Instructionally Valid Assessment Within Response to Intervention

Jim Ysseldyke | **Matthew K. Burns**

Sarah E. Scholin | **David C. Parker**

Federal legislation permitting the use of student response to intervention (RTI) has increased interest in data-driven instructional decision making and renewed the debate about what constitutes instructionally relevant data. What are the purposes and goals for assessment within RTI? What makes data instructionally relevant? Research on the nature of learning, learning disabilities, and assessment suggests that the three are dynamic and strongly influenced by the student's individual ecology—so, data within RTI should accurately reflect the student ecologically and include rates of growth. Instructionally relevant assessments must be precise, frequent, and sensitive to change. There are several different assessment tools that meet these requirements and can inform the instructional and diagnostic decisions made within an RTI framework.

Special educators work daily to enhance the competence of individual students by taking them from where they are instructionally and moving them toward desired educational goals or outcomes. In so doing, these educators typically make decisions about *what* to teach (instructional level or content) and *how* to teach (the particular instructional approach, content, instructional strategies, or tactics that work best). Tests are helpful tools in making decisions about what to teach, but they do not help us decide how to teach (Salvia, Ysseldyke, & Bolt, 2010). The best way to make decisions about how to teach is to teach, and gather data on the relative effectiveness of alternative instructional approaches or interventions. Essentially, the task is to monitor progress and use data to make instructional modifications; such is the thinking that underlies precision teaching (Lindsley, 1964), data-based program modification (Deno & Mirkin, 1977), and RTI (Burns, Deno, & Jimerson, 2007).

Professionals traditionally have demonstrated instructional validity by attempting to show that performance on ability or achievement measures interacts with specific treatments or instructional interventions to produce differential outcomes. These interactions are variously referred to as aptitude-by-treatment interactions (ATIs) or trait-by-treatment interactions (Cronbach & Snow, 1977). Put simply, the assumption is that test performance predicts the success of differing interventions, and that educators' knowledge of how students perform on tests helps them differentiate instruction. However, decades of research have yet to identify ATIs (Kavale & Forness, 2000) and the passage of the Education of All Handicapped Children Act (EHA) in 1975 made prediction a moot issue. We entered a zero-demission era in which our focus shifted from making predictions about students' lives to making a difference in their lives (Reschly & Ysseldyke, 2002).

EHA's successor, the Individuals With Disabilities Education Act (IDEA, 2008), permits schools to use student response to validated interventions (i.e., RTI) to identify learning disabilities. This has resulted in intensified interest in gathering assessment information that suggests valid approaches to instruction, and in heightened attention to data-driven instructional decision making. At the same time, it's spurred renewed debate about what constitutes instructionally relevant data (Batsche, Kavale, & Kovaleski, 2006). It is important to clarify both the definition and use of instructionally relevant data within RTI. As the unique and complex situations educators encounter

TEACHING Exceptional Children, Vol. 42, No. 4, pp. 54-61.

in practice can often lead to confusion about valid assessment procedures, it's also important to clarify the purposes and goals for assessment within successfully implemented RTI. We suggest that the primary purpose of assessment within RTI models is to facilitate decision making that leads to instructional practices that help children with unique needs learn.

Purpose of Assessment

Assessment data within an RTI model are used to identify the need for an intervention and which intervention is necessary (Witt, Daly, & Noell, 2000), whether or not an intervention resulted in adequate response (Gresham, 2002), and whether students are eligible for special education (Burns, Jacob, & Wagner, 2008). Assessment procedures commonly used in schools today are often inconsistent with the purposes of RTI and the characteristics of valid diagnostic paradigms (Cromwell, Blashfield, & Strauss, 1975). Published norm-referenced tests provide scores that reflect students' knowledge relative to other students (i.e., indices of relative standing), but provide limited information about how well a student is responding to instruction—and even this is dependent on the extent of overlap between the curriculum and test content (Shapiro, 2004). Historically, assessment of students with unique learning needs almost always has been focused on making entitlement (eligibility) decisions, and the orientation was toward making predictions, comparisons, and placements. Tests were used to ascertain the extent to which students gave evidence of the characteristics of federally named (or legally named) categories and conditions, and only secondarily in efforts to decide what and who to teach.

