The Treatment Utility of Assessment

A Functional Approach to Evaluating Assessment Quality

Steven C. Hayes
Rosemary O. Nelson
Robin B. Jarrett
University of Nevada—Reno
University of North Carolina at Greensboro
University of Texas Health Science Center at Dallas

ABSTRACT: In practical terms, the sine qua non of the modes, methods, devices, strategies, and theories of clinical assessment is their contribution to treatment outcome. The importance of this contribution has often been noted, but under many different labels and rationales. The resultant conceptual confusion has considerably restricted the visibility and frequency of research in this critical area. In this article we propose a name for the impact of assessment on treatment outcome: the "treatment utility of assessment." Some of the questions that can be asked about the treatment utility of assessment are described, and methods appropriate for asking them are examined. Finally, the implications of this kind of utility for other approaches to evaluating assessment quality are analyzed.

Clinical assessment is an important and fertile area of psychology, and yet there is general agreement that it has not been in a state of continuous and healthy growth (e.g., Bersoff, 1973; Korchin & Schulberg, 1981; Rorer & Widiger, 1983). Compared to the early days of clinical psychology, there has been a decline in the emphasis on clinical assessment both in training and in practice (Garfield & Kurtz, 1973, 1976; Levitt, 1973; Shemberg & Keeley, 1970).

One reason may be that clinical assessment has not yet proven its value in fostering favorable treatment outcomes. As clinical psychologists have devoted more and more time to treatment activities, the practical value of assessment has come under closer scrutiny (McReynolds, 1985). Unfortunately, experienced clinicians have not always found assessment data to be useful in treatment (Adams, 1972; Daily, 1953; Meehl, 1960; Moore, Bobbitt, & Wildman, 1968). Even the proponents of clinical assessment admit that "we do not believe that the current [high] status of testing is due to its demonstrated value" (Kaplan, Colarelli, Gross, Leventhal, & Siegel, 1970, p. 15).

The lack of empirical evidence for the practical value of assessment has long been noted. In 1959, Meehl pointed out that, even if an assessment procedure is reliable and valid, a clinician might "be seriously in error if he concluded therefrom that his tests were paying off in practice. On this question there is . . . no published empirical evidence" (p. 117). Twenty-two years later, Korchin and Schulberg (1981) were still worried that "clinical assessment may not provide the kind of information needed by therapists. Objective evidence is slim" (p. 1154). More recently, McReynolds (1985) asked "Are tests helpful to the therapist? Amazingly, there has been little research on this crucial question" (p. 10). The purpose of the present article is to consider how to make better progress in understanding the role of assessment in successful treatment.

Any lack of evidence on the clinical value of assessment is not caused by a lack of appreciation for its ultimate practical purposes. Korchin (1976) defined clinical assessment as "the process by which clinicians gain understanding of the patient necessary for making informed decisions" (p. 124). Thus, the "basic justification for assessment is that it provides information of value to the planning, execution and evaluation of treatment" (Korchin & Schulberg, 1981, p. 1154). Wiggins (1973) said, "Although measurement and prediction may be evaluated by formal psychometric criteria, such as reliability and validity, the outcomes of [assessment] decisions must be evaluated in terms of their utilities for individuals and institutions within our society" (p. 272, emphasis in original). Meehl (1959) phrased it as follows: "In what way and to what extent does this . . . information help us in treating the patient?" (1959, p. 117). He called this question "ultimately the practically significant one by which the contributions of our [assessment] techniques must be judged" (p. 116).

Definition of the Treatment Utility of Assessment

The impact of clinical assessment on treatment outcome has been discussed under a wide variety of labels. Among other terms, it has been viewed as a matter of incremental validity (Mischel, 1968), concurrent validity (Meehl, 1959), construct validity (Edwards, 1970), predictive validity (Lord & Novick, 1968), discriminative efficiency (Wiggins, 1973), and utility (Cronbach & Gleser, 1965). There is considerable confusion about the concepts relevant to the measurement of pragmatic clinical value.

We propose to use the phrase the treatment utility of assessment to refer to the degree to which assessment is shown to contribute to beneficial treatment outcome. ¹

¹ Earlier (Nelson & Hayes, 1979), we had used the term treatment validity. Although utility issues can indeed be considered an aspect of validity, the present term seems more direct.
An assessment device, distinction, or strategy has this kind of utility if it can be shown that treatment outcome is positively influenced by this device, distinction, or strategy. The treatment utility of assessment deserves to be termed a type of utility because it relates closely to the functional thrust of that psychometric term. The need to qualify the word utility with the adjective treatment is justified by two facts. First, utility has been almost exclusively evaluated in terms of personnel decisions (e.g., Wiggins, 1973). The issues and methods involved in demonstrating the impact of assessment on treatment outcomes differ significantly from the methods appropriate to the analysis of personnel decisions. Second, in personnel matters the concept of utility has come to refer primarily to the cost–benefit ratio of assessment strategies. This is why it was originally distinguished from predictive validity (Mischel, 1968). The treatment utility of assessment is not primarily a matter of cost–benefit analysis but of the demonstration of a particular type of benefit.

