DOCUMENT RESUME

ED 153 093   CG 012 336

AUTHOR   Kelley, John M.
TITLE   Developing Performance-Based Assessment Techniques
                Using the Performance Descriptor Approach.
PUB DATE   Jul 77
NOTE   48p.; Paper presented at the Annual Conference on
                Alcohol and Drug Treatment Evaluation (Nashville,
                Tennessee, September 26-27, 1977)

EDRS PRICE   MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS   Adults; *Counselor Performance; Employees;
                *Evaluation Methods; *Job Skills; *Measurement
                Instruments; Models; *Occupational Tests; Personnel
                Selection; Program Descriptions; *Skill Analysis;
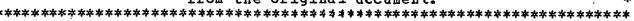   *Task Performance
IDENTIFIERS   *Process Mapping Model

ABSTRACT
        This paper describes in some detail Siegfreid's
Process Mapping Model for describing human performance. The model's
application to task analysis and the advantages of using it when
constructing assessment strategies and instruments are discussed.
Examples of the model's use in building portfolios, behavioral rating
scales, and skill demonstration packages are set forth. Such skill
demonstration packages can, when properly constructed and organized,
identify those individuals whom one can conclude are able to perform
various activities related to a job. They will not, however, predict
the extent to which these individuals will perform adequately on the
job. The authors suspect that skill demonstration packages may be
most useful in diagnosing areas in which a given counselor needs to
improve skill, as opposed to being used in predicting a counselor's
level of effectiveness. (Author)

Developing Performance-Based Assessment Techniques

Using the Performance Descriptor Approach

Presented at the Twelfth Annual Conference of the
Canadian Foundation on Alcohol & Drug
Dependencies

Winnipeg, Canada   July, 1977

By

John M. Kelley, Coordinator of Research/Evaluation
Medical College of Pennsylvania
Division of Social Sciences
1308 Wendover Avenue
Rosemont, Pa. 19010

TASK ANALYSIS AND THE COMPETENCE APPROACH

Competency-based undertakings typically seek to gain a handle on "what" performances or behaviors, skills and knowledges are needed for the efficient accomplishment of specified human endeavors. Once such behaviors, skills and knowledges are agreed upon, they can constitute the raw materials for constructing assessment approaches. One way of organizing performances that has been increasingly used over the last thirty years has been generically labelled "task analysis". This approach inherently views work as composed of a finite number of distinct, identifiable tasks. Such tasks are first specified and described. Oftentimes skills, knowledges, and attitudes requisite to performing a given task successfully are abstracted for each task and consolidated across tasks. Assessment procedures and instruments are then designed directly from this "task data". The information gained from administering these assessments devices is then used as an aid in determining an individual's level of job competence. Proponents of such an approach proclaim that a noticeable advantage is found in the fact that strict adherence to job-related behavior minimizes the chances that one will be assessed according to behavior or qualities which are less relevant to the work at hand.

Like many models that have been advanced over the last few years, the task analysis model has been implicitly used for centuries in selecting and classifying individuals for key roles. Spartan warriors and Renaissance sculptors as well as Cub Scouts and top-level executives have all been the object of such an assessment approach. A good example of the informal use

of this approach is the "try out" for a Little League baseball team. Based on experience and analysis, baseball coaches realize that successful player performance includes such tasks as: 1) throwing a ball in the intended direction; 2) sharply striking a thrown ball with a bat; and 3) catching a ball that has either been thrown or struck with a bat. An assessment situation is ordinarily set up wherein youngsters who seek to try out for the team(s) show up at the baseball field at a predetermined hour. Each youngster is afforded a chance to demonstrate competence by attempting to catch a few balls that are hit to him/her, hit a few balls, and throw a few balls. The coaches (the assessors) either store this data in their memories or make notes on the candidates' proficiency. The youngsters are then judgementally classified according to their ability. Coaches from the same team often huddle and compare observations to come up with consensual choices. Selections are then made based on the information and judgements of a certain youngster's ability.

The "try out" analogy contains all the essential steps common to competence-based, task analysis models.

Step 1: a cluster of essential tasks related to adequate performance was identified.

Step 2: the tasks were agreed upon or consensually validated as important or successful performance.

Step 3: criteria for demonstrating competence in the tasks were specified (In our example this was probably done implicitly to include such criteria as catching as opposed to fumbling; throwing accurately as opposed to erratically; hitting the ball crisply as opposed to feebly or not at all).

Step 4: an assessment strategy was constructed to test candidates ability to perform tasks.

Step 5: the information gathered from the assessment was used in making decisions regarding competence.

The recent approaches to task analysis have served to add precision, organization and uniformity to a process that was previously executed with little standardization or definition. Such a lack of definition lent a quality of arbitrary subjectivity to the selection process which made for a rather whimisical application of fuzzy performance standards.

Thus, in the previous example, one Little League coach may value a certain quality termed "hustle". If this value were not explicitly communicated as being part of the coach's selection criteria, the candidate may overlook demonstrating "hustle". Furthermore, the coach might assess according to standards not shared by co-assessors and not known to candidates. The result would be an inconsistent, inequitable, and possibly inaccurate evaluative process. This example points up two prime assets of task analysis methodologies:

1) tasks which may be implicity and inconsistently thought to be essential are explicitly identified and subjected to consensual judgement.

2) tasks are described in such a way as to produce descriptions of the behaviors involved in efficient task accomplishment.

Such a clear identification and description of tasks is the bedrock for evolving sound assessment techniques which reflect objectivity, fullness, and work relatedness.

TOWARDS A TAXONOMY OF TASK ANALYSIS METHODS

A number of task analysis methods have been advanced and are in
current usage. When a technique boasting novel dimensions was introduced,
it was not rare that the originators of the new approach relentlessly
critiqued alternative predecessor methods in an effort to clearly demon-
strate the superiority of the new method. In actuality, it appears more
historically accurate that one technique was built incrementally upon its
predecessors rather than emerging as an unrelated species independently
generated. Furthermore, different approaches were created for different
purposes. To repudiate a specific methodology for lacking certain features
that were not part and parcel of its original purpose is akin to denying
the value and usefulness of a screwdriver because it does not pound in nails
as well as a hammer or scorn a quater horse for making a poor showing in a
mile and a half race.

The goal of this section will be to evaluate several methods of task
analysis according to a single criterion: their amenability to the con-
struction of assessment instruments that are high in content validity. It
is proposed that using this approach permits a logical hierarchical ordering
of methods according to a single criterion (content validity) that is essen-
tial to performance-based assessment. Actually, there is nothing very com-
plicated about this approach. Content validity is determined by the faith-
fulness with which the assessment mechanism:

  1) Covers content which directly reflects the behavior being assessed.

2) Covers content which is a representative sample of the behavior being assessed. (cf., Anastasi, 1968; Thorndike and Hagen, 1969).

To demonstrate good content validity, an assessment strategy must not only facilitate the evaluation of behavior which is connected to the task(s) at hand, but should facilitate evaluating behavior that systemmatically stresses the important dimensions of the task while minimizing attention devoted to insignificant behaviors. A task analysis procedure which fosters the accurate selection of job-related behaviors in their proper proportion is deemed most desirable for assessment purposes.

