Human cognitive abilities

A survey of factor-analytic studies

JOHN B. CARROLL
Professor of Psychology Emeritus
University of North Carolina at Chapel Hill

CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Victoria 3166, Australia

First published 1993

Printed in the United States of America

Library of Congress Cataloging-in-Publication Data

Carroll, John Bissell, 1916–
Human cognitive abilities : a survey of factor-analytic studies / John B. Carroll.
p. cm.
Includes bibliographical references and indexes.
ISBN 0-521-38275-0
1. Cognition. 2. Psychology – Mathematical models. 3. Factor analysis. I. Title.
BF311.C288 1993 92-12453
153–dc20 CIP

A catalogue record for this book is available from the British Library.

ISBN 0-521-38275-0 hardback
ISBN-0-521-38712-4 paperback

Contents

Preface

This book is in a sense an outcome of work I started in 1939, when, as a graduate student at the University of Minnesota, I became aware of L.L. Thurstone's research on what he called "primary mental abilities" and undertook, in my doctoral dissertation, to apply his factor-analytic techniques to the study of abilities in the domain of language. Over the years of my career as a specialist in psychometrics, educational psychology, and the psychology of language I tried to keep abreast of both methodological and substantive developments in factor analysis, and from time to time I found it useful to conduct factor-analytic studies on topics of particular interest to me. Increasingly, however, I sensed the field's need for a thoroughgoing survey and critique of the voluminous results in the factor-analytic literature on cognitive abilities. It was not until 1974, when I came to the University of North Carolina at Chapel Hill as director of the L.L. Thurstone Psychometric Laboratory, that I felt that an appropriate use of my time would be to plan and execute such a survey.

A sabbatical year in 1979–80 supported by the Kenan Fund at the University of North Carolina and by the James McKeen Cattell Fund enabled me to start compiling materials systematically and make visits to prominent investigators in the U.S. and Europe. The major efforts starting in 1983, after I retired from the university, were supported chiefly by grant BNS-82-12486 from the National Science Foundation.

The book has three parts. Part I (Chapters 1–4) is introductory, historical, and methodological. Part II consists of chapters covering each of a number of domains of ability, ending with Chapter 15 on higher-order factors of ability, including *g* or general intelligence. In Part III, I consider more general issues about abilities. In Chapter 16 I propose a three-stratum theory of cognitive abilities. In Chapter 17 I outline the implications of such a theory for problems of nature and nurture, and more generally for cognitive psychology. In Chapter 18 I make recommendations for future research, for the application of

1 *The Study of Cognitive Abilities*

Our work is primarily with the grade book of the psychologist.
Clark Wissler (1901)

SOME PROBLEMS OF DEFINITION

A predominant and recurring concern throughout this book is the identification and description of cognitive abilities. I had better be clear, at the outset, on what I mean by *ability*, *cognitive ability*, and related terms.

Ability

Although the term *ability* is in common usage both in everyday talk and in scientific discussions among psychologists, educators, and other specialists, its precise definition is seldom explicated or even considered. It is a word that seems to be accepted as a sort of conceptual primitive, and in fact it is intimately related to such commonly used words as *able* and the simple modal auxiliary *can*. It is sometimes used to characterize material objects, as in the sentence "This bullet has the ability to penetrate a wooden board three inches thick." More frequently, however, it is used to characterize attributes of human individuals, as in expressions like *athletic ability*, *musical ability*, and (in the context of this book) *cognitive ability*. It expresses a kind of *potential*, a term which has merited the attention of philosophers of education (Scheffler, 1985).

Oddly enough, dictionaries are of little help in developing an exact, analyzed meaning of the term. The *American Heritage Dictionary*, for example, defines ability as "the quality of being able to do something; physical, mental, financial, or legal power to perform." In the present context, of course, we can lay aside concern with financial and legal powers, but mental powers, and possibly physical powers, remain of interest. Dictionary definitions often have an air of circularity, as is the case here: *ability* is defined in terms of "being able to perform something" but *able* is defined as meaning "having sufficient ability." Dictionaries of psychology might be more useful, but it happens that the word *ability* does not appear either as an entry term or in the index of a recently issued *Encyclopedic Dictionary of Psychology* (Harré & Lamb, 1983), although it is used there in numerous contexts, for example, in defining intelligence as "the all-round mental ability (or thinking skills) either of human or of lower animal species" (p. 313). In older dictionaries of psychology, considerable attention is devoted to defining ability and related terms. English and English (1958), for example, define ability as "actual power to perform an act, physical or mental, whether or not attained by training and education." They continue:

> GENERAL ABILITY is concerned with all sorts of tasks, but especially those of a cognitive or intellectual sort. *Syn. intelligence.* SPECIAL ABILITY has to do with a defined kind of task. Each special ability should, when possible, be so defined as not to overlap with other special abilities.

Curiously, none of these definitions contains, in any explicit way, the notion that there can be variations in ability over individuals.

It seems, therefore, that we must pursue a bit of logical and semantic analysis to arrive at a more precise definition of the term *ability* as it is to be used in this book. Some issues to be addressed are: In what sense does ability imply "potential"? Is ability a matter of degree, and if so, to what extent can its degree be quantified? To what extent may ability vary within an individual and across different individuals? How general is ability, that is, does it apply only to single performances, to some class or classes of performances, or to all possible performances? To what extent is an ability to be construed as a "trait" of an individual? Let us first consider these questions in the case of the physical ability of strength, because this case affords a concrete, easily grasped context in which to do so – a context more easily handled than if we were to consider a mental ability of some kind.

Every ability is defined in terms of some kind of performance, or potential for performance. Physical strength would have to be defined in terms of a performance that would require physical strength. Lifting 100 pounds of weight on a barbell would be one such performance. Suppose that an individual is characterized as possessing the ability to lift a barbell with 100 pounds of weight on its ends. This implies that the individual has the *potential* of doing so *if conditions are favorable to it* – that is, if a 100-pound barbell is available, if the individual is fully conscious and attentive, is willing and motivated to do so if asked, and can assume an appropriate position and a good grip on the barbell to perform the task. Nevertheless, even the concept of potential has to be thought of in probabilistic terms. If the individual were tested for lifting a 100-pound barbell on 100 different occasions, he or she might succeed, say, on only 95 of these; various unforeseen or unknown conditions might prevent the person from performing on the other five trials. Still, we would be inclined to admit, on this basis, that the individual *does* have the ability to lift a 100-pound barbell. In fact, if the individual succeeded on only *one* occasion, we might still be inclined to ascribe that ability to the individual. Note that holders of world records in athletics often attain their records on only one or a very few out of many trials;

yet, we are willing to grant these people ability to attain the world record even if success is attained only occasionally. In such cases ability is defined in terms of maximal performance. As we shall see, however, this is not the way in which ability is best defined in psychometric terms.

Thus far we have considered physical strength only in terms of a single, narrowly defined task – lifting a 100-pound barbell. But this would hardly be a particularly difficult task for many individuals, although it might be difficult for many others (say, young children). Therefore, ability or lack of ability to lift a 100-pound barbell might tell us little about an individual's physical strength. If we want to ask "how strong is this individual?" a more informative procedure would be to give the individual trials with barbells of different weights, using both light ones and heavier ones. This means that physical strength ability would now be defined in terms of a *class* of highly similar tasks, less narrowly defined than before, but still restricted to the barbell-lifting task. The measure of physical strength would come out of finding at what weight, in this series or class of tasks, the individual would start to have difficulty, in the sense of having a less than 100% probability of being able to lift that weight. (We might have to control for fatigue effects, by randomizing the trials with respect to the weights used, but for now let us ignore this problem.) There are considerations, from the technical discipline known as psychophysics, whereby it turns out that the most accurate or reliable measure of individual differences in strength would be at that weight where the individual has just a 50% probability of being able to lift the weight, for example, 225 pounds for a certain individual. From a series of trials with different weights, it would be possible, at least in principle, to estimate the point on the weight scale where an individual's probability of success would be 50%, and this would be the most reliable *quantified* measure of barbell-lifting ability, even though it does not indicate the *maximal* weight the individual might be able to lift.