Barriers to Research on the Treatment Utility of Assessment

Theorists have long believed that research on the treatment utility of assessment should be feasible. "It is well within the capacity of available research methods and clinical facilities to determine what, if any, is the pragmatic advantage of a personality assessment" (Meehl, 1959, p. 125). Why then has there continued to be so little research?

There are several possible reasons. First, because of conceptual confusion about the psychometric concepts relevant to the treatment utility of assessment, little has been written about the kinds of methods appropriate to treatment utility questions. In the present article, we present a taxonomy of treatment utility methods in the hope of alleviating this problem.

Second, "Clinical psychologists . . . often make a sharp cleavage between their roles as diagnostician and therapist" (Blatt, 1975, p. 336). Assessment is often not integrated into the therapy process. When the two are disconnected, the value of assessment seemingly turns on the question, Is this diagnosis correct? not Is this assessment useful in treatment? Some clinicians even fear that the assessment process is negatively intrusive on the therapeutic alliance. By focusing on the contribution of assessment to treatment outcome, treatment utility provides an approach for the testing of such concerns and may itself help integrate assessment and treatment roles.

Another part of the problem may lie in the belief that complete psychometric purity is necessary before the treatment utility of assessment can be shown or even examined. Wiggins (1973), who has emphasized the practical importance of assessment more than most psychometric theorists, noted the possibility that "concern with the technical problems of measurement . . . has resulted in a relative neglect of the broader context in which such problems arise, namely, the optimal assignment of men to jobs or treatments" (p. 272). As we will show, there seems to be little reason to delay treatment utility research until psychometrically perfect devices are developed.

Finally, research on the treatment utility of assessment did not have a proper intellectual context until the last few years. Two trends in particular have now come together to make such research both possible and important: the increased specificity of treatment and of assessment, and the emphasis on more theoretically oriented therapy research.

Treatment and Assessment Specificity

In earlier eras of clinical psychology, most psychotherapists practiced according to general schools of therapy. When clinicians apply the same general approach to most clients, assessment data can have few treatment implications (e.g., Wallen, 1956). With the arrival of more specific interventions, however, the need for guidance by assessment data becomes more obvious. Paul's (1969) well-known admonition that one should seek to identify which treatment works best for whom now carries a sense of contemporary urgency.

Compared to the early days of clinical psychology, today there are literally hundreds of competing types of psychosocial interventions . . . what qualities of this person, or aspects of his or her life situation and environment, recommend that he or she be seen in what type of therapy, with what kind of therapist, using what kind of technique? This is the question facing us today. (Korchin & Schulberg, 1981, p. 1155)

Similarly, the early eras of clinical assessment tended to emphasize more global measures of personality. The treatment implications of such measures are often correspondingly vague, especially because general personality characteristics may be apparent to the clinician without an elaborate assessment process. This may be one reason that traditional assessment rarely alters the diagnostic judgments of clinicians, much less their treatment plans (Adams, 1972). Many contemporary assessment devices, however, are now quite specific, measuring specific behavior patterns, specific cognitions, and so on. With specific assessment devices, more specific treatment implications can be explored.

Theory-Oriented Treatment Research

Psychotherapy researchers now agree that simply showing that therapy works is relatively uninteresting. As the recent special issues of the American Psychologist and the Journal of Consulting and Clinical Psychology (Kazdin, 1986; VandenBos, 1986) show, psychotherapy researchers are adopting a more conceptually oriented process approach. The renewed emphasis on the role of theory in therapy research (e.g., Forsyth & Strong, 1986) enhances the im-

The authors would like to thank Paul McReynolds for his helpful comments on an earlier draft of this article.

Correspondence concerning this article should be addressed to Steven C. Hayes, Department of Psychology, University of Nevada–Reno, Reno, NV 89557-0062.
portance of the treatment utility of assessment. As we will discuss later, the theoretical importance of treatment research tends to emerge at the interface between assessment and treatment. In addition, there seems to be agreement that "future psychotherapy outcome research should, at a minimum, . . . [explore] the relative advantages and disadvantages of alternative treatment strategies with different specific psychological and behavioral difficulties" (VandenBos, 1986, p. 111), moving away from "undifferentiated treatment [as] the status quo" (Miller & Hester, in press; see also McLean, 1981; Shaw, 1981). Thus, the field may now be intellectually ready to tackle the treatment utility of assessment.

Several Beginning Examples

The questions that can be asked about the treatment utility of assessment span the range of those relevant to assessment in general. Several short examples may serve as an introduction to the flexibility of the approach.

Target Behaviors and Classification

Research on the treatment utility of assessment can address which target behaviors should be the focus of treatment in specific patient subtypes. In one empirical example, patients with social skills problems were divided into two groups: those with apparent skills deficits and those with apparent suppression of skills because of social anxiety (Trower, Yardley, Bryant, & Shaw, 1978). This had been a popular distinction in the social skills literature (e.g., Hersen & Bellack, 1977), but the relationship between identifying skills deficits or social anxiety and finding differential treatment outcomes had not been shown. Trower et al. (1978) showed that socially deficient patients showed greater improvement with social skills training than with systematic desensitization, whereas socially anxious patients showed equal improvement under either treatment. Thus, the assessment distinction between skills deficits and social anxiety can be said to have a degree of treatment utility, but only for the skills deficit group.

