9
Assessing Student Response to Intervention

Stephanie C. Olson, Edward J. Daly III, Melissa Andersen, April Turner, and Courtney LeClair

Stephanie C. Olson, MA, is a graduate student in Educational (School) Psychology at the University of Nebraska-Lincoln. scolson@bigred.unl.edu
Edward J. Daly III, PhD, is Associate Professor of Educational (School) Psychology at the University of Nebraska-Lincoln. edaly2@unl.edu
Melissa Andersen, MA, is a graduate student in Educational (School) Psychology at the University of Nebraska-Lincoln. melissandersen@gmail.com
April Turner, MA, is a graduate student in Educational (School) Psychology at the University of Nebraska-Lincoln. april_d_turner@yahoo.com
Courtney LeClair, MA, is a graduate student in Educational (School) Psychology at the University of Nebraska-Lincoln. cleclai1@bigred.unl.edu

Seen by many as a significant educational innovation with far-ranging implications for how school districts respond to the needs of their students, the notion of "Response to Intervention" (RTI, upper case) has taken on immense proportions; justifiably so, in our view. RTI will directly affect the educational experience of millions of students nationwide. School districts are revamping their processes for classifying students with learning disabilities. Educators are now investing significant time, effort, and resources in screening processes to identify students' risk status. School personnel are combing the intervention literature to find strategies that can be implemented locally. Administrators are stuttering like David Bowie when considering the "*ch-ch-ch-changes*" that need to take place in their schools to live up to this new mandate.

The importance of these events for the overall integrity of RTI as a broad innovation cannot be overstated. Yet, if we lose sight of the elegant simplicity of the fundamental rationale, logic, and methods associated with RTI, there is a risk of drifting off course and forgetting the purpose of these changes. The pattern is clear and has been established through many cycles of educational reform: innovations have a tendency to eventually become simply a series of procedural steps that represent nothing more than an "add on" to existing, ineffective educational practices (Fullan, 2001). Someone somewhere will make up a checklist that fulfills RTI requirements and haggard-looking former visionaries will resign themselves to routinely complying so as to dig themselves out from under the overwhelming case loads that snuffed out their spark.

At the risk of oversimplifying the many complex dimensions of RTI, this chapter will unfold the basic concept of response to intervention (lowercase) as an organizing rubric for the activity of assessment. Our goal is to bring clarity to how practitioners conceptualize and carry out their assessment role in the RTI process as it relates first of all to student learning. After all, the primary purpose of assessment should always be improving student learning. However, the data generated through these assessments will likely provide a database for categorical decisions, like eligibility for special education, as schools move toward full-scale implementation of RTI. Therefore, we hope that the principles and practices described in this chapter will also help to improve the quality of the databases that will be used for high-stakes decisions like eligibility for special education when administrative action is in order. The purpose of this chapter is to provide a conceptual map for assessment activities to guide the questions that are asked and how one goes about answering those questions within RTI.

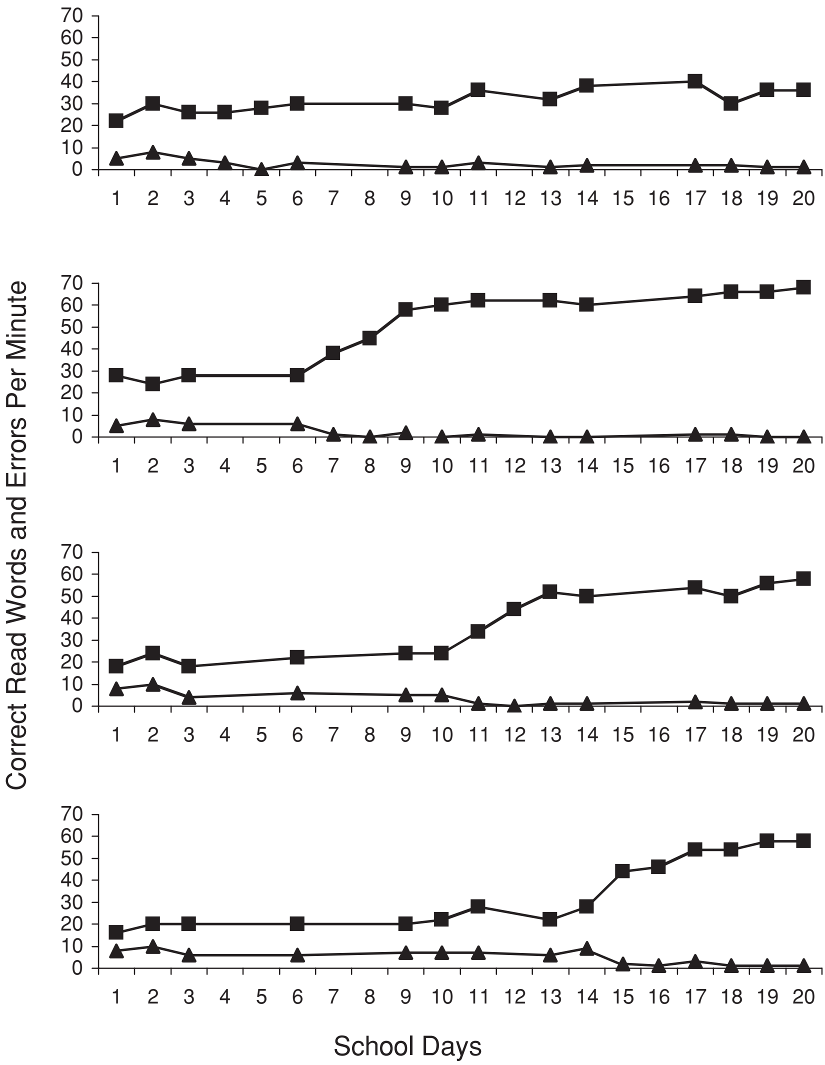

FIGURE 9.1. Hypothetical examples of fluency data (corrects and incorrects) displayed in single-case, A–B designs.

