

# FOCUS ON EXCEPTIONAL CHILDREN

## **Formative Evaluation: Developing a Framework for Using Assessment Data to Plan Instruction**

*John L. Hosp*

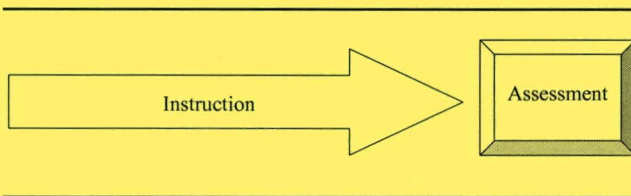
Assessment, a systematic process of collecting information, is a core activity in education, and its importance has grown over the past decade (Shepard, 2008). Advances in the development of standards for comparison, purposes and uses of assessment data, and methods and tools to collect data have come fast and furious as developers strive to capitalize on new technologies and refine old ones (Salvia, Ysseldyke, & Bolt, 2012). Although the core purposes of assessment have changed little over the past 40 years for educators working with students with disabilities, the number of methods and tools has expanded dramatically. This makes staying current on advances in assessment an increasingly daunting task.

The Council for Exceptional Children (CEC), the National Association of School Psychologists (NASP), the American Speech and Hearing Association (ASHA), and many other organizations have standards for training and certification that include assessment as a core component. Both the National Council on Teacher Quality (NCTQ, 2012) and the National Comprehensive Center for Teacher Quality (NCCTQ; Hosp, 2010) have recently focused on the need for assessment in preservice training programs for all educators. Yet training in assessment for inservice teachers has not nearly been such a focus.

Part of the reason for a lesser focus on assessment for inservice teachers may be the plethora of demands on their time. The core activity of all teachers is rightly ... teaching. Traditionally this has meant the act of instruction—presenting knowledge or skills or guiding activities designed to impart knowledge or skills (Darling-Hammond & Bransford, 2006). In this context, assessment was considered something that was done once the teacher was finished with instruction, something done “to determine how much students have learned, with little or no emphasis on using results to improve learning” (McMillan, 2007, p. 7). Figure 1 illustrates this unidirectional relation between instruction and assessment. Although aligning assessment and instruction in order to use assessment data to inform instruction has always been considered to some degree, it has received a stronger emphasis for the past 15 years in the form of formative assessment.

---

Dr. Hosp is an associate professor in the Department of Teaching and Learning at the University of Iowa.



**FIGURE 1.**

**Old Model of the Relation between Assessment and Instruction wherein Assessment Occurs once Instruction Is Completed**

The term “formative” is sometimes credited as being first applied to education by Scriven (1967) in his descriptions of educational evaluations that were summative or formative in nature. He used it to describe the process of collecting data while an educational program was still “malleable” and therefore subject to adjustment to alter the outcomes. In 1969, Bloom made the change to formative assessment, and

it is this term that seems to have gained traction (although there are distinctions between the two, which will be discussed later). Formative assessment has also been referred to as “assessment for learning” rather than assessment of learning (Torgesen & Miller, 2009) or “prospective” rather than retrospective (Wiliam, 2007). Figure 2 illustrates this recursive relation between instruction and assessment.

Although it may be “cast in the role of life preserver ... as a strategy to fend off pervasive accountability pressure” (Popham, 2008, p. 2), the renewed emphasis on using assessment data to inform instruction in a closely aligned and interconnected fashion is a good one, given the evidence that it can improve student performance (Black & Wiliam, 1998). However, it still leaves us with the question of why we assess.

**WHY DO WE ASSESS?**

We assess for many reasons. There are often numerous legal and ethical bases for the assessment we conduct (Hosp, 2010). Legally, every educator is working under various federal and state laws that dictate certain actions of assessment. The accountability provisions of the Elementary and Secondary Education Act (ESEA; even before in its incarnation as No Child Left Behind) dictated students in grades 3–8 taking a high-stakes test in reading and math. Race to the Top (RTT), in addition to containing assessment provisions, included a subcompetition, the RTT Assessment Program, for two consortia to develop comprehensive assessment systems for accountability. Individual states have also passed laws requiring various end of year (EOY), end of course (EOC), or exit exams for various services or content courses. Ethically, professional standards govern various educational disciplines. Particularly when these organizations are responsible for certifying or recertifying educators, it is important to adhere to their standards. However, for educators the most important reason for assessing is generally practical: it is the need to gather information in order to guide our instructional practices.

