8
Decision-Making Validity in Response to Intervention

David W. Barnett, Renee Hawkins, David Prasse, Janet Graden, Melissa Nantais, and Wei Pan

David W. Barnett, PhD, is Professor of School Psychology at the University of Cincinnati. david.barnett@uc.edu
Renee O. Hawkins, PhD, is Assistant Professor of School Psychology at the University of Cincinnati. renee.hawkins@uc.edu
David P. Prasse, PhD, is Professor and Dean, School of Education, Loyola University Chicago. dprasse@luc.edu
Janet L. Graden, PhD, is Professor of School Psychology at the University of Cincinnati. janet.graden@uc.edu
Melissa Nantais, PhD, is Educational Consultant at the Southwest Ohio Special Education Regional Resource Center. nantais_m@swoserrc.org
Wei Pan, PhD, is Assistant Professor of Quantitative Educational Research at the University of Cincinnati. wei.pan@uc.edu

Validity can be defined as the "approximate truth of an inference" (Shadish, Cook, and Campbell, 2002, p. 33). *Decision-making validity* can be viewed as the process of marshaling and weighing evidence to support actions (Messick, 1995). At first glance, these definitions alone do not sound too bad as criteria for professional decisions, but in considering response to intervention (RTI) we would need to include the validity of prevention efforts, measures and approaches to student selection, interventions in appropriate intensity sequences, and outcomes, among other variables, since we make inferences (i.e., conclusions) about all of these. Perhaps not surprisingly, there is a vast amount of literature that applies to the discussion of *decision-making* and *validity* that communicates both the strengths and weaknesses of human choice, the challenges of intervention evaluation, and, therefore, the many possible vulnerabilities of professional roles.

Professionals are valued when they assist with the highly challenging decisions needed to promote positive outcomes for individuals. To prepare for this role of decision-making consultant, professionals do the best they can by reviewing intervention research, applying problem-solving steps, and teaming. Consumers expect that professionals have mastered decision skills as well as validity ideals and apply them in a way that approximates perfection when offering advice, making instructional decisions, and intervening with children. Decision-making validity addresses this tension in RTI practice through examining prior and ongoing evidence of effectiveness.

This chapter provides both a general discussion of issues relevant to decision-making validity and more specific recommendations for strengthening validity arguments when implementing an RTI model. The first part of the chapter provides an overall context for decision-making validity in RTI, highlights the importance of establishing validity as a way to improve confidence in decisions, and examines the types and sources of validity evidence. The second part of the chapter offers suggestions for ways to build validity arguments.

8.1 Overview of Decision-Making Validity Issues

8.1.1 Context for Validity of Decision-Making within Response to Intervention

Decisions made within an RTI model operate from a different set of assumptions, practices, and areas of focus relative to traditional decisions made in a test-based model, so it is important to first recognize some of these critical differences. Foremost is the emphasis in RTI on demonstrated student need

based on outcomes using time-series data. This ongoing consideration and use of data differs from the traditional approach that focuses on eligibility determination based on child disability conditions at a single point in time. Consistent with recommendations from the President's Commission on Special Education (2002) and the National Association of State Board of Directors of Special Education (Batsche et al., 2005), and as described in other chapters in this book, we rely on a tiered model for RTI implementation. Decision-making validity is central in this model, with emphasis on determining child service needs using scientifically based and empirically demonstrated instruction and intervention, making decisions based on time-series data across tiers of varying intensity of services, and using important, or socially valid, child outcome data to judge success or need for instructional or intervention changes. In some well-developed RTI models (see other chapters), decisions are solely based on need for services, with no categorical differentiation, which is most consistent with the emphasis of RTI. Specific issues for decision-making validity within this approach will be highlighted throughout this chapter.

8.1.2 A Primer on Intervention Decision-Making

Many human information variables have been studied in decision-making that impact professional behaviors (Hastie and Dawes, 2001; Kahneman, 2003). Among them include time pressures, the types and amount of information available, qualities of information displays, and the order in which information becomes available (Barnett, 1988). Even simply recasting the descriptions of children's behavior may significantly influence judgments (i.e., "a child is aggressive towards peers" versus a replacement behavior such as "we need to increase successful play bouts") (e.g., Hall, Ashley, Bramlett, Dielmann, and Murphy, 2004). A complicated array of data can increase feelings of confidence in decisions ("looks like we have everything well covered") while potentially increasing actual error (i.e., a critical variable is more likely to be obfuscated by a clutter of data).

Examples of judgment errors include (1) diagnosing and intervening based on ideas that come easily to mind (*availability*), (2) limiting goals that may be set for children due to preconceived ideas about what they may be able to learn (*anchoring*), and (3) maintaining these initial biases, even despite data to the contrary, in favor of the preconceived ideas (insufficient *adjustment*) (Kahneman, 2003). In fact, individuals may not handle ambiguity and uncertainty all that well but these are pervasive characteristics of problem situations (Kahneman, 2003). Professionals may find themselves offering interventions that have been reinforcing to them because of past successes.