Five methods of task analysis which seem to be representative of the spectrum of existing methods shall now be discussed. These methods and their originators are:

1) Functional Job Analysis - Sidney J. Fine

2) Critical Incident Technique - John C. Flanagan

3) Health Services Mobility Study - Eleanor Gilpatrick

4) Stimulus-Response Format - David Barbee

5) Process Mapping Model - Robert J. Siegfried

1. Functional Job Analysis (FJA)

Functional Job Analysis is one of the most widely utilized methods. According to its author, Sidney J. Fine, FJA grew out of experience in job family and test development work at the United States Training and Employment Service (USTES) before and immediately after World War II and in the analysis and classification of all military occupational specialties during the war (1971, p. 82). FJA received its prominence as a tool for developing occupational classification systems. In fact the Dictionary of Occupational

Titles, Department of Labor, represents Fine's FJA model. Fine notes (1971):

> Functional Job Analysis is probably three things: (1) a conceptual system defining dimensions of work activity and thus a way of conceiving the world of work, (2) an observational method of analysis--of evaluating the design of work and its performance. (p. 77)

Fine's assertion that FJA is a method of evaluating work performance is questionable from the perspective of constructing tests from FJA data. The difficulty lies in Fine's limiting his collection of raw data to "the task". According to Fine, "task" is defined as follows:

> A task is an action sequence grouped through time designed to contribute a specified end result to the accomplishment of an objective and for which functional levels and orientations can be reliably assigned. (p. 9)

What this boils down to in Fine's method is listing first a worker's action and its expected outcome along with any pertinent tools or instructions.

By way of example, Fine notes the following task:

> Asks client questions, listens to responses and writes answers on standard intake forms, exercising leeway as to sequence of questions, in order to record basic identifying information (items 1-8). (p. 12)

Tasks are then rated on several scales to determine their orientation to people, data, things and their functional skill level. From a test construction viewpoint, Gilpatrick's criticism (1976) seems applicable to FJA:

> Traditional methods which use task data as inputs to proficiency test construction provide no clear-cut guidelines for content selection and rely on experts in the occupation who are not trained either in job analysis or test construction.

> Traditional task data are usually task inventories and not task descriptions; traditional task scaling usually refers to abstract qualities of "difficulty" or "responsibility" and not a taxonomy of skills and knowledges which are scalable. (pp. 66, 67)

For purposes of creating job descriptions, occupational classifications and pay scales, Fine's method is superlative. However, as Gilpatrick points out, there is minimal description of how tasks are performed. The example given above does not give information as to how one "questions, listens, writes or exercises leeway". It does not describe effective ways to complete the task or describe difficulties commonly encountered along with ways in which such difficulties might be met. As Gilpatrick noted, descriptions of tasks, not just inventories, are needed for content selection of specific test items.

2. The Critical Incident Technique (CIT)

This technique grew out of John C. Flanagan's research with the Air Force during World War II. According to Flanagan (1954), the Critical Incident Technique:

> . . . consists of a set of procedures for collecting direct observations of human behavior in such a way as to facilitate their potential usefulness in solving practical problems and developing broad psychological principles. The Critical Incident Technique outlines procedures for collecting observed incidents having special significance and meeting systemmatically defined criteria. (p. 327)

CIT calls for persons engaged in a job, for example, supervisors, to provide factual descriptions of specific instances in which they observed workers commit particularly effective or ineffective acts.

CIT has been criticized for its vulnerability to subjective values, biases and beliefs concerning what is "critical" to either good or poor performance (Pottinger, 1976). However, a review process could refine CIT data, screening out inappropriate material. Nevertheless, the CIT method yields atomistic, discrete descriptions of work and may neglect tasks which do not happen to be perceived as critical dimensions over the observation period.

Thus, in terms of yielding a representative sampling of job performance the CIT is questionable. However, its accent on producing models of particularly effective and ineffective actions constitutes an extremely valuable resource since these descriptions could be used as models from which assessment instruments could be developed.

3. Behavior Functions Analysis

Eleanor Gilpatrick's work may be seens as combining the precise task identification procedures of Fine with the enlightening descriptive processes of Flannagan. Her work, which is at a highly commendable level of specificity, provides the following information for each task: concrete outcomes of the task, materials used in performing the task, a comprehensive description of what the worker does in performing the task. Furthermore, Gilpatrick pays close attention to the external cues that a worker monitors and which cues indicate actions to be taken. The following brief excerpt which describes a portion of a task is illustrative. The task is part of a radiologic technologist's work in preparing x-ray film processing equipment for use:

> If not already there, peformer goes to darkroom; does not enter while red light is on; if open, knocks to make sure that room is empty or can be entered. Makes sure that no unexposed film is in the open before turning on lights other than safelight. (Gilpatrick, 1976, p.6)

It is apparent that this type of descriptive data is highly adaptable to the formulation of specific assessment strategies. Even the small amount of the material just quoted is sufficient to indicate how easily one could construct cognitive and simulation tests from this type of descriptive data. As Gilpatrick states, there is inherent content validity "because there is nothing you can choose for test items that is not relevant to the work" (Gilpatrick, 1976, p. 68). It would appear that the only area which might

need improvement in this system is the specification of import for various

components of this material to "make certain that all major aspects are

adequately covered by the test items, and in the correct proportions."

(Anastasi, 1968) In other words, any assessment item built from Gilpatrick's

descriptions are "represented" in the task as described, but the format is

not necessarily conducive to insuring that such items are "representative"

of proportionate criticality. It may be apparent to the reader that selecting

assessment material which is "representative" is a judgment stemming from

someone(s) reviewing and comparing the assessment device and the task data.

The quality of "representativeness" is not inherent in the data per se

but is rationally inferred. In fact, since content validity itself is

essentially the result of rational and judgmental analysis, it has sometimes

been termed "rational" or "logical validity." It seems consistent then that

any model which employs a process and display format that increases the

ease of the rational process involved in developing representative assessment

items maintains an advantage over other methods whose display and/or process

makes such a decision more difficult.

4. The Stimulus-Response Format

David Barbee has recently developed an approach which describes a worker's

performances as consisting of a series of perceived stimuli with each stimulus

indicating an appropriate action. The following excerpt deals with a counselor's

task of reducing anxiety and building trust. The data was collected and

validated in a project sponsored by the State of Oregon.

S1.0 If client's non-verbal behavior indicates anxiety,discomfort (e.g. no eye contact, fidgeting, etc.)

R1.0 Then counselor initiates conversation not related to purpose of visit, offers coffee,etc. to give client time to be comfortable.
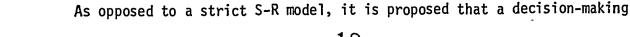
S2.0 <u>If</u> client's anxiety remains <u>high</u>

R2.0 <u>Then</u> counselor gives client verbal feedback (e.g. "you seem anxious") and continues direct verbalization regarding anxiety until diminished.

.   .   .   .   .   .

S4.0 <u>If</u> client anxiety and defensiveness are low

R4.0 <u>Then</u> counselor initiates introduction and discussion regarding purpose of visit

This approach provides a consistent and organized format. Obviously this makes it easier to systematically construct test items from each S-R pair. Segmenting each step makes it simpler to focus and select items that are representative and proportionally distributed in terms of import. The coding system itself is a valuable tool since, when dealing with great volumes of data, each assessment item can be coded in such a way that retrieving the task S-R data that underlies any item is rendered a simple process. It merits noting that the S-R format is merely a way in which work is described, not a dictum as to how work is actually performed.

Nevertheless, this model does not clearly or consistently enough document internal processes. Thus, in the case of judging a client to be anxious certain behavioral indicants are noted; namely, "fidgeting," "no eye contact." In the case of low anxiety, no behavioral indicants are documented. Such indicants are valuable because their perception is a key part of performance and as such deserve systematic betrayal in the performance. Although one could subdivide each stimulus into a condition and its behavioral indicants, it seems more advantageous to employ a model which gets this data systematically within a process flow as opposed to adding categories to a model which was not built to include such categories explicitly. For example, all stimuli are not conditions--some are behaviors. How are these to be logged?