Every act of assessment should begin with a question— not “why am I doing this?” or “is this really necessary?” but a substantive question that is directly relevant to a decision

# FOCUS ON EXCEPTIONAL children

ISSN 0015-511X

FOCUS ON EXCEPTIONAL CHILDREN (USPS 203-360) is published monthly except June, July, and August as a service to teachers, special educators, curriculum specialists, administrators, and those concerned with the special education of exceptional children. This publication is annotated and indexed by the ERIC Clearinghouse on Handicapped and Gifted Children for publication in the monthly *Current Index to Journals in Education* (CIJE) and the quarterly index, *Exceptional Children Education Resources* (ECER). The full text of *Focus on Exceptional Children* is also available in the electronic versions of the *Education Index*. It is also available in microfilm from Serials Acquisitions, National Archive Publishing Company, P.O. Box 998, Ann Arbor, MI 48106-0998. Subscription rates: individual, \$50 per year; institutions, \$68 per year. Copyright © 2012, Love Publishing Company. All rights reserved. Reproduction in whole or part without written permission is prohibited. Printed in the United States of America. Periodical postage is paid at Denver, Colorado. **POSTMASTER:** Send address changes to:

Love Publishing Company  
Executive and Editorial Office  
P.O. Box 22353  
Denver, Colorado 80222  
Telephone (303) 221-7333

**CONSULTING EDITORS**

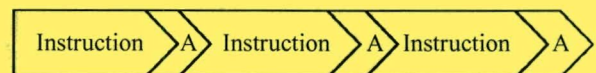
Steve Graham  
Vanderbilt University

Ron Nelson  
University of Nebraska–Lincoln

Eva Horn  
University of Kansas

Carrie E. Watterson  
Senior Editor

Stanley F. Love  
Publisher



**FIGURE 2.**

**Current Model of Interactive Assessment (A) and Instruction**

you need to make about a student or group of students. Questions should involve what to teach, how to teach, or whom to teach. The level of specificity of the question can certainly vary as well as the content and focus, but if you cannot readily answer the what, how, and whom questions for a process of assessment, then it might be necessary to revisit why you are doing it. Assessment should always be included in a process of evaluation.

The answer to the question of why we assess is “to evaluate” because that is the heart of linking assessment and instruction. Whereas assessment is often defined and used to describe a systematic process of collecting information, evaluation is the extension of that process whereby we use the information to make important educational decisions (Hosp, 2011). Research on decision making shows a few interesting trends. First, inferences are generally more accurate when made at a broader or more general level than when made about specifics (Kruglanski, 1990). Second, teacher inferences about what a student is having difficulty with (i.e., the gap between expected performance and actual performance) are more accurate than their decisions about *what to do with the information* (i.e., instructional planning to address the achievement gap; Heritage, Kim, Vendlinski, & Herman, 2009). This suggests that educators are generally proficient in assessment but have difficulty in putting their assessment data to good use to make educational decisions—often referred to as evaluation. Given that educators make approximately 9–14 instructional decisions per lesson (Morine-Dershimer & Vallance, 1975), make decisions more frequently when collecting systematic assessment data (Fuchs, Fuchs, Hamlett, & Stecker, 1991), and effect greater student achievement when making decisions based on data (Stecker & Fuchs, 2000; Wesson, 1991), it is crucial to increase educators’ ability to make these decisions.

## MAKING DECISIONS

The number and variety of decisions educators routinely make are staggering, but even when focusing solely on decisions that are significant and directly relevant to instruction, they may average 10 per hour (McKay, 1977). It is no wonder then, that, given the amount of data available, to make this many decisions can be, at times, overwhelming and lead to what is referred to as “decision fatigue” (Vohs et al., 2008). Essentially the processes used to make choices, ranging from seemingly simple (“Should I have a second cup of coffee?”) to more complex (“How will I fit all these tasks into a busy schedule for the day?”), are similar to using a muscle. Practice making decisions can be useful by making one more adept at it, but the more decisions we make, the more those processes get fatigued and the more difficulty we have making decisions.

Making decisions can be broken down into three components: weighing the options, selecting one, and implementing it (Vohs & Baumeister, 2004). Clearly these must be implemented in order. By the time we get to the point of implementation, we have already expended a good deal of effort—more effort needing to be expended when weighing many options or having to consider many factors when selecting one. In fact, having too many options or components to the decisions to be made can even lead to “decision paralysis” wherein the individual makes no choice at all because of the overwhelming nature of the task of choosing (Iyengar & Lepper, 2000). This might explain the finding that the process of using assessment data to guide instruction more often breaks down at the implementation phase (Heritage et al., 2009). However, we can take steps to decrease the effect of decision fatigue.

## STRATEGIES FOR AVOIDING DECISION FATIGUE

Although research into strategies to reduce decision fatigue specifically is still in its infancy, there are some promising steps one can take. Nothing can be done to completely avoid decision fatigue, but it can be lessened. Like any intervention, they will be differentially effective for different people, at different times, and for different purposes. As with instruction, it is important to have multiple strategies available to you (see Figure 3).

**Make important or complex decisions early in the day.** The effort of decision making has a cumulative effect on an individual, and, therefore, we are more likely to be “fresh” in the morning. One hypothesis for this early freshness is that, like any exertion, decision making uses up a lot of glucose, and the body has better reserves in the morning (Gaillot et al., 2007). Adding healthy snacks that increase one’s glucose at strategic times during the day also has been found to help provide energy for decision making (Gaillot et al., 2007).