9.1 Use of an Evaluation Design in Response to Intervention

Psychologists get themselves into trouble when they fail to use predetermined evaluation criteria for important decisions about human problems (Dawes, 1994). Having a strong evaluation design reduces the likelihood of cognitive heuristics, post-hoc explanations, and other judgment errors (Barnett, 1988). The standard that has developed for data-based problem solving is the use of single-case accountability designs, like the A–B design (Barnett, Daly, Jones, and Lentz, 2004). Hypothetical examples of intervention outcomes arranged as a series of A–B designs appear in Figure 9.1. In the RTI process, the evaluation design involves repeated measures across different phases of instruction, each of which includes an assortment of instructional and/or motivational variables that reflect elements of the natural environment. In Figure 9.1, each graph has a baseline which serves as the point of comparison for an intervention that is applied repeatedly over time. Each graph in the example also has an intervention phase in which some planned modification of the environment is carried out. The result is that projections are made about the trajectory of student learning under various instructional conditions.

Decisions about intervention effectiveness are based on visual analysis of level, trend, and variability in the data across and within phases (Kazdin, 1982; Parsonson and Baer, 1992). Structured criteria for visual inspection which determine statistical significance and which do a good job of controlling decision errors have been recently developed and could be used as well (Fisher, Kelly, and Lomas, 2003). Problems with academic skills are most frequently behavioral deficits. Interventions, therefore, are expected to lead to increases in performance over time (i.e., changes in level and trend). An unsuccessful intervention phase would lead to results that do not differ from the baseline phase.

Comparisons between conditions are planned to test hypotheses about when a student is more or less likely to respond to some kind of environmental arrangement. When these comparisons are done within an adequate evaluation design with variables that reflect elements of the natural setting, generalizations (i.e., inferred meaning) are stronger because competing explanations have been ruled out and the results have direct implications for students' instructional needs. Although it is the student's responding that is being measured, it is the *instruction* that is being scrutinized (Englemann, Granzin, and Severson, 1979). Therefore, assessment is essentially a process of testing instruction through *response-guided experimentation* (Barlow and Hersen, 1984). Changes in student responding (or a lack thereof) within and across phases of instruction serve as feedback about the effectiveness and appropriateness of the instruction. The evaluation design and the data in the graphs are used as a basis for determining what should be done next. Within this model of assessment, an accurate description of the relationship of student responding to instruction is vital to guiding how instruction should be changed over the course of time. If student responding does not improve as expected following instruction, then subsequent instruction should increase in intensity and/or be differentiated in some way from previous instruction. The process is iterative until a solution is achieved. If student learning does not improve before the process is terminated, then *we* are the nonresponders (and not the students). The discussion will return to the examples in Figure 9.1 several times as assessment questions and practices are addressed. The graphs within the figure will be labeled in different ways throughout the chapter for purposes of illustration across examples. In addition, data from an actual case will be presented to round out illustration of many of the points.

9.2 Using Skills Assessment to Describe Problems with Student Responding

"Assessing student response to intervention" is a fitting description of the underlying purpose of assessment. Assessments are designed to detect responses which are presumed to have significance that transcends their measured occurrence. A student's response is the focal point of inference about the "meaning" of assessment results. The meaning directs the evaluator's decisions and future actions regarding the student. For example, student responses are used daily by evaluators across the country to deduce disabilities and risk status. However, the response is loaded with implications that may escape the attention of the evaluator if they fail to take note of the events that precede a measured response. The evaluator can avoid speculation about the meaning of student responses and instead make those responses all the more significant when he or she purposefully arranges or manipulates the events that precede student responding during assessment. By intentionally investigating how student responding changes as a function of instructional materials and demands, the evaluator enhances the meaning of assessment (Barnett et al., 2004).

The most natural starting point for assessing student learning in a classroom or curriculum is to note whether student responses to instructional tasks are correct or incorrect. Obviously, over time, correct responses should increase and incorrect responses should decrease as a function of instruction. More specifically, the nature of academic responding is such that correct responses should increase in frequency, rapidity, and consistency across instructional tasks. An observer will note that a response that was not initially in the student's repertoire may begin to increase in frequency (as errors decrease) when the student is presented with an instructional item. In other words, the learner's responses become *accurate* when presented with the instructional task. As accuracy improves, responses become more rapid and *fluency* develops. *Consistency* in responding emerges when the student answers correctly when presented with similar instructional items and/or instructional items that require the same or a similar response. For instance, a second grader who has "mastered" double-digit addition with regrouping can presumably calculate any combination of numbers, even combinations not directly taught by the teacher. Consistency also is a factor when the skill is used to accomplish a larger task that requires a broader repertoire of skills; the skill is used in conjunction with other skills to achieve an overarching outcome. For example, this same student should also eventually be able to use their computation skills to accomplish other tasks, like completing a science experiment. In this case, the double-digit addition with regrouping skill is one *component* of a *composite* skill that requires multiple component skills (e.g., reading the science text, following directions in order).

Student response to intervention, therefore, is the degree to which responding changes in terms of accuracy, fluency, and consistency within and across a variety of tasks, with improvements in *all* of these areas being critical to successful student performance. A deficiency in any of these areas signals that there is a problem and a need for further investigation. At the risk of overstating the obvious, it is worth noting that students are referred to evaluators because they exhibit fewer correct responses than desired or expected and a more systematic evaluation of student responding is necessary. Fortunately, there are highly developed, standardized procedures for directly assessing accuracy and fluency of basic skills. Curriculum-based measurement (CBM; Shinn, 1989) provides information regarding rate of responding, which reflects a combination of both accuracy and fluency of responding. CBM is widely popular and has become a standard practice in graduate training programs in school psychology (Shapiro, Angello, and Eckert, 2004). Given that fluency is an indicator of both accuracy and speed of responding and that it is a better measure of response strength than accuracy alone (Binder, 1996), assessments of basic skills should measure fluency.