Errors of inference may be ubiquitous in decision-making and thus are not necessarily stamped out by RTI and problem solving or by another method. Answering questions addressed by validity is a major way to achieve confidence in decisions. However, new validity territory is introduced by RTI by shifting the focus to child outcomes and, therefore, raising questions of how to sequence interventions. A poorly planned sequence will consume unnecessary resources (if too intense), or unnecessarily keep a child in a prolonged failure experience (not sufficiently intense), or lead to erroneous conclusions (eligible as a child with a disability versus poor intervention sequence). Strategies to help reduce errors of intervention judgment include (1) applying a keystone target variable selection strategy functionally linked to success in typical environments and base rate information (Kame'enui, Good, and Harn, 2005; VanDerHeyden and Witt, 2005), (2) creating a range of plausible interventions based on prior research linked to the targeted variables, child characteristics, and supported by contextual or setting variables (Lentz, Allen and Ehrhardt, 1996), (3) communicating uncertainty, in that interventions result in likely *patterns* of outcomes and not specific outcomes, and (4) graphing student response data and applying valid decision rules to interpret data.

8.1.3 Confidence in Decision-Making

Practical validity questions for RTI have a dual role. First, practitioners will need to monitor progress at the system level to know that the overall RTI model is healthy and is doing its job in the best way possible. Outcomes supporting RTI validity indicating system health include reduced risk for children (e.g., improved reading outcomes, improved behavioral

outcomes), as well as satisfaction from consumers and participants, and indicators of sustainability. For example, system (school or district) data would indicate increased reading performance in third and fourth grade as a result of K-1 early literacy skill screening and interventions (Tier I). Since RTI will continue evolving with regard to research on instructional and social interventions, interpreting and implementing research are significant examples of decision-making validity. Validity checks will lead to ongoing RTI design modifications with new research.

Second, RTI requires monitoring decisions made for selected groups of children and individual children. Decision-making validity includes questions about the psychometric adequacy and utility of measures and criteria (benchmarks) selected for RTI use. Measures need to be correctly selected and accurate, and, when interpreted by teams, they need to link children to the most promising instructional or intervention alternatives. Decision-making through development of rules for selecting students for interventions and determining adequate progress for students receiving interventions, and problem solving, are used to satisfy the objectives pertaining to group and individual outcome determination. In summary, validity evidence for measures, selection procedures, and intervention sequences stand at the center of the RTI decision process.

8.1.4 Validity Evidence

8.1.4.1 Reliability Jumpstarts Validity

Many measures may be used throughout RTI in order to create data for decision-making. Reliability, typically defined as the consistency of measurement, has a direct relationship to RTI validity evidence as it connotes the allowable confidence in scores or observations used for decisions (Nunnally and Bernstein, 1994). Reliability facets would include internal consistency and accuracy of administration and scoring (before starting), as well as consistency in measuring a set of skills, behavior, or performance during intervention (ongoing). Decision confidence increases with the number of observations, items, or scores, but not justifiably if measures are not reliable, valid, or well sampled. Error rates for combined facets of reliability (e.g., scoring, internal consistency, and retest) that mirror natural decision-making, in that error sources are simultaneously active, are likely to be much higher than typically represented in test manuals (Macmann and Barnett, 1999).

Beyond instrument reliability, *procedural reliability* and the subset of intervention measurements known as fidelity, integrity, or adherence also underlie what can be said with confidence about intervention outcomes. Measures include not only student skill, performance, or behavior, but also include RTI model adherence and, ostensibly, instructional quality indicators from a verifiable model of instruction (e.g., Barnett, Ihlo, Nichols and Wolsing, 2006; McCardle and Chhabra, 2004; Twyman, Layng, Stikeleather, and Hobbins, 2005). In intervention research, low procedural reliability creates greater variability in outcomes that cannot be directly attributed to the intervention. By doing so, low procedural reliability creates lower effect sizes in research (Cohen, 1988) and questionable ethics in practice if decisions are made *as if* the intervention were carried out as planned (Gresham, 2004).

