Chapter 1

Construct Modeling: The "Four Building Blocks" Approach

1.0 CHAPTER OVERVIEW AND KEY CONCEPTS

construct modeling
the "four building blocks"
construct map
items design
outcome space
measurement model

This chapter begins with a description of what is meant by *measurement* in this book. The remainder of the chapter then outlines a framework, which I call *construct modeling*, for understanding how an instrument works by understanding how it is constructed. Construct modeling is a framework for developing an instrument by using each of four "building blocks" in turn. This chapter summarizes all four building blocks, and the following chapters describe each in detail. In this volume, the word *instrument* is defined as a technique of relating something we observe in the real world (sometimes called *manifest* or *observed*) to something we are measuring that only exists as part of a theory (sometimes called *latent* or *unobserved*). This is somewhat broader than the typical usage, which focuses on the most concrete manifestation of the instrument—the items or questions. Because part of the purpose of the book is to expose the less obvious aspects of measurement, this broader definition has been chosen. Examples of types and formats of instruments that can be seen as coming under the "construct mapping" framework are shown in this and the next few chapters. Generally, it is assumed that there is a *respondent* who is the object of measurement, and there is a *measurer* who seeks to measure something about the respondent. While reading the text, the reader should mainly see him or herself as the measurer, but it is always useful to assume the role of the respondent as well. The next four chapters explain each of the four building blocks in turn, giving much greater detail, many examples, and discussion of how to apply the ideas to instrument development.

1.1 WHAT IS MEASUREMENT?

In some accounts, *measurement* is defined as the assignment of numbers to categories of observations. The properties of the numbers become the properties of the measurement—nominal, ordinal, interval, ratio, and so on. (Stevens, 1946).[1] Assigning numbers to categories is indeed one feature of the account in this book; correspondingly, those numbers have certain properties. Yet that is only one aspect of the process of measurement—there are steps preceding the assignment of numbers that prepare the ground for measuring, and there are also steps after the assignment of numbers that (a) check that the assignment was successful, and (b) make use of the measurements.

[1]In Stevens' (1946) classic account, measures are classified into successively more number-like categories as follows: (a) when the objects of measurement can be placed into (unordered) categories, the measurement is *nominal*; (b) when the objects can be placed into ordered categories, the measurement is *ordinal*; (c) when the objects of measurement can be labeled with numbers that can be added and subtracted, the measurement is *interval*; and (d) when the objects of measurement can be labeled with numbers that can be used as divisors, the measurement is *ratio*.

The central purpose of measurement, as interpreted here, is to provide a reasonable and consistent way to summarize the responses that people make to express their achievements, attitudes, or personal points of view through instruments such as attitude scales, achievement tests, questionnaires, surveys, and psychological scales. That purpose invariably arises in a practical setting where the results are used to make some sort of decision. These instruments typically have a complex structure, with a string of questions or tasks related to the aims of the instrument. This particular structure is one reason that there is a need to establish measurement procedures. A simpler structure—say just a single question—would allow simpler procedures. However, there are good reasons that these instruments have this more complex structure, and those reasons are discussed in the following chapters.

The approach adopted here is predicated on the idea that there is a single underlying characteristic that an instrument is designed to measure. Many surveys, tests, and questionnaires are designed to measure multiple characteristics—here it is assumed that we can consider those characteristics one at a time so that the real survey or test is seen as being composed of several instruments, each measuring a single characteristic (although the instruments may overlap in terms of the items). This intention, which is later termed the *construct*, is established by the person who designs and develops the instrument. This person is called the *measurer* throughout this book. The instrument, then, is seen as a logical *argument* that the results can be interpreted to help make a decision as the measurer intended them to be. The chapters that follow describe a series of steps that can be used as the basis for such an argument. First, the argument is constructive—that is, it proceeds by constructing the instrument following a certain logic (this occupies the contents of chaps. 2–5). Then the argument is reflective, proceeding by gathering information on whether the instrument did indeed function as planned (this occupies the contents of chaps. 6–8). The book concludes with a discussion of next steps that a measurer might take. This lays the groundwork for later books.

In this book, the concept being explored is more like a verb, *measuring*, than a noun, *measurement*. There is no claim that the procedures explored here are the only way to measure—there are other approaches that one can adopt (several are discussed in chaps. 6 and 9). The aim is not to survey all such ways to measure, but to lay out one particular approach that the author has found successful over the last two decades in teaching measurement to students at the University of California, Berkeley, and consulting with people who want to develop instruments in a wide variety of areas.