As opposed to a strict S-R model, it is proposed that a decision-making

process model might be more advantageous in describing behavior in a more consistent fashion while maintainig the clarity and organization of the S-R format.

5. The Process Mapping Model

Robert Siegfried,in his work on the Skills Matrix Project which is being conducted by the Medical College of Pennsylvania for the National Institute on Drug Abuse, has developed the Process Mapping Model. Since this model will be a major focus for the remainder of this paper, I shall describe it in some detail. The reader who is interested in learning more of this model as well as the use of models in general and their application to competency-based undertakings is referred to an earlier paper in this series by Robert Siegfried, entitled "Process Mapping: A Technique for Program Improvement." (1977)

As Siegfried (op.cit.) has noted, the Process Mapping Model consists of the application of three distinct meta-models for describing and enriching human performance:

1. The Performance Descriptor Model (formerly referred to as the Task Descriptor Model);

2. The Language Enrichment Model (an adaptation of a model presented by Bandler and Grinder, 1975);

3. A Performance Specification Model.

Each of these three models is specifically treated in Siegfried's paper (op. cit.). The Performance Descriptor Model and the Performance Summary Model will be dealt with here.

THE PERFORMANCE DESCRIPTOR MODEL

Basic to describing human performance is some view or model of what happens when people work. Consistent application of such a model assures that various descriptions of work are homogeneously and systemmatically recorded. Thus, if one were to se'ect two descriptions of different work functions, their portrayal in similar model format aids in comparison and cross-analysis, and helping one avoid errors that may occur in comparing "apples to oranges".

After exploring a number of existing models, Siegfried developed a model specifically for use in the Skills Matrix Project. It is important to note that from its inception, this model was developed with one of the end-goals for its usage being the development of assessment strategies. Thus, with each step in the development and refinement of the model, its pragmatism in the assessment process was one of the checkpoints. Having assessment in mind during the inception and growth of a task analysis model is somewhat unique. Most such models were built for purposes such as: defining occupational duties, problem solving, and developing training curricula. The use of these models to develop assessment strategies was frequently an afterthought or at best a secondary objective. As a result, these models are well-suited to the accurate and smooth development of assessment techniques.

In reality, the Performance Descriptor Model is a meta-model applicable to all forms of goal-oriented human behavior. With respect to work performances,

it is assumed that work is not performed whimsically, but that an individual can "consciously attribute purpose to the activity involved" (Siegfried, op. cit., p.11). In this context, it is further assumed that work activity can be described in terms of:

- The information a worker acquires based on his/her awareness of the needs present in the work context;
- The judgements the worker arrives at based on this information;
- The procedures or actions the worker intitates based on these judgements.

An individual may not actually work in this fashion, but important aspects of what a worker does may be described by using these categories.

Several examples may help in depicting this model. Assume that you spill a cup of coffee on your desk. Your next activity would probably be called something like "cleaning up the mess". We might describe one procedure a person would follow in cleaning up the mess as: Having seen the coffee moving across the desktop, and having concluded that the papers resting on the desk would become wet, then the worker moves the papers. This single sequence is termed a "Performance Descriptor", and is displayed in a three-column format.

SAMPLE PERFORMANCE DESCRIPTOR

COLUMN I INFORMATION	COLUMN II JUDGEMENT	COLUMN III PROCEDURE
Having seen the coffee moving across the desktop	Having concluded that the papers resting on the desk would become wet and stained	Then the worker moves the papers.

Performance descriptors can be strung together in a logical and/or chronological order until the work function under focus has been described from start to finish. For example, the next performance descriptor in this sample may be something like: <u>Having seen</u> that the papers have been placed on a chair, and <u>having concluded</u> that the papers are safe from the coffee, <u>then</u> the worker procures some paper towels and lays them on the spilled coffee. Here is an example using this format which is closer to the kinds of descriptions the Skills Matrix Project is generating:

COLUMN I INFORMATION	COLUMN II JUDGEMENT	COLUMN III PROCEDURE
1.0 <u>Having seen</u> the client fidget, not looking at the worker, sitting or standing rigidly	<u>Having concluded</u> that the client is anxious	<u>Then</u> the worker offers the client coffee and gives the client time to be comfortable.
2.0 <u>Having seen</u> the client continue to fidget and avoid looking at the worker and sitting/ standing rigidly	<u>Having concluded</u> that the client's anxiety level has remained high in spite of efforts to reduce it	<u>Then</u> the worker gives verbal feedback by paraphrasing the client's state (e.g., "You seem anxious"; using warm, understanding tone).

Performance descriptors are written at a highly specific level. Descriptors may be obtained through interviews with workers, self-report, brain storming, or translating existing materials into this format. No hard and fast rules or formulae can be set up to determine how specific a Performance Descriptor should be, as the level of specifity is determined by the purpose to which the data will be applied. It is stressed that descriptors be written in a clear fashion, that the relationship across columns and between descriptors be amenable to devising assessment items. No attempt is made to structure information, judgements or procedures into pre-exisiting categories. While it is recognized that much information is accessed through

sensory channels, (sight, hearing, touch, smell, taste), it is recog-
nized that information may be accessed through internal feelings, recall.
as well as by reading, counting and so forth. "The model makes no
attempt to limit the experiences which can be considered information."
(Siegfried, op. cit., pp. 12, 13). The same holds true for, "carrying
out a procedure". As Siegfried notes:

> "No attempt is made in the meta-model to categorize
> the kinds of procedures which the performer carries
> out. The processes in which the performer engages
> at this point may be cognitive, psychomotor or
> affective in nature. The only assumption is that the
> procedure taken is identifiable. Often the procedure
> will be merely to mentally record the judgment and the
> context in which it was made and return to the monitor-
> ing of that context. In other instances, the performer
> may engage in making a choice between two procedural
> alternatives. When this occurs, the context of the
> performer's activity is now in a new one with a new
> purpose, and it too can be made explicit and mapped,
> (described), in another performance descriptor."(op. cit.
> p. 13).

While it very well may occur that categorization schema for in-
formation, judgments and procedures will be discovered which can aid
in the model's better describing work in a useful fashion, it is

currently held that such extensive classification would be counter-
productive to the purpose at hand; namely, a clear, usable description
of work from which assessment strategies may be effectively built. In
experiments with such classification schema, (e.g. dividing information
into an organized array of sources, such as sensory track, recall, etc.
or dividing procedures into behavioral taxonomies,) it was felt that
the phenomenological flow was broken and that workable description
became bulky, complicated analysès that were not amenable to practical
data collection and review.

Once a series of Performance Descriptors are strung together or
assembled to describe a particular work function, the resultant set is
called a Process Map, (or Unit of Work when the function at hand is
a work performance).

The following process map or unit of work is a sample, adapted
from Barbee's work (1976) with a group of Oregon drug counselors. (The
map is found on the following page).

Once a Process Map or Unit of Work is developed, it is then sub-
jected to a "language enrichment process" to systematically identify
and reduce ambiguously worded statements so that potentially different
interpretations of the same wordings are minimized. Siegfried has
described this process as the, "Language Enrichment Model", (op. cit.,
p.p. 17-24).

The Process Map is then circulated to a group of workers for
validation purposes. This is a standard procedure in most task analysis
processes aimed at securing assurance that the description of work is

accurate, that no important pieces have been omitted and that no extra-
neous or highly subjective (non-representative) pieces have been
included. Based on feedback from this review, the Process Map is
rewritten and re-reviewed until consensual validity is, hopefully,
achieved; that is, until workers can consensually agree that the
Process Map accurately portrays the performance being described. Once
this has been achieved, the Map can be utilized as a model for con-
structing assessment approaches. It bears noting that this, "con-
sensual validation", process is often a time consuming, difficult ordeal.
Furthermore, even when consensus is reached, no pretense is made that a
final, absolute model has been attained, rather it is urged that such models
be treated as "open, dynamic systems" to be revised based upon new information
and needs.