There are a number of other examples of treatment utility research examining specific patient subtypes, although they are surprisingly rare given their obvious importance. For example, studies have shown the treatment utility of assessment distinctions between types of insomnia (Borkovec, Grayson, O'Brien, & Weerts, 1979), anxiety (Altmair, Ross, Leary, & Thombrough, 1982; Elder, Edelstein, & Fremouw, 1981; Ost, Jerremalm & Johanson, 1981; Shahan & Merbaum, 1981), dysmenorrhea (Chesney & Tasto, 1975), and depression (Paykel, Prusoff, Klerman, Haskell, & DiMascio, 1973).

Note that in studies of this type, treatment utility is based on a nexus of assessment devices, theoretical distinctions, and treatment approaches. An assessment distinction can originally have no treatment utility, but may acquire it when theories and treatments that make effective use of the distinction are later developed. Thus, the treatment utility of assessment reflects not only on the assessment device, but also on its theoretical and practical context. Mischel (1968) noted this point clearly:

It is hard to separate the validity of a test from the validity of the personality construct it supposedly measures. The utility of statements about . . . constructs of psychodynamic theory, for example, depends on the value of the predictions and treatment decisions to which they lead in the individual case. The utility of the results reflects both on the theoretical constructs and on the test procedures from which these results are generated. (p. 104)

Devices and Instruments

Another question that could be addressed from a treatment utility approach is whether given assessment devices are useful. Apparently, no studies have yet been conducted that are designed to measure directly the treatment utility of given devices, but in principle the strategy seems straightforward. For example, suppose one wished to assess the treatment utility of using self-monitoring diaries to guide treatment planning, above and beyond that provided by questionnaires. All subjects could complete both questionnaires and self-monitoring diaries, but self-monitoring data for only half the subjects would be made available to therapists designing their treatment programs. If these subjects improved more than those whose treatment plans were based on questionnaires alone, then the treatment utility of using self-monitoring data would be established for this particular disorder. In a similar fashion, the treatment utility of interviews, role-plays, projective tests, and other devices could be evaluated.

Strategies

The treatment utility of general assessment strategies can also be assessed. For example, the treatment utility of a functional analysis of the individual client's problems can be determined by comparing the progress of clients with individually designed treatment programs to those receiving a standard treatment package for a given disorder. A host of similar questions can be asked in the same way. Does objective assessment aid in effective treatment to a greater extent than projective assessment? Does the use of single-subject evaluation designs improve treatment more than not using such designs? Is the use of psycho metrically pure instruments generally more valuable in treatment planning than the use of other instruments? Many of the most central questions in clinical assessment are of this type, and yet often the field has generated virtually no data directly relevant to them. Research on the treatment utility of assessment provides a method for answering such questions.

A Methodological Typology

Research on the treatment utility of assessment requires considerable attention to the question being asked. It seems useful to describe a typology summarizing the kinds of treatment utility studies that can be done and to discuss ways in which methodological difficulties can be avoided. Some of the various areas of assessment that can be evaluated using treatment utility studies are summarized in Table 1. Each row describes a specific kind of study and the question it asks. The last three columns briefly describe the methods appropriate to each question,

November 1987 • American Psychologist 965
### Table 1
**Types of Treatment Utility Studies, Questions Asked, and Methods Used**

<table>
<thead>
<tr>
<th>Type of study</th>
<th>Question</th>
<th>Typical group comparison</th>
<th>Time-series (single case)</th>
<th>Main question between subject</th>
<th>Main question within subject</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><strong>Post hoc studies</strong></td>
<td>Pre–post correlational</td>
<td>Time-series design, then cor rational</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manipulated assessment</td>
<td>What is the effect of the administration of, or data from, different assessment devices or methods on treatment outcome?</td>
<td>Two or more groups randomly assigned. Assessment taken or made available differs. Use of information in treatment stays the same.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manipulated use</td>
<td>What is the effect of different uses of available assessment data on treatment outcome?</td>
<td>Two or more groups randomly assigned. Assessment the same. Use of assessment in treatment differs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Obtained differences</td>
<td>What is the relation between distinct patient types and treatment outcome?</td>
<td>Two or more known groups based on pretreatment differences. Same treatment.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manipulated assessment/</td>
<td>What is the effect of different assessment devices or methods when the information from them is used in different ways to design treatment?</td>
<td>Factorial groups randomly assigned. Assessment taken or made available differs. Use of assessment data in treatment differs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>manipulated use</td>
<td>A priori single dimension studies</td>
<td>Two or more groups randomly assigned. Assessment taken or made available differs. Use of information in treatment stays the same. Treatment assessed in series of time-series designs.</td>
<td>Assessment taken or made available differs. Use of assessment data in treatment differs. Treatments compared within subject using time-series designs.</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>A priori multiple dimension studies</td>
<td>Two or more known groups based on pretreatment differences. Treatment assessed in series of time-series designs.</td>
<td>Assessment the same. Use of assessment in treatment differs. Treatments compared within subject using time-series designs.</td>
<td>Two or more known groups based on pretreatment differences. Treatment assessed in series of time-series designs.</td>
<td>Not applicable.</td>
</tr>
</tbody>
</table>