With this procedure, it would be possible to compare different individuals for barbell-lifting ability, and to form a statistical distribution of the measurements for different samples of individuals – individuals of different ages, genders, states of health, etc. The measurements, incidentally, would be given in absolute terms, that is, on what Stevens (1951) called a *ratio scale*; it would be reasonable, for example, to call an individual who has a 50% chance of being able to lift a 400-pound barbell twice as strong as one who has a 50% chance of being able to lift a 200-pound barbell.

It would be of interest, also, to plot curves of individuals' probabilities of successful weight-lifting performance as a function of the weights. A sample series of such curves for a few individuals might look something like those in Figure 1.1. I call such curves *person characteristic functions* (PCFs; see Carroll, 1990). The 50% points on these curves, referred to the baseline, are the measurements of weight-lifting ability; these may be called *liminal* or threshold levels. Thus, Curve A is for a comparatively weak individual, who has a 50% chance of lifting a

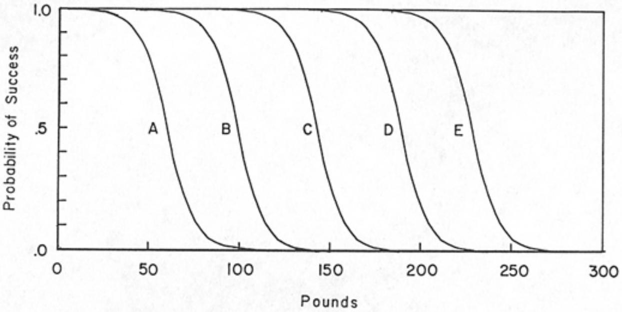

Figure 1.1. Illustrative Person Characteristic Function (PCF) curves for a barbell lifting task (hypothetical).

62-pound barbell and whose liminal level of weight-lifting ability is therefore 62 pounds. Curve E is for a comparatively strong individual – at least if we consider strength or weakness with respect to performance on the barbell-lifting task. This individual has a 50% chance of lifting a 228-pound barbell. The slopes of these curves are shown as fairly steep only because I assume they would have that degree of steepness; exactly how steep they could be is determined only by empirical investigation. In any case, the steepness of the curves would be associated with accuracy of measurement. If the curves were much flatter, this would imply that individuals' performances would be much more variable from trial to trial, even, possibly, that our barbell test of strength is for some reason not a good or appropriate one. Regardless of the steepness of the curves, they are in principle *monotonic descending*; that is, they always descend and never rise. It is unreasonable to expect that an individual who has a low probability of being able to lift a relatively light weight would nevertheless have a higher probability of being able to lift a heavier one.

The question now arises: How general is physical strength ability? It could be objected that barbell-lifting might be only a very specific skill. Common observation suggests, however, that a person who is very strong in the barbell-lifting task would also be strong in other physical strength performances, such as push-ups or pull-ups. This hypothesis could be examined by testing a group of people not only on the barbell task but also on a series of these other tasks. The group would have to exhibit some variability of performance on the barbell task; that is, it would have to include people whose abilities range from weak to strong on this task. The generality of physical strength ability would be judged from the correlations among the various tasks, that is, from the degree to which the individuals' measurements on the different tasks correspond to each other, or show similar rankings. If the correlations were all high – particularly if they were about as high as they could be in view of the reliabilities[1] of the measurements – we would conclude that physical strength ability is an ability that generalizes perfectly over the series of physical strength tasks that were employed. If the correlations were not significantly different from zero, despite high reliabilities of the measures, we would conclude that each task measures a different ability. If the correlations were significantly positive, but still somewhat lower than the maximal values they could attain in view of their reliabilities, we would probably conclude that there is a general physical strength ability measured to some extent by each task, but that each task measures, in addition, some rather restricted special ability. That is, we might infer that doing well in each task requires, in addition to a general physical strength ability, some special skill.

I am not aware that any thoroughgoing investigation of physical strength abilities like these is available (but see Fleishman, 1964). My guess is that the most likely outcome would be the last one mentioned in the preceding paragraph. Namely, there is a general physical strength ability measurable in a wide variety of physical tasks, but some tasks would require abilities more or less unique to themselves, possibly reflecting strengths or weaknesses in particular groups of muscles used in the tasks, or special strategies in performing particular tasks.

Still other questions would arise. How fixed is the ability? Does it vary over time? In the case of physical strength ability, common observation would suggest that over some short period of time – say a day or a week, or even longer – physical strength ability would not vary much at all, unless an individual takes special steps (like doing exercises) to improve his ability, or becomes subject to some debilitating conditions. Degree of physical strength ability can thus be regarded as a characteristic or trait of the individual, measurable at any particular point of time. It might vary somewhat more if measured at long intervals – in each of a number of years, say, in a mature adult. If considered developmentally – that is, if measured at each of a series of ages in a child – it would tend to increase, but the ranking of a number of children on this ability might tend to stay relatively the same, in which case we might infer that some developmental parameter of physical strength ability would be characteristic of each child. Such a developmental parameter could be calculated either on the basis of a series of measurements over a number of years, or possibly on the basis of noting a child's standing relative to those of a population of children of comparable age, physical size, or other attributes. In any case, such a developmental parameter would be a secondary, derived type of measurement, to be clearly labeled as such.

We have assumed here that an ability can be regarded as a *trait* to the extent that it exhibits some degree of stability or permanence even over relatively long periods of time. Many abilities do show this kind of stability. If an ability is found to be highly variable over time, a particular measurement of it would be best regarded as reflecting a *state* rather than a trait, just as a measure of a person's temperature might indicate presence of a fever.[2]

We are now in a position to define *ability* in a more precise way than before: *As used to describe an attribute of individuals,* ability *refers to the possible variations over individuals in the liminal levels of task difficulty (or in derived measurements based on such liminal levels) at which, on any given occasion in which all conditions appear favorable, individuals perform successfully on a defined class of tasks.*

In this definition, levels are specified as *liminal* (threshold) values in order to take advantage of the fact that the most accurate measurements are obtained at those levels.

Something needs to be said about the concepts of "task" and "defined class of tasks." Dictionary definitions of the word *task* do not adequately convey the characteristics and structure of what is intended here, and some connotations of the word (its association with the notion of work, the assignment of tasks by superiors, and the difficulty, tediousness, and/or aversiveness of tasks) are irrelevant. We may define a *task* as *any activity in which a person engages, given an appropriate setting, in order to achieve a specifiable class of objectives, final results, or terminal states of affairs.* It is to be understood, however, that "finality" is only relative; the end result or terminal state may only lead to another task, either a repetition of the same task or a different one. The specifiability of the end result of a particular task is crucial, however, because the individual performing the task must have some notion of what type of end result is to be attained and possibly of the criterion or criteria by which attainment of the end result is to be assessed. Many tasks are imposed by others, as when an individual is asked a question, presented with an item on a psychological test, or requested to perform some action. Many other tasks, however, are self-imposed, as when an individual decides to write a letter, sing a song, memorize a poem, or seek some action on the part of another person or a group. It can be the case that some kind of ability, and its level, could be inferred from an individual's successful or unsuccessful performance on any of such tasks, whether self-imposed or imposed by another.

By a *class of tasks*, we mean a group or series of possible tasks that have at least some identical or similar attributes. These attributes may refer to the kinds of stimuli that must be dealt with, the kinds of actions that must be performed, or the means by which those actions can be performed. The greater the similarities, it may be assumed, the more likely it is that the same or similar abilities of performance are involved. In the illustration used above, a highly similar group of tasks was that represented in the series of tasks utilizing barbells of different weights; the only difference among the tasks, ideally, would be the weights used. A less similar group of tasks was that represented by the barbell task, the push-up task, and the pull-up task, taken together; their similarity consists mainly in the fact that they require some kind of muscular strength to achieve the desired end results.

