4 Speed of information processing and population differences

Arthur R. Jensen

The purpose of this chapter is to propound the potential contribution of mental chronometry to the study of abilities in cross-cultural psychology. It informs researchers in this field of some of the techniques which have already proved useful in the study of individual differences within culturally homogeneous groups.

I use the term *cross-cultural* to refer to populations which differ in their symbolic systems, beliefs, values, and customs, without making any assumptions concerning the degree to which such cultural differences play a causal role in the variety of ability differences observed between particular populations. That is a question for empirical research. I would reject the assumption, which seems implicit in much cross-cultural research, that all behavioural differences between culturally different groups are attributable to, and wholly explainable in terms of, their cultural differences per se. The scientifically most defensible working hypothesis, I believe, is that the study of *all* human differences, in mental as well as physical characteristics, should be approached from a genetic–environment interactionist position. The culture of a population and its genetic structure are most plausibly a two-way process, each shaping the other in complex ways. It is difficult to imagine how cultural differences can be properly studied except within the broad framework of behaviour-genetic analysis, if our purpose is to go beyond the merely descriptive. Description of cultural environments and objective assessment of behaviour, however, remain crucial aspects of cross-cultural research. Chronometric techniques lend themselves to the assessment of virtually all variables that fall under the heading of "mental abilities." The choice of variables and techniques would depend upon the investigator's purpose.

To keep this chapter within the assigned limits, I shall focus on only the theoretical purposes of cross-cultural mental testing. The problems in this sphere are much more difficult than those involved in the practical use of tests, as in educational and personnel selection. The question of validity generalisation, when tests devised in one culture are applied in another, is a nontheoretical, empirical matter, if all we are concerned with is achieving practical predictive validity. The practical usefulness of a given test across different cultural groups can be evaluated by the standard psychometric methods, or a test may need to be markedly revamped in order to have practical validity in a different culture. Conceptually, the methods of applied psychometrics in test construction and validation remain fundamentally the same in different cultures, although many of the tests found to be most useful for similar purposes in different cultures may show very little superficial resemblance in form or content.

We face the really difficult, and largely unsolved but theoretically more interesting, problems when we focus on *construct validity* in cross-cultural testing. Essentially, this is the question of whether the *same ability* is measured in two (or more) different cultural groups, by whatever means are most appropriate in the particular culture. An even more difficult, but subsidiary, question is whether the same ability can be measured on directly *comparable scales* in different cultural groups. If so, it would mean that the scaled differences between individuals from two *different* populations are equivalent to the same-sized differences between individuals *within* either population. Both conditions are necessary if cross-cultural comparisons of abilities are to attain construct validity, rather than just practical predictive validity for a particular criterion. Confidence that the same ability is being measured in different populations is more easily achieved than confidence in the comparability of scales across populations. The latter condition is dependent on the former, but not vice versa.

One theoretical purpose of cross-cultural research, from the standpoint of differential psychology, is to discover those aspects of human behaviour, or the theoretical constructs "underlying" certain forms of behaviour, which are *invariant* across different cultures. By "invariant," in this context, I mean *structural invariance*, that is, invariant patterns of relationships among parameters in a limited behavioural domain, although the parameters' absolute scale values may vary between individuals or between populations. Human anatomy, for example, is replete with structural invariance across different racial groups, although numerous features statistically show *dimensional variation*. The concepts indicated by these two terms – *structural invariance* and *dimensional variation* – are also applicable to cross-cultural psychology, particularly in the study of mental abilities. Just as these properties are found in gross anatomy, they probably also apply to the fine structures of the brain. Hence, it is plausible to hypothesise that structural invariance and dimensional variation in the neural basis of behaviour are manifested also at the behavioural level. However, analysis of the brain–behaviour relationship is generally beyond the powers of unaided observation and must depend upon a number of technical developments to secure the necessary data.

The data derived from ordinary psychometric tests are often criticised, when used in cross-cultural research, on the ground that such tests reflect only the end *products* rather than the *processes* of problem solving, and that any given product or level of performance can arise in many different ways. This, in fact, is a common criticism of traditional IQ tests, even as they are used within a culturally homogeneous population. Obviously, ordinary test scores per se are too gross and far removed from the various cognitive processes that resulted in the scores to permit conclusions about structural invariance in cross-cultural studies. But analysis of item characteristics, such as the rank order of item difficulty, may reveal a considerable degree of invariance across populations. If many diverse test items maintain the same rank order of difficulty in two populations, it is presumptive evidence that the same process or processes involved in item responses are operating in both populations. If mean (or median) latencies (i.e., response times) of individual item responses also show the same rank order in both populations, and if item latencies are correlated with item difficulties in both populations, the presumption is greatly strengthened that the same cognitive processes are operating in both populations. These are examples of two types of evidence for structural invariance.

Two populations which are invariant for relative item difficulties, however, still might differ in the average efficiency of the hypothesised processes involved in item responses, which would be manifested as consistent differences in item difficulty between the two populations and, of course, in their overall test scores. It should be clear that at this level of analysis we are dealing entirely with descriptive phenotypic variables, which do not warrant inferences concerning the causes of the apparent information processing differences between populations. Special behaviour–genetic designs, such as cross-cultural adoptions and separated monozygotic twins, would be required for evidence of the relative influences of genetic and cultural factors (and their interaction) on the test performance in question.

Experimental studies in which certain abilities are trained up to asymptotic levels of performance may also be used to examine the claim that two given groups differ in some particular task performance entirely because of cultural difference in the amount of prior experience with the task. Asymptotic training is unfeasible for the kinds of complex items typically used in psychometric tests, and besides, asymptotic performance, if ever achieved by most persons, would result in near-perfect test scores by everyone, thereby precluding the possibility of measuring individual or group differences by means of psychometric test scores. This problem can be overcome by the measurement of response latencies to very simple items, performance on which is nevertheless significantly correlated with scores on complex psychometric tests. The items can be so simple that there is zero variance when performance is scored in terms of number of right or wrong responses. The only source of variance remaining is in the response latencies to each item, and these may be trained up to asymptotic levels in a fairly short time, possibly revealing stable individual or group differences in asymptotic performance – differences which may be correlated with differences on complex psychometric tests. Simple addition of pairs of integers is one example of a task which permits chronometric analysis at asymptotic levels of performance (Groen & Parkman, 1972).

Opposing hypotheses

Much of the cross-cultural research on abilities is influenced, explicitly or implicitly, by one or the other of two fundamentally different notions concerning the nature of human abilities. I shall refer to them as the *specificity-learning hypothesis* and the *structural-process hypothesis*, or, for short, the *learning* and *process hypotheses*.

The *learning hypothesis* holds that there is really only a single basic mental ability with which all humans are genetically endowed, namely, learning ability. (Whether individuals or populations are thought to be equally or unequally endowed is a separate issue; it is not intrinsic to either the learning or process hypothesis.) The apparent variety of abilities observed in performance and revealed by factor analysis of diverse tests is viewed as entirely a product of learning. According to this view, the structure of abilities discerned by factor analysis reflects only the structure of the environment in which learning occurs, rather than anything intrinsic to the learner. Original human nature, psychologically, consists of a homogeneous learning ability which gradually acquires whatever structure (i.e., interrelated skills and information) that is shaped by living in a given physical and cultural environment. Whatever invariance in cognitive structures there is across culturally different groups is the result of certain common cultural experiences and the many universal features of the physical environment. The specific-learning hypothesis thus implies virtually unlimited possibilities for the variety of abilities and their factor structure, just as a lump of clay can be moulded into innumerably different shapes in the hands of different sculptors. If abilities are entirely the result of context-specific learning, evidence of cross-cultural invariance in the structure of abilities must be viewed as due either to happenstance or to common cultural elements, rather than to common species-specific neural structures that have emerged in the course of human evolution. If various cognitive structures are not intrinsic to the human organism but are arbitrarily fashioned by the demands of a particular culture, we can take a completely relativistic view of mental abilities. "Intelligence," for example, would not be seen as a particular universal characteristic of humans, but only as whatever learned behaviours are most esteemed in any given culture. It has been suggested by some psychologists that intelligence, for Australian Aborigines, would be defined in terms of such specific adaptive skills as seeking sources of water in arid terrain, or tracking wallaby and felling them with a boomerang. As assessed by these criteria, Shakespeare and Newton might well be found mentally deficient. If we believe that context-specific learning is the whole basis of cognitive development, there would be little motivation in looking for invariance in cognitive structures across different cultural groups. Cross-cultural research would need only to describe similarities and differences in cultural environments. The only fundamental question of psychological interest, then, would be whether various populations differ in the primal learning ability itself.

The *structural-process hypothesis* agrees with the learning hypothesis only with respect to the *informational content* of learning. Of course, that is necessarily context-specific. It is like distinguishing between the specific words and syntax of different languages and the deep grammatical structures common to all languages. The main difference between the two hypotheses is this: The process hypothesis views learning as just one aspect of a number of cognitive processes which are built in or intrinsic to the organism. These distinct processes are a part of humans' "original nature," to use an old-fashioned term which is regaining currency. Just as the physical organism is not a homogeneous blob of protoplasm waiting to be shaped by the environment, so, too, mental abilities are not shaped out of a homogeneous capacity for learning. Rather, the evolution of the brain has resulted in a system of differentiated structures for information processing. In this view, individual differences and group differences can arise not only from differences in learning ability or differences in experience, but also from differences in the speed or efficiency of various component processes, such as stimulus encoding, short-term memory capacity, retrieval of information from long-term memory, and so forth. The process hypothesis invites investigation of the degree of invariance of processes (and their interrelationships) across populations from widely differing cultures. Does the set of processes that underpins an Ability factor (in the psychometric-factor analytic sense) in a given culture show up in the same configuration in a different culture? If the configuration of processes is invariant across cultures, we can then examine possible differences between populations in the efficiency of a given process or system of interrelated processes. How much of the variance in psychometric factors, such as the General Intelligence factor, or g, and in Verbal and Spatial factors, can be accounted for in terms of the efficiency or speed of the elementary cognitive processes that are the operational basis of these broad ability factors? Chronometric analysis of abilities lends itself to answering such questions. However, chronometric methods have been scarcely used by cross-cultural researchers, who have generally relied either on traditional psychometric methods or on experiments with specially devised tests or laboratory tasks which do not incorporate chronometric techniques.

General intelligence

Of all the factors identified by the factor analysis of various tests of mental abilities, the General Intelligence factor, which Spearman labelled g, has been the most prominent and the most important, theoretically and practically. It appears in virtually every battery of diverse tests when they are administered to samples that fairly represent the range of talent in the general population. The *g* factor shows more substantial correlations with other variables of importance in the real world than does any other single variable known to psychology. Its practical predictive power has been amply demonstrated, for example, in education and in personnel selection in business, industry, and the military (Jensen, 1984a). Two centuries ago, Adam Smith argued that the wealth of nations depends on the abilities and knowledge of their people. The *g* factor is central in this proposition. The perception of its importance for the economic welfare of a nation is seen in the recent establishment of a ministry of intelligence by the government of Venezuela, its mission being to raise the people's level of educability. This action suggests an awareness that educational achievement and its hoped-for effects on economic development and quality of life depend on something more than just the availability of schooling. The benefits of education depend heavily on characteristics of the pupils themselves, the prime national resource being reflected by the distribution of *g* in the pupil population. Other developing countries may well follow the lead of Venezuela in this concern.

The practical implications of cross-cultural investigations of *g* theory seem obvious. The prospect has probably been deterred by the close historical connection of *g* with traditional culture-loaded psychometric tests developed for use in western industrialised nations. Nevertheless, it remains a question of prime importance, theoretically and practically, whether the *g* construct is an identifiable characteristic of all *Homo sapiens* in *every* culture (Jensen, 1986, 1987a). The hypothesis that *g* is an intrinsic human trait is countered by the plausible claim that *g* is a cultural artifact: *g* reflects the all-positive correlations among diverse psychometric tests and this pattern of correlations reflects only the particular abilities, skills, and achievements that are most valued in modern industrialised societies and are inculcated by their educational and other cultural institutions. This hypothesis of *g*-as-artifact has gained popularity in present-day psychology. The hypothesis directly challenges the theory which originated with Francis Galton, that of intelligence as a general mental ability, a product of human evolution as a fitness character in the Darwinian sense. It should also be a challenge to cross-cultural researchers to test this theory. To prove that one and the same *g* can be identified in the populations of widely differing cultures would be a remarkable achievement. Whether proved or disproved, the knowledge gained in the attempt would profoundly advance our understanding of the nature of human abilities. For example, where do various abilities stand on the continuum between cultural specificity and species universality? Such research could obviously not be accomplished if we were limited to our ordinary psychometric tests. Chronometric data derived from various specific tasks devised so as to be equally appropriate within different cultures, however, may provide the needed common metric for cross-cultural analysis.