THE PERFORMANCE SPECIFICATION MODEL

When the Process Mapping model is applied to work, the first step is
an identification of the on-the-job functions performed by the worker.
Functions are seen as statements of procedures or work activities (e.g., pre-
pare client treatment plans; help client recognize personal defense mechanisms).
This results in an extensive list of worker functions or activities.
Important or critical functions are then identified for additional analysis.
Certain work activities may need to be further broken down into manageable
components which can be mapped or described according to the Performance Des-
criptor Model just discussed. After the activity is reduced to a set of per-
formance descriptors and a Process Map or Unit of Work is thereby formed, some
supplementary information is gathered: criticality of the activity, the work
context, materials or other needed resources used in performing the activity.

This supplementary material as well as the Process Map is reviewed and a "Performance Specification Statement" is then written for the activity (see Siegfried's treatment of the "Performance Specification Model"; op. cit., pp. 24-26). The Performance Statement includes the following elements:

1. The context in which a performance was or is to be performed;

2. The purpose of the performance being specified;

3. The nature of the performance.

If an assessment approach(es) has been identified, a fourth element is included in the statement enumerating the nature of the process(es) by which the performance is to be judged for adequacy.

Siegfried presents the following example:

Context:	IN a rural drug abuse program in which the performer is the only counselor,
Purpose:	IN ORDER TO reduce the anxiety and build the trust of a white male, formerly drug dependent client,
Performance:	USING AS EXTERNAL RESOURCES the client's body movements, other changeable bodily characteristics and the client's verbal statements,
	THE PERFORMER reflects the client's feelings, paraphrases his statements, confronts the client with contradictions and shares his own feelings,
	WITH BEHAVIOR THAT is empathic, sensitive to the client and consistent,
Judgement Process:	AS JUDGED BY the performer's supervisor,
	COLLECTING INFORMATION BY observing the interaction between worker and client in three counseling sessions,
	AND COMPARING THE INFORMATION COLLECTED TO THE FOLLOWING MODEL(S) FOR ADEQUACY IN THIS PERFORMANCE:

1. Carkhuff and Berenson, Beyond Counseling and Therapy;

2. Bandler and Grinder, The Structure of Magic;

3. Skills Matrix Unit of Work: <u>Reducing Anxiety</u>
    <u>and Building Trust.</u>

The structure of this model is not unique. Similar forms have been
offered by others (e.g., Mager, 1975; Fine and Wiley, 1971; Kibler, Barker,
and Miles, 1970). Its usefulness for our purpose. is to provide organizing,
summary statements for work activities and Process Maps. Other applications,
such as writing specific job descriptions, training objectives, and so forth,
have demonstrated the soundness of this model.

To summarize this section, the Process Mapping Model seems to hold the
following advantages over other task analysis models:

1) It "gets at" the worker's internal processes (information, monitor-
    ing and judgement) in a methodical fashion.

2) It systemmatically displays descriptions in a format which makes
    it easy for appropriate individuals (workers, assessors) to choose
    representative behavior and to weight such.

In addition, the model boasts:

1) A task identification capacity on a level with Fine's FJA,

2) A task descriptor capacity on a level with Flanagan's CIT and Gil-
    patrick's model, and

3) A unified format on a level with Barbee's S-R model.

Based on these observations, the Process Mapping Model appears most .
facilitative of the criteria involved in achieving content validity.

SOME COMMENTS ON ASSESSMENT

The term "assessment" is a relatively new one appearing first in the psychological literature of World War II. Over the past thirty years "assessment" has remained a very general term which Kelly (1967) defines as "any procedure for making meaningful evaluations or differentiations with respect to any characteristic or attribute." (p. 1). This paper is concerned with a <u>particular procedure</u> (the Process Mapping Model) for describing performances applicable to <u>making evaluations</u> and differentiations among a <u>particular group of human beings</u> (drug abuse workers) with respect to a <u>particular characteristic</u> (competent job performance). Some general comments on assessment are useful in providing a context for the remarks which follow.

One way of viewing assessment is to conceptualize it as a process for gathering evidence for decision-making. The type of decision to be made guides the type of evidence required. Frequently several distinct forms or samples of evidence are concomitantly needed to make a sound decision. For example, suppose I intend to drive across the country in order to attend a particular function and am deciding which route to follow. It may be helpful for me to gather evidence about the speed, safety, scenery, and projected lodging stops for various potential routes. Once I have gathered the needed information and judged it validity. I have in hand evidence upon which I can base my decision and subsequently carry it out.

Difficult questions arise, however, when our deicisions involve complex human behavior, such as deciding an individual's level of work proficiency or competence. Precisely what evidence will be acceptable in making such a decision? In order to obtain evidence, what information must be gathered and what judgments must be made? What resources must be expended to gather the evidence? Does the evidence merit the resources expended? These are questions which must be

raised. It is not uncommon that strikingly different forms of evidence are used by different persons in order to make similar types of decisions. Consider the examples of evidence utilized singularly or in different combinations with varying weights in order to select job candidates: letters of recommendation, resumes and portfolios, non-structured as well as structured interviews, rating scales, vocational interest inventories, intelligence and personality tests, performance tests. It is readily apparent that each technique yields evidence that varies significantly from the other in terms of important factors such as reliability, objectivity, and relationship to work. Although there seems to be consensus that performance tests (wherein the applicant demonstrates skill by actually "doing" selected facets of the work) supply the most valid evidence for skill assessment, even these tests are problemmatic. Performance tests are often too expensive for practical use. In addition, performance tests that deal with complex behaviors of an interpersonal or affective nature frequently require human raters, which opens the possibility of erroneous ratings due to the subjective perceptions, values, biases, and assumptive worlds of the judges. With respect to counseling, authentic performance tests in which an applicant would be assigned to work with a real client are generally dismissed on ethical grounds in that such a procedure could very well be counterproductive to the client's development and violate confidentiality. Furthermore, even performance tests are vulnerable to inaccuracy stemming from the fact that all tests sample a small bit of an individual's behavior. Perceiving the evidence from such a limited behavioral sample as equalling "typical" performance is unwarranted due to vicissitudes o the applicant and the extra strain that may be present due to the applicant's awareness of the fact that he or she is in a test situation. Other threats to the validity of even the most rigorously constructed performance tests could be singled out; however, the point at hand is that even evidence gathered from those procedures

which boast the best validity must be viewed with some caution since error

could result if one were to identify unequivocally the assessment performance

with typical work performance. The fact is that no existing assessment

strategy or combination of such strategies which evaluate complex behaviors

yields information which correlates exactly with, and accurately predicts, job

performance. Given the fact that an error potential is present it is incumbent

upon assessors to select those procedures which yield the best evidence for the

decision in question. In order to accomplish this goal decisions which are

based on value as opposed to scientifically verified evidence must be addressed

more directly. Decision-makers must expend great effort in discovering hidden

values, explicitly verbalizing them, and exploring whether such values are

held in common be co-decision makers. If those behaviors that are consensually

valued can be clearly mapped or described, models of the valued behaviors can

be developed and promulgated. W.thin the context of work performance such

behavioral models can be the basis for more equitable and effective evaluation

since assessment strategies can be built from a common base of valued and

desired behavior that has been explicitly described and communicated to those

being evaluated.

USING THE PROCESS MAPPING MODEL FOR COMPETENCE IDENTIFICATION, DEMONSTRATION, AND ASSESSMENT

Performance Identification and Selection

If decision-makers (e.g. certifiers, credentialing boards, teachers,

supervisors) purport to assess an individual's competence by means of a

performance-based model, it seems consistent that they must first select those

performances an individual must demonstrate in order to be considered competent.

