



Data Collection: Monitoring Children's Treatment Progress

Lesley B. Olswang

University of Washington, Seattle

Barbara Bain

Idaho State University, Pocatello

For many of us, not having data concerning a client's progress during treatment is tantamount to being unprepared for a lecture, or showing up at a birthday party without a present—totally unthinkable. This zealous position is based on the assumption that clinical decisions regarding treatment efficacy should be based on data. Data, in this case, refers to both quantitative and qualitative information that provides evidence for deciding the course of treatment. This is not to say that

data can and should be collected on every aspect of the clinical process. Indeed, intuitive decision-making on the part of the speech-language pathologist is often warranted. But in general, a series of decisions regarding whether or not treatment is working, can and should be based on data. This article examines the ways in which we can measure treatment progress, and provides guidelines for the reader in the use of a data-based, decision-making model.

The Value of Data

What may seem to be the most obvious reason for data collection, that is, accountability, is not the driving force behind our rationale. Accountability has increasingly come to mean the involvement of a third party; that is, evaluating productivity or documenting the efficacy of the clinical process for the purposes of convincing others about progress during treatment. Data collection clearly plays a significant role in this scenario, whereby the data are used to justify treatment decisions that have been made. However, data collection can be viewed as having a more integral part in the actual decision-making process—not serving after-the-fact to justify decisions, or prove effectiveness, but rather, serving as the means for making ongoing clinical decisions. In this view, data are a necessary part of the clinical process: without the data, informed decisions concerning the client's management cannot be made.

This particular perspective is based on two assumptions. First, the goal of treatment is to provide focused, intensive stimulation to alter a specific behavior or sets of behaviors. This target attack approach to intervention is based on the speech-language pathologist's observations that specific behaviors are missing, limited, or deviant, and that without intervention targeting these behaviors, continued progress

in language learning will be hindered. We realize that children with language disorders may have many needs and that intervention can be more global in design, attempting to stimulate language learning in general. However, our experience has suggested that at some time during the intervention process, particular behaviors become a focus of concern, deserving attention and intervention. The second assumption is that data are only as good as they are designed to be. That is to say, good clinical decisions can only be made from good data; bad data yield bad decisions. We are certainly not advocating an approach that urges the use of data at all costs.

Given our proactive perspective in the use of data, and with our stated caveats in mind, we will be discussing data collection for the purpose of making ongoing clinical decisions concerning client progress in treatment. Specifically, the following clinical questions seem to be critical to all speech-language pathologists, and further, these questions seem to be amenable to data-based decision-making: Is the client responding to the treatment program? Is significant, important change occurring? Is treatment responsible for the change? How long should a therapy target be treated? Although a host of other clinical questions could be raised for which data collection would be an appropriate vehicle for decision-making, these will not be included in the present discussion. The decision to limit this discussion is purely a pragmatic one, based on space limitations.

¹Brenda Y. Terrell served as the Associate Editor for the review of this article.

Clinical Decisions

Is the client responding to the treatment program? This is probably the first clinical question that is raised by the speech-language pathologist. With the best professional judgment and personal intentions, treatment approaches are selected and implemented with our clients. However, as with all helping professions, this decision is neither foolproof nor permanent. Thus, the speech-language pathologist needs to have some indication as to whether the selection of the treatment target(s) and treatment techniques(s) was reasonable and whether the client is responding to the therapy. These are basic questions, considered preliminarily in evaluating efficacy. The data for addressing these questions need to be sensitive to a variety of aspects of client involvement in the treatment process, including session-to-session interest, motivation, success, etc.

Is significant, important change occurring? As the client becomes more involved in the treatment program, the speech-language pathologist must determine whether or not the client is changing. Change needs to be examined along a variety of dimensions, including rate of change, magnitude of change and extent of change. The client's progress needs to be evaluated to determine whether the rate of change in the targeted behaviors is faster than what would have occurred without intervention. Further, the magnitude of change in performance needs to be sufficiently impressive to convince the clinician that treatment is working. And finally, change needs to be evaluated in terms of its importance for the client's overall wellbeing. Judging significant change is a formidable task; readers are urged to see Bain and Dollaghan (1991) for a practical review of this topic. The data needed to evaluate significant change must be broad enough to encompass the various dimensions of progress. This necessitates the use of multiple measures in evaluating change.

Is treatment responsible for the change? Acknowledging that this question has an accountability quality, it should also be viewed as an integral part of the speech-language pathologist's ongoing decision-making. The key issue here is whether any other force could be responsible for the client's change other than the treatment itself. In research, these forces are known as "threats to validity"; they are the factors, such as maturation, history, instrumentation, and so forth, that can influence the direction of the results. These threats absolutely exist in the clinical world and need to be examined as they interface with treatment progress. The speech-language pathologist's task is to determine that the treatment efforts are responsible for the client's change; if not, ethical issues concerning service delivery would certainly be raised. The data needed to address this clinical question must evaluate the specific threats to validity.

How long should a therapy target be treated? Intervention may serve to facilitate or trigger the emergence of a behavior, induce the emergence and mastery of a behavior, or maintain a behavior at its most sophisticated, complex level (Olswang & Bain, 1991). The speech-language pathologist must first determine which of these outcomes

is anticipated, and then determine whether or not the selected target has reached the desired goal. Treatment on any given behavior or sets of behaviors should only be provided as long as necessary, given the expectations and realities about the occurrence of the behavior(s) and the client's characteristics. Determining the goal for a treatment target involves not only knowledge about the client, but also about his/her environment. As with the other clinical questions, determining how long a therapy target should be treated raises complex issues. Campbell and Bain (1991) have attempted to address these issues in their recent publication. The data for determining whether treatment can be terminated must be sensitive to a client's performance, with various amounts and levels of contextual support.