Time as a psychological variable

Every mental act, even the simplest imaginable, takes a finite amount of time, and the time taken is surprisingly long as compared to the amount of time taken by the sense organs and sensorimotor neural conduction per se. There are highly reliable time differences for various simple tasks which, subjectively, all seem of equally trivial difficulty. For example, it takes people significantly longer to add 5 + 3 than to add 5 + 2, and it takes university students, on average, more than 100 milliseconds (msec) longer to name a *glove* than to name a *chair*, and about 200 msec longer to name an *anvil* than to name an *anchor* (Oldfield & Wingfield, 1965). Such differences between the average speeds for naming various familiar objects are highly consistent across persons. (The differences in naming latencies are attributable not to word length or number of syllables but to the accessibility of the names in long-term memory, which is closely related to their frequency of use.) Moreover, there are highly reliable individual differences, even among young university graduates, in the average speed with which highly familiar objects can be named.

It is of considerable theoretical interest that virtually every type of measure of mental speed that has ever been used in experimental psychology has shown a correlation (in the expected direction) with psychometric intelligence in those studies which have included both chronometric variables and psychometric tests. This holds, however, only for chronometric tasks that are quite simple, in the sense of evoking response latencies not much greater than 1,000 msec. We are dealing here with extremely short time intervals that require great precision of measurement. They should not be trivialised by the fact that the differences between such brief time intervals are well below the threshold of our subjective awareness. One of the great advantages of mental chronometry is that it permits highly reliable measurement of sources of variance in performance, both between different tasks and between persons, that could not be discerned by any other means. Items requiring only the simplest forms of information processing would reveal absolutely no reliable variance whatsoever if performance on them were scored in the gross fashion of typical psychometric tests. Yet they are capable of yielding highly reliable measures of individual differences in terms of average response latencies.[1] For example, subjects are allowed 2 seconds

[1] Early studies of reaction time (RT) showed quite a low reliability of RT measurements for individuals. This was mainly because too few trials were given. The internal consistency reliability of RT, measured within a single test session, increases with the number of trials, closely in accord with the Spearman–Brown prophecy formula, and any desired degree of reliability, less than 1, can be obtained by increasing the number of trials. Test–retest reliability, when test sessions are one or more days apart, is generally lower than the internal consistency reliability, and differs according to the complexity of the RT tasks. Median RT (over trials), being less affected by outliers, is more reliable than mean RT. With 50 or more trials, median RT approaches the reliability of conventional psychometric tests. Internal consistency and test–retest reliability data on simple RT and choice RT (for varying numbers of choices) are reported by Jensen (1982b; 1987b; 1987c).

to read a brief statement on a video screen, such as "*A* after *B*," and then must respond with either *true* or *false* as quickly as possible upon the presentation of the letters *AB* (or *BA*). There are highly reliable individual differences in response latency, averaging about 500 to 600 msec in university students, and the latencies are substantially correlated with scores on untimed paper-and-pencil IQ tests (Paul, 1985). Chronometric paradigms do not depend on such crude indices of item performance as *can* versus *can't* or *right* versus *wrong*, but depend only on the time required to produce the correct response. An essential task requirement is that the correct response should be easily within the capability of all subjects for whom the test is intended. This is a crucial advantage in the study of the most fundamental components of information processing, measured by *elementary cognitive tasks* (ECTs), because these components are presumably possessed by all subjects and, of course, must be fully evoked by the ECT in order for their latencies to be measured at all.

Our chief interest in response latencies to various ECTs is that individual differences in the latencies, which can be measured with satisfactory reliability, are found to be correlated with various mental test scores and especially with psychometric *g*, the paramount factor in all IQ tests. A question of crucial theoretical importance is why reaction time (RT) to very simple tasks is correlated with performance on IQ tests comprising relatively complex items of knowledge, reasoning, and problem solving. The correlation seems counterintuitive to many psychologists. We naturally think of intelligence as being qualitatively different from, and of a higher order of mentality than, any such simple processes as could be reflected by reactions taking less than 1 second. Hence, we tend to appeal to superficial explanations, almost as if to ward off the possibility that mental speed is intrinsic to intelligence.

One such commonsense suggestion, for example, is that psychometric tests often have time limits or are "speeded," and therefore a Speed of Work factor enters into both the psychometric test and the RT task, making for the correlation between them. But there are several difficulties with this idea of a general Work-speed factor being responsible. Timed or speeded tests show no higher correlations with various RT tasks than do tests given with no time limit, in which subjects are urged to attempt every item and are encouraged to take all the time they need to finish the test (Vernon, 1983, 1985). The speed with which subjects perform on tests of intelligence is not correlated with their intelligence scores on the same or on other tests; thus work speed in the test situation is not a correlate of IQ as is RT to ECTs. Also, not all aspects of response latency in ECTs are correlated with IQ or with each other. We have found for some ECTs, for example, that when total response time is divided into RT (or *decision* time) and movement time (MT), the RT is more highly correlated with scores on an untimed intelligence test (*Raven's Progressive Matrices*) than with MT. Another line of evidence comes from looking at RTs on tasks that vary in complexity, such as *simple* and *choice* RT (SRT and CRT). In SRT, the subject responds to the occurrence of a single stimulus. In CRT, the subject makes different responses to the occurrence of two (or more) different stimuli. SRT reflects individual differences in sensory lag and motor skill common to *both* tasks, as well as differences in general effort, attention, and overall speediness of responding. Subtracting SRT from CRT gets rid of the time taken up by the noninformational aspects of task performance; the *difference* between CRT and SRT, therefore, reflects essentially the time required for processing one additional bit of information. There are reliable individual differences in the time increment of CRT minus SRT, and these differences are also correlated with IQ. All these findings are inconsistent with the idea of a general speediness factor in all types of performance, as the cause of correlations between RT and IQ. It is also noteworthy that measurements derived from the amplitude of the average evoked potential (see Chapter 3 by Eysenck), which is an involuntary cortical reaction to a "click," obtained while the subject is relaxed in a reclining chair, is correlated both with response latencies in ECTs and with psychometric *g* (Jensen, Schafer, & Crinella, 1981). In other words, response latencies that are correlated with various intelligence test scores are also correlated with a measurement of brain activity that bears absolutely no resemblance to the notion of work speed.

Another seemingly plausible explanation of the correlation between RT and IQ is in terms of speed–accuracy trade-off. Even some of the simplest ECTs can have a very low error rate (such as missing the response button or hitting the wrong button), amounting to as much as 4% or 5% of all trials. It has been shown in RT experiments that when subjects are specially instructed to maximise their speed, they do, in fact, make more errors in responding, and when instructed to maximise accuracy, they react less quickly. So it is quite conceivable that the more intelligent subjects may decide that the optimal strategy for performing an ECT is to sacrifice accuracy for speed. Brighter subjects would thereby show a higher error rate but faster RT than less intelligent subjects. The trouble with this plausible conjecture is that it turns out to be completely false. Brighter subjects, in fact, show both *faster* RT and a *lower* error rate. That is, individual differences in quickness and accuracy of response are *positively* correlated. We have never found an exception to this rule in any of our own RT studies or in any studies reported in the literature.

The evidence forces us to distinguish clearly and dissociate *speed of information processing* from speediness in the ordinary sense of working fast or hurrying to get things done in limited time, or deftness and quickness in overt behaviour (Jensen, 1984b). Individual differences in speed of information processing are probably not observable at the level of people's overt behaviour and can be detected only by means of chronometric techniques. Confusing speed of information processing with quickness of speech and other casually observable gross behaviours is a decided hindrance to understanding the relationship between mental processing speed and IQ or psychometric *g*.

The limited-capacity trace-decay theory of mental speed and psychometric performance

The most basic concept in understanding the correlation between mental speed and psychometric intelligence is the severe limitation of so-called working memory, or "short-term memory." STM has a quite limited capacity for processing incoming information or information retrieved from long-term memory (LTM). Without continuous rehearsal, the limited information in STM rapidly decays beyond retrieval, and must be replaced by future input. Manipulating information held in STM usurps some of its capacity for processing incoming information. Every mental operation takes up a certain amount of time, and if common processes are involved in two or more different operations, these must be performed successively to avoid interference with successful execution of the operations. Overloading the capacity of the system causes shunting or inhibition of the information input or a momentary breakdown in internal operations. All these effects have been demonstrated experimentally in numerous studies and are now generally acknowledged as well-established phenomena in experimental cognitive psychology (e.g., Posner, 1966, 1978, 1982).

How do these limitations of working memory figure in the observed correlation between mental speed in various ECTs and performance on untimed psychometric tests? A faster speed of mental processing, such as encoding stimuli, chunking, transformation, and storage of incoming information and retrieval of information from LTM, permits the system to overcome its limited capacity, by allowing critical operations to occur before the *decay* of information (or its memory trace) in STM. If the trace decays before solution is achieved, repetition of the information input is required until the correct response can occur. The memory span for recalling digits backward, for example, is smaller than the span for digits forward, because the operation of reversing the digits takes a certain amount of time, during which the information in STM decays. Hence, subjects who can recall 7 digits forward can usually recall only 5 digits backward. Beyond some optimal point, which varies across individuals, the average being 7 digits, the greater the number of digits presented, the smaller the number of digits recalled in correct order, because of overload and decay of memory traces. Forward and backward digit span are correlated with psychometric *g*, and are often included in IQ tests such as the *Stanford–Binet* and the *Wechsler* scales. Backward digit span, because of its greater processing demands, consistently shows a higher *g* loading than forward digit span.

Similarly, the correct responses to all mental test items depend on various elementary cognitive processes, the more complex items making the greater processing demands in terms of information storage, operations performed, information retrieved from LTM, and so forth. The more complex the information and the operations required on it, the more time that is required, and consequently the greater the advantage of speed in all the elementary processes involved. Loss of information due to overload interference and decay of traces that were inadequately coded or rehearsed for storage or retrieval results in "breakdown" in grasping all the essential relationships required for arriving at the correct answer. Speed of information processing, therefore, should be increasingly related to success in dealing with cognitive tasks to the extent that this informational load strains the individual's limited working memory. The most discriminating test items are those that "threaten" the processing system at the threshold of breakdown, beyond which erroneous responses occur. In a series of graded complexity, this breakdown would occur at different points for various individuals. If individual differences in the speed of the elementary components of information processing could be measured in tasks that are so simple as to rule out breakdown failure, it should be possible to predict the individual differences in the point of breakdown for more complex tasks, such as *Raven's Progressive Matrices* items or other items typically found in IQ tests. This is the hypothesised basis for the observed correlations between RT variables and scores on complex *g*-loaded tests.

This hypothesis is consistent with the following observed phenomena: (1) The mean response latencies to correctly answered *Raven* items (in a group of subjects) are highly correlated with the item difficulties (i.e., percentage of subjects failing an item); (2) simple true–false test items which are so easy as to elicit 100% correct answers among university students when the items are given as an untimed paper-and-pencil test, however, show highly reliable differences in average response latencies among the items when they are administered to university students as a reaction time test. The mean RTs to the items, as obtained in the university sample, are highly correlated with the *item difficulties* obtained in a group of primary school children, aged 7 to 9 years, to whom the simple true–false items were administered as an untimed paper-and-pencil test. In other words, there is a close relationship between the complexity of processing required by the items, as indicated by the items' average RTs among university students and the average item difficulties when the items are used as a typical psychometric test among primary school children. It should be noted that the items of this test (the *Semantic Verification Test* described later in this chapter) were so simple for university students that the response latencies were generally less than 1 second.

Processes and factors

The study of individual differences in mental abilities is largely based on the analysis of correlations. When more than two variables are involved, the

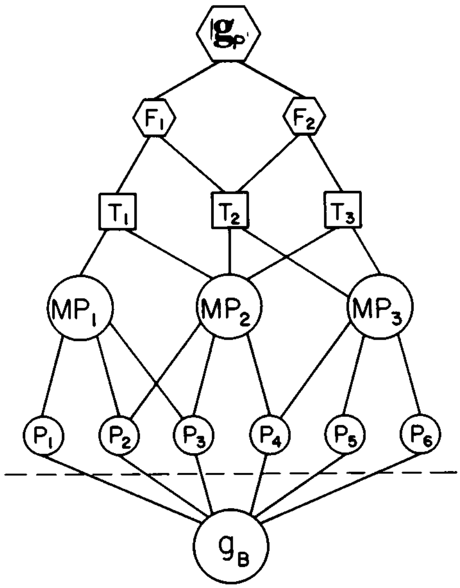

Figure 4.1. Simplified representation of relationships among processing and psychometric variables.

most commonly used methods of correlational analysis are partial correlation, multiple correlation, canonical correlation, and factor analysis of various types. All of these methods have been used to advantage in studying the relationships among various chronometric tasks designed to measure certain elementary cognitive processes and psychometric tests.

