3. THE IDEA OF A VARIABLE

When we make measures, we do so with the intention of being accurate enough for our practical purpose. We do not expect absolute precision. Our notion of "accurate" does not imply "perfect." Instead it implies "close enough to be useful." We record a person's height to some useful approximation like the nearest half-inch. This is sufficient for most practical purposes. More precision, such as to the nearest eighth or sixteenth of an inch, is rarely necessary and we would not ordinarily expect it to be given.

This example reminds us that while we want to be accurate there is always an implied, if not explicit, tolerance in our measures. Unless height measures require some particular accuracy, it is not necessary, and without scientific instrumentation, impossible, to make measures of height more accurate. However, we are not at all frustrated by our lack of absolute precision because "to the nearest half-inch" is practical and useful. We make measures which are good enough for the occasion, good enough to satisfy our practical requirements.

Measures are based on observations. Observations are essentially qualitative. To make measures we develop rules by which to control how these observations are best made. These rules include specifying the degree of accuracy that we want. When measuring height, for example, we ask people to remove their shoes, stand straight and not wiggle in order to standardize the observations. Then we observe which marks on our yardstick they exceed and which they fail to exceed. We find the marks closest to the top of their head. We pick the mark that looks closest and call their height the calibration of that nearest mark. These rules provide the level of accuracy we need in order to make useful measures of a person's height. We use the constructed functional unidimensionality of the yardstick to bring out and record the single dimensioned height of the multidimensional person.

Measuring "ability" is analogous to measuring "height." First we bring to the fore our idea of the variable we want to measure. Next we determine what observations it will be useful to consider as informative manifestations of that variable. Then we construct agents, write items, intended to elicit singular instances of this "made-to-be" unidimensional "ability" variable.

The idea of a variable can be visualized as a line that has direction. When we think about "length" we think about a line that is as long as necessary for our work. This idea is manifest in a one-foot ruler when we expect measures to be 1 to 12 inches, in a yardstick for 1 to 36 inches, on a surveyor's tape for longer distances and so on. In each of these instances the agent of measurement is a focused manifestation of our infinite linear image of the variable "length".

With these simple ideas of measurement in hand, let us turn to the problem of measuring an ability. Consider arithmetic ability and, more specifically, the computation skills needed for the whole number operations of addition, subtraction, multiplication and division.

We imagine a line of arithmetic items progressing from left to right with each successive item harder than the previous one. A few items will suffice for our example. Additional items are added to the line by designing them to fit between any two items that we have already placed upon the line and then verifying their location by observing student responses.

The idea of a line upon which to position arithmetic items provides us with a picture of the arithmetic variable and shows us how to proceed in the construction of tests to measure along that variable. We use our knowledge of arithmetic to position items along the line. Theoretical locations for the items can be hypothesized initially by teaching experts who have experience with students learning arithmetic. Later we can add and reposition items on the line as we observe how well students actually answer these items.

CONSTRUCTING THE LINE OF THE VARIABLE

We begin with a single item and position it on the line of the arithmetic variable:

$$\text{-------}\quad \begin{array}{r}5\\ \underline{+7}\end{array} \quad \text{-----------------}\!\!>$$

Can an easier item be constructed? Yes, and so we will position it somewhere to the left. A harder item will be positioned to the right. Hence:

$$\text{-}\ \begin{array}{r}2\\ \underline{+2}\end{array} \quad \text{----}\quad \begin{array}{r}5\\ \underline{+7}\end{array} \quad \text{-----------------}\!\!>$$

Now we have three items and the process of constructing new ones only requires hypothesizing their expected positions among the existing items and then estimating item positions empirically by collecting responses to them. The critical decision to make at this point is where each item belongs, in our best judgment, relative to the items already positioned on the line. There is also no reason why items cannot be repositioned according to better information from teachers about their difficulty relative to other items. Item construction thus proceeds in an orderly fashion guided by the idea of a line and the successive placement of items on the line according to our best expectations of their relative difficulty. These items now serve as the agents designed to evoke manifestations of our arithmetic variable.