Rather than each group of decision-makers going through their own extended

a priori or empirical analysis in order to "come up with" desirable performances,

it seems more efficient to work from a list of performance statements. Efficiency is not the only advantage to working from an array of statements and maps. Indeed the selection of desirable behavior is replete with value judgments as opposed to verified absolutes. Selecting statements and maps serves the earlier mentioned purpose of "consensualizing" assessors and providing behavioral descriptions of the desired performances.Such performance statements pertinent to the work performed in the drug abuse field will be generated by the Skills Matrix Project, along with a number of sample Process Maps for certain performance statements. Should a performance statement be missing its companion Process Map(s), pertinent maps could be drawn up and consensually validated. It seems practical for such statements and maps to be centrally maintained through a system parallel to the several "task banks" that are currently in use. Armed with statements and process maps, decision-makers would have performance statements before them to select the desired behaviors as well as behavioral descriptions (the maps) for each performance.

It is further postulated that various Performance Descriptors from different Process Maps could be analyzed and clusters of related information, judgments and procedures from across maps could be abstracted. For example, the "accurate interpretration of non-verbal behavior" is a judgment that appears in many maps on many job levels. Abstracting and combining similar information, judgment, and procedural components, may serve to make the identification and selection of desired performance more manageable. For example, should the desired performance statements and maps become too extensive or bulky, it may be more feasible to work with certain "collapsed" units evidencing frequent occurence and commonality across maps.

In essence, the desired performances can be selected in two ways:

1. By selecting specific Process Maps or Units of Work; and,

2. By selecting common information, judgments and/or procedures from across many maps or units of work.

Each approach is valuable in "getting at" different kinds of information. The specific map approach can elicit information about a worker's ability to perform a specific series of activities and leads itself to determining competence in specific work performances. The cluster approach is useful in eliciting information about the extent to which a worker can demonstrate skills and knowledge more generic to the job as a whole. Each approach has advantages and disadvantages. Because each approach elicits different kinds of information -- both of which may be important in assessing competency -- a combination of the two approaches may be necessary for an adequate selection of desirable performances. (See Skills Matrix Project, Administrative Report, January, 1977.)

For example, here is a list of performance statements with respect to counseling which was shaped primarily from work conducted by the Human Resources Development Center dealing with substance abuse counselors in Maine (1977).

Context: IN a typical outpatient drug abuse program,

Purpose: IN ORDER TO help the individual come to a better understanding of his/her problems, develop better coping abilities and reduce the desire for substance abuse,

Performance: THE COUNSELOR:

1) Listens to the client
2) Reflects feelings of the client
3) Encourages the client to respond spontaneously
4) Explores areas of concern
5) Offers understanding and advice
6) Identifies problems
7) Suggests solution alternatives and the merits/demerits of each
8) Perceives verbal/non-verbal cues
9) Facilitates client understanding of how his/her behavior and feelings affect his/her relationship with others
10) Reduces anxiety and builds trust*

*Source: Oregon task data.

Suppose maps similar to the one on the following page were included for each performance. With this data in hand, assessors would be in the position to select specific performances and maps which are valued as indicators of competence. Given such a reference base, the identification and development of additional performance statements and mapd describing behaviors deemed important for demonstration could be generated.

Performance Demonstration

When the performances deemed as valuable benchmarks of competence have been identified and selected, a situation must be identified or developed in which the person will actually demonstrate the desired performances. A number of options are available. Information can be collected about a person's performances:

1) On the job (past and present)
2) In a classroom
3) In a training experience
4) In simulations of on-the-job situations
5) On tests, and so forth
6) In creating products (e.g., samples of work)

While on-the-job behaviors appear to be the most reliable source of information, collecting such information may at times be impractical, impossible or irrelevant. For example, it is impossible to collect such data on non-experienced candidates seeking to enter a field; it may be irrelevant to collect such data for persons seeking to demonstrate competence in new job-positions (e.g., a move from counselor to administrator involves a significant transition in tasks).

Basically, it is up to the decision makers to determine which demonstrated performances will be acceptable as evidence for competence. Indeed, it seems advantageous to map this vital process (i.e., the process of selecting which performances are to be demonstrated and acceptable as evidence)

TASK DESCRIPTOR FORM

UNIT OF WORK:	Reducing Anxiety and Building Trust
JOB TITLE:	Treatment - One to One Counseling
CAREER AREA:	Counselor

(Information Column) "HAVING"--	(Judgment Column) "and, HAVING CONCLUDED THAT"--	(Action/Procedure Column) "THEN"--
1.0 SEEN the client 1.1 fidget 1.2 not able to look at the worker 1.3 sitting or standing ridigly	1.0 the client is anxious	1.0 the worker 1.1 offers coffee 1.2 gives client time to be comfortable
2.0 SEEN the client continue to 2.1 see 1.1; 1.2; 1.3 (H)	2.0 the client's anxiety remains high	2.0 give the client verbal feedback by paraphrasing the client's anxiety; e.g., "You seem anxious" and continue to verbalize about the client's anxiety until anxiety is diminished.
3.0 HEARD the client's responses	3.0 the client's responses are defense mechanisms to avoid anxiety 3.1 projecting 3.2 denying 3.3 intellectualizing 3.4 rationalizing 3.5 evidencing reaction-formation	3.0 the worker spends time interacting with client by paraphrasing statements 3.1 show client that worker is aware of what client is saying 3.2 reflects back client's behavior or statements of emotional upset
4.0 ------------ ------------>	4.0 the client does not trust worker	4.0 worker "shoots the breeze" for several sessions to develop sense of "I want to get to know you as a person".
5.0 HEARD the client's responses	5.0 client lacks relationships (intimacy) other than with substance abuse	5.0 see 4.0 (T).
6.0 HEARD client responses	6.0 the client's motivation is low	6.0 see 4.0 (T).
7.0 HEARD client responses	7.0 client anxiety and defensiveness is low 7.1 trust is developing in therapy 7.2 has appropriate affect 7.3 client is ready to work on problem(s)	7.0 worker has client identify 7.1 events, people, things that sound as if they hurt or bother client
8.0 SEEN body posture e.g., head leans slightly forward, arms and legs slightly limp or casually crossed	8.0 client body language is relaxed	8.0 worker confronts client with contradictions and asks 8.1 logically present the ignored, undistorted data 8.2 strongly request that client responds to clarified information 8.3 challenge client only on those behaviors in which positive change is feasible.
9.0 HEARD client responses	9.0 the client trusts worker but shows 9.1 verbal/non-verbal inconsistencies 9.2 illogical thought patterns 9.3 avoidance of feelings 9.4 denial of internal and external data	9.0 see 8.0 through 8.3 (T).
10.0 -------------------------->	10.0 the client trusts worker (see 7.0) and elicits strong positive or negative feelings in worker	10.0 worker shares feelings with client if worker is reasonably sure sharing will achieve more positive interactions from client OR will lead to more recognition and possible positive change in client.

NAME AND ADDRESS: This sample was derived from performances described in data generated in an earlier, related project conducted in Oregon under the direction of Dr. David E. Barbee.

TASK DESCRIPTOR FORM

UNIT OF WORK:

JOB TITLE:

CAREER AREA:

(Information Column) "HAVING"--	(Judgment Column) "and, HAVING CONCLUDED THAT"--	(Action/Procedure Column) "THEN"--

- 6 -

NAME AND ADDRESS:

HAVING --	HAVING CONCLUDED --	THEN --

1.0   seen the group members sitting and looking at me
        -AND-
observed that the group members have ceased talking among themselves
	the group is giving me their attention	I introduce myself to group
		        -AND-
		describe the objectives for the session
		        -AND-
		ask if anyone has any questions

I have just completed a Task Descriptor. As is evident, the Task Descriptor Model has three components:

1. Having Column: The information I use to make judgments. Such information may arise from a sensory or external source -- what I see, hear, touch, read, count, smell, taste -- or an internal source -- what I recall.

