DOCUMENT RESUME

ED 072 108   TM 002 349

AUTHOR   Katz, Martin R.
TITLE   Evaluating Guidance--Why, What, and How.
INSTITUTION   Educational Testing Service, Princeton, N.J.
REPORT NO   ETS-RM-72-10
PUB DATE   Aug 72
NOTE   31p.; Paper presented at the Rutgers Guidance
              Conference (New Brunswick, New Jersey, October 20,
              1971)

EDRS PRICE   MF-$0.65 HC-$3.29
DESCRIPTORS   *Educational Accountability; Educational Counseling;
              Educational Guidance; *Evaluation Criteria;
              *Evaluation Methods; *Guidance Counseling; *Guidance
              Objectives; Guidance Programs; Occupational Guidance;
              Program Evaluation; Secondary Grades

ABSTRACT
              The 3 major questions about evaluating guidance are:
(1) Why should we evaluate? (2) What are we evaluating? (3) How do we
evaluate? School guidance counselors must first define their goals in
order to evaluate their performance and results. This is part of the
counselor's accountability to himself and to others. And by
communicating their objectives, the counselor can influence the
evaluation others make of his work. The success of a guidance program
is difficult to evaluate because defining and measuring behavior
objectives are not adequate for evaluation. One can raise scores on a
criterion measure without affecting the actual success of the program
being evaluated. Longitudinal evaluation studies are difficult, and
few have been conducted. Many variables and a considerable time-lag
are involved in identifying wise decisions. And the tendency to
generalize from results can be overdone. In real decision-making,
students do not simply choose from alternatives; they can often
create their own options. We can't define wisdom merely in terms of
outcomes. Students make decisions after they have examined competing
values and formed their own value systems. Without directing the
content of an individual's choice, we can help him in the process of
choosing. The major methods of evaluation are really inadequate
because they fail to take account of human differences and their
interactions with environmental circumstances. But we must, through
evaluation, provide students with a model of decision-making
behavior. (KM)

RM-72-10

RESEARCH
MEMORANDUM

EVALUATING GUIDANCE--WHY, WHAT, AND HOW

Martin R. Katz

Paper presented at the 28th Rutgers Guidance
Conference, New Brunswick, New Jersey, Wednesday,
October 20, 1971.

Educational Testing Service
Princeton, New Jersey
August 1972

EVALUATING GUIDANCE--WHY, WHAT, AND HOW

My talk today will deal with three major questions about evaluation of school guidance programs: Why should we evaluate them? What, exactly, are we evaluating; and, finally, How can we go about doing evaluations?

1. The first question, Why evaluate, is the easiest one to answer. One reason is to do our work better. Another reason is to convince others that our work is worth supporting. A third reason I will hold in suspension until after we have talked some more about why, what, and how.

The first reason--self-improvement--recognizes that a counselor is accountable to himself. Evaluation by the counselor in terms of his own standards, expectations, and concepts is a continuous feedback loop of the sort we all use to monitor most of our efforts and try to improve them. The emphasis in this evaluation for improvement is on processes and short-term products. At a very primitive level, the counselor watches what he is doing while he is doing it, makes some observations about immediate effects, and takes corrective action as indicated. There's no sense in waiting for long-term outcomes if you know you have to make a change now. For example, a speaker on a platform--like this--can sense whether anyone is listening to him or not. If not, he has to do something different--right away--talk louder or softer, speed up his delivery or slow it down, say something new or maybe something more familiar.

This is evaluation of micro-actions--before they have aggregated and agglutinated into a macro-program. It deals with the "necessary but not sufficient" conditions for success. If there is anything that can possibly be accomplished by speaking, it can't be accomplished unless someone is listening.

If a counselor announces office hours and sits back to wait for
students to come in, and no one comes in, then he knows it's time to try
something else. If he puts occupational information into a file, and no
one uses it, then he knows it's not doing any good, and a new approach is
required.

This mention of "something else" or a "new approach" suggests that
evaluation involves a choice between alternatives. It helps to have a
big pool of alternatives available. If there truly were only one way to
accomplish an objective, evaluation of process would be futile. Occasionally,
we keep on doing something--even though it doesn't work--because we can't
think of an alternative. Creativity in counselors may often take the form
of seeking and finding other ways to do things when our current way doesn't
work. Creativity may be fostered and stimulated as counselors make these
ongoing informal evaluations more systematic, more structured, more explicit.
My colleague, Henry Dyer, calls this kind of simple systematic investigation
'shirtsleeves research." It may consist of no more sophisticated data
collection than counting. For example, how many students used the occupa-
tional information file? Just formulating the question may be enough to
indicate what kind of data are needed and how they should be interpreted,
at least for this most primitive level of process evaluation.