Having a fluency score is only a part of knowing what the score means, however. An analysis of student responding will be incomplete if it does not account for the type of academic material given to the student as a part of the assessment. The material used for assessment will reflect the type of consistency and generality of responding being investigated. For instance, repeatedly assessing responding in materials instructed by the teacher yields information regarding consistency over time, referred to as *response maintenance*. Graphs B, C, and D in Figure 9.1 might reflect outcomes of instruction or a planned intervention across three different passages used by the teacher. Results are staggered because the teacher instructs the stories sequentially. If the teacher stopped instruction in earlier stories when moving on to subsequent stories, the latter data points in each graph would reflect maintenance once instruction was withdrawn.

An actual example of maintenance data from a reading intervention done with a ninth-grade student appears in Figure 9.2. Intervention was carried out over several days and the results were measured across three conditions: reward, instruction/taught materials, and instruction/untaught materials (these conditions will be described in more detail below). In order to measure response maintenance, the student was assessed two more times in each condition the week following withdrawal of the intervention. For the instruction/taught materials condition (the top data series in Figure 9.2), the data reveal that the student improved significantly throughout the intervention phase, and the performance leveled off following withdrawal of the intervention. It can be concluded that the student maintained his improvements because the performance during the maintenance phase was close in level to the performance during intervention. In the instruction/untaught materials condition (the middle data series), minor improvements in performance occurred during intervention. During the maintenance phase, there was a drop in performance. Finally, in the reward condition (the bottom data series), which served as a type of control condition, no performance increases were witnessed. These maintenance data reveal that the student's performance in the instructional conditions led to performance improvements that persisted when instruction was terminated.

A complete assessment should go beyond merely measuring what has been taught. The curriculum material taught by the teacher really only represents a subset of material in which the student should show improvement (Alessi, 1987). In an earlier example of consistency of responding, we pointed out that it is highly unlikely that the teacher would directly teach the student every possible number combination for double-digit addition with

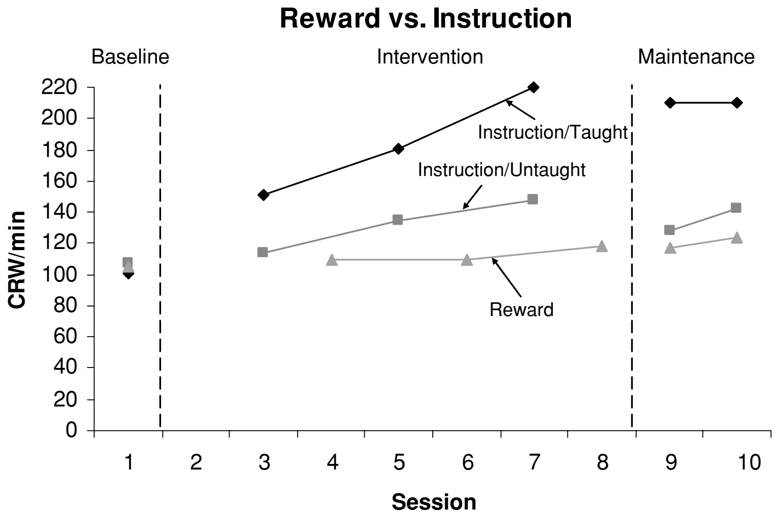

FIGURE 9.2. Example of correct read works per minute (CRW/min) in reward versus instruction conditions and in taught versus untaught passages.

regrouping problems. Curriculum material can be divided into *directly taught* tasks, or instructional materials, and *untaught* (but presumably equal difficulty level) tasks, or instructional materials. Assessing skill proficiency in untaught tasks provides information about the degree to which student responding is generalizing to similar instructional items. For example, an evaluator who assesses a second-grade student's reading fluency in a second-grade reading series that differs from the one used by the teacher in the classroom is assessing generalization of reading fluency across second-grade passages. The evaluator might choose to sample student performance over time using classroom instructional materials, which could be represented by the results in graph A, and separately sample student performance over time using an independent reading series, which could be represented by the results in graph B. Graph A provides information about changes in student responding in directly taught materials. Graph B provides information about how well the student is generalizing to untaught but equivalent difficulty level material. This information is probably even more important than the information in graph A, because it reveals how broad the effects of instruction are.

Figure 9.2 also displays an example of generalization to untaught materials. The ninth-grade student participated in a fluency intervention, in which he worked one-on-one with an experimenter, repeatedly reading a particular passage and receiving corrective feedback on his performance. Following practice, assessment data were collected by having the student read two different passages: the instructional passage, which was practiced as part of the intervention (instruction/taught materials); and another passage, which had not been practiced during the intervention (instruction/untaught materials). One would expect to see large increases in the instruction/taught materials condition, since the student practiced with those exact materials. In addition, one would hope to see increases in the instruction/untaught materials condition, since that would suggest overall improvements across grade-level materials (generalization). However, it would be expected that such improvements would be modest and gradual. Figure 9.2 reveals that these expectations were, indeed, met: the student demonstrated significant improvements in the taught materials and modest improvements in the untaught materials, suggesting some degree of generalization.

If a student improved in taught materials but did not improve in untaught materials, then the teacher's job is not done. The student is likely to struggle if he is moved up in the curriculum before consistency in responding across grade-level instructional tasks is achieved. For these reasons, priority should be given to measuring responding to untaught material over time as a basis for judging whether the effects of instruction have generalized sufficiently for the student to be ready to move on in the curriculum.

Consistency of responding is also vital to skill use when the skill is a necessary part of a larger repertoire of skills which are coordinated into a composite skill. For example, a student may be able to pronounce phonemes (sounds) when presented with letters on flashcards (e.g., pronouncing "b") and even be able to blend those phonemes to form words that the student was previously unable to read (e.g., "tab" and "cab"). However, the student still needs to be able to read those words in connected text and even blend untaught phonemes when he encounters an unfamiliar word in text. An assessment that evaluates skill proficiency in the context of critical composite skills produces valuable information about the student's ability to generalize the skill (and hence about its consistency in the presence of new and more complex problems or tasks). In this case, the results of graph A might reflect outcomes of phoneme blending assessments (in which fluency with phoneme tasks is repeatedly assessed with words) and the results of graph B might reflect oral reading fluency outcomes in phonetically regular passages that contain phonemes instructed in isolation by the teacher. As in the prior example, graph A indicates progress in the taught skill and graph B indicates progress in use of the skill when applied to a composite skill that appears as a later objective in the curriculum.