**Predetermine standards for comparison.** Rather than taking a student’s raw performance on a task to be analyzed

- 
1. Make important or complex decisions earlier in the day
  2. Predetermine standards for comparison
  3. Impose a structure to complex decision-making tasks
  4. Routinize low-stakes or frequent decisions
  5. Get input or feedback
  6. Do not wait for perfection
- 

**FIGURE 3. Strategies for Managing or Avoiding Decision Fatigue**

in its entirety (i.e., considering many aspects of performance simultaneously), make sure to have pre-established standards to which to compare the student's performance. Ideally these have been empirically validated to relate to or predict a meaningful outcome (Joint Committee on Standards in Educational and Psychological Testing, 1999). Using a standard for comparison reduces the informational input to a dichotomous (e.g., proficient/nonproficient) or ordinal (e.g., high/medium/low) scale from the full body of raw information. It is a form of summarizing or aggregating data to make the most of it. Using this approach also requires knowing beforehand what characteristic or characteristics of the performance are most important to compare and summarize.

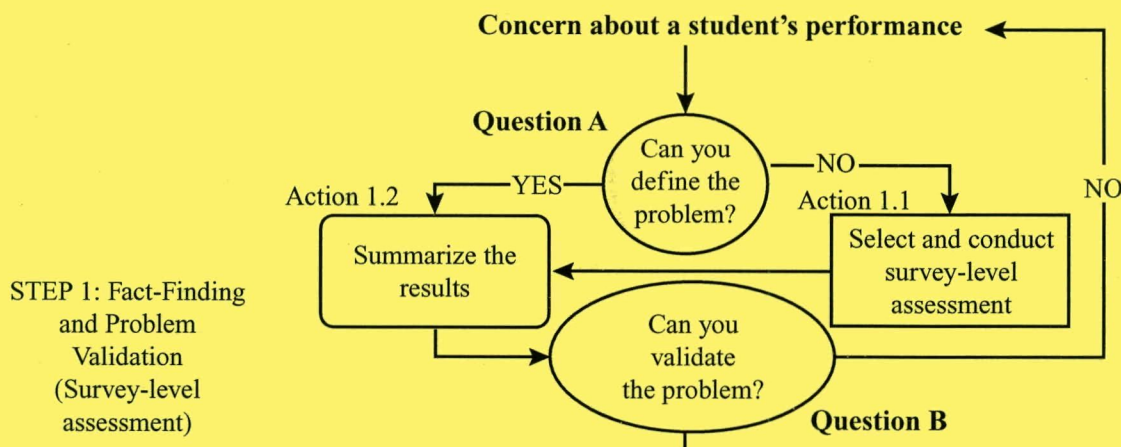
#### Impose a structure to complex decision-making tasks.

When working with large amounts of information or attempting to make complex decisions (i.e., ones that have many facets to consider), the process of decision making can be especially effortful. Having a predetermined structure can remove some of the process decisions that would otherwise need to be made. The type of decision that needs to be made determines the structural alternatives. The most common of these in education are flowcharts, multiple gating procedures, and innovation configurations. Flowcharts offer a structure to the decision-making process by ordering the content to focus on or the decision to be made in a sequential fashion (Howell, Hosp, & Kurns, 2008). As is shown in Figure 4, the process of decision making moves from one area to the next, and it is only once a decision has been made about one step that the process moves on. In this manner,

eventually all the decisions within the complex decision making process can be addressed.

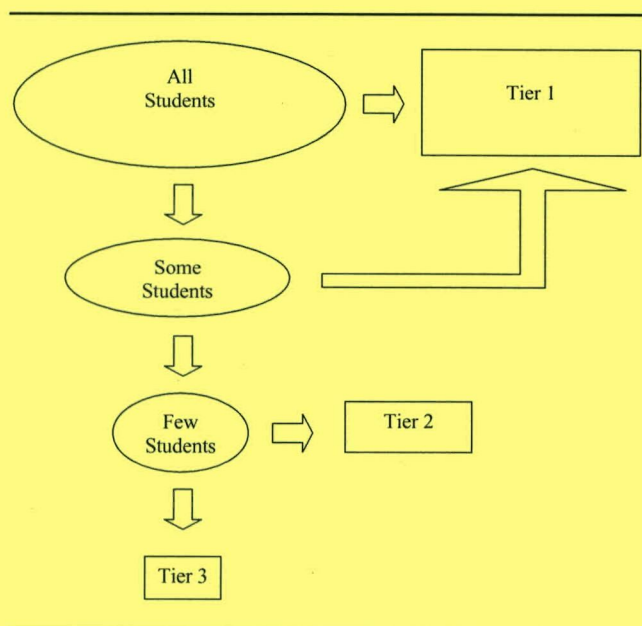
Figure 5 shows a multiple gating procedure wherein a complex decision is broken into a series of intermediate steps. These are most often used to pare alternatives down to the most relevant, if not the best, choice. With decision making, potential options must be classified into groups, and the groups that most clearly do not meet the criteria for selection are then ruled out. This makes the process simpler than having to eliminate each option individually. Multiple gating procedures are often used in screening decisions where educators need to decide which students in a school or grade level are in need of which interventions or levels of service (e.g., Walker et al., 2009; Fuchs et al., 2011). The first gate might be good for identifying those with the greatest or least needs; from there, subsequent gates are designed to collect additional information only about those students for whom a decision cannot yet be made until decisions can be made about all students.