1.2 THE CONSTRUCT MAP

An instrument is always something secondary: There is always a purpose for which an instrument is needed and a context in which it is going to be used (i.e., involving some sort of decision). This precipitates an idea or a concept that is the theoretical object of our interest in the respondent. Consistent with current usage, I call this the *construct* (see Messick, 1989, for an exhaustive analysis). A construct could be part of a theoretical model of a person's cognition—such as their understanding of a certain set of concepts or their attitude toward something—or it could be some other psychological variable such as "need for achievement" or a personality variable such as a bipolar diagnosis. It could be from the domain of educational achievement, or it could be a health-related construct such as "Quality of Life" or a sociological construct such as "rurality" or migrants' degree of assimilation. It could relate to a group rather than an individual person, such as a work group or sports team, or an institution such as a workplace, or it could be biological phenomena such as a forest's ability to spread in a new environment. It could even be a complex inanimate object such as a volcano's proclivity to erupt or the weathering of paint samples. There is a multitude of theories—the important thing here is to have one that provides motivation and structure for the construct to be measured.

The idea of a *construct map* is a more precise concept than *construct*. We assume that the construct we wish to measure has a particularly simple form—it extends from one extreme to another, from high to low, small to large, positive to negative, or strong to weak. There may be some complexity in what happens in between these extremes, but we are primarily interested in where a respondent stands on this range from one extreme to the other. In particular, there may be distinguishable qualitative levels between the extremes—these are important and useful in interpretation. At this point, it is still an idea, latent rather than manifest. Although qualitative levels are definable, we assume that the respondents can be at any point in between—that is, the underlying construct is continuous. In summary, a construct map can be said to be a unidimensional latent variable. Many constructs are more complex than this. For example, they may be multidimensional. This is not a barrier to the use of the methods described in this book—the most straightforward thing to do is tackle each dimension one at a time—that way they can each be seen as a construct map. There are also constructs that are quite different from those that can be well described by a construct map. For example, suppose the construct consists of two different groups, say those who are likely to immigrate and those who are not. This construct is not much like that of a construct map and, hence, is not likely to be well represented by one.

In this chapter, the four building blocks are illustrated with a recent example from educational assessment—an assessment system built for a high school chemistry curriculum, "Living by Chemistry: Inquiry-Based Modules for High School" (Claesgens, Scalise, Draney, Wilson, & Stacey, 2002). The Living by Chemistry (LBC) project at the Lawrence Hall of Science was awarded a grant from the National Science Foundation in 1999 to create a year-long course based on real-world contexts that would be familiar and interesting to students. The goal is to make chemistry accessible to a larger and more diverse pool of students while improving preparation of students who traditionally take chemistry as a prerequisite for scientific study. The focus is on the domain knowledge they have acquired during instructional interactions in terms of how the students are able to think and reason with chemistry concepts.

The set of constructs on which both the LBC curriculum and its assessment system (an application of the *BEAR Assessment System*; Wilson & Sloane, 2000) are built is called "Perspectives of Chemists." Three variables or strands have been designed to describe chemistry views regarding three "big ideas" in the discipline: matter, change, and stability. *Matter* is concerned with describing atomic and molecular views of matter. *Change* involves kinetic views of change and the conservation of matter during chemical change. *Stability* considers the network of relationships in conservation of energy. The matter progress variable is shown in Fig. 1.1. It describes how a student's view of matter progresses from a continuous, real-world view to a particulate view accounting for existence of atoms and molecule, and then builds in sophistication. This progression is conceptualized as being reflected in two substrands within matter: visualizing and measuring.

Levels of success	ATOMIC AND MOLECULAR VIEWS	MEASUREMENT AND MODEL REFINEMENT
5 Integrating	bonding and relative reactivity	models and evidence
4 Predicting	phase and composition	limitations of models
3 Relating	properties and atomic views	measured amounts of matter
2 Representing	matter with chemical symbols	mass with a particulate view
1 Describing	properties of matter	amounts of matter
	A Visualizing matter	B Measuring matter

Fig. 1.1 A construct map for the Matter strand from LBC.