Nature of the Data

The variety in clinical questions demands different kinds of data for formulating answers. Data can be either quantitative or qualitative; the value of each needs to be recognized and weighed when designing data collection procedures. Quantitative data refers to objective data, where behaviors can be operationally defined for observation and measurement. The critical feature of quantitative data is that a behavior can be described so precisely that at least two independent observers could observe and count occurrences of the behavior, for example, production of plural /s/ in obligatory contexts. Qualitative data, in contrast, refers to subjective data. These are data from observation, interviews, and other sources, as diaries, official documents, photographs, etc. Qualitative data bear the mark of the data collector, in that they reflect what the collector sees, that is, behaviors in context, such as a teachers' attitude about appropriate turn-taking behavior. Collecting quantitative data requires that a taxonomy of behaviors be defined before data collection; the examiner knows what he/she is looking for. Qualitative data requires an open mind and eye; no a priori decisions are made. Rather, the examiner describes whatever appears pertinent to the question being raised. In such a case, data consist of behavioral description and interpretation. In the discussion that follows, the use of both quantitative and qualitative data will be considered. In addition to this classification schema, clinical data will be further categorized into three general types: treatment data, generalization probe data, and control data. The identification of these three data types is necessary for us to address the different clinical questions.

Treatment data. Treatment data are those data gathered while treatment is being conducted. They describe the client's performance during the teaching paradigm; that is, what the client does during treatment, in response to the speech-language pathologist's antecedents and consequences. Traditionally, these have been quantitative data, the pluses and minuses that clinicians often tally as they are providing their treatments. An example of quantitative treatment data would be the number of times a child correctly produces a target phoneme in syllables when provided with a direct model by the clinician. A qualitative

data example would be the clinician's observation of a child's interest and motivation as documented by notes concerning the child's on-task behaviors. These data are useful for determining if the client is responding to the treatment program; they will reveal whether the client is attending and participating. These data are limited to describing learning in the teaching context; they are not useful for providing information as to whether learning is generalizing. Treatment data reflect performance colored by the clinician's most accommodating prompts and cues for teaching purposes. As such, treatment data reflect a restricted view of learning, yet an important one.

Generalization probe data. In contrast to treatment data, probe data are not collected during treatment. Rather, these data are gathered "outside" of the teaching paradigm. Their purpose is to describe the client's learning beyond the treatment context; that is, the client's generalized performance. Using the behavioral terms of "stimulus" and "response" generalization, the probe data are designed to sample the client's ability to generalize learning to new stimuli and responses. These data typically examine the client's performance on trained items presented with new materials, new people, or new settings (stimulus generalization), and the client's performance on untrained items (response generalization). An example of quantitative generalization probe data would be the number of times a child correctly produces a target phoneme in conversation with the clinician when playing with new materials (that is, those not previously used in treatment). A qualitative data example would be the parent's perception of a child's intelligibility change as rated over the course of treatment. When considering whether the client is making significant change in treatment, the speech-language pathologist must gather data to reflect the multidimensional aspects of learning. Therefore, generalization probe data always consist of measures of a variety of behaviors related to the goals of treatment.

Control data. Unlike treatment data and probe data, control data are not measures of behaviors that are expected to change as a result of treatment. Rather, these data reflect behaviors that could change as a result of other "cosmic occurrences," but their change would not be considered directly tied to treatment effects. In selecting behaviors that will be used for control data, the speech-language pathologist must consider what behaviors would be considered unrelated in terms of response classes, but related in terms of developmental expectations. These are behaviors that could change because of maturational influences, but would not be expected to change because of the treatment influences. An example of quantitative control data would be the measurement of change in liquids /r, l/ when sibilants /s, z, ʃ, ʒ/ are targets for treatment.

Principles of Data Selection and Collection

The essence of using data collection for making informed clinical decisions is that the speech-language pathologist should be able to observe behaviors changing over time in conjunction with the implementation of

treatment. As treatment is implemented, behaviors expected to change (those measured by the treatment and generalization probe data) should, in fact, change correspondingly, and those not expected to change (control data), should not. For the data to be useful for interpretation in this way, they must be appropriate for frequent collection over time. In this regard, the data and collection procedures must adhere to three basic principles: the data must be valid, they must be reliable, and they must be collected in a reasonable fashion.

Validity. Validity refers to the truthfulness of the data; that is, do the data accurately measure the phenomenon of interest. To ensure valid data, the speech-language pathologist must observe behaviors that will representatively reflect change. Measuring change, or learning due to treatment, is difficult in part because of the complex nature of communication. Communication comprises several domains (syntax, pragmatics, semantics, and phonology) and includes both comprehension and production processes. As such, valid data must be conceived of broadly. A child's treatment, and thus learning, may be restricted to one domain, but so too might learning be broader, crossing domains. We need to think about what we are teaching, what we hope will change, but also how other aspects of the child's development could be influenced by the treatment. A holistic view of children is mandatory as we consider our data, specifically our generalization probe data. So too must we have a broad perspective as we attempt to monitor change in the child due to forces other than the treatment. Valid control data will reflect changes in development (within or across domains) due to normal maturation, versus changes due to the benefits of treatment. Both qualitative and quantitative data must be valid. To ensure validity, we must have adequate amounts of data, adequate variety of data, and adequate confirming and disconfirming evidence to demonstrate plausibility. The bottom line is that the data and data analysis must reflect what we know and believe to be true about the developmental and clinical processes. Ultimately, the confidence we have in our clinical decisions directly reflects our confidence in the truthfulness of our data.

Reliability. Reliability refers to the trustworthiness of our data. This means that the data must be amenable to collection over time without concern for variability in performance other than what is "true" for the client, versus being in the "mind of the beholder." The speech-language pathologist must be able to trust the data to be credible over time, truly fluctuating as a result of the client's changing abilities. For quantitative data, reliability is ensured in part by having independent observers sample the collection of the same data. For qualitative data, credibility is ensured by having different sources of data yield the same conclusions (Lincoln & Guba, 1985). The bottom line for reliability and credibility is that the data do not merely reflect what is in the clinician's mind, but rather that others can trust in the way the data have been collected and analyzed.

Collection. Finally, having identified valid and reliable data, the speech-language pathologist must be comfortable in the reasonableness of the collection task. The realities of

a clinical practice in any setting place data collection in real jeopardy, no matter what the benefits, unless the task is feasible. To ensure this, several guidelines are suggested. First, only the most relevant data should be collected. We will urge for the evaluation of multiple measures, and the use of quantitative and qualitative data; we will also encourage speech-language pathologists to collect only the most necessary data. Treatment data, generalization probe data, and control data provide different kinds of information; the selection of the kinds of data to be collected should be made wisely. Second, not all kinds of data need to be collected as frequently. Consider the rate of change expected in different behaviors, in different domains, and make measurements at appropriate intervals given these expectations. Third, data may be collected in a variety of ways, some of which are more convenient than others. Consider different ways in which to collect data; this will include variations in setting, context, people, time, etc. For example, perhaps a brief, 5-minute sample of conversational speech in the waiting room or at home will provide sufficient data for making a judgment about a client's progress on generalization and control measures. Or perhaps an interview done monthly will yield the necessary evidence for documenting change as viewed from the parent's perspective. The sampling condition needs to maintain the validity and reliability of the measure, but allowing for those constraints, the possibilities for actual collection are endless.