This document is copyrighted by the American Psychological Association or one of its allied publishers. This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.
<table>
<thead>
<tr>
<th>Type of study</th>
<th>Question</th>
<th>Typical group comparison</th>
<th>Main question between subject</th>
<th>Main question within subject</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manipulated assessment/obtained differences</td>
<td>What is the effect of different assessment devices or methods on treatment outcome for two or more distinct patient types?</td>
<td>Groups randomly assigned within known groups. Assessment taken or made available differs. Use of data in treatment stays the same.</td>
<td>Groups randomly assigned within known groups. Assessment taken or made available differs. Use of data in treatment stays the same.</td>
<td>Assessment taken or made available differs. Use of data in treatment stays the same. Treatments compared within subject using time-series designs in each of two or more known groups.</td>
</tr>
<tr>
<td>Manipulated use/obtained differences</td>
<td>What is the effect of different uses of available assessment data on treatment outcome for two or more distinct patients types?</td>
<td>Groups randomly assigned within known groups. Use of assessment data differs. Assessment taken or made available stays the same.</td>
<td>Groups randomly assigned within known groups. Use of assessment data differs. Treatment assessed in series of time-series designs.</td>
<td>Use of assessment data differs. Treatments compared within subject using time-series designs in each of two or more known groups.</td>
</tr>
<tr>
<td>Manipulated assessment/ manipulated use/obtained differences</td>
<td>What is the effect of different types and uses of available assessment data on treatment outcome for two or more distinct patient types?</td>
<td>Groups randomly assigned within known groups. Use of assessment data differs. Assessment taken or made available differs.</td>
<td>Groups randomly assigned within known groups. Use of assessment data differs. Treatment assessed in series of time-series designs.</td>
<td>Nature and use of assessment data differs. Treatments compared within subject using time-series designs in each of two or more known groups.</td>
</tr>
<tr>
<td>Obtained differences/two or more treatments</td>
<td>What is the effect of different treatments on outcome for two or more distinct patient types?</td>
<td>Two or more known groups of subjects randomly assigned to two or more treatments.</td>
<td>Two or more known groups of subjects randomly assigned to two or more treatments. Treatment assessed in series of time-series designs. Each subject receives one type of treatment.</td>
<td>Two or more known groups of subjects. Two or more treatments compared within subject using time-series designs.</td>
</tr>
</tbody>
</table>
organized by types of design: group comparison and time-series (with the main question either being asked within or between subjects). In this section, we will briefly elaborate on some of the questions that can be asked about the treatment utility of assessment and the methods that can be used.

**Post Hoc Identification of Dimensions**

One approach to the treatment utility of assessment is to administer treatment to many persons and then to identify aspects of the assessment that help predict who will respond to therapy. This is a well-known strategy with a long tradition in clinical science (Beutler, 1979; Miller & Hester, in press). For example, a review of research in depression by Bilski and Friedel (1976) found that higher social class, insidious onset, anorexia, weight loss, middle and late insomnia, and psychomotor disturbance predicted a favorable response to tricyclic medication. In contrast, neurotic, hypochondriacal, and hysterical traits; multiple prior episodes; and delusions predicted a poor response.

Research of this sort often correlates patient outcome with such measures as the Minnesota Multiphasic Personality Inventory (MMPI; e.g., Shealy, Lowe, & Ritzler, 1980; Strassberg, Reimherr, Ward, Russell, & Cole, 1981), locus of control devices (e.g., Poole, Dunn, Sanson-Fisher, & German, 1981; Woodward & Jones, 1980), self-efficacy measures (e.g., DiClemente, 1981; Williams, Dooseman, & Kleifeld, 1984), or demographic characteristics (e.g., Dumas & Wahler, 1983; Ellsworth et al., 1979). These measures are probably popular because they are commonly collected at the beginning of treatment and thus provide ready data bases for post hoc examination.

At an applied level, these reported relations at times are not useful. They may identify variables that are difficult or impossible to change, such as global personality characteristics or the chronicity of the disorder itself (e.g., Bollard, 1982; Brown & Lewinsohn, 1984; Miller & Joyce, 1979). Factors may be identified that predict poor treatment response, without corresponding knowledge about how to produce a beneficial response in these same individuals. Furthermore, research of this sort has often failed to find consistent or strong correlations between assessment measures and outcomes. It is common to see such studies reporting correlations that account for 5% to 10% of the variance in outcome (e.g., Luborsky, Mintz, & Christoph, 1979). Despite the many studies of this type, there is little evidence that they are leading to increasingly effective treatments, though they may help in resource allocation within institutions. There are several possible sources of the limited value of post hoc analyses to treatment outcome.