Assessments carried out in pilot studies of this variable show that a student's atomic views of matter begin with having no atomic view at all, but simply the ability to describe some characteristics of matter, such as differentiating between a gas and solid on the basis of real-world knowledge of boiling solutions such as might be encountered in food preparation, for instance, or bringing logic and patterning skills to bear on a question of why a salt dissolves. This then became the lowest level of the matter variable. At this most novice level of sophistication, students employ no accurate molecular models of chemistry, but a progression in sophistication can be seen from those unable or unwilling to make any relevant observation at all during an assessment task on matter, to those who can make an observation and then follow it with logical reasoning, to those who can extend this reasoning in an attempt to employ actual chemistry knowledge, although they are typically done incorrectly at first attempts. All these behaviors fall into Level 1, called the "Describing" level, and are assigned incremental 1– and 1+ scores, which for simplicity of presentation are not shown in this version of the framework.

When students begin to make the transition to accurately using simple molecular chemistry concepts, Level 2 begins, which is called the "Representing" level. At Level 2 of the matter progress variable, we see students using one-dimensional models of chemistry: A simple representation or a single definition is used broadly to account for and interpret chemical phenomena. Students show little ability to combine ideas. Here students begin extending experience and logical reasoning to include accurate chemistry-specific domain knowledge. In the conceptual framework, this is when students begin to employ definitions, terms, and principles with which they later reason and negotiate meaning. At this level, students are concerned with learning the language and representations of the domain of chemistry and are introduced to the ontological categories and epistemological beliefs that fall within the domain of chemistry. Students may focus on a single aspect of correct information in their explanations, but may not have developed more complete explanatory models to relate to the terms and language.

When students can begin to combine and relate patterns to account for (e.g., the contribution of valence electrons and molecular geometry to dissolving), they are considered to have moved to Level 3, "Relating." Coordinating and relating developing knowledge in chemistry becomes critical to move to this level. Niaz and Lawson (1985) argued that without generalizable models of understanding, students choose to memorize rules instead, limiting their understanding to the Representing level of the perspectives. Students need a base of domain knowledge before integration and coordination of the knowledge develops into understanding (Metz, 1995). As they move toward the Relating level, students should be developing a foundation of domain knowledge so that they can begin to reason like chemists by relating terms to conceptual models of understanding in chemistry, rather than simply memorizing algorithms and terms. Students need to examine and connect ideas to derive meaning in order to move to the Relating level.

The LBC matter strand is an example of a relatively complete construct map, although as yet untested at the upper end: These cover college and graduate levels—those interested in the upper levels should contact the LBC project at Lawrence Hall of Science. When a construct map is first postulated, it is often much less well formed than this. The construct map is refined through several processes as the instrument is developed. These processes include: (a) explaining the construct to others with the help of the construct map, (b) creating items that you believe will lead respondents to give responses that inform levels of the construct map, (c) trying out those items with a sample of respondents, and (d) analyzing the resulting data to check whether the results are consistent with your intentions as expressed through the construct map.

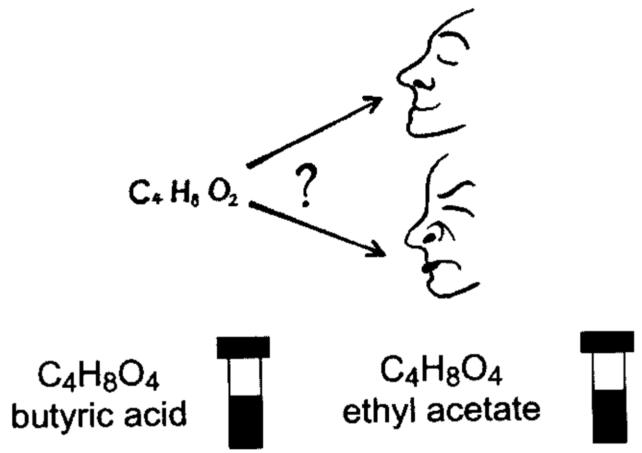
1.3 THE ITEMS DESIGN

Next the measurer must think of some way that this theoretical construct could be manifested in a real-world situation. At first this will be not much more than a hunch, a context that one believes the construct must be involved in—indeed that the construct must play some determining role in that situation. Later this hunch will become more crystallized and will settle into a certain pattern. The relationship between the items and the construct is not necessarily one way as it has just been described. Often the items will be thought of first and the construct will be elucidated only later—this is simply an example of how complex a creative act such as instrument construction can be. The important thing is that the construct and items should be distinguished, and that *eventually* the items are seen as realizations of the construct.