Similarly, reading comprehension could be viewed as a composite skill requiring the component skill of reading fluency (e.g., Pinnell et al., 1995). Thus, improvements in reading fluency may contribute to improved comprehension outcomes. For example, the ninth-grade student was asked to practice reading high-school-level passages. The student repeatedly practiced the first third of the passage. Fluency was then assessed in the second third of the passage, and comprehension was assessed through a cloze procedure in the final third of the passage. In the cloze procedure, every sixth word was replaced with a blank, and the student was instructed to provide words to replace the blanks. Figure 9.3 displays the results for two separate passages. The data indicate that improvements in reading comprehension correspond to improvements in generalized reading fluency. Indeed, the comprehension data show similar trends and changes in level as the fluency data.

9.3 Arranging Assessment Conditions to Figure Out What to Do About the Problem

Evaluating accuracy, fluency, and consistency/generalization may not be very satisfactory if assessment information is not related in some way to what can be done about the problem. Fortunately, the evaluation of these various dimensions of responding can also guide assessors in determining what to do about the problem. If one treats the assessment process as an opportunity to ask a series of questions, then assessments of skill fluency can be designed as mini-experiments that shed light on potentially effective and ineffective interventions that can be examined over time (Daly, Witt, Martens, and Dool, 1997). A series of questions is proposed that can be readily answered through planned instructional trials and ongoing fluency assessments. Our recommendation is to examine simple solutions first and progress to more complex interventions only as necessary.

A relatively simple initial question about how to change student performance is whether it can improve with rewards (Daly et al., 1997). If responding improves with rewards contingent on prespecified goals, then additional instructional support may be unnecessary to promote accurate, fluent, and consistent responding (Duhon et al., 2004; Eckert, Ardoin, Daisey, and Scarola, 2000; Eckert, Ardoin, Daly, and Martens, 2002; Noell et al., 1998). The advantage is that demands on those responsible for the intervention are minimized. For example, Duhon et al. (2004) developed a simple strategy for examining whether rewards or additional instruction were necessary to improve the performance of four students who had been referred for writing or math difficulties. A 2-min math calculation probe and a 3-min writing probe were administered to an entire class that included the four referred students. Brief, individual assessments were then conducted with each of the four students. During these assessments, performance goals were communicated to the students and rewards were offered for meeting the performance goals. Two of the students significantly improved their scores with rewards only. The other two students did not respond to rewards and required additional instructional assistance. Extended analyses of results confirmed the

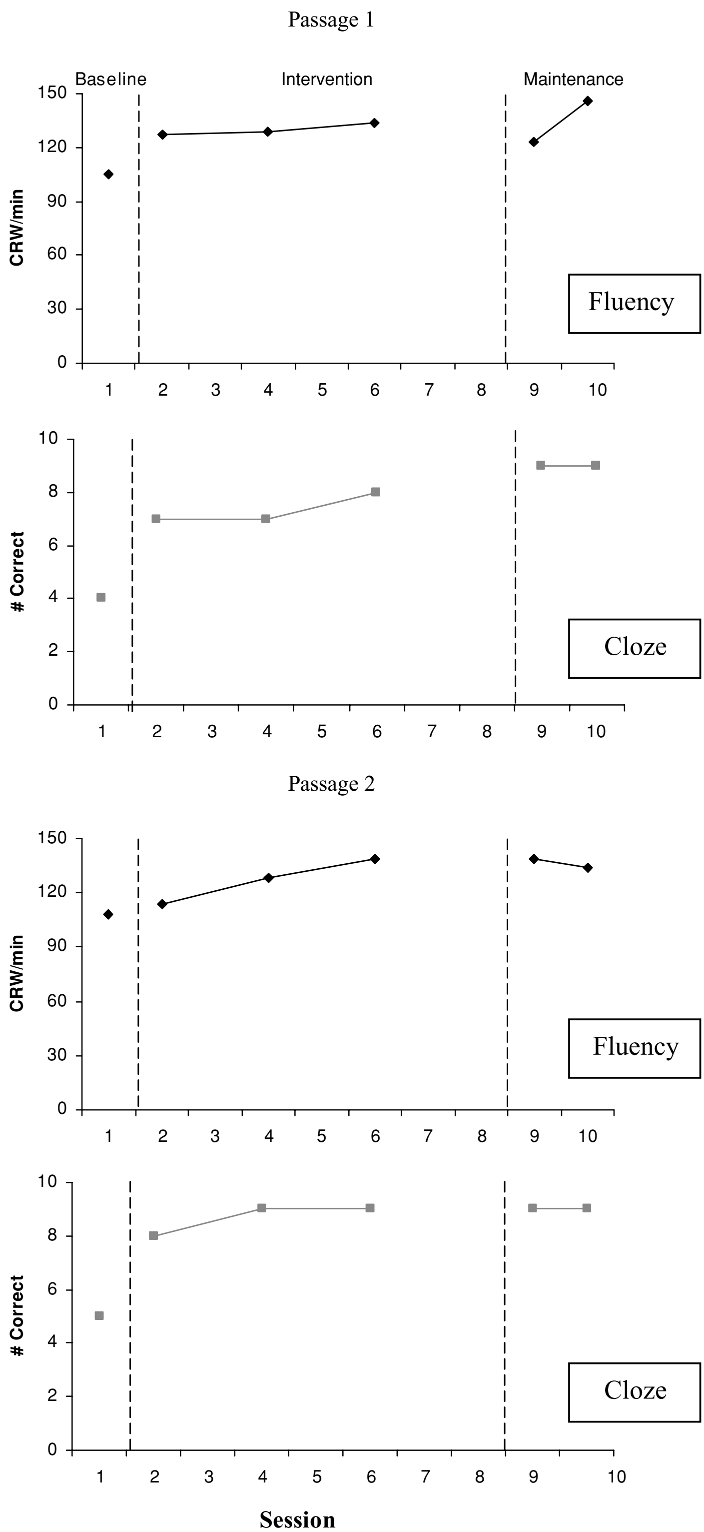

FIGURE 9.3. Relationship between fluency and comprehension.

conclusions of the initial assessment. In this study, Duhon et al. (2004) expanded standard CBM procedures by adding procedures to determine whether students would respond to a motivational strategy or an instructional strategy.