**The problem of the individual.** In a typical post hoc approach to treatment utility, large numbers of subjects have been assessed pretreatment and posttreatment on the target behavior. The degree of improvement is then correlated with other assessment measures, and certain subject characteristics are identified that predict treatment outcome.

Unfortunately, in this type of study, treatment responsivity at the level of the individual is not determined. Improvement could be due to extraneous variables (real but uncontrolled sources of changes), to measurement inconsistencies (variation in the measurement process itself), or to treatment effectiveness. It is generally assumed that extraneous variables and measurement inconsistencies are evenly distributed and are reflected in interindividual variability. Intergroup variation that is large relative to intragroup variation is thus thought to be treatment related. At the level of the individual, however, knowledge about treatment-produced successes and failures (widely agreed to be important, see Gendlin, 1986) cannot be identified. Thus, any correlations found between the improvement of individuals and their pretreatment characteristics mix all three sources of variability in the correlation (treatment, extraneous variables, and measurement inconsistency). It should be no surprise that such correlations tend to be low and inconsistently obtained.

One solution to this problem is to identify treatment responsiveness at the level of the individual, that is, to use individual time-series designs (Hayes, 1987). This approach will probably require just as many subjects as the previous approach, but because individual assessment results will now be correlated with individually identified treatment effects, stronger external validity for post hoc correlations seems likely. Even so, the other problems with post hoc studies seem likely to continue.

**Selection of measures.** "Going fishing" with many premeasures and a mountain of data from a given treatment program is unlikely to yield highly useful relationships. Post hoc analyses of theoretically derived premeasures are more likely to be useful because it is unlikely that chance correlations will emerge in systematic accord with theoretical expectations. When such measures are available, however, there may be enough conceptual clarity about the issues to warrant more powerful a priori research strategies.

**Inconsistencies between the studied group and application.** The post hoc approach is sensitive to any inconsistencies between the original study and formal cross-validation studies or the clinical environment itself. Thus, treatment utility cannot be assumed with post hoc studies until additional research or application has been done. In contrast, if a treatment utility relation is predicted a priori and successfully shown, then some degree of robustness (external validity) in our conceptualization has already been documented (Hayes, 1987). It is to such a priori treatment utility strategies that we now turn.

**A Priori Group Comparison Approaches: Single Dimension Studies**

In a priori treatment utility studies, the assessment question is known before the data are collected, and the methods are designed accordingly. There are several distinct types. We will first describe studies that can be conducted on single dimensions with group comparison approaches. Single dimension studies assess whether a specific assessment difference relates to treatment outcome. Most ex-
perimental research that has been done on the treatment utility of assessment fits into this category, although few of these examples used the obtained data deliberately to evaluate assessment. There are three subtypes: manipulated assessment, manipulated use, and obtained differences.

**Manipulated assessment.** In this type of study, assessment data are at issue. A single group of subjects is randomly divided into two or more subgroups, and either the collection or availability of assessment data is varied systematically. The therapist then designs and implements treatment in accord with the data available. For example, data from a particular device may be made available for only one subgroup. Differential outcomes between groups confirm the treatment utility of assessment characteristics manipulated. Manipulated assessment studies could address such questions as the treatment utility of the MMPI, multimodal assessment, projective tests, and so on.

**Manipulated use.** These studies are somewhat similar in design to manipulated assessment studies, but they differ fundamentally because the same assessment information is available for all subjects. The researcher manipulates the way in which the assessment information is used. A study we conducted (Jarrett, Nelson, & Hayes, 1981) is one of a handful of examples and contains an instructive flaw. We attempted to compare an idiographically based treatment of depression with a nomothetic approach. In the idiographic group, subjects received treatment modules that matched their specific weaknesses. For example, a subject with frequent irrational thoughts received a cognitive module, subjects with poor social skills received a social skills module, and so on. In the nomothetic group, subjects received modules yoked to the idiographic subjects.

The results showed no differential effect for treatment matching. In this study an interesting, and for our purposes, illuminating methodological problem was missed, however. Unfortunately, many in the yoked treatment group also (by chance) received needed treatment, as indicated by their pretreatment assessment. A subsequent studies could address such questions as the treatment utility of the MMPI, multimodal assessment, projective tests, and so on.

**Obtained differences.** In this third type of a priori study, subjects are divided into groups nonrandomly, based on assessment differences. Subjects then receive one type of treatment. If outcome differs between the groups, the treatment utility of subject differences is shown. This is a more common approach than the preceding two. It is the simplest experimental version of post hoc studies that seek to identify treatment responders and nonresponders. For example, Keller (1983) found that a group of depressed community volunteers with low pretreatment scores on the Dysfunctional Attitude Scale (DAS; Weissman, 1980) showed a greater response to cognitive therapy than did their counterparts with high scores. Keller argued that these data contradict the notion that cognitive therapy is more "relevant" (better matched) to subjects with high scores on the DAS.