It may be assumed that tasks vary in *difficulty*, that is, in the probabilities that individuals will be able to perform them. When the tasks vary in only one parameter, like the series of tasks involving weights on a barbell, it is possible to determine a liminal probability by giving trials with different values of the parameter. When the tasks differ in their parameters, but still can be found to tap the same ability, a measure of an individual's ability could be obtained only by somehow aggregating the measurements on a series of tasks. For example, if general physical strength were to be measured with three tasks – a barbell task, a push-up task, and a pull-up task, – the final measurement could be obtained from some function (for example, a weighted sum) of the scores on these three tasks.

It is on the basis of these concepts that we can begin to see how various human abilities may be defined and measured. In common parlance we may speak, for example, of musical ability, athletic ability, and learning ability. In each case, it is presupposed that a particular class of tasks is involved. Nevertheless, it would be recognized that some abilities could be more narrowly defined, with a corresponding restriction of the class of tasks to which they apply. For example, although there might be a rather general musical ability exhibited by relatively good performance in a wide variety of musical activities – singing, performing on one or more musical instruments, reading music at sight, composing music, etc. – one might recognize special abilities in each of these activities; that is, there is a special class of tasks called "singing," a special class of tasks called "playing classical music on the piano," and so forth, so that "singing ability" and "classical piano-playing ability" could be defined as somewhat separate abilities. Similarly, it is commonly recognized that there are different types of athletic abilities – in distance running, playing football, playing basketball, etc. – and that these abilities are only loosely related. People who are good distance runners are not necessarily good basketball players, and vice versa.

As we shall see, the investigations dealt with in this book can be regarded as attempts to identify abilities by systematically classifying different tasks with respect to the abilities they appear to require.

Cognitive Ability

Since this book is concerned with a class of tasks designated as *cognitive*, I must specify what I refer to by this word as used in the expression "cognitive ability."

Insofar as we have defined *task* as any activity that a person may engage in (or be made to engage in) in order to achieve a specifiable class of terminal states of affairs, and insofar as it may be assumed that the person must have a notion of what is to be performed, one might conclude that *any* task is automatically a *cognitive* task – even the task of lifting a barbell, or of digging a hole in the ground. By using the adjective *cognitive*, however, I mean to limit the range of cognitive tasks to those that centrally involve mental functions not only in the understanding of the intended end results but also in the performance of the task, most particularly in the *processing of mental information*. That is, a cognitive task is one in which suitable processing of mental information is the major determinant of whether the task is successfully performed. Although barbell lifting may involve certain kinds of processing of mental information (kinesthetic perceptions of the barbell's balance and one's grip on it, for example), successful performance of the task is determined mainly by the physical strength of the muscles involved, and thus we would not call it a cognitive task. In contrast, the task of repeating a series of digits (as in a memory-span test) is a cognitive task because it requires storing the digits and their order in short-term memory, and retrieving them, in addition to chunking or otherwise manipulating the materials to be repeated. I define a cognitive task, therefore, as *any task in which correct or appropriate processing of mental information is critical to successful performance*. A cognitive ability is *any ability that concerns some class of cognitive tasks, so defined*. At many points in this book we will be concerned with what kinds of mental information have to be processed or operated on in the classes of tasks associated with particular cognitive abilities. Here it is necessary to consider what *cognitive processes* are, and what kinds of processes are involved in mental information processing.

Cognitive Process

In general, a *process* refers to any action or series of actions by means of which something is operated on to produce some result. A *cognitive process* is therefore one in which mental contents are operated on to produce some response. These mental contents may be representations or encodings either of external stimuli or of images, knowledges, rules, and similar materials from short-term or long-term memory. The response may be either covert (generally unobservable) or overt (observable). In the context of mental testing, only observable responses are admissible as data, although it may be useful, to explain such data, to develop hypothetical constructs concerning covert responses.

Many cognitive tasks are complex, but can often be analyzed into distinct processes, stages, or components. Sternberg (1977), for example, has provided one possible way in which to analyze typical analogies tasks found on many intelligence and scholastic aptitude tests. He proposes (pp. 135–137) that such tasks (like evaluating the correctness of the verbal analogy *red*:*stop*::*green*:*go*, symbolized as A:B::C:D) can be analyzed into the following components:

Encoding: The process of translating each stimulus into an internal representation upon which further mental operations can be performed.

Inference: The process of discovering a rule, X, that relates the A term of the analogy to the B term, and storing the rule in working memory.

Mapping: The process of discovering a higher-order rule, *Y*, that relates the *A* term to the *C* term, and storing the result in working memory.

Application: The process of generating a rule, *Z*, that forms an image of the correct answer and tests it against the *D* term of the analogy.

Justification: The (occasionally necessary) process of deciding whether the *D* term of the analogy is sufficiently close to the image formed by the application process to be regarded as correct.

Preparation-response: The (control) processes of preparing to solve the analogy, monitoring the solution process, and translating the solution into a response.

Sternberg also developed experimental operations whereby the time taken by each of these processes, and the correctness or accuracy of the response, could be observed or estimated. In a series of experiments, he applied these operations to several types of analogy tasks. In one of his experiments, 16 subjects were given these analogy tasks along with several psychometric tests of reasoning and perceptual abilities. Although the data are limited to only 16 individuals and two tasks, I have pointed out (Carroll, 1980c, pp. 16–17) that the data suggested that these processes might generalize in certain ways over the tasks and psychometric ability measures. For example, measures of the encoding speed component and of the preparation-response component generalized over a verbal analogy task and a "People Piece" analogy task, as well as over three psychometric tests of reasoning. From evidence of this kind, it appears that cognitive abilities can sometimes be more sharply defined by associating them with classes of particular task components.

In pursuing the notion of cognitive task and attempting to analyze such tasks into processes that might be referred to particular abilities, it is convenient to define what I call an Elementary Cognitive Task (ECT), as follows:

> An elementary cognitive task (ECT) is any one of a possibly very large number of tasks in which a person undertakes, or is assigned, a performance for which there is a specifiable class of "successful" or "correct" outcomes or end states which are to be attained through a relatively small number of mental processes or operations, and whose successful outcomes depend on the instructions given to, or the sets or plans adopted by, the person.[3]

Any ECT, as studied in the laboratory, consists not only of the operations performed by the subject but also of a series of distinct operations performed by the experimenter: giving the subject instructions and any preparatory practice, the presentation of the stimuli in a specified procedure (including a specified time-schedule), and the observation, clocking, and evaluation of the responses. One aspect of a task's being cognitive is that the nature of the task can depend on the instructions given, even when stimulus presentations and other objective events remain the same. A simple example is the memory-span task: With the same stimuli (a series of spoken digits presented one per second, say), the outcome can differ depending on whether the subject is asked to repeat them in forward or in backward order, or with a numerical constant added to each digit. Strictly speaking, each such instruction makes for a different task.

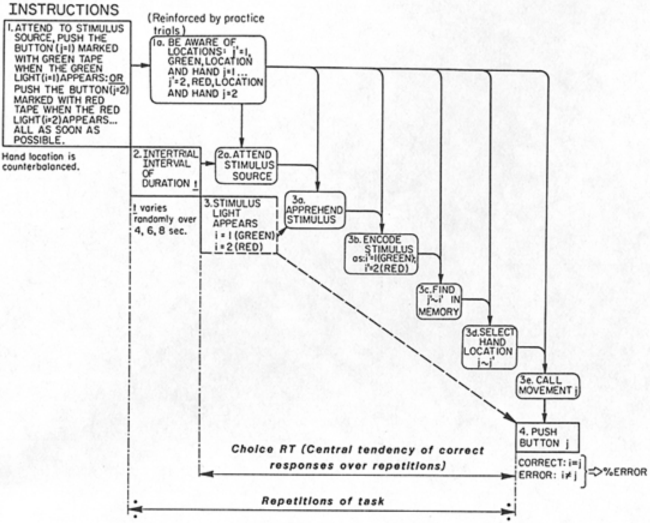

Figure 1.2. Dual time representation (DTR) for Keating and Bobbitt's (1978) choice RT procedure. Reproduced by permission from J. B. Carroll, 1980a, Figure 1, p. 15, *Individual difference relations in psychometric and experimental cognitive tasks.* Chapel Hill, N. C.: L. L. Thurstone Psychometric Laboratory Report No. 163.