At present, there is no generally agreed-upon model for representing the relationships between these two domains. I believe that some form of hierarchical model, however, best serves as a basis for theoretical speculation and the generation of important questions for research. One of the principal aims of theory and research in this field is to understand how and why very simple chronometric variables are related to broad factors of mental ability measured by highly complex tasks. The literature, as yet, provides no complete or compelling account. But a tentative model may help to summarise prevailing questions and focus further empirical enquiry. The simple hierarchical model shown in Figure 4.1 is one possible representation of the most prominent variables and constructs of present concern.

The horizontal dashed line in Figure 4.1 separates the behaviourally or psychologically measurable variables (above the line) from those that are measurable only physiologically, such as evoked brain potentials, or are inferred physiological processes, such as cortical conductivity (Klein & Krech, 1952), synaptic errors (Hendrickson, 1982), neural oscillation (Jensen, 1982b), and the like. The physiological level is represented as one general factor, g_B (*B* for "biological"), but in our present state of knowledge this level could just as well be represented as several distinct physiological processes or as correlated processes, because they share one common process, that is, g_B. The nature of this physiologic underpinning of human abilities is a major focus of Eysenck's (1982) theorising about the findings of correlations between features of the average evoked potential and psychometric *g*, or g_P, which is depicted in the hexagon at the top of the hierarchy in Figure 4.1. All of the solid lines in the figure represent correlations.

The various elementary cognitive processes (*P*) are correlated through sharing common physiological processes. Different parts of the brain or different neural assemblies are presumably specialised for various aspects of information processing. These processes are described as follows: stimulus apprehension; iconic memory; stimulus encoding; short-term memory (STM); rehearsal of encoded STM traces; memory scanning; retrieval of information from long-term memory (LTM); transfer; discrimination; generalisation; transformation of encoded information; mapping of relations; visualisation and mental rotation of figures in two- or three-dimensional space; and response execution. The *processes* (*P*) in this model – which are depicted here as all being closely connected with some biological substrate – can be measured by means of chronometric tasks, either directly or through derived scores, by subtraction of response latencies of simple tasks from those of more complex tasks, in order to measure the additional processes involved in the latter, or by the use of partial correlations, or by factor analysis of a combination of various tasks intended to tap different processes. The methodology of RT studies has been explicated in detail elsewhere (Jensen, 1985a).

Different sets of elementary processes (*P*) can be utilized by a given *metaprocess* (*MP*). The metaprocesses are further removed from the biologic substrate and are probably mainly products of learning and practice. Their connection to the biologic substrate is via the elementary processes which enter into the metaprocesses. Metaprocesses consist of strategies for selecting, combining, and using elementary processes, problem recognition, rule application, planning, allocation of resources, organisation of information, and monitoring one's own performance. Different metaprocesses are intercorrelated because they share certain processes in common and also because the experiential factors which inculcate metaprocesses are correlated in the educational and cultural environment. It is probably at the level of metaprocesses that cultural differences have their primary impact.

The processes and metaprocesses enter into performance on complex *psychometric tests* (*T*). Even a single complex test item may depend upon a number of *P*s and *MP*s for correct performance. Various tests are intercorrelated because they share certain common *P*s and *MP*s and also because they may share common information stored in long-term memory. Note that at each level in this hierarchy, something new is added in terms of environmental inputs. The cumulation of these acquired elements is at its maximum at the level of single items in psychometric tests. Item variance is largely *specificity*, a technical term in factor analysis, referring to a source of variance which is peculiar to a particular item (or a particular test) and is not shared in common with other variables. Specificity may arise from individuals' idiosyncratic experiences, making for unique and uncorrelated bits of information, or from complex and unique interactions among the *P* and *MP* demands and the informational content of a particular test item. In fact, all primary psychological measurements are infested with task-specific variance. Chronometric measurements of elementary processes are no exception. Specificity, which is the bane of individual differences research, can be reduced only by using composite scores or factor scores (which are a particular weighted composite of the component scores) derived from a number of varied tasks or tests, thereby "averaging out" the specificity of the individual tasks.

The top part of the hierarchy in Figure 4.1, including *T*, *F*, and g_P, comprises the realm of traditional psychometrics, including various test scores and hierarchical factors extracted by factor analysis. Here, for the sake of simplicity, are represented only two first-order factors (*F* and F_2) and one second-order factor, psychometric *g*, or g_P. The most general factor, of course, may emerge as a third-order or other higher-order factor. Each successively higher factor level excludes some source of variance. The primary factors, for example, exclude the test-specific variance, and the second-order factors exclude the variance that is peculiar to each primary factor, and so on. The most general factor, g_P, is the variance common to all the sources below it in the hierarchy.

Some homogeneous tests, such as *Raven's Progressive Matrices*, contain relatively little specificity and are therefore quite good measures of g_P. Other tests, such as the *Wechsler* scales, although containing quite heterogeneous items and subtests with considerable specificity, yield composite scores from which, in effect, the specificity is averaged out, providing a good measure of g_P.

Superficially very different tests, such as *Verbal Analogies, Digit Span*, and *Block Designs* are intercorrelated presumably not because of common content or correlated educational experiences, but because they have a number of elementary processes and metaprocesses in common. Because the more superficial differences between tests contribute mainly to their specificities, they are not reflected in g_P. Hence, it has been found that *g* factor scores are more highly correlated with chronometric measures of elementary processes than are any particular types of tests. Vernon (1983), for example, found that all of the correlations between the *Wechsler Adult Intelligence Scale* (*WAIS*) and a number of reaction time measures (from several elementary cognitive tasks) were due to the general factor of the *WAIS*. When the *g* factor was partialled out, none of the *WAIS* subtest scores correlated in the least with the RT measures. Thus, although g_P and P_1, P_2, and so on, appear widely separated in the schematic hierarchy, they actually seem to have greater variance overlap, as shown by the correlation, than do some of the more proximal variables. This picture may also help to elucidate the otherwise surprising finding that, although g_P is derived from factor analysis of psychometric tests which bear virtually no superficial resemblance in format, content, or method of administration to the RT techniques used in elementary cognitive tasks, g_P shows almost as large correlations with ECTs as with the psychometric tests from which g_P is derived.

One of the crucial theoretical questions, with reference to Figure 4.1, regarding which there is presently little consensus, is whether more of the variance in psychometric g (g_P) is attributable to the processes (P) or to the metaprocesses (MP). The learned information content in the psychometric tests (T) can already be virtually ruled out as an important source of g variance, because tests that differ extremely in their information content, such as vocabulary and matrices, are nevertheless highly saturated with one and the same g. The multiple correlation of several simple ECTs (which would tend to limit the role of complex metaprocesses) with g_P has been so substantial in some studies as to suggest that perhaps 50% or more of the g_P variance is accounted for by individual differences in elementary cognitive processes (e.g., Vernon, 1983). If task specificity were further minimised in such studies by using at least three or four different techniques for measuring each of the elementary processes which have already been shown to yield substantial correlations, it seems likely that even as much as 70% of the g variance would be associated with the processing variables. Also, the existing studies have not taken sufficient account of the reliability of these processing measures. Proper corrections for attenuation might appreciably raise the correlations between ECTs and g_P. Split-half or other internal consistency estimates of the reliability of ECTs usually overestimate the test–retest reliability, and it is the test–retest reliability which should be used in correcting correlations for attenuation when the correlated measurements were obtained in different test sessions, such as on different days or even at different times of the same day, say, before and after lunch. Some ECT measurements are so highly sensitive to an individual's fluctuating physiological state from morning till night and from day to day as to have quite low test–retest reliability as compared with most psychometric tests. Theoretical interest, of course, focusses on the *true-score* multiple correlation between g_P and the elementary cognitive processes. A conceivable goal of this research would be to determine the relative proportions of variance in g accounted for by each of a number of clearly identifiable processes and metaprocesses.

The model as presented here is admittedly a reductionist one, in the sense that g variance accounted for at the level of processes is subtracted from that accounted for at the level of metaprocesses. That is, the sources of individual differences in g_P are sought working from the bottom toward the top in Figure 4.1. This approach is arguable, of course. But it seems more plausible that elementary processes such as speed of stimulus apprehension and speed of encoding could affect performance on complex g-loaded tasks such as matrices or vocabulary than that the high-level reasoning skills and specific knowledge tapped by the psychometric tests would affect, say, choice RT to a pair of lights or speed of recognising whether two letters have the same or different names (e.g., *Aa* or *Ab*).

It also seems reasonable, at least in theory, to argue that the broad heritability of g_P sets the lower limit to the proportion of variance in g_P that is attributable to the biological substrate of intelligence (g_B in Fig. 4.1). Estimates of the broad heritability of highly g-loaded tests fall mostly in the range from about .4 to .8, with a central tendency close to .7 when corrected for attenuation. Broad heritability is defined as the proportion of variance in a trait that is attributable to all of the genetic factors that condition the development of the phenotype; it is the squared correlation between genotypes and phenotypes in the population. (For a fairly comprehensive review, see Scarr, 1982.) In what is perhaps the first published study of the heritability of reaction times in a number of ECTs and their correlations with several psychometric factors, based on twins reared apart, McGue, Bouchard, Lykken, and Feuer (1985) conclude, "The results reported here support the existence of a general speed component underlying performance on most experimental cognitive tasks which is strongly related to psychometric measures of g, and for which there are substantial genetic effects." But even if most of the g variance is ultimately traceable to inherited physiological mechanisms, it would not diminish the importance of deciphering the intervening processes and metaprocesses through which these mechanisms find expression at the behavioural level in people's performance on psychometric tests and in all the "real life" manifestations of intelligence. It is a task for behaviour–genetic research to discover the extent to which education and experience influence individual differences at each level of the hierarchy. It is also important to discover the amenability of processes and metaprocesses to specific training and the extent to which the effects of such training are reflected at the various psychometric levels of the hierarchy.

Factor analysis can also be applied to ECTs. This has not yet been attempted on a large enough scale to gain a clear picture of the factorial structure of a wide variety of ECTs. In several multivariate studies (e.g., Keating & Bobbitt, 1978; Vernon, 1983; McGue et al., 1985; Vernon & Jensen, 1985) that I have seen, however, one feature is quite clear: There is always a large General Speed factor along with other relatively smaller factors associated with particular processes, such as Stimulus-encoding Speed and Memory-scanning Rate. Thus the P's in Figure 4.1 are seen as all being highly intercorrelated because of a General Speed factor, yet they are differentiated by some variance unique to each of the processes. It is a reasonable hypothesis that the differentiated structure of abilities revealed by the factor analysis of psychometric tests is derived in part from the unique variance in different processes, which enter into performance on various psychometric tests in different combinations and degrees. Aside from a general Speed-of-Processing factor, for example, different processes are called upon in a vocabulary test (and in the acquisition of word knowledge) than in a test of arithmetic reasoning (and in the acquisition of quantitative skills).

How do seemingly small individual differences in the rate of information processing, as revealed, for example, by the difference between choice RT and simple RT, eventuate in large individual differences in performance on psychometric tests and in scholastic achievement? As indicated previously, because of limited channel capacity for information processing, and rapid decay of information held in working memory, speed of processing is most advantageous when the operations required for successful solution of a test problem are sufficiently complex to strain the subject's working memory. Thus, some part of the correlation between mental speed as measured by ECTs and the general ability measured by psychometric tests is attributable to conditions intrinsic to the *complexity* of the test item, quite apart from any information *content* required by the item. "Culture-free," "culture-fair," or "culture-reduced" types of tests depend mainly on the *complexity* of the mental operations demanded by the items rather than on their *content*, that is, the specific acquired knowledge or skills called for by a particular item. The research on ECTs proves, if nothing else, that it is possible to measure psychometric *g* by means which depend scarcely at all on individual differences in *content*.

But what about those test items which depend on knowledge content, such as the Vocabulary and General Information subtests of the *Wechsler* scales? These tests are highly *g* loaded and are correlated with RT parameters of ECTs. A reasonable explanation of the correlation is that persons with faster rates of information processing acquire more information per unit of time from their experiences than do persons with slower processing rates. Even small individual differences in processing rate, when multiplied by considerable lengths of time, can eventually result in surprisingly large differences in amount of acquired knowledge and skill. A car that on average takes 22.0 msec to travel a foot and a car that takes 22.7 msec will be a mile apart after only an hour's travel. Similarly, we have found that groups with average differences in information-processing rates of only five to ten bits of information per second (i.e., when bits/sec is measured by the reciprocal of the difference between simple and choice RT) differ by one standard deviation or more in tests of scholastic aptitude and achievement (e.g., Jensen, 1982b, Table 1).