Note that these three types of studies all examine the interface between assessment and treatment by producing or observing differences in only one aspect of the assessment–treatment equation. In manipulated assessment studies, assessment or available assessment information varies, but ways in which available assessment information are used are the same. In manipulated use studies, assessment information is the same, but the link between assessment and treatment varies. In obtained difference studies, assessment outcomes are observed to vary, but both assessment methods and treatment are the same for all.

**A Priori Group Comparison Approaches: Multiple Dimension Studies**

More complex questions can be asked about the treatment utility of assessment by crossing the simpler types of questions just described (see Table 1). We term these "multiple dimension" studies. For example, groups developed as in a manipulated assessment study can then be matched with treatment in two or more distinct ways. Such a manipulated-assessment/manipulated-use study asks if difference studies, assessment or available assessment information has varying effects on two or more methods of using assessment information. A practical example might be a study that asks if an individual functional analysis is more successful when direct measures of overt behavior are available compared to when they are not. Similarly, in an obtained differences/manipulated-use study, the use of available assessment information can be manipulated within two patient groups. A practical example might compare the effects of treating the most severe problem first in both depressed and agoraphobic populations. Other possible types of multiple dimension studies are shown in Table 1.

Frequent examples exist for only one type of a priori multiple dimension study. One way obtained differences studies are elaborated is to cross distinct patient groups with two or more distinct treatment approaches. The resultant obtained differences/two-or-more-treatments study asks if two or more types of treatment interact with two or more types of patients. This approach is particularly elegant when each patient group theoretically implies and is crossed with a distinct treatment approach. The resultant factorial design tests not only the treatment utility of the patient group distinctions but also the conceptual and theoretical distinctions that gave rise to them and their implied therapies.

Simons, Lustman, Wetzel, and Murphy (1985) have provided an illustration. These investigators divided the sample of depressed patients according to their pretreatment scores on Rosenbaum's (1980) Self-Control Scale.
Subjects were treated (based on random assignment within patient types) with either cognitive therapy or a regimen of nortriptyline. Subjects with high scores on the Self-Control Scale showed a better response to cognitive therapy than those with low scores. In contrast, subjects with low scores on the Self-Control Scale showed a better response to nortriptyline than those with high scores. This is a particularly sophisticated example because the study was based on a theoretically significant distinction between these two types of patients and a similar distinction between the two treatments used. Thus, the results validate not only the usefulness of the Self-Control Scale, but also, in part, the theoretical distinction on which the study was based. Other examples of this type of study have been provided by Abramowitz, Abramowitz, Roback, and Jackson (1974); Meichenbaum, Gilmore, and Fedoravicius (1971); Prusoff, Weissman, Klerman, and Rounsaville (1980); and Safran, Alden, and Davidson (1980).

**A Priori Time-Series Approaches: Main Question Between Subject**

In the various group comparison approaches to treatment utility, the relation between outcome and assessment is addressed at the level of the group. Unfortunately, as argued earlier, most group comparison studies do not allow identification of those individuals who actually improved because of treatment per se. Indeed, the statistical measurement of outcome is a controversial matter in its own right (e.g., Cronbach, Gleser, Nanda, & Rajaratnam, 1972), and the concept of treatment utility does not solve the researcher of the need to defend the statistical methods used to measure outcome.

The need to identify individual treatment responders can be solved by using time-series designs with all individual subjects (Barlow, Hayes, & Nelson, 1984). These individual analyses can be collapsed into conditions or groups. Researchers can then conduct the same types of studies outlined above, but in a more precise manner.

For example, in a manipulated assessment study, some subjects would receive treatment based on different assessment information than others, but all treatments would be analyzed within subject (e.g., through an A-B-A-B or other time-series design). By collecting many such subjects the researcher can (a) still analyze in a pre-post fashion, if desired; (b) identify at the level of the individual the successes and failures caused by treatment in each group; and (c) begin to correlate observed differences with treatment success and failure, given each assessment approach. The main questions, however, remain between-subject questions because the manipulations are done between subjects.

**A Priori Time-Series Approaches: Main Question Within Subject**

Questions about the treatment utility of assessment can also be asked using within-subject designs, except for those single dimension studies involving obtained differences. Such research usually requires the use of a design that enables two or more treatments to be compared within subject. Many such designs exist (Barlow et al., 1984; Hayes, 1981), but perhaps the most powerful is the Alternating Treatment Design (ATD: Barlow & Hayes, 1979). The ATD consists of the random or semirandom alternation of two or more conditions where alternations are possible about once per meaningful unit of measurement. If measures are most meaningful when grouped by the week, for example, conditions should be capable of alternating about once each week (determined randomly or semirandomly). This allows differential treatment effectiveness to be determined within an individual.