Evaluation of process leads naturally to evaluation of product. If
the students are reading the occupational information material, what good
is it doing them? Are they learning something important and useful for
their career decision-making? If they are coming in to talk to the counselor,
what difference is the talk making? What contribution is each procedure or

facility making to some outcome, or product? To gain some objective, how
much time should a given student spend on each procedure? How much of the
talk or of the reading could you delete without affecting outcomes? John
Wanamaker, the department store merchant, once said, "I know that half the
money I spend on advertising is wasted. The trouble is I don't know which
half." Maybe half the time we spend talking to kids is wasted. (Runkel,
1962, did a study in Illinois high schools that showed no relationship
between such process variables as frequency of student-counselor talks
and such criteria as students' information about chosen occupation and
the appropriateness of curriculum choices to occupational choices.) Can
we devise studies that open up the "black box" of the counseling interview
and tease out the elements that are effective?--effective, that is for
which students under which circumstances in accomplishing which objectives?
We probably can't do this in "shirt sleeves"--we need to put on the research
specialist's coat for evaluations of that complexity.

But the first step in such evaluations--the step in which purposes are
stated--is one that the counselor can take and should certainly want to take.
Others may not agree with his goals--with what he says should be the product
of his work. But he has to spell out what he is trying to do if he is
going to be accountable to himself for outcomes. I am not saying that he
will readily find the opportunity to check out how well he accomplishes his
long-range goals, his ultimate product. But at least he has to have the
long-range objectives conceptualized in order to define intermediate and
short-range objectives that are logically aligned with them. In brief,
the improvement of processes implies that purposes and goals are known.

others may not necessarily agree with them—but the counselor should say
what his goals are.

This first step in the counselor's accountability to himself is also
the first step in his accountability to others. The more explicit he can
make his own objectives for guidance and his standards for judging accom-
plishment, the more clearly he can perceive the demands, expectations, and
standards of others to whom he is accountable—students, their parents,
administrators, other school staff, the community. Enlarging in this way
his consciousness of agreements and differences between his own concept
of his role and the concepts others have of it, he is better prepared to
negotiate with these others—to build on areas of agreement and to try to
reconcile differences, or at least increase understanding and tolerance of
differences on all sides. By knowing and communicating his objectives, he
can influence the nature of the accounting system which others may use in
evaluating his work. He can help direct the traffic, not just stand there
and maybe get run over.

In the speech he would have delivered this morning, Dr. Allen used a
different metaphor to express his high hopes for public accountability. He
called public accountability "the most promising cure for many of education's
most serious ills." He warned, however, that the public is becoming
"sophisticated and able to detect any attempts to substitute more of the
same old brew in new bottles." This expression brings to mind an episode
described by my wife on her return from the weekly shopping she does every
day at the supermarket. In the parking lot, she saw a woman she knew to
be pregnant suddenly slump over the steering wheel. Fearing an "emergency,"

she ran to offer help and found the woman doubled up not in labor but in laughter. It seems she was en route to visit her obstetrician, who had told her to bring a urine specimen. The only container available at home had been an empty whiskey bottle. While she was in the store, some one had stolen her whiskey bottle. Our moral is that the new bottle-labeled accountability will not fool many people for very long, if the contents are the same old bleep--which has so frequently been used in evaluations-- counselor-student ratios, or hours of graduate study completed by counselors, or size of the occupational library.

Going from scatology to eschatology, we must expect--as Dr. Allen has warned--that public evaluative judgments will be made of guidance programs. Since the beginnings of NDEA, guidance has enjoyed a favored status. Under NDEA support, guidance programs were established at many schools that had previously had none. But after the mid-60's, NDEA support fell off, and the burden fell heavier on local school districts. In his recent book Eli Ginzberg (1972, p. 305) recommends cutting off all mandated federal aid to guidance. He urges that the issue of support for guidance programs be decided "not in the halls of Congress but closer to home." In other words, he would put guidance needs in the pit with other educational needs. The magic claimed for accountability is that in Lessinger's words, 'resources and efforts are related to results in ways that are useful for policy-making, resource allocation, or compensation." Thus the decision- makers at federal and local levels want to examine cost-effectiveness so that they can make decisions about deployment of resources. The present commissioner of education, Dr. Marland, has recently made a commitment to

support model career development programs with a strong guidance component--
implemented by a $9 million allocation for 1972. His directive requires emphasis
on "careful measurement of student outcomes in relation to the treatments."
It also requires cost information on each component. Finally, it calls for
"third party evaluation." So we see that even the supporters of guidance
do not exempt guidance programs from judgments. These judgments, however,
are lower case and plural. They should not be mistaken for the Judgment Day,
when presumably the purpose of the evaluation will be perfectly clear, the
criteria sharply defined, and the measures absolutely reliable and valid.
The present-day judgments, in contrast, will be fallible: we see no clear
consensus on purposes, there are sharp disagreements on fuzzy criteria, and
measures that have been developed so far appear to have validities that are,
at best, indeterminate or modest.

II. This brings us to the question of what we are trying to evaluate.
There seems little prospect in the immediate future of convincing the public--
or even yourself--that any of the following direct questions can be answered
definitively: Does guidance work? Does it achieve its goals? How well is
the guidance program at your school doing? Are children getting good
guidance? What difference is it making in their careers? Are the programs
worth what they cost? Is the money they cost being used efficiently? Should
the guidance programs continue to do what they are doing?

Tumin (1970) has called evaluative questions like these the "fool's
questions"--"because they are absolutely right to ask and impossible to
answer as put." These are the big questions that research and evaluation
studies have never been able to answer. At least not unless one fragments

each of these questions into subquestions, defines each fragment in operational terms, samples from the new sets of questions that are thus generated, and identifies relevant observations or measures with the expectation that enough such observations can be combined to represent a facet of each little question, and that enough answers to little questions eventually allow us to assemble some kind of inference about one of the fragments of a big question-- and so on.

