

Measuring Instructional Results

or Got a Match?

Second Edition

Robert F. Mager

The bottom of the cover features a series of horizontal stripes in dark blue and black, creating a decorative border.

Measuring Instructional Results

or Got a Match?

Second Edition

Robert F. Mager

Lake Publishing Company
Belmont, California

BOOKS BY ROBERT F. MAGER

Preparing Instructional Objectives, *Revised Second Edition*

Measuring Instructional Results, *Second Edition*

Analyzing Performance Problems, *Second Edition*
(with Peter Pipe)

Goal Analysis, *Second Edition*

Developing Attitude Toward Learning, *Second Edition*

Making Instruction Work

Developing Vocational Instruction
(with Kenneth Beach)

Troubleshooting the Troubleshooting Course

Copyright © 1984 by Lake Publishing Company, 500
Harbor Boulevard, Belmont, California 94002. All rights
reserved. No part of this book may be reproduced by any
means, transmitted, or translated into a machine language
without written permission from the publisher.
ISBN-1-56103-340-5

Library of Congress Catalog Card Number: 83-60502
Printed in the United States of America

Preface

Once upon a time, as the crow flies, the king of Hairmania decided to shave off his beard.

“It is an event that will bring attention and fame—not to mention tourists,” he beamed. “Bring the Royal Barber.”

“But sire,” lamented his advisor, “there isn’t one. No one has shaved for a hundred years.”

“Hairesy!” exploded the king. “No wonder we’re so crowded. Sally ye forth, therefore, and find me the best in all the land.”

Which he did. And when at last the most famous barber was found, he was sent to the Royal Three Committees for the Royal Testing.

“Tell us about the history of barbering,” asked the first committee.

And he did.

“Tell us about the importance of barbering,” asked the second committee.

And he did.

“Tell us what instruments you would use to shave the king,” asked the third committee.

And he did.

Whereupon they draped his neck with their Medallion of Approval and led him before the king. Wasting no time, the barber prepared his tools and spread his cloth. But when he picked up his razor with a swirling flourish—he accidentally sliced a piece off the king’s ear.

“Gadzooks!” cried the king. “You’ve cut off my royal ear!”

“Ooops,” chorused the nine voices of the Royal Three Committees.

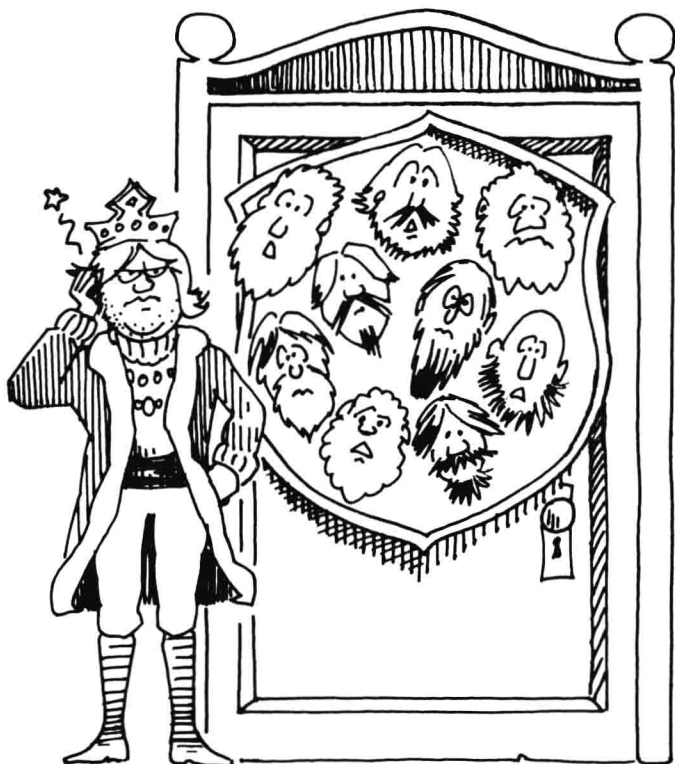
“Oops?” astonished the king. “I ask for skill and you give me oops?”