Substantial efforts have gone into deciding specifically who is experiencing academic learning difficulties, who is "learning disabled," and who is entitled to special education services. Numerous assessment strategies and tools have been developed to answer these questions, and do so with considerable success (see box, "Internet Resources for Information About Assessment Within RTI"). However, the process underlying RTI is more concerned with questions about how to instruct those for whom previously attempted strategies are ineffective. Related to this goal is the fact that understanding learning, learning disabilities, and the assessment thereof is actually an interactive, dynamic process; one that is strongly influenced by the student's individual *ecology* (Dean, Burns, Grialou, & Varro, 2006; Simos et al., 2002; Swanson & Howard, 2005). Thus, the first requirement of an

Internet Resources for Information About Assessment Within RTI

Florida Assessment for Instruction in Reading (FAIR)
http://www.fcrr.org/FAIR_Search_Tool/FAIR_Search_Tool.aspx

Intervention Central
http://www.interventioncentral.com

National Association of State Directors of Special Education
http://www.nasdse.org/Projects/ResponsetoInterventionRtIProject/tabid/411/Default.aspx

National Research Center on Learning Disabilities
http://www.nrcld.org/

Renaissance Learning-Advanced Technology for Data-Driven Schools
http://www.renlearn.com/

Research Institute on Progress Monitoring
http://www.progressmonitoring.net/

RTI Classification Tool and Resource locator
http://www.rtictrl.org/resources/

RTI Action Network
http://www.rtinetwork.org/

System to Enhance Educational Performance (STEEP) and RTI
http://www.joewitt.org/steep.html

effective RTI assessment system is that data accurately reflect ecological information, including what was actually taught. Contextual information (e.g., direct observations) leads to more valid diagnostic decisions (Steinhausen & Reitzle, 1996) and is more sensitive to behavioral change (Someki & Burns, 2009). To illustrate, decisions made within an RTI framework should reflect rate of growth. Approaches that use single assessment administrations are likely to yield different outcomes than decisions based on frequent assessments that reflect growth (Gresham, 2002).

Questions about who is experiencing learning disabilities and who is learning disabled have too often led school personnel to spend too much time collecting historical data on alleged internal and unalterable student characteristics and not enough time on assessment for student learning (Stiggins, 2005). For decades researchers have failed to identify the clinical utility of various classification systems (Kavale & Forness, 1999). Conversely, recent evidence suggests that diagnostic decisions can and should be based on student response to evidence-based interventions within the RTI model (Burns, Jacob et al., 2008; Fuchs & Fuchs, 1998), which is consistent with the Cromwell et al. (1975) classic criteria requiring assessment paradigms that collect data leading to predictable outcomes.

The first requirement of an effective RTI assessment system is that data accurately reflect ecological information, including what was actually taught.

Generally, the most meaningful and effective data are likely direct and frequent samples of the behavior in question before, during, and after implementing interventions. Such data provide information about how the student is responding to specific interventions and about the extent to which instructional changes are necessary. Data obtained through systematic direct observations of behavior are most sensitive to behavioral interventions because they are designed to identify changes following an intervention; when interventions are measured with teacher rating scales, only small to moderate effects are observed (Stage & Quiroz, 1997). This difference is likely due to the fact that commercial rating scales are generally constructed to measure stable child characteristics with less emphasis on environmental factors. Similarly, special education services determined using single-session, high-stakes summative assessment procedures to identify learning disabilities led to a mean effect size of .20 for educational outcomes (Kavale & Forness, 1999), whereas instruction determined by frequent, formative evaluation led to a moderate to large mean effect size of .71 (Fuchs & Fuchs, 1986).