Let's take an illustration. We ask a big question, Are high school students getting good guidance? Let's define a subquestion: Are they making their career decisions wisely? This subquestion must be sliced up into smaller and smaller questions before we can begin to answer it. Recent studies have attempted to elaborate a construct called "vocational maturity," and ask whether students have gained in vocational maturity. One indicator of vocational maturity might be, are they seeking occupational information? One of many ways in which they might seek occupational information is through reading printed materials in the occupational information library or files. Aha, now we have something we can observe or measure. We can count the uses made of these materials, we can ask students what use they make of them, we can test students on the information contained in them. Does this kind of observation or measure tell us whether they are making career decisions wisely, and whether they are getting good guidance? How many little questions like this must we answer in order to make an inference about the big question. "Are they making career decisions wisely?" or "Are they getting good guidance?"

Am I lacking in the reverence that is usually given by evaluators to "behavioral objectives"? Do I imply that defining and measuring behavioral

objectives is not adequate for evaluation? Just so. Focusing exclusively

on behavioral objectives can lure us into rationalizing the inclusion of

behaviors just because they are easy to measure.[1] Often the use of such

behaviors and their measures in evaluation tends to impoverish rather than

to enrich practice. Teaching to the test makes us lose sight of the big

question, the "fool's question." Guidance is not the only field in which this

problem occurs. Even the "hard curriculum" areas face it. For example,

Sheldon Myers has criticized current statements of behavioral objectives

for mathematics in elementary grades on the basis of their "great specificity.

The unfortunate consequence of this atomization is that interrelatedness of

mathematical concepts is lost and the statement is a tedious list of very

trivial low-level skills [Myers, 1970]."

Lee Cronbach has pointed out that specific behaviors can and should be

used as indicators of constructs but not as the definers of those constructs.

It is the constructs, the network of relations or characteristics, that are

crucial to evaluation--not a single specific incident of behavior. "The

operationists who want to equate each construct with one indicator," he says,

"...are advocating that we restrict descriptions to statements of tasks

performed or behavior exhibited and are rejecting construct interpretations....

The writers on curriculum and evaluation who insist that objectives be defined

in terms of behavior...are denying the appropriateness and usefulness of

constructs [Cronbach, 1969]."

Let's point this problem up by assuming that you are working under a

performance contract. You are to be paid according to the "results" you get.

Now how are results to be measured? You name one objective of guidance as

helping students make career decisions wisely. You invoke the construct of

vocational maturity. You assume that information plays a role in this. You

may reason, as I wrote some years ago:

> Decision-making . . . may be regarded as a strategy for
> acquiring and processing information. If a decision is truly
> to be made, if it is not a foregone conclusion, it must involve
> some novel elements. The person confronted with the problem of
> decision-making either does not know what information he needs,
> does not have what information he wants, or cannot use what
> information he has. Thus, the pressure for making a decision
> creates a discrepancy between the individual's present state
> of knowledge [or wisdom] and the state that is being demanded
> of him.

> The role of guidance should be to reduce the discrepancy
> between a student's untutored readiness for rational behavior
> and some hypothetical ideal state of knowledge and wisdom.
> So the appropriate criteria for a given program designed to
> retail information might be: (1) Do students know what infor-
> mation they need? (2) Can they get the information they want?
> (3) Can they use the information they have? [Katz, 1966].

But when all this language gets translated into specific measurable

behaviors for a performance contract, the contract may call for a question-

naire to be given students on the extent to which they use occupational

information materials, or a count of such uses, or a test of knowledge of

facts about occupations. Would you as the contractor then attempt to develop

in students a general competency in the strategy of information-processing?

Or would you--as the Texarkana contractors are alleged to have done--find

a more direct route to raising scores on the criterion measure? After all,

students can be induced in many ways to take materials out of a library, or

to respond in a certain way on a questionnaire, or even to memorize some

facts. They would not need a "guidance program" for this--just, if we

wanted to be crass about it, a little coaching. One can raise scores on such

criterion measures without affecting the outcome that is of real concern.

Such an increase in scores would be no more valuable than, in Thorndike's phrase, boiling the thermometer to heat the house.

The ripple effect of studies that use such measures of specific behaviors is another problem. By the time the study report gets cited in the literature, the specific behaviors and measures that underlie the findings are often forgot . A verbal summary of the conclusions is quoted and requoted: "This treatment significantly increased information-seeking behavior of students and thereby contributed to an improvement of wisdom in decision-making and a gain in vocational maturity." The indicator has now become a definer. The network of lines from specific measures to constructs has been short-circuited.

So the question what to measure leaves us in a dilemma. On the one hand we don't want to swamp our evaluative enterprise with meaningless rhetoric about goals that give us no clue to measurement of progress. On the other hand, we don't want to limit our observations to trivial and low-level behaviors that are directly coachable under such conditions as performance contracting.

So where, we must ask, is the middle ground between what Tumin calls "trivial precision and apparently rich ambiguity"? Let's see whether we can find it in any of the criteria that have been kicking around for some years.