Figure 6 shows an innovation configuration, a matrix that is designed to describe and be used to evaluate components of a specific practice within a professional field (Hosp, 2010). Most innovation configurations are designed to evaluate training within a field but can also be used to evaluate implementation of an instructional plan by focusing on essential components and degree of implementation (Roy & Hord, 2004). One promising aspect of their use is the current emphasis on fidelity of implementation of instruction and intervention in education (Sanetti, Dobey, & Gritter, 2012).



Source: Excerpted from "Best Practices in Curriculum-Based Evaluation." by K. Howell, J. Hosp, & S. Kurns, 2008. Copyright 2008 by National Association of School Psychologists. Excerpted with permission.

**FIGURE 4.**  
Example of a Flowchart to Systematize Decision Making in Education



**FIGURE 5.**  
**Example of a Multiple Gating Process to**  
**Systematize Educational Decision Making**

**Routinize low-stakes or frequent decisions.** As the above examples provide a structure for complex decisions in order to reduce the decision-making load, similar approaches can be used with simple, low-stakes, or frequent decisions. These include checklists and division of responsibilities. Checklists have enjoyed a burgeoning renaissance in recent years (Gawande, 2009). They can provide a quick and effective way of preventing simple errors, particularly for routine tasks that are often performed automatically or when one needs to divert attention to perform more complex decisions or tasks (Pronovost et al., 2006). However, checklists are not ideal for complex decisions because they tend to become a “fill-in-the-blanks” approach that de-emphasizes consideration of disconfirming data (Groopman, 2007). It is important to align the intervention to systematize decision making with the level of complexity of the decisions just as it is important to align instruction to student needs.

Another way to routinize low-stakes or frequent decisions is to have one person responsible for certain decisions or components of decisions. If a paraprofessional is present in a classroom or grade level, this could be an ideal job (Stephens & Woodbury, 2011). If not, other options might be a parent volunteer, or, in higher grades, students assigned to take responsibility for certain “jobs” within the classroom.

**Get input or feedback.** Input from others can be one way to move past decision paralysis by providing a fresh

perspective on the situation. Input from someone who has not been involved in the deliberations can also be assistive by weeding out extraneous information that creeps into decision-making processes, as tangential information may become more prominent as the ability to prioritize information becomes more difficult.

**Do not wait for perfection.** Especially when devoting a great deal of time to a complex decision or when the stakes are high, a tendency can be to attempt to make the perfect decision that we are absolutely certain is correct. In addition to devoting time and energy to an unreasonable standard, the final choice can become a victim of “choice-supportive bias” in that if we have invested that much into it and assume it is the perfect choice, evidence showing that it is not perfect (which is bound to arise) is more likely to be rationalized away or ignored (Mather, Shafir, & Johnson, 2003). One way to combat the aim for perfection is to be attentive to saturation—the theoretical point when there is enough information to make a good decision and gathering more information will not increase our odds of making a better decision (Hosp, 2008).

#### THE SUMMATIVE/FORMATIVE CONTINUUM

Another strategy for reducing decision fatigue is having a clear structure and framework for the decisions that need to be made—a clearer identification of the roles and relations between assessment and evaluation for different purposes. The term formative, as mentioned above, is often used as if it is one half of a dichotomy, diametric to the term “summative.” In this framework formative assessment is assessment *for* learning, whereas summative assessment is assessment *of* learning (Torgesen & Miller, 2008); formative is prospective, whereas summative is retrospective (William, 2007). This is really a somewhat artificial dichotomy, the best evidence for which is the attention being given to a third grouping, “interim,” which is intended to bridge the divide between the two (Perie, Marion, & Gong, 2007). Figure 7 shows how this trichotomy is conceptualized using a school year as a timeframe. Formative assessments are administered frequently throughout the year to inform instruction, interim assessments are administered periodically, and summative assessments are administered at the end of the year. A reason this dichotomy/trichotomy exists may be the treatment of summative and formative as *assessment* (i.e., instruments used to collect information) rather than *evaluation* (i.e., a process of making educational decisions). It provides a framework for test developers to classify and market their products, but it oversimplifies the needs of educators when making instructional decisions.