8.1.4.2 Construct Validity Connects the Dots

Construct use is pervasive in RTI, in that socially derived constructs are found in many areas of prevention, risk, and disability status, even though RTI constructs may not be recognized as such. Construct validity includes bigger ideas, such as academic achievement, social/behavioral risk, learning disability (chiefly because of its use as a federal category, although implementation varies by state – Iowa, for example, bases decision on need, not category), and RTI itself (e.g., Fuchs, Fuchs, and Speece, 2002). Construct validity also includes narrower domains or associated variables, such as reading (e.g., Fletcher and Francis, 2004; Kame'enui et al., 2005), academic or social engagement (e.g., Greenwood, Delquadri, and Hall, 1984), and intervention intensity (e.g., Barnett, Daly, Jones, and Lentz, 2004; Daly, Witt, Martens, and Dool, 1997; Gresham, 1991). From these examples, constructs are supported by networks of measurement as evidence. In RTI, child outcomes, the instructional environment, and interventions are measured. Construct validity gives this process of aligning measurement of construct-guided variables (i.e., risk), selection of children, and intervention the possibility of coherent analyses for cause-and-effect relationships. Construct validity addresses the unifying links and evidence,

including sampling adequacy, what is measured, how the data are interpreted (decision rules), how interventions are designed and evaluated, and how the next decisions are made. In other words, in intervention outcome research, construct validity provides the conceptual basis and foundation for understanding change based on measurement *and* intervention (Kazdin, 1998; Shaddish et al., 2002).

For example, a new instructional intervention may not only (a) provide creative and engaging lessons, but also (b) add considerable opportunities to practice the skill, (c) teach self-graphing to children for progress monitoring, and (d) provide additional rewards for improvements (i.e., reinforcement for increasing rates of fluency). In addition, the selection of certain children for the intervention is a critical part of analysis. Inadequate attention to selection may minimize or possibly exaggerate results (make outcomes difficult or relatively easy to achieve). Also, what is measured and how measurement samples are obtained allow different views of intervention outcome. Ideally, the intervention construct would include all key intervention facets (with corresponding measurement) as possible active ingredients in change. *Internal validity* provides arguments for attributing change to the intervention (cause and effect). *Statistical conclusion validity* addresses the analyses of any differences that might be found, but the processes or variables that *explain* change are questions of construct validity and would require ongoing measurement of relevant variables (Shaddish et al., 2002). In this example, significant independent variables, if measured, could include the engagement value of lessons (i.e., stimulus or conditions sampling), practice opportunities, scheduling and type of performance monitoring, self-graphing (i.e., accuracy of procedures, etc.), and reinforcement procedures (i.e., functional), plus undoubtedly other variables as well. Dependent variables could include different aspects of reading behavior if the focus is on students (e.g., Kame'enui et al., 2005) and instructional variables (i.e., changes in the qualities of practice) as the focus shifts to teachers or curriculum.

In summary, construct validity is used to help design and interpret studies through the selection and measurement of dependent and independent variables, and samples of students and teachers. The interventions are expected to move the children's performance measures consistent with measures used to select children and assign them to the appropriate intervention, and to help select criteria to judge outcomes. Interventions are construct linked, in that the children, measures, and interventions selected fit some conception of prevention, risk, or disability that could be used to explain change.

8.1.4.3 It Looks Like a Great Intervention, But Will it Work in My School?

Questions addressed by construct validity also help answer the questions of *external validity* or the degree that causal relationships are upheld over different settings, students, and other implementation variables (Shaddish et al., 2002). Selection of children for research may create samples that are quite different than child populations that professionals may face in schools. The best that we can do in most practice situations is "logical generalization" based on similarities between the research and our practice objectives, settings, and participants (e.g., Edgington, 1966; Hayes, Barlow, and Nelson-Gray, 1999).

8.1.4.4 Efficacy and Effectiveness Research

What validity evidence would support intervention practices? There have been a number of influential position papers addressing this question (Chambliss and Hollon, 1998; Kratochwill and Stoiber, 2002). *Efficacy* research shows the potential intervention outcomes under carefully controlled conditions. These conditions include screening and selecting participants, randomly assigning participants to groups (control and experimental, often not feasible in educational practice), and ensuring adherence to research protocols. Rather than comparing a new intervention with no intervention, comparisons with the best available rival intervention make efficacy studies critically important (Chambliss and Hollon, 1998). *Effectiveness* research looks at how well the intervention of proven efficacy can work in actual or more natural conditions. Questions include generalization, feasibility, and cost effectiveness, setting the bar quite high for researchers (Chorpita, 2003).

8.1.4.5 Single-Case Designs and Validity Evidence

Single-case designs provide a flexible and valid methodology for empirically evaluating interventions (Horner, Carr, McGee, Odom, and Wolery,

2005) and allow educators to assess the effectiveness of interventions for individual students, classes, and school systems in natural settings (Skinner, 2004). Both internal and external validity can be established through the use of single-case designs. Withdrawal, multiple-baseline, and changing-criterion designs allow for the repeated demonstration that an intervention systematically changes a given target variable (Barlow and Hersen, 1984). As a functional relation between intervention and behavior change is demonstrated and replicated, internal validity is established and the intervention becomes a plausible cause of behavior change. In practice, the internal validity of single-case designs can be strengthened by using control conditions and interventions with an empirical evidence base (Barnett et al., 2004). Designs such as alternating treatments also enable the rapid comparisons of alternative interventions to evaluate the most promising for a child (discussed later).