For example, the LBC items often began as everyday events that have a special significance to a chemist. Typically, there will be more than one real-world manifestation used in the instrument; these parts of the instrument are generically called *items*, and the format in which they are presented to the respondent is called the *items design*. An item can also take on many forms. The most common ones are probably the multiple-choice item from achievement testing and the Likert-type item (e.g., with responses ranging from *strongly agree* to *strongly disagree*) from surveys and attitude scales. Both are examples of the forced-choice type of item, where the respondent is given only a limited range of possible responses. There are many variants on this, ranging from questions on questionnaires to consumer rankings of products. The respondent may also produce a free response within a certain mode, such as an essay, interview, or performance (such as a competitive dive, piano recital, or scientific experiment). In all of these examples so far, the respondent is aware that they are being observed, but there are also situations where the respondent is being observed without such awareness. The items may be varied in their content and mode: Interview questions typically range over many aspects of a topic; questions in a cognitive performance task may be presented depending on the responses to earlier items; items in a survey may use different sets of options; and some may be forced-choice and some free-response.

In the case of LBC, the items are embedded in the instructional curriculum, so much so that the students would not necessarily know that they were being assessed unless the teacher tells them. An example LBC item is shown in Fig. 1.2. This task was designed to prompt student responses that relate to the lower portions of the matter construct described in Fig. 1.1. (An example of student response to this task is shown later in Fig. 1.6.)

The initial situation between the first two building blocks can be depicted as in Fig. 1.3. Here the construct and items are both only vaguely known, and there is some intuitive relationship between



Both of the solutions have the <u>same molecular formulas</u>, but <u>butyric acid smells bad and putrid while ethyl acetate smells good and sweet</u>. Explain why these two solutions smell differently.

FIG. 1.2 An example LBC item.

FIG. 1.3 A picture of an initial idea of the relationship between construct and item responses.

them (as indicated by the dotted line). Causality is often unclear at this point, perhaps the construct "causes" the responses that are made to the items, perhaps the items existed first in the developer's plans and hence could be said to "cause" the construct to be developed. It is important to see this as an important and natural step in instrument development—a step that always occurs at the beginning of instrument development and can need to recur many times as the instrument is tested and revised.

Unfortunately, in some instrument development efforts, the conceptual approach does not go beyond the state depicted in Fig. 1.3, even when there are sophisticated statistical methods used in the data analysis. This unfortunate abbreviation of the instrument development typically results in several shortcomings: (a) arbitrariness in choice of items and item formats, (b) no clear way to relate empirical results to instrument improvement, and (c) an inability to use empirical findings to improve the idea of the construct. To avoid these problems, the measurer needs to build a structure that links the construct closely to the items—that brings the inferences as close as possible to the observations.

One way to do that is to see causality as going from the construct to the items—the measurer assumes that the respondent "has" some amount of the construct, and that amount of the construct is a *cause* of the responses to the items in the instrument that the measurer observes. That is the situation shown in Fig. 1.4—the causal arrow points from left to right. However, this causal agent is latent—the measurer cannot observe the construct directly. Instead the measurer observes the responses to the items and must then *infer* the underlying construct from those observations. That is, in Fig. 1.4, the direction of the *inference* made by the measurer is from right to left. The remaining two building blocks embody two different steps in

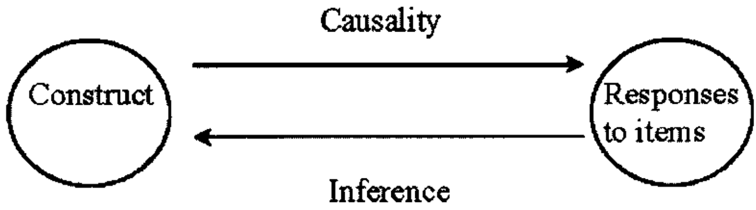


FIG. 1.4 A picture of the construct modeling idea of the relationship between degree of construct possessed and item responses.

that inference. Note that the idea of causality here is an assumption; the analysis does not prove that causality is in the direction shown, it merely assumes it goes that way. In fact the actual mechanism, like the construct, is unobserved or latent. It may be a more complex relationship than the simple one shown in Fig. 1.4. Until research reveals the nature of that complex relationship, the measurer is forced to act as if the relationship is the simple one depicted.