2. Having Concluded Column: Judgments I make on the above information.

3. Then Column: Procedures I follow resulting from my judgments.

STEP 7.   Identify the very last or terminal action which completes the unit of work and write a Task Descriptor at the bottom of a separate Task Descriptor Form. For example, my group training session might end with me thanking the group for their coming and dismissing the group.

HAVING --	HAVING CONCLUDED --	THEN --

1.0   seen that there are no responses to my question: "Are there any additional questions?"
        -AND-
seen general head nodding
        -AND-
heard trainees answer, "Yes." to my question: "Do you have a good grasp on the material?"
	that trainees do have a good grasp on the material
	        -AND-
	that no further work or comment is needed
		I thank the group for coming and dismiss the group.

STEP 8.   We've now set some limits around the unit of work, namely, a beginning and an end. The final step is to fill in Task Descriptors (TD's) that occur between the initial and terminal TD's. Task Descriptors proceed in a logical order. The action of Column III often stimulates new information --> judgment --> action. For example, my first TD ended with the action of me "asking if anyone had questions". This action gives rise to new information and so forth.

31

HAVING --	HAVING CONCLUDED --	THEN --
1.0 seen the group members sitting and looking at me -AND- observed that the group members have ceased talking among themselves	the group is giving me their attention	I introduce myself to the group -AND- describe the objectives for the session -AND- ask if anyone has any questions.
2.0 seen a trainee raise his/her hand	that he/she wished to ask a question	I look at the trainee and say, "You have a question?"
3.0 heard the question	I can answer the question effectively	I answer the question and ask trainee if answer was satisfactory.

STEP 9. Logically record Task Descriptors until you arrive at your previously determined terminal descriptor. Number each Task Descriptor as you go along. You have now finished your Unit of Work. Congratulations!

STEP 10. Read over your Unit of Work to make sure it is complete, accurate and that there is a logical flow.

STEP 11. Mail your Unit of Work to us and we'll be happy to review it and personally write you the results of the review.

It is apparent that this approach yields very specific data in terms of information, judgments and actions relating to job performance. It is our objective to review data collected by project task force members, identify areas of commonality and explore whether or not there is professional consensus around specified informational, judgmental and procedural areas. In those areas where consensus is achieved, a variety of job performance-based tools can be constructed: supervisory and evaluative checklists, assessment instruments, training curricula and feedback instruments. Further Administrative Reports will address specific ways of shaping Task Descriptor data into helpful tools.

o - o - o

BIBLIOGRAPHY

A brief bibliography of literature pertinent to the project and to the development of competency-based assessment materials is available upon request. This bibliography identifies approaches to the definition of work and the development of skill demonstration materials, as well as literature which contributed to the Skills Matrix Project's Task Descriptor Model.

o - o - o

so that this process itself may be refined, promulgated and improved upon.
Finally, it bears emphasizing that in selecting performances which best
reflect real work behaviors are the soundest evidence for competence demon-
stration. In other words, whenever there must be assurance of the ability to
to do a particular operation rather than to talk about or select the appro-
priate, reliance should be placed on behaviors which closely mirror the
performance in question.

Performance Assessment

Once desired behaviors have been selected and acceptable demonstrations
have been agreed upon, some form of evaluating the performances must be de-
vised. This subsection will deal with three modes of assessment that are
currently being utilized: the portfolio, the behavioral rating scale and
the test. Actually these modes are not distinct but may overlap consider-
ably; nevertheless, for purposes of analysis, each mode shall be reviewed
separately. The review will not dwell upon the pros and cons of each mode
from a measurement perspective*, but will focus upon utilizing the Process
Mapping Model in the construction of instruments pertinent to each mode.

The Portfolio

The portfolio can be conceived of as an organized package which docu-
ments the job-related experiences and skills of an individual and presents
evidence that the person has demonstrated these skills. Evaluators may look
over the portfolio and assess competence based upon the evidence provided.

*The reader interested in exploring the measurement perspectives (e.g.,
 the validity, reliability) of each mode will find pertinent material in
 most standard texts dealing with psychological texts and measurement such
 as Anastasi, 1968; Cronbach, 1970).

As Ziener (1977) recently wrote with respect to drug counselors, "Unlike the artist, a counselor cannot pack a successful client into a portfolio." (P. 4). This raises the question of justwhat goes into a portfolio and how is it assessed. Portfolios used in the academic sector commonly include such components as: a resume, work experience descriptions, education, training, travel, hobbies, work samples and other documentations of life experience. The burden for putting together a portfolio usually falls upon the worker or candidate. If guidelines are not well laid out, the possibility of bias exists wherein the worker with better editorial skills and skills in portfolio construction will be favored over workers who may perform the job better but are not endowed with the skills required in portfolio development. Thus, the validity question is raised: Are we measuring what we intend to measure, namely, job performance capability? In the Process—Mapping approach, a standardized list of performance statements with maps could be assembled thereby listing the desired performances. A uniform list of acceptable ways in which the behaviors can be demonstrated could also be built. Finally, acceptable assessment techniques for the performances could be identified. The following example lists these three elements in a simple three column break-down.

DESIRED PERFORMANCES	ACCEPTABLE DEMONSTRATIONS	ACCEPTABLE ASSESSMENT MODES
A. Listens to client	on-the-job	supervisory ratings
B. Reflects feelings of	in a classroom	peer ratings
C. Encourages spontaneous client responses	in a training experience	self-ratings
D. Explores ares of concern	in a simulation situation	teacher ratings
E. Offers understanding and advice	in work samples	trainer ratings
F. Identifies problems	on tests	test scores

G. Suggests solution
alternatives   ----------   ----------

H. Perceives verbal/non-
verbal cues   ----------   ----------

I. Facilitates client's
self-understanding   ----------   ----------

J. Reduces anxiety and
builds trust   ----------   ----------

K. Writes psycho-social
history   ----------   ----------

L. Completes intake forms   ----------   ---------- *

Units of work

If performance statements and maps were constructed for each desired

performance, individuals would be provided with a clear description of what

evidence must be gathered and assessors would be provided with clear pic-

tures as to what behaviors were to be judged. Thus, a portfolio face sum-

mary page(s) could be built which would consist of a series of items similar

to the following one:

PERFORMANCE STATEMENT	CONTEXT OF BEHAVIOR DEMONSTRATED	ASSESSMENT MODE
1. During a one-to-one counseling session in a typical agency setting (context), in order to help a client better understand his/her problems (purpose), the counselor <u>reduces client's anxiety and builds trust</u> in a manner similar to Map 1 (See pg. _Unit work #1_	(circle ones included in portfolio - see key on following page) A B C D E F G If G was circled, explain here _____	(circle ones included in portfolio - see key on following page) 1 2 3 4 5 6 7 8 If 8 was circled, explain here _____

*This example is meant to be explanatory, not exhaustive.

KEY

Context	Assessment Mode
A = On-the-job	1 = Supervisor ratings
B = Classroom	2 = Peer ratings
C = Training	3 = Self ratings
D = Simulation	4 = Teacher ratings
E = Work sample	5 = Trainer ratings
F = Tests other than simulation	6 = Simulation tests
G = Other	7 = Other tests
	8 = Other

The rest of the portfolio could sequentially be composed of presenting

the evidence indicated on the summary page, as well as any supplementary

information which may be required. The organization of material according

to such a format also provides the ultimate assessors of the portfolio with

a systematic document that facilitates the review process. Much more could

be said here, such as the need for assessors to achieve consensus should they

value and hence score one form of assessment or behavior context over another.