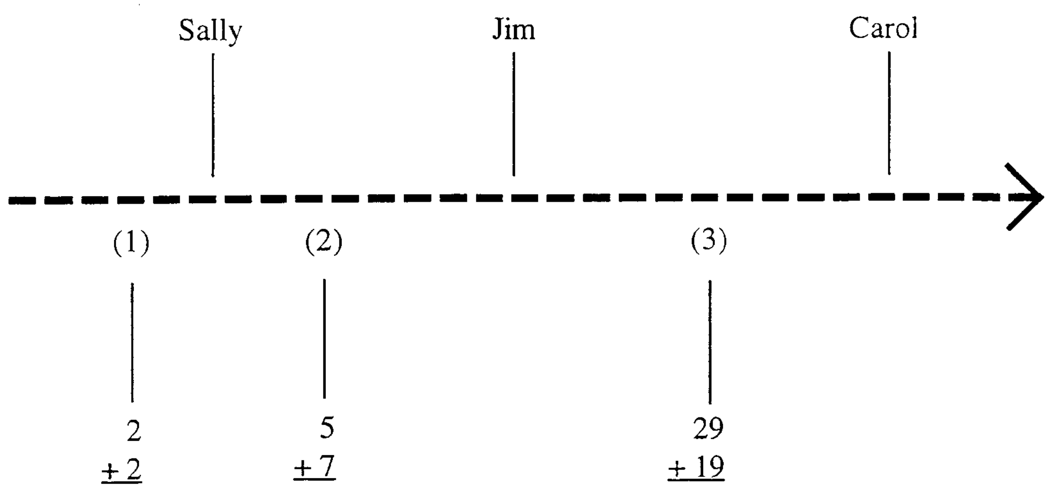
The line of our variable can be made as long as necessary to describe the variable. It can be divided into segments for ease in handling. It can be abbreviated according to our practical needs for administration in exactly the same way that we partition our idea of length into measuring tools - rulers, yardsticks, tapes of various extension - all to facilitate the measurement of various lengths.

The idea of a line helps us to determine item positions by considering each item relative to the items already positioned on the line. This determination can be done by comparing pairs of items with respect to their relative difficulties along the line. Each successive item position as we move to the right indicates "more" of the variable to be measured.

The idea of the variable becomes defined by the construction of the items which work to elicit indications of the variable. As we define the variable with more and more instances, using more and more items, our work of building the variable proceeds in a logical manner and our conceptualization of the variable becomes ever more clearly defined.

Once the variable is constructed by the line of items, we can proceed to position students on this same line. Their probable positions can be specified initially by our best guess as to their ability to correctly answer the items which define the variable. The line of our variable shows both the positions of items and the positions of students. Eventually the positions of students will become more explicit and more empirical as we observe what items they correctly answer.

Consider this picture:



Sally's position on the variable is indicated by an expected correct response to Item 1 but an expected incorrect responses to Items 2 and 3. Her differing responses to Items 1 and 2 locate her on the variable between two items that describe her ability in arithmetic computation. She can add 2 and 2 but not 5 and 7.

Jim's position is between Items 2 and 3 because we expect him to answer Items 1 and 2 correctly but not Item 3. In Jim's case we have somewhat less precision in determining his arithmetic ability because of the lack of items between Items 2 and 3. If we had additional items in this region, we could obtain a more accurate indication of Jim's position on the variable as defined by his responses to these additional items.

Carol solves all three problems. Her ability is "above" Item 3. But what her position is beyond Item 2 remains unknown. We cannot position her more precisely on the variable because we do not know whether her true position is only slightly above the position of Item 3 or far beyond it. If we had her responses to additional items on the variable above Item 3, Carol's position might be indicated more exactly.

Now we give a more specific example of how to construct a variable for arithmetic.

First, we choose 17 items and arrange them on a test form in what we expect to be their approximate order of difficulty.

Then we administer this test form to 270 students in Grades one to six in order to obtain actual data with which to calibrate these 17 items objectively.

Next we calibrate these items and determine person measures for this sample. (See Wright & Stone, 1979, for the details of how to do this. It is not hard to do.)

The calibrations of the 17 items are used to map the items in Table 3.1. The variable line goes down in difficulty from hard at the top to easy at the bottom. On the left side of Table 3.1 is the person count for this sample of 270 students at every raw score position, then comes the raw score, the measure implied by each raw score and the associated estimation error. Items are identified by their item number and text and positioned according to the difficulties calibrated from the observations gathered from our sample.

We have constructed an arithmetic computation variable and located items and students along it from our observations of how these students were able to answer these items.

Our development of this emerging variable defined by items and students provides an operational definition. The variable's limits are bounded only by the range of agents (items) and objects (students) that we can position along the line. We can make variables of interest as dense as we need. The tests which implement these variables can be sparse for rough screening or dense for more specific pinpointing.