It is now well established that time is a critical factor in scholastic achievement or in any type of cognitive learning which progresses from simple to more complex in a cumulative fashion, thereby continually making demands on the learner's working memory. A large part of the function of working memory in the educative process is the encoding and storage of new information into long-term memory and the retrieval of information from LTM, as well as performing a number of other operations on retrieved information, such as transformation, mapping, and transfer, in accord with task requirements (Anderson, 1983). Individual differences in the total time needed for learning scholastic material to a given criterion are correlated with IQ, or *g*. Time-to-learn (TTL) ratios, comparing the slowest to the fastest learners are typically of the order of 2:1 to 7:1 in most studies, depending on the range of talent in the sample and the complexity of the material to be learned (Gettinger, 1984). The ratios of low- to high-IQ groups in rates of information processing, as measured in terms of differences between simple and choice RT, fall within a somewhat narrower range, about 1:2 to 1:4. In comparing average or above-average groups with the mildly retarded (IQs 60 to 80), one does not find average subjects whose median choice RT (CRT) is as slow as the mean of retarded subjects, but there are a few retarded persons whose median CRT is faster than the mean CRT of the average group. The relationship between CRT and mental retardation could be stated as follows: Fast CRT is necessary but not sufficient for average or above-average intelligence, whereas slow CRT is sufficient but not necessary for mental retardation.

The speed–complexity "paradox"

The magnitude of correlation between individual differences in a reaction time task, or ECT, and psychometric *g* is related to task complexity. The relationship is not linear, however, but rather appears as an inverted-U function. That is, the correlation increases, going from very simple tasks to more complex tasks; but beyond some optimal level of complexity the correlation gradually decreases. Response latencies to highly complex items such as those in *Raven's Progressive Matrices*, which often require a minute or more to solve, show close to zero correlation with *g*. It may seem paradoxical that the *Raven*, which is highly *g* loaded when given as an untimed test, is not at all correlated with *g* in terms of the subject's mean response latencies to the correctly answered items. We have found that the range of task complexity (as indicated by mean latencies) for which the latencies show significant correlations with *g* is quite narrow: tasks with mean latencies between 300 and 1,000 msec, for university students. The limits of this optimal range of RTs probably differs according to the average level of ability of the group in which the RT × *g* correlation is computed, the optimal latencies being longer for less able groups and for children. In a series of 14 RT tasks of varying complexity, with mean response latencies ranging between about 400 and 1,400 msec (for university students), we have found the maximum correlation (about −.50) with *g* (*Raven's Advanced Progressive Matrices*) for those tasks with mean latencies close to 700 msec.

This finding suggests the hypothesis that a task becomes too complex for the optimal correlation with *g* when the solution time exceeds the decay time of the information in working memory. At that point, different complex strategies are invoked for rehearsing information in short-term memory and for getting information into LTM, or for searching LTM for relevant information for problem solution. In other words, when the mental operations required for successful performance of a problem cannot be performed, either simultaneously or sequentially, within the time constraints of working memory imposed by rapid decay of memory traces, other, more complex metaprocesses come into play, and the total response latency then is not as pure a reflection of the efficiency of the elementary cognitive processes. A highly complex problem is, in effect, divided into a number of subproblems for solution and, depending on the subject's strategy, time is allocated differently to the various subproblems. Personality variables and individual differences in other noncognitive factors, such as involuntary rest pauses, enter into the subject's performance, severely attenuating sheer response latency as a measure of *g*. It appears as if the metaprocesses or strategies are more susceptible to idiosyncratic interactions between subjects and tasks than are the elementary processes. This interactive strategy variance is, in effect, "averaged out" when response latencies to individual items are averaged over subjects, revealing a high correlation between *average* item latencies and the difficulty levels (i.e., percent failing) of the items when administered as an untimed psychometric test. One of my graduate students, Steven Paul (1985), has recently obtained data showing clearly that the mean latencies of simple test items that have short latencies in the range of optimal correlations with *g* in university students are highly correlated with the mean item difficulty levels (percent failing) in school children of ages 7 to 9 years, who took the items in the form of an untimed paper-and-pencil test. These are extremely simple items for university students, with average latencies mostly below 1,000 msec, and they reflect the efficiency of hardly more than elementary cognitive processes. This suggests to me that the elementary processes may be more consistently related to *g* than are the higher executive functions, or metaprocesses.

There remains a crucial question that has not yet been definitively answered by any research that I can find in the literature. Is the correlation of various ECTs with psychometric *g* attributable entirely to the common factor (Mental Speed) among the ECTs? Or do the different processes (encoding, retrieval, etc.) involved in various ECTs also contribute to their correlation with *g*? In terms of multiple regression, we know that successively adding ECTs to the regression equation for predicting *g* increases R^2, that is, the proportion of variance in *g* accounted for by the ECTs (e.g., Keating & Bobbitt, 1978). But this is an ambiguous finding, as it stands. Does the R^2 increase because independent processes are being successively added to the regression equation? Or does the R^2 increase merely because, by adding more ECTs, we are increasing the reliability of the Common Speed factor, and could just as well produce the same increments in R^2 by adding in repeated testings on one and the same ECT? This question can be answered only by obtaining highly accurate estimates of the test–retest reliability of each of the ECT variables and correcting all of the zero-order correlations for attenuation before calculating R^2. Then, if the common factor among the ECTs is solely responsible for their correlation with psychometric *g*, a step-wise multiple regression (entering the most highly predictive ECT first) should not show significant increments in R^2 by the addition of successive ECTs after the first one in the regression equation. This procedure has not yet been done.

A variety of elementary cognitive tasks

In the past few years, considerable research findings have accrued to a number of ECTs. The substantive and methodological aspects of much of this research have been reviewed in detail elsewhere (Carroll, 1980; Berger, 1982; Jensen, 1982a, 1982b, 1985a; Vernon, 1985). My purpose here is to provide very brief descriptions of the chronometric paradigms which have already shown dependable correlations with psychometric *g*, and to indicate in summary fashion the most salient findings and unresolved research questions associated with each paradigm.

The stimuli and task requirements of most of these chronometric paradigms are so universally available to experience, and are so simple and relatively free of intellectual and cultural content, that it seems they could be easily adapted to cross-cultural testing.

All of these paradigms, being chronometric, have the methodological and quantitative advantage of measurement on a ratio scale, being absolute measurements of real time, expressed in decimal fractions of a second, which are standard units in the universally adopted Système Internationale for all physical and scientific measurements. However, it is most important to note that the absolute time values obtained in any chronometric study, however reliable they may be, are a function not only of subject factors but also of apparatus characteristics. Seemingly slight variations in apparatus or procedures can result in quite marked differences in the absolute values of RT measurements (Jensen, 1985a). If a given paradigm is to be used in cross-cultural comparisons, it is essential that either the identical apparatus (i.e., the stimulus response console with which the subject interacts) or highly standardised replicas, as uniform in every respect as the technology of manufacture will permit, be used for all groups. *Relationships* among chronometric variables and between chronometric and psychometric variables remain quite stable across considerable variations in apparatus and testing procedure, but the *absolute values* of the chronometric variables are remarkably sensitive to even slight variations in these features.

Inspection time

The research on IT has been reviewed by Brand and Deary (1982), Brand (1984), and Nettelbeck (1987). IT measures speed of apprehension of visual or auditory stimuli. Brand (1984) describes IT as "a ready ability to apprehend the most simple perceptual realities that constitutes one major psychological and ontogenetic basis for the development in intelligence." The basic idea originated about a century ago with Galton, who argued that "the only information that reaches us concerning outward events appears to pass through the avenue of our senses; and the more perceptive the senses are of difference, the larger is the field upon which our judgment and intelligence can act" (Galton, 1883, p. 19).

In the simplest form of visual IT, two vertical parallel lines, differing about 30% in length, are exposed in a tachistoscope, followed, after a brief interval, by a masking stimulus. The subject reports whether the longer line appeared on the right or the left of the shorter line. (Their positions are randomised across trials.) There is no time constraint on the subject's verbal response per se. The interstimulus interval (i.e., time between exposure of the vertical lines and exposure of the mask) is varied systematically until that interval is found at which the subject responds correctly on 95% of the trials. This interval, measured in milliseconds, is the subject's IT. Individual differences in IT range widely, between about 20 and 700 msec (see Brand & Deary, 1982, Table 1).

Correlations between IT and IQ also range widely, depending mainly on the level and range of IQs in the sample. There is no simple way to summarise the IT × IQ correlations in the literature. But two generalisations seem warranted at present: (1) Almost all the IT × IQ correlations are in the expected direction, that is, slower IT is associated with lower IQ; (2) the overall results are highly significant, rendering the null hypothesis definitely untenable; (3) correlations are much higher in samples that fall in the lower half of the IQ distribution (i.e., IQs < 100) than in the upper half. IQ 85 seems to be the critical threshold; samples that include roughly equal numbers of subjects who are above and below IQ 85 show the most impressive IQ × IT correlation, and the inclusion of mentally retarded subjects with IQ below 70 increases the correlation. In samples of above-average IQ, correlations are typically around −.25. In samples covering the full range of IQ, correlations are typically above −.50. Brand (1984) suggests there is a linear relation between IT and IQ up to about IQ 110, but little relation beyond that point. Vernon (1983) found zero correlation between IT and IQ in a group of 100 university students with a mean *WAIS* IQ of 122 and in which the lowest IQ was 110; and in a factor analysis including nine RT variables with loadings ranging from +.51 to +.91 on the first principal factor, IT had a nonsignificant loading of −.17. Brand (1984) suggests an explanation of this apparently nonlinear relation of IT to intelligence throughout the full range of IQ with an analogy:

> The relation between mental intake speed and intelligence may resemble the relation between income and patterns of investment and expenditure. Across the lower ranges of income there are fairly predictable relations between a person's income and his possessions; but, as the higher ranges of income are reached, big individual differ-

ences arise in the disposal of income – into luxuries, education, health care, addictions, and so on.

There are considerably more complex elaborations of the basic IT paradigm, however, which have shown quite impressive correlations (about .6 to .7) with IQ and scholastic aptitude scores in university students, despite the fact that the IT tasks involved no intellectual content (Livson & Krech, 1956; Raz, Willerman, Ingmundson, & Hanlon, 1983). In the study by Raz et al. (1983), an auditory recognition test (identifying a target tone followed by a masking tone as either "high" or "low") showed mean differences of more than two standard deviations (*SD*s) between a group of university students with total *Scholastic Aptitude Test* scores above 1,200 and a group with total scores below 800 (a mean difference of between 1 and 3 *SD*s).

Simple and choice reaction time

Simple reaction time (SRT) shows slight correlations with *g*, usually not much above −.10. Choice reaction time (CRT) involves some uncertainty as to which one of *n* possible response alternatives will be called for when one of *n* stimulus alternatives occurs. CRT almost invariably shows higher correlations with IQ or *g* than does SRT, the correlation increasing (up to a point) as the number of alternatives, and hence the amount of uncertainty, increases. RT increases as a linear function of the logarithm of the number of alternatives, a relationship now known as Hick's law. In information theory, the unit of information, known as a *bit* (for binary digit), is defined as the amount of information needed to reduce uncertainty by one-half. Accordingly, amount of uncertainty (and conversely, the quantity of information required to reduce it one-half) can be defined as the logarithm to the base 2 of the number (*n*) of alternatives, or choices. Thus a *bit* = $\log_2 n$.

These relationships, which I shall henceforth refer to simply as "the Hick paradigm," have been implemented for the study of individual differences in RT by means of the apparatus shown in Figure 4.2, called the reaction time–movement time, or RT–MT, apparatus. The number of light/button combinations used for any given RT task can be varied by the use of overlays, which may expose any number from 1 to 8 lights/buttons. A trial begins with the subject holding down the "home" button with the index finger of his preferred hand. A preparatory stimulus ("beep") sounds; after a random interval of 1 to 4 seconds, one light (the "reaction stimulus") goes on. The subject's task is to turn off the light as quickly as possible by touching the sensitive microswitch push-button adjacent to the light. Trials are spaced 5 to 10 seconds apart. Typically, 15 or 20 trials are given at any one level of difficulty. The most commonly used levels of difficulty are 1, 2, 4, and 8 choice alternatives (*n*) which correspond to 0, 1, 2, and 3 bits of information.