A subsequent paper to be produced by the Skills Matrix Project of the Medical

College of Pennsylvania will address the more specific issues in portfolio

assessment. It suffices here to point out that the Process Mapping Model is

advantageous in its focusing of desired behaviors and thereby providing a

fairly specific guide to evidence gathering.

The Behavioral Rating Scale

Rating scales of one form or another are frequently used to provide more

standardized frames of reference for persons evaluating behavior. While

there are numerous problems in constructing and using rating scales

(Thorndike and Hagen, 1969), such scales are often the most practical and

efficient way of assessing performance. The Performance Descriptor

Model fosters the clear systematic identification of behavior which can

be built into such scales. As previously noted three types of behavior are
included in the model: information accessing, making judgments and implementing
procedures.

Consider the following performance descriptor excerpted from a map or
unit of work dealing with a group counselor's initiating a relaxation exercise:

Having (Information)	and Having Concluded (Judgment)	Then (Procedures)
Seen group members manifest: - rapid eye movement - rigid postu. `s - taut facial expressions - clenched hands	That group members are tense	The worker decides to implement a relaxation exercise.

An item for rating could easily be constructed for each column. For example:

Rate the ability of counselor X to:
1. Pick up cues that group members are tense (e.g. rapid eye
   movements, rigid postures, taut facial expressions, clenched
   hands);
2. Make accurate judgments on such cues,(e.g. judge group members
   to be tense as opposed to attentive);
3. Make appropriate and timely decisions as to when a relaxation
   exercise should be implemented with a group.

Our concern here is not with the psychometric principles and procedures
underlying scale construction; rather, it is with the ability of the Process
Descriptor Model to yield systematic content for items that tap the internal
behaviors of information accessing and judgment as well as procedures. In
this sense the model has the advantage of not only offering a systematic guide
to item construction which diminishes the threat of non-consistency, but also
providing a means of assessing the three distinct behaviors bound up with
executing a specified element of work performance. It depends on the purpose
and goal of the assessment process to determine which behavioral mode
(i.e. information, judgment, procedure) or combination of modes one should

rate and which particular performances within each mode are most important and, thereby, merit rating. These are judgments which pertinent decision-makers must formulate. It is contended that a further asset of the model is that its clear, triadic lay-out of the data can provide organized information of a high level of specificity which can aid decision-makers choose where to hone in on desirable performances. A sample rating scale which focused on the procedural behaviors of the unit of work "Reducing Anxiety and Building Trust" is presented in the appendix.

The Test or Skill Demonstration Package

At the present time, the Skills Matrix Project is working with the con-struction of four (4) types of tests or "Skill Demonstration Packages" (SDP's). Each type of SDP deals with a distinct kind of "synthetic" work situation:

1) Worker Interaction Simulations - in which the person INTERACTS with people, information or things in a simulated work setting (e.g., role play simulation).

2) Worker Response Simulations - in which a person RESPONDS to people, information or things which are depicted in a situation which simulates work (e.g., responding to a video-tape of a "typical client's" dilemma).

3) Case Descriptions - in which a person responds. to a description, problem or problems encountered in a typical work situation by describing the information, judgments or procedures he/she would collect, form or follow (e.g., a written situational test).

4) Information-Recall Prompts - in which the person answers, ques-tions or responds to statements which are related to aspects of a work situation but do not describe the exact work situation itself (e.g., tests of work-related knowledge).

For purposes of illustration, approaches and/or items illustrating these four SDP types will be described using the following set of performance

descriptors pertaining to a group counselor's initiating a relaxation

exercise.*/**

HAVING-- (Information)	and, HAVING CONCLUDED- (Judgement)	THEN-- (Procedures)
1.0 SEEN group members manifest: -rapid eye movement -rigid postures -taut facial expressions -hands clenched	1.0 that group members are tense	1.0 worker decides to implement relaxation exercise.
2.0 RECALLED several relaxation exercise	2.0 that a basic exercise requiring: -deep breathing -closed eyes -tightening/untighten of muscles is non-threatening and appropriate	2.0 worker asks if group would like to join in a simple relaxation exercise.
3.0 SEEN group members nod heads AND/OR HEARD group members say "yes"	3.0 that the group is willing to participate.	3.0 worker gives clear instructions for the exercise using relaxed verbal tone while demonstrating steps of exercise.

*The Appendix contains samples of an Information-Recall and Case Study SDP dealing with reducing anxiety and building trust. The reader may wish to compare these SDP's with their map or unit of work (page   ), to see how items were built directly from performance descriptors.

**This section is taken from an earlier paper of the Skills Matrix Project, presented at the National Institute on Drug Abuse Conference, San Francisco, 1977, entitled "Toward the Identification, Demonstration and Assessment of Drug Abuse Worker's Skills".

If one accepts that these Performance Descriptors accurately portray work performance, many assessment items can be constructed from the Performance Descriptor data. Beginning with the information-recall level, one could construct several types of paper and pencil items which assess whether or not a person can bring to mind and write down appropriate material. Items of this nature tap into:

1) Recalling information

   Sample: List three non-verbal behaviors that commonly signify that a client(s) is tense.

2) Making judgements

   Sample: Given a counseling situation wherein members are exhibiting the following behaviors: rapid eye movements, rigid posture, nervous laughter, taut facial expressions and clenched fists, how would you describe the emotional state of the group?

3) Formulating procedures

   Sample: If during a group session you judge the group to be tense, describe a reasonable course of action that you as a counselor would take to relax the group.

The particular Performance Descriptor component (recalling information, making judgements, formulating procedures) which one chooses to address would depend upon the purpose of the testing and the job performance data which is sought.

It is perhaps apparent that information-recall type items, although derived from job performance descriptors, are not extremely congruent with the context in which work is performed or with the actual on-the-job behaviors. One way of increasing the SDP's approximation of reality is to design items which are of a case study nature.

   Sample Case Study Item: Picture yourself as a group counselor. This is the third meeting of the group. Members have gotten to know one another but have avoided any problem sharing or confrontation. Right now the group appears to be very tense.

> You decide to try a relaxation exercise in hopes that tension
> will be reduced and cohesion will be increased.
>
> Please describe in the space below:
>
> 1)   The relaxation exercise that you would elect, and
> 2)   How you would go about implementing the exercise.
>
> Be as specific as possible.

Such a case study item could be portrayed in a written or video-tape format. While the video-tape offers a visual picture which more closely resembles work than simply a written description, both the SDP stimulus and the response behavior diverge quite a bit from the actual work performance executed in conducting a relaxation exercise. Test validity and reliability may be jeopardized since factors such as reading skills, cognitive recall and writing skills may actually be measured, as opposed to real work performance.

A popular technique which solicits performance more closely consonant with on-the-job functioning is the role-play simulation. A group of people playing the part of clients could be trained to manifest the tension symptoms found in the Performance Descriptor. Individual case study scripts could easily be constructed for each actor in the client group. The examinee would also receive information regarding the actor clients and the clinical context. Additional prompts could even be relayed to the examinee during the simulation exercise. For example, after a few minutes of working with a "tense" group, the examinee may receive a message requiring him/her to initiate a relaxation exercise. Trained judges would be required to rate the examinee on behaviors derived from the Performance Descriptor data and shaped into a checklist.

Sample Checklist Items

Please rate the candidate on the following behaviors:

1) Determining that the group is tense
2) Giving clear instructions for relaxation exercise
3) Using relaxed tone in explaining relaxation exercise

The same checklist and rating procedure used in a simulatioi. often may be directly applicable by supervisors or other appropriate persons in rating someone's job performance relative to selected Units of Work.