First, we must face the problem of long-range vs. short-range criteria. Unfortunately, this has been a very slippery problem. Like a fussy fisherman who cannot eat what he can catch and cannot catch what he could eat, the would-be evaluator has found angling for data on long-range outcomes overtaxes his patience and resources, while the short-term data that are easily netted

often lack nourishment or flavor and may well be thrown back. The ultimate
criteria for judging effectiveness of a full-scale vocational guidance program
have been elusive. What many want to know is: Does guidance make a differ-
ence in people's careers? What kind of occupational success, adjustment,
and satisfaction do they achieve? What contributions do they make to society?
To fish for answers to such questions takes time, money, and control of many
variables.

Precious few have even tried to conduct longitudinal evaluation studies
ranging over a period of years. Rothney's (1963) follow-up of experimental
and control groups beyond high school is a notable exception. He used many
criteria, such as amount of post-secondary education, achievement in college,
promotions in jobs, satisfaction with current status and with intervening
decisions and actions. (In general, differences between the experimental
and control groups were small and not significant. But even if there had
been significant differences, would the time-lag and changing conditions
permit assurance that the same treatment would have equally favorable out-
comes today?) At any rate, most evaluators of guidance, like those who
evaluate other areas of the school curriculum, settle for the kind of
criteria they can net more readily. A comprehensive search of their creels
over the last 35 years discloses, most commonly, such criteria as student
satisfaction with counseling; persistence in school; comparisons of students'
self-ratings with test scores; judges' ratings of "realism" or "appropriate-
ness" of "preferred occupations" named by students; the proportion of a
class expressing an occupational goal; the constancy of expressed occupational
choice over a period of time (say, from ninth to twelfth grade); the relation-
ship between proportion of a high school class expressing preference for

each occupation and the latest census count showing proportion of working force in each occupation in the community; expressions of counselee satisfaction; improvement in counselees' school marks; etc. (Incidentally, guidance has rarely made a significant difference in these variables. There is no clear reason why it should.)

Notwithstanding consistent negative results, these criteria may have had some utility for the objectives of guidance that were widely accepted up to about 1950. The increasing acceptance of recent developments in guidance theory, however, has made the digestibility of such criteria increasingly dubious. Today, such data seem hardly worth pulling from the stream; the would-be evaluator must find other fish to fry. It is evident that the construct represented by all these long-range and short-range criterion variables was whether students had learned to make wise decisions. That is, were the outcomes better for the experimental group than for the control group?

But to evaluate the long-term outcomes of decisions is not only difficult: it is presumptuous.[2] Tennyson wrote, "No man can be more wise than destiny." I would feel more comfortable if we changed the criterion from "Making Wise Decisions" to "Making Decisions Wisely." This shifts the emphasis from content to process. "Wise decisions" implies an understanding of outcomes and a mastery over events to which we cannot aspire. "Making decisions wisely," on the other hand, implies an understanding of self and a mastery over processes which may be more attainable. It is in this sense of wisdom that Tennyson is contradicted by the old Latin motto ("Fato prudentia major") "Wisdom is stronger than fate."

Suppose you were counseling students in the late 50's or early 60's, and heeded the goal supported by Congress, under NDEA, to identify able students and encourage them to continue with their education and prepare for certain high-level occupations. Of course, NDEA owed its existence partly to the shock of sputnik--so you might feel particularly effective if, with your guidance, one of your brightest and ablest students decided to become an aerospace engineer. How gratifying for you to have done your duty by Congress and your profession! But now your former student is unemployed. Was his decision a wise one? Was your guidance good?

The problem in identifying wise decisions, however, is not _just_ the time-lag between the choice-point and one judgment day--the day when all the evidence on consequences of the choice is in. Nor is it just a matter of insufficient predictive validity. Predictive data are really historical data, and our predictions are manifestations of what we have learned from history. Thus, if our predictors had perfect validity, we could extend the aphorism "Those who do not learn from history are condemned to repeat it," by adding "_and_ those who _do_ learn from history are also condemned to repeat it." But in fact we don't repeat history, even when events materialize as we have predicted. For there is always a surplus of events--there are more events than predictions. The outcomes of decisions exceed the purposes of decision-makers. Any decision that is not trivial has ramifications without end. Each outcome then may generate new purposes and new decisions, leading in turn to more outcomes, and so on ad infinitum. Thus the original purposes and predictions may be buried under this landslide of outcomes and decisions and outcomes.

Consider, as a somewhat painful example, the decision of the U. S. govern-
ment to intervene in Vietnam. One could argue, and indeed the government has
argued, that this was a wise decision in that the purposes of this decision
were (and are being) fulfilled as predicted. But surely the government does
not maintain that all the outcomes of that decision were predicted, and it
has built no granaries for storing the surplus events until such time as we
need them--or at least are better able to cope with them. As the Pentagon
papers have made clear, the fault in the decision to intervene in Vietnam
was in the process, not just in the outcome. Suppose the outcome had been
somewhat different: suppose we had had a great military success there--had
"brought the coonskin home and nailed it to the wall." Would that military
success have wiped the slate clean of the flaws in decision-making? Would
it have justified our decision? Perhaps it would have prevented the moral
questions from being raised--as when we intervened in the Dominican Republic--
although it is unlikely that we could have "won" in Vietnam that fast, or
with less publicity and condemnation than the Russian interventions in Hungary
and Czechoslovakia. At least a few voices--voices like Jim Allen's--would
have cried out in the wilderness about the moral issues. But a victorious
outcome would have prevented widespread popular concern. My Lai and tiger
cages and one-man election races would never have plagued us, and the whole
incident would have soon blown over in the media and the public consciousness.
Would that military success have made the decision a wise one? Would the
decision have been made any more wisely?