Two basic time measurements (in msec) are recorded on each trial: RT, or the interval between the onset of the light and the subject's release of the home button; and movement time (MT), the interval between release of the

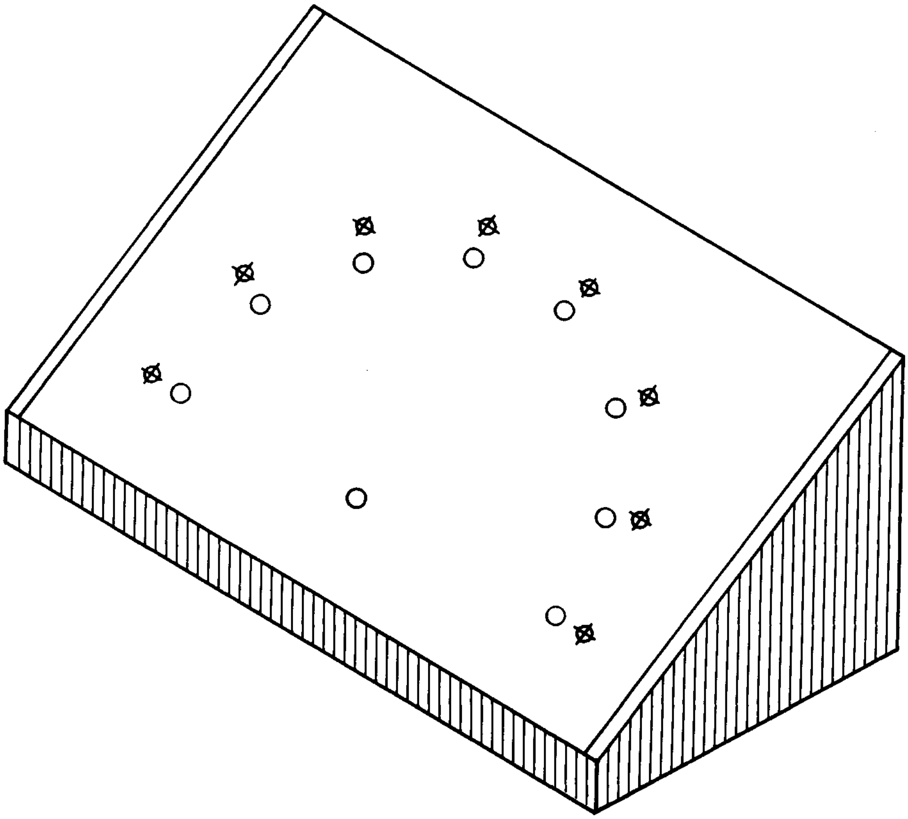

Figure 4.2. Subject's console of the RT–MT apparatus. Push-buttons are indicated by circles; green jewelled lights, by crossed circles. The "home" button is in the lower centre, 6 inches from each response button.

home button and touching the button which turns off the light. Older RT procedures did not distinguish between RT and MT; the two were confounded in a single measurement. The mixing of RT and MT in a single measure, however, attenuates its correlation with any other variable, such as *g*, if RT and MT (as here defined) are differentially correlated with the other variable. RT and MT evidently measure different processes. Their intercorrelation *within* subjects is zero; *between* subjects, the *r* is about +.30. Therefore, adding individual differences in MT to individual differences in RT is almost equivalent to adding random error to RT. It is likely that the failure of many older studies to show significant correlations between response latency and intelligence is the result of not separating RT from MT.

Results of research with this procedure have been comprehensively reviewed elsewhere (Jensen, 1982a, 1982b, 1987b). Typical features of RT and MT are shown in Figure 4.3. So far, MT has been of lesser interest than

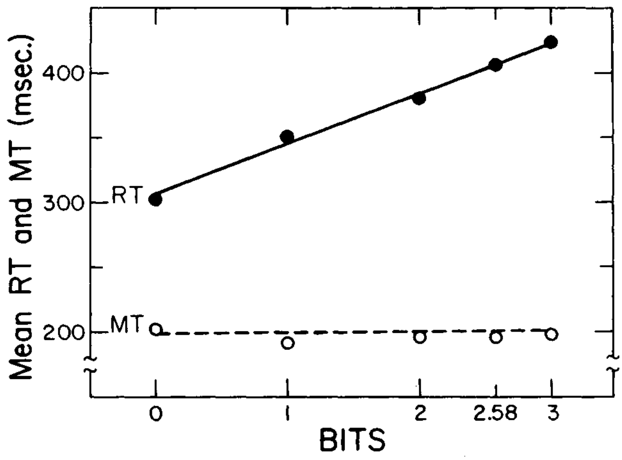

Figure 4.3. Mean RT and MT as a function of bits in 160 schoolchildren in grades four to six. The correlation (r) between mean RT and bits is +.996.

RT. Usually, RT shows larger and more consistent correlations with IQ and mental age. Also, RT is always highly related to differences in task complexity, such as number of choices, whereas MT is relatively constant across variations in task complexity. The three most important individual difference RT parameters derivable from this paradigm are the RT intercept, slope of RT (as a function of bits), and intraindividual variability, labelled SDRT (the average standard deviation of RT over all trials at each level of bits). The *intercept* is complexly determined, reflecting sensory and motor lag, peripheral nerve conduction, apprehension and encoding of the stimulus, and preparation and initiation of the response. The *slope* reflects central processes: discrimination, comparison, choice, and response selection. The reciprocal of the slope can be interpreted as *rate* of information processing, in number of bits per millisecond. Intercept, slope, and SDRT are all correlated with g. Slope shows the lowest correlations, unless they are corrected for attenuation – a dubious procedure when test–retest reliability is quite low, as is usually true for the slope parameter. SDRT has shown the consistently highest correlations, despite its having lower reliability than the intercept, which is by far the most stable of the three parameters.

The simple "lawfulness" or regularity of the phenomena found in this paradigm, as seen in Figure 4.3, is always striking and consistent in every study and even for individual subjects. For example, the correlation between mean RT and bits is .996; it averages close to .97 for individual subjects. SDRT is equally regular, except that it is an exponential rather than linear function of bits, as seen in Figure 4.4. However, when SDRT is plotted as a function of number of response alternatives (n), the relationship is just as

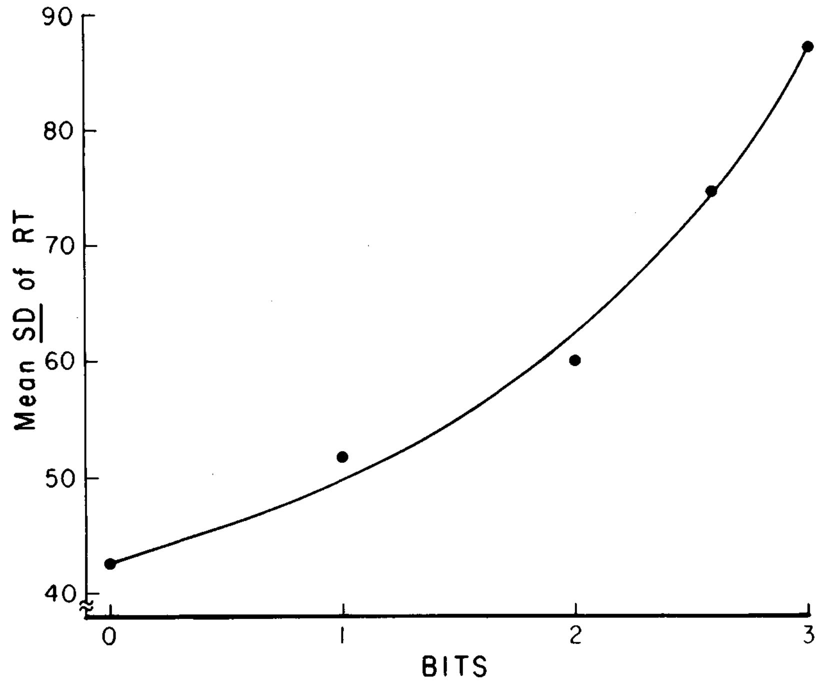

Figure 4.4. Mean intraindividual variability (measured by the standard deviation of RTs in milliseconds on 30 trials) as a function of bits, for 160 schoolchildren in grades four to six. When this RT parameter is plotted as a function of number of alternatives (n), rather than bits (i.e., $\log_2 n$), it is almost perfectly linear, showing a correlation (r) of +.996 between n and *SD* of RT.

perfectly linear as the regression of mean RT on bits. Linearity is confirmed by correlations of +.996 in both cases.

There is one other set of relationships, not heretofore mentioned in the literature, which is no less striking and in need of theoretical explanation. This is the simplex pattern of intercorrelations among the RTs at each level of difficulty (i.e., n = 1, 2, 4, 6, or 8 lights/buttons). These correlations, based on the same data shown in Figures 4.3 and 4.4, are given in the area above the diagonal of the matrix in Table 4.1. All the correlations have been corrected for attenuation (based on Spearman–Brown boosted split-half reliabilities). The pattern of these correlations is an almost perfect simplex, that is, the correlations systematically decrease with each step that they are removed (in either direction) from the principal diagonal of the matrix. Interestingly, the identical pattern can be generated from a simple common-elements model of correlation. By this model, the correlation between RTs for $n = 1$ and $n = 2$ is $r_{1,2} = \sqrt{1/2}$. Similarly, $r_{1,4} = \sqrt{1/4}$, $r_{1,6} = \sqrt{1/6}$, and $r_{6,8} = \sqrt{6/8}$, and so on. The matrix of correlations below the diagonal in Table 4.1 was generated by this model. These generated correlations are correlated .997 with the corresponding empirical correlations (above the

Table 4.1. *Observed correlations (above diagonal) between RTs for different numbers of alternatives* (n), *and theoretical correlations predicted by a common-elements model (below diagonal)*

n	1	2	4	6	8
1		.90	.80	.76	.72
2	.71 (.90)[a]		.90	.80	.77
4	.50 (.79)	.71 (.90)		.96	.90
6	.41 (.74)	.58 (.83)	.82 (.96)		.99
8	.35 (.71)	.50 (.79)	.71 (.90)	.87 (.98)	

[a] Correlations in parentheses are a linear transformation (transformed $r' = .5305 + .5185r$) of the theoretical correlations predicted by the overlap model. The correlation between the observed and predicted correlations is +.997.

diagonal). The fact that the generated correlations are of overall lesser magnitude than the observed correlations is of no theoretical consequence, because each set of correlations is merely a linear transformation of the other set. For example, a linear transformation of the model-generated correlations (shown in parentheses in Table 4.1) makes them highly similar to the obtained correlations in absolute magnitudes. The correlation of .997 between the model-generated and empirically observed correlations, of course, remains unchanged by any linear transformation of the correlations. These regularities in RT data derived with the Hick paradigm look more like the data of physics than of psychology, and they invite theoretical efforts to construct a model of the brain mechanisms that could cause such lawful phenomena (Jensen, 1982b).

I emphasise these striking regularities in the Hick paradigm here, however, because of their importance for cross-cultural research. The linear relations of RT to bits, the linear relation of SDRT to n, the simplex pattern of intercorrelations of RTs at different levels of n – if all of these regularities appear in data obtained in two (or more) different cultural groups, it seems a safe presumption that the Hick paradigm measures individual differences in the very same processes that are responsible for these regularities in both groups. Differences between the groups in mean RT, or in any other parameters of the RT paradigm, could hardly be explained in terms of the task's measuring different processes or ability factors in the different cultures.

Short-term memory scan

This paradigm, introduced by S. Sternberg (1966), measures the time required to retrieve an item of information from short-term memory (STM). A set of digits (termed the *positive set*), varying in length from 1 to 5 or 7 digits, is presented for a sufficient time for the subject to memorise the series. After a 2-sec blank interval, a single probe digit is presented. On a random half of the trials, the probe is a member of the positive set and on half of the trials it is not. The subject's task is to respond as quickly as possible *Yes* or *No* as to whether the probe was or was not a member of the positive set. Response is made by pressing buttons labelled *Yes* and *No*. RTs consistently show two features: (1) They are longer for *No* than for *Yes* responses, and (2) they are a positive linear function of the *number* of digits in the positive set. Individual differences in the intercept and slope of this function have been found in several studies to be correlated with *g* (for reviews, see Jensen, 1982a, 1982b, 1987c). No one has yet explained why RT is a linear function of $\log_2 n$ in the Hick paradigm, but is a linear function of n in the Sternberg paradigm. There is an obvious need for a theoretical model that could explain this difference.