Of great potential importance for practice is the usefulness of single-case designs to address the actual application from the external validity evidence of interventions. Single-case designs provide a method for determining the generalizability of findings from controlled experimental studies to specific populations and individuals under applied conditions (Gresham, 2004). As procedures from efficacy research are replicated in natural settings, intervention effects in less-controlled environments can be evaluated.

8.1.4.6 Social Validity

Social validity evolved from single-case research (Wolf, 1978) to help evaluate intervention research through an expanded evaluation (i.e., participants, consumers, potential consumers) of satisfaction, appropriateness, and effectiveness of intervention goals, procedures, and outcomes. Methods for social validation include use of rating scales by teachers/parents to judge social validity, comparisons with various norms (i.e., peer comparisons), and evidence of sustainability (Kennedy, 2005). Social validity addresses many aspects of RTI, including the viability of the goals and methods of an intervention program prospectively and the viability of the goals, methods, and outcomes once the process is underway (Schwartz and Baer, 1991).

8.2 Response to Intervention Decision-Making Validity

RTI involves ongoing decision-making regarding instruction and intervention. Each decision affects the next as the process unfolds. Permanent product documentation, including graphs of universal screening results, group as well as individualized intervention outcomes (demonstrated through single-case design graphs), and decision rule use, is critical for decision confidence based on a comprehensive and cumulative record of the process.

8.2.1 Examining Validity Evidence

8.2.1.1 Target Variable Selection

Before selecting and implementing intervention procedures, a target variable is selected. There should be documented evidence that the variable targeted for intervention is appropriate. Data collected on the target variable must be evaluated by members of the problem-solving team to ensure that it is a direct measure of the problem, can be reliably measured over time, and will be sensitive enough to detect change resulting from the intervention (Macmann et al., 1996). Indirect measures (e.g., interviews, questionnaires) have generally not been shown to meet these criteria (i.e., reliable measurement over time, sensitive to growth), but may be used to generate a broader picture of the problem situation.

8.2.1.2 Instruction and Interventions, Vetting Criteria, and Sources

Once the target variable for change is clear, appropriate instruction and interventions need to be identified. There are numerous web resources available describing instruction and interventions for school-based problems (Table 8.1), but it is still necessary to be cautious with regard to evaluating effectiveness research and generalizing research to one's school and students. The challenges lie in determining which intervention will be most effective, most positive and natural, least costly, and least time consuming at a given point of time. Potential instructional approaches and interventions should be evaluated to determine (1) if they are appropriate and acceptable for universal (Tier 1), selected (Tier 2) or

TABLE 8.1. Examples of vetting sources.

University of Oregon	http://reading.uoregon.edu/curricula/index.php
Florida Center For Reading Research	http://www.fcrr.org/FCRRReports/index.htm
National Registry of Effective Practices	http://www.modelprograms.samhsa.gov/
What Works Clearinghouse	http://www.whatworks.ed.gov/
Intervention Central	http://www.interventioncentral.org
Edformation	www.edformation.com
Institute for the Development of Educational Achievement (IDEA)	www.idea.uoregon.edu
Positive Behavioral Interventions and Supports	http://www.PBIS.org

intensive (Tier 3) implementation (Gresham, 2004), (2) if they are designed to improve selected target variable performance, (3) if they are appropriate for the age and skill level of the students, and (4) if the school system has the resources to support proper implementation. If a chosen instructional approach or intervention is poorly matched on these criteria, then it is unlikely to have the desired effects on student performance and may lead to invalid decisions about the need for additional services. Also, although an intervention is empirically supported by efficacy research, the effectiveness of the intervention may still need to be determined in a natural setting, and these studies are rare (Chorpita, 2003). Interventions may need to be adjusted to meet the needs of a student or the resources of the system without losing effectiveness. Single-case methods may be used to provide answers about the feasibility of an intervention in a real-life situation and empirically "fine-tune" interventions to fit ecologies and children's needs.

8.2.1.3 Criteria for Judging Research Outcomes

Consumers of research need to judge the adequacy of the research design and procedures, statistical significance (the degree the results might be chance related?), size of effect (amount of change?), and social or clinical significance of the outcomes (Cohen, 1988; Foster and Mash, 1999; Kazdin, 1999; Wolf, 1978). Effect sizes estimate the amount of change measured in standard deviation units. An effect size of "1" means that data points represented in the intervention condition improved by one standard deviation over the control condition. Social validity includes broad methods relating change back to societal functioning.