1.4 THE OUTCOME SPACE

The first step in the inference is to make a decision about which aspects of the response to the item will be used as the basis for the inference, and how those aspects of the response are categorized and then scored. This I call the *outcome space*. Examples of outcome spaces include: The categorization of question responses into "true" and "false" on a survey (with subsequent scoring as, say, "1" and "0"); the question and prompt protocols in a standardized open-ended interview (Patton 1980) and the subsequent categorization of the responses; and the translation of an educational performance into ordered levels using a so-called *rubric*, more plainly called a *scoring guide*. Sometimes the categories are the final product of the outcome space, sometimes the categories are scored so that the scores can (a) serve as convenient labels for the outcomes categories, and (b) be manipulated in various ways. To emphasize this distinction, the outcome space may be called a *scored* outcome space. The resulting scores play an important role in the construct mapping approach. They are the embodiment of the direction of the construct map (e.g., positive scores go upwards in the construct map).

The outcome space is usually implemented by a person who rates the responses into certain categories—I call the person in this role the *rater* (sometimes also called a *reader* or *judge*). The rater might also be a piece of software as is needed in an intelligent tutoring system (ITS), or it can be a fully automated rule, as in a multiple-choice item. The distinction of the outcome space from the items design is not always obvious mainly due to the special status of the two most common item formats—the multiple-choice item and the Likert-style item. In both of these item formats, the item design and outcome space have been collapsed—there is no need to categorize the responses because that is done by the respondents. In most cases, the scores to be applied to these categories are also fixed beforehand. However, these common formats should really be seen as "special cases"—the more generic situation is that of free-responses—this becomes clear when one sees that the development of these fixed-choice item formats (properly) includes an iteration that is in the free-response format (this point is returned to in Section 3.3).

The outcome space for the LBC matter constructs is summarized in Fig. 1.5—it is divided into ordered categories because the LBC curriculum developers see the underlying latent construct as a dimension—that is, as they see the students as progressing from little of it at the beginning of the year, and (if the curriculum developers and teachers have been successful) to having more at the end. This scoring guide allows a teacher to score student responses to the questions related to the matter constructs into the six different levels. Level 1, "Describing," has been further differentiated into three ordered sublevels—similar differentiation is planned for the other levels where it is found to be appropriate. Note how the scores (even the + and –) relate the categories to the desired direction of student progress. As well as the scoring guide in Fig. 1.5, teachers have available to them examples of student work (called *exemplars* in LBC), complete with adjudicated scores and explanations of the scores. An example is shown in Fig. 1.6. A training method called *moderation* is also used to help teachers be accurate raters and interpret the results in the classroom (see Wilson & Sloane, 2000, for a discussion of this). Really, it is the sum of all these elements that is the true outcome space, Fig. 1.5 is just a summary of one part of it. What we get out of the outcome space is a score, and for a set of tasks it gives a set of scores.

X. No opportunity.

There was no opportunity to respond to the item.

0. Irrelevant or blank response.

Response contains no information relevant to item.

1. *Describe the properties of matter*

The student relies on macroscopic observation and logic skills rather than employing an atomic model. Students use common sense and experience to express their initial ideas without employing correct chemistry concepts.

1 −	Makes one or more macroscopic observation and/or lists chemical terms without meaning.
1	Uses macroscopic observations/descriptions and restatement AND comparative/logic skills to generate classification, BUT shows no indication of employing chemistry concepts.
1+	Makes accurate simple macroscopic observations (often employing chemical jargon) and presents supporting examples and/or perceived rules of chemistry to logically explain observations, BUT chemical principles/definitions/rules cited incorrectly.

2. *Represent* changes in matter with chemical symbols

The students are "learning" the definitions of chemistry to begin to describe, label, and represent matter in terms of its chemical composition. The students are beginning to use the correct chemical symbols (i.e. chemical formulas, atomic model) and terminology (i.e. dissolving, chemical change vs. physical change, solid liquid gas).

2-	Cites definitions/rules/principles pertaining to matter somewhat correctly.
2	Correctly cites definitions/rules/principles pertaining to chemical composition.
2+	Cites and appropriately uses definitions/rules/principles pertaining to the chemical composition of matter and its transformations.

3. *Relate*

Students are relating one concept to another and developing behavioral models of explanation.

4. *Predicts how the properties of matter can be changed.*

Students apply behavioral models of chemistry to predict transformation of matter.

5. *Explains the interactions between atoms and molecules*

Integrates models of chemistry to understand empirical observations of matter/energy.