Instructionally Relevant Assessment

Labels tend to carry real meaning, and assessment that is instructionally relevant results in more functional labels than does classification based on disability terms. For example, commonly used terms within RTI such as *targeted need*, *intensive need*, *student needing alternative programming*, and *low fluency* may overtake terms such as *learning disabled*, *emotionally and behaviorally disordered*, and *mentally retarded*. To construct the meaning behind these new labels, RTI needs to lead to positive learning outcomes. To achieve positive learning outcomes, instructionally relevant assessment needs to be precise, frequent, and sensitive to change.

Precise

Considerable precision is needed in measurement of student progress toward instructional goals or outcomes. For example, in reading it is becoming relatively common to monitor progress with general outcome measures (e.g., oral reading fluency), which may provide as much data as are needed for most students. However, more precise measures (mastery of specific objective or skills) may be needed for students experiencing reading difficulties. Specifically, fluency data can provide information about whether or not a child is struggling with reading fluency, but do not inform us of specific skill deficits that are contributing to poor reading fluency or how to guide intervention. After determining that a child's fluency within a particular subject area is below grade-level expectations, educators should conduct measures of specific subskills. If a child is struggling with reading, assessments of phonemic awareness and phonetic skills can help determine the appropriate intervention (Burns, VanDerHeyden, & Boice, 2008). Specific measures needed for mathematics skills could include knowledge of numeracy, basic fact fluency, and sign-to-operation correspondence (Fuchs et al., 2003). Finally, written expression includes correct formation of letters, production (amount of words written), spelling and rules, vocabulary, and organization (Bradley-Johnson & Lesiak, 1989), all of which can be individually assessed.

It may be beneficial to examine student attainment of learning objectives that are specific to the learning unit. Researchers have found that modifying instruction with assessment data from specific objectives and subskills increased mastery of those skills (Burns, 2007), and also increased performance on outcome measures and the rate with which students obtained passing scores on state accountability tests (VanDerHeyden & Burns, 2005). Each of these examples illustrates the utility of calibrating the precision of assessment within RTI to the unique needs of the learner.

Frequent

Researchers recurrently have discussed frequency of assessment for reliable decision making. Christ (2006) found that when using data to estimate rates of student growth, at least 8 to 10 weeks of twice-weekly assessments were needed to obtain data that were sufficiently reliable for instructional decisions. Moreover, there is a reasonable concern that annual assessments alone do not provide instructionally relevant information (Shepard, 2000). Specifically, criterion-referenced annual assessments fail to account for environmental influence on students' scores and assess a limited range of subskills, resulting in reduced sensitivity to growth (Shapiro, 2004). Thus, an assessment model within RTI should contain periodic and continuous assessments in addition to annual assessments. Periodic assessments are often referred to as *benchmark assessments* or *interim assessments* in which general outcome data are collected for every child three to five times each year.

Other forms of data are collected more frequently (e.g., moment-by-moment, daily, once or twice each week) and are used to assess specific skills rather than general outcomes, a process typically referred to as *progress monitoring*. Although most or all children require annual and periodic data, continuous data are usually only collected for students whose needs warrant them. However, preliminary research found that data derived from continuous monitoring of students' daily performance was highly predictive of their performance on annual high-stakes state tests (Ysseldyke, 2006), suggesting that continuous assessment of all students can be useful. Classroomwide continuous assessment can be achieved by newly developed technology-enhanced assessment systems (e.g., Accelerated Math™; Renaissance Learning, 1999). Continuous monitoring systems facilitate progress monitoring by providing daily feedback on student performance and allowing for easy data collection and analysis (Ysseldyke & McLeod, 2007). And, the addition of continuous progress monitoring with feedback to standard instruction significantly enhances student mastery of subject matter content (Bolt, Ysseldyke, & Patterson, in press; Ysseldyke & Bolt, 2007).

Sensitive to Change

The basis of much of the criticism of traditional assessment practices centers on a lack of instructional utility because of two factors: (a) a lack of overlap between assessment and curriculum, and (b) insensitivity to changes in behavior. Only 29% to 71% of the skills included in common curricula appeared on standardized norm-referenced measures (Martens, Steele, Massie, & Diskin, 1995). Moreover, student performance on norm-referenced tests is interpreted in comparison to a norm group, not used for obtaining changes in scores between test administrations.