In representing and analyzing ECT's (or, for that matter, any cognitive task), I have found it useful to develop a special kind of chart that I call Dual Time Representation (DTR). One such chart appears as Figure 1.2. It shows, along the diagonal, the objective events of the task, and in the space of the upper right triangle, the assumed cognitive processes performed by the subject, based on a logical and cognitive analysis of the requirements of the task. This particular task was one used by Keating and Bobbitt (1978) to measure choice reaction time. The subject was to press either a left-hand or a right-hand button depending on whether a red or a green light was the stimulus. I assumed, in this case, that the subject had to make a mental translation of the color code of the light to a code for the button position to be pushed, and that this mental translation process accounted for part of the reaction time. This mental translation process is shown in box 3c of the DTR chart. Note also that the measure of choice reaction time was taken from the onset of the stimulus to the onset of the button press.

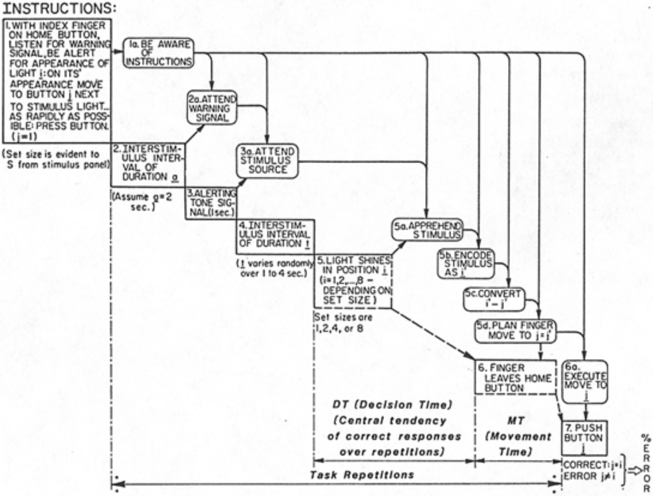

Figure 1.3. Dual time representation (DTR) for Jensen's (1980) RT-MT procedure. Reproduced by permission from J. B. Carroll, 1980a, Figure 2, p. 18, *Individual difference relations in psychometric and experimental cognitive tasks.* Chapel Hill, N. C.: L. L. Thurstone Psychometric Laboratory Report No. 163.

This chart is to be compared with a DTR chart (Figure 1.3) for a choice reaction-time task studied by Jensen (1980, pp. 688ff.) There are two important differences between the Jensen and the Keating and Bobbitt tasks, even though both were designed to measure choice reaction time. One difference has to do with what cognitive processes may be assumed to operate in the respective tasks. In Jensen's task, the subject has to move his/her finger to press a button just next to the *position* of a light, there being different numbers of light positions in the several experimental variations of the task. If there was any cognitive process in translating a light position to a hand movement code, it was very simple. Nevertheless, it is shown as box 5c of the DTR chart. In the Keating and Bobbitt task, however, choice of position depends on a translation of a color code into a position code, and it may be assumed that this translation would take more time than simply moving toward a light. The second difference between the tasks consists in the method of measuring times. In Jensen's task, reaction time was taken from the onset of the stimulus to the subject's finger leaving a "home" button, and was what may be called *decision time. Movement time* to move from the home button to a light was separately measured. In Keating and Bobbitt's task, reaction time, measured from the onset of the stimulus to the button-pushing, included both decision and movement time. Jensen was able to measure individual differences in both decision and movement time – largely unrelated, while Keating and Bobbitt were able only to measure individual differences in the sum of decision and movement times.

While it is debatable whether it is appropriate to distinguish decision and movement times (see, e.g., Smith and Stanley, 1983), or more precisely, whether "decision time" is really a valid measure of how long it takes an individual to make a decision, separate measurement of decision and movement times is nevertheless often operationally useful in order to refine the definition of whatever cognitive ability is being measured. The construction of a DTR chart for any task being studied is also useful because it prompts the investigator to specify the cognitive processes that might be involved in ability measurements.

Most measurements of human ability are based on performances of individuals on psychological or educational tests.[4] Each such test, as I have pointed out (Carroll, 1976a), can be regarded as a collection of relatively similar cognitive tasks, many of which can also be considered elementary cognitive tasks because they involve a relatively small number of cognitive processes. The cognitive processes can, however, take many forms. Some of these processes are involved in understanding, or learning to understand, the requirements of a task – the types of stimuli to be presented, what is to be done to them, what kinds of responses are to be made, and (sometimes) the time allowed for performance and how the responses are to be scored or otherwise evaluated. These and still other processes are involved in the actual performance of each task; some of these can be very elementary processes such as attending to stimuli, encoding or recognition of stimuli, comparison of stimuli, transforming representations of stimuli, retrieving associations, executing responses, and monitoring one's performance. Some processes may depend on prior experience, learning, or knowledge of particular facts, procedures, algorithms, or rules. Particular abilities in the execution of different processes or in the knowledge of particular facts or procedures may be the source of individual differences in any one or more of these processes, but ordinarily the tasks are constructed in such a way as to emphasize only one or a few of such processes and abilities. That is, some of the required processes may be so simple as not to reflect possible individual differences, while other aspects of the task, such as variations in the familiarity of the stimuli, may cause striking individual differences to appear. Analysis of individual differences revealed by psychological or educational tests must appeal to detailed considerations of the types of processes and learnings called for by a test.

A psychological or educational test is meant to be administered in some uniform fashion to all individuals who are to take the test. That is, the procedure is meant to be applied according to some consistent set of rules for presenting tasks – even if different tasks are administered to different individuals depending on their responses, as can occur in what has been called "adaptive" or "tailored" testing (Lord, 1980; Weiss, 1983). Many tests are administered with time limits for the total test or for subtests; specified time limits constitute part of the standard procedure for administering the test, even if this means that there will be variations in what tasks or items different examinees attempt. A problem with this widely used procedure is that the scores may not adequately distinguish between rate of work (speed in task performance) and overall level of mastery (how accurate the examinee would be if allowed to attempt every item, without time constraints) (see further discussion in Chapter 11).

Some measurements of cognitive ability – appearing as variables in certain investigations – are not based on psychological or educational tests; instead, they can be based on judgments or ratings (either self-ratings or ratings by others), or on counts or evaluations of outputs (e.g., by noting the number of pages per year a writer publishes). In any case, the measurements can be viewed as resulting in some way from the performance of cognitive tasks, even if the tasks are self-imposed and far from "standardized" like a psychological test. Under many conditions, such measurements can have at least some validity in assessing given aspects of ability.

Most measurements are of ability assessed *at a particular point of time.* Nothing can be said about the fixity or over-time stability of an ability unless appropriate information is assembled on this matter. Even for a measurement taken at a particular point of time there must be concern with reliability, that is, the accuracy of a score. This is a matter that is dealt with in classical theory of measurements and will not be discussed here because it is well treated in many standard textbooks. It may be pointed out, however, that unless reliability is determined with "split-half" or internal consistency measures, as opposed to the use of equivalent measures or "alternate forms," coefficients of reliability inevitably have some implications for the possible over-time stability of ability scores, because the equivalent measurements must be taken at different points of time.

Analysis of the cognitive tasks presented in psychological ability measures shows that many of the processes involved in them also occur in various "real-life" situations. It is for this reason that measurements from psychological ability tests often show correspondences with individuals' performances – successful or unsuccessful – in various real-life situations, including schooling, occupation, and career. Some abilities appear to be more important or crucial in real life than others. This book deals with a very wide variety of abilities – that is, all that can be demonstrated from empirical studies, regardless of whether their importance in real life can be shown. This is because the book attempts to further the science of human abilities. That is, it seeks to contribute to the program of studies recommended by Carroll & Horn (1981), and not merely to investigate the more obviously important abilities. We cannot adequately appraise the importance of different human abilities until we have mapped the whole spectrum of those abilities.