8.2.1.4 Replicated Studies

Replications of efficacy and effectiveness strengthen intervention validity evidence and, thus, the validity of decisions to implement those procedures. Even when an intervention has been investigated through the primary methodology of efficacy research (i.e., randomized experiments), replicated studies of intervention effectiveness are especially important (Chambliss and Hollon, 1998; Horner et al., 2005; Stoiber and Kratochwill, 2000) to estimate and to purposefully influence external validity.

An intervention should not be overlooked as a potential solution to problem behavior for the sole reason that it has not been investigated through a randomized experiment. Single-case design researchers consider within- and between-series replication (i.e., ABAB, multiple baseline across participants, behavior, settings), not random assignment, to be the *sine qua non* of valid design, with replication across different participants and researchers building justifiable confidence in conclusions even further. Horner et al. (2005, pp. 175–176) suggest that an intervention may be considered *evidence based* by using single-case designs when:

> (a) a minimum of five . . . studies that meet minimally acceptable methodological criteria and document experimental control have been published in peer-reviewed journals, (b) the studies are conducted by at least three different researchers across at least three different geographical locations, and (c) . . . a total of at least 20 participants [are included across studies].

Practitioners in the field are continuously developing new and effective interventions to address student problems but are unable to establish cause–effect relationships for a variety of reasons (i.e., limited resources, teacher/parent preference not to return to baseline) (Skinner, 2004). However, by

sharing data on interventions developed and applied in the field, practitioners can begin the process to more extensively replicate procedures in order to establish the relationship and boundaries between the intervention and behavior change. From the parents' view, permissions and informed consent for services at Tier 3 would be based on estimates of established empirical confidence in the intervention, or an agreement to try newer procedures based on full knowledge of intervention alternatives.

8.2.1.5 Researched Principles of Learning

Familiarity with basic principles of learning also can help with the process of sorting through research to find the most appropriate intervention. Although there are numerous empirically supported interventions for school-related academic and behavior problems in research, many of these interventions share key components. Recognizing these important principles of learning can help when modifications of interventions are deemed necessary or when judging newer interventions. Common features of effective academic or social interventions include clarifying objectives, practice, feedback, and reinforcement principles (e.g., Shapiro, 2004; Sugai et al., 2000).

8.2.1.6 Decision Rules

Problem-solving teams should have data to support pre-established (nonarbitrary) decision rules that will be used to determine when adjustments to intervention protocols are needed. Empirically set decision rules are based on generalizations from past research with specific interventions (e.g., significant characteristics of sessions usually needed to produce effects; how long to keep a child in an intervention without making changes). As part of the permanent product record of the RTI process, these data provide evidence of the validity of decision rules for new student groups or individuals. Graphs of established benchmarks or local norms can provide a point of comparison as student response to intervention is monitored. Decision rules should also take into account base rate data (VanDerHeyden and Witt, 2005). Recognizing the prevalence of reading or social problems within a school or school system can inform decisions about what level of intervention support is needed (e.g., school-wide versus small group). In summary, by pre-establishing decision rules, decision-making validity can be examined.

8.2.2 Ongoing Assessment of Validity Evidence

8.2.2.1 Intervention is Implemented as Intended

The validity of decisions made through the RTI process can be significantly threatened if interventions are not implemented accurately (Gresham, 1989). "Accuracy" should include adherence to procedures and appropriate schedules of contact between student and the intervention (i.e., "dose" of intervention). The use of intervention scripts helps address this issue by providing a detailed outline of how intervention plans are to be implemented (Ehrhardt, Barnett, Lentz, Stollar, and Reifin, 1996). Scripts provide the individuals responsible for intervention implementation with a step-by-step contextual and natural guide, increasing the likelihood that the intervention will be implemented as intended. The scripts can also be used to document intervention adherence by providing a checklist of the implementation steps completed and occasions of use. Thus, scripts can be completed by the individual responsible for the intervention as a guide and used by an individual observing the intervention being implemented. Adherence data provide evidence that an intervention was implemented accurately and that change in behavior was likely due to effects of the intervention. These data are particularly important when a student is not making desired progress. Without evidence that the intervention was implemented accurately and as scheduled, it will be unclear as to whether a student's failure to make desired levels of progress is an indicator that they need additional intervention supports or an artifact of a poorly implemented intervention protocol. This evidence would need to be included in some format (e.g., co-plotted or referenced on a progress-monitoring graph; scripts with completed items checked off) as a permanent product in the intervention file (Ehrhardt et al., 1996). It is worth noting that some curricula (e.g., direct instruction approaches) have built-in methods to determine and provide a record of implementation. In general, evidence suggests that teachers may need considerable support for implementation (e.g., Noell et al., 2000).

