# QUANTITATIVE AND QUALITATIVE PROPERTIES OF AN INTELLIGENCE TEST: SERIES COMPLETION

LAZAR STANKOV
THE UNIVERSITY OF SYDNEY

ANNE CREGAN
THE UNIVERSITY OF SYDNEY

**ABSTRACT:** This article employs conjoint measurement and correlational procedures in order to study measurement properties of intelligence. The theoretical background of this article derives from recent developments in cognitive psychology and in psychometrics. In particular, the choice of independent variables—changes in working memory requirements and motivation —was influenced by the theory of processing resources and a view which sees intelligence as a conglomerate of elementary cognitive abilities. The three dependent variables (number correct, time to complete the test, and a derived measure representing the rate of working through the test) show support for the hypothesis that intelligence has quantitative structure. Therefore, the use of parametric statistics with these measures is justified. Correlations between the Raven's Progressive Matrices test and number correct scores from the Letter Series test show a systematic increase with working memory requirements (but not with motivation). This finding supports the view that task complexity depends on the amount of processing resources demanded by the Letter Series task.

In intelligence testing, test outcomes are usually expressed as number correct or accuracy scores. With the use of computers for test administration, speed of answering data is increasingly available. Scores that combine correctness and speed information can thus be calculated, providing a new type of measure of performance. What are the measurement properties of accuracy, speed and these new 'rate' scores? According to psychometric measurement theory, the

**Direct all correspondence to:** Lazar Stankov, Department of Psychology, The University of Sydney, Sydney, 2006, Australia.
**Learning and Individual Differences,** Volume 5, Number 2, 1993, pages 137–169.  ISSN: 1041-6080

justification for the use of test scores is that there is a psychological trait or *ability* which determines performance. Although the use of numerical test results presumes that the ability in question is a quantitative variable—i.e. that we can add different magnitudes of ability—the proof of this assumption is lacking. Using measures of intelligence, can we prove that intelligence is a quantitative variable?

Before investigating quantity, it is necessary to distinguish it from the qualitative nature of a variable. Within psychology, qualitative differences among variables can be detected by examining correlations. Thus, it has become acceptable to say that if variables have loadings on the same factor, they are of the same kind. The attribute of quantity is distinguished from quality by virtue of the fact that entities that differ in quality cannot be manipulated arithmetically—added, divided, etc.—because they differ in kind. Questions about the quantitative structure of intelligence thus cannot be answered in isolation from the understanding of its quality—either unitary or multifaceted.

Theories of cognition provide the conceptual framework for the following investigation of quantitative and qualitative properties of intelligence.

## COGNITIVE THEORY, QUANTITY/QUALITY AND INTELLIGENCE

Contemporary views of intelligence can be seen as stemming from one or the other of two positions that crystallized during the first half of this century. In the wholistic view of Spearman, Thurstone, Jensen, and Eysenck, intelligence and/or cognitive ability is a unitary trait. The alternative, atomistic view of Thomson, Thorndike, Humphreys, and Carroll is premised on numerous elementary cognitive operations that are sampled by psychometric tests. Although these two views are conceptually different, they provide equally acceptable accounts of many of the findings of psychology. Both, for example, can be used to argue for the existence of a general factor; Spearman's wholistic notion of mental energy can be re-interpreted as the sum-total of elementary cognitive abilities. The atomistic view, however, is reflected more strongly in the general thrust of recent work in cognitive psychology, and it also provides an easier approach to the questions of quantity and quality of intelligence.

In the atomistic theory, the number of elementary cognitive abilities is large; some believe a million or more (see Horn 1988). An attempt to enumerate those that have been subjected to experimental study was made by J.B. Carroll (1980). Although at one level of analysis all these elementary abilities have the same status, they are also categorized as sets of abilities that belong to distinct domains. Thus the perceptual processes of various sensory modalities, processes of memory, language, and components of analogical reasoning, etc. are seen as belonging to distinguishable groups of abilities. These groups are in turn grouped as broader organizations that encompass an even greater number of elementary abilities. Fluid and crystallized intelligence (Gf and Gc; see Cattell 1971; Horn 1988) are examples of such organizations. People differ in terms of the number of abilities they possess and/or in terms of the number of abilities that they are motivated to activate in the course of performance on any cognitive task. Every cognitive task, however, must demand a definite selection of elementary abilities for its execution, and it is reasonable to assume that typical intelligence test items will draw upon not only a number of elementary abilities, but also upon several groupings of abilities.

In the construction of cognitive tasks, it is possible to vary task demands in a systematic way so that they become either more or less demanding. This implies that we can *order* values of a treatment variable (task demands) in relationship to the dependent variable (test scores). This ordering, of itself, cannot ensure that either dependent or independent variable have quantitative properties. However, the theory of conjoint measurement shows that a cross of two independent variables can provide not only information about the order, but also about the *additivity* of values on all variables involved (i.e. the two independent variables and a dependent variable). Thus, satisfaction of the conjoint measurement assumptions provides evidence that variables are quantitative (see Michell 1990).

As mentioned earlier, the qualitative identity of a variable can be established through inspection of the correlations that different levels of the variable have with external measures. In theory, these external measures should be the ". . . entire repertoire of acquired skills, knowledge, learning sets, and generalization tendencies considered intellectual in nature and available at any one period in time . . ." , as Humphreys (1979) defines intelligence. Experience in mental testing has taught us that it is possible to choose a single test—e.g. Raven's Progressive Matrices test—that has a tendency to correlate with a wide range of tasks of fluid intelligence and use it as a proxy for the "entire set". From the point of view of psychometric theory, this approach is not entirely satisfactory since it can tell us only about relative changes in common and specific variances; there is no possibility of identification if some other factor that is not measured by the Raven's test should gain in prominence with changes in levels of difficulty.[1]

From the point of view of cognitive functioning, changes in the correlation of tasks with measures of intelligence reflect the fact that there are individual differences in the presence of elementary cognitive abilities that are relevant to the tasks and can be mobilized to cope with the tasks' demands. Stankov (1988) refers to experimental manipulations that lead to changes in mean levels of performance *and* in correlations as 'complexity' manipulations. Where no systematic changes in correlations take place only 'difficulty' manipulations are involved. Both difficulty and complexity are important concepts in this article. Difficulty (or structure on arithmetic means) provides the basis for quantitative considerations in conjoint measurement analyses. It is reasonable to assume that changes in task difficulty place demands on abilities from the same domain.