If a student does not improve responding following rewards, then instructional strategies should be investigated as a next step. For example, use of rewards might lead to limited outcomes like those presented in graph A. Instructional strategies might then be applied to other instructional materials, leading to results like those depicted in graphs C and D. Strategies that improve accuracy and fluency include modeling, practice, error correction, and performance feedback (Daly, Lentz, and Boyer, 1996a; Eckert et al., 2002).

Instructional strategies can be directly compared with a reward condition. As alluded to previously when discussing Figure 9.2, the study involved comparison of the student's performance in instruction and reward conditions. The instructional package included practice, error correction, and performance feedback, and fluency effects were assessed in taught and untaught materials. In addition, another passage was used to assess fluency improvements when the student did not receive instruction and was instead offered a reward for improving upon his previous score. The data which appear in Figure 9.2 indicate that the student hardly improved in the reward condition but did better in the instructional condition in both taught and untaught passages. Therefore, an effective intervention for that student would clearly require the use of instructional strategies.

Teachers, parents, and students themselves have been taught to use reading fluency interventions, such as listening passage preview (modeling fluent reading for the student), repeated readings (having the student repeatedly practice a passage), phrase drill error correction (having the student repeatedly practice phrases with error words), and performance feedback (telling the student how accurately and fluently they read the passage) (Bonfiglio, Daly, Persampieri, and Andersen, 2006; Daly, Persampieri, McCurdy, and Gortmaker, 2005; Gortmaker, Daly, McCurdy, Persampieri, and Hergenrader, in press; Persampieri, Gortmaker, Daly, Sheridan, and McCurdy, 2006). These strategies can be examined individually (Daly, Martens, Dool, and Hintze, 1998; Jones and Wickstrom, 2002) or in combination with one another. For example, Daly, Martens, Hamler, Dool, and Eckert (1999) systematically evaluated combinations of intervention strategies by sequentially adding treatment components. The results suggested that some students required simpler interventions and some required more complex intervention packages. For example, if the strategy used for intervention in graph C was procedurally simpler than the strategy used for intervention in graph D, then the former strategy is preferred for that student. Each of these strategies is directly applicable to any reading text and easily tested out in a single or a small number of sessions (Daly, Chafouleas, and Skinner, 2005).

Figure 9.4 also illustrates this point. In this case, the ninth-grade student with deficits in reading fluency was exposed to two different intensities of intervention; both involved repeated readings, phrase drill error correction, and performance feedback, but one was very brief, lasting about 5 min (low-intensity condition), while the other was more time consuming, lasting about 25 min (high-intensity condition). Assessment data were collected immediately following intervention in the same passage in which intervention occurred. The data indicate that the high-intensity intervention led to greater improvements than the low-intensity intervention. This could especially be seen in the maintenance data. Although the high-intensity intervention was more effective, strong effects were also seen with the low-intensity intervention, suggesting that it could be an appropriate replacement if significant time constraints were present.

In some cases, generalization to untaught material might not be observed. Therefore, as a next step, rewards should be combined with instructional strategies. Daly, Bonfiglio, Mattson, Persampieri, and Yates (2005) improved generalized reading fluency when instructional strategies like listening passage preview, repeated readings, and error correction were carried out prior to offering a reward for meeting performance goals. The instructional strategies were applied to different passages from those in which rewards were promised. What both types of passage shared in common were many of the same words (written in a different order). Therefore, combining instructional and reinforcement strategies may produce generalized word reading in some cases, especially when generalized improvements are reinforced.

If student responding still does not improve, then two strategies should be tried. First, consider reducing the difficulty level of instructional material

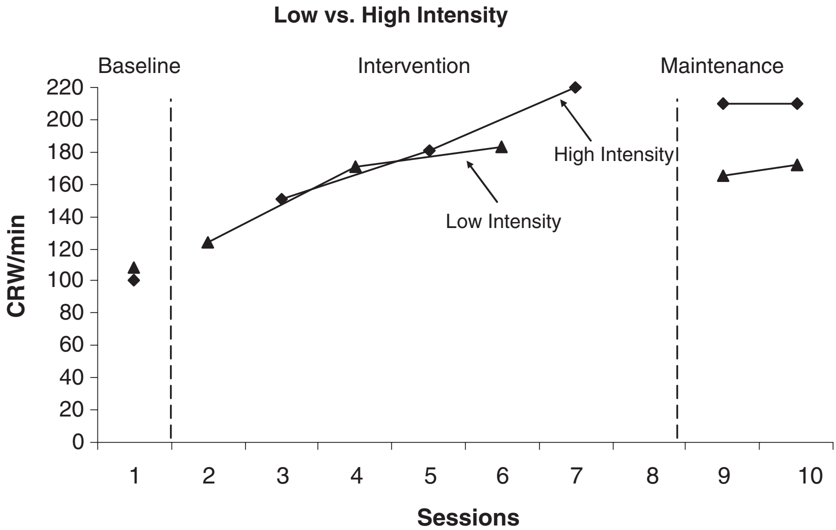

FIGURE 9.4. Example of correct read works per minute (CRW/min) in low-intensity versus high-intensity instructional conditions.

by moving down in the curriculum. Daly, Martens, Kilmer, and Massie (1996b) found greater generalization of reading fluency when difficulty level was better matched to students' instructional level (i.e., the materials were not too hard). Similar results were found by VanAuken, Chafouleas, Bradley, and Martens (2002). A second strategy is to teach responding in isolation first (e.g., by using flashcard exercises for word reading or math problems) before having the student practice in context (Daly et al., 1996b). For example, the teacher may have the student practice difficult words from texts on flashcards before having the student practice reading the story that contains those words.