Abilities, Aptitudes, and Achievements

I regard the term *ability* as entirely neutral and even uninformative as to whether any given ability is an "aptitude" or an "achievement." It is a term that refers only to variations in individuals' potentials for present performance on a defined class of tasks.[5] It is obvious that performance on any task, at any stage of life except possibly very early infancy, is affected to some extent by prior learning and experience. In this sense, any measurement can be regarded, trivially, as a measure of some kind of achievement, that is, as a measure of the extent that certain behaviors have been learned. On the other hand, level of performance, as reflected in a cognitive ability measurement, may be affected (in childhood and adolescence) by the individual's level of constitutional maturation or (at any age, but particularly in later adulthood) by the health of the individual's central nervous system. Further, it is possible that levels of performance can be associated to some extent with the genetic constitution of the individual. Behavioral geneticists (e.g., Plomin, DeFries, & McClearn, 1990) cite evidence that genes are implicated in "general intelligence" or general mental ability, and possibly in other abilities.

To the extent that cognitive abilities are at least relatively stable and relatively resistant to attempts to change them through education or training, and at the same time are possibly predictive of future success, they are often regarded as *aptitudes*. (Here, I use the term *aptitude* in a relatively narrow sense, i.e., to refer to a cognitive ability that is possibly predictive of certain kinds of future learning success; I exclude the notion of aptitude as interest in and motivation for a particular activity.) Conditions for regarding an ability as an aptitude are specified in Figure 1.4. That is, an ability is clearly a measurement of aptitude for some particular future learning success if, in a sample of individuals tested both in aptitude and in achievement in some specified learning or training activity at two points of times, once before training (time *A*) and once after training (time *B*):

1. There is reliable variance in the measure of aptitude at time *A*.
2. There is no reliable variance in achievement tested at time *A*, because no learning has occurred.
3. As a consequence of condition 2, above, there is no significant correlation of aptitude and achievement at time *A*.
4. No significant change in aptitude is observed from time *A* to time *B*.
5. Significant change in achievement is observed from time *A* to time *B*, with reliable variance in achievement at time *B*.
6. There is a significant correlation between aptitude measured at time *A* with achievement at time *B* (trivially, this will be the same as the correlation between aptitude and achievement both measured at time *B*).

Cases in which such conditions obtain, at least approximately, in the use of aptitude tests for predicting success in foreign language learning have been

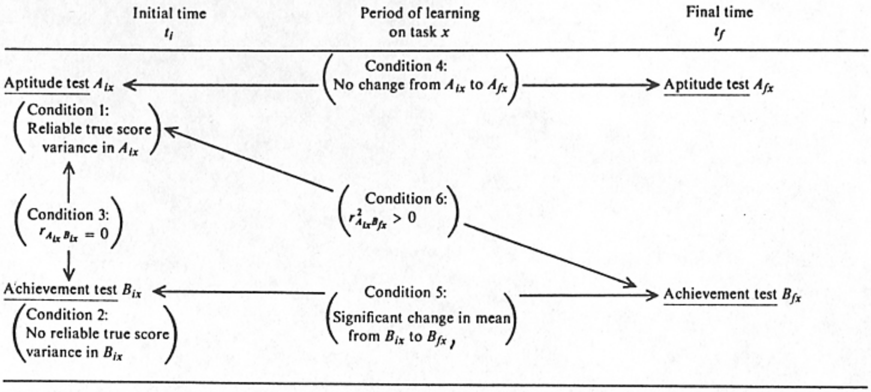

Figure 1.4. Ideal conditions for an aptitude test. Reproduced by permission from Carroll, 1974, Figure 9.1, p. 289, in D. R. Green (Ed.), *The aptitude–achievement distinction.* Copyright (C) 1974 by CTB-Macmillan/McGraw-Hill, Monterey, CA.

discussed by Carroll (1974). Cases are also to be found in studies of musical aptitudes (Stanton & Koerth, 1930), where it is shown that musical training does not significantly affect the aptitude measures even though they are significantly predictive of success in musical training.

Because it is seldom that *all* these conditions obtain, we can relax them somewhat, still retaining condition 5, by asserting that an ability measured at time *A* in a suitable sample of individuals is an aptitude if it contributes significantly to the prediction of achievement at time *B* over and above the prediction obtainable from a measurement of achievement at time *A*. That is, an ability is an aptitude if it helps in predicting degree of learning beyond a prediction from degree of prior learning. (It is sometimes said, in textbooks on psychological measurement, that a particular measure can be either a measure of aptitude or a measure of achievement, depending upon its use, but the account just given provides, in my view, a more accurate means of identifying an aptitude.)

Generally, I reserve the term *achievement* to refer to the degree of learning in some procedure intended to produce learning, such as a formal or informal course of instruction, or a period of self-study of a topic, or practice of a skill. Tests designed primarily to measure such degree of learning are measures of achievement, in addition to being measures of ability. (When *aptitude* is used to refer to any measure that is predictive of future performance, measures of achievement, in the narrow sense, could also be viewed as measures of aptitude to the extent that they might be found to be predictive of future learning progress.[6]

THE ISOLATION AND IDENTIFICATION OF ABILITIES

Given the definitions and discussion of the previous section, we can now ask what methods might be used to discover, identify, and differentiate human cognitive abilities.

Suppose we knew nothing about cognitive abilities or their measurement. A start toward identifying such abilities could presumably be made by inventing a wide variety of cognitive tasks. In view of the theory of ability offered above, it would probably be most effective to establish sets of tasks that would be highly similar within sets and in many ways different across sets. Within each set of tasks, tasks would differ only in attributes that might make them of differential difficulty – in the sense that different numbers of persons in typical samples would be expected to perform them successfully or with high levels of performance. The goal would be to establish that each given group of tasks would measure a single ability (or cluster of abilities). Administration of these tasks to typical samples of individuals would generate data analyzable in various ways. A preliminary sifting of the data would determine the difficulty levels of the tasks and permit elimination of any tasks that for some reason failed to show adequate discrimination among individuals with different degrees of average success on the tasks. An attempt would then be made to demonstrate that the data for a given set of tasks conform to a model whereby the tasks in that set could be regarded as measuring a single ability. One way of doing this might be to construct person-characteristic functions (PCF's) for groups of persons of different average levels of performance; if these curves showed characteristic shapes (like those illustrated in Figure 1.1) and at the same time differed markedly in their positions along the task difficulty baseline, the tentative conclusion might be reached that all the tasks in the set measure the same ability. Preliminary information on the nature of that ability would arise from an analysis of what task attributes are similar across all tasks in the set and what attributes of the tasks are associated with different levels of task difficulty.

Suppose, now, that we have established a series of task sets such that the tasks in each set conform to a model of a single ability. The question now arises: Do all these sets measure different abilities, or do they at least to some extent measure certain abilities in common? This question cannot be definitively answered solely from examination or logical analysis of the tasks in these sets. It can be answered only by analysis of empirical data obtained by administering all the task sets to a substantial representative sample of individuals and investigating the extent to which performances on different task sets correspond to each other. Ideally, the individuals ought to be comparable or homogeneous in any attribute, such as age, sex, or amount of education, that might be conceived to be associated with levels of performance in a similar way across task sets. (Otherwise, corresponding levels of performance across task sets might indicate only the common influences of such variables.) If performances on different task sets show no systematic correspondence, the conclusion would be that each task set measures a different ability from the abilities measured by the other task sets. If performances on different task sets show correspondences, the conclusion would be that there are one or more abilities that are common to two or more task sets. Just what abilities are measured by the several task sets would have to be determined by more detailed analysis.

Actually, the above paragraphs represent an attempt to give a simple, nontechnical description of the operations that investigators have been carrying out for some decades to discover and identify different abilities. It may be convenient to list here for the reader the correspondences between the nontechnical phrases we have used and the more technical terms that will generally be employed subsequently:

Nontechnical phrase	*Technical term or phrase*
Task sets	Psychological tests
Establishment of task sets measuring single abilities	Construction of psychological tests
Preliminary sifting of data	Item analysis
Correspondences between data from different task sets	Correlations among scores on different psychological tests

Procedures of test construction and of correlational analysis have become increasingly technical over the years. It is beyond the scope of this volume to elaborate on procedures of test construction, but certain suggestions arising from our analyses are made about these procedures in Chapter 18. Correlational analysis is treated in considerable detail in Chapter 3, but at this point it is useful to explain certain points about its logic.