Accuracy (i.e. reliability) of student position is given by the standard error associated with each measure. The unit of measurement used in this table is the logit expressed as a decimal centered on 0.0 for this set of 17 items. Observe that the standard errors are smallest (most precise) where items are most dense and we have the most information about the measure and largest (least precise) at the extremes where items are least dense and we have the least information.

Table 3.1 can be examined to determine where we have gaps between items, where there are too many items at a particular position and where more items are needed to extend the variable above and below the items already calibrated. (For an example of this kind of variable building, see Wright & Stone, 1979, pp. 83-93). The map of the variable is a picture of the extent to which we have accomplished the task of variable construction. The map also shows us what to do next.

Variable maps begin by showing item positions along the line of the variable as shown in Table 3.1. We can also add students along the line of the variable and index their positions on the map by name, grade, gender or other student characteristic. As we add to the map we enrich our picture of the variable and increase its utility.

The construction of an empirical variable map enhances the value of testing. A good variable map is self-explanatory because the visualization of the variable makes explicit what the variable represents. The interpretation of test results is facilitated because all items calibrated and all students measured are positioned together on the same variable - along with whatever additional information has been added to make the map more useful.

CRITERION REFERENCING

A variable map is automatically criterion-referenced by the relative positions of item content. The texts of the items in their positions along the variable describe in detail the explicit hierarchy of content and hence the construct implied by the variable. This item-by-item criterion referencing of the variable applies to any measure subsequently derived from any test composed of some items which have been calibrated on this variable. Thus, criterion referencing is complete and the evidence of content and construct validity is explicit.

NORM REFERENCING

Personal and demographic characteristics of any and all students tested can be added to the variable map at the measured positions of these students. This provides as extensive and versatile norm referencing as the use of tests based on items calibrated on this variable can provide. Thus, norm referencing is also as complete as possible with the data available.

Table 3.1

The Item Map of the Arithmetic Variable

STUDENT COUNT	RAW SCORE	MEASURE SCALE*	STANDARD ERROR	ITEM NUMBER	ITEM TEXT
		6.50		#17	$7\frac{1}{6} - \frac{3}{4}$
		6.30			
17	16	6.10	1.35		
		5.90			
		5.70			
		5.50			
		5.30			
		5.10			
		4.90			
22	15	4.70	1.11	#16	$\frac{1}{3} + \frac{1}{3}$
		4.50			
		4.30			
		4.10		#15	$536\overline{)4762}$
		3.90			
24	14	3.70	0.99		
		3.50			
		3.30			
		3.10			
		2.90			
23	13	2.70	0.88		
		2.50		#12	42×29
		2.30			
22	12	2.10	0.81		
		1.90			
		1.70		#9, #14	837×7; $5204 - 530$
24	11	1.50	0.76		
		1.30		#11	$31\overline{)62}$
18	10	1.10	0.74		
		0.90			
		0.70		#13	$9\overline{)72}$
11	9	0.50	0.74		
		0.30		#10	23×3
		0.10		#8	$84 - 36$
17	8	-0.10	0.75		
		-0.30			
23	7	-0.50	0.79		
		-0.70		#6	$45 + 16 + 27$
		-0.90			

Table 3.1 (Continued)

The Item Map of the Arithmetic Variable

STUDENT COUNT	RAW SCORE	MEASURE SCALE*	STANDARD ERROR	ITEM NUMBER	ITEM TEXT
23	6	-1.10	0.85	#7	14 - 8
		-1.30		#5	67 +4
		-1.50			
		-1.70			
15	5	-1.90	0.99		
		-2.10			
		-2.30			
		-2.50			
		-2.70			
		-2.90			
18	4	-3.10	1.15		
		-3.30			
		-3.50			
		-3.70			
		-3.90			
		-4.10			
9	3	-4.30	1.04		
		-4.50			
		-4.70		#2	6 + 7
		-4.90			
3	2	-5.10	0.98	#1, #4	7 + 4; 8 - 3
		-5.30			
		-5.50		#3	6 - 4
		-5.70			
		-5.90			
1	1	-6.10	1.16		
270				17	

MEAN ABILITY OF PERSONS = 1.03 LOGITS
STANDARD DEVIATION = 2.56 LOGITS

* Should we wish a numbering system simpler than the decimal logits, a linear conversion can be made to positive whole numbers. See Wright & Stone, 1979, pp. 191-209.

MEASUREMENT ESSENTIALS

2nd Edition

BENJAMIN WRIGHT

MARK STONE

WIDE RANGE, INC.
Wilmington, Delaware