“We’re very sorry,” apologized the Royal Three Committees. “We must have lost our heads.”

“A capital idea,” rejoiced the king, and sprang himself forth to make it permanently so.

And ever since and forever more,
There hang nine heads on the Royal Door.
For this was the fate of the Committees Three . . .
May it never befall such as me . . . or thee.

And the moral of this fable is
HE WHO ASKS WRONG QUESTIONS MAY LOSE
MORE THAN FACE.



When it is important that one's instruction be successful, it would also seem important to bestir oneself to determine whether one has succeeded as intended. The measurement of instructional success is accomplished mainly through the development of test items (situations, performance items, simulations, role-play, paper-pencil, etc.) that precisely match each objective in scope and intent. This book is designed for those who want to know how well their instruction works and how to develop the basic tools with which to measure instructional results. It shows how to recognize or prepare test items through which one can determine whether an instructional objective has been achieved.

We do not weigh steam with a yardstick or evaluate music with a bathroom scale. Such measures would be irrelevant to the things being measured. Similarly, we do not measure achievement of objectives with test items that do not match those objectives. If you can identify a test item that is right for testing an objective, you will have most of the skill needed to prepare your own items. My intent is to offer you a procedure and practice items that will help you to do just that. Specifically, the objective of the book is this:

Be able to discriminate (select, point to) the test items that are appropriate (i.e., items that match the objective in performance and conditions) for testing the achievement of an instructional objective, when given (1) an objective, (2) one or more allegedly suitable test items, and (3) the Objective/Item Checklist.

The checklist just mentioned is provided as an aid to carrying out the matching procedure, so it will not be necessary to memorize the steps of that procedure (see page 59 and the card inserted between the last page and the back cover). Once you know *how* each step works, the checklist can be used to remind you *when* each may be applicable.

Three comments before we begin. First, most of the objectives you will encounter will need at least some repair before you can select or prepare items to match. Therefore, examples

of these kinds of objectives are included to make the practice items more useful.

Second, we are concerned here only with matching objectives and test items, and consequently we will not deal with the characteristics of either one that aren't relevant to the job at hand. For example, we will not deal with the usual test item construction issues of item difficulty or the structure of multiple-choice or true-false items. At that, the book offers more of a beginning than an end—the intent is more to get you started than to try to solve the advanced problems. But a beginning is important. After all, it is useless to complain that the advanced problems haven't been solved while one is still screwing up the fundamentals.

Finally, when an objective is being drafted, it is not always known whether it will ultimately be judged important enough to be achieved. For that reason, many objectives come into being that aren't really important. And when you see an objective that isn't important, that is, when it doesn't matter whether it is achieved, you should not only not bother to write or select test items that are appropriate to the objective—you shouldn't test at all.

And so, to begin.

ROBERT F. MAGER

Carefree, Arizona
January, 1984

Contents

Preface *v*

- 1 What It's All About 1**
- 2 Distinctions 7**
- 3 Decoding the Objective 15**
- 4 Matching the Performances 51**
- 5 Matching the Conditions 71**
- 6 Peripheralia 91**
- 7 A Pride of Items 105**
- 8 Got a Match? 135**

Trial Summary 163

Selected References 166

Checklist/Flowchart *Insert Card*

I || What It's All About

Suppose you had been working hard during a course to achieve this objective handed you by the instructor on the first day:

On a level paved street, be able to ride a unicycle one hundred yards without falling off.

Suppose you had strengthened your thighs with deep-knee bends, and had practiced riding until you could mount and ride with relative ease for at least *two* hundreds yards. And suppose that when testing time came around, your instructor asked you to get out pencil and paper and answer the following questions:

1. Define *unicycle*.
2. Write a short essay on the history of the unicycle.
3. Name at least six parts of the unicycle.
4. Describe your method of mounting a unicycle.

What would be your reaction? How would you feel if you had been told to learn one thing and were then tested on another?

Equally important, perhaps, would the instructor find out whether the objective had been achieved?

Suppose the instructor “justifies” this situation to you with one or more of the following comments. How would you feel?

“We don’t have the facilities to give performance tests.”