Instruments are the tools or procedures we use to collect information that can include review procedures, interview or

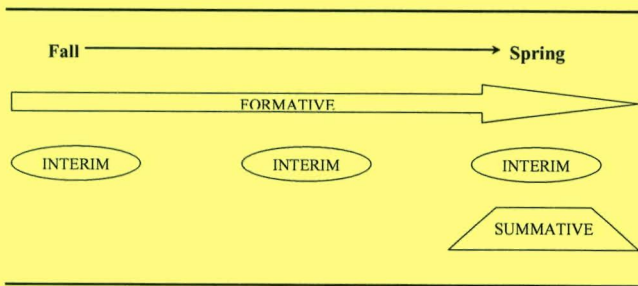
Essential Components	Degree of Implementation				
	Code = 0	Code = 1	Code = 2	Code = 3	Code = 4
<p><i>Instructions: Place an X under the appropriate variation implementation score for each course syllabus that meets each criterion specified from 0 to 4. Score and rate each item separately.</i></p> <p><i>Descriptors and/or examples are bulleted below each of the components.</i></p> <p><b>Fundamentals of Assessment</b></p> <ul style="list-style-type: none"> <li>• Reliability—definition and types commonly used to judge educational assessments (e.g., test–retest, interrater)</li> <li>• Validity—definition and types commonly used to judge educational assessments (e.g., criterion-related, content)</li> <li>• The use and interpretation of score scales (e.g., percentiles, standard scores, systematic observation metrics)</li> <li>• Legal provisions of assessment (e.g., NCLB, IDEA)</li> <li>• Issues of cultural and linguistic bias and fairness</li> <li>• Accommodations and modifications for students with disabilities or English learners</li> <li>• The educational decisions that assessment data can be collected to help in making (i.e., screening, progress, broad and targeted diagnostic, and outcome)</li> </ul>	<p><b>No evidence</b> that the concept is included in the class syllabus</p>	<p>Syllabus <b>mentions</b> content related to the concept</p>	<p>Syllabus mentions the concept <b>with required readings</b> on the topic</p>	<p>Syllabus mentions the concept, requires readings, and has <b>an assignment, project, or test</b> on the topic</p>	<p>Syllabus mentions the concept; requires readings; has an assignment, project, or test; <b>and has supervised practice</b> related to the topic through fieldwork</p>

Source: From "Linking Assessment and Instruction: Teacher Preparation and Professional Development," by J. Hosp, December 2010, TQ Connection Issue Paper. Copyright 2010 by the National Comprehensive Center for Teacher Quality. Reprinted with permission.

### FIGURE 6. Example of an Innovation Configuration to Systematize Educational Decision Making

observation protocols, tests, rating scales, and many other forms of assessment (see Hosp, 2011, for a more thorough discussion). A key point in this sort of definition is that instruments provide information. That information can be used in a variety of ways. Some instruments are designed to provide information that is more useful (i.e., more valid) for making certain decisions than others, and several other

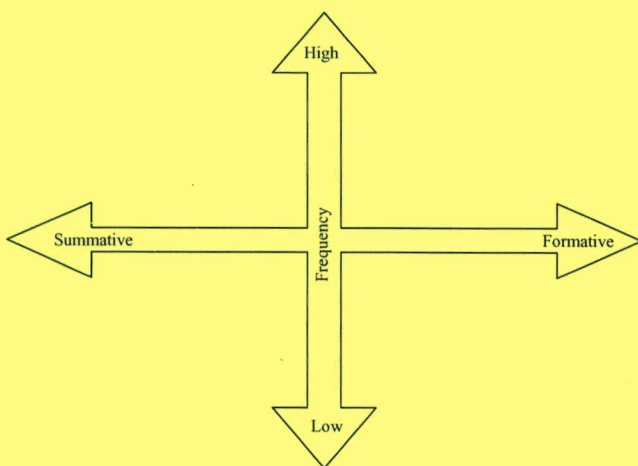
aspects enter into consideration: how often an instrument can be administered, how long it takes to administer the instrument, the stakes of the decision, what type of information is needed (e.g., relative standing, prescriptive information about specific skills), how the information will be aggregated or reported, and many more. Instruments are the tools to get at the really important things—decisions.



**FIGURE 7.**  
**Conceptualization of Formative (F), Interim (I),  
 and Summative (S) Assessments as  
 Discrete Instruments or Activities**

Rather than conceptualizing summative and formative evaluation as discrete types of assessment (with interim added to the mix), it can be helpful to conceptualize them as ends of a continuum with instruments providing information that can be used both summatively and formatively (see Figure 8). In this framework, information is collected with varying frequency (the vertical axis) and results from different types of assessment procedures (review, interview, observation, test) contain a mixture of information that can be used for summative or formative purposes. These purposes should be aligned with the decisions educators need to make.