9.4 Instructional Validity: Directly Assessing Instruction

We have emphasized how assessment of student responding repeatedly over time is the measure of instructional effectiveness. Unfortunately, however, intervention plans are seldom followed consistently in the absence of some type of direct observation and follow-up on the part of a consultant (Noell et al., 2005). Therefore, a critical step in the process of assessing a student's response to intervention is documenting the validity of instruction. Even the most carefully and systematically chosen and tested intervention is limited by the extent to which it is delivered frequently and reliably and reflects sound principles of instruction.

Direct observation of instruction can be facilitated (and structured) by the use of the Instructional Validity Checklist that appears in Appendix A. Academic engagement improves student learning (Greenwood, 1994) and can serve as a valid indicator of student response to instruction (Barnett et al., 1999). Part I of the Instructional Validity Checklist allows for the collection of momentary time sampling data. Results can be summarized to indicate the percentage of time the student is actively engaged during instruction.

The assessment team should directly observe the instructional methods and behavior management strategies used by the teacher. Part II of the Instructional Validity Checklist contains principles of effective teaching that have been shown to be related to student outcomes (Witt, Daly, and Noell, 2000). These behaviors are listed on the second form of the Instructional Validity Checklist in a rating-scale format to guide your observation. Strong instruction is responsive to the student's responding and each of the strategies on the checklist should be used by the teacher *as necessary* to facilitate student engagement with the instructional task. The items on the checklist can serve as a point of departure for analyzing the quality of instruction or intervention episodes. It is important for teachers to be clear and direct when explaining and giving directions for a task. Eliminating ambiguity increases the amount of time a student can spend correctly engaging in the task rather than figuring out what they are to do or practicing incorrectly. Teachers can bolster clear and direct explanation through modeling the task and prompting the correct answer, two critical strategies for increasing response accuracy and fluency. Praising and rewarding student effort, good behavior, and even the smallest of successes are an important part of instruction because they increase student motivation and effort and help decrease behavior problems. Watching student practice increases both teacher awareness of the student's progress and the opportunities to provide positive and corrective feedback for the student's errors, which in turn facilitates accurate practice to increase correct responding. Note also items related to student proficiency, difficulty level and relevance of instruction to the student's problem, all of which address the appropriateness of instructional match. Finally, it is worth observing whether misbehavior is a problem that needs to be addressed directly during instructional sessions. If misbehavior seems to be a problem, then the reader is referred to Witt, VanDerHeyden, and Gilbertson (2004) for guidelines about what to do.

TABLE 9.1. Implications for Practice

1. Choose an evaluation design *prior* to student evaluation.
2. Plan repeated fluency assessments that examine consistency and generality of student responding.
3. Implement a planned intervention.
   a. Try rewards. If that doesn't work:
   b. Try instructional methods like modeling, practice, error correction, and performance feedback. If that doesn't work:
   c. Try combining rewards and instruction. If that doesn't work:
   d. Try reducing difficulty level of the material and/or teaching component skills in isolation.
4. Evaluate the instructional validity of intervention sessions by examining both student engagement and the teacher's instructional behaviors.
5. Use outcome data to validate or sequentially modify instruction until student responding reaches desired level of performance.

9.5 Implications for Practice

Table 9.1 includes a summary of implications for practice for each of the topics discussed in this chapter. The evaluation design should be chosen before anyone engages in assessment. It establishes the rules for determining the effectiveness of instruction in advance and acts as a deterrent to the temptation to make post-hoc judgments about effectiveness. Post-hoc judgments are notoriously biased and inconsistent (Dawes, 1994), which will have an adverse effect on student outcomes. Use of single-case designs for evaluation and accountability purposes is strongly advised.

With respect to assessments to be conducted, an evaluator may spend almost as much time planning what will be assessed as actually conducting the assessments. Assessments should reflect important dimensions of student learning. Simple fluency assessment procedures for basic skills are readily available to educators, with CBM being the most prominent version. What requires careful deliberation is the selection of assessment tasks that produce information about the consistency and generality of students' skill proficiency. At a minimum, evaluators should routinely sample untaught but equal-difficulty-level tasks over time to check for generalization of responding. Failure to assess generalization provides an incomplete account of learning and may have a negative effect on the student if they are moved on prematurely in the curriculum. If instruction is targeted toward component skills (e.g., letter reading, phoneme blending) that should contribute to composite skills (e.g., reading words in text), then serious consideration should be given to monitoring the composite skill to determine whether instruction is impacting a student's ability to generalize use of the skill.

A wide variety of interventions appear in the literature and it is beyond the scope of this chapter to go into much detail on this topic. However, we have tried to provide guidelines in principles of instruction that are derived from our understanding of how accuracy, fluency, and consistency develop and which are broadly generalizable across skill types. For example, modeling of reading and modeling of math calculations look different in many ways. Functionally, though, they reflect the same principle of learning. The other guiding principle that we recommend is to begin with simpler interventions and increase in complexity only as necessary. It is important to keep in mind that the task may be a new one for the person who is responsible for implementing the intervention and also that they probably have other demands going on at the same time. A simpler intervention is more likely to be carried out (Lentz, Allen, and Ehrhardt, 1996).