“We don’t have enough unicycles to go around.”

“This is an educational institution, not a training institution.”

“It doesn’t matter how well you can ride; if you don’t know anything about the unicycle you can’t really appreciate it.”

“I’m teaching for transfer.”

“It’s too easy to learn to ride a unicycle; I have to add some harder items so I can grade on a curve.”

“If everybody learned to ride, I’d have to give everybody an A.”

“I like to vary the type of items I use to make my tests interesting.”

“I want my tests to be a learning situation.”

“I’m teaching creativity and insight.”

“I have to design my tests so they can be machine scored.”

“Students should learn by discovery.”

Regardless of the truth or falsity of the comments listed above, the fact remains that you *cannot* find out whether a person can ride a unicycle unless you or someone else watches that person ride one. You cannot find out if the objective is achieved unless you use items that ask the student to perform whatever the objective is about. If you use items that aren’t “right” for an objective, not only will you *not* find out if your objective is achieved, you may fool yourself into thinking it is. That’s not so bad when an objective isn’t very important, but when it *matters* whether it is achieved, you run a risk by using poorly conceived or inappropriate items. If it matters whether the patient’s temperature is less than 100 degrees, you’d better use a thermometer to measure his or her temperature rather than a yardstick . . . or an essay. If it matters whether a student pilot can react quickly and accurately in a stall emergency, then you’d better use a reaction-producing item rather than a yardstick . . . or an essay. If it matters whether a student is able to read at least two hundred words per minute, then you’d better find out if that skill can be performed, so you can respond with

more instruction when it can't or with applause when it can, rather than with merely a student label (i.e., a grade).

THE VIEW FROM THE TOP

In theory, the development of items that test for an objective is straightforward and a simple matter for those whose objectives are derived from task analyses and are well stated. (Those who have objectives of this kind may not realize that they are among a small minority, and may well wonder how so much can be said about what to them is only "common sense.") One merely has to prepare items that ask the student to demonstrate the performance called for by the objective, under the conditions called for by the objective. In other words, one prepares items in which the performance and conditions match those of the objective to be assessed. In practice, this preparation is only a little more difficult than the telling.

However, the main difficulties do not arise from problems in item writing. Most difficulties are caused by the objectives. Often, when test items are drafted to match an objective, the objective itself exists in only a crude or sloppy state. If, for example, no performance at all is mentioned in the statement *called* an objective, test items cannot be written until the objective is repaired.

Another important obstacle to easy preparation of test items arises from the tendency of instructors to consider the student fair game for almost any kind of test. This tendency somehow gives instructors a feeling of uneasiness when they construct test items strictly according to the objective. "These items don't cover enough ground," the feeling says. "These items are too easy," it tells them and makes them conveniently forget that the object is not to develop a variety of items that only half the students can master, but to *prepare items that will reveal which students can perform as desired*. And the feeling goes on to say, "Well, maybe students can perform as well or better than expected, but they won't *really* understand it unless . . ." and then urges instructors to add all sorts of test

items having little or no relation to the objective. Finally, this funny feeling, in a last desperate bid for survival, says, "Well, maybe all the students *have* achieved the objective, but you need to add some harder items so you can spread them out on a curve."

Because the objectives we see and use vary considerably in their precision, I will be using comparable statements (of varying precision) in the examples that follow. We need to learn to handle the world as it is; it is hardly useful to learn to handle an ideal that seldom appears.

As for that funny feeling that urges us to use almost any form of test item so long as it "covers" the material presented, the best I can do to help you avoid that feeling is to describe the reasons for the present approach and to try to help you develop skill in implementing it. After all, *people who know how to do something are more likely to do it than those who don't!!* And if they know *why* they are doing it they may feel more at ease when they *are* doing it.

WHAT'S TO COME

The next chapter will describe a few distinctions in terms, so we will have a better basis for communication. After that there will be discussion and practice in locating and interpreting ("decoding") the key characteristics of an objective. Then there will be discussion and practice in matching performances, followed by practice in matching conditions. Next, if you like, you can practice repairing some items to match the objectives they are supposed to be related to, before going on to some practice in the entire skill. Finally, if you should desire to test *my* skill, there is a set of test items that allegedly matches the objective of this book.