## EDUCATIONAL DECISIONS



**FIGURE 8.**  
**The Continuum of Summative and  
 Formative Purposes**

Every educational decision has two dimensions: domain (or level of aggregation) and focus (or type of decision). Consideration of the first dimension as evaluation domains puts educational decisions in a framework of how proximal the decision is to the learning event (Hosp, 2011). Decisions about characteristics of the learner (i.e., a specific student) are the most proximal, as it is the student's learning that is the most direct and relevant outcome of instruction. The next, slightly less proximal, domain is the instruction provided to foster that learning. This is the "how to" teach. Third is the curriculum, or the "what to" teach. Last, setting characteristics can be important facilitators or inhibitors of teaching and learning. When a student is not learning sufficiently, the breakdown might occur in any one or a combination of domains. Decisions need to be made about the relevant characteristics within each domain, as well as interactions across domains, to ensure that barriers are minimized. For example, even though the instruction might be considered generally effective, it is important to consider that this instruction is both effective for the curriculum content being delivered and well aligned with learning and setting characteristics.

Another way of thinking about the domains is to consider them as levels of aggregation: the learner is the individual student level; instruction is aggregated at the classroom; curriculum is aggregated at the grade level (either within a school or district); and setting would include classroom, school, and district levels. In this way, the interactions between and among domains can be subsumed into levels of aggregation. For example, considering aggregation of information at the classroom level would include interactions among the setting, curriculum (which may be determined at the district or state level), instruction, and multiple learners.

The second dimension is the focus, or type, of decision. A common framework in current use is that of screening, progress, diagnostic, outcome (SPDO). These decisions are in relation to answering specific types of questions such as "Which students?" (screening), "How much progress?" (progress), "Where is the problem?" (diagnostic), and "How well?" (outcome; Hosp, 2011).

Screening often occurs with medium to high frequency. Along the summative end of the continuum, screening decisions would be less frequent (one to three times a year), as the purpose is to select and classify students into instructional groups or tiers of service delivery (i.e., what is often described as interim or benchmark assessments). This can be accomplished via direct route screening or multiple gating procedures. Direct route screening requires instruments with strong technical adequacy (i.e., reliability and validity) that are administered to all students (i.e., they are universal). The intent is to make the screening decisions once and place students into groups or tiers based on the information. Although the grouping can be flexible and may be revisited,

the decision occurs as a single step. Multiple gating screening (as described above) generally starts with a universal step that includes a decision of which students are clearly proficient or not in need of additional intervention. Additional information is collected about the remaining students, and a decision is made to sort them into higher and lower performing groups or tiers of differentially intensive instruction. This continues until all students are reliably classified or sorted into the predetermined number of groups or tiers (see Figure 5). Along the formative end of the continuum, screening would include pretests or previews of lesson or unit content to determine what the students in a class or group already know. Conducting this type of screening allows the teacher to adjust instruction so that known information is reviewed and reinforced, while the unknown information is taught more explicitly and comprehensively—requiring a different emphasis and different instructional methods (Archer & Hughes, 2010).

Progress decisions generally occur with higher frequency but may be included within screening, diagnostic, or outcome decisions for more summative purposes. Along the summative end of the continuum, progress decisions are included within a dual discrepancy framework for response to intervention (RtI; Fuchs & Fuchs, 1998) or growth models based on large-scale assessment (LSA; Elliott, Kurz, & Neergaard, 2011). Dual discrepancy decisions compare both a student's level of performance and rate of progress to standards for proficient performance, whereas LSA growth models use a student's year-to-year performance to make normative comparisons over time. Along the formative end of the continuum, ongoing progress monitoring toward year-end goals or mastery of a topic or unit, such as curriculum-based measurement (CBM) approaches (Hosp, Hosp, & Howell, 2007) would be included. Informal measures used repeatedly within a lesson would also serve a formative progress function.

Diagnostic decisions tend to occur with low frequency for summative purposes but high frequency for formative purposes. Along the summative end of the continuum, diagnostic decisions are similar to medical decisions, often referred to as a trait profile differences approach (Nitko, 1989). In it, the focus is almost entirely on learner characteristics with the purpose of finding those characteristics that represent a discrepancy from typical or a standard or those that match a documented pattern that has been judged to represent a deficiency. Along the formative end of the continuum, diagnostic decisions are about collecting information that can be used to plan, deliver, and evaluate teaching. Formative diagnostic decisions often use a combination of formal and informal measures across a variety of assessment procedures (i.e., review, interview, observation, test) and are conducted with varying frequency. The research into teacher decision

making (cf. Jackson, 1968; Morine-Dersheimer & Vallance, 1975) often focused on instructional decisions or instructional planning of this type. Today, the focus of the formative assessment movement is on making instructionally relevant decisions of this type (cf. Heritage, 2010; Popham, 2008; McManus, 2008).

Outcome decisions are generally low frequency—occurring once in the appropriate timeframe. Because of their evaluative nature, serving to compare performance to a standard in order to make determinations about effectiveness or proficiency, outcome decisions have most often been equated with summative assessment—assessing to “summarize” performance. This represents their function along the summative end of the continuum, whereby they serve as a final evaluation. This would include high school exit exams, end of course exams, end of year exams, and other high-stakes purposes—those that are most often associated with accountability (Elliott et al., 2010). Along the formative end of the continuum, outcome decisions can occur at the end of a smaller portion of time, often called “granularity” or “grain size” (Heritage, 2010). Grain sizes for outcome decisions with formative components might include a lesson, a unit, a week, a month, or a quarter, reporting period, or semester. In this sense, information about student performance can be collected at the end of a lesson to decide to what degree the student mastered the material (a summative purpose), yet the same material can be useful in planning the next lesson within the scope and sequence of the unit or in relation to the curricular standards for that grade (a formative purpose).