The slope of the linear function relating RT to the number of items that must be scanned in STM may be regarded as the speed of STM scanning in msec/item. Cavanagh (1972) has shown a Pearson correlation of +.9975 between this measure of mean STM scanning speed (msec/item) and the reciprocal of the mean memory span (for group data) for different types of items (digits, colours, letters, geometric shapes, words, random forms, and nonsense syllables). This striking discovery is highly consistent with the theory of trace decay in working memory as the basis for the relationship between RT and psychometric test performance, described earlier in this chapter. The greater the complexity or information load of the items that must be processed, the slower is the STM scan rate and the shorter is the memory span for the items. Memory span is limited by the number of items that can be reported before the STM trace decays. Memory span is correlated with *g* and has long been included among the items of individual IQ tests, most notably the *Binet* and *Wechsler* scales.

The Sternberg paradigm, in connection with Cavanagh's discovery, would lend itself nicely to the search for cross-cultural invariance of cognitive processes. A wide variety of stimulus items selected for their high familiarity in different cultures could be used to plot the Cavanagh function relating speed of processing to memory span in age-matched subject samples from the different cultures, in order to determine if the items selected from different cultures fall on a common regression line and in the same rank order in the subject samples from each culture.

Long-term memory retrieval paradigms

The purpose of these paradigms is to measure the time it takes to retrieve highly overlearned items of information from LTM. This is achieved by comparing RT on a simple discriminative task which involves an LTM component with RT on a task that makes virtually identical sensory-discrimination and response demands but does not require access to LTM. The simplest of such procedures, originated by Posner (1978, chap. 2), measures the time to access the names of highly overlearned single letters of the alphabet. Pairs of letters, printed in either upper or lower case, in which the letters in each pair are either the *same* or *different* in letter *name*, are presented, and the subject responds by pressing buttons labelled *Same* (*S*) and *Different* (*D*) in terms of whether the letters have the same or different names. For example, *Aa* (*S*) *AA* (*S*), *Ab* (*D*), *AB* (*D*). This task condition is known as *name identity*, or NI. The comparison task is *physical identity*, or PI. Here the very same letter pairs are presented, but the subject responds *Same* or *Different* on the basis of whether the letters are physically the same or different, thus: *Aa* (*D*), *AA* (*S*), etc. Average RT is longer for NI than for PI, and the difference, NI − PI, is a measure of the time required to access letter names in LTM. The difference, NI − PI, amounts to about 75 msec in university students, which is the same as their difference in RTs between 0 and 3 bits of information in the Hick paradigm. NI − PI is modestly correlated (about −.25) with IQ, and appears to be correlated with a verbal ability factor in addition to *g*. The NI task, however, involves so very little LTM search, at least in the case of university students, as to not disciminate individual differences in intelligence with as much precision as can be achieved with more complex processing tasks.

To increase the complexity and LTM search demands of the basic NI − PI paradigm, whole words have been substituted for single letters. Corresponding to the NI condition is the *Synonyms–Antonyms* (*SA*) *Test*, in which pairs of short, high-frequency (*AA* in the *Lorge–Thorndike* word count) words are presented. The paired words are of either similar or opposite meaning, and the subject responds on push-buttons labelled *Same* or *Different*, for example, *big–large* (*S*), *hot–cold* (*D*). The comparison test, corresponding to PI, is called the *Same–Different* (*SD*) *Test*. It consists of pairs of words (comparable in length and frequency to those used in the NI condition) which are either identical or unrelated. The subject responds *Same* or *Different* on the basis of physical identity, for example, *cow–cow* (*S*), *hot–table* (*D*). The mean RTs for *SA* − *SD* also measure time for retrieving information from LTM. This difference is about 300 msec for university students. This paradigm has shown quite substantial correlations with IQ (Vernon, 1983; Vernon & Jensen, 1984). An essential requirement for this procedure is that the stimulus words should all be highly familiar to all subjects. It should be a test not of knowledge but of speed of access to well-learned, highly available information stored in LTM. This requirement can be tested by first giving the items as a paper-and-pencil test to all subjects. In the version of the *Synonyms–Antonyms* test used in our studies, the error rate for university students is zero; it is only 10% for average 8- and 9-year-old schoolchildren. This test could be used in cross-cultural research, with the appropriate words or other symbols which have similar or opposite meanings in a given culture. Choice of words could be matched in terms of frequency within each language or culture.

Another speed of LTM retrieval task is *category matching*. It is easier (by about 90 msec) than the *Synonyms–Antonyms* task, for university students. The subject is presented successively 40 pairs of words, in which the first word is the name of one of five categories (animals, clothing, fruits, furniture, sports) and the second word of the pair is either a member of the named category or of one of the other four categories. The subject responds (by push-button) whether the second word is or is not a member of the named category. RT on this task is correlated about +.70 with RT on the *Synonyms–Antonyms* test, and with IQ (Vernon, 1985).

Semantic Verification Test (SVT)

This test, originally suggested by Baddeley (1968), measures the amount of time it takes for a person to decide whether a physical stimulus does or does not correspond to a brief "sentence" describing it (see also Clark & Chase, 1972). We have devised two chronometric forms of the SVT, differing in complexity. The simpler form involves only two capital letters, *A* and *B*. A "sentence" such as one of the following is presented visually for 3 seconds:

A before *B*
A after *B*
A not before *B*
A not after *B*

After a 1-sec blank interval, a pair of letters appears, for example, *AB*, and the subject responds on push-buttons labelled *True* or *False* according to whether the letter positions correspond to the descriptive statement. RT is the interval between onset of the letters and the subject's releasing the home button in order to touch the *T* or *F* button. The task is subject-paced, each trial initiated by the subject's depressing the home button. RT and MT are recorded on each trial.

The more complex form of the SVT uses three letters (*A*, *B*, *C*) and other descriptors besides *before* and *after*: *first*, *last*, *between*. "Sentences" in positive and negative forms are composed of every possible permutation of these descriptors, and the reaction stimuli consist of all permutations of *ABC*, half of them true and half of them false with respect to the given statements. Using *ABC*, instead of only *AB*, markedly increases the processing demands of the task, mainly because the "sentences" in the simple *AB* condition *always* permit the subject to form a mental image of the order of the letters that will appear as the reaction stimulus, whereas the "sentences" in the complex *ABC* condition allow more "degrees of freedom" for the reaction stimulus; the order of the three letters cannot be invariably or completely anticipated.

Before being given the chronometric form of the SVT, subjects are given the items as a true–false paper-and-pencil test to provide familiarity and practice on this paradigm and to ensure that all subjects are capable of errorless performance under nonspeeded conditions. RT on the SVT is correlated about −.50 with *Raven's Advanced* scores in university students. When corrected for attenuation and restriction of IQ range, in the university sample, this correlation comes close to −.80. The various sentences of the complex form of the SVT show large, reliable differences in mean RT, ranging from about 300 to 1,300 msec. MT, in contrast, is constant across all sentence conditions and shows negligible and nonsignificant correlation with the *Raven* (Paul, 1985).

Dual tasks

These tasks (also termed *competition tasks*) are a means for measuring storage/processing trade-off in working memory. The more of the capacity of working memory that is used for STM storage of information, the less there is available for other forms of processing information. A dual task thereby puts a greater strain on the storage and processing capacity of working memory. As a consequence, dual tasks show higher correlations with *g* than either of the component tasks given singly. In a classic set of experiments, Baddeley and Hitch (1974) showed that when persons were given a short string of digits to memorise, followed by a simple reasoning task, and then had to recall the previously memorised digits, their performance increasingly deteriorated as the number of digits to be remembered approached the subject's memory span. Following this lead, Stankov (1983) has made the important discovery that performances on a variety of ECTs are more highly intercorrelated, and are therefore more heavily *g* loaded, when they are presented in the dual task paradigm than when presented as single tasks. Also, Stankov distinguishes between the *active* and *passive* aspects of working memory, corresponding to the *processing* and *storage* of information. Stankov claims evidence that the active component of working memory is more highly correlated with fluid *g* than is the passive component and that "operations performed on information in working memory are more indicative of fluid intelligence than is the ability to hold this information in working memory" (Stankov, 1983, p. 51). This observation is very similar to Jensen's (1974) distinction between Level I and Level II abilities as *encoding and retention* of stimulus input (Level I) and *mental manipulation* of encoded material (Level II).

An obvious advantage of dual tasks in cross-cultural research is that, because dual tasks are more highly *g* loaded, the increments in RT produced by dual versus single tasks, when the component tasks are identical, are a content-free measure of mental efficiency.

In our laboratory we have used two dual tasks which yield four RT measures. They are composed of the Sternberg (1966) *Digit Scan test* and the *Same–Different* (SD) word pairs and *Synonyms–Antonyms* (*SA*) word pairs described previously. A single trial of each dual task can be summarised as follows:

Dual Task 2 (DT2)
Digit Series (2 sec)
Same–Different words – RT
Probe Digit – RT
Dual Task 3 (DT3)
Digit Series (2 sec)
Synonym–Antonym – RT
Probe Digit – RT

The subject is presented with a string of 1 to 7 digits (for 2 seconds), which he is told to rehearse for later recall. When the digits leave the screen, they are replaced by a pair of words to which the subject responds *Same* or *Different* on the push-button console, and RT is recorded. After responding, the subject presses the home button again, which triggers the appearance of a single probe digit; the subject then responds *Yes* or *No* as to whether the probe was a member of the string of digits that appeared before the words, and RT is recorded. In the three studies in which these dual tasks have been used, they have shown significant correlations with *g*, and the correlations are generally higher than for the same tasks given singly (Vernon, 1983, 1985; Vernon & Jensen, 1984).

Chronometric studies of population differences

Black and white (i.e., negroid and caucasoid) populations, in any part of the world, whenever they have been given psychometric tests, have been found to differ statistically, on average, at least one standard deviation (σ) in every kind of test that has a large *g* loading when it is factor analysed among any large and diverse collection of mental tests. In fact, Spearman (1927, p. 379) conjectured that the magnitude of the black–white difference on various tests is directly related to the tests' *g* loadings. This hypothesis is borne out by a number of factor-analytic studies and has so far been contradicted by none (Jensen, 1980a, pp. 535–539, 732–735; 1985b; 1985c; Naglieri & Jensen, 1987). Understanding the nature of this difference at a more basic level of analysis than the factor analysis of psychometric tests is a long-standing desideratum of differential psychology. Chronometric techniques would seem to afford one potentially fruitful approach toward this aim.

Several studies have used one or more of the chronometric techniques previously described in comparing black and white groups in Africa and America. No entirely consistent pattern of results emerges, and there is not yet enough data derived from sufficiently similar chronometric techniques or sufficiently large, representative, and comparable samples across studies to warrant any worthy general conclusions. In each study, however, statistically significant effects have been found, which suggests that further studies in this vein, more systematically designed and theory guided, and with proper replication, using standardised techniques and procedures, should lead to consistent, theoretically interpretable results. The few scattered studies reported thus far are reviewed here briefly. All differences between sample means are reported here in standard deviation (σ) units, where σ is the square root of the N-weighted mean of the variances of the two samples.

Bligh (1967), as reported in Poortinga (1971), compared 26 white European and 26 black African subjects on simple RT, using auditory ("click") and visual (flash of light) stimuli. Bligh used the peculiar procedure of having the subjects keep their eyes closed during presentation of the stimuli. Poortinga (1971, p. 70) suggests that the more heavily pigmented eyelids of the black subjects might have affected the results, causing the group difference in visual RT to be greater than that for auditory RT. The black–white differences for visual RT and auditory RT were $.90\sigma$ and $.57\sigma$, respectively.

Poortinga (1971) measured simple and choice visual RT and auditory RT in 40 white and 40 black African university students between the ages of 18 and 24 years. Poortinga claims that "there seems to be no reason why the [black] African sample cannot be considered to be representative of all African students in South Africa" (p. 24). On three psychometric tests, the black–white differences were (from Poortinga, 1971, Table 24, p. 72): *Mental Alertness* (2.31σ); *Raven's Advanced Progressive Matrices* (2.16σ); *Blox* (spatial) (1.52σ). Poortinga also used a click and a flash as reaction stimuli, but his procedure was different from Bligh's. For *simple RT*, clicks and flashes were presented in alternative order and the subject responded to each stimulus as fast as possible by pressing a button. For *choice RT*, the click and flash occurred in random order, and the subject responded to each stimulus by pressing one of two designated buttons. On one-fifth of the choice RT trials, the click and flash were presented simultaneously and the subject was instructed to press the click or the flash button according to which stimulus seemed to occur first. The black–white differences, all of them nonsignificant, were as follows (from Poortinga, 1971, Table 17, p. 66);

	Auditory	Visual
Simple RT	$.13\sigma$	$.02\sigma$
Choice RT	$-.03\sigma$	$.00\sigma$

Poortinga also gave a more complex 4-choice RT task in visual and auditory modes to the same groups. The visual stimuli were four coloured lights (green, yellow, blue, red); each appeared in the same aperture. The auditory stimuli were four highly distinct sounds (Wundt hammer, buzzer, hooter, bell). Subjects responded on four push-buttons, using the index and middle fingers of each hand. Finally, an 8-choice task was given, consisting of the four auditory and four visual stimuli presented in a random order. Because stimulus–response compatibility (i.e., the physical proximity or spatial correspondence of the response buttons to the alternate stimuli) was quite low in this arrangement, there was a pronounced practice effect, amounting to about 10 to 15 msec, over the course of 100 trials. Hence, the task involves some degree of learning as well as reaction time per se. Internal consistency reliabilities were high (.71–.90), however, and permit correction of the mean group differences for attenuation. The black–white differences (with disattenuated differences in parentheses) are as follows (from Poortinga, 1971, Tables 7 and 10, pp. 47 and 50). All of the differences are significant beyond the .05 level.