The two aforementioned criticisms can be best addressed by deriving

Table 1. Assessment for Instructional Relevance Within an RTI Framework

RTI Tier	Level of Precision	Frequency	Sensitivity to Change	Examples
Tier 1	GOM	3–5 times per year	Low	CBM and/or group achievement test (e.g., Measures of Academic Progress, Northwest Evaluation Association, 2003)
Tier 2	GOM and categorical SMM	At least once every other week	Medium	CBM and CBA for instructional design or a phonic screener
Tier 3	GOM and very precise SMM	At least once a week	High	CBM and nonsense-word fluency

Note. RTI = response to intervention; GOM = general outcome measures; SMM = subskill mastery measures; CBM = curriculum-based measurement; CBA = curriculum-based assessment.

assessments from the curriculum and using fluency-based raw-score measures that allow monitoring of student progress with a constant metric throughout the school year (Elliott & Fuchs, 1997). Curriculum-based assessments (CBA) and curriculum-based measurements (CBM) provide this level of sensitivity while also supplying an almost unlimited supply of alternate forms with which to frequently test student skills (Burns, MacQuarrie, & Campbell, 1999). CBAs, for example, are taken from the school's curriculum and are used to assess the obtainment of short-term instructional objectives. CBMs can be applied as part of the aforementioned progress monitoring process, and their sensitivity to change is enhanced through repeated administration of equivalently difficult probes (or, in math, probes that represent similar tasks). Educators can be confident that improved performance on the CBM is indicative of skill acquisition.

Assessment Within RTI

Instructional Decisions

As a student's unique needs increase, the levels of precision, frequency, and sensitivity of assessment correspondingly increase. The RTI model can be conceptualized as a resource allocation model (Burns & VanDerHeyden, 2006; Tilly, 2008), and can thus be seen as a framework for understanding how the different characteristics of assessment increase with students' needs. As a student moves up from Tier 1, assessment resources should be allocated accordingly. Table 1 shows the characteristics of instructionally relevant assessment for reading at each tier of intervention, using the criteria of precision, frequency, and sensitivity.

At Tier 1, students are assessed for benchmark reading scores three times each year in order to provide data for two purposes. First, quality core instruction is ensured by evaluating classwide reading scores, suggesting systemic interventions if necessary. Second, individual students may be identified as needing Tier 2 supports. Local norms can assist with the second purpose because grade-level criteria can be used to identify students who need additional support (e.g., students who score at or below the 25th percentile on grade-level local norms).

General outcome measures are collected at Tier 2 as well, but subskill mastery measures are used in addition to identify broad targets for intervention. For example, if the student needs a fluency-based intervention, CBA can be used to ensure that the student is instructed within his or her instructional level, for which research has shown positive effects (Burns, 2007). Alternatively, a phonics measure such as nonsense-word fluency might demonstrate that the student requires a phonics intervention to be successful in the general curriculum. Thus, in the latter example, progress might be monitored with both oral reading (general outcome measure) and nonsense word (subskill mastery measure) fluencies.

Students receive a Tier 3 intervention when the interventions attempted at Tier 2 are not successful. Assessment data used within Tier 3 should inform an in-depth problem analysis procedure such as the review, interview, observe; and test/instruction, curriculum, environment, and learner matrix (Heartland Area Education Agency, 2002) in which multiple measures are used to assess multiple sources of a child's learning environment (Howell, Hosp, & Kurns, 2008). In summary, assessment data must be highly precise so as to test related hypotheses about student difficulties, but must also be collected frequently and address variables that can be modified through intervention (Hosp, 2008).