Various measures of statistical association and dependence are available (Carroll, 1961). The Pearsonian correlation coefficient is the most frequently used measure, however, and this discussion assumes its use. It is often said that this coefficient is an indication of the degree to which two variables measure "the same thing," i.e., the same ability or the same combination of abilities. This statement would be completely correct only if the coefficient takes an absolute value of 1.00. If the coefficient is other than this value, a more accurate statement would be that it indicates the degree to which there is similarity or overlap in the abilities called upon for different members of the sample on which the correlation is computed.

An illustration of this can be offered in the following context. Suppose we have six measures:

1. A measure of visual acuity (such that an individual with 20/20 vision would get a high score, while blind individuals would receive a very low score, say zero).
2. An alternate measure of visual acuity that would be very highly correlated with variable 1.
3. A measure of hearing ability (such that an individual with no hearing loss of any

kind would receive a high score, while deaf individuals would receive a very low score.

4 An alternate measure of hearing ability that would be very highly correlated with variable 3.

5. A measure of performance success in a situation in which it is required to detect and locate stimuli with both visual and auditory components – for example, detecting and locating a bird singing and flying around high up in the trees in a forest. Thus, it would be possible to get a high score *either* by having excellent vision *or* by having excellent hearing, but not necessarily both. A low score would be obtained by individuals who have *both* poor vision and poor hearing. Let us call this measure "Bird task *A*."

6. An alternate measure similar to variable 5, such that its correlation with variable 5 is very high. Call this measure "Bird task *B*."

Suppose, now, that we obtain measures on all six variables on a sample of people 25% of whom have both vision and hearing (though with varying degrees of these abilities), 25% have vision but are deaf, 25% have hearing but are blind, and 25% are both deaf and blind. Further, we find the following correlations among these measures:

Variable		1	2	3	4	5	6
Vision *A*	1	1.00	.95	.00	.00	.50	.50
Vision *B*	2	.95	1.00	.00	.00	.50	.50
Hearing *A*	3	.00	.00	1.00	.93	.48	.48
Hearing *B*	4	.00	.00	.93	1.00	.48	.48
Bird task *A*	5	.50	.50	.48	.48	1.00	.80
Bird task *B*	6	.50	.50	.48	.48	.80	1.00

The .95 correlation between variables 1 and 2 could be regarded as an indication of the accuracy with which vision is measured by either variable.[7] Similarly, the correlation between variables 3 and 4 could be taken as an indication of the accuracy with which hearing ability is measured by either of these variables. These high correlation coefficients establish that vision and hearing are well measured by the respective variables. Cross-correlations between variables 1 and 2, on the one hand, and variables 3 and 4, on the other, are all zero, indicating that in this sample there is no correlation between vision and hearing abilities, or more precisely, that there is no overlap between the abilities measured by the two sets of variables. The correlations of .50 between variables 1 and 2, on the one hand, and variables 5 and 6, on the other, indicate some degree of overlap between the abilities measured by the two sets of variables. One might conclude that at least some of the sample perform the bird-detection task by virtue of their vision, whatever their degree of visual acuity, and regardless of their hearing ability. Similarly, the correlations with values of .48 between variables 3 and 4, on the one hand, and variables 5 and 6, on the other, show that at least some of the sample perform the bird-detection task by virtue of their hearing. Now consider the correlation, .80, between variables 5 and 6. From one point of view, it could be taken to be the reliability of either of these measures. Possibly the tasks on which they are based are constructed in such a way that whether the singing bird is detected is to some extent a matter of chance, regardless of the abilities required or used, thus contributing to a lack of perfect reliability. From another point of view, the correlation might be considered as indicating the extent to which the tasks measure the same ability or complex of abilities for every individual in the sample. Obviously they do not, however. What ability or abilities the tasks measure depends upon what combination of abilities a given individual possesses. If the individual has visual acuity ability, it may measure that ability if the individual detects the stimulus by seeing it. Similarly, if the individual has hearing ability, it may measure that ability if the individual uses it in performing the task. For individuals who are reasonably good in both vision and hearing, performance on the task provides no information on whether the task is performed by vision or by hearing. The correlation provides only an indication of the extent to which the task provides equal opportunities for the individual to use whatever abilities he/she possesses. This characterization of the meaning of a correlation may be generalized to all correlations between individual measures, *a fortiori* to all correlations in the above correlation table. Thus, the correlation of zero between variables 1 and 3 says that tasks 1 and 3 do not provide equal opportunities for a subject to use the same ability in performing them; in fact, it says that they possibly require different abilities. The correlation of .50 between variables 1 and 5 indicates that there is only partial equivalence between tasks 1 and 5 in providing for use of the same ability, and then possibly only for some individuals in the sample.

For this reason, identifying abilities from correlational data is often a complex problem, particularly when the tasks investigated permit use of different abilities in different combinations for different individuals. However, when relatively "pure" measures like tasks 1–4 in the above example are included in the analysis, tasks that might be demonstrated to measure single abilities by various operations of test construction and analysis, such as analysis of person-characteristic functions, the abilities become more apparent. (Thus, for task 1 the PCF's would have high slopes, and their positions on the ability difficulty scale for different individuals would be demonstrated to be dependent only on certain relevant visual attributes of the stimuli.)

In general, sophisticated correlational analysis can make use of the technique of factor analysis. Discussions of this technique and its development are to be found in Chapters 2 and 3. It may be noted at this point, however, that even factor analysis is not invulnerable to difficulties of the type found in the illustrative example shown above. What the example shows is that correlational analysis, including factor analysis that is based on correlations, should be regarded as a classificatory procedure, in that it classifies tasks with respect to the abilities that are called on *for at least some members of the sample.* If a factor analysis were performed on the correlations in the example, it would show two factors – a vision factor and an audition factor. Variables 5 and 6 would have substantial loadings on both of these factors, but such a result should not be interpreted to mean that these variables measure both vision and audition equally for all members of the sample. On the other hand, it would be possible from the data to determine separate scores on audition and vision for each individual, and to make a more precise determination of the meaning, for any given individual, of performance on the bird-detection tasks.

Abilities, Factors, and Latent Traits

Implicit in the above discussion is the assumption that an ability can be measured by any of a number of substantially correlated measures. Thus, in the example presented above it was assumed that visual acuity could be measured by either of two measures, and that this ability could underlie, at least to some extent, performance on the bird-detection task. If there are individual differences in performance on a task, or better, a set of tasks that conforms to a model whereby a single ability is measured, the individual differences can be said to be the immediate manifestation of individual differences in an underlying ability or *latent trait.* Further evidence for the existence of a latent trait derives from a demonstration that a number of similar task sets are highly correlated, or in factor-analytic terms, have weights on the same *factor.* A factor, if it is well established in a number of empirical investigations, is in essence a latent trait reflecting differences over individuals in ability characteristics or potentials.

It has been argued by some writers (e.g., Gould, 1981) that to speak of an ability, or a factor of ability, is to reify it. It is not clear what is meant by reification in this case, or what is supposed to be objectionable about it, although it appears that Gould is uncomfortable with the notion of ranking that accompanies it. Apparently the transgression imputed to psychometricians is that of assuming that ability is a thing or entity that somehow resides in the individual. No such assumption need be made. It would in fact be highly naive to make such an assumption, because to do so would entail an overly simplistic view of how the organism functions.

Consider, for example, the notion of physical strength ability, which we have mentioned as a possible latent trait or factor of ability. From what we know about the functioning of the human body and its parts, individual differences in physical strength ability have their source in the fact that individuals differ in conditions of their body build and musculature. It would not be necessary to specify the exact nature of these conditions, although physiologists and sports medicine specialists probably find it useful to do so. Physical strength ability, considered as a latent trait, functions at least as an *intervening variable*

(MacCorquodale & Meehl, 1948) that is useful as a parameter for describing and predicting individual differences in performances of various physical tasks.