8.2.2.2 Is the Intervention Having the Desired Effect?

As intervention procedures are implemented, the effects must be continuously monitored and documented. Graphs of student progress over time that include goal lines, aim lines, and conditions, can be used to provide evidence of intervention effectiveness or ineffectiveness. Pre-established decision rules from past research provide guidelines for data interpretation and when adjustments to interventions should be made. The frequent collection of progress-monitoring data will be needed to inform the ongoing evaluation process.

8.2.2.3 Intervention Components and Sequences

Comprehensive, multifaceted intervention packages have proven to effectively address the needs of students at high risk of school failure due to poor academic performance and/or highly disruptive behavior. However, all components of an intervention package may not be necessary for individual students to demonstrate progress and may unnecessarily and inefficiently use system resources. In addition, the more time consuming and difficult that intervention procedures are to implement, the less likely they will be implemented as designed (Gresham, 1989). If an intervention package includes components unnecessary for student progress and unlikely to be implemented accurately, then the validity of the decision to continue providing such services is significantly threatened.

Intervention sequence data can help problem-solving teams determine the level of support necessary for student success by examining the effects of increasing and decreasing the intensity of intervention designs (Barnett et al., 2004). Increasing-intensity designs start with the least intensive instructional intervention and add additional intervention components as necessary based on progress-monitoring data. Alternatively, decreasing-intensity designs start with more comprehensive interventions and elements of the intervention are systematically withdrawn. With both increasing- and decreasing-intensity designs, the goal is to ensure that intervention procedures are at the minimum level necessary to achieve desired levels of student performance. The data collected on the effects of systematically adding or withdrawing intervention components provides empirical evidence to validate decisions about necessary services for students.

8.2.2.4 Which Intervention is Best?

Using well-established methods, the validity of alternative interventions or reinforcers for individual students can also be established by behavioral assessments and single-case designs (e.g., Steege, Wacker, Berg, Cigrand, and Cooper, 1989). Referred to as *brief experimental analysis* (or *brief trial* designs in Table 8.2), exposures to alternative interventions that are pre-planned with regard to prior

TABLE 8.2. How to address decision-making validity in RTI.

Unit of analysis	Prior validity evidence	Ongoing validity evidence
RTI model	Validity evidence for model or at least components in reducing risk, etc.; social validity	Progress monitoring for key service delivery "events"; outcomes show reduced risk; social validity
Universal screening	Reliability and validity of measures, cut scores, or benchmarks	Progress-monitoring data leads to accurate decisions about risk
Universal prevention and intervention	Efficacy and effectiveness research on construct (academic or social risk prevention)	Progress-monitoring data leads to accurate conclusions about risk reduction
Target variable selection	Research in academic achievement and social behavior and its measurement	Progress-monitoring data allows evaluation of intervention effectiveness
Targeted interventions	Efficacy and effectiveness research; replicated single-case designs; replicated principles of learning	Single-case research, "brief trial" or accountability designs
Individualized intervention	Replicated single-case designs; replicated principles of learning	Single-case research or accountability designs; functional assessment and analysis
Eligibility for special services	Validity of specialized services	Validity of intervention "intensity" variables

efficacy for specific targeted behaviors may rapidly yield validity evidence. Applied to academic skills problems, brief experimental analysis has been used to test various empirically supported individualized interventions to improve reading performance (e.g., Daly and Martens, 1994; Daly, Martens, Dool, and Hintze, 1998). Students are exposed to different hierarchically arranged intervention conditions for a few sessions. Brief withdrawals and replications are then used to validate the most effective interventions (Daly et al., 1997). The alternating treatment design can show the relative effectiveness of two candidate interventions (Steege et al., 1989).

8.2.2.5 Functional Analysis

Rather than trying out likely interventions even briefly, functional analysis allows an understanding of a behavior by first examining hypothesized functional relationships, or patterns of behavior, that vary systematically by antecedents (or predictors) and/or consequences of behavior. First, teams hypothesize and establish the function, and then design the intervention based on function. *Brief functional analysis* procedures include brief exposure to manipulated conditions with replication of results (Steege and Northup, 1998). Crone and Horner (2003) provide decision rules to guide the levels of functional assessment and analysis based on risk appraisals for highly concerning behaviors. The primary objective is increasing the validity of an intervention design by establishing its function and, through the design, making the problem behavior irrelevant, inefficient, or ineffective. A primary example is functional communication training (Carr and Durand, 1985; Horner et al., 2005).