Progress monitoring is a critical aspect of any RTI model. Student progress is monitored at Tier 1 with the three benchmark assessments, but weekly to twice-weekly measures are used for Tiers 2 and 3. Students receiving a Tier 2 intervention should probably be assessed weekly, because the more frequently data are collected the faster instructional decisions can be made. The progress of students receiving a Tier 3 intervention should be monitored at least once each week. Data collected to monitor student progress are used to determine if the current intervention is being effective and should be based on general outcome measures such as CBM. Other types of data could be instructionally useful, but resource allocation decisions should be made with grade-level measures obtained from psychometrically sound procedures. Students who score below a criterion for skill level (e.g., at or below the 25th percentile on a local norm) and who are not making sufficient growth (e.g., below one standard deviation of the grade-level mean rate of growth) require additional support than what they currently receive (Fuchs, 2003). For example, a student receiving a Tier 2 intervention for sufficient time and with acceptable implementation integrity whose postintervention score remains below criterion and whose rate of growth deviated from a typical rate of growth would receive a Tier 3 intervention.

Diagnostic Decisions

The IDEA regulations require "a full and individual initial evaluation" (34 CFR § 300.301) prior to providing special education services—this could include health, vision, hearing, social and emotional status, general intelligence, academic performance, communicative status, and motor abilities if appropriate. Some have suggested that a comprehensive evaluation should include measures of cognitive processing (Hale, Naglieri, Kaufman, & Kavale, 2004). The lack of instructional utility of standardized norm-referenced cognitive measures is well documented (Charlesworth, Fleege, & Weitman, 1994; Reschly, 1996; Salvia et al., 2010), and there is a long history of research finding mostly small effects for interventions derived from cognitive processing data (Kavale & Forness, 2000). Moreover, basing specific learning disability (SLD) identification decisions on cognitive processing data is a practice that is susceptible to many of the same criticisms as using a discrepancy model and research has not supported the idea (Dean & Burns, 2002). Nor has research supported efforts to base instructional interventions on profile analyses or diagnostic identification of process or aptitude deficits (Arter & Jenkins, 1979; Ysseldyke, 1973; Ysseldyke & Salvia, 1974).

The collection of RTI data is not in and of itself a comprehensive evaluation, but additional data are only collected when appropriate. Additional data could include measures of instructional environments (including classroom, home, and home-school partnership aspects of the instructional environment; Ysseldyke & Christenson, 2002), current academic skills (using broad band or diagnostic achievement tests), or functional analyses of behavior. Discussion about IDEA included in the 2006 *Federal Register* stated that,

> The Department does not believe that an assessment of psychological or cognitive processing should be required in determining whether a child has an SLD. There is no current evidence that such assessments are necessary or sufficient for identifying SLD. Further, in many cases, these assessments have not been used to make appropriate intervention decisions. (p. 46651)

Thus, comprehensive evaluations for SLD eligibility within RTI will have more to do with precise, frequent, sensitive assessments of instructionally relevant variables and less to do with cognitive processes.

Final Thoughts

RTI can be conceptualized as the use of assessment data to systematically and efficiently allocate resources for the purpose of improving learning for all students (Burns & VanDerHeyden, 2006). Thus, those of us who are passionate about assessment see this as an opportunity to finally use school-based assessment data to their fullest potential. This lofty goal can be best accomplished with an assessment model that examines the context of learning, directly samples the skill and curriculum, uses periodic and continuous assessments with tools that are sensitive to change, and provides data that lead to interventions. Even to us this seems like a great deal of work. Marzano (2003) points out that even when schools do use instructionally sensitive assessments, they often lack a system for interpreting the data. Therefore, systems with which to collect, analyze,

and use data are necessary in order to allow classroom teachers to easily use data to enhance instruction. Fortunately, many technology enhanced assessment systems already exist, such as STAR Early Literacy (Renaissance Learning, 2004), Measures of Academic Progress (Northwest Evaluation Association, 2003), Accelerated Math (Renaissance Learning, 2004), and Yearly Progress Pro (McGraw-Hill Digital Learning, 2004). To thoroughly discuss the different technology enhanced assessment systems would exceed the scope of this article; readers are referred to Ysseldyke and McLeod (2007) who provided a comprehensive discussion of these various tools. These technological advancements are indicative of the ongoing work to lessen the difficulty behind gathering and interpreting the data necessary for making instructionally relevant decisions in RTI frameworks.