A similar logic would apply in the case of cognitive abilities of various kinds. The definition of an ability arises from systematic observations of individual differences in performances on defined classes of tasks. These observations constitute the empirical basis of ability measurement. They require no assumptions or exact knowledge about neurophysiological functions that might be responsible for performance levels, although specialists outside the strict field of psychometrics may find it possible and useful to seek such knowledge. In any case, reification is not an essential or characteristic feature of the process of defining a cognitive ability. For our purposes, a cognitive ability can be viewed as an intervening variable, i.e., a calculational convenience, as it were, in linking together a particular series of observations.

Relations between Tasks and Abilities

The performance of any task, it can be assumed, calls on whatever ability or abilities it requires. Most, or perhaps all tasks, require more than one ability. For example, in a printed English vocabulary test, a subject might be asked to check "Yes" or "No" in response to being asked whether two words are opposites. Successful performance might be analyzed as depending on at least four abilities: (1) being able to read and recognize the words; (2) knowing the meaning of the words; (3) being able to evaluate their oppositeness, and (4) being able to make a check mark with a pencil. One could conceive of individuals with any of the $2^4 = 16$ combinations of these abilities, each ability being considered to have values of 0 or 1. A speaker of Chinese learning English, for example, might have no ability to read the words, yet know the words (in spoken form), be able to evaluate their oppositeness, and be able to make a checkmark. An English-speaking fourth-grader, in contrast, might be able to read and recognize the words, and be able to make checkmarks, yet not know their meanings well enough to evaluate whether they are opposites. Both individuals, one would assume, would fail to perform the task correctly, except possibly by guessing. Aside from the possibility of guessing, the only individuals performing the task correctly would be those having scores of 1 in all four abilities. In practice, however, only certain abilities are crucial to correct task performance, in the sense that individual differences in those abilities would be likely to be present in typical samples of individuals for whom the test is applicable. It is unlikely that a printed English vocabulary test would be presented to speakers of Chinese who had not learned to read English words; ordinarily it would be presented only to speakers of English who had learned to read and recognize words. It could be assumed that the ability to evaluate the oppositeness of the words would be closely associated with knowledge of the meanings of the words. Thus, in practice the only ability crucially required by the task would be that of knowing the word meanings (in this case, knowing both word meanings). All the operations of item (task) construction, item analysis, and determining ability dimensionality would be based on the assumption that knowledge of word meanings is the only ability in which individuals tested or likely to be tested differ significantly. Of course, if this assumption does not hold, the test might still require abilities other than knowledge of word meanings. This type of possibility – that a test may measure abilities others than those intended – is always worthy of consideration in interpreting test data.

At the same time, confusion can arise if one focuses more on the actual tasks assembled in a test than on the underlying ability. This is an issue that bears on matters of test use in predicting educational or occupational success. Suppose, for example, that it has been determined that vocabulary knowledge or general verbal ability is significantly predictive of success as a business executive. Such a determination would have been made, let us say, by establishing substantial correlations between a well-constructed vocabulary test and ratings of success in business administration. The vocabulary test would have been found to conform to criteria for measuring a single ability, being composed of a series of items varying in difficulty and producing satisfactory person characteristic functions.

Suppose, however, that a critic were to charge that this test was invalid or inappropriate because it contained items that required knowledge of words that are never needed or encountered in business administration. Such a critic would be confusing knowledge of particular items with the underlying ability being measured by the test. It is the underlying *ability* that is relevant to success in business administration, not the knowledge of particular items. When it is necessary, as it often is, to measure the upper reaches of vocabulary knowledge with words rarely encountered in business administration, this is a consequence of the way in which people acquire and organize their knowledge of words and their meanings; it has little if anything to do with the use of those words in business administration.

Similar logic applies to the use, in scholastic aptitude examinations, for example, of tasks that may seem to have little direct relevance or use in the college/university learning situations for which such examinations are claimed to be predictive. Assuming the appropriate evidence has been assembled (as it has for the College Board Scholastic Aptitude Examination, by Donlon, 1984), the underlying ability or abilities measured by these tasks are what is relevant to the predictiveness of the examinations, not the particular tasks contained in them.

Perhaps this point can be clarified and reinforced by noting that nobody would challenge the use of opticians' Snellen letter charts in appraising visual acuity, even though performance in reading such letter charts is unlikely to be directly involved in occupations in which high visual acuity is required.

THE IMPORTANCE OF COGNITIVE ABILITY STUDIES

The history of the study of cognitive abilities is presented in Chapter 2, but it is useful to consider here why cognitive ability studies are important in larger contexts of society.

For several thousand years – even in classical Greek and Roman times, and among the ancient Chinese (DuBois, 1970) – it has been recognized that there are individual differences in cognitive abilities, and that these differences have something to do with the roles and behaviors of individuals in society.

In the sixteenth century, the Spanish scholar Juan Huarte de San Juan (Huarte, 1575) examined the concept of intelligence. According to Linden and Linden (1968, p. 2), Huarte "invested the term with what today might be called *productive imagination*" and distinguished three characteristics of intelligence: "(1) docility in learning from a master, (2) understanding and independence of judgment; and (3) inspiration without extravagance." Huarte also recognized that some special aptitudes exist. According to Franzbach (1965), Huarte's ideas on the nature of genius influenced the thinking of various German writers and philosophers in the eighteenth and nineteenth centuries, including Goethe, Herder, and Schopenhauer, and his speculations about physical chararactistics of persons with different abilities and character traits found some echoes in the work of the early twentieth-century psychologist Kretschmer.

In England of the late nineteenth century, Galton (1869) revived interest in individual differences in intelligence that he thought were reflected in the different achievements of geniuses and of persons of lesser talent. Galton inspired a long tradition in British psychology to which such figures as Spearman (1904b, 1927) and Burt (1940) made contributions. Burt devoted many of his writings to correlations of intelligence with educational progress and success, occupational status, juvenile delinquency, and other phenomena of importance in society. To varying degrees, Galton, Spearman, and Burt were all convinced that intelligence was largely a matter of innate hereditary differences; that is, they believed that "nature" was more important than "nurture." This conviction crossed the Atlantic to become a dominant belief among many early American psychologists, like Terman (1916) and E. L. Thorndike (Thorndike et al., 1926). Purely as a scientific question, the nature–nurture issue has had its ups and downs throughout the history of psychology and the social sciences generally (Cravens, 1978). Already toward the close of the nineteenth century, the sociologist Cooley (1897) questioned Galton's conclusion that genius alone is sufficient to cause a person to rise to fame, but he appeared not to reject Galton's notion that genius is hereditary. The issue first became prominent in the 1920s, with Walter Lippmann's (1922) critique of intelligence testing, and again in the late 1930s with the claim of a group of University of Iowa psychologists (Skeels & Dye, 1939; see also Woodworth, 1941) that intelligence could be increased by special training and environmental adjustments. Most recently, the issue was again brought to the fore by Arthur Jensen's (1969) emphasis on the probable large role of nature in individual differences, particularly its possible role in white–black IQ differences. During the 1970s the so-called IQ controversy raged between adherents of either side (Block & Dworkin, 1974, 1976; Kamin, 1974; Eysenck & Kamin, 1981). At the present writing, the controversy seems to have come to a draw, neither side being willing to make concessions. Persons like myself, who are not directly engaged in active research on the nature–nurture issue, continue to believe that both genetic and environmental factors are important, but that it may be difficult to assess their relative importance with any exactness. In any event, the issue is of obvious social significance, because whatever the truth may be, it impinges on our concepts of equality of opportunity and social justice. The matter is so sensitive, Herrnstein (1982) claims, that the press and the media tend to suppress scientific findings that appear to favor hereditary factors in intelligence and cognitive abilities, and to feature any evidence favoring the role of training. Nevertheless, the press and the media have in the last several years shown keen interest in Bouchard's (1984) studies of twins reared apart – studies that now appear to support a strong genetic component in cognitive abilities (Bouchard, Lykken, McGue, Segal, & Tellegen, 1990). (For further discussions of the IQ controversy, see Aby, 1990; Cronbach, 1975; Kaplan, 1985; Snyderman & Rothman, 1988).