8.3 Conclusions

RTI is construct-linked with regard to theories of prevention (achievement and social risk), interventions ordered by intensity for struggling children, and decisions for special services eligibility for challenging-to-serve children. Many types of research are needed to support RTI, not only large-scale and single-case intervention research, but also research addressing measurement, selection, progress monitoring, and outcome evaluation. All of these involve complex decision processes and, thus, vulnerabilities to inaccurate decision-making. A strong model, procedures, and validity evidence for procedures are ways to improve decision-making. We have stressed validity evidence for interventions. While not meant to be inclusive, Table 8.2 is organized by the roles of examining existing and ongoing sources for validity evidence for RTI for practice.

On the surface, validity discussions look like they are for professionals and researchers. However, if one considers the consequences of decisions made, then RTI validity evidence is relevant to parents and any stakeholders who are invested in attaining positive outcomes for individual and groups of children. Such evidence will permit stakeholders to make informed choices about available services as much as it will help researchers and practitioners to evaluate potential RTI models and formatively enhance existing ones.

Acknowledgments. This work was partially funded by an Ohio Board of Regents grant.

References

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

Barnett, D. W. (1988). Professional judgment: a critical appraisal. *School Psychology Review, 17*, 656–670.

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

Barnett, D. W., Ihlo, T, Nichols, A., & Wolsing, L. (2006). Preschool teacher support through class-wide intervention: a description of field initiated training and evaluation. *Journal of Applied School Psychology, 23*, 77–96.

Batsche, G., Elliott, J., Graden, J. L., Grimes, J., Kovaleski, J. F., Prasse, D., et al. (2005). *Response to Intervention: Policy Considerations and Implementation.* Alexandria, VA: National Association of State Board of Directors of Special Education.

Carr, E. G. & Durand, V. M. (1985). Reducing problem behaviors through functional communication training. *Journal of Applied Behavior Analysis, 18*, 111–126.

Chambliss, D. L. & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology, 66*, 7–18.

Chorpita, B. F. (2003). The frontier of evidence-based practice. In A. E. Kazdin & J. R. Weisz (Eds.),

Evidenced-based Psychotherapies for Children and Adolescents (pp. 42–59). New York: Guilford.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Crone, D. A. & Horner, R. H. (2003). *Building Positive Behavior Support Systems in Schools: Functional Behavioral Assessment.* New York: Guilford.

Daly III, E. J. & Martens, B. K. (1994). A comparison of three interventions for increasing oral reading performance: application of the instructional hierarchy. *Journal of Applied Behavior Analysis, 27*, 459–469.

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., 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.

Edgington, E. S. (1966). Statistical inference and nonrandom samples. *Psychological Bulletin, 66*, 485–487.

Ehrhardt, K. E., Barnett, D. W., Lentz, F. E., Jr., Stollar, S. A., & Reifin, L. H. (1996). Innovative methodology in ecological consultation: use of scripts to promote treatment acceptability and integrity. *School Psychology Quarterly, 11*, 149–168.

Fletcher, J. M. & Francis, D. J. (2004). Scientifically based educational research. In P. McCardle & V. Chhabra (Eds.), *The Voice of Evidence in Reading Research* (pp. 59–80). Baltimore: Brookes.

Foster, S. L. & Mash, E. J. (1999). Assessing social validity in clinical treatment research: Issues and procedures. *Journal of Consulting and Clinical Psychology, 67*, 308–319.

Fuchs, L. S., Fuchs, D., & Speece, D. L. (2002). Treatment validity as a unifying construct for identifying learning disabilities. *Learning Disabilities Quarterly, 25*, 33–44.

Greenwood, C. R., Delquadri, J. C., & Hall, V. R. (1984). Opportunities to respond and student academic performance. In W. L. Heward, T. E. Heron, D. S. Hill, & J. Trap-Porter (Eds.), *Focus on Behavior Analysis in Education* (pp. 58–88). Columbus, OH: Merrill.

Gresham, F. (1989). Assessment of treatment integrity in school consultation and prereferral intervention. *School Psychology Review, 18*, 37–50.

Gresham, F. (2004). Current status and future directions of school-based behavioral interventions. *School Psychology Review, 33*, 326–343.

Gresham, F. M. (1991). Conceptualizing behavior disorders in terms of resistance to intervention. *School Psychology Review, 20*, 23–36.

Hall, J. D., Ashley, D. M., Bramlett, R. K., Dielmann, K. B., & Murphy, J. J. (2004). ADHD assessment: a comparison of negative versus positive symptom formats. *Journal of Applied School Psychology, 21*, 163–173.

Hastie, R. & Dawes, R. M. (2001). *Rational Choice in an Uncertain World: The Psychology of Judgment and Decision Making*. Thousand Oaks, CA: Sage.

Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The Scientist Practitioner: Research and Accountability in the Age of Managed Care.* Boston: Allyn & Bacon.

Horner, R. G., Carr, E. G., McGee, G., Odom, S., & Wolery, M. (2005). The use of single subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165–179.