FIG. 1.5 The LBC outcome space, represented as a scoring guide.

1.5 THE MEASUREMENT MODEL

The second step in the inference is to relate the scores to the construct. This is done through the fourth building block, which is traditionally termed a *measurement model*—sometimes it is also called a *psychometric model*, sometimes a *statistical model*, although the

A response at the Representing Level:

"They smell differently b/c even though they have the same molecular formula, they have different structural formulas with different arrangements and patterns."

Analysis: Appropriately cites principle that molecules with the same formula can have different arrangements of atoms. But the answer stops short of examining structure-property relationships (a relational, level 3 characteristic).

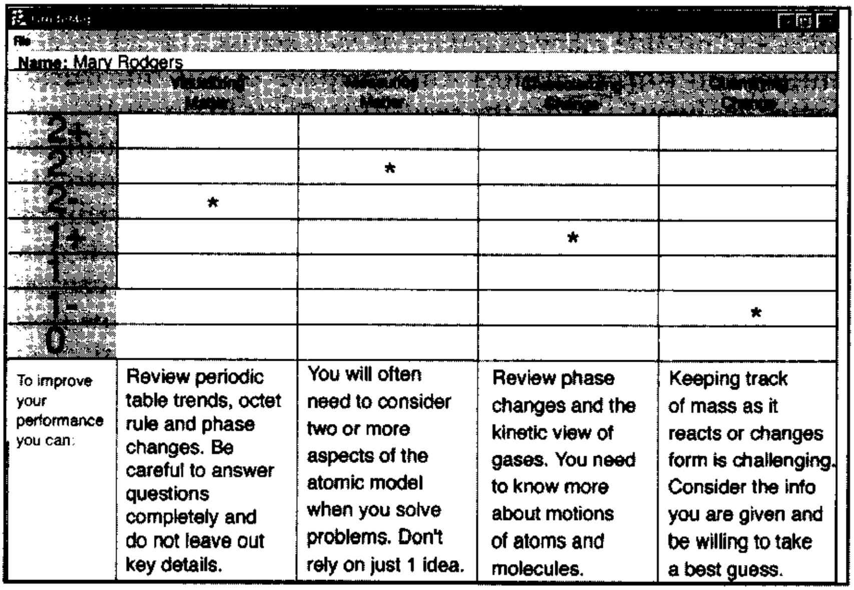
FIG. 1.6 Student respose to the item in Fig. 1.2.

conceptualization used in this chapter does not require that a statistical model be used, hence it might also be termed an *interpretational model* (National Research Council, 2001). The measurement model must help us understand and evaluate the scores that come from the item responses and hence tell us about the construct, and it must also guide the use of the results in practical applications. Simply put, the measurement model must translate scored responses to locations on the construct map. Some examples of measurement models are the "true-score" model of classical test theory, the "domain score" model, factor analysis models, item response models, and latent class models. These are all formal models. Many users of instruments (and also many instrument developers) also use informal measurement models when they think about their instruments.

The interpretation of the results is aided by graphical summaries that are generated by a computer program (*GradeMap*; Wilson, Kennedy, & Draney, 2004). For example, a student's profile across the four constructs is shown in Fig. 1.7—this has been found useful by teachers for student and parent conferences. Other displays are also available: time charts, whole-class displays, subgroup displays, and individual "fit" displays (which are displayed and described in later chapters).

Note that the direction of inference in Fig. 1.8—going from the items to the construct—should be clearly distinguished from the direction of causality, which is assumed to go in the opposite direction. In this figure, the arrow of causality does not go through the outcome space or measurement model because (presumably) the construct would have caused the responses regardless of whether the measurer had constructed a scoring guide and measurement model.

This sometimes puzzles people, but indeed it amply displays the distinction between the *latent* causal link and the *manifest* inferential link. The initial, vague link (as in Fig. 1.3) has been replaced in Fig. 1.8 by a causal link and several inferential links.


File

Name: Mary Rodgers

	[illegible]	[illegible]	[illegible]	[illegible]
2+				
2		*		
2-	*			
1+			*	
1				
1-				*
0				
To improve your performance you can:	Review periodic table trends, octet rule and phase changes. Be careful to answer questions completely and do not leave out key details.	You will often need to consider two or more aspects of the atomic model when you solve problems. Don't rely on just 1 idea.	Review phase changes and the kinetic view of gases. You need to know more about motions of atoms and molecules.	Keeping track of mass as it reacts or changes form is challenging. Consider the info you are given and be willing to take a best guess.