The role of mental abilities in education seems always to have been recognized, in the sense that learners have been classified as "apt" and "fast" or "inept" and "slow," and dealt with accordingly. Until modern times, those who appeared unable to keep up with instruction were often simply dropped from school. The work of the French psychologist Alfred Binet in introducing a series of intelligence scales (Binet & Simon, 1905) for assessing children's chances of school success is well known; this work was echoed in Germany, Spain, and many other countries, including the U.S. (Goddard, 1910; Terman, 1916). Volumes have been written about the role of mental abilities in education. The recent interest in issues such as the declines of mean scholastic aptitude test scores over the past several decades (Lipsitz, 1977; Advisory Panel on the Scholastic Aptitude Test Score Decline, 1977) and the levels of the nation's literacy (NAEP/ETS, 1985; Kirsch & Jungeblut, 1986) indicate public awareness of the importance of cognitive abilities in education and in the functioning of society and the economy.

The fairly strong correlation of intelligence with occupational status has been documented many times: on the basis of World War I mental testing data, by Yerkes (1921) and Fryer (1922); and on the basis of World War II data, by Stewart (1947). Gottfredson (1984) has made detailed analyses of the roles of intelligence and education in the division of labor. She presents evidence to show that (1) occupations differ in the general intellectual difficulty of the tasks they require workers to perform on the job, (2) the occupational prestige hierarchy reflects an ordering of occupations according to that intellectual difficulty level, (3) jobs that are higher in intellectual difficulty are more critical to the employing organization, and (4) large differences in intelligence in the population are evident by early school years and this distribution is not substantially changed by school or work environments. Recently (Gottfredson, 1986b), she has argued that it is virtually hopeless to expect that such differences can be circumvented in employee selection, even by extensive training programs, because while it may be possible to teach lower-ability persons certain job skills and knowledges (if enough time is taken to do so), it is practically impossible to teach the skills of good judgment and decision making that depend on level of intellect.

According to a reviewer in *Science*,

> The relation of innate intelligence and crime is a central concern in *Crime and Human Nature* [Wilson & Herrnstein, 1985]. That IQ scores are correlated with delinquency and with school performance has long been known, though often ignored or interpreted by criminologists as reflecting social factors of class and culture and responses to school contexts. After describing studies of the relations between IQ test scores and other indices of behavior, such as school performance, the authors conclude that intelligence plays a prior and independent role in association with crime.

This is later explained:

> Intelligence affects crime in that the individual of low intelligence is less aware of long-run consequences, less willing to defer present gratifications, and less able to restrict impulsivity (Gusfield, 1986, p. 413).

Undoubtedly there are many other areas of societal importance in which individual differences in cognitive abilities may be implicated to some extent, for example, consumer behaviors and extent and kind of television viewing, but any review of investigations of these topics is beyond the scope of this volume.

Nearly all studies of the societal correlates of cognitive abilities have focused on a very general mental ability, called *g* by Spearman (1927), and measured – though with some variations in exactly what is measured – by a wide variety of intelligence or IQ (intelligence quotient) tests. Likewise, most studies of the heredity–environment issue have utilized measures of general intelligence. It is the thesis of this book that there exist a substantial number of distinguishable and important mental abilities – as many as thirty or more. While it may well be the case that general intelligence – a recognized higher-order factor of cognitive abilities – is the most weighty element in all these relationships (Hunter, 1986), the possible importance of more specialized abilities cannot and should not be ignored. The predictive validity of several specialized abilities in military and occupational selection settings has often been demonstrated (Guilford & Lacey, 1947; United States Employment Service, 1970; Ghiselli, 1966.) My own work in foreign language aptitude strongly suggests that specialized abilities beyond general intelligence play an important role in learning a foreign language (Carroll, 1981b). It is commonly recognized that various specialized abilities in music and the arts are largely independent of general intelligence. Some evidence exists that there are genetic determinants for some special abilities, independent of any for general intelligence (Vandenberg, 1962; DeFries, Vandenberg, & McClearn, 1976).

Some special abilities are of such social importance that they have attracted the attention of historians. For example, in a book entitled *The Intelligence of a People*, Calhoun (1973) traces the development and importance of several special abilities – verbal and spatial-mechanical – in American social history since colonial times. Cohen (1982) has undertaken a similar task with respect to what she calls "numeracy," and Soltow and Stevens (1981) have chronicled the rise of literacy in America.

From all these considerations, it should be clear that the study of cognitive abilities is important from several social and practical standpoints. This volume attempts to present what is now known about cognitive abilities and its scientific basis. Social and educational policies need to take account of what cognitive abilities exist, how they are best measured, how they are formed, how they normally develop and change over the life span, how amenable they are to improvement, and by what means they can be improved, if that is possible.

NOTES

1. The reliability of a measurement is an index of its "accuracy," that is, the degree of agreement between successive applications of a measurement procedure. It is often expressed as the correlation, in a typical sample of cases showing dispersion of measurements, between scores or values obtained on two successive occasions, or from presumably equivalent measurement procedures. The correlation between any two variables A and B is reduced (in absolute magnitude) to the degree that the measurements of the variables lack perfect reliability. In theory, the maximum correlation that could be expected between variables A and B is the square root of the product of their respective reliability coefficients. See standard textbooks on psychological measurements for further details, particularly with respect to corrections for "attenuation."
2. According to the entry in Harré and Lamb (1983, p. 641), "a trait is a characteristic of a person... which varies from one individual to another.... Traits are conceived as reasonably stable and enduring attributes, distinguishing them from *states*, which are temporary behavioral predispositions." In personality theory, the concept of trait has been debated; some authorities reject the concept altogether, on the ground that behavior is so much situationally determined that there can be no stable traits. There is much evidence, however, that cognitive abilities exhibit a great degree of stability, even in diverse situations, and can therefore be regarded as traits. See Anastasi (1983) for further discussion. Note that the concept of trait as applied to cognitive abilities has no necessary connection with the concept of trait employed in genetics in discussing, e.g., dominant and recessive traits.
3. Some authorities dislike the expression "elementary cognitive task" because in their opinion it wrongly suggests that cognitive tasks can be "elementary" and simple when they are actually probably very complex, or that tasks can be decomposed into simple components when they are actually not amenable to such decomposition. In my usage of this expression, I only mean to suggest that some tasks are simpler than others; being "elementary" is a matter of degree. For example, a reaction time task studied in the laboratory is undoubtedly simpler than the task of writing a letter or essay. Whether

tasks can actually be decomposed into component processes is a matter for empirical determination.

4. Note that I imply a very broad definition of "test." The class of things that can be called *tests* includes not only the typical paper-and-pencil multiple-choice test that can be administered to groups of individuals at the same time, but also many types of individually administered examination schedules such as the Stanford–Binet intelligence scale or the Wechsler Intelligence Scale for Children (WISC). In his textbook on psychological testing, Cronbach (1990, p. 32) suggests that a test be defined as any "systematic procedure for observing behavior and describing it with the aid of numerical scales or fixed categories."
5. A reviewer has pointed out that the term *potential* implies future performance, and questions whether it is appropriate in this context. I use it only to refer to the idea that an individual's degree of ability on a particular dimension of ability implies a certain probability of success – that is, potential for success, in attempting *any* particular task involving that ability, depending, of course, on the level of difficulty of the task.
6. I strongly favor a distinction between aptitude and achievement, in opposition to a considerable body of opinion in the psychometric community to the effect that there is no useful distinction to be drawn. The issue was discussed extensively in a symposium whose proceedings were edited by Green (1974). In that symposium, I – along with several others – promoted the idea that aptitudes are basic characteristics of individuals that control rates of learning, while achievements are merely products of learning. Thus, in this context the distinction between *aptitude* and *achievement* bears on the nature of these constructs. This is in contrast to contexts in which variables – whether they be measures of aptitude or measures of achievement in the senses just mentioned – are considered to measure aptitudes to the extent that they are useful in predicting future performance.
7. If variables 1 and 2 were scores obtained by successive application of the same measurement procedure, their intercorrelation could be considered as a measure of the reliability of that measurement procedure. In this example, however, I envision two different measurement procedures, which could still have a high intercorrelation indicating that they tend to measure the same thing, namely, visual acuity.