five stages or levels of learning that students go through before mastering a task or skill. As a student begins to learn a task, he/she is in the acquisition phase. This is marked by the student becoming increasingly accurate at performing the task. Once achieving accuracy of 90-100%, the student moves into the proficiency or fluency stage, which is marked by high accuracy as well as an increasing rate of performing the task (i.e., being able to perform the task more quickly while maintaining high accuracy). Next the student enters the stage of maintenance, which is marked by retention of high rate and accuracy before moving onto the next stage, generalization. This is marked by the student beginning to transfer performance of the task to new settings or applications. Last the student enters the stage of adaptation, wherein she/he is able to capitalize on the knowledge and use that knowledge to solve problems in various settings—particularly using new or novel applications of the task. One reason this is important to consider in assessment is that if the student is in the accuracy stage of learning and can perform the task with 70% accuracy, yet the instrument being used to measure the student’s performance requires performance at rate (i.e., at the proficiency/fluency stage), the assessment results might suggest that the student cannot perform the task, when in reality the student can perform the task, but at a different level of learning. It is especially important to consider when the assessment requires a late stage of demonstration (i.e.,
generalization or adaptation) and the student is in the early stages of learning (i.e., acquisition or proficiency).

If the student has an Individual Education Plan (IEP), a section 504 plan, or any other document that explicitly determines educational goals and objectives, it is important to ensure that the methods of assessment align with the student’s goals and objectives. Preservice teachers should learn what types of these documents might exist for their students and where they should go to find them. They also should know which other specialists in their school would be primarily responsible for these plans or documents, if they are not the ones responsible. Some of the laws or regulations guiding development of these documents are state- or district-specific, so situating the preservice training (or inservice professional development, particularly for new teachers coming from out of state or district) in the laws and regulations specific to that state and district is important.

During the course of a typical school day, students generate a lot of work—some of it transitory (e.g., oral responses to questions that are not recorded or written down) and some of it permanent (e.g., written, audio, or video recorded work). Good teachers are always looking at (or listening to) student work with an evaluative focus to judge the sufficiency of the student’s performance. Much of the time this evaluation is informal, including subjective judgments of quality, inferences about the difficulty of the task for the child, and determinations of whether or not the work was completed within the allotted time. While all of these on-the-spot evaluative judgments may be incorporated into the teacher’s overall impression of the student’s performance, sometimes it is important to use more standard judgments of student work in order to include the permanent products into the student’s cumulative folder, or to share it with others who are involved in decision making about the student (e.g., parents, related service personnel, administrators).

As an example of linking assessment to identification of content to teach, consider the case of a 10th grade student (Hubert) having difficulty in an American History class. At the beginning of the year, the teacher (Ms. Washington) gave a test of content from the year’s curriculum to all students to determine their prior knowledge of the material. This would be screening decisions using both summative and formative evaluation characteristics. It gives Ms. Washington an idea of what the students in her class already know, but also a guide for what she will need to teach during the year. During the first month of school, Ms. Washington gives weekly quizzes to all students to monitor their progress and make formative decisions about the effectiveness of her instruction. She notices Hubert does not participate in class and has failed every weekly quiz. Ms. Washington decides to review Hubert’s records to evaluate his prerequisite skills that she determines are important. She finds that his vision and hearing are both excellent and that his attendance is good. She also evaluates related skills that might impact his performance. She notes that no previous teacher has documented a difficulty with attention or focus and that his reading skills (particularly comprehension) are good.

At this point, Ms. Washington decides to examine the specific subskills she has been focusing on in the American history class. There have been two main foci: facts such as names, dates, and locations of colonial America as well as concepts such as colonialism. On measures of American history facts, Hubert scores above 90%, can recall the facts at rate, and is doing so for the facts from the prior units. This suggests to Ms. Washington that Hubert’s learning of these facts is at a maintenance stage of learning and is where she expects it to be (it is similar to that of other students in the class). On measures of the conceptual information, Hubert has difficulty identifying core characteristics of the
concepts as well as providing nonexamples. This suggests to Ms. Washington that Hubert is having difficulty acquiring the conceptual knowledge that she is teaching. She next wants to determine whether this difficulty represents a skill or performance deficit, so she uses can’t do/won’t do procedures with Hubert to determine that his difficulties arise from a skill deficit—he is having trouble grasping the concepts involved. Now she understands that she needs to provide Hubert with additional instruction in acquiring the concept of colonialism.