Kahneman, D. (2003). A perspective on judgment and choice. *American Psychologist, 58*, 697–720.

Kame'enui, E. J., Good III, R. G., & Harn, B. A. (2005). Beginning reading failure and the quantification of risk: reading behavior as the supreme index. In W. L. Heward, T. E. Heron, N. E. Neef, S. M. Peterson, D. M. Sainato, G. Cartledge, et al. (Eds.), *Focus on Behavioral Analysis in Education: Achievements, Challenges, and Opportunities* (pp. 68–88). Upper Saddle River, NJ: Pearson.

Kazdin, A. E. (1998). *Research Design in Clinical Psychology* (3rd ed.). New York: Harper & Row.

Kazdin, A. E. (1999). The meanings and measurement of clinical significance. *Journal of Consulting and Clinical Psychology, 67*, 332–339.

Kennedy, C. H. (2005). *Single-Case Designs for Educational Research.* Boston Allyn & Bacon.

Kratochwill, T. R. & Stoiber, K. C. (2002). Evidence-based interventions in school psychology: conceptual foundations of the Procedural and Coding Manual of Division 16 and the Society for the Study of School Psychology Task Force. *School Psychology Quarterly, 17*, 1–55.

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

Macmann, G. & Barnett, D. (1999). Diagnostic decision making in school psychology: understanding and coping with uncertainty. In C. R. Reynolds & T. B. Gutkin (Eds.), *Handbook of School Psychology* (3rd ed., pp. 519–548). New York: Wiley.

Macmann, G. M., Barnett, D. W., Allen, S. J., Bramlett, R. K., Hall, J. D., & Ehrhardt, K. E. (1996). Problem solving and intervention design: guidelines for the evaluation of technical adequacy. *School Psychology Quarterly, 11*, 137–148.

McCardle, P. & Chhabra, V. (Eds.) (2004). *The Voice of Evidence in Reading Research.* Baltimore: Brookes.

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and

performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741–749.

Noell, G. H., Witt, J. C., LaFleur, L. H., Mortenson, B. P., Ranier, D. D., & LeVelle, J. (2000). Increasing intervention implementation in general education following consultation: a comparison of two follow-up strategies. *Journal of Applied Behavior Analysis, 33*, 271–283.

Nunnally, J. C. & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill.

President's Commission on Excellence in Special Education. (2002). A new era: revitalizing special education for children and their families. Retrieved August 2, 2002 from http://www.ed.gov/inits/commissionsboards/whspecialeducation/reports.html

Schwartz, I. S. & Baer, D. M. (1991). Social validity assessment: is current practice state of the art? *Journal of Applied Behavior Analysis, 24*, 189–204.

Shaddish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* Boston: Houghton Mifflin.

Shapiro, E. S. (2004). *Academic Skills Problems* (3rd ed.). New York: Guilford.

Skinner, C. H. (2004). *Single-Subject Designs for School Psychologists*. Binghamton, NY: Haworth.

Steege, M. W. & Northup, J. (1998). Brief functional analysis of problem behavior: A practical approach for school psychologists. *Proven Practice: Prevention and Remediation Solutions for Schools, 1*, 4–11, 37–38.

Steege, M. W., Wacker, D. P., Berg, W. K., Cigrand, K. K., & Cooper, L. J. (1989). The use of behavioral assessment to prescribe and evaluate treatments for severely handicapped children. *Journal of Applied Behavior Analysis, 22*, 23–33.

Stoiber, K. C. & Kratochwill, T. R. (2000). Empirically supported interventions and school psychology: Rationale and methodological issues. Part I. *School Psychology Quarterly, 15*, 75–105.

Sugai, G., Horner, R. H., Dunlap, G., Hieneman, M., Lewis, T. J., Nelson, C. M., et al. (2000). Applying positive behavioral support and functional behavioral assessment in schools. *Journal of Positive Behavioral Interventions, 2*, 131–143.

Twyman, J. S., Layng, T. V. J., Stikeleather, G., & Hobbins, K. A. (2005). A nonlinear approach to curriculum design: the role of behavior analysis in building an effective reading program. In W. L. Heward, T. E. Heron, N. E. Neef, S. M. Peterson, D. M. Sainato, G. Cartledge, et al. (Eds.), *Focus on Behavioral Analysis in Education: Achievements, Challenges, and Opportunities* (pp. 55–68). Upper Saddle River, NJ: Pearson.

VanDerHeyden, A. M. & Witt, J. C. (2005). Quantifying context in assessment: Capturing the effect of base rates on teacher referral and a problem solving model of identification. *School Psychology Review, 43*, 161–183.

Wolf, M. M. (1978). Social validity: the case for subjective measurement or how applied behavior analysis is finding its heart. *Journal of Applied Behavior Analysis, 11*, 203–214.