## CONCLUSION

Although educational trends come and go, and using assessment data to make decisions is no exception, the current age of accountability has been fairly consistent and even increasing in its focus on data-based decision making. With the advances in technology that make the collection, organization, and use of data ever easier, it seems highly unlikely that the current emphasis on using assessment data to plan, deliver, and evaluate instruction will ever go away. What is likely to continue is that educators will gain access to increasing amounts of information and be expected to make even more decisions than they currently make. While this is good from a student achievement standpoint because it is likely to lead to improved student outcomes, it does increase the potential for decision fatigue, which might be counterproductive to the need to use data to make decisions.

Educators need to be aware of the potential for decision fatigue and take active steps to prevent or lessen it. Having a clear set of goals and questions to be answered is one simple way to keep the decision-making process from becoming



too cumbersome. Particularly when used within a framework for decisions or a structure such as a multiple-gating process or flowchart, an external scaffold is in place to guide one's thinking and help prevent common errors of omission or commission. Making decisions early in the day, getting input or feedback from colleagues, and not waiting to make that perfect decision can all help avoid decision fatigue and make the process of formative evaluation efficient and effective.

## REFERENCES

- Archer, A., & Hughes, C. (2010). *Explicit instruction: Effective and efficient teaching*. New York: Guilford Press.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Educational Assessment: Principles, Policy, and Practice*, 5, 7–74.
- Council for Exceptional Children (CEC). (2003). *What every special educator must know: Ethics, standards, and guidelines for special education* (5th ed.). Arlington, VA: author.
- Darling-Hammond, L., & Bransford, J. (Eds.) (2006). *Preparing teachers for a changing world: What teachers should learn and be able to do*. San Francisco: Jossey-Boss.
- Elliott, S., Kurz, A., & Neergaard, L. (2011). Large-scale assessment for educational accountability. In K. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook*. (Vol. 3, pp. 111–138). Washington, DC: American Psychological Association.
- Fuchs, L., Compton, D., Fuchs, D., Hollenbeck, K., Hamlett, C., & Seethaler, P. (2011). Two-stage screening for math problem-solving difficulty using dynamic assessment of algebraic learning. *Journal of Learning Disabilities*, 45, 372–380.
- Fuchs, L. S., & Fuchs, D. (1998). Treatment validity: A unifying concept for reconceptualizing the identification of learning disabilities. *Learning Disabilities Research and Practice*, 13, 204–219.
- Fuchs, L. S., Fuchs, D., Hamlett, C. L., & Stecker, P. M. (1991). Effects of curriculum-based measurement and consultation on teacher planning and student achievement in mathematics operations. *American Educational Research Journal*, 28, 617–641.
- Gaillot, M., Baumeister, R., DeWall, C., Maner, J., Plant, A., Tice, D., ... Schmeichel, B. (2007). Self-control relies on glucose as a limited energy source: Willpower is more than a metaphor. *Journal of Personality and Social Psychology*, 92, 325–336.
- Gawande, A. (2009). *The checklist manifesto: How to get things right*. New York: Metropolitan Books.
- Groopman, J. (2007). *How doctors think*. Boston: Houghton Mifflin.
- Heritage, M. (2010). *Formative assessment: Making it happen in the classroom*. Thousand Oaks, CA: Corwin.
- Heritage, M., Kim, J., Vendlinski, T., & Herman, J. (2009). From evidence to action: A seamless process in formative assessment? *Educational Measurement: Issues and Practice*, 28(3), 24–31.
- Hosp, J. (2008). Best practices in aligning academic assessment with instruction. In A. Thomas & J. Grimes (eds.) *Best practices in school psychology V* (pp. 363–376). Bethesda, MD: National Association of School Psychologists.
- Hosp, J. (2010, December). *Linking assessment and instruction: Teacher preparation and professional development* (Issue paper). Washington, DC: National Comprehensive Center on Teacher Quality.
- Hosp, J. (2011). Using assessment data to make decisions about teaching and learning. In K. Harris, S. Graham, & T. Urdan (Eds.), *APA Educational psychology handbook*. (Vol. 3, pp. 87–110). Washington, DC: American Psychological Association.
- Hosp, M., Hosp, J., & Howell, K. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York: Guilford Press.
- Howell, K. W., Hosp, J. L., & Kurns, S. (2008). Best practices in curriculum-based evaluation. In A. Thomas & J. Grimes (eds.) *Best practices in school psychology V* (pp. 349–362). Bethesda, MD: National Association of School Psychologists.
- Individuals with Disabilities Education Improvement Act, Pub. L. No. 108-446, § 1400, 118 Stat. 2647 (2004).
- Iyengar, S., & Lepper, M. (2000). When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79, 995–1006.
- Jackson, P. (1968). *Life in classrooms*. New York: Holt, Rinehart, & Winston.
- Joint Committee on Standards in Educational and Psychological Testing. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Kruglanski, A. (1990). Conditions for accuracy: General or specific? *Advances in Psychology*, 68, 15–34.
- Mather, M., Shafir, E., & Johnson, M. (2003). Remembering chosen and assigned options. *Memory & Cognition*, 31, 422–433.
- McKay, A. (1977). The Alberta studies of teaching: A quinquennial search of some sailors. *CSSE News*, 3, 14–17.
- McManus, S. (Ed.) (2008). *Attributes of effective formative assessment*. Washington, DC: Council of Chief State School Officers.
- McMillan, J. (2007). *Assessment essentials for standards-based education* (2nd ed.). Thousand Oaks, CA: Corwin.
- Morine-Dershimer, G., & Vallance, E. (1975, November). *A study of teaching and pupil perceptions of classroom interaction* (Technical report 75-11-6). San Francisco: Beginning Teacher Evaluation Study, Far West Laboratory.
- National Council on Teacher Quality. (NCTQ). (2012, March). *What teacher preparation programs teach about assessment and assessment data* [draft report]. Washington, DC: Author.
- Nitko, A. (1989). Designing tests that are integrated with instruction. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 447–474). New York: Macmillan.
- No Child Left Behind Act, Pub. L. No. 107-110, § 101, 115 Stat. 1439 (2002).
- Perie, M., Marion, S., & Gong, B. (2007). *Moving towards a comprehensive assessment system: a framework for considering interim assessments*. Dover, NH: The National Center for the Improvement of Educational Assessment, Inc. Retrieved from www.ncea.org
- Popham, W. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Pronovost, P., Needham, D., Berenholtz, S., Sinopoli, D, Chu, H., Cosgrove, S., ... Goeschel, C. (2006). An intervention to decrease catheter-related bloodstream infections in the ICU. *New England Journal of Medicine*, 355, 2725–2732.
- Race to the Top Fund, 74 Fed. Reg 59,836 (Nov. 18, 2009).
- Roy, P., & Hord, S. M. (2004). Innovation configurations chart a measured course toward change. *Journal of Staff Development*, 25, 54–58.
- Salvia, J., Ysseldyke, J., & Bolt, S. (2012). *Assessment in special and inclusive education* (12th ed.). Florence, KY: Cengage Learning.
- Sanetti, L., Dobey, L., & Gritter, K. (2012). Treatment integrity of interventions with children in the *Journal of Positive Behavior Interventions* from 1999 to 2009. *Journal of Positive Behavior Interventions*, 14, 29–46.