For the sake of argument, let us suppose that we have predicted and can
evaluate the ramified outcomes of this decision to intervene as in some sense
superior to those which would have been produced by any alternative decision.

Even then, what would the substantive payoff of this decision reinforce?
The content of this decision itself? But this same decision is not likely
to come up again. We only pass this way once. Then we would be hard put
to claim an increment in wisdom from the content of this decision. The
content of a single wise decision is not likely to be transferable to the
next decision, and the next.

In fact, what one learns from the multitude of real-life outcomes may
or may not be relevant to wisdom. Like Mark Twain's cat, who learned from
sitting on a hot stove never to sit on any stove again, we may learn from
these outcomes more "wisdom" than is in them. For example, the current
overflow of outcomes from the Vietnam decision might teach us to revert to
isolationism (in contradiction to the "lessons" from previous decisions and
outcomes). The little boy who is spanked for turning the faucets on full
blast and flooding the bathroom may learn not to wash his hands and face.

It is these tendencies to "generalize" that lead the behaviorists to
concern themselves with what Skinner calls "contingencies" in their schedules
of reinforcement. Or as O. H. Mowrer once put it, in a classroom discussion
of one of his learning experiments, "You've got to be smarter than the rat."
Well said, since such an approach to defining wisdom in terms of outcomes
requires that wisdom reside in the experimenter--or counselor, not in the
subject--or student. But this is where the presumption comes in: do we
as counselors know which decisions are wise?

Here one may object, are there not "universally desired" outcomes that
represent a cultural consensus or folk wisdom for which the counselor may
serve as spokesman? Let us grant this, while noting that we may retain some

squeamishness about our ability to identify such universals even in retro-
spect, let alone in advance. Presumably, we can teach students to make
these decisions that lead--with a high degree of probability and low risk--
to universally desired outcomes.

But when we have identified such universals and induced students to
learn them, we are not really concerned with decision-making--or with guid-
ance. Then we are concerned with indoctrination. A large part of an
individual's schooling consists of such indoctrination. The distinctive
concern of guidance, however, is not with the universals, but with the
"alternatives"--toward which the culture tends to be more permissive.

However, I must express some dissatisfaction with the term "alternatives."
The individual is not always constrained to choose from clearly shaped alterna-
tives that are already "there" like the options in a multiple-choice test.
He often has some opportunity to construct, or create, his own options--in
the sense that the poet creates his verses, perhaps creates alternative verses,
before choosing the ones he wants. He is not merely choosing alternatives
from his total vocabulary, any more than the painter is merely choosing colors
and lines from an existing pool of options. He does not find his new and
unique combinations, variations, and transformations by considering all
possible permutations. Fifty chimpanzees typing for fifty years might
compose the complete works of Shakespeare, but they wouldn't know how to
write a new work of similar quality. In terms of content and outcomes, they
might have made "wise" decisions, and yet they would be none the wiser. As
critics, we can evaluate the poet's decisions, recognize them as creative,
or wise, and teach someone to memorize them. We can even derive and apply

rules for transfer of content. For example, we can analyze a line like "Now is the winter of our discontent" and recognize an association between emotion or state of mind and a season of the year, in which season is used to represent feeling. No doubt, a computer could be programmed to ring the changes on this kind of association, with such results as "Now is the summer of my happiness," "Now is the spring of my joy," "Now is the autumn of my melancholy," etc. ad nauseam. But could it ever make the long leap from this last to reach "my way of life is fallen into the sere, the yellow leaf ..."? This illustrates, I think, the gap between recognition of a creative, or wise, decision and the ability to make one. How often the best and wisest decision is not to choose between historically "given" alternatives, but to construct a new option. Like able students who squirm at being forced to choose the best of five bad options on a multiple-choice test question, our wisest decision-makers can sometimes think of a better response than any given.

I hope that all this suggests an "alternative" to defining wisdom in terms of outcomes. How a choice comes out, and even how one chooses between alternatives, may be less important than how one constructs alternatives. In this view, wisdom derives not from the outcome of a decision but from the process of decision-making. And our greatest folk-wisdom, our most compelling "universal," may apply most directly to the process of constructing and choosing alternatives.

For may we not regard democracy itself as an evolving process of decision-making? It is its processes, not the content of any one policy decision, that make it distinctive.