Although previous research in assessment and learning has substantially informed practice, more work is clearly needed. Most educators know what data are needed in order to

Even when schools do use instructionally sensitive assessments, they often lack a system for interpreting the data.

conduct assessment *for* learning (Stiggins, 2005), and how those assessments should be conducted and the data used. However, despite technological advances there is less certainty as to how to get schools to implement with integrity a contemporary assessment model. More important, once implemented, what is the best way to sustain implementation? Dramatic reforms in assessment practices within special education have occurred over the past 30 years, but the practice has not yet caught up to research. Perhaps the growing trend of implementing RTI within the schools will lead to increased use of instructionally relevant assessment data. Regardless, using assessment to guide instructional decision making is the only means to assure individualized instruction to meet the unique learning needs of a child with, or without, a disability.

References

Accelerated Math [Computer software]. (1999). Wisconsin Rapids, WI: Renaissance Learning.

Arter, J., & Jenkins, J. (1979). Differential diagnosis—Prescriptive teaching: A critical appraisal. *Review of Educational Research, 49,* 517–556.

Batsche, G., Kavale, K. A., & Kovaleski, J. F. (2006). Competing views: A dialogue on response to intervention. *Assessment for Effective Intervention, 32,* 6–19.

Bolt, D. M., Ysseldyke, J. E., & Patterson, M. J. (in press). Students, teachers, and schools as sources of variability, integrity and sustainability in implementing progress monitoring. *School Psychology Review*.

Bradley-Johnson, S., & Lesiak, J. L. (1989). *Problems in written expression: Assessment and remediation.* New York, NY: Guilford.

Burns, M. K. (2007). Reading at the instructional level with children identified as learning disabled: Potential implications for response-to-intervention. *School Psychology Quarterly*, 22, 297–313.

Burns, M. K., Deno, S. L., & Jimerson, S. R. (2007). Toward a unified response-to-intervention model. In S. R. Jimerson, M. K. Burns, & A. VanDerHeyden (Eds.), *Handbook of response to intervention* (pp. 428–440). New York, NY: Springer.

Burns, M. K., Jacob, S., & Wagner, A. (2008). Ethical and legal issues associated with using response-to-intervention to assess learning disabilities. *Journal of School Psychology, 46*, 263–279.

Burns, M. K., MacQuarrie, L. L., & Campbell, D. T. (1999). The difference between curriculum-based assessment and curriculum-based measurement: A focus on purpose and result. *Communiqué, 27,* 18–19.

Burns, M. K., & VanDerHeyden, A. M. (2006). Using response to intervention to assess learning disabilities: Introduction to the special series. *Assessment for Effective Intervention, 32,* 3–5.

Burns, M. K., VanDerHeyden, A. M., & Boice, C. H. (2008). Best practices in delivering intensive academic interventions for individual students. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed, pp. 1151–1162.). Bethesda, MD: National Association of School Psychologists.

Charlesworth, R., Fleege, P. O., & Weitman, C. J. (1994). Research on the effects of group standardization testing on instruction, pupils, and teachers: New direction for policy. *Early Education and Development, 5,* 195–212.

Christ, T. J. (2006). Short-term estimates of growth using curriculum-based measurement of oral reading fluency: Estimating standard error of the slope to construct confidence intervals. *School Psychology Review, 35,* 128–133.

Cromwell, R., Blashfield, R., & Strauss, J. (1975). Criteria for classification systems. In N. Hobbs (Ed.), *Issues in the classification of children* (pp. 4–25). San Francisco, CA: John Wiley & Sons.

Cronbach, L., & Snow, R. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York, NY: Irvington.

Dean, V. J., & Burns, M. K. (2002). Inclusion of intrinsic processing difficulties in LD diagnostic models: A critical review. *Learning Disability Quarterly, 25,* 170–176.