Identification of Student Response

Assessment for identification of the content to teach is primarily about determining a student’s level of performance in different areas or on different skills. While comparing this performance to various standards, the assessment is typically collected at a single point in time to describe the student’s performance. However, it is also important for teachers to consider how the student’s performance changes over time. This is done through assessment of student progress. Progress decisions (or progress assessment) is one of the types of decisions (J. Hosp, in press) included in comprehensive frameworks for assessment and decision making in education that are typically included with the fundamentals of assessment (see Salvia, Ysseldyke, & Bolt, 2010). Monitoring student progress and making progress decisions is a core feature of response to intervention (RtI: Reschly & Wood-Garnett, 2009).

The specific choices of instruments to use to collect information on student progress will differ by content area (reading, math, etc) and by the grade or age level of the student (pre-k, elementary, secondary). However, there are many common attributes the preservice or inservice teachers should be aware of when selecting instruments (J. Hosp, in press). The first consideration is about the core characteristics about the instrument—its reliability, validity, and non-discrimination against subgroups of students (i.e., no statistical bias and general fairness)(see National Center for Response to Intervention, 2010 for a review of progress measures). The second consideration is efficiency. Instruments for progress monitoring should be quick and easy to administer and score (Deno, 2003). In general, if a progress measure takes more than 3-5 minutes per student to administer and score, it will take up too much instructional time to be useful for progress decisions. It is useful to present and consider progress measures as dynamic indicators of growth over time (Shinn, 2008) similar to how at every visit to the doctor’s office the patient’s temperature, weight, height, and blood pressure are measured. These are quick, efficient indicators of overall health rather than in-depth assessment of specific issues. Part of the efficiency is consideration for interpretation and communication of performance. Many progress measures achieve this through the use of graphs. In particular, line graphs are useful for showing change over time (see Figure 1 for an example). With inclusion of a standard for comparison (the rate of growth that is expected to meet a later goal), interpretation of how the student’s progress compares is simple.

Another consideration for selection of progress instruments is consistency of administration, scoring, and materials. This is also referred to as standardization. The use of standardized directions and scoring rules is how most instruments are able to demonstrate good reliability and validity. It is similar to how weights and measures have standard definitions—if the length of a foot were allowed to vary among rulers or tape measures, it would be nearly impossible to build things or communicate dimensions of objects. Use of consistent materials ensures that when the teacher measures growth in
student performance, it is due to learning and not to changes in the materials. This is especially important with progress assessment because the instruments must be able to be administered frequently to the same student. Achieving consistency is possible by using the exact same materials, but only if the student is not expected to learn from or remember the specific materials (otherwise his/her growth could represent a practice effect; O’Connor, White, & Swanson, 2007). In most academic areas, consistency of progress materials is achieved through the use of alternate, parallel forms—versions of the same task that include the same form of task at an equivalent difficulty, but with different specific items included. In math operations, this would include the same types of problems (e.g., multiplication facts) but with different numbers. In reading and content areas, it would include the same difficulty of the content, but a different focus (e.g., one story on the life of sea turtles, another on whether or not bears hibernate).

Another benefit of this consistency is that if all students in a class or grade level are doing the same task, under the same conditions, with the same scoring, those data can be used to make multiple decisions. The data can be used to make decisions about that individual student, but they can also be aggregated to make decisions about the progress of small groups of students (e.g., different reading groups, English learners), the classroom as a whole (e.g., is the instruction effective at increasing everyone’s performance?), or an entire grade level across the school or district (e.g., to judge the adequacy of the curriculum). While decisions at larger levels of aggregation might be beyond the control of preservice or inservice teachers, they are important considerations and ones to which the teacher can then contribute.