	Auditory	Visual
4-choice RT	1.36σ (1.70σ)	1.53σ (2.06σ)
8-choice RT	1.30σ (1.71σ)	1.26σ (1.69σ)

Thus it appears, from Poortinga's study, that the black–white difference is nonexistent or negligibly small for quite simple visual and auditory RT tasks involving no more than 0 or 1 bit of information (i.e., simple RT and 2-choice RT). But when the RT task is more complex, involving 4-choice and 8-choice RT (or 2 and 3 bits), quite marked differences appear, averaging about 1.4σ (1.8σ corrected for attenuation), a difference about two-thirds as large as the groups' difference on *Raven's Advanced Progressive Matrices*, a highly *g*-loaded test.

The correlations between Poortinga's various RT tasks and the *Raven's*, however, are inconsistent, as shown below (from Poortinga, 1971, Tables 25 and 26, pp. 73–74):

		Black	White
Simple RT	Click	−.05	+.06
	Flash	−.08	+.18
2-choice RT	Click	+.17	+.09
	Flash	+.28	+.08
4-choice RT	Auditory	−.07	−.45*
	Visual	−.05	−.16
8-choice RT	Auditory	−.05	−.38*
	Visual	−.02	−.17

(*$p < .01$)

These mostly nonsignificant correlations, and their lack of consistency between the two samples, bring into question whether these RT tasks (or the *Raven*) are measuring any significant component of *g* in both racial samples. They pose an insoluble problem for interpretation, given only the results of Poortinga's study. Do Poortinga's RT results show any consistency with other RT studies? To some extent, yes.

Noble (1969) tested representative samples of white and black grade-school children (106 in each group) in rural Georgia, matched for age and sex, on a 4-choice RT test. As in Poortinga's study, there was a low degree of stimulus–response compatibility in Noble's RT task, so that part of the subject's task consisted of learning the multiple-choice connections between the four reaction stimuli (coloured lights) and the correct motor responses (pushing toggle switches). RTs therefore showed gradual, negatively accelerated improvement with practice over the course of 160 trials, but the practice curves trend toward significantly ($p < .01$) different asymptotes for blacks and whites. The results, when plotted in terms of response speed

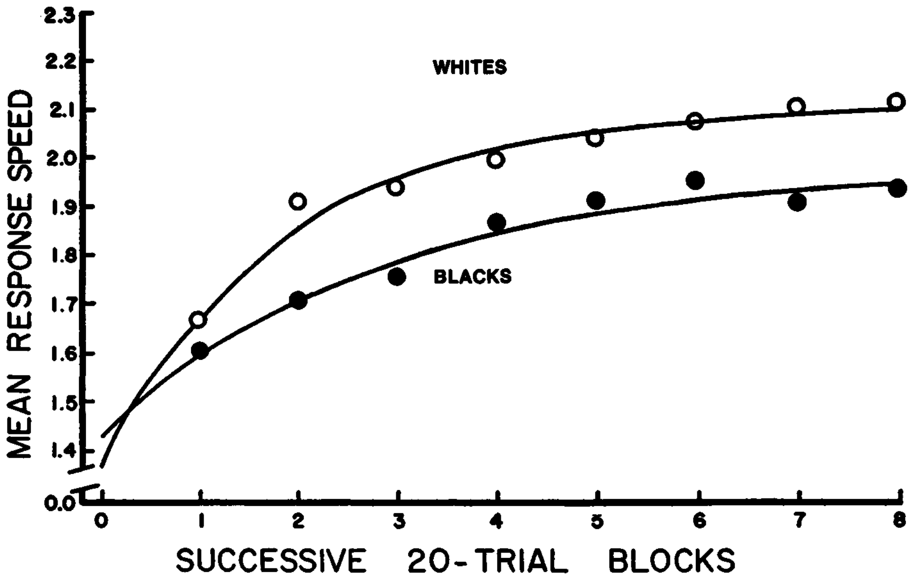

Figure 4.5. Mean response speed (reciprocal of RT in sec) in successive 20-trial blocks on a 4-choice reaction time test. Each curve is based on 106 children, with groups matched for age, sex, and pretest practice condition (from Noble, 1969).

(i.e., the reciprocal of RT in sec), are shown in Figure 4.5. Thus Noble's study is consistent with Poortinga's in showing a significant black–white difference in 4-choice RT under conditions of low S–R compatibility.

A study by Borkowski and Krause (1983) compared 20 black and 29 white children (grades two and three, ages 8 and 9 years) on simple and 2-choice RTs by means of a Gerbrand's reaction timer. The stimuli were red and green lights presented in the same aperture. This apparatus does not permit distinguishing between RT and MT. The black–white differences on the *WISC* IQ (based on Information and Vocabulary subscales) and the *Raven's Progressive Matrices* are .78σ and 1.55σ, respectively. These may serve as a basis for comparison with the black–white differences in RT which are shown below. As Borkowski and Krause have provided the split-half reliabilities of these RT measures, it is possible to correct the σ differences for attenuation, by dividing the σ difference by the square root of the reliability coefficient. The corrected values are shown in parentheses.

Simple RT:	.62σ (.67σ)	$p < .05$
2-choice RT:	.22σ (.28σ)	n.s.
Choice − simple:	−.35σ (−.61σ)	n.s.

Strangely, the only significant difference is on simple RT. The black–white difference's being larger on simple than on choice RT is at odds with the other studies reviewed here and has no obvious explanation. The black–

white difference of .62σ on SRT is considerably larger than the mean difference of .20σ between second and third graders, who differ 1.25 years in average chronological age. In this study, the correlations between RT and *Raven* scores were as follows (disattenuated values in parentheses):

	Black	White
Simple RT	−.58 (−.62)	+.43 (+.46)
Choice RT	−.60 (−.77)	(not given)

The apparent, inexplicable inconsistencies in these correlations may be attributable to sampling error. The *positive* correlation between SRT and psychometric *g* (*Raven* score) in the white sample is the only positive correlation between RT and psychometric *g* that I have come across in the entire RT literature. The correlations between intraindividual variability (*SD* of an individual's RTs over trials) in simple RT and *Raven* scores are striking: −.40 for black children and −.70 for white children. But the negative correlation in the white sample implies that there must be a negative correlation between RT and intraindividual variability in RT, which is opposite to every other correlation between these variables reported in the literature. These results of Borkowski and Krause should be shown to be replicable in another, larger study before anyone could reasonably attempt a theoretical interpretation of these peculiar findings. A detailed critique of this study appears elsewhere (Jensen, 1985a).

Studies of black–white differences in our laboratory have permitted distinguishing between RT and MT by using the RT–MT apparatus described earlier (see Fig. 4.2). In the first study, 99 black and 119 white first-year male students, ages 18 to 19, in a vocational college were given 30 trials at each of three levels of complexity on the RT–MT apparatus, 1, 4, and 8 light/buttons, corresponding to 0, 2, and 3 bits of information. The results are shown in Figure 4.6. On RT, the groups do not differ significantly in intercepts (a difference of only 3 msec at 0 bits), but the difference in slopes is significant ($t = 3.13$, $p < .01$). The black–white difference increases by approximately 10 msec per bit. The black–white difference of .63σ in mean median MT is significant ($t = 4.44$, $p < .001$). Also, intraindividual (i.e., between trials) variability (σ_i) of RT was significantly ($t = 3.50$, $p < .001$) greater for blacks than for whites, a difference of .54σ. A limitation of this study is that the groups are not representative of the white population or especially of the black population. Both groups have a restricted range in psychometric *g* (as indicated by the *Scholastic Aptitude* scores used for selecting vocational college applicants), with no very low or very high IQs. The white group is just slightly above the population average IQ for whites whereas the black group is about a standard deviation above the mean IQ of the general population of blacks. Under these conditions, the black sample's RT means should be expected to regress further away from the white group's RT means as the task complexity (i.e., number of bits of information) increases, as seen in Figure 4.6. These results may be interpreted as being

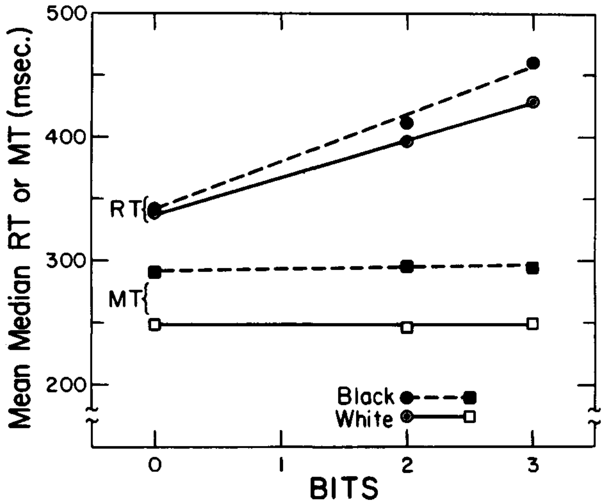

Figure 4.6. Mean median RT and MT of black ($N = 99$) and white ($N = 119$) vocational students, ages 18–19, tested on the RT–MT apparatus, shown in Figure 4.2.

consistent with those of Poortinga (and Noble) in showing a significant black–white difference on 8-choice RT, but not on simple RT. (The difference on 4-choice RT is nonsignificant, two-tailed $t = 1.71$, $p < .10$.) However, more problematic is the fact that these black–white differences in the RT–MT paradigm did not replicate in our laboratory in another study with similar samples.

Vernon and Jensen (1984) tested 50 black and 56 white male vocational college students on eight different speed-of-information-processing paradigms, including the RT–MT apparatus (Fig. 4.2). The groups were selected from a subject pool similar to that of the previous sample, but in this study the black and white groups differed about two-thirds of a standard deviation in scholastic aptitude. Both groups, however, were above the average of their respective populations. None of the parameters of the RT–MT paradigm showed significant correlations with the general factor of the *Armed Services Vocational Aptitude Battery (ASVAB)* or significant differences between the black and white groups. There is no plausible explanation, other than sampling error, for the complete failure of the second study to replicate the results of the previous study with the RT–MT apparatus, as the very same apparatus and procedures were used in both studies, with the exception that only half as many trials (15) were given in the second study instead of all 30 trials given in the first study. Examination of the data with respect to this procedural difference shows it as unable to explain the different outcomes. Statistical details of both studies are provided in Jensen (1987b).

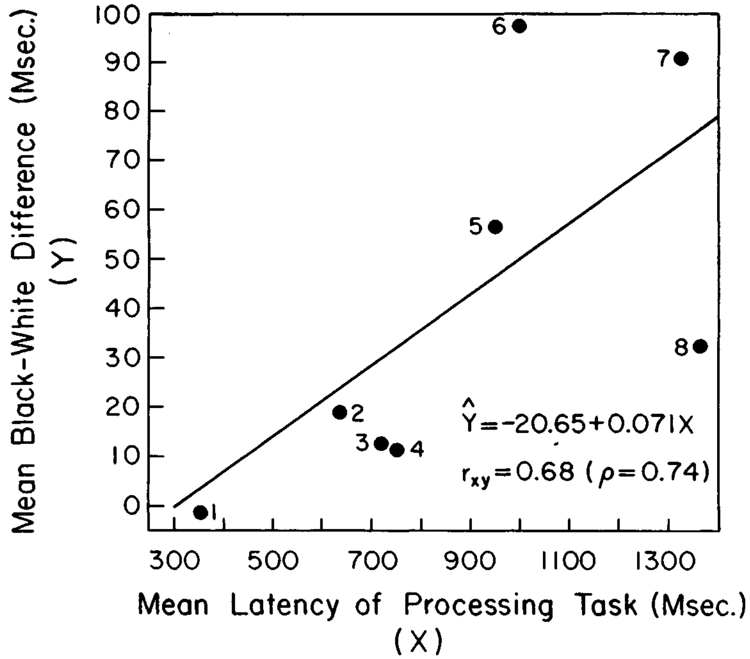

Figure 4.7. Mean black–white difference (msec) in RT to various processing tasks as a function of task complexity as indicated by mean of response latency on each task in the combined groups (from Vernon & Jensen, 1985).