If a student's data do not improve during an intervention phase, then it may be that the intervention was not carried out as planned. We provided an observation format that will allow documentation of instructional validity across a broad range of interventions. The best thing to do is to organize intervention steps into a step-by-step protocol (Witt et al., 2000). However, there will be many cases in which it might not be possible to do this. Evaluating student engagement and the teacher's instructional and management behaviors will provide some information about the quality of the intervention. Documenting intervention episodes over time may provide information about the frequency of intervention use, which is also likely to be an important factor in intervention effectiveness.

The real strength of this model is how data help educators be responsive to students' instructional needs. Outcome data will always indicate one of two things: (a) validation of intervention effectiveness (the desired outcome), or (b) the need for modification of the intervention plan. Fortunately, even when modification is necessary, previous intervention phases can be very "instructive" to educators, who often see ways that the intervention can be changed to promote better engagement and learning. In principle, the process repeats itself until student performance reaches an acceptable level. If we accept anything short of this, then we may be depriving a student of his or her right to a free and appropriate public education. Furthermore, this kind of direct, professional engagement with student outcomes is more likely to fan the spark of visionaries into a burning flame.

Appendix A

A.1 Instruction Validity Checklist Part I: Student Behavior

Student: Date: Instructor/Tutor: Time of Day: Length of Lesson:

Lesson Topic(s): (Check all that apply)	Phonemic awareness	Phonics/word study	Reading fluency	Comprehension	Spelling/writing

Active Student Engagement: Record student behavior at 10 second intervals using momentary time sampling; includes reading aloud, answering an academic question, asking an academic question, writing in response to teacher request, and silent reading (eye movements indicate student is scanning text).

Total for Row:

10	20	30	40	50	1 min	10	20	30	40	50	2 min	10	20	30	40	50	3 min	↓
10	20	30	40	50	4 min	10	20	30	40	50	5 min	10	20	30	40	50	6 min	
10	20	30	40	50	7 min	10	20	30	40	50	8 min	10	20	30	40	50	9 min	
10	20	30	40	50	10 min	10	20	30	40	50	11 min	10	20	30	40	50	12 min	
10	20	30	40	50	13 min	10	20	30	40	50	14 min	10	20	30	40	50	15 min	
10	20	30	40	50	16 min	10	20	30	40	50	17 min	10	20	30	40	50	18 min	
10	20	30	40	50	19 min	10	20	30	40	50	12 min	10	20	30	40	50	21 min	
10	20	30	40	50	22 min	10	20	30	40	50	23 min	10	20	30	40	50	24 min	
10	20	30	40	50	25 min	10	20	30	40	50	26 min	10	20	30	40	50	27 min	
10	20	30	40	50	28 min	10	20	30	40	50	29 min	10	20	30	40	50	30 min	

Minutes of Observation $\times 6 =$ _____ Observation Intervals Sum of Last Column = _____

Sum of Last Column/Observation intervals = _____% Active Student Engagement for the observation.

A.2 Instruction Validity Checklist Part II

Student: Date: Instructor/Tutor:

To be filled out during or immediately after observation of student engagement during instruction.

Instruction: Record the degree to which each of the teaching behaviors was observed to occur.

Explaining task/Giving directions	☐ Not done at all	☐ Done some of the time	☐ Done consistently, as necessary
Modeling/Demonstrating	☐ Not done at all	☐ Done some of the time	☐ Done consistently, as necessary
Prompting correct answer	☐ Not done at all	☐ Done some of the time	☐ Done consistently, as necessary
Praising and/or rewarding	☐ Not done at all	☐ Done some of the time	☐ Done consistently, as necessary
Watched student practice	☐ Not done at all	☐ Done some of the time	☐ Done consistently, as necessary
Corrected student errors	☐ Not done at all	☐ Done some of the time	☐ Done consistently, as necessary

How **proficient** was the student with assigned work during instruction?	Student answers were often incorrect (inaccurate)	Student answers were often accurate but slow (accurate but not fluent)	Student had difficulty giving correct answers across instructional tasks (generalization problem)

Was the task at an appropriate **difficulty level** for the student? YES NO

Was instruction stopped more than once to correct **misbehavior**? YES NO

Was **instruction relevant** to the student's skill problems? YES NO

Notes and Observations: _______________

____

____

References

Alessi, G. (1987). Generative strategies and teaching for generalization. *The Analysis of Verbal Behavior, 5*, 15–27.

Barlow, D. H. & Hersen, M. (1984). *Single Case Experimental Designs: Strategies for Studying Behavior Change* (2nd ed.). Boston: Allyn & Bacon

Barnett, D. W. (1988). Professional judgment: a critical appraisal. *School Psychology Review, 17*, 658–672.

Barnett, D. W., Bell, S. H., Gilkey, C. M., Lentz Jr., F. E., Graden, J. L., Stone, C. M., et al. (1999). The promise of meaningful eligibility determination: functional intervention-based multifactored preschool evaluation. *The Journal of Special Education, 33*, 112–124.

Barnett, D. W., Daly III, E. J., Jones, K. M., & Lentz Jr., F. E. (2004). Empirically-based special service decisions from increasing and decreasing intensity single case designs. *Journal of Special Education, 38*, 66–79.

Binder, C. (1996). Behavioral fluency: evolution of a new paradigm. *The Behavior Analyst, 19*, 163–197.

Bonfiglio, C. M., Daly III, E. J., Persampieri, M., & Andersen, M. N. (2006). An experimental analysis of the effects of reading interventions in a small group reading instruction context. *Journal of Behavioral Education, 15*, 92–108.

Daly III, E. J., Bonfiglio, C. M., Hauger, T., Persampieri, M., & Yates, K. (2005). Refining the experimental analysis of academic skill deficits, part I: an investigation of variables affecting generalized oral reading performance. *Journal of Applied Behavior Analysis, 38*, 485–498.

Daly III, E. J., Chafouleas, S. M., & Skinner, C. H. (2005). *Interventions for Reading Problems: Designing and Evaluating Effective Strategies*. New York, NY: Guilford Press.

Daly III, E. J., Lentz, F. E., & Boyer, J. (1996a). The instructional hierarchy: a conceptual model for understanding the effective components of reading interventions. *School Psychology Quarterly, 11*, 369–386.