While we have already discussed standards for comparison, it is an important consideration for selection of progress measures. The use of norms or benchmarks (which are criterion-referenced) differs on the purpose of comparison. Many progress measures use benchmarks that have been empirically derived in order to reliably predict proficient performance on a meaningful or important outcome measure such as the state’s high stakes accountability measure. If the progress measure is being used to ensure that each student’s growth keeps them on track for proficient performance at the end of the year, benchmarks would be a good standard to use. Norms would be useful when attempting to compare a student’s performance to his/her peers. If the progress measure were being used to determine when a student receiving special education services can be reasonably reintegrated into the general classroom, or when he/she should be exited from special education services, the use of norms allows a comparison of that student’s performance to that of his/her peers. This is an important consideration for integrating or exiting a child (Powell-Smith & Ball, 2008). In standards for comparison of progress, ipsative standards can be used to compare the student’s current progress to prior progress; however, important to note is that this is only appropriate if the student’s prior progress was sufficient or if the comparison is to determine how much change (in rate of progress) has occurred as a result of an instructional change (M. Hosp, Hosp, & Howell, 2007).

As an example of how identification of student response aids in linking assessment and instruction, consider the case of a 3rd grade student (Marina) having difficulty with reading. Marina’s teacher, Mr. Jones, uses a standardized measure of reading for both screening and progress monitoring. He chose published materials created in the vein of curriculum-based measurement (CBM; Deno, 1985; 2003) because of their good reliability, validity, and ability to predict mastery on the end of the year state-mandated test. For reading, he is using a measure of oral reading fluency (ORF) because it is
efficient (taking only one minute per student per week) and consistent in that the publisher has 30 alternate forms available so that he can use a different one each week. Through the other data he has collected, Mr. Jones knows that Marina is having great difficulty with reading. He has identified the specific areas in which she needs help and planned the instruction to provide her with the skills she is missing. Mr. Jones is measuring Marina’s response to the instruction that he is providing.

Mr. Jones first identifies Marina’s current level of performance. This is indicated by the three stacked points to the left of Figure 1. He also identifies the end of year goal for Marina and marks it on the chart. He then draws a line to connect these because that line shows the average weekly rate of progress that Marina needs to demonstrate in order to meet the end of year goal. Mr. Jones begins implementing the additional instruction that he is providing to Marina and once per week he has her read aloud from one of the passages and counts the number of words she reads correctly in that minute. As the weeks go on, he can see how she is responding to his instruction. After six weeks, Mr. Jones sees that Marina’s reading is not progressing at the rate she needs to be successful by the end of the year. He draws an intervention line to indicate that he made an instructional change. Important to note is that the assessment data do not tell Mr. Jones what to change or how to change it—he needs to use his professional judgment, expertise, and other sources of information to make that decision. Once he does, he implements that instruction and continues to monitor Marina’s progress to ensure that she is on track to meet her goal.

A few other points to make about Mr. Jones and Marina. If Mr. Jones had not been monitoring her progress weekly, at the end of the year (given her rate of progress and poor response to his instruction) he would have found that Marina was even farther behind than she was at the beginning of the year. Unfortunately it would have been too late for him to do anything about it. It would be frustrating for him and demoralizing for her and her parents. Teachers who use measures that meet the standards of reliability, validity, efficiency, and consistency, have been shown to make more frequent instructional decisions (L. Fuchs & Fuchs, 1986) and effect greater student learning (Black & William, 1998) than teachers who do not use such data to make decisions. However, it is not just the act of collecting information that effects greater student learning. Teachers need to actively use the information to critically evaluate their instruction in order to determine how it could be changed to better meet the student’s needs (L. Fuchs, Fuchs, Hamlett, & Stecker, 1991).

Recommendations

This section provides a few recommendations for how to integrate the components of the Linking Assessment and Instruction Innovation Configuration into a program of study for preservice teachers or professional development for inservice teachers.