We recognize, as a crucial characteristic for the processes by which we ideally make national policy decisions, that our society is pluralistic. On

every issue competing interests and pressure groups are heard. Sometimes they differ on predictions of outcomes--for example, the effects of a tax increase on the economy. More often, and more significantly, they have different definitions of desirability, different objectives, even when they agree on predictions of outcomes. How do these differences get resolved "wisely"? The necessary condition, we believe, is freedom--the open marketplace of ideas, in which every voice can be heard and judged. Out of this confrontation of competing values, the legislative or executive can find--or claim to find--a consensus for decision, to be translated into a mandate for action. But it does not stop there. The process is ongoing, permitting revision of content in accordance not just with outcomes, but also with changes in values and objectives. This provision for change, this ability to accommodate to new situations and circumstances, has perhaps insured the survival of democracy, up till now, through many vicissitudes. (Our ability to reverse our decision on Vietnam is a sign of strength, not of weakness.)

Need I belabor the analogy with individual decision-making? The individual too recognizes that ne must choose between competing values. How then does he make order out of the rabble of impulses that beset him? They should be neither suppressed nor blindly obeyed, but brought under the rule of reason, each given "equal time" and attention. The individual, like the nation, must hold himself open and receptive to different values, allowing each to speak to him as loudly as the others. This process involves active and systematic examination and exploration of competing values.

One way in which he can examine values is to study their sources. Here we see a nice articulation of education and guidance. If a major purpose of

education is to transmit the culture, an important purpose of guidance is to help the individual come to terms with the culture--that is, the choices he makes will indicate how he sees himself in the culture. But first he must see the culture in himself. So his first question must be, where have my values come from? Then he will be better prepared to ask, where are they taking me?

When the student has taken full cognizance of the range of values in the culture, and has formulited his own value system quite explicitly, he will be ready to lay his values on the line in making a decision. The specifics of a strategy for accomplishing this I have described elsewhere and will not have time to go into now. But I want to emphasize that with the individual, as with the nation, decision-making should be an ongoing process, subject to continual revision. Otherwise, he may run afoul of the warning that "tne only thing worse than not getting what you want is getting it."

In shunning a definition of wise decisions in terms of content, or predicted outcome, I have assumed that experience does not teach us what will be best for the individual (or society) except freedom to work things out. Thus, I have defined the best choice as the choice that is most nearly free. But I do not define freedom as complete laisser-faire. Rather, it is the freedom (expressed by Shaw in the preface to Man and Superman and quoted by Freud in contrasting his "reality principle" with his "pleasure principle") "to be able to choose the line of greatest advantage instead of yielding in the path of least resistance." So without directing the content of an individual's choice, we do think we can help him in the process

of choosing. This emphasis on process does not pretend to insure the "right" choice--except insofar as the right choice is defined as an informed and rational choice. Our bias--our conviction--is that in education enlightened processes are intrinsically important. Therefore, we bend our efforts to increase the student's understanding of the factors involved in choice (imperfect though our own understanding may be) so that he can take responsibility for his own decision-making, examine himself and explore his options in a systematic and comprehensive way, take purposeful action in testing hypotheses about himself in various situations, and exercise flexibility in devising alternate plans.

In short, we don't want to play the decision-making game for him. We want to help him master the strategies for rational behavior in the face of uncertainty (which may be the nearest he can get to wisdom) so that he can play the game effectively himself.

Horace, in one of his satires, asked "Who then is free?" and answered "The wise man who can govern himself."

Let me make "free" with Horace, and interchange the descriptors, to ask, "Who then is wise?" and answer "The man who can govern himself freely."

III. So now at last we move on to the question of how evaluations can be made. In an interesting paper, Hartnett (1971) has pointed out some of the weaknesses of the classic model of evaluation, which involves such elements as (1) behaviorally defined objectives, (2) the random assignment of subjects to treatments, (3) clearly differentiated treatments, and (4) criterion measures chosen or developed on the basis of the behavioral objectives. He suggests that dissatisfactions with this model are leading

to two important changes: a concern for the consequences, not just the
objectives, of a treatment, and a style of inquiry which is exploratory in
nature rather than attempting to apply in life situations the kinds of controls
and manipulations that are feasible only in the scientific laboratory.

Pace (1969) has typified the new models in this way: "The spirit of
the evaluator should be adventurous. If only that which could be controlled
or focused were evaluated, then a great many important educational and social
developments would never be evaluated...that would be a pity."

In guidance, this exploratory set must be emphasized. We have no neat
evaluation packages all wrapped up and ready to use. For example, a number
of people have developed what purport to be measures of "Vocational Maturity."
Can any of these measures be recommended for use?

One of the best known measures is John Crites' VDI, an inventory keyed to
the responses of 12th-graders. Extensive research has been done on this
instrument--for example, on elimination of variance attributable to acquiescent
response set (Crites, 1971). Yet the instrument has been criticized on just
these grounds: Vocational Maturity, as defined by VDI, means saying no.
(There we see it again, the instrument taken as the definer rather than the
indicator of a construct.) Another criticism involves the use of 12th-graders'
responses as the keyed responses: a group of 10 counselor educators and
vocational psychologists disagreed with the keys for a number of items.