Daly III, E. J., Martens, B. K., Dool, E. J., & Hintze, J. M. (1998). Using brief functional analysis to select interventions for oral reading. *Journal of Behavioral Education, 8*, 203–218.

Daly III, E. J., Martens, B. K., Hamler, K., R., Dool, E. J., & Eckert, T. L. (1999). A brief experimental analysis for identifying instructional components needed to improve oral reading fluency. *Journal of Applied Behavior Analysis, 32*, 83–94.

Daly III, E. J., Martens, B. K., Kilmer, A., & Massie, D. (1996b). The effects of instructional match and content overlap on generalized reading performance. *Journal of Applied Behavior Analysis, 29*, 507–518.

Daly III, E. J., Persampieri, M., McCurdy, M., & Gortmaker, V. (2005) Generating reading interventions through experimental analysis of academic skills: demonstration and empirical evaluation. *School Psychology Review, 34*, 395–414.

Daly III, E. J., Witt, J. C., Martens, B. K., & Dool, E. J. (1997). A model for conducting a functional analysis of academic performance problems. *School Psychology Review, 26*, 554–574.

Dawes, R. M. (1994). *House of Cards: Psychology and Psychotherapy Built on Myth.* New York: The Free Press.

Duhon, G. J., Noell, G. H., Witt, J. C. Freeland, J. T., Dufrene, B. A., & Gilbertson, D. N. (2004). Identifying academic skills and performance deficits: the experimental analysis of brief assessments of academic skills. *School Psychology Review, 33*, 429–443.

Eckert, T. L., Ardoin, S. P., Daisey, D. M., & Scarola, M. D. (2000). Empirically evaluating the effectiveness of reading interventions: the use of brief experimental analysis and single-case designs. *Psychology in the Schools, 37*, 463–474.

Eckert, T. L., Ardoin, S. P., Daly III, E. J., & Martens, B. K. (2002). Improving oral reading fluency: an examination of the efficacy of combining skill-based and performance-based interventions. *Journal of Applied Behavior Analysis, 35*, 271–281.

Englemann, S., Granzin, A., & Severson, H. (1979). Diagnosing instruction. *The Journal of Special Education, 13*, 355–363.

Fisher, W. W., Kelley, M. E., & Lomas, J. E. (2003). Visual aids and structured criteria for improving visual inspection and interpretation of single-case designs. *Journal of Applied Behavior Analysis, 36*, 387–406.

Fullan, M. G. (2001). *The New Meaning of Educational Change* (3rd ed.). New York, NY: Teachers College Press.

Gortmaker, V. J., Daly III, E. J., McCurdy, M., Persampieri, M. J., & Hergenrader, M. (in press). Improving reading outcomes for children with learning disabilities: Using brief experimental analysis to develop parent tutoring interventions. *Journal of Applied Behavior Analysis.*

Greenwood, C. R. (1994). Confirming a performance-based instructional model. *School Psychology Review, 23*, 652–668.

Jones, K. M. & Wickstrom, K. F. (2002). Done in sixty seconds: Further analysis of the brief assessment model for academic problems. *School Psychology Review, 31*, 554–568.

Kazdin, A. E. (1982). *Single-Case Research Designs: Methods for Clinical and Applied Settings.* New York: Oxford.

Lentz, F. E., Allen, S. J., & Ehrhardt, K. E. (1996). The conceptual elements of strong interventions in school settings. *School Psychology Quarterly, 11*, 118–136.

Noell, G. H., Gansle, K. A., Witt, J. D., Whitmarsh, E. L., Freeland, J. T., LaFleur, L. H., et al. (1998). Effects of contingent reward and instruction on oral reading performance at differing levels of passage difficulty. *Journal of Applied Behavior Analysis, 31*, 659–664.

Noell, G. H., Witt, J. C., Slider, N. J., Connell, J. E., Gatti, S. L., Williams, K. L., et al. (2005). Treatment implementation following behavioral consultation in schools: A comparison of three follow-up strategies. *School Psychology Review, 34*, 87–106.

Parsonson, B. S. & Baer, D. M. (1992).The visual analysis of data, and current research into the stimuli controlling it. In T. R. Kratochwill & J. R. Levin (Eds.), *Single-Case Research Design and Analysis* (pp. 15–40). Hillsdale, NJ: Lawrence Erlbaum Associates.

Persampieri, M., Gortmaker, V., Daly III, E.J., Sheridan, S. M., & McCurdy, M. (2006). Promoting parent use of empirically supported reading interventions: Two experimental investigations of child outcomes. *Behavioral Interventions, 21*, 31–57.

Pinnell, G. S., Pikulski, J. J., Wixson, K. K., Campbell, J. R., Gough, P. B., & Beatty, A. S. (1995). *Listening to Children Read Aloud.* Washington, DC: Office of Educational Research and Improvement, US Department of Education.

Shapiro, E. S., Angello, L. M., & Eckert, T. L. (2004). Has curriculum-based assessment become a staple of school psychology practice? An update and extension of knowledge, use, and attitudes from 1990 to 2000. *School Psychology Review, 33*, 249–257.

Shinn, M. R. (1989). *Curriculum-based Measurement: Assessing Special Children.* New York: Guilford Press.

VanAuken, T., Chafouleas, S. M., Bradley, T. A., & Martens, B. K. (2002). Using brief experimental analysis to select oral reading interventions: an investigation of treatment utility. *Journal of Behavioral Education, 11*, 163–181.

Witt, J. C., Daly III, E. J., & Noell, G. H. (2000). *Functional Assessments: A Step-by-Step Guide to Solving Academic and Behavior Problems.* Longmont, CO: Sopris West.

Witt, J. C., VanDerHeyden, A. M., & Gilbertson, D. (2004). Troubleshooting behavioral interventions: a systematic process for finding and eliminating problems. *School Psychology Review, 33*, 363–381.