14. VARIABLE MAPPING

This primer shows how variable mapping is fundamental to measurement. Mapping is where test development begins (as an idea) and mapping represents its realization (through empirical validation) and its actionable interpretation.

The map of the variable begins as the blueprint for the design of a test. When the map is well conceived, its test design implementation will be a straightforward representation in ordered items. Later, as the map is empirically verified by candidates' responses, it results in the successful implementation and realization of an idea. The map of the variable pictures the idea and its realization.

Alfred Binet is better known for the American versions of his tests than for his ideas on test construction and measurement. But Binet's work in test development represents an excellent example of what is implied in mapping. The first (1905) edition of his test consisted of thirty items arranged in difficulty order. This item arrangement enabled measuring ability by locating a child "along" the ordered item scale.

Gould (1981) faults Binet for a "hodgepodge of diverse activities" in item selection. But, however diverse, Binet's items were not haphazard.

> First of all, it will be noticed that our tests are well arranged in a *real order of increasing difficulty.* It is *as the result of many trials,* that we have established this order; we have by no means imagined that which we present. If we had left the field clear to our conjectures, we should certainly not have admitted that it required the space of time comprised between four and seven years, for a child to learn to repeat 5 figures in place of 3. Likewise we should never have believed that it is only at ten years that the majority of children are able to repeat the names of the months in correct order without forgetting any; or that it is only at ten years that a child recognized all the pieces of our money (Binet, 1905, p. 185).

Binet relied upon "numerous" replications of his ordered items to give him the measurement accuracy he desired.

> "One might almost say, "It matters very little what the tests [items] are, so long as they are numerous'" (1911, p. 329).

Furthermore, Binet writes that his scale:

> "properly speaking, does not permit the measure of intelligence, because intellectual qualities are not superposable, and therefore cannot be measured as linear surfaces are measured" (1905, p. 40).

In recognizing this deficiency in his test, Binet shows that he knew that linearity was necessary for measurement.

Binet's essential ingredients for the construction of measurement were:

1. Item arrangement by difficulty order
2. Numerous items to insure precision
3. Need for, but recognized difficulty in producing, linear measures.

How else could one build a test? There is no other way except to begin as Binet did: with an idea for a variable illustrated by items arranged by intended difficulty, and measures of persons according to their locations among the items along the variable. Every attempt at test construction is made along such lines whether successful or not.

The hallmark of Binet's efforts is his attempt to benchmark items and persons. The idea of a benchmarked line of increasing amounts is fundamental in constructing a variable, and the map of intentions is the blueprint for item selection and/or item construction. Its realization gives us a picture of the variable and a means for seeing the locations of items and persons along the variable.

A contemporary example is WRAT3 (Wilkinson, 1993). This test implements the achievement measures: (1) word naming, (2) arithmetic computation and (3) spelling from dictation. The item arrangement is developmental and indicates the sequence of instruction and learning. The items for each scale proceed from items at the most elementary levels to those of increasing difficulty at higher levels.

Arrangement of items in difficulty order is exactly what we want in any measuring tool. The locations of items along the variable are determined by teacher judgment, curriculum, and learning experts. Validation is rendered by subjecting the initial item arrangement to empirical testing. When the arrangement of items is sensible it will be supported by data gathered from students' response to these items.

The arrangement of items should also correspond to the arrangement of persons. Less able persons should be located below more able persons. A hierarchical correspondence between items and persons will show easy items in company with less able persons, more difficult items associated with more able persons.

Determining the calibrations of items and the measures for persons will either substantiate the original item placement or suggest revisions. This leads to a continuous dialogue between the idea for the variable and its data. A good initial plan generally results in fewer cycles between idea and data before an acceptable definition of the intended variable is achieved.