Course Structure

There are many ways to structure a series of preservice training courses or inservice activities to cover the range of topics to link assessment and instruction. Three are highlighted here: the sequential, infused, and hybrid methods. With explicit use of cognitive maps or scope and sequences of the
interrelated nature of topic (Darling-Hammond, Hammerness, Grossman, Rust, & Shulman, 2005) any of these methods could meet the “connected and coherent” criteria of effective teacher preparation programs (Zeichner & Gore, 1990). No one of these methods has been shown to be better than the others, so it is up to the program members to determine which fits best into other requirements as well as the needs of the program and students—keeping a consistent focus on the core conceptual ideas and practical skills required (Wideen, Mayer-Smith, & Moon, 1998).

The sequential method of training or development for linking assessment and instruction involves separate courses or activities for different areas. For example, a preservice program of study might involve an introductory assessment course (to cover the fundamentals), a separate course on decision making (or an advanced assessment course to cover application and implementation), and then coursework that focuses on content area instructional methods. An elementary education program may have separate methods courses for reading/language arts, math, science, social studies, etc. Secondary education programs will generally be more content specific (e.g., science, math) unless the degree is for a more general focus such as special education. One benefit of the sequential method is that the coursework can clearly build on prior courses which provides the repeated practice that effects deeper learning and development of expertise (Gick & Holyoak, 1983). A potential disadvantage is when integration of the courses becomes more difficult due to fragmented structure or inconsistent faculty message (Gore & Zeichner, 1991) or explicit application within the content methods courses is lacking (Ericsson, Krampe, & Tesch-Romer, 1993).

The infused method of training or development for linking assessment and instruction involves not having a stand-alone course for assessment or decision making, but rather infusing that information into the content methods courses. One benefit of this method is that the examples used and practice activities can be specifically aligned with that content area and practice can be used to reinforce the concepts both of assessment and of instruction in order to align them. It also aligns well with Bruner’s notion of a spiral curriculum that returns to emphasize basic ideas repeatedly and in different contexts to promote a deeper understanding of the material (1960/1977). However, a disadvantage of this approach is that the core assessment information (i.e., the fundamentals) often must be repeated across the methods courses or included with a single course from which it then diverts valuable instructional time (such that that area does not get equal coverage as the others that do not include assessment fundamentals).

The hybrid method involves aspects of both the sequential and infused methods. In this method, there is a stand-alone assessment course to cover the fundamentals of assessment. This is often used as a prerequisite for the instructional methods courses. Then students take the instructional methods courses in which the decision making and application instruction of assessment and its linking with instruction is infused. Ideally, it provides a spiral curriculum (Bruner 1960/1977) with repeated opportunities for practice (Gick & Holyoak, 1983) provided there is consistent structure (Zeichner & Gore, 1990) and a consistency of message (Wideen, 1998).

**Use a Variety of Practice Activities**
Practice is an important part of any effective training or professional development (Darling-Hammond, Hammerness, Grossman, Rust, & Shulman, 2005). When providing preservice teacher training or inservice teacher professional development on linking assessment and instruction, it is important to include a variety of practice activities that are appropriate for the skills being covered (Ball & Cohen, 1999). This is in addition to other evaluative activities to determine if the preservice or inservice teachers have learned the factual information about the fundamentals of assessment. Practice should cover at least four areas: selecting instruments, administration, scoring, reporting/interpreting results to parents or other professionals.

Practice selecting instruments is often aided by providing a checklist for the preservice/inservice teachers to use in order to ensure that they are considering the most relevant characteristics the need to make accurate decisions (see next section). This can also include activities such as locating and researching different instruments that are available and accessible to provide information aligned with making the decisions they need to make. Providing them with potential scenarios or allowing them to use scenarios that arise in their classroom, practicum, or student teaching site allows for practice that will be relevant to the decisions they need to make. Sharing among groups or individuals also allows them to build a sort of toolbox expanding on each others’ work.