It is consistent with test theory, especially within an occupational context, that the closer the behaviors demonstrated and the testing environment approach job behavior and performance, the more lid the information elicited. In the examples just mentioned, the role-play simulation ana the rating of on-the-job behavior appear to be more appropriate than written or oral question-and-answer techniques. Of course, where actual job behavior is of a writing nature, written tests may best mirror the work situation.

Finally, it bears noting that companion SDP's are easily developed if one wishes to focus upon the information and judgement components implicit in a live interactional SDP. For example, after a role-play the examinee could be asked to answer such questions as: What specific non-verbal cues led you to judge that the group was tense?

CONSTRUCTING SDP's FROM ACROSS UNITS OF WORK

In reality, it may be very impractical to construct individual SDP's for each component of each Unit of Work performed in the course of one's job. A given worker may perform hundreds of functions which could be accurately documented only through thousands of Performance Descriptors. For example, the seemingly simple function of driving a car has been analyzed to include over 1.700 tasks!

Rating Units of Work with respect to their job criticality and frequency is helpful in isolating those Units of Work descriptive of tasks and functions considered by workers

as most essential to job performance and the delivery of quality care. In addition, both logical and preliminary project data support the hypothesis that different Units of Work evidence similar generic behaviors. Consider the following Performance Descriptor:

HAVING-- (Information)	and, HAVING CONCLUDED-- (Judgement)	THEN-- (Procedures)
1.0 SEEN the client -clench teeth -pound desk forcefully	1.0 the client is angry	1.0 the worker (using relaxed tone) asks client to talk about what he/she is feeling.

Both this example and the previous one dealing with a relaxation exercise require the worker to perform similar behaviors. On a very basic level, the worker is required to visually perceive client body movements such as posture, clenched hands and teeth and so forth. "Seeing body movements" is common to each example. While "seeing" may be assessed by simple vision tests, a behavior more germane to the counselor and evident in both examples is "the interpretation of non-verbal behavior."

Information about a person's ability to interpret non-verbal behavior may be a valuable indicant of competence. An SDP could be constructed to assess a person's skill at interpreting messages which are transmitted non-verbally by clients. For example, a series on non-verbal client scenarios could be video-taped and an examinee could be instructed to interpret the scenarios by means of a paper and pencil instrument. It must be stressed that prior agreement is imperative if certain SDP responses are to be scored on a "right-wrong" continuum. The Medical College's approach to achieving such agreement will be to attempt to acquire consensual validation of sample responses from workers in the field and from experts they have identified.

Both the "relaxation exercise" and the "anger" Descriptors depicted the worker as communicating to clients with a relaxed verbal tone. It is possible that responding to clients who manifest high degrees of negative emotion (tenseness, anger, anxiety) in a relaxed tone may be viewed as a generic skill for which SDP items would then be constructed.

Our approach to consolidating across Units of Work will consist of a magnification of the process just described. Performance Descriptors from various Units of Work will be analyzed to determine whether or not and to what degree common behaviors are evident in the areas of:

-Monitoring informations
-Making judgements
-Selecting procedures

Clusters of generic behaviors will be reviewed by state task forces involved in the project, and sample SDP's will be constructed for a number of generic behaviors. Formal validation of SDP's is not in the project's scope of work.

SUMMARY

This paper has attempted to describe in some detail Siegfried's model for describing human performance, the Process Mapping Model. The model's application to task analysis and the advantages of using it when constructing assessment startegies and instruments were discussed. Examples of the model's use in buidling portfolios, behavioral rating scales, and Skill Demonstration Packages were set forth.

It is important to note, however, what we believe Skill Demonstration Packages based on Process Mapping data can and cannot do. They can, when properly constructed and organized, idenitfy those individuals whom one can conclude are able to perform various activities related to a job. They will not predict the extent to which these individuals will perform adequately on the job. Too many other factors extraneous to the work activities themselves influence the worker's abilities to perform those activities. Before we can predict on the job performance, these extraneous factors must be identified, their impact on a particular worker's performance in a particular context made manifest and some means identified or developed for recognizing them. The Skills Matrix Project is not addressing itself to these extraneous factors. It is our suspicion, in fact, that SDP's may be most useful in diagnosing areas wherein a given counselor may need to improve skill, as opposed to being used in predicting a counselor's level of effectiveness.

.Finally, while <u>Process</u> <u>Mapping</u> can be seen as useful in evaluative efforts, it.is also a useful tool for use in areas such as:

-- Clinical Supervision

-- Curricula Development

-- Program, Planning

-- Documentation of Competence and Credentialing Workers

It is our hope that this model will be a step in the direction of identifying, assessing and improving competent work performance. More competent work performance, we hope, will lead to better services to the clients we serve. Your comment is welcome.

REFERENCES

Anastasi, A., Psychological Testing, Third Edition, New York: MacMillan Co., 1968.

Bandler, R. and Grinder, J., The Structure of Magic. Palo Alto, California: Science and Behavior Books, Inc. 1975

Barbee, D., "Task Analysis Data for Oregon Drug Counselors", 1976.

Carkhuff, R.R. and Berenson, B.G., Beyond Counseling and Therapy. New York: Holt, Rinehart, and Winston, 1967.

Cronbach, L.J., Essentials of Psychological Testing, Third Edition, New York: Harper and Row, 1970.

Fine, S. and Wiley, W.W., An Introduction to Functional Job Analysis, Kalamazoo, Michigan: Upjohn Institute for Employment Research, 1971.

Flanagan, J.C., "The Critical Incident Technique", Psychological Bulletin, 1954, 51, 327-358.

Gilpatrick, E., "Role Definitions and Specifications", in Proceedings of a National Conference for Evaluating Competence in the Health Professions, New York: sponsored by the Professional Examination Service, November, 1976.

Human Services Development Institute, "Substance Abuse Counselor Task Analysis Question-aire for Content Validation", Portland, Maine: University of Maine, May, 1977.

Kelley, J.M. and Siegfried, R.J., "Towards the Identification, Demonstration and Assess-ment of Drug Abuse Workers' Skills", paper presented at the National Drug Abuse Conference, San Francisco, May, 1977 by Skills Matrix Project, Medical College of Pennsylvania, Department of Psychiatry and Social Sciences.

Kelley, E. Lowell, Assessment of Human Characteristics. Belmont, California: Brooks/Cole Co., 1967.

Kibler, R., Barker, L.L., and Miles, D.T., Behavioral Objectives and Instruction. Boston: Allyn and Bacon Co., 1970.

Mager, R.F., Preparing Instructional Objectives, Second Edition. Belmont, California: Fearon Publishers, 1975.

Pottinger, P., "Concepts and Issues Related to the Identification, Measurement, and Validation of Competence", Institute for Competence Assessment, Boston, Mass.: McBer and Co., 1976.

Siegfried, R.J., "Process Mapping: A Technique for Program Improvement", pre-publication draft, Department of Psychiatry and Social Sciences, Medical College of Pennsylvania, presented at a Competence Symposium sponsored by the Southern Regional Education Board, May, 1977.

Skills Matrix Project, Medical College of Pennsylvania, Department of Psychiatry and
    Social Sciences, Administrative Report, January, 1977.

Thorndike, R.L. and Hagen, E., _Measurement and Evaluation in Psychology and Education_,
    Third Edution, New York: John Wiley and Sons, 1969.

Ziener, G.H., "Portfolios: An Alternative Approach to the Demonstration of Competencies",
    paper presented at the National Drug Abuse Conference, San Fransicso, Calif-
    ornia, May, 1977.