Dean, V. J., Burns, M. K., Grialou, T., & Varro, P. J. (2006). Comparison of ecological validity of learning disabilities diagnostic models. *Psychology in the Schools, 43,* 157–168.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual.* Reston, VA: Council for Exceptional Children.

Elliott, S. N., & Fuchs, L. S. (1997). The utility of curriculum-based measurement and performance assessment as alternatives to traditional intelligence and achievement tests. *School Psychology Review, 26,* 224–233.

Fuchs, L. S. (2003). Assessing intervention responsiveness: Conceptual and technical issues. *Learning Disabilities: Research & Practice, 18,* 172–186.

Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children, 53,* 199–208.

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., Prentice, K., Burch, M., Hamlett, C. L., Owen, . . . Jancek, D. (2003). Explicitly teaching for transfer: Effects on third-grade students' mathematical problem solving. *Journal of Educational Psychology, 95*, 293–305.

Gresham, F. M. (2002). Responsiveness to intervention: An alternative approach to the identification of learning disabilities. In R. Bradley & L. Danielson (Eds.), *Identification of learning disabilities: Research to practice. The LEA series on special education and disability* (pp. 467–519). Mahwah, NJ: Lawrence Erlbaum.

Hale, J., Naglieri, J. A., Kaufman, A. S., & Kavale, K.A. (2004). Specific learning disability classification in the new Individuals With Disabilities Education Act: The danger of good ideas. *The School Psychologist, 58*, 6–13.

Heartland Area Education Agency. (2002). *Improving children's educational results through data-based decision making*. Johnston, IA: Author.

Hosp, J. L. (2008). Best practices in aligning academic assessment with instruction. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 363–376). Bethesda, MD: National Association of School Psychologists.

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* (5th ed., pp. 349–362). Bethesda, MD: National Association of School Psychologists.

IDEA Regulations, 34 C.F.R. § 300 (2006).

Individuals With Disabilities Education Act, 20 U.S.C, §§ 1400 *et seq*. (2008).

Kavale, K. A., & Forness, S. R. (1999). Effectiveness of special education. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 984–1024). New York, NY: John Wiley & Sons.

Kavale, K. A., & Forness, S. R. (2000). Policy decisions in special education: The role of meta-analysis. In R. Gersten, E. P. Schiller, & S. Vaughn (Eds.), *Contemporary special education research: Synthesis of the knowledge base on critical instructional issues,* (pp. 281–326). Mahway, NJ: Lawrence Erlbaum.

Lindsley, O. R. (1964). Direct measurement and prosthesis of retarded behavior. *Journal of Education, 147*, 62–81.

Martens, B. K., Steele, E. S., Massie, D. R., & Diskin, M. T. (1995). Curriculum-bias in standardized tests of reading decoding. *Journal of School Psychology, 33*, 287–296.

Marzano, R. (2003). Two wrongs and a right. *Educational Leadership, 60*(5), 56–60.

Northwest Evaluation Association. (2003). *Technical manual for the NWEA measures of academic progress and achievement level tests.* Lake Oswego, OR: Author.

Reschly, D. J. (1996). Functional assessments and special education decision making. In W. Stainback & S. Stainback (Eds.), *Controversial issues confronting special education: Divergent perspectives* (2nd ed., pp. 115–128). Boston, MA: Allyn & Bacon.

Reschly, D. J., & Ysseldyke, J. E. (2002). Paradigm shift: The past is not the future. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (4th ed., pp. 3–21). Bethesda, MD: National Association of School Psychologists.

Salvia, J., Ysseldyke, J., & Bolt, S. (2010). *Assessment: In special and inclusive education* (11th ed.). Boston, MA: Houghton-Mifflin.

Shapiro, E. S. (2004). *Academic skills problems: Direct assessment and intervention* (3rd ed.). New York, NY: Guilford Press.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher, 29*(7), 4–14.