Implementing Data Collection

Actually implementing data collection involves three preliminary steps: deciding what to measure, deciding how to measure, and deciding when (or how frequently) to measure.

What to Measure

Deciding what to measure requires that we revisit our discussion of quantitative versus qualitative data, and consider how these different kinds of data can appropriately contribute to clinical decision-making.

Quantitative data. Quantitative data, or objective behavioral measurement, are overt, countable data. Collecting quantitative data requires that the clinician have clear ideas about which behaviors are important to observe. These include behaviors that have been selected as treatment targets, behaviors that are related to the targets and thus expected to change as a result of treatment (generalization behaviors), and behaviors that are unrelated to the targets, and thus not expected to change with treatment (control behaviors). Quantitative data have been used in our profession for many years as a means for documenting change. Their origin, in behavior modification and quantitative research, is bound in a perspective of objectively gathering facts to prove or disprove a hypothesis. Behaviors are described sufficiently such that they can be counted, yielding frequency and duration measures. Most clinicians have been reared on the use of quantitative data, but unfortunately, quantitative data may be viewed as a burden for clinical application. We urge clinicians to

reconsider their knowledge in the use of quantitative data; without a doubt, objective data can be informative for making clinical decisions.

When using quantitative data, deciding what to measure includes determining the treatment targets and the generalization and control behaviors, and operationally defining these behaviors. The generation of the operational definitions requires an understanding of the characteristics of those behaviors: are they definable as discrete events (e.g., correct /s/) or are they behaviors that are sustained over time (e.g., joint attention); are they behaviors that can be judged correct or incorrect, occurring or not occurring (e.g., plural allomorphs /s, z, əz/ in obligatory contexts), and/or appropriate or inappropriate (e.g., requesting or commenting)? These characteristics need to be defined so that the clinician can decide whether frequency of occurrence measures, latency or duration measures, response rate measures, or percent accuracy measures might be most useful in monitoring progress. (The reader is referred to Appendix A for a brief summary of such measures and a list of relevant readings regarding use of quantitative data.) In deciding what to measure, countable behaviors will be important to include, but they are not the only source of data.

Qualitative data. Qualitative data, or subjective measurement, reflect interpretation. They form a description of the client in context, and as such reflect a socially constructed reality, a view of the client and treatment process from an insider's point of view (i.e., the insider being the clinician, the client, the parents, the teachers, etc.). Gathering qualitative data acknowledges that the variables surrounding target behaviors are complex, interwoven and difficult to measure, and thus quantitative, objective data, alone are inappropriate or insufficient. The complexity of communication is not only acknowledged, but appreciated. As such, data collection involves immersing oneself in the setting and lives of the client and his/her significant others, and using multiple means to gather data. Participant observation and fieldnotes, interviews, and diaries all become tools for data collection. The clinician's task is to systematically identify behaviors that appear important, and to determine the best way to appraise these behaviors. The behaviors may be client behaviors (e.g., communication attempts during group activities), or the behaviors and thoughts of relevant others (e.g., sibling's view of communication efforts). Behaviors deemed critical are decided by what the clinician hears, sees, and interprets as important in the context of interpersonal communication. These may be a client's newly acquired linguistic structure, or the client's attitude about talking, or even the client's parent's perception of change. Correspondingly, data collection tools can be observation of behaviors in different settings, questionnaires about attitudes, or interviews about development.

Qualitative data can contribute important information to evaluating client progress and the success of the clinical process. They can be useful as part of treatment data, generalization probe data, and control data. As a profession, we have often used qualitative data, via parent and teacher report and informal observation, to support our intuitions about client progress. However, when document-

ing clinical effectiveness, or when being urged to be accountable, quantitative data is often regarded as preferable. We urge clinicians not to limit themselves in the kinds of data they use, but rather to employ different kinds of data, appreciating how each can contribute to the ultimate goal of documenting communication change.

Qualitative data, collected and analyzed systematically, can be extremely valuable. Observational data reveals ways in which the client interacts with the environment, allowing for the determination of how specific contextual variables influence performance. Often clinicians feel at odds with quantitative data because they are forced to isolate behaviors and define them operationally for observation. They may see particular communication behaviors as inseparable from the environment. This perspective leads appropriately to the collection of qualitative data using fieldnotes, interviews, questionnaires, and other personal and public documents. Observational fieldnotes, interviews, and questionnaires will probably be the primary tools used to gather qualitative data. Fieldnotes are the clinician's attempt to create a written account of what he/she hears, sees, experiences, and thinks in the course of observing the client in a particular context. Interviews consist of a dialogue between the clinician and the client, or significant others. Interviews are often recorded and later transcribed for data analysis; they also can yield observational data as the clinician reflects and records fieldnotes of the interview experience post hoc. Finally, questionnaires are a means of gathering subjective data about opinions and attitudes. They can be totally qualitative, or be numeric ratings that yield quantitative data. Describing all forms of qualitative data is beyond the scope of this paper; instead, the reader is referred to some key terms and references provided in Appendix B. We will however, illustrate the value of this kind of data collection as we return to our discussion of clinical decisions.

How to Measure

Planning for the implementation of data collection also requires that clinicians decide how to collect data. Data collection strategies can be viewed on a continuum from naturalistic to highly structured.

Naturalistic data collection. At the naturalistic end of the continuum, the clinician imposes no structure while observing the child. The child is in a familiar environment and directs the nature of the interaction and activity. The clinician acts as a passive observer, or participant observer. Data collected in a naturalistic manner are considered to be representative of a child's typical status. Both quantitative and qualitative data are appropriately gathered in naturalistic contexts. By and large, this is the ideal environment for collecting qualitative data, as the essence of observational data is that it reflects the client's performance within the social context. A limitation of naturalistic collection of quantitative data is that the opportunity for the behavior of interest may not occur or may occur only infrequently during the observation period. Thus, truly naturalistic data collection may be problematic for generating sufficient numbers of exemplars of the target behavior for evaluating competency.

An example of naturalistic data collection might be observing a child in the home. Assume the child's target is requesting and the clinician decides to collect generalization data while observing the child in play with a parent. For qualitative data, the clinician might wish to observe the parent and child at play, allowing the parent to determine the activities. Fieldnotes would reveal types of activities selected, which ones were most engaging for the child and adult, and which ones allowed for the most requesting. These data would contribute to the clinician's knowledge of whether requesting was generalizing to the home setting, and provide insight to the nature of the generalization, including those aspects of the context that were interconnected with the requesting behavior. For quantitative data, the clinician might be more focused on the child and his/her actual request productions. The data would include counting the frequency of occurrence of child requests. Such a focus requires that the clinician ensures the observation is of an activity that inherently allows requesting. The quantitative data are being collected to document competency in requesting. Competency can only be measured by having sufficient exemplars to prove productivity.

Thus, when collecting data in a naturalistic setting, the clinician must be mindful of the purpose. Is it to observe and sample a communication behavior as it naturally occurs, thus allowing, and even desiring, all kinds of variation in occurrence? Or is it to observe and sample competency as measured by performance of the target behavior in a naturalistic context that ensures sufficient opportunities for measurement? The former encourages, actually requires, unobtrusive sampling of behaviors as they occur. The emphasis is on observing the target behavior in social context, where the variables defining the social context are of as much interest as the target behavior itself. The latter is examining and judging competency in using the target behavior. To judge competency, sufficient opportunities must be available from which to sample occurrence of the target behavior. In both cases, the naturalistic context in general is appropriate, but the nature of the information being sought is different and the kinds of data being collected are different.

Structured data collection. At the structured end of the continuum, the clinician actively manipulates the activity and allows for multiple opportunities for the behavior of interest to occur in a time-efficient manner. Elicited probes are a form of structured data collection. For example, assume a child's target goal is correct production of fricatives in conversational speech. The clinician may have selected 10 initial /f/ words and 10 initial /s/ words to use during the treatment program. In order to determine if the child was beginning to generalize and produce fricatives, the clinician identifies an additional set of 5 initial /f/ and 5 initial /s/ words that were not treated. Once a week, the clinician asks the child to name the pictures of the untreated /f/ and /s/ words. In this example, the time necessary to obtain the data is short (probably less than 2 minutes) and the clinician knows that several opportunities have occurred for the behavior of interest. However, because the context is so structured, it may closely re-

semble the treatment situation, thus allowing for only a limited view of generalization. For example, the child may produce acceptable fricatives in the untrained words and yet not be producing untrained fricative words in conversation, as would be revealed by a more naturalistic sampling.

Qualitative data may also be collected in a structured manner. For example, a critical observer might be asked periodically to complete a questionnaire regarding a child's communication performance. Such a questionnaire might be broad in its view of communication, asking the observer to make general judgments about ability. Or the questionnaire could be more narrow, and specific, asking the observer (e.g., clinician, parent, teacher) to rate a child's correct production of a target phoneme. In this case, the observer would be making general judgments about phoneme use: for example, never correctly uses, sometimes correctly uses, half of the time uses, almost always uses, and always uses. These are qualitative, subjective data, but they are collected in a structured manner.

The advantage of structured probes is that they can be collected in a time-efficient manner. Further, they can be structured in such a way as to reflect a variety of performances from the child. If the purpose of your data collection system is to reflect a child's typical performance, the structure needs to be implemented in a way that does not call attention to the production of the target. For example, a clinician can obtain measurements using untrained elicited probes at the beginning of a session, prior to "conducting therapy," by asking the child to look at and name some pictures and objects as the clinician and child interact in a play situation. Thus the data are collected in a structured and time-efficient manner, yet in a way that allows the data to reflect the child's typical performance. Contrast this scenario with one where the clinician asks the child to produce the untrained exemplars but does so while in the same setting and with the same degree of attention that is employed during the actual treatment process. These data will reflect the child's performance that might be viewed as "optimal" rather than "typical," and may thus overestimate the child's capability.

Thus, structured data collection can vary in type of data collected, and in the amount of structure provided. The clinician must be clear about the nature of the data desired and the way in which the data will be used. Again, multiple measures may be indicated, and include quantitative and qualitative, naturalistic and structured collections. The clinician has many options as to what types of data to collect, where, and by whom. The value of multiple measures is that a fuller picture of communication change can be obtained. The clinician's task is to decide what the data are to demonstrate and what is practical for data collection.

When to Measure

When to measure involves determining the frequency with which specific measurements will be obtained. The critical element to consider in deciding when to measure, or how frequently to measure, is whether the data are being collected often enough to judge ongoing change. The

problem, of course, is practicality. Clinicians want to be time efficient and not collect data if useful information will not be obtained. Multiple measures, and the manner in which data are collected, might be considered when determining how frequently to obtain measurements. For example, structured probes are time efficient and do not involve lengthy analysis. Accordingly, they might be collected fairly frequently—weekly or biweekly. Structured probes, however, are limited in scope, so the clinician might want to obtain an additional naturalistic measure of a child's performance, albeit less frequently, to verify or expand the information gained from the structured probes. A suggested guideline is that when a child begins to perform the behavior of interest on structured probes, the clinician obtain a more naturalistic measure in order to confirm the consistency of the behavior. Transcribing and analyzing language samples and fieldnotes can be a time-consuming procedure, and the clinician would not want to do this on a weekly basis. By combining structured probes with naturalistic sampling, quantitative data with qualitative data, the clinician can devise a valid yet efficient data collection system.

Clinical Decisions Revisited

We return now to our clinical questions to illustrate how the data collection procedures might be implemented for making informed decisions and evaluating the treatment process.

Is the child responding to the treatment program? One common type of treatment approach involves the clinician planning and implementing some sort of hierarchical remediation program that is designed to change specific communication behaviors. The remediation program often involves the clinician in providing salient and relevant antecedents, consisting of cues and prompts, to elicit behaviors that can frequently and appropriately be reinforced in the beginning stages of treatment. Then as the child acquires the target behavior, the antecedents and consequences become less intense and focused. For this type of therapeutic approach, treatment data are useful in aiding the clinician to decide if the child is responding to the treatment program. Treatment data can consist of treatment probe data and the child's actual responses to the ongoing treatment program.

First, the clinician employs the treatment probe data to identify the step in the program that is appropriate for a child during a specific treatment session. The treatment program devised by the clinician is an attempt to identify and plan relevant steps that may be needed by the child. We have all experienced the child who needed additional steps from what we planned (branch steps) or the child who did not need to experience each and every step originally identified by the clinician. Treatment probe data allow the clinician to modify the program and implement the program based on the child's needs for each session. To obtain treatment probe data, the clinician asks the child to attempt steps of the program not previously experienced by the child. For example, if a child's target is production of final consonants, the treatment program may consist of a

hierarchy of antecedent steps proceeding from a direct model, to an indirect model, to an elicitation question. Assume that the child during the first session was able to produce targeted final consonants in words when provided with a direct model approximately 50% of the time. At the beginning of the second session, the clinician would probe and see if the child could produce targeted final consonants with an indirect model. By probing steps of a program not previously experienced, the clinician can determine the most advanced level at which a child is able to function, and modify or adjust the treatment program based on the ability and need of a child during a specific session. By not employing probes, the clinician assumes that each and every step planned is needed by that child. Thus, treatment probe data allow the clinician to determine the specific steps needed based on child performance, and to omit those steps not needed.

Treatment data also refer to the child's actual responses to the treatment program. The clinician usually employs either frequency of event recording or calculating a percentage of correct responses. Although some clinicians may record every response of the child's, some of us find that this data recording process interferes with the clinician-child interaction and places more emphasis on the data collection than on the intervention process to enhance communication. We suggest that clinicians employ a sampling process and record data on a particular step in treatment program for five or ten responses or in two 5-minute time periods. The results can then be used to infer how the child responded to the session in general. The sampling procedure allows the clinician to document that the child was successful at particular steps and yet the emphasis of the intervention can be on the communication between the child and the clinician.

Qualitative treatment data can also be quite informative for determining a child's session-to-session interest and motivation, along with performance success. Observation and documentation of highly enthusiastic participation, along with a description of the context, can be used to speculate about salient, important learning moments, and thus be used for planning future sessions. These observations, in conjunction with interviews, can be quite revealing, and as they accompany the quantitative data, can begin to create a package of valuable evidence for treatment benefits.

Is significant, important change occurring? A clinician usually employs a treatment program as a means to change a child's communication in a general sense. The end goal of treatment is to help a child be more proficient in communicating with others in everyday activities. To assess "Is significant, important change occurring?" a clinician must collect generalization data. Responsiveness to the treatment program is judged as effective only if stimulus and response generalization occur.

Stimulus generalization data involve assessing whether the child is using the treatment exemplars in other untrained situations, with other persons not involved with treatment or with materials not employed during treatment. For some treatment goals, stimulus generalization is the most that can be expected. For example, vocabulary and

certain pragmatic goals may be restricted to stimulus generalization; that is, teaching a list of vocabulary words in one setting, hoping for generalization to another. Similarly, for some children with certain characteristics, such as a child with severe mental retardation, stimulus generalization may reflect optimal performance (e.g., a child learning a requesting routine with the clinician and parent, and generalizing to the babysitter). The clinician can assess stimulus generalization by identifying materials, settings, or persons not involved in the treatment process and then periodically assessing if the child has generalized the use of the treatment target. For time efficiency, elicited probes are often employed to assess stimulus generalization. A minimum of three data points should be collected before the initiation of treatment to determine if the baseline performance is stable. After treatment is initiated, stimulus generalization should be collected periodically depending on how the child is performing in treatment. The clinician will most likely set a minimum time period, for example, every 3 weeks or until highly consistent levels of performance are reached in the treatment program. As the child nears the end of the treatment program, the frequency with which stimulus generalization data are collected should increase. In order to verify the clinician's data, information from other persons should also be obtained. The clinician can ask others (parents or teachers) informally how the child is doing, or the clinician might devise a rating scale and obtain more specific information regarding the child's performance in other settings or with other persons. These more naturalistic measures can then be used to confirm and expand the results of the elicited probe data.

In most clinical situations, the clinician hopes that a child will learn a linguistic or communication rule. In these situations, significant and important changes are assessed through response generalization. For many children, response generalization should receive the emphasis in data collection rather than stimulus generalization. Many of us are of course pleased when a child begins to use the treatment items in other settings, but a more sensitive indication of "rule learning" is when the child begins to generate new untrained utterances or behaviors that are similar to, yet different from, those emphasized during treatment. Response generalization is identified when a child uses responses similar to, yet different from, the ones employed during training. For example, if a child's treatment target is production of final consonants in conversational speech, the clinician will likely focus treatment on a few consonants, hoping that other words and untreated final consonants will also be produced. When this happens, the clinician assumes the child has learned the rule of production of final consonants.

Clinicians are urged to think broadly as they are attempting to document a client's generalization. The clinician's task is to consider the phenomenon of interest, and to decide how the child might demonstrate competence. Another example considers a clinician wishing to determine whether facilitated communication has been a viable means for aiding an autistic child to express him/herself. The clinician needs to look for response and

stimulus generalization evidence. This could involve identifying examples of communication behavior using facilitated communication that have not been observed before. Further, if these new behaviors are observed with different facilitators and different communication partners, the data become even more supportive. And if the child seeks to use facilitated communication in other than planned events, the evidence mounts even further. Multiple exemplars and multiple sources of data contribute to documenting significant change, and all avenues for data need to be explored.

Is treatment responsible for the change? When a child is enrolled in treatment, demonstration of significant change in communication is in itself insufficient for documenting efficacy. Children tend to change and improve as a function of maturation, preschool and school experiences, as well as other factors. In today's world of accountability and limited financial resources, clinicians have an obligation and responsibility to demonstrate that the change was related to the treatment delivered. Clinicians of course want natural change to occur (i.e., change due to maturation, school, etc.); however, they have an ethical responsibility to document changes due to their treatment. Data collection to demonstrate that the significant change was related to treatment requires that the clinician monitor a behavior that is not treated, but which is developmentally similar to the target behavior, yet assumedly not related to the treatment target. The clinician then compares the child's performance on this untreated, control behavior to the child's performance on the generalization probes. If the child produces more of the generalization probes than the control probes or produces the generalization probes earlier than the control probes, the change is assumed to relate to the treatment, at least primarily, and not some other factor. The change that occurs due to treatment should be at a faster rate, or a higher, more advanced level than the change that occurred due to other, natural factors.

Selecting control behaviors can sometimes be quite straightforward, and at other times more tricky. For example, if a clinician were treating a child with a phonological disorder for "gliding," "cluster reduction" can be monitored as the control behavior. Or if a clinician were treating a toddler with a language disorder for "verbal requesting," the clinician might monitor vocabulary growth as a control behavior. These choices are relatively straightforward, assuming that they comprise problem areas for the children, and that the target and control behaviors are developmentally equivalent. Consider, however, a treatment target such as requests for information, or polite forms, finding compatible control behaviors might not be so easy. The clinician must look first at the child's communication/linguistic repertoire and identify missing, deviant, or otherwise problematic behaviors. These become the pool from which to select the control. We acknowledge that this is not an easy task, but the effort is well worth the trouble; use of control data allow for increased confidence in judging treatment effectiveness.

How long should a therapy target be treated? Typically, the ultimate goal of treatment is generalization of the

treatment target. However, a critical question is how long do we treat a particular target and how much evidence of generalization is needed before moving on to a new target(s). Generalization data are suggested as the basis for determining how long to treat a specific target. Treatment data are of little value in knowing if the child is beginning to generalize. Many of us have encountered the child who functioned at the most advanced level of a treatment program, required few clinician cues, yet failed to demonstrate generalization.

The goal in determining how long to treat a specific target requires frequent monitoring so that as soon as generalization is demonstrated, treatment on a particular target can be withdrawn. Concurrently, the child's performance continues to be monitored to determine if indeed the treatment was sufficient. Elicited generalization probes are efficient for this purpose because administration time is brief. The clinician administers the probes routinely as described in the section discussing significant change. Similarly, the clinician will also want to obtain a more naturalistic measure of generalization to confirm that the child's performance was representative of the child's usual communicative functioning.

Olswang & Bain (1985) investigated withdrawing treatment on specific sounds with four preschool children with phonological disorders. They established three criteria at which treatment on a specific target was withdrawn. The criteria were based on quantitative data using elicited response generalization probes collected weekly (untrained exemplars of the target sound). The criteria included the following: 100% correct performance during one session, 2 sessions between 75% and 99% correct, or 3 sessions between 30% and 74%. The results from this study indicated that for three of the four subjects, treatment on the target sounds could be terminated when productions on the generalization probes met any one of the three criterion levels. Although treatment was withdrawn, the children continued to make progress on the target sounds. These results were replicated more recently by Hallaway & Bain (1992) on four preschool children with a language disorder who were being treated for syntactic targets. Treatment on specific targets was terminated when the performance on the structured generalization probes reached one of the three criterion levels described above. All four subjects continued to generalize without additional treatment on the specific targets. Contrast these results to those reported by Diedrich and Bankert (1980) and Eger (1988). These investigators found that clinicians continued treatment on a specific target until the children performed at a correct level of 90% and 100% in 3 minutes of conversation for a period of 3 months. Using generalization probes as a means to decide when treatment on specific targets can be terminated allows clinicians to measure and monitor progress in a systematic manner and to be more efficient in service delivery.

Monitoring a child's progress in treatment: An example. Table 1 provides a detailed example of how quantitative and qualitative data might be used to make informed clinical decisions. This example attempts to illustrate the features of the model we have been discuss-

TABLE 1. Data collection and clinical decision-making: An example for the production of final consonants with /m, p, f, d/ as target sounds.

Clinical Decisions	What to Measure?	How to Measure?	When to Measure?	Results and Implications	
1) Is the child responding to the treatment program?	<ul style="list-style-type: none"> • Determining what cue and response levels are to be used for a particular day. 	<ul style="list-style-type: none"> • Treatment probes—Quantitative—Productions in response to direct and indirect models of treatment words. Percent correct at each cue level. 	Structured	At beginning of each session	<ul style="list-style-type: none"> • Child produces no final consonants in 4 words with an indirect model and produces 2 out of 4 final consonants with a direct model. Treatment session starts with clinician using direct model.
	<ul style="list-style-type: none"> • Determining whether the child meets criterion to advance in program. 	<ul style="list-style-type: none"> • Treatment response—Quantitative—Response to 5 treatment words for each final sound /m, p, f, d/ with a direct model. Percent correct obtained. 	Structured	During each session	<ul style="list-style-type: none"> • Child produces 80% of the /m, p, d/ treatment words with a direct model and 60% of the /f/ words. Clinician moves to indirect model for /m, p, d/ words and continues direct cue level for /f/ words.
		<ul style="list-style-type: none"> • Treatment response—Qualitative—Observation of child's attention to tasks as a reflection of motivation and effort. 	Somewhat structured	Each session	<ul style="list-style-type: none"> • Child's lack of attention indicates activities need to be changed.
		<ul style="list-style-type: none"> • Treatment response—Qualitative—Observation of child's effort in production of final sounds in target words. 	Somewhat structured	Each session	<ul style="list-style-type: none"> • Child expends considerable effort producing final consonant in target words, which provides information as to whether more "practice" may be needed, even if criterion of 80% correct is obtained at a given cue level.
	<ul style="list-style-type: none"> • Treatment response—Qualitative—Parent observation of child's willingness to come to treatment. 	Unstructured—Interview, questionnaire, or log	Weekly	<ul style="list-style-type: none"> • Child shows eagerness about coming to treatment. 	
2) Is significant and important change occurring?	<ul style="list-style-type: none"> • Examining performance on stimulus and response generalization probe measures. 	<ul style="list-style-type: none"> • Stimulus generalization data—Qualitative—Observation of child using treatment target words in conversation with parent. 	Somewhat naturalistic	After child has progressed to advanced levels in treatment program.	<ul style="list-style-type: none"> • Child begins to use treatment targets elsewhere, which indicates response generalization probes should be administered.
		<ul style="list-style-type: none"> • Response generalization probes—Quantitative—Production of 5 untrained words for each target sound /m, p, f, d/. Percent correct for each sound obtained. 	Structured—Naming pictures	Three sessions before beginning of treatment, then minimum of every 4 weeks or when stimulus generalization is observed or reported.	<ul style="list-style-type: none"> • Child produces /m, p, d/ in untrained exemplars at the 60% level, and /f/ at 40% level. Clinician obtains a brief conversational sample to confirm production observed in generalization probes.
		<ul style="list-style-type: none"> • Response generalization data—Qualitative—Parent observation of child's use of target sounds in new words and with other people. 	Naturalistic—Interview, questionnaire or log	Monthly	<ul style="list-style-type: none"> • Parent reports child uses final target sounds in variety of words with different people.
	<ul style="list-style-type: none"> • Response Class Generalization probes—Quantitative—Production of words containing final consonants related to but different from target sounds /n, b, s, t, k/ (5 words for each sound). Percent correct for each sound obtained. 	Structured—Naming pictures	Three sessions before beginning of treatment, then after child has produced some of the untrained treatment exemplars.	<ul style="list-style-type: none"> • Child correctly produces 40% of the /n, b, t, k/ words and 20% of the final /s/ words. Child begins to mark final consonants. Clinician obtains a brief conversational sample to verify production in a more natural situation. 	

TABLE 1 (cont'd). Data collection and clinical decision-making: An example for the production of final consonants with /m, p, f, d/ as target sounds.

Clinical Decisions	What to Measure?	How to Measure?	When to Measure?	Results and Implications
	<ul style="list-style-type: none"> Response Class Generalization—Parent observation of child's use of new sounds at the end of words in conversation. 	Naturalistic—Interview, questionnaire, or log	Monthly	<ul style="list-style-type: none"> Parent reports most words have a final sound even if not produced correctly, and that overall intelligibility is improving.
3) Is treatment responsible for the change?	<ul style="list-style-type: none"> Control Probes—Quantitative—Child velar assimilates in the initial position; this serves as the control behavior. Compare percent correct on untrained exemplars /m, p, f, d/ to percent correct production of initial velars in words ending with velars (5 each of k and g). 	Structured—Naming pictures	At same time and with the same frequency as response generalization probes.	<ul style="list-style-type: none"> Child correctly produces final consonants at a higher rate (60% for /m, p, d/ and 40% for /f/ than for the initial velars (10% /k/ and 0% /g/). Treatment appears to be primarily responsible for production of final consonants rather than other factors, such as maturation.
4) How long should a therapy target be treated?	<ul style="list-style-type: none"> Quantitative—Production of words in response generalization probes (i.e., untrained exemplars of /m, p, d, f/. Percent correct obtained. Quantitative—Response generalization on untrained exemplars /m, p, d, f/ in conversational speech. Qualitative—Parent observation of child's use of sounds at the end of words at home and in other settings. 	<ul style="list-style-type: none"> Structured—Naming pictures Somewhat naturalistic—materials selected to ensure sampling target sounds Naturalistic—Interview, questionnaire, or log 	<ul style="list-style-type: none"> As discussed in decision two. Weekly Monthly 	<ul style="list-style-type: none"> Child correctly produces more than 40% of probe items. Treatment is discontinued on this target. Child's productions of final consonants are monitored to ensure continued progress. If decrease is noted, treatment would be reinstated on final consonants. Parent reports child is correctly producing final consonants in the home and other settings.

ing. For each of the four clinical decisions, we have indicated what to measure, how and when. Under Results and Implications, *we have hypothesized* how a child might perform and what the performance would mean in regard to the clinical decision. This example is meant to bring together the pieces, and offer a gestalt of a data-based approach to clinical decision-making.

Conclusions

Data provide essential information for making well-informed treatment decisions. Adding data collection to the speech-language pathologist's duties need not be a burden. Indeed, data collection can become an integral part of the treatment process. The critical features of designing a viable data collection plan include selecting the right measures, and determining how and when to best make the measurements. Multiple measures are available for answering a variety of clinical questions. Multiple sources of data are available for appropriately answering questions and for increasing the validity of the findings. And finally, multiple schedules of collection are available and appropriate for easing the onus of the task. Consider the behavior

change being documented, consider how to best reflect that change and finally, consider what is reasonable for implementation. The effort placed in planning immediately pays off as clinicians find themselves being able to rely on their data for making well-informed, well-supported treatment decisions.

Acknowledgment

This article was supported in part by the NIDCD Grant R29-DC00431, "Predicting the Benefits of Treatment."

References

- Bain, B., & Dollaghan, C. (1991). Treatment efficacy: The notion of clinically significant change. *Language, Speech, and Hearing Services in Schools, 22*, 264–270.
- Campbell, T., & Bain, B. (1991). Treatment efficacy: How long to treat: A multiple outcome approach. *Language, Speech, and Hearing Services in Schools, 22*, 271–276.
- Diedrich, W., & Bangert, J. (1980). *Articulation learning*. Houston, Texas: College-Hill Press, Inc.
- Eger, D. (1988). Accountability in action: Entry, measurement, exit. *Seminars in Speech and Language, 9*, 299–319.
- Holloway, A., & Bain, B. (November, 1992). *Treatment efficacy:*

Determining when to withdraw treatment. Annual Convention of the American Speech-Language-Hearing Association, San Antonio, Texas.

- Lincoln, Y., & Guba, E.** (1985). *Naturalistic inquiry*. Beverly Hills, CA: Sage Publications.
- Olswang, L., & Bain, B.** (1991). Intervention issues for toddlers with specific language impairments. *Topics in Language Disorders, 11*, 69–86.
- Olswang, L., & Bain, B.** (1985). Monitoring phoneme acquisition for making treatment withdrawal decisions. *Applied Psycholinguistics, 6*, 17–37.

Received June 15, 1993
Accepted March 16, 1994

Contact author: Lesley B. Olswang, PhD, Department of Speech & Hearing, Sciences JG-15, University of Washington, Seattle, WA 98195

Key Words: measurements, intervention, speech and language, treatment efficacy

Appendix A

Quantitative Data

Quantitative data can yield event, duration, or interval recording measures.

Event recording. Event measures reflect “something countable” about discrete behaviors. One can count the frequency of occurrence of particular behaviors, or the frequency of occurrence of the behaviors in a particular, predetermined, constant time frame. Some behaviors can be judged as correct or incorrect, appropriate or inappropriate, or performed or not performed, and thus lend themselves to percent correct measurements. In any case, the clinician can set up situations to sample behaviors, and measure the event according to frequency of occurrence, response rate, or percent correct/appropriate/performed.