FIG. 1.7 A student's profile on the LBC constructs.

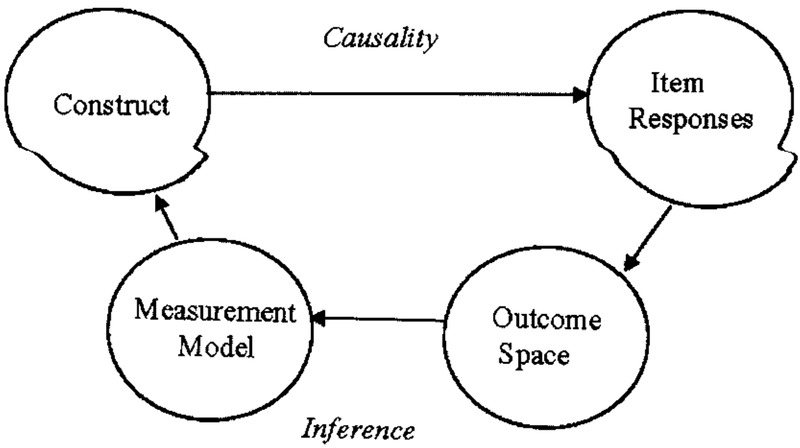


FIG. 1.8 The "four building blocks" showing the directions of causality and inference.

1.6 USING THE FOUR BUILDING BLOCKS TO DEVELOP AN INSTRUMENT

The account so far, apart from the LBC example, has been at quite an abstract level. The reader should not be alarmed by this because the next four chapters are devoted, in turn, to each of the four building blocks and provide many examples of each across a broad range of contexts and subject matters. One purpose of this introductory chapter has been to orient the reader to what is to come.

Another purpose of this chapter is to start the reader thinking and learning about the practical process of instrument development. If the reader wants to learn to develop instruments, it is obvious that he or she should be happy to read through this section and carry out the exercises and class projects that are described in the chapters that follow. However, even if practical experience about how to develop instruments is not the aim of the reader, then this section, and later sections like it, should still be studied carefully and the exercises carried out fully. The reason for this is that learning about measurement without actually developing an instrument leaves the reader in an incomplete state of knowledge—it is a bit like learning how to ride a bike, cook a soufflé, or juggle by reading about it in a book without actually trying it. A great deal of the knowledge is only appreciated when you experience how it all works together. It can be difficult to actually carry out the exercises, and certainly it takes more time than just reading the book, but carrying out these exercises can bring its own sense of satisfaction and will certainly enrich the reader's appreciation of the complexity of measurement.

The four building blocks provide not only a path for inference about a construct, but they can also be used as a guide to the construction of an instrument to measure that construct. The next four chapters are organized according to a development cycle based on the four building blocks (see Fig. 1.9). They start by defining the idea of the construct as embodied in the construct map (chap. 2), and then move on to develop tasks and contexts that engage the construct—the items design (chap. 3). These items generate responses that are then categorized and scored—that is the outcome space (chap. 4). The measurement model is applied to analyze the scored responses (chap. 5), and these measures can then be used to reflect back on the success with which one has measured the construct—which brings one back to the construct map (chap. 2), so this se-

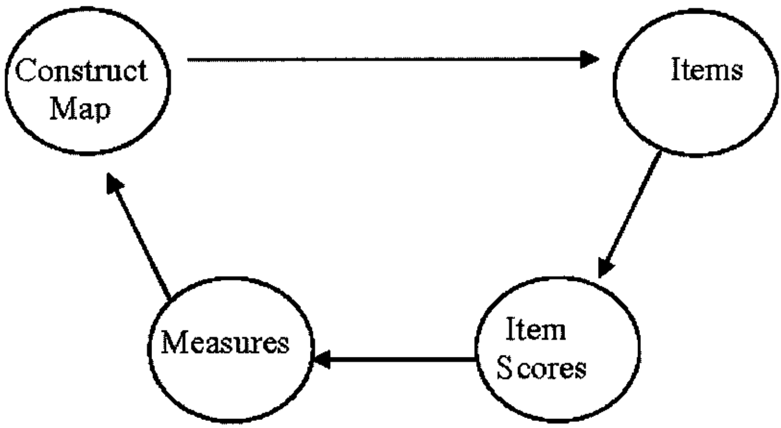


FIG. 1.9 The instrument development cycle through the four building blocks.

quence through the building blocks is actually a cycle—a cycle that may be repeated several times. The following three chapters (6, 7, and 8) help with this appraisal process by gathering evidence about how the instrument works: on model fit, reliability evidence, and validity evidence, respectively.