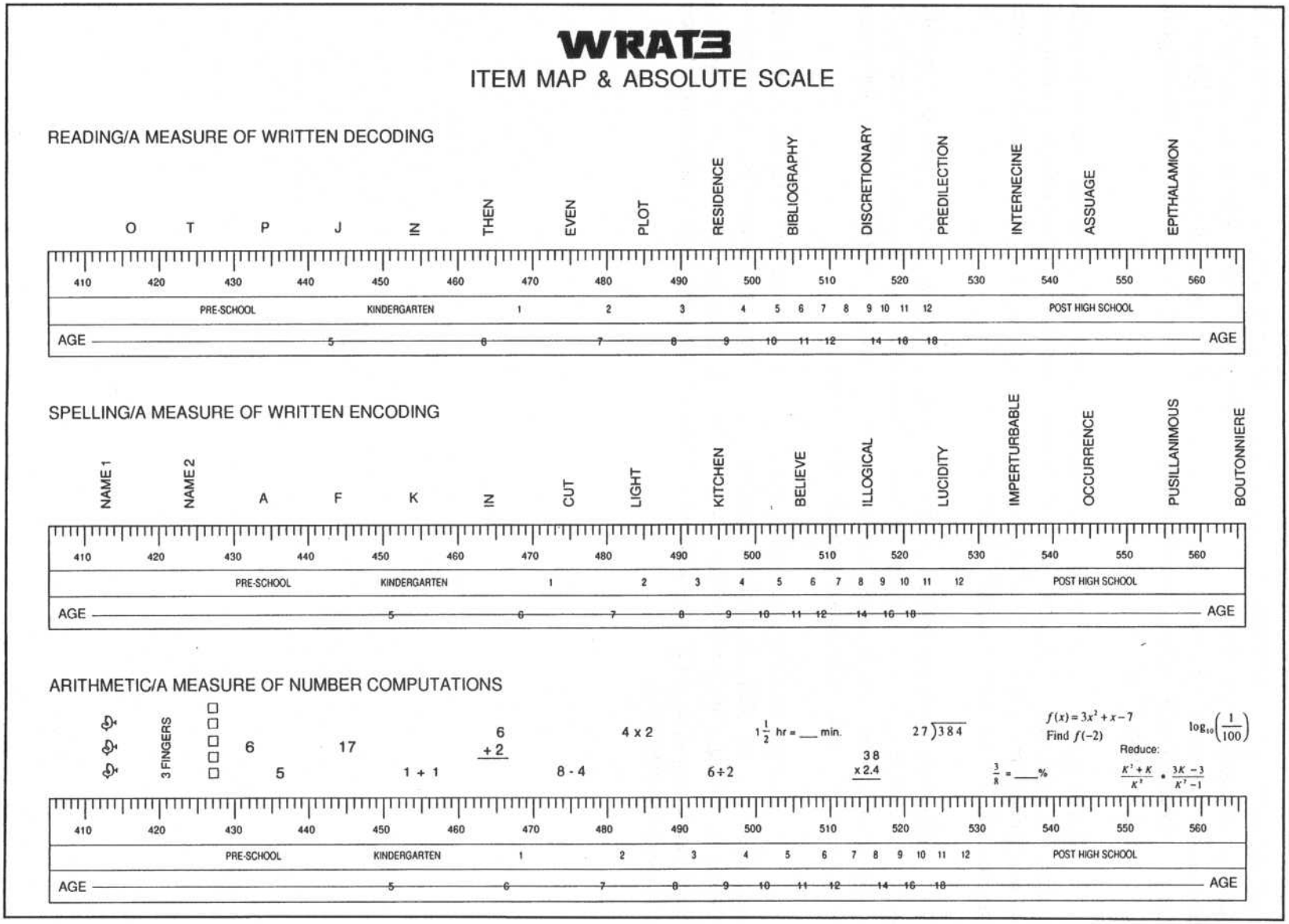
Successful item calibration and person measurement leads to a map of the variable. The resulting map is no less a "ruler" than one constructed for measuring length. It can be applied in a similar way to produce measures as useful as those of any yardstick.

Figure 14.1 is a map of the WRAT3 variables for word reading, arithmetic computation and spelling from dictation. It proceeds from left to right in a progressive order of difficulty for items and ability for persons.

The map of each variable gives sample items showing their progressive difficulty. Below the items is an absolute equal interval scale providing measures. The location of average grade and age are also given.

Figure 14.1

Variable map.



The map has immediate appeal and application. Like the marks of increasing height of a child on the door jamb, the map can show student progress on these three scales. Especially helpful is the overall view that the map provides, giving a sense of order and coverage to the entire variable of interest. The map shows the order implied in the variable and it can be used to show the location and subsequent progress of pupils along the ruler. The absolute scale gives values useful in data analysis. The grade and age norms show the progress we expect to see at increasing grade and age levels. The wider spacing observed on the left of the ruler compared to the right indicates the accelerated growth occurring among younger children.

Gathering data on all three scales across persons allows comparisons to be made between scales. We notice that successfully reading the word "residence" corresponds in general to spelling the word "kitchen" from dictation and computing the problem "6-2= ." These comparisons enable further diagnosis and make the map a useful diagnostic tool.

Although each ruler appears orderly, that order is in need of continual reappraisal and revalidation. Variable definitions are never finished. While there is consistency and order to the WRAT3 scales, a fact demonstrated by five successive editions, it remains necessary to monitor continuously the variable in order to keep the map coherent and up-to-date.

Continuous monitoring is required for any variable. Concerns for reliability and validity do not rest in historical coefficients, but in continuing successful demonstrations that can be referenced by test consumers in order to determine the extent to which the test is relevant to their intended application. Such indications of applicability must be continuously provided in order to maintain the variable map and assure its relevancy.

MEASUREMENT ESSENTIALS

2nd Edition

BENJAMIN WRIGHT

MARK STONE

WIDE RANGE, INC.
Wilmington, Delaware