Back in the 1950's, I developed an objective test that I am not particularly
proud of. It attempted to find out whether students had mastered certain
concepts involved in self-appraisal, getting and using information, and
decision-making (Shimberg & Katz, 1962). At the same time, and in connection

with the same project, we commissioned Warren Gribbons to develop an interview schedule, known as Readiness for Vocational Planning, to see whether students were actually applying those concepts to their own educational and occupational decisions (Gribbons, 1960). We were evaluating a work text for group guidance, and found highly significant differences between experimental and control groups--for example, experimental students scored very significantly higher on the test and also showed very significantly greater awareness of their own values, better ability to define their values and to describe the role their values play in their decision-making, and so on. A group of professionals in guidance, listening to tapes of the interviews without knowledge of the scales or scores, ranked the students in the same order on "vocational maturity" as the total scores did. Gribbons and Lohnes have now converted the interview schedule into a questionnaire form called Readiness for Career Planning.

Super and his colleagues have recently developed a Career Questionnaire that also purports to measure vocational maturity. It includes scales called Concern with Choice, Acceptance of Responsibility, Occupational Information, Work Experiences, Crystallization of Interests, and so on--rubrics derived from the Career Pattern Study.

Westbrook has been developing a series of Vocational Maturity Tests, including some of the items from my old test. The items tap various kinds of information, Course and Curriculum Selection, Planning, Goal Selection, etc.

These are the major standardized efforts I know of to get at the construct, vocational maturity, and they are all well conceived; they are good

tries. I am not damning them with faint praise. I just want to forewarn
you that you may be disappointed when you see the actual instruments and
study them item by item. You will agree, I am sure, that even though they
may be indicators of vocational maturity, they are not definers of it.

The questions getting at facts about specific occupations hardly seem
appropriate for students who may have had no interest whatsoever in those
occupations. Then, too, a number of the items depend on occupational prefer-
ences expressed by the students--for example, Super is concerned with "Wisdom
of the Vocational Preference" and with "Consistency of Preference."

The title of an occupation, however, is probably a poor indicator of
what choosing an occupation means to an individual. More relevant questions
might be, In his view, how important an element of his life is represented
by occupation? What kinds and amounts of satisfaction does he hope to derive
from it? What differentiations does he discern between occupations in
capability of providing such satisfactions? How much control over his choice
and responsibility for his choice does he appear to exercise? What role to
predictive data play in his choosing?--does he consider them? is he dominated
by them? What risks is he willing to take to achieve the occupational
satisfactions he says he wants? What decision rules does he employ? What
resources does he use? What reality tests of his perceptions and predictions
has he made, or does he plan to make? How has he coped--how will he cope--
with obstacles and difficulties? Has he formulated viable alternative plans?
How explicit and consistent is his reasoning about these questions?

Once we have probed beneath the surface of choice to get at such underly-
ing perceptions, attitudes, and rationales, we may find ourselves with much

richer criteria of growth and vocational development. Dr. Bingham's efforts
to get at the dimensions along which individuals construe occupations--using
an adaptation of Kelly's Role Concept Repertory test--is a step in this
direction. Some of my associates and I have developed and used, in an
exploratory way, interview schedules to get at students' occupational
constructs (Katz, Norris, & Kirsh, 1969). Examples of some of the more
productive questions we asked were:

> Now sit back and turn your imagination loose. Try to describe,
> as fully as you can, what you would regard as an ideal or
> "dream" occupation. It can be a real occupation, or one you
> invent.

> In view of what you've said about an ideal occupation, why
> didn't you decide to become a _____ instead
> of (preferred occupation choice)?

> Now reverse your field and think of the worst occupation
> you can. If the other was a "dream," this would be a
> "nightmare." Describe it.

Of course the interview itself had its effects. One probably cannot measure
the status of an individual's decision-making without influencing it. For
instance, at the end of interviews with junior college students we got
comments like this: the interview "extended my ideas about what to look
for in an occupation," "made me think about why I was making my choice,"
and so on. For example, it seemed to have a particularly strong impact on
one student who had appeared especially firm and specific in his plan to
become a chemical engineer. Working as a draftsman after his graduation
from high school (where he said he had been "pushed into" a vocational
curriculum by his guidance counselor), he had had a particularly good
opportunity to observe chemical engineers at work and had an unusually
thorough knowledge of their work activities. His perceptions (in the

comparisons of occupations) seemed fixed almost exclusively on one construct:

whether an occupation offered an outlet for scientific interest and inventive-

ness, or not. The sole deviation involved a discrimination between occupations

in terms of altruism--opportunity to help others. The systematic exploration

and examination that accompanied his scaling of values brought out more

explicit recognition of Altruism as a value of some importance to him. With

this discovery, other values of which he had not been fully aware also came

into focus as quite important to him: notably Variety and Autonomy. At

the end he said that the interview had "brought to the surface values I've

held but never recognized. That shakes me. ...If I had two lives to lead,

for one of them I'd go into the Peace Corps as soon as I finished college.

Maybe then I'd try to become a high school teacher or counselor, or a

community worker. But I came up the hard way. There are things I see now I

want to do, but I can't do them until I get firm ground under me. I'm still

determined to become a chemical engineer. Not like a machine, though, but

like a person."

If you can't measure a condition without changing it, does that mean you

should not try to measure it? No, not even if it is a differential influence,

affecting different students in different ways. After all, people encounter

many common experiences that have differential effects, and this attempt at

measurement is only one of such an unknown number. The differential effect

may indeed be part of the substance of what we are trying to investigate.

Samuel Messick has pointed out that traditional questions in education and

psychology have frequently spawned answers that are either downright wrong,

in that they summarize findings "on the average" in situations where a

hypothetical "average person" simply doesn't exist, or else are seriously

lacking in generality, in that they fail to take account of the multiplicity

of human differences and their interactions with environmental circumstances.

An example is the "horse race" question typical of much educational

research of past decades: Is treatment A better than treatment B? Such

questions are usually resolved by comparing average gains in achievement

for students receiving treatment A with average gains for students receiving

treatment B. But suppose treatment A is better for certain kinds of students

and treatment B better for other kinds of students? A completely different

evaluation of the treatments might result if some other, more complicated

questions had been asked, such as "Do these treatments interact with differ-

ences in personality and cognitive characteristics of students--or with

differences in their educational history, or family background, or community,

or culture--to produce differential effects upon achievement?"

Hard upon this warning of the complexity of evaluation in guidance,

let me quote again from Henry Dyer (1970):

> The term educational accountability, as used most recently
> by certain economists, systems analysts, and the like, has
> frequently been based on a conceptualization that tends, by
> analogy, to equate the educational process with the type of
> engineering process that applied to industrial production....
> It must be constantly kept in mind that the educational process
> is not on all fours with an industrial process; it is a social
> process in which human beings are continually interacting with
> other human beings in ways that are imperfectly measurable or
> predictable. Education does not deal with inert raw materials,
> but with living minds that are instinctively concerned first
> with preserving their own integrity and second with reaching
> a meaningful accommodation with the world around them. The
> output of the educational process is never a "finished product"
> whose characteristics can be rigorously specified in advance;
> it is an individual who is sufficiently aware of his own in-
> completeness to make him want to keep on growing and learning
> and trying to solve the riddle of his own existence in a world
> that neither he nor anyone else can fully understand or predict.

Despite these problems, evaluate we must. And so I come back, in conclusion, to my third reason for why we evaluate.

My third reason for evaluation, despite all its snarls and pitfalls, is simply this. If we believe in trying to help students make career decisions wisely--that is, make rational and informed decisions--then we must also, in all honesty, believe that guidance practitioners should make their professional decisions wisely. We have to provide students with a model for decision-making behavior--and that is just what an evaluation process is. It is a commitment to use of information and reason, to rational behavior under conditions of uncertainty. So--like the students--we must take responsibility for evaluation. We must make our professional values explicit, examine and explore them. We must formulate hypotheses about the effects of our activities, and try to get feedback. We must revise our hypotheses, plans, and activities in the light of new information.

When we evaluate, we commit ourselves to a continuous process of decision-making. It is a commitment we should welcome. The methods and the product may leave much to be desired. But let us realize that commitment to the process itself may be a powerful indicator of how good a school guidance program is.

References

Crites, J. The maturity of vocational attitudes in adolescence. Washington, D. C.: American Personnel and Guidance Association, 1971.

Cronbach, L. Validation of educational measures. In Proceedings of the 1969 Invitational Conference on Testing Problems. Princeton, N. J.: Educational Testing Service, 1969.

Dyer, H. Toward objective criteria of professional accountability in the schools of New York City. Phi Delta Kappan, 1970, 52(4), 206-211.

Ginzberg, E. Career guidance. New York: McGraw-Hill, 1972.

Gribbons, W. Evaluation of an eighth-grade group guidance program. The Personnel and Guidance Journal, 1960, 38, 740-745.

Hartnett, R. Accountability in higher education: A consideration of some of the problems of assessing college impacts. New York: College Entrance Examination Board, 1971.

Katz, M. Criteria for evaluation of guidance. In A. Martin (Ed.), Occupational information and vocational guidance for noncollege youth. Pittsburgh: University of Pittsburgh, 1966.

Katz, M. Learning to make wise decisions. Research Memorandum 68-14. Princeton, N. J.: Educational Testing Service, 1968.

Katz, M., Norris, L., & Kirsh, E. Development of a structured interview to explore vocational decision-making. Research Memorandum 69-3. Princeton, N. J.: Educational Testing Service, 1969.

Myers, S. Comments on behavioral objectives in education. Memorandum for the record. Princeton, N. J.: Educational Testing Service, November 1970.

Pace, C. R. An evaluation of higher education: Plans and perspectives.

   CSE Report No. 51, Center for the Study of Evaluation. Los Angeles:

   UCLA Graduate School of Education, January 1969. (Mimeographed.)

Rothney, J. Educational, vocational, and social performances of counseled

   and noncounseled youth ten years after high school. Madison: University

   of Wisconsin Press, 1963.

Runkel, P. The effectiveness of guidance in today's schools: A survey

   in Illinois. Unpublished report, 1962.

Shimberg, B., & Katz, M. Evaluation of a guidance text. Personnel and

   Guidance Journal, 1962, 41, 126-132.

Tumin, M. Evaluation of effectiveness of education: Some problems and

   prospects. Interchange, 1970, 1(3).

Footnotes

[1]This discussion of the disadvantages of exclusive reliance on "behavioral objectives" is indebted to Rodney Hartnett's (1971) recent publication, <u>Accountability in Higher Education</u>.

[2]This section on wisdom in career decision-making is derived from an earlier paper (Katz, 1968).