Every new instrument (or even the redevelopment or adaptation of an old instrument) must start with an idea—the kernel of the instrument, the "what" of the "what does it measure?", and the "how" of "how will the measure be used?" When this is first being considered, it makes a great deal of sense to look broadly to establish a dense background of knowledge about the content and uses of the instrument. As with any new development, one important step is to investigate (a) the theories behind the construct, and (b) what has been done in the past to measure this content—in particular, the characteristics of the instrumentation that was used. Thus, a *literature review* is necessary and should be completed before going too far with other steps (say, before commencing the activities discussed in chap. 3). However, a literature review is necessarily limited to the insights of those who previously worked in this area, so other steps also have to be taken.

At the beginning, the measurer needs to develop a small set of informants to help with instrument design. They should be chosen to span as well as go slightly outside the usual range of respondents.

Those outside the usual range would include (a) professionals, teachers/academics, and researchers in the relevant areas; as well as (b) people knowledgeable about measurement in general and/or measurement in the specific area of interest; and (c) other people who are knowledgeable and reflective about the area of interest and/or measurement in that area, such as policymakers, and so on. At this point, this group (which may change somewhat in nature over the course of the instrument development) can help the measurer by discussing experiences in the relevant area, criticizing and expanding on the measurer's initial ideas, serving as guinea pigs in responding to older instruments in the area, and responding to initial item formats. The information from the informants should overlap that from the literature review, but may also contradict it in parts.

1.7 RESOURCES

For an influential perspective on the idea of a construct, see the seminal article by Messick (1989) referenced earlier. A contemporary view that builds on that perspective, and one that is similar in a number of ways to the current account, is given in Mislevy, Wilson, Ercikan, and Chudowsky (2003), and a similar one is found in Mislevy, Steinberg, and Almond (2003).

The link between the construct map and measurement model was made explicit in two books by Wright (Wright & Stone, 1979; Wright & Masters, 1981), which are also seminal for the approach taken in this book.

The BEAR Assessment System (Wilson & Sloane, 2000), which is based on the four building blocks, has been used in other contexts besides the LBC example given earlier (Claesgens, Scalise, Draney, Wilson, & Stacey, 2002). Some are: (a) SEPUP's IEY curriculum (see Wilson & Sloane, 2000), and (b) the Golden State Exams (see Wilson & Draney, 2000).

A closely related approach is termed *Developmental Assessment* by Geoffery Masters and his colleagues at the Australian Council for Educational Research—examples are given in Department of Employment, Education and Youth Affairs (1997) and Masters and Forster (1996). This is also the basis of the approach taken by the Organization for Economic Co-operation and Development's (1999) PISA project.

Many examples of construct maps across both achievement and attitude domains are given in the series of edited books called "Objective Measurement: Theory into Practice" (see Engelhard & Wilson, 1996; Wilson, 1992a, 1992b, 1994a, 1994b; Wilson & Engelhard, 2000; Wilson, Engelhard, & Draney, 1997). Further examples can be found among the reference lists in those volumes.

1.8 EXERCISES AND ACTIVITIES

1. Explain what your instrument will be used for and why existing instruments will not suffice.
2. Read about the theoretical background to your construct. Write a summary of the relevant theory (keep it brief—no more than five pages).
3. Research previous efforts to develop and use instruments with a similar purpose and ones with related, but different, purposes. In many areas, there are compendia of such efforts—for example, in the areas of psychological and educational testing, there are series like the Mental Measurements Yearbook (Plake, Impara, & Spies, 2003)—similar publications exist in many other areas. Write a summary of the alternatives that are found, summarizing the main points perhaps in a table (keep it brief—no more than five pages).
4. Brainstorm possible informants for your instrument construction. Contact several and discuss your plans with them—secure the agreement of some of them to help you out as you make progress.
5. Try to think through the steps outlined earlier in the context of developing your instrument, and write down notes about your plans, including a draft timetable. Try to predict problems that you might encounter as you carry out these steps.
6. Share your plans and progress with others—discuss what you and they are succeeding on and what problems have arisen.