Here and there along the way you will be encouraged to skip the sections that describe things you may already know. There is no sense in spending time going over what you can already do.

The use of inappropriate test items is a widespread phenomenon and is, in my opinion, a practice (malpractice?) most

urgently in need of improvement. When we deceive the student by discrepancies between our words and our deeds, both sides are the losers. Putting it more plainly, when we cheat students, they generally find a way to cheat back.

2 || Distinctions

A number of distinctions will serve as the basis for the following chapters, and it will be useful to describe these distinctions at least briefly. If we both use words in a similar manner, the words won't get in the way of the ideas; these descriptions won't take long to go through, and they should help to put a few ideas in their proper place.

ITEMS AND TESTS

Perhaps the main distinction to make is between tests and test items. If you've ever spent much time in a school you couldn't avoid either one, so this distinction may appear obvious. It is mentioned, however, as a way to remind you that this book is concerned mainly with test *items* rather than with tests.

Item: A test item calls for a single response or set of responses to a single stimulus or stimulus pattern. It is one sample of a behavior or performance. That performance may be simple, as when asking someone to write the answer to an addition problem, or it may be complex, as when asking someone to perform an appendectomy, analyze a problem, or compose a sonata.

Test: A test is an event during which someone is asked to demonstrate some aspect of his or her knowledge or skill. Though a test can consist of a single test item, a test generally consists of several items.

MEASUREMENT, EVALUATION, AND GRADING

Measurement: The process of measurement determines the extent of some characteristic associated with an object or person. For example, when we determine the length of a room or the weight of an object, we are measuring.

Evaluation: The act of evaluation compares a measurement with a standard and passes judgment on the comparison. We are making evaluations when we say things like—it's too long, it's too hot, he's not motivated, she's too slow, he's honest. We have noted the extent of some characteristic, compared it with some standard, and then passed judgment on the comparison.

A test is a way of *measuring* some desired characteristic. We pose a situation or problem and note a response. The shape or extent of the response provides the basis for the measurement. Thus, when we say that "seven out of ten of the student's responses were correct," we are noting the extent of an ability to respond to that kind of question or problem. We are using a test as a measuring device. On the other hand, when we then say things such as

"She passed"

"He flunked"

"She's not working up to her potential"

"He's good enough"

we are making evaluations. We have compared the results of the measurement with some standard (real or imagined, stable or shifting) and have made a judgment.

The difference between measurement and evaluation can be illustrated with this two-line dialogue:

"Hey, these watermelons are three feet long." (measurement)
"Wow!" (evaluation)

Or it can be illustrated with this example:

“This student can type thirty words per minute.” (measurement)

“That’s too slow.” (evaluation)

Grading: A grade is a label representing an evaluation. Sometimes that evaluation is based on measurement and sometimes on guesses, intuition, expectation, or bias. Traditionally, a grade has intended to say something about how well a student has performed (or tried) *in relation to his or her peers*; it has meant that the student is very good, pretty good, about the same as, not as good as, or poorer than those who happen to be his or her classmates. Also traditionally, the student has seldom been informed of the precise basis for the grade.

In this book we are concerned with the issue of measurement, with offering a sound basis for determining whether an objective has been accomplished.

NORM-REFERENCED AND CRITERION-REFERENCED EVALUATION

Norm-referenced: When the performance of one student is compared with that of other students, and a judgment is passed on the basis of that comparison, a norm-referenced evaluation is being made. Thus, when we say that student X is above average and that student Y is below average, we are making a norm-referenced evaluation. When we rank-order students on the basis of their performance in *reference to each other*, we are making norm-referenced evaluations. Grading on a curve is such an evaluation. So is the assignment of IQ.

If we have five automobiles, none of which runs, we might measure the extent of their defects and say “Automobile G ‘doesn’t run’ the best. None of them goes, but G is the best of non-goers.” That would be a norm-referenced evaluation. If we then said, “Give that car an A+,” we would be grading on the basis of a norm-referenced evaluation.