Practice administering instruments is often best achieved through different levels. First it is important for preservice/inservice teachers to administer an instrument to others in the training and to receive feedback from both the instructor and the other preservice/inservice teachers. This allows them to not only get the perspective of others (the instructor and their peers), but to have the chance to watch others administer the instrument and compare their performance to the standardization rules. Use of fidelity of implementation checklists (which are often available with published instruments and can be made for instruments without implementation checklists) is an easy, structured way to make sure that everyone is looking for the same characteristics while still allowing space for personal observations such as quality of implementation and aspects that are performed particularly well. Next, practice administering the instrument to a student for whom the data are not needed in an important step. It is important to not practice administering the instrument to a student who has recently taken the measure or will in the near future as this may affect his/her results when it matters. When practicing with a student, it is important that the results are not shared with the student or his/her teachers or parents because they are for training purposes only. If the preservice/inservice teacher has a teaching certification or his/her work is being specifically observed and checked by a certified teacher, some programs and districts will allow their use. When in doubt, it is preferable to err on the side of caution and differentiate administration for practice with administration for actual data collection and decision making. Over time, supervision and scaffolding of administration and scoring support can be gradually released (Lampert, 2001).

Practice scoring instruments is often useful to start with simulated (or sample) results that the preservice/inservice teacher does not have to collect him or herself. This allows the instructor to calculate reliability among the preservice/inservice teachers (which can be a useful exercise in demonstrating the importance of standardization as well as the concept of error in measurement). Once the preservice/inservice teachers have administered the instruments to each other or students (for
practice), these results can also be scored. Having the teachers exchange the raw results to rescore each others’ work can also be used to check reliability and consistency of use of standardized scoring rules.

Practice reporting or interpreting results is the last step, but is certainly important as a component of practice. In their classrooms, the teachers will be required to share assessment results with parents and other educators. Practicing aspects of presenting the results will facilitate this. This includes describing the assessment tasks, explaining how the results are reported, explaining the standards for comparison, using graphs and charts as much as possible, and explicitly detailing how the results allow him/her to make instructional decisions about that student. It is also important to practice asking for feedback and interpretations about the results from others. These practice activities should begin with presentation to each other. Ideally, it also includes presenting to individuals without the same training or experience (e.g., parents who are not educators), but there are concerns about confidentiality that must be navigated. It is important to remember that practice activity results should not be presented to parents as they are for practice purposes and not actual decision making about the student.

Develop/Use a Checklist to Complete When Selecting or Administering an Instrument

Much as structured decision making guidelines ease professionals into complex processes, having a structured checklist to complete when selecting or administering an instrument can be a useful tool for the preservice or inservice teacher who is not yet, fully proficient at these activities. A checklist should cover the general topics and principles outlined in this document: fundamentals of assessment (e.g., is the measure sufficiently reliable, valid for this purpose, and appropriate for this population?), standards for comparison (e.g., which type of standard is appropriate and where can I find them?), considerations for decision making (e.g., what is the purpose for which I need it, does this fit into my decision making framework?), assessment procedures (e.g., are there other ways I could collect this information?), identification of the content (e.g., ensuring that the measurement tasks align with those expected to be taught, measurement of skill v. performance deficits), and identification of student progress (e.g., will this provide a level of performance only, or can it also be used to index growth over time?). Before attempting to select an instrument, it is also important for the preservice/inservice teacher to ask questions such as “why am I giving this instrument?”, “is there a more efficient way to get this information?”, and “will this lead to better instruction and outcomes for this student?” If the purpose for administering an instrument cannot be explicitly and emphatically stated before selecting it, there are other questions that need to be answered rather than whether or not it is reliable.

It is also useful to provide preservice/inservice teachers with a checklist for preparing to administer or score an instrument. Sometimes these are similar to the implementation checklists for specific measures (see Good & Kaminski, 2002) and sometimes there are other resources that are specific to an instrument (see M. Hosp, Hosp, & Howell, 2007). When there is not, other general checklists for setting up and preparing to administer an instrument with a student are also available (see M. Hosp & Hosp, 2000).
Conclusion

Assessment and instruction are two key components to effective teaching and therefore necessary components to preservice teacher training and inservice teacher professional development. They are also two components that should be intricately linked. Although there is great variation in the details of how information is collected, what it is used for, and the effect it has, research has consistently shown that teachers who base their instructional decisions on assessment data effect greater student learning (L. Fuchs & Fuchs, 1986; Black & William, 1998). Not all components of this document or the Linking Assessment and Instruction Innovation Configuration will be equally important for all training activities, but they are important concepts and skills for all teachers and educators to have. As the field of education moves increasingly to evidence-based practice, the role of teachers as scientist-practitioners will also increase. Through a detailed understanding and applied use of linking assessment and instruction, teachers will be well-situated for this role.
References