For a psychologist interested in individual differences an important question is: What task properties lead to increased correlations? The answer to this question can provide a clue about the nature of intelligence. It is possible that manipulations of difficulty, including those of the present article, will not lead to changes in correlations. We may postulate that changes in complexity imply the engagement of elementary abilities from the broader grouping of fluid intelligence, and that persons with higher fluid intelligence scores can bring more elementary abilities into play.

From a statistical point of view, changes in correlations are convenient indicants of changes in complexity under the assumption of equal variances and equal reliabilities across the levels of experimental manipulations. If the assumption of equal variances cannot be sustained, covariances in addition to (or instead of) correlations can be analyzed. If inequalities in error variances are a matter of concern, the use of a confirmatory factor-analytic or structural equation model instead of raw correlations (or covariances) can eliminate at least some alternative explanations. The aim of this work is to show that within-subjects manipulations can lead to systematic changes in between-subjects variability. A judicious use of different statistical procedures can illuminate these relationships.

The recent development in psychometrics and cognitive psychology means that both *quantitative and qualitative* properties of psychological variables *can now be tested empirically.* Three areas of psychology are relevant to the issues raised in this article: (1) The theory of conjoint measurement; (2) The concepts of limited processing capacity and working memory; (3) The psychometric concept of intelligence and issues related to the use of accuracy and speed scores in the measurement of intelligence.

In the following sections of this article we outline the general theory of conjoint measurement and develop formulations specific to the experimental conditions employed in our study; describe the experimental method and provide rationales for our choice of dependent and independent variables; and finally, we consider our empirical data in the light of conjoint measurement assumptions and the notion of quality.

## QUANTITATIVE PROPERTIES

Theoretical discussion of 'scales of measurement' and 'permissible statistics' initiated by the writings of S.S. Stevens (1946, 1951), has highlighted a fundamental difference between measurement in the natural sciences (e.g. physics) and the social sciences (e.g. psychology). The most cherished measures in physics employ either interval or ratio scales. They are quantitative in the sense that their values are ordered and they possess additive structure (see the Appendix for a more formal definition of these terms).

The nature of the majority of psychological measures, whether quantitative or otherwise, has not been established. This is because until recently the only known procedure for verifying quantitative structure in the measurement of a variable required the existence of a suitable concatenation operation. For example, the variable "length" can be shown to have additive structure through the concatenation operation of joining rods end to end. This procedure is not generally practicable in the measurement of psychological phenomena. There is some disagreement (e.g. Lord 1953) with Stevens' assertion that scale-type determines what statistics can be employed in order to make legitimate inferences and this issue is still being debated (see Michell 1986). Nevertheless, the value of showing a variable to be quantitative is undeniable, as it resolves the problem of permissible statistics by bypassing it altogether.

Recent developments in psychological scaling known in the literature as "conjoint measurement" (see Luce & Tukey 1964), provide an alternative procedure for establishing whether a given variable is quantitative. This is logically equivalent to the concatenation operation. Essentially, it involves the setting up of experimental conditions in accordance with a set of conjoint measurement assumptions, and examining whether the resulting structure on the means satisfies further conjoint measurement conditions. If it does, the property in question, say ability as measured by the performance scores of an intelligence test, is quantitative, making all operations of parametric statistics acceptable.

Although formal proof exists in Krantz, Luce, Suppes, and Tversky (1971) that satisfaction of the conjoint measurement assumptions implies that dependent and independent variables possess quantitative structure, we follow Michell's (1990) exposition, which is in tune with the present approach.

## ESSENTIAL ASPECTS OF CONJOINT MEASUREMENT

Conjoint measurement allows for the assessment of the structure of a variable L which, in turn, is a non-interactive function of two other variables M and W.[2] It applies specifically to those situations where neither L, M, or W is already quantified. We can view M and W as two crossed independent variables and L as a dependent variable. For the purposes of this article, let us also assume that both M and W have at least three levels each (i.e. $m_1$, $m_2$, $m_3$ for M, and $w_1$, $w_2$, $w_3$ for W). Figure 1 illustrates a cross of M and W. Each cell within this cross can be identified with respect to its marginal values. Thus, cell $m_1w_1$ is the top left-hand cell. Since in the present context L refers to the dependent variable, a measure of performance—say, arithmetic mean over many subjects—can be employed in order to test the conjoint measurement assumptions.

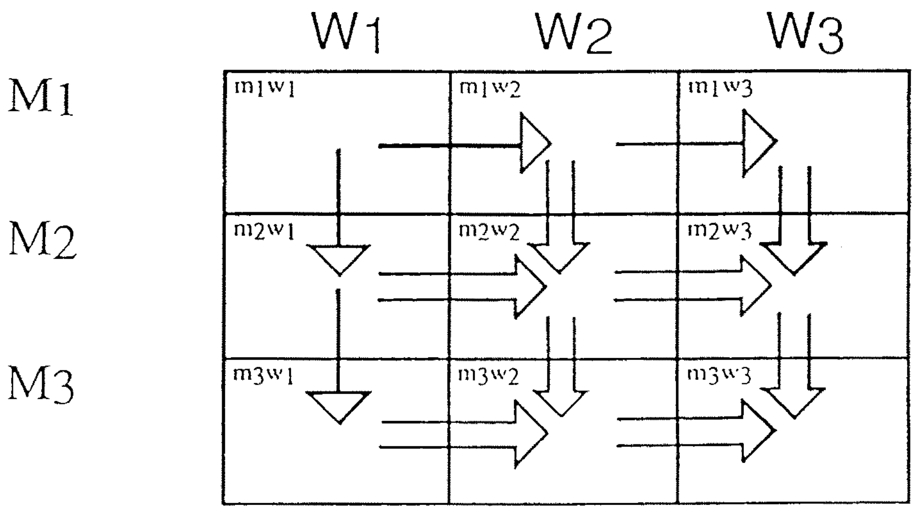
Assuming that the variable L possesses an infinite number of values, the main requirements for conjoint measurement are that:

L = f(M, W) i.e. L is some (say, additive) mathematical function of M and W;

There is a simple order $\geq$ upon the values of L (i.e. L is ordinal);

Values of M and W can be identified (i.e. objects may be classified according to the value of M and W they possess), and manipulated independently of each other. In other words, the value of each component can be realized without influencing the value for the other.