Duration recording. Quantitative data may also lend themselves to duration recording. In this measurement type, the clinician attends to how long a behavior lasts (duration) or the latency of time lapsing before the initiation of a particular behavior (latency). Many behaviors lend themselves to duration and latency measurement. This is usually true for behaviors that are not exactly discrete, or for which the time dimension is critical to examine. For example, attending is such a behavior. The clinician may be interested in increasing a child’s attending behavior, not in its discrete occurrence, but rather the length of time the child can maintain focused attention. Similarly, the clinician might be interested in increasing a child’s ability to immediately begin attending following an instruction. In this latter case, a latency measure would be useful, where the amount of time between clinician instruction and child’s initiation of attending is measured.

Interval Recording. This measurement employs sampling of behaviors in prespecified time frames. Interval recording can be either *whole interval*, *partial interval*, or *time sampled*. In whole interval recording, a particular time frame is divided into equal portions, for example, a 30-minute class is divided into 10 3-minute segments. During each 3-minute segment, the clinician judges whether the behavior of interest occurred *throughout* the segment. In partial interval recording, the setup is exactly the same, but now the clinician judges whether the behavior occurred *at all* during the segment. Whole interval recording is a more conservative

measure, for it requires occurrence of the behavior throughout the time segment. Partial interval recording requires only the occurrence of the behavior at some point during the segment. Lastly, time sampled recording also involves dividing a large block of time into equal intervals and judging the occurrence of previously defined behaviors. However, in time sampling, the judgment is made *at the onset* of the time segment; for example, every 3 minutes, the clinician would judge whether the target behavior was occurring or not occurring at that particular moment in time. These interval recording methods allow for breaking down the observation task. In event and duration recording, considerable expenditure of time and effort is required, because all occurrences of behaviors are counted. Interval recording allows for breaking down the task by sampling occurrence of behaviors, not all occurrences, in prespecified time intervals.

Key References for Quantitative Data

- Hegde, M. N.** (1993). *Treatment procedures in communicative disorders: 2nd edition*. Austin, Texas: PRO-ED.
- Hegde, M. N.** (1994). *Clinical research in communicative disorders: Principles and strategies*. (Second Edition). Boston: College Hill Publication, Little, Brown and Company.
- Hegde, M. N., & Davis, D.** (1991). *Clinical methods and practicum in speech-language pathology*. San Diego: Singular Publishing Group, Inc.
- Leith, W.** (1993). *Clinical methods in communication disorders: 2nd edition*. Austin, Texas: PRO-ED.
- McReynolds, L., & Kearns, K.** (1983). *Single-subject experimental designs in communicative disorders*. Baltimore: University Park Press.
- Mowrer, D.** (1988). *Methods of modifying speech behaviors: Learning theory in speech pathology*. Prospect Heights, Ill.: Waveland Press.

Appendix B

Qualitative Data (based on Bogdan & Biklen, 1992)

Observation–Fieldnotes. Observations may involve the clinician as a participant observer or a passive, nonparticipant observer. As a participant observer, the clinician is part of the communication interaction with the child. The clinician is an important, viable part of the context. As such, the clinician's observations will reflect this involvement. As a passive, nonparticipant observer, the clinician is not part of the interaction with the child, and as such is attempting to observe without intruding. Being nonobtrusive is difficult, and the clinician will always have to acknowledge and evaluate the impact of his/her presence on the child in the observational context.

Fieldnotes have two components, descriptive and reflective. The descriptive component represents the clinician's best efforts to record what is occurring in the observational context. The clinician should be concrete in his/her description, attempting to describe accurately what is occurring. The clinician should describe the child at the moment of the observation, including physical appearance, temperament, attitude, etc. Other individuals present in the context should also be described in as much detail as possible. The physical setting itself should be described or drawn, including where the clinician is in relation to others in the context. Events and activities should be described. And, of course, the behavior of the child and relevant others should be detailed, including transcriptions of dialogue or aspects of dialogue.

The reflective part of fieldnotes include the clinician's personal observations and interpretations of what is occurring. Here, the more subjective side of the data is recorded. The clinician should include ideas, feelings, speculation, hunches, etc. Bogdan and Biklen (1992) have suggested that the reflective part of fieldnotes might include the following: reflections on analysis, reflections on method, reflections on ethical dilemmas and conflicts, reflections on the observer's frame of mind, and points of clarification. This aspect of fieldnotes is extremely important to creating rich observation, and the clinician is urged to be reflective about all aspects of the observation, including him/herself.

Interview. Data can come from a transcript of an interview or notes. Tape recorded and transcribed interviews obviously provide more thorough data. Qualitative interviews are by definition supposed to be open-ended, but certainly structure of varying amounts can be imposed. The interview is regarded as a more personal way of gathering information about attitudes and beliefs; of course, the success of any interview is dependent on the clinician's skills at preplanning and implementation.

Personal Documents. This category refers to any "first-person written narrative that describes an individual's actions, experiences, and beliefs" (Bogdan & Biklen, 1992, p. 132). The personal document allows for a view of the "actor's" perceptions of situations. Diaries are included in this category of qualitative data.

Official Documents. This category refers to data obtained from official documents, including school reports, evaluations, memos, statements of policy and philosophy, etc. They can be materials relevant to a particular individual or group of people. Individual files on children often contain anecdotal comments from teachers that can be informative. As Bogdan and Biklen (1992) point out, "by and large, qualitative researchers approach student records not for what they tell about the child, but rather for what they reveal about the people who keep the records (psychologists, administrators, teachers). In this framework, the information the files contain—the letters, the teacher's comments, the test scores—represent perspectives on the child. They present one side of the picture" (p. 137).

Photographs. Photographs provide a unique form of qualitative data. They can provide information based on objective description or subjective evaluation. Photographs can be taken by others (including the client) or the clinician, and as such will reveal different information. Photographs taken by others reveal how they view a given situation, or what they believe to be important. Photographs taken by the clinician are often used to supplement observation, either for expansion or clarification of a point.

Key References for Qualitative Data

- Bogdan, R., & Biklen, S.** (1992). *Qualitative research for education: An introduction to theory and methods*. Boston: Allyn and Bacon.
- Erickson, F.** (1986). Qualitative methods in research on teaching. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). New York: Macmillan.
- Glesne, C., & Peshkin, A.** (1992). *Becoming qualitative researchers: An introduction*. White Plains, N.Y.: Longman.
- Lund, N., & Duchan, J.** (1993). *Assessing children's language in naturalistic contexts*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
-