Race to the Top Fund, 74 Fed. Reg 59,836 (Nov. 18, 2009)


Appendix A

**Linking Assessment and Instruction Innovation Configuration: Essential Components for Teacher Competency**

<table>
<thead>
<tr>
<th>Essential Components</th>
<th>Degree of Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Instructions:</strong> Place an X under the appropriate variation implementation score for each course syllabus that meets each criterion specified from 0 to 4. Score and rate each item separately.</td>
<td>Code = 0</td>
</tr>
<tr>
<td>No evidence that the concept is included in the class syllabus</td>
<td>Syllabus mentions content related to the concept</td>
</tr>
<tr>
<td><strong>Descriptors and/or examples are bulleted below each of the components.</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Fundamentals of Assessment</strong></td>
<td></td>
</tr>
<tr>
<td>• Reliability—definition and types commonly used to judge educational assessments (e.g., test-retest, interrater)</td>
<td></td>
</tr>
<tr>
<td>• Validity—definition and types commonly used to judge educational assessments (e.g., criterion-related, content)</td>
<td></td>
</tr>
<tr>
<td>• The use and interpretation of score scales (e.g., percentiles, standard scores, systematic observation metrics)</td>
<td></td>
</tr>
<tr>
<td>• Legal provisions of assessment (e.g., NCLB, IDEA)</td>
<td></td>
</tr>
<tr>
<td>• Issues of cultural and linguistic bias and fairness</td>
<td></td>
</tr>
<tr>
<td>• Accommodations and modifications for students with disabilities or English learners</td>
<td></td>
</tr>
<tr>
<td>• The educational decisions that assessment data can be collected to help in making (i.e., screening, progress, broad and targeted diagnostic, and outcome)</td>
<td></td>
</tr>
<tr>
<td><strong>Standards for Comparison of Performance</strong></td>
<td></td>
</tr>
<tr>
<td>• Norm-referenced (i.e., comparison to age or grade similar peers)</td>
<td></td>
</tr>
<tr>
<td>• Criterion-referenced (i.e., comparison to empirically-derived level of proficiency)</td>
<td></td>
</tr>
<tr>
<td>• Ispasive standards (i.e., comparison to prior</td>
<td></td>
</tr>
</tbody>
</table>
### Considerations for Decision Making
- Usage of and differentiation between the terms assessment and evaluation
- Comparisons of inside and outside purposes of decision-making
- Definitions and comparisons between summative and formative evaluation (and possibly interim assessments)
- Use of structured decision-making frameworks (e.g., curriculum-based evaluation)

### Assessment Procedures
- Review of prior records
- Interview relevant individuals
- Observe performance in appropriate settings
- Administration and interpretation of test results
- Selection of assessment procedures that provide the information needed to make instructional decisions

### Identification of Content to Teach
- Consideration of and focus on broad areas (e.g., reading, math)
- Consideration and focus on specific subskills, important prerequisites, or related skills (e.g., phonological segmenting, understanding of place value)
- Addressing alignment of assessment/instruction on different forms of knowledge (i.e., facts, concepts, strategies)
- Working within the instructional hierarchy (accuracy, fluency, generalization, adaptation; Haring & Eaton, 1978)
- Consideration of difficulties arising from skill or performance deficits
- Alignment with or writing of goals and objectives
- Analysis of student work
### Identification of Student Response

- Characteristics of good formative measures
  - Ease of administration and scoring
  - Representation of performance with graphs
  - Aggregation of data to make individual/small group, classwide, and school/district-wide decisions

- Standards for comparison of performance
  - Norms
  - Benchmarks
  - Prior progress

<table>
<thead>
<tr>
<th>Column Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Add Figure 8.3 from ABCs of CBM.

*Figure 1.* A sample line graph showing student results on a curriculum-based progress measure.

![Sample CBM graph with each part labeled.](image)