As an example, a manipulated assessment study could be done examining the impact of popular projective tests. Two sets of assessment data would be developed for each of several clients. One set could be based on a comprehensive assessment battery without projectives whereas the second set would include projectives. Expert clinicians skilled in these types of assessment would devise a treatment program for each client, based on each of the two sets of data. If the two suggested programs were distinct, they both could now be delivered within subject in an ATD fashion. In this way one could determine the differential impact of these types of assessment data on treatment outcome. A wide variety of assessment devices, approaches, or packages could be assessed in this manner.

As another example, consider obtained differences—two or more treatments studies. In this type, subjects may be divided into types of subjects based on assessment data. Unlike a purely post hoc approach, however, relatively pure types can be sought and obtained, which allows for a less confounded examination of the relation between obtained differences and outcome. Two or more treatments are then applied within subject, and differential effects are assessed. An example is the study by McKnight et al. (1984). Subjects were obtained who were depressed and either had (a) poor social skills but few irrational cognitions, (b) irrational beliefs but adequate social skills, or (c) problems in both areas. After a short baseline, all subjects received both social skills training and cognitive therapy in an ATD fashion. The degree of self-reported depression is shown in Figure 1. As can be seen, the subjects with only social skills problems improved most following social skills training, whereas the reverse was true for subjects with problems only in the cognitive area. Subjects with problems in both areas improved as rapidly in either condition. Thus, the suspected relation between these obtained differences and two types of therapies was confirmed.

**Pseudo Treatment Utility: Manipulated Target Studies**

There is a final and important type of study that does not truly evaluate the treatment utility of assessment, but appears to do so. It seems important to distinguish the treatment utility of assessment from therapeutically useful strategies in general. Often a given syndrome or disorder will contain several possible target behaviors. For example, in depression one could target overt social behavior, cognitive behaviors, and so on. If treatments that target cognitions are consistently superior to those that target social
behavior, one might be tempted to say that selecting cognition as a target behavior in depression has "treatment utility." Note, however, that this is not an evaluation of assessment because assessment is not used to determine the specific targets or the treatment received. In such studies, researchers may monitor the targets themselves, but they do nothing with this information other than to evaluate the impact of treatment. Manipulated target studies, although valuable, are not studies about the treatment utility of assessment.

The Theoretical Impact of Treatment Utility Research

Treatment utility research sets the stage for important theoretical development, in part because it points out differences that are functionally important and thus require theoretical explanation. For example, there are undoubtedly many functionally distinct subtypes of patients currently mixed together in popular diagnostic systems. Yet even when researchers think they have reliably identified a distinct subtype, they do not know if the distinction is trivial until its treatment utility is established. Indeed, differential response to treatment is itself a good way to identify important patient subtypes in the first place.

Until the sources of treatment utility are identified, however, practices can have treatment utility and still not have much theoretical significance. Suppose, for example, that repeated assessment throughout the course of treatment can be shown to have some degree of treatment utility. Researchers would then have to identify the variables that control or maximize the contribution that repeated assessment makes to treatment effectiveness. They might examine whether it increases the probability that the clinician will engage in other behaviors (e.g., praising improvements), which in turn produce the therapeutic advantages. Researchers might see if the repeated assessment increases a client's contact with relative improvements, serving as an incentive for effective action. Thus, over time, treatment utility research should fuel conceptual as well as technological progress as the processes responsible for treatment benefits are identified.

The advantage of treatment utility research in this regard is twofold. First, it naturally guides researchers in the field toward theorizing based on nontrivial differences (i.e., those that actually make an impact on treatment outcome). Second, researchers examining a particular link between assessment and treatment usually will have some theoretical analysis that leads them to believe that there might be such a link. If it is found, the viability of the theoretical construction increases incrementally. If not, that is also revealing. The few available treatment utility studies already suggest that previous conceptualizations are often only partially confirmed (e.g., Trower et al., 1978), or fail to be confirmed at all (e.g., Keller, 1983). Thus, in addition to enhancing treatment outcome, treatment utility research can improve our understanding of behavior and contribute to clinical science.

Psychometric Criteria and Treatment Utility

As noted earlier, the slow pace of research on the treatment utility of assessment may be due in part to the belief that such questions cannot be asked until all psychometric hurdles are cleared. Psychometrics is based fundamentally on the concepts of reliability and validity. Reliability refers to the consistency of scores obtained by the same persons when examined with a given assessment device under systematically changing circumstances (e.g., different times, different items, different examiners, different internal divisions of the test). Validity refers to what an
the actual test report was at issue, the test also could not giving the test, watching the patient respond to it). There-
have treatment utility because the same test performance
Such a test could not have reliability or validity. If only
in the way they interpreted a particular projective test.
are not readily established (Blatt, 1975; Zubin, Eron, &
invalid and still have treatment utility on that basis. This
does not mean that clients with high degrees of en-
vironmental stress will not respond differently to treat-
This does not mean that clients with low levels of environ-
mental stressors present in a client's life. There
is no reason in general to suppose that clients who just
lost a spouse will also have just been fired. Thus, high
internal consistency should not necessarily be expected.
This does not mean that clients with high degrees of en-
vonmental stress will not respond differently to treat-
ment than clients with low levels of environmental stress.
Many kinds of psychometric criteria are not sufficient,
and at times may not even be necessary, for treatment
utility.