Simos, P. G., Fletcher, J. M., Bergman, E., Breier, J. I., Foorman, B. R., Castillo, E. M., . . . Papanicolaou, A. C. (2002). Dyslexia-specific brain activation profile becomes normal following successful remedial training. *Neurology, 58,* 1203–1213.

Someki, F., & Burns, M. K. (2009). Methods for measuring student response to stimulant medication: A meta-analytic review. *Psychology in the Schools, 46,* 388–396.

Stage, S. A., & Quiroz, D. R. (1997). A meta-analysis of interventions to decrease disruptive classroom behavior in public education settings. *School Psychology Review, 26,* 333–368.

STAR Early Literacy [Computer software]. (2004). Wisconsin Rapids, WI: Renaissance Learning.

Steinhausen, H. C., & Reitzle, M. (1996). The validity of mixed disorders of conduct and emotions in children and adolescents: A research note. *Journal of Child Psychology and Psychiatry*, 37, 339–344.

Stiggins, R. (2005). From formative assessment to assessment FOR learning: A path to success in standards-based schools. *Phi Delta Kappan, 87*, 324–328.

Swanson, H. L., & Howard, C. B. (2005). Children with reading disabilities: Does dynamic assessment help in the classification? *Learning Disability Quarterly, 28,* 17–34.

Tilly, W. D., III. (2008). The evolution of school psychology to science-based practice: Problem solving and the three-tiered model. In A. Thomas & J. Grimes (Eds.), *Best practices in school psychology* (5th ed., pp. 17–36). Bethesda, MD: National Association of School Psychologists.

VanDerHeyden, A. M., & Burns, M. K. (2005). Using curriculum-based assessment and curriculum-based measurement to guide elementary mathematics instruction: Effect on individual and group accountability scores. *Assessment for Effective Intervention, 30*(3), 15–29.

Witt, J. C., Daly, E. M., & Noell, G. (2000). *Functional assessments: A step-by-step guide to solving academic and behavior problems.* Longmont, CO: Sopris West.

Yearly Progress Pro [Computer software]. (2004). Columbus, OH: McGraw-Hill Digital Learning.

Ysseldyke, J. (1973). Diagnostic-prescriptive teaching: The search for aptitude–treatment interactions. In L. Mann & D. Sabatino (Eds), *The first review of special education* (pp. 1–37). New York, NY: Grune & Stratton.

Ysseldyke, J. (2006, April). *Using continuous progress monitoring data to make accountability decisions.* Paper presented at the annual meeting of the National Association of School Psychologists, Anaheim, CA.

Ysseldyke, J., & Bolt, D. (2007). Effect of technology-enhanced continuous progress monitoring on math achievement. *School Psychology Review, 36,* 453–467.

Ysseldyke, J., & Salvia, J. (1974). Diagnostic-prescriptive teaching: Two models. *Exceptional Children, 41*, 181–186.

Ysseldyke, J. E., & Christenson, S. L. (2002). *Functional assessment of academic behavior.* Longmont, CO: Sopris West.

Ysseldyke, J. E., & McLeod, S. (2007). Using technology tools to monitor response to intervention. In S. Jimerson, M. K. Burns, & A. M. VanDerHeyden (Eds.), *Handbook of response to intervention: The science and practice of assessment and intervention* (pp. 396–407). New York, NY: Springer.

Jim Ysseldyke *(CEC MN Federation), Birkmaier Professor of Educational Psychology;* **Matthew K. Burns** *(CEC MN Federation), Associate Professor of Educational Psychology;* **Sarah E. Scholin***, Graduate Fellow, Educational Psychology; and* **David C. Parker***, Research Assistant, Educational Psychology, College of Education & Human Development, University of Minnesota-Twin Cities.*

Address correspondence to Jim Ysseldyke, Educational Psychology, 342 EdSciB, 56 East River Road, Minneapolis, MN 55455 (e-mail: jim@umn.edu).

TEACHING Exceptional Children, *Vol. 42, No. 4, pp. 54–61.*