- Scriven, M. (1967). The methodology of evaluation. In R. W. Tyler, R. M. Gagne, & M. Scriven (Eds.), *Perspectives of curriculum evaluation, Volume I* (pp. 39–83). Chicago: Rand McNally.
- Shepard, L. (2008). A brief history of accountability testing 1965–2007. In K. E. Ryan & L. A. Shepard (Eds.), *The future of test-based educational accountability* (pp. 25–46). New York: Routledge.
- Stecker, P. M., & Fuchs, L. S. (2000). Effecting superior achievement using curriculum-based measurement: The importance of individual progress monitoring. *Learning Disability Research and Practice, 15*, 128–134.
- Stephens, T., & Woodbury, C. (2011). Paraeducators. In C. G. Simpson & J. P. Bakken (Eds.), *Collaboration: A multidisciplinary approach to educating students with disabilities* (pp. 71–85). Waco, TX: Prufrock Press.
- Torgesen, J. K., & Miller, D. H. (2009). *Assessments to guide adolescent literacy instruction*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.
- Vohs, K., & Baumeister, R. (2004). Understanding self-regulation: An introduction. In R. Baumeister & K. Vohs (Eds.), *Handbook of self-regulation: Research, theory, & applications* (pp. 1–12). New York: Guilford Press.
- Vohs, K., Baumeister, R., Schmeichel, B., Twenge, J., Nelson, N., & Tice, D. (2008). Making choices impairs subsequent self-control: A limited-resource account of decision making, self-regulation, and active initiative. *Journal of Personality and Social Psychology, 94*, 883–898.
- Walker, H., Seeley, J., Small, J., Severson, H., Graham, B., Feil, E., ... Forness, S. (2009). A randomized control trial of the first step to success early intervention: Demonstration of program efficacy outcomes in a diverse, urban school district. *Journal of Emotional and Behavioral Disorders, 17*, 197–212.
- Wesson, C. L. (1991). Curriculum-based measurement and two models of follow-up consultation. *Exceptional Children, 57*, 246–257.
- Wiliam, D. (2007). Keeping learning on track: Classroom assessment and the regulation of learning. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: A project of the National Council of Teachers of Mathematics* (pp. 1053–1098). Greenwich, CT: Information Age.