To take this point to its logical extreme, it is useful
to recognize that there may even be times when an as-
essment process could have treatment utility without the
assessment report having any reliability or validity what-
soever. All psychometric evaluations are of assessment
results (e.g., reports, diagnoses, scores, profiles) and not of the actual process of conducting an assessment (e.g.,
giving the test, watching the patient respond to it). There-
fore, if this assessment process can have treatment utility,
then some tests or procedures may be psychometrically
invalid and still have treatment utility on that basis. This
controversial claim can perhaps be seen in the belief,
common among many clinicians, that projective tech-
niques may be useful even if their reliability and validity
are not readily established (Blatt, 1975; Zubin, Eron, &
Schumen, 1965). Suppose that clinicians varied widely in
the way they interpreted a particular projective test. Such a test could not have reliability or validity. If only
the actual test report was at issue, the test also could not have treatment utility because the same test performance
would lead to different scores, diagnoses, or profiles and thus to different treatment recommendations. Even so,
the process of giving this "unreliable" test could have treatment utility. For example, in giving the test and in
watching the client respond, the clinician may unwittingly
learn more about the client's personality. The clinician
might then behave more effectively in treatment even
though he or she may be unable to verbalize the relation
between the test, the client's personality, and the clinical
decisions. In some ways, this possibility seems to match
what many clinicians say about the use of projective de-
vices in experienced clinical hands. Whether the claim
could be supported is an empirical matter, but it is one
that traditional psychometric evaluation cannot deter-
mine. It is a treatment utility question that could be ad-
dressed using a manipulated assessment study.

What this shows is that the relation between psy-
chometric criteria and treatment utility is often open to
question. Thus, there seems to be little reason to insist
in principle that treatment utility must be the last step
in the evaluation of a given assessment device, distinction,
or approach. Treatment utility is shown by a well-designed
study, regardless of the state of previous psychometric
evaluations.

This is not to say that reliability and validity should
be ignored in treatment utility research. Suppose a cli-
nician devised a test, proceeded directly to an assessment
of its treatment utility, and failed to find any. The failure
could be due to poor reliability of various kinds, the lack
of appropriate treatment, a lack of treatment integrity,
or a host of other possible problems other than the prac-
tical uselessness of the dimension the clinician was trying
to measure. Thus, psychometric evaluations limit the
possible sources of a failure to find treatment utility. It
also seems reasonable to expect measures with good re-
liability and validity to be more likely to have treatment
utility, though this is largely an empirical question. Fur-
thermore, it is impossible to build a science of assessment
totally on treatment utility alone. Knowing the effect on
treatment outcome itself requires assessment, and that
means that treatment utility must ultimately be based on
measures that are themselves not validated in this manner.

Once an aspect of assessment (e.g., a device, ap-
proach, or distinction) has been shown to have treatment
utility, it is also then possible to use traditional psycho-
metric evaluations to expand the value of the finding. For
example, suppose that a given measure has been shown
conclusively to distinguish two subgroups of patients who
respond differently to various treatments. If a second
measure separates patients into virtually the same
subgroups, then it must also have treatment utility. A
problem comes in, however, when the relation between
the two measures is only moderate. Then the treatment
utility of the new measure cannot be assumed because it
is not certain that the area of shared variability is itself
responsible for the treatment utility shown with the first
measure.

Treatment utility is not the only functionally based
means by which to evaluate assessment. It is possible for

2 The authors thank Marilyn Demerest for suggesting this example.
assessment to produce descriptions that currently have no treatment utility, but contribute to the understanding of a given disorder. The degree to which assessment contributes to the knowledge base may be thought of as an issue of the “conceptual utility” of assessment. Treatment utility and conceptual utility are undoubtedly related, but they are distinguishable. For example, a device may contribute to treatment outcome (it has treatment utility), but for theoretically unclear reasons (it does not yet have conceptual utility). Similarly, a device may be theoretically useful and as yet be useless in terms of treatment outcome. For present purposes, the point is simply that there are important goals for assessment other than treatment outcome. Treatment utility should not, therefore, be thought of as the only functional grounds for evaluating assessment quality.

Summary

Treatment utility provides the potential for a science of clinical assessment that is built on evidence of successful contributions to treatment. As such, it is an important method of examining the quality of assessment—one in which the word quality is directly related to a central function of clinical assessment. To date, the role of treatment utility has been buried by conceptual confusion, poorly articulated methods, and inappropriate linkage to structural psychometric criteria. The present article was designed to help in these areas. Treatment utility research has also been hindered by nonspecific types of therapy, nonspecific types of assessment, unnecessary divisions between the role of assessor and therapist, and an overemphasis on technique over conceptual advancement. The trends in the field all seem positive in these areas.

Thus, the time now seems ripe for a vigorous expansion of research on the contribution of assessment to treatment outcome. Because treatment utility provides the practical basis for a concern with clinical assessment, it seems important to proceed rapidly to its demonstration.

REFERENCES