The other, more complex, processing tasks used in the second study, however, showed highly significant effects. The entire battery of eight tasks yielded a shrunken multiple correlation with psychometric *g* factor scores of .47 in the combined groups, and the black–white mean difference on the general factor of the processing tests was .21σ as compared with a mean difference of .69σ on the *ASVAB*, a battery of ten paper-and-pencil tests heavily dependent on scholastic achievement. The processing tasks were RT–MT, *Digit Scan*, *Same–Different Words*, *Synonyms–Antonyms*, and *Dual Tasks* comprising Digits, Words, and Synonyms, as described in the previous section. The correlation of each task with the *ASVAB g* factor scores was quite closely related to the complexity of the task's processing demands, as indicated by the mean latency of response. A Pearson correlation of +.98 (rank–order correlation = +.93) was found between task complexity (i.e., mean RT) and the task's correlation with *g*. The black–white difference in mean RT on the various tasks is also related to task complexity, with a Pearson *r* of +.68 (rank–order correlation = +.74), as shown in Figure 4.7. These results are highly consistent with the general impression emerging from all these studies that the locus of the black–white difference exists only at the more complex levels of information processing. It is noteworthy that the two largest black–white differences occurred on the *Dual Tasks*, the largest RT difference ($.40\sigma$) occurring on Task No. 6, the *Dual Task* involving *Digit Scan* and *Same–Different Words*. The black–white differences on these two tasks when they were presented singly were considerably smaller ($.11\sigma$ and $.27\sigma$, respectively), suggesting that the black–white difference resides more in the active, or processing, aspect of working memory than in the passive, or storage, aspect.

Summary

This chapter has summarised how recently developed analytical techniques of "mental chronometry" can advance our understanding of the nature of the abilities that underlie individual differences and cultural group differences in performance on conventional tests of scholastic aptitude and achievement. Scores on such tests are believed to reflect some unknown amalgam of more elementary abilities. But test scores on such global measures of ability as IQ and scholastic achievement afford virtually no possibility of making the kind of analytical assessments called for by modern theories of ability, which attempt to understand aptitude and achievement in terms of information processing. Individual differences in proficiency of information processing can result from differences in any one or a number of different processes that cannot be separately assessed by the kinds of complex tasks that ordinarily compose the items of conventional psychometric tests of intelligence and scholastic aptitude.

Because general ability, or intelligence, as measured by conventional tests, is now viewed by most researchers in the field of human abilities as a composite effect of a number of distinct cognitive processes, research is aimed towards investigating whether individual differences in each of these processes can be reliably measured by means of chronometric techniques. These techniques are based on the measurement of a person's reaction time (RT) to simple tasks, which are specially devised to engage only particular processes. Most such elementary tasks devised to measure the most fundamental cognitive processes are so simple that they are within the capability of every person who is without marked sensory or motor handicaps. Therefore, individual differences do not consist of whether some individuals can and some cannot perform the tasks, as is the case with ordinary test items. Individual differences in these tasks can be measured only in terms of the *speed* with which the underlying processes occur, as represented by reaction times under varying task conditions.

The three main classes of basic processes reflected in performance on most conventional aptitude tests are:

1. *Apprehension*, *discrimination*, and *encoding* of stimuli. *Apprehension*, or speed of awareness, of a stimulus varies according to the degree of uncertainty of the nature of the stimulus or of the exact time or location of

its occurrence. *Discrimination* consists of responding to a stimulus that is distinguished from others by some particular attribute, when the stimulus occurs simultaneously with one or more other stimuli. *Encoding* is the attachment of a particular meaning, label, interpretation, or classification to the stimulus, such as responding (vocally or subvocally) with the appropriate utterance when, say, the symbol *A* occurs, or the numeral *5* or the word *cat*, or the colour *red*.

2. *Short-term memory capacity*, or "working memory." This component of the information-processing system is reflected in the amount of information that can be held or manipulated simultaneously (or within a brief time period) in full awareness. Associated processes are the *speed of search and retrieval* of an item of information held in short-term memory.

3. The *store of acquired knowledge*, *strategies* for dealing with specific types of problems, and other *learned complex skills stored in long-term memory*. Associated processes are the *speed of search and retrieval* of task-relevant information in long-term memory. Learned cognitive strategies also include the "*executive*" *processes*, which govern the deployment of the most appropriate routines for problem solving, monitoring one's own performance, and planning a course of action.

All of these types of information processes are involved in such complex tests as defining words in a vocabulary test, reading comprehension, and solving arithmetic problems. And they are involved in learning new skills. Individual differences and population differences in proficiency of performance in these complex tasks can be analysed and described in terms of the more elemental cognitive processes that underlie such complex abilities. Because the chronometric variables derived from a variety of elementary cognitive tasks reflect mainly cognitive *processes* rather than cognitive *content*, they would seem an especially valuable technique in the investigation of mental ability differences between populations that vary racially and culturally.

References

Anderson, J. R. (1983). Retrieval of information from long-term memory. *Science, 220*, 25–30.

Baddeley, A. D. (1968). A 3 min reasoning test based on grammatical transformation. *Psychonomic Science, 10*, 341–342.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. Bower (Ed.), *The psychology of learning and motivation* (Vol. 8). New York: Academic Press.

Berger, M. (1982). The "scientific approach" to intelligence: An overview of its history with special reference to mental speed. In H. J. Eysenck (Ed.), *A model for intelligence*. New York: Springer-Verlag.

Bligh, M. N. (1967). *A comparative study of visual and auditory responsiveness in two ethnic groups*. Unpublished master's thesis, University of Witwatersrand, Johannesburg.

Borkowksi, J. G., & Krause, A. (1983). Racial differences in intelligence: The importance of the executive system. *Intelligence, 7*, 379–395.

Brand, C. R. (1984). Intelligence and inspection time: An ontogenetic relationship? In C. J. Turner (Ed.), *The biology of human intelligence*. London: Eugenics Society.

Brand, C. R., & Deary, I. J. (1982). Intelligence and "inspection time." In H. J. Eysenck (Ed.), *A model for intelligence*. New York: Springer-Verlag.
Carroll, J. B. (1980). *Individual difference relations of psychometric and experimental cognitive tasks*. Chapel Hill: L. L. Thurstone Psychometric Laboratory, University of North Carolina.
Cavanagh, J. P. (1972). Relation between the immediate memory span and the memory search rate. *Psychological Review, 79*, 525–530.
Clark, H. H., & Chase, W. G. (1972). On the process of comparing sentences against pictures. *Cognitive Psychology, 3*, 472–517.
Eysenck, H. J. (1982). The psychophysiology of intelligence. In C. D. Spielberger & J. N. Butcher (Eds.), *Advances in personality assessment* (Vol. 1). Hillsdale, NJ: Erlbaum.
Galton, F. (1883). *Inquiries into human faculty and its development*. London: Macmillan.
Gettinger, M. (1984). Individual differences in time needed for learning: A review of the literature. *Educational Psychologist, 19*, 15–29.
Groen, G. J., & Parkman, J. M. (1972). A chronometric analysis of simple addition. *Psychological Review, 79*, 329–343.
Hendrickson, A. E. (1982). The biological basis of intelligence. Part I: Theory. In H. J. Eysenck (Ed.), *A model for intelligence*. New York: Springer-Verlag.
Jensen, A. R. (1974). Interaction of Level I and Level II abilities with race and socioeconomic status. *Journal of Educational Psychology, 66*, 99–111.
Jensen, A. R. (1980a). *Bias in mental testing*. New York: Free Press.
Jensen, A. R. (1980b). Chronometric analysis of intelligence. *Journal of Social and Biological Structures, 3*, 103–122.
Jensen, A. R. (1982a). The chronometry of intelligence. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, NJ: Erlbaum.
Jensen, A. R. (1982b). Reaction time and psychometric *g*. In H. J. Eysenck (Ed.), *A model for intelligence*. New York: Springer-Verlag.
Jensen, A. R. (1984a). Test validity: *g* versus the specificity doctrine. *Journal of Social and Biological Structures, 7*, 93–118.
Jensen, A. R. (1984b). Mental speed and levels of analysis. *The Behavioral and Brain Sciences, 7*, 295–296.
Jensen, A. R. (1985a). Methodological and statistical techniques for the chronometric study of mental abilities. In C. R. Reynolds & V. L. Willson (Eds.), *Methodological and statistical advances in the study of individual differences*. New York: Plenum.
Jensen, A. R. (1985b). The nature of the black–white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences, 8*, 193–219.
Jensen, A. R. (1985c). The black–white difference in *g*: A phenomenon in search of a theory. *Behavioral and Brain Sciences, 8*, 246–263.
Jensen, A. R. (1985d). Race differences and Type II errors: A comment on Borkowski and Krause. *Intelligence, 9*, 33–39.
Jensen, A. R. (1986). *g*: Artifact or reality? *Journal of Vocational Behavior 29*, 301–331.
Jensen, A. R. (1987a). The *g* beyond factor analysis. In J. C. Conoley, J. A. Glover, & R. R. Ronning (Eds.), *The influence of cognitive psychology on testing and measurement*. Hillsdale, NJ: Erlbaum.
Jensen, A. R. (1987b). Individual differences in the Hick reaction time paradigm. In P. A. Vernon (Ed.), *Intelligence and speed of information processing*. Norwood, NJ: Ablex.
Jensen, A. R. (1987c). Process differences and individual differences in some cognitive tasks. *Intelligence, 11*, 107–136.
Jensen, A. R., Schafer, E. W. P., & Crinella, F. M. (1981). Reaction time, evoked brain potentials, and psychometric *g* in the severely retarded. *Intelligence, 5*, 179–197.
Keating, D. P., & Bobbitt, B. (1978). Individual and developmental differences in cognitive processing components of mental ability. *Child Development, 49*, 155–169.
Klein, G. S., & Krech, D. (1952). Cortical conductivity in the brain-injured. *Journal of Personality, 21*, 118–148

Livson, N., & Krech, D. (1956). Dynamic systems, perceptual differentiation, and intelligence. *Journal of Personality, 25*, 46–58.

McGue, M., Bouchard, T. J., Jr., Lykken, D. T., & Feuer, D. (1984). Information processing abilities in twins reared apart. *Intelligence, 8*, 239–258.

Naglieri, J., & Jensen, A. R. (1987). Comparison of black–white differences on the WISC-R and the K-ABC: Spearman's hypothesis. *Intelligence, 11*, 21–43.

Nettelbeck, T. (1987). Inspection time and intelligence, In P. A. Vernon (Ed.), *Intelligence and speed of information processing*. Norwood, NJ: Ablex.

Noble, C. E. (1969). Race, reality, and experimental psychology. *Perspectives in Biology and Medicine, 13*, 10–30.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17*, 273–281.

Paul, S. (1985). *Speed of information processing: The Semantic Verification Test and general mental ability*. Unpublished doctoral dissertation, University of California, Berkeley.

Poortinga, Y. H. (1971). Cross-cultural comparison of maximum performance tests: Some methodological aspects and some experiments with simple auditory and visual stimuli. *Psychologica Africana* (Monograph Supplement No. 6).

Posner, M. I. (1966). Components of skilled performance. *Science, 152*, 1712–1718.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.

Posner, M. I. (1982). Cumulative development of attentional theory. *American Psychologist, 37*, 168–179.

Raz, N., Willerman, L., Ingmundson, P., & Hanlon, M. (1983). Aptitude-related differences in auditory recognition masking. *Intelligence, 7*, 71–90.

Scarr, S. (1982). Genetics and intelligence. In R. Sternberg (Ed.), *Handbook of human intelligence*. Cambridge University Press.

Spearman, C. E. (1927). *The abilities of man*. New York: Macmillan.

Stankov, L. (1983). The role of competition in human abilities revealed through auditory tests. *Multivariate Behavioral Research* (Monograph No. 83–1).

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*, 652–654.

Vernon, P. A. (1983). Speed of information processing and general intelligence. *Intelligence, 7*, 53–70.

Vernon, P. A. (1985). Relationships between speed-of-processing, personality, and intelligence. In C. Bagley & G. Verma (Eds.), *Personality, cognition, and values: A cross-cultural perspective of childhood and adolescence*. Calgary, Alberta: University of Calgary Press.

Vernon, P. A., & Jensen, A. R. (1984). Individual and group differences in intelligence and speed of information processing. *Personality and Individual Differences, 5*, 411–423.

Vernon, P. A., & Kantor, L. (1986). Reaction time correlations with intelligence test scores obtained under either timed or untimed conditions. *Intelligence, 10*, 315–330.