**FIGURE 1**
**Single cancellation conditions. Column labels W1 to W3 represent levels of the working memory placekeepers (WMP) factor, and row labels M1 to M3 stand for the motivation (M) factor. Cells within the cross are defined in terms of the marginal labels. These same labels are used in Tables 2 to 5 in order to describe particular tests of single and double cancellation. Lines with arrowheads within the cross of W and M illustrate tests of single cancellation. Premises are indicated by the lines and the conclusion is represented by a double line. Basically, satisfaction of the single cancellation condition establishes that cells in all rows and in all columns are ordered in exactly the same way.**




Then, if the simple order $\geq$ on L satisfies the conditions of:

1. *Single cancellation* (also known as *Independence.*): Using $m_1$, $m_2$ to designate any levels of M, and $w_1$, $w_2$ to designate any levels of W, a relation $\geq$ on L satisfies single cancellation condition if and only if:

$m_1w_1 \geq m_2w_1$ then $m_1w_2 \geq m_2w_2$ for any $w_1$; and
$m_1w_1 \geq m_1w_2$ then $m_2w_1 \geq m_2w_2$ for any $m_2$.

In words, single cancellation means that for any two rows, if a cell in a particular row is greater than or equal to a corresponding cell in the other row then all other cells in that row should be greater than or equal to their corresponding cells in the other row. Similarly, the orderings between the columns of a conjoint matrix should be the same regardless of which row is referred to. For a $3 \times 3$ cross, a list of tests of single cancellation for rows and columns is provided in Figure 1 and in the middle sections of Tables 2, 3, and 4.

Single cancellation requires that all rows and all columns show *exactly* the same ordering of the cells.

2. *Double cancellation:* A relation $\geq$ on L satisfies double cancellation if, for every $m_1$, $m_2$, $m_3$ in M, and $w_1$, $w_2$, $w_3$ in W,

if $m_2w_1 \geq m_1w_2$
and $m_3w_2 \geq m_2w_3$
then $m_3w_1 \geq m_1w_3$.

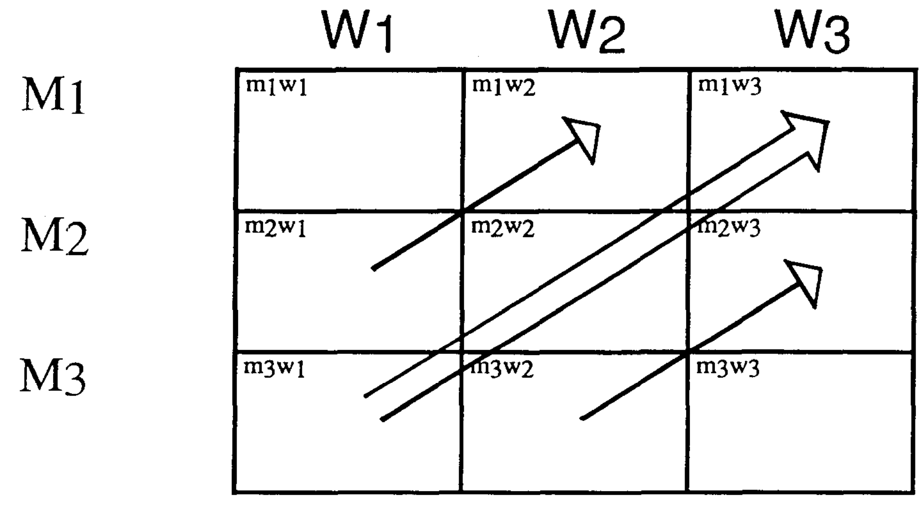
This test of double cancellation is depicted in Figure 2.

Then, for a suitably chosen set of dependent (L) and independent (M and W) variables, the satisfaction of the single (i.e. independence) and double cancellation conditions is supportive of a quantitative hypothesis. For the special case of a 3 by 3 data matrix, satisfaction of both single and double cancellation are sufficient for the existence of an additive representation (see Michell 1988).

## EXPERIMENTAL APPLICATION OF CONJOINT MEASUREMENT

The procedure of conjoint measurement is sufficiently general to be applicable to many areas of psychology. The main interest in any application of the conjoint measurement procedure resides in the structure of the dependent variable, *L*. We should note, however, that the success of this application depends on both the inherent nature of the variables and the actual measurement operations (chosen levels of independent and dependent variables). This may limit the domain of generalizability of our conclusions. For example, we may be interested in the quantitative structure of intelligence and, although it may be the

**FIGURE 2**
**Double cancellation condition. Column labels W1 to W3 represent levels of the working memory placekeepers (WMP) factor, and row labels M1 to M3 stand for the motivation (M) factor. Lines with arrowheads within the cross of W and M illustrate a test of double cancellation. Premises are indicated by the single lines and the conclusion is represented by a double line.**



case that intelligence itself is quantitative, if the operationalized index of intelligence used in the experiment does not reflect this quantitative structure, the overall result will be inconclusive. Prior knowledge of and careful choice of independent and dependent variables is therefore of crucial importance.

## DEPENDENT VARIABLES (L): THE OPERATIONALIZATION OF INTELLIGENCE

Since the use of intelligence tests and, as a rule, the scores as dependent variables in experimental studies of one kind and another involve statistical analyses depending on the calculation or arithmetic means, variances, and correlations, assurance of the quantitative nature of these measures of intelligence is highly desirable. As a modest beginning in the task of gaining this assurance, we chose the Letter Series, a much used test from L. Thurstone's studies of primary abilities for closer examination.

The general framework for our study is provided by the theory of fluid and crystallized intelligence (Gf/Gc) developed by Cattell (1971) and Horn (1988). This theory is based in part on hierarchical factor analysis. We should think of Gf and Gc as broad organizations of abilities, perhaps as factors that emerge when one carries factor analysis on correlations between primary abilities. In this context, the Letter Series is a good measure of the Inductive Reasoning primary factor and of fluid intelligence at the second order. In addition to the Letter Series test we also employed the Raven's Progressive Matrices test which is a measure of the Figural Relations primary factor and also a well-known measure of Gf at the second order. In the present study, this matrices test serves a control purpose—it can tell us whether the Letter Series test behaves as a proper measure of Gf in this sample of subjects (see Myors, Stankov, & Oliphant 1989; Stankov & Myors 1990). We did not use measures of any other broad abilities from the theory of fluid and crystallized intelligence (i.e. Gc, Gv, Ga, SAR, and TSR), since the theoretical links between these and the main dependent variables of this study are not strong. Furthermore, in previous work we used a battery of tests that covered a broad range of abilities and found that both easy and difficult items of the Letter Series test measure the same abilities—i.e. they do not differ in quality (Myors et al. 1989). This test is therefore a convenient vehicle for our purposes here.

In this article we explore the structure of three measures of performance on the Letter Series test: a. Number-correct; b. Time taken to work through all items of the test; and c. Rate i.e. number-correct score divided by the time (in minutes) taken to complete the test (a/b). Studies in our laboratory (e.g. Spilsbury, Stankov, & Roberts 1990) have indicated that rate may possess the desirable properties of more evenly spaced points along its scale of measurement than either of the component measures and a clear trend in increasing correlations with a separate measure of fluid ability with increasing difficulty. The Raven's Progressive Matrices scores allow us to compare the merits of all three measures. It will be of particular interest to replicate Spilsbury et al.'s (1990) findings with the 'rate' measure.

All three dependent measures satisfy the primary conjoint measurement conditions as they possess, at least in theory, an infinite number of values, and satisfy a simple order on these values. Our interest lies in establishing if all measures also satisfy the additional conjoint measurement assumptions of single and double cancellation. To check this, we need to select two other variables whose values can be identified (and for the sake of convenience, ordered) and independently manipulated, so that checks of single and double cancellation, as required by the conjoint measurement theory, can be applied.

## INDEPENDENT VARIABLES: WORKING MEMORY PLACEKEEPERS (W)

Simon and Kotovsky (1963) and Kotovsky and Simon (1973) devised a notation for describing both Letter and Number Series Completion Tests which can be implemented as a computer model, generating items such as those employed in studies of primary mental abilities by Thurstone (1938). A more recent work in this area was carried out by Butterfield, Nielsen, Tangen, and Richardson (1985). This notation provides a systematic way (through a set of well-defined rules) of exploring which aspects of the series completion items are responsible for the pronounced individual differences in performance.

The rules consist of constants (denoted $C$), variables (denoted $X$), and operators (denoted $+N$ ( )), enclosed between square brackets which correspond to the cycle length or period of the item. The values of the variables, once initialized, change from one cycle to the next according to the operator. For example, the following rule:

$$[X_1, + 2(X_1), C_1,]$$

generates Letter Series $M, B, O, B, Q, B, S, (B)$ if $X_1$ is set to $M$ and $C_1 = B$. It is easy to generate analogous series by choosing different values for $X_1$ and $B$. The work of Holzman, Pellegrino, and Glaser (1983) and Myors, Stankov, and Oliphant (1989) show that the difficulty of the series completion items depends on the operators used within a given period of a rule. Since it is assumed that the number of operators which can be monitored for a given item depends on the individual's working memory capacity, the number of "working memory placekeepers" (WMP) or "operators" can be interpreted as a theoretically-defined measure of the working memory requirements of a Letter Series item. The work of Myors et al. (1989) shows that as the number of WMPs (or operators) increases, the difficulty of the items increases. This, then shows that the values of this independent variable are ordered. Table 1 provides the list of rules for generating items with WMP1, WMP2, and WMP3 requirements used in this study. Although this manipulation places emphasis on 'correct' scores, we expect a decrement in speed as well, since more difficult items will require a longer time to complete.

**TABLE 1**
**Rules for Generating Series Completion Items.**[a]

| WMP | Rule | Example |
|---|---|---|
| 1 | $[X_1, X_1, +(2X_1)]$ | P,P,R,R,T,T,V, (V) |
| 1 | $[X_1, +2(X_1)]$ | F,H,J,L,N,P,R,(T) |
| 1 | $[X_1, +2(X_1), C_1]$ | J,Q,L,Q,N,Q,P,Q,R,Q, (T) |
| 1 | $[X_1, +2(X_1), C_1, C_2]$ | C,J,S,E,J,S,G,J,S,I,(J) |
| 2 | $[X_1, +2(X_1), X_2, +2(X_2)]$ | V,L,X,N,Z,P,B, (R) |
| 2 | $[X_1, +2(X_1), C_1, X_2, +2(X_2), C_2]$ | U,L,G,W,N,G,Y,P,G,A,(R) |
| 2 | $[X_1, X_1, +2(X_1), X_2, +2(X_2)]$ | F,F,O,H,H,Q,J,J,S,L,L, (U) |
| 2 | $[X_1, +2(X_1), X_2, +2(X_2)C_1]$ | L,Y,R,N,A,R,P,C,R,R,(E) |
| 3 | $[X_1, +2(X_1), X_2, X_2, +2(X_2), X_3, +2(X_3)]$ | Z,I,I,G,B,K,K,I,D,M,M,K,F,O,(O) |
| 3 | $[X_1, X_1, +2(X_1), X_2, +2(X_2), X_3, +2(X_3)]$ | W,W,P,F,Y,Y,R,H,A,A,T,J,C,C,(V) |
| 3 | $[X_1, +2(X_1), X_2, +2(X_2), X_3, +2(X_3), C_1]$ | T,Z,G,O,V,B,I,O,X,D,K,O,Z,F, (M) |
| 3 | $[X_1, +2(X_1), X_2, +2(X_2), X_3, +2(X_3)]$ | D,H,U,F,J,W,H,L,Y,J,(N) |

[a]Notation used in this Table is a modified version of Holzman et al. (1983).

## CHANGES IN MOTIVATION THROUGH INSTRUCTIONS (M)

Some recent theories of the processes underlying intelligence emphasize the availability of processing resources in performance on intelligence tests. A refinement of these theories provides for a distinction between data-limited and resource-limited processes (see Norman & Bobrow 1975). Since resource-limited processes are deemed to be of central importance for fluid intelligence, we wished to identify the task under study as resource limited. To do this we employed a manipulation of motivation as the second independent variable in this study. Basically, if performance on a task can be improved by inducing a subject to invest more mental effort in his or her problem-solving activity, the task is said to be resource-limited. If there is no improvement despite increased effort, the task is data-limited. Thus, if motivational manipulation produces changes in the level of performance, we have some assurance that we are dealing with central intellective resource-limited processing.

Our way of inducing an increased effort consists in asking subjects to perform as correctly as possible at their own pace, and then to try to do the same number of items as correctly as possible in under 75%, and then, in under 50%, of their own original time. Although this manipulation emphasizes speed, it should, as a result of a trade-off, affect accuracy as well. In order for it to be appropriate to compare number-correct scores across the three conditions of instruction, we allow our subjects to work past their time-limit until the completion of the test.

Our evidence to date (Stankov & Crawford 1993) has indicated that explicit instruction to work as fast as possible while maintaining the accuracy of the work does not change the qualitative nature of number-correct scores. That is, the scores tend to have the same correlation with the external variables under both care- and speed-emphasis conditions.

***The Issue of Interaction between Independent Variables.*** Since the application of conjoint measurement requires an absence of interaction between the independent variables, it is necessary to consider whether the relationship between motivation (M) and task difficulty (WMP) may involve interaction. The meaning of interaction here differs from that of experimental design literature. In the present context, an absence of interaction is defined as satisfaction of the single cancellation conditions. If all rows and columns of a matrix show the same ordering, variables M and W are non-interactive in regard to their effects on L. Obviously, it is possible for an interaction term in ANOVA to be significant in experimental design terms, but of no consequence for conjoint measurement. However, a significant cross-over interaction in ANOVA is an instance in which both experimental design and conjoint measurement procedures recognize an interaction. The appearance of such interaction in our experiment would be undesirable.

A possible confounding effect in a study of the relationship between difficulty and motivation is that an increase in task difficulty may have a motivating effect in itself. As the task becomes more difficult (e.g. the number of WMPs increases), a person may tend to invest more effort in its execution. Evidence of this phenomenon in cognitive tasks was reviewed by M. Eysenck (1982). Although we do not have prior information on the Letter Series task, it is rather unlikely that an interaction would produce cross-over effect—i.e. it is unlikely that subjects could, as a result of increased investment of effort in the difficult task, obtain higher scores on the difficult than on the easier task. Nevertheless, in order to diminish the likelihood of the appearance of a significant cross-over interaction we employed randomization of item presentation, such that on a given trial an item of any difficulty level may occur. In the absence of cues to the difficulty of the forthcoming item, it can be expected that approximately the same amount of effort will be invested across all items.[3]

**Identification of a Quantitative Variable.** The focus of our analysis is the postulate that one of our three dependent variables—the Letter Series test—is a measure of intelligence. Our choice of supplementary variables necessarily reflects current scientific understanding of the phenomena of interest. The W (working memory) and M (motivation) variables were chosen simply as known effective means for generating changes in levels of performance on the Letter Series test. If conjoint measurement assumptions hold, all three variables involved are quantitative, and therefore the independent variable that 'generates' scores on the Letter Series test, postulated to be intelligence, is quantitative.

Our argument, further, is that an atomistic theory of intelligence provides an appropriate theoretical underpinning of this experiment. The concept of working memory and its operationalization, the difficulty manipulations through the WMPs, accord with the limited capacity interpretation of intelligence. Manipulation of motivation (M), too, is an accepted means of establishing the presence of a resource-limited process. These two independent variables are construed as manipulations of test demands on elementary cognitive abilities that define intelligence. Since performance on the Letter Series test is a function of these two variables (and also a measure of intelligence), if conjoint measurement assumptions are satisfied, we can claim that central capacity, identified with intelligence is a quantitative variable.[4]

## QUALITATIVE PROPERTIES

**Expectations.** The pattern of correlations between Raven's Progressive Matrices test and scores on the Letter Series test under different treatment conditions can reveal two things. First, correlations with the Raven's Progressive Matrices test can inform us about the relationship between that accepted measure of intelligence and the three dependent variables of this study: number correct, speed of test-taking, and rate. On the basis of our previous results, we expect higher correlations with the number correct and rate scores and lower correlations with the speed scores. Second, it can show trends in correlations that are indicative, or not, of complexity as defined previously. In particular, within the range of the WMPs used in this study, we expect that as the number of WMPs increases, the correlation between Raven's Matrices and number correct scores of the Letter Series will also increase. Correlations of the time to complete the test with Raven's scores should either increase or remain the same across the WMP levels (Spilsbury et al. 1990; Stankov & Crawford 1993). As regards the independent variable of motivation, our expectations cannot be precise due to a lack of experience with this variable. Stankov and Crawford (1993) did not find much change in correlation under increasing time-pressured conditions, which suggests that motivation represents a difficulty rather than a complexity manipulation.

## METHOD

### SUBJECTS

Participants in this study were 99 (74 female) First Year students from the University of Sydney who took part in this experiment in order to obtain course credit. This population of students was tested with WAIS-R in previous work, the average Performance IQ score being about 110 and Verbal IQ average around 120.

### MATERIALS AND INSTRUCTION

The materials used were the Ravens Advanced Progressive Matrices Sets I and II, the Eysenck Personality Questionnaire comprising the four subscales of psychoticism, extraversion/introversion, neuroticism and lie scale, and Self-Description Questionnaire III (see Marsh 1987), designed for use with late adolescents and young adults, and measuring self-concept on the 13 scales of maths, religion, general, honesty, opposite sex, verbal, emotional, parent, academic, problem solving, physical appearance, same sex and physical ability. We do not present the data on personality measures in this paper (see Note 1).

In addition to this, a computer program was written to conduct the series completion task using letters under the various experimental conditions described below. Four set of 45 items were given under different instructions.

Initially the subjects were told:

> "You will be doing a task called 'letter series'. It is part of an intelligence test inventory. You will see a series of letters appear on the screen. These letters will be related by some rule, and it is your job to work out what letter should come next in the series."

Preceding the practice set the subjects were given the instructions:

> "We are going to start with a set of practice items. The important thing is to learn the method of doing the problems."

With the display of each item is the question, "What is the next letter in the series?" During each set, the item is displayed until the subject responds, and "Number of items to go" is displayed and updated after the subject's response to each item.

Subjects were then told, in the instructions preceding Set 1:

> "Now is the time to start the real test. The test is in three parts. In the first part, there will be 45 items. You must do these as accurately and as quickly as possible. Your performance will be timed."

During the first set, time taken in seconds was displayed and updated after each item as well as "number of items to go". This was recorded as motivation level 1 ($M_1$).

The instructions before Set 2 were:

> "In the second part of the test you will again have to complete 45 items, but you will only have 75% of the time that it took you in the last section. That is, you will only have ______ seconds to do 45 problems. The computer will display how much time you have left. Do each question as accurately as you can, but you must finish the 45 questions in the time available. Therefore, you must work more quickly than before in order to finish."

During this set, the "time to go" in seconds was displayed and updated after each item as well as "number of items to go". If the subject had not finished when the time limit had elapsed, the message: "You have run out of time!!—You must finish as quickly as you can" appeared at the bottom of the screen, and the subject was allowed to continue working until all items in the set had been attempted. This message was present until the set was completed. This is motivation level 2 ($M_2$).

Note that the subjects continued working until all 45 items had been attempted, so that the only difference between sets is the instructions and messages displayed on the screen.

The third set of items was exactly analogous to the second set except that subjects were told they had only 50% of their original time. Again subjects continue working until the set was completed. This is motivational level 3 ($M_3$).

Within each set of 45 items there were 15 items at each of the three WMP levels: 1, 2 and 3—the levels of our W variable. Within each WMP level, three different rules (see Table 1) were used with five items corresponding to each. Note that within each set of 45 items, the items were randomly ordered so that WMP levels and rules were mixed throughout the set.

In the construction of items, variables and constants were randomly initialized and in all cases the operator was held constant at +2 ( ) (i.e. the series always increased by 2 letters at a time) and the period was held constant at 3.5. Although holding the period constant results in series items of different lengths, series length has been shown to be an unimportant variable (Holzman et al. 1983).

## PROCEDURE

The experiment was conducted in two stages each of approximately one hour duration, no more than four weeks apart. In the first stage subjects were administered the Ravens Advanced Progressive Matrices Set I and II and completed the Eysenck Personality Questionnaire (EPQ) and Self Description Questionnaire (SDQ) III. The Ravens Advanced Progressive Matrices Test was given with standard instructions except that due to time constraints only 30 minutes were allowed to complete Set II rather than the usual 40.

The second stage of the experiment consisted of subjects being tested with the letter series program on Commodore 64 computers. It was emphasised to each subject that during part of the experiment there would be a time limit and it was important to try to do the task within the time limit. Subjects were allowed to adjust the screen if necessary and were encouraged to put on the headphones provided if distracted by noise. The experimenter was on hand in case of any queries or difficulties.

## DATA MANIPULATION

All data was scored on Commodore 64s. For each subject, we collected the information regarding her or his answer and also the time that elapsed between the item presentation and the press of the RETURN key. These were added to obtain the number correct score and also the time score. Number correct scores were recorded for both the time-limit and for the total-time of working on the test. All analyses of the present article are based on the total-time scores.

## RESULTS

### QUANTITATIVE PROPERTIES

In this section we consider the three dependent variables—number correct scores, time to complete the test, and 'rate' of performance—in the context of conjoint measurement. We treat each one of them in turn as the L variable, performing parallel analyses on each. These consist of traditional repeated measures analyses of variance with MANOVA, and paired t-tests for conjoint measurement comparisons. Working memory placekeepers and motivation through instructions are the two independent variables—the W and M variables respectively—in this study.

### NUMBER CORRECT SCORES

The top part of Table 2 presents arithmetic means and standard deviations for the "number-correct" scores under different combinations of treatment conditions. The first row of arithmetic means in Table 2a (and also in Tables 3a and 4a) contains information from the practice trials. The practice trials, like the actual test trials, show a clear decrement in performance over different W conditions, confirming that our procedure for affecting difficulty in the Letter Series test was successful.

The data relevant for checks of single and double cancellation conditions are in the nine cells representing the cross of M and W variables. It can be observed that the highest value in the table of arithmetic means is in the cell which is labelled m1w1 in Figure 1 and the lowest value is in the m3w3 cell. All other cells have values in between the two extremes. Since the number of items under each condition is fifteen, it is apparent that the easiest version of the Letter Series test was performed at the ceiling level for most subjects.

Prior to considering the results of the repeated measures analysis, it is necessary to note that the values of the standard deviations are "correlated" with the means—i.e. large means are accompanied by small standard deviations and small means have large standard deviations. A significant chi-square test result was obtained using Bartlett's test for the homogeneity of variances. (The same finding was obtained with the other two dependent variables—the 'time' and 'rate' scores.) This is to be expected given the ceiling level which was reached with the easy test items and raises questions regarding both the appropriate statistical analysis and substantive interpretation of outcome. We shall leave the latter for the discussion section of this article. The former determined our choice

**TABLE 2**
**Testing Conjoint Measurement Assumptions With the Number Correct Measure.**

*a. Arithmetic Means and Standard Deviations Under Different Levels of W and M*

| | | The Number of Working Memory Placekeepers | | |
|---|---|---|---|---|
| | | $w_1$ | $w_2$ | $w_3$ |
| Practice | | 14.09(1.60) | 12.14(3.12) | 10.67(3.89) |
| 'Normal' Speed | $m_1$ | 14.31(1.44) | 13.21(2.65) | 12.01(4.08) |
| 75% Time | $m_2$ | 13.88(1.98) | 12.54(3.04) | 11.44(4.15) |
| 50% Time | $m_3$ | 13.34(2.38) | 12.62(3.51) | 11.04(4.04) |

*b. Tests of Single Cancellation (Independence)*

| Rows: | t-test | Columns: | t-test |
|---|---|---|---|
| $m_1w_1 > m_1w_2$ | 4.987** | $m_1w_1 > m_2w_1$ | 2.856** |
| $m_1w_2 > m_2w_3$ | 6.759** | $m_2w_1 > m_3w_1$ | 3.518** |
| $m_2w_1 > m_2w_2$ | 5.940** | $m_1w_2 > m_2w_2$ | 3.813** |
| $m_2w_2 > m_2w_3$ | 4.446** | $m_2w_2 = m_3w_2$ | −.363 |
| $m_3w_1 > m_3w_2$ | 3.327** | $m_1w_3 > m_2w_3$ | 2.785** |
| $m_3w_2 > m_3w_3$ | 7.143** | $m_2w_3 > m_3w_3$ | 1.893* |

The ordering of cells is the same across each row:
$m_1w_1 > m_1w_2 > m_1w_3$

The ordering of columns:
Column 1: $m_1w_1 > m_2w_1 > m_3w_1$
Column 2: $m_1w_2 > m_2w_2 = m_3w_2$
Column 3: $m_1w_2 > m_2w_2 > m_3w_2$

*c. Tests of Double Cancellation*

| Conditions | t-test |
|---|---|
| $m_2w_1 > m_1w_2$ | 3.377** |
| $m_3w_2 > m_2w_3$ | 4.313** |
| $m_3w_1 > m_1w_3$ | 4.387** |

of MANOVA rather than the repeated measures univariate ANOVA for statistical analysis.

Basically, homogeneity of variance and, more generally, what is known as a "compound symmetry" condition needed to be satisfied for the appropriate use of ANOVA. Otherwise the traditional F-test is known to be too liberal—i.e. it tends to provide too high values. One way around this problem is to use MANOVA which does not depend on the compound symmetry assumption. This is the procedure we follow in the present article, and employ Pillai's F-values as the test statistic. Under these circumstances, it is important to note that it is impossible to use a single estimate for the error term for all 15 comparisons required by the conjoint measurement. Since within the theory of conjoint measurement these comparisons are defined in terms of the local levels only, we use a simple one-directional paired t-tests (with df = 98) for checking single and double cancellation requirements.

The aim of MANOVA analysis was to establish the significance of the experimental treatments and their interactions. This information, however, is not crucial for the conjoint measurement procedure. Both main effects in this analysis proved significant. Thus, for the number of working memory placekeepers (W), the Pillai's $F$-statistic $F(2,97) = 31.735$; $p < .01$; and for the motivation (M) independent variable, $F(2,97) = 18.222$; $p < .01$. Since the interaction term was not significant (Pillai's $F(4,392) = 1.807$; $p > .05$), we can conclude that there is an essential parallelism of lines connecting arithmetic means for the three levels of M over the three levels of W. As mentioned earlier, even the presence of significant interaction in the experimental design sense, would not be necessarily in conflict with the independence condition of the conjoining measurement theory. It is re-assuring, of course, that this interaction is in fact not significant.

The middle section of Table 2 contains tests of single cancellation for both rows and columns. As summarized at the bottom part of Table 2b, all rows clearly show the same ordering on arithmetic means. Orderings within columns are the same for WMP levels 1 and 3. The ordering for column 2 differs. Performance under the 75% of time condition (level *m*2) is statistically the same as performance under the 50% of time condition (level *m*3). Because of that difference in orderings between columns, we can conclude that the arithmetic means for the number correct scores do not support the single cancellation condition. The bottom part of Table 2 presents the tests relevant to the double cancellation. Since both premises of double cancellation show significant t-tests, and the direction of the conclusion section is congruent with the premises and is significant, we can conclude that the data support the double cancellation condition.[5]

These data do not unequivocally support the hypothesis that intelligence, as measured by the number correct scores of the Letter Series test, has quantitative structure. Since number correct scores are the most important, and often the only type of score derived from intelligence tests it is necessary to consider the contradiction in the results in more detail. First of all, from the purely statistical point of view, we should note again that the MANOVA interaction term is not significant, which means that for the overall analysis the essential equality of performance under $m_2w_2$ and $m_3w_2$ conditions is not sufficient to offset the conclusion that the lines connecting the arithmetic means presented in Table 2 are parallel. Also, even though the $m_2w_2$ is slightly higher than the $m_3w_2$ mean, the disconfirmation of the conjoint measurement requirement is due to equality, not to a significant opposing effect. These facts create uncertainty regarding the statistical basis for the conclusion that intelligence as measured by number correct scores is not quantitative. Secondly, from the substantive point of view it is hard to understand why the problem appeared only at the WMP level 2 and not at the other two levels. Our choice of particular WMP and motivation levels may be responsible—more pronounced differences between the levels of W and M could have led to a less ambiguous result. Finally, the relatively small number of items (15) under each condition may have contributed to the outcome. Therefore it seems reasonable to conclude that the quantitative structure of number correct Letter Series scores is essentially supported by the data.

## TIME TO PROVIDE ANSWERS TO ALL ITEMS

We carried out the same analyses using the time for responding as the dependent variable—i.e. the total time in seconds it took to answer all 15 items within a given block of items of the same WMP level. It has been our experience (Spilsbury et al. 1990; Stankov & Crawford 1993) that, as the items of a test of fluid intelligence increase in difficulty, the time needed to provide answers also increases. Since one of our experimental manipulations—i.e. change in motivation through instruction to work faster (M)—can be reflected in the time measure rather than in number correct it is valid to perform a parallel analyses using the time scores as the dependent variable.

Table 3a contains arithmetic means for the times under different treatment conditions. The general trend in this Table indicates that the fastest average speed to answer 15 items of the Letter Series test occurred in the $w_1m_3$ condition

**TABLE 3**
**Testing Conjoint Measurement Assumptions With the Time Measure (Seconds).**

*a. Arithmetic Means and Standard Deviations Under Different Levels of W and M*

| | | *The Number of Working Memory Placekeepers* | | |
|---|---|---|---|---|
| | | $w_1$ | $w_2$ | $w_3$ |
| Practice | | 205.34(63.70) | 465.61(177.23) | 672.66(313.668) |
| 'Normal' Speed | $m_1$ | 111.14(37.57) | 215.11(77.85) | 292.04(129.20) |
| 75% Time | $m_2$ | 97.49(31.71) | 167.63(51.16) | 237.12(88.27) |
| 50% Time | $m_3$ | 77.71(23.98) | 131.14(42.31) | 190.53(75.91) |

*b. Tests of Single Cancellation (Independence)*

| Rows: | t-test | Columns: | t-test |
|---|---|---|---|
| $m_1w_1 < m_1w_2$ | 16.318** | $m_1w_1 > m_2w_1$ | 4.734** |
| $m_1w_2 < m_1w_3$ | 9.295** | $m_2w_1 > m_3w_1$ | 9.148** |
| $m_2w_1 < m_2w_2$ | 16.166** | $m_1w_2 > m_2w_2$ | 8.429** |
| $m_2w_2 < m_2w_3$ | 11.198** | $m_2w_2 > m_3w_2$ | 11.768** |
| $m_3w_1 < m_3w_2$ | 17.436** | $m_1w_3 > m_2w_3$ | 5.828** |
| $m_3w_2 < m_3w_3$ | 10.604** | $m_2w_3 > m_3w_3$ | 8.043** |

The ordering of cells is the same across each . . .
Row: $m_iw_1 < m_iw_2 < m_iw_3$ Column: $m_1w_j > m_2w_j > m_3w_j$

*c. Tests of Double Cancellation*[a]

| | t-test |
|---|---|
| $m_2w_3 > m_1w_2$ | 2.892** |
| $m_3w_2 > m_2w_1$ | 8.530** |
| $m_3w_3 > m_1w_1$ | 10.621** |

[a]This test is based on the assumption that single cancellation holds and that means in cells of section (a.) of this Table are ordered so that they decrease as one moves from the top left-hand corner down and towards the right-hand side.

which involves the smallest number of working memory placekeepers, under the requirement to work twice as quickly as under "normal" conditions. Cell $m_1$ $w_3$, on the other hand, shows the longest average time.

As for the number correct scores, the two-way repeated measures MANOVA indicates that both motivation (M) and the number of working memory placekeepers (W) main effects are significant. Thus, changes in the number of working memory placekeepers (W) produce Pillai's $F(2,97) = 211.667$; $p < .01$, and changes in motivation induced through instructions (M) has Pillai's $F(2,97) = 157.169$; $p < .01$. With the time measure, however, the M $\times$ W interaction term is also significant (Pillai's $F(4,95) = 23.027$; $p < .01$. This latter outcome is largely due to the fact that the difference between performances under different motivational levels is not as pronounced at the lowest level of difficulty—at the $w_1$ level—as it is at the more difficult W levels. Plot of the cell means shows a fanning-out, not cross-over, effect. This causes a lack of parallelism between the lines connecting mean performance levels at different values of the M and W variables and therefore a significant interaction. We may note again that the significant interaction does not necessarily have a bearing on conjoint measurement.

Tests of the single cancellation condition for the time measure are provided in Table 3b. The order of cells is the same for all rows and t-tests are significant. Similarly, the order of cells is the same for all columns. It follows that the single cancellation condition is satisfied for the dependent measure of time. Tests of double cancellation are presented in Table 3c. We can see from this Table that all three t-tests have the same direction and that all of them are significant. This means that the double cancellation condition is satisfied in these data. We can conclude that our data support the conclusion that time needed to provide answers to the Letter Series problems is a measure that displays quantitative structure.

## THE RATE MEASURE

Justification for the use of a derived measure—in this case rate—has to come from evidence that this measure has useful properties over and above those of its components. In this study "rate" represents the number of correctly solved items per one minute of time. A useful property of the rate measure derives from the fact that in studies involving manipulations of difficulty, ceiling and floor effects may lead to small variability of number correct scores at the extremes. In these circumstances, the rate measure may provide a meaningful transformation leading to greater homogeneity of variances. Also, Spilsbury et al. (1990) found that the rate measure provides a more sensitive scale than simple number correct scores. These authors point out additional useful properties of this measure deriving from its correlations with the intelligence test scores which we shall consider in the next section of this article. Finally, the rate measure in particular should be considered if planned experimental manipulations affect the speed-accuracy trade-off, as is the case in the present study.

Table 4a contains the arithmetic means for 'rate' measures of the Letter Series test over different combinations of working memory placekeepers (W) and motivation (M) treatment conditions. The ordering of means in this Table is such that the highest value occurs under the $m_3w_1$ condition—i.e. when the difficulty is low and subjects are encouraged to work at 50% initial rate, they solve the largest number of items within a one-minute time period. The lowest arithmetic mean occurs under the $m_1w_3$ condition—i.e. when the test items are difficult and the subjects are not asked to work under specific time pressure. All other means are in between these two extremes.

As with the 'number correct' and 'time' measures, the MANOVA produced significant main effects for both the M and W variables. Thus, for the W variable Pillai's $F(2,97) = 429.242$; $p < .01$ and for the M variable, Pillai's $F(2,97) = 212.761$; $p < .01$. As with the 'time' measure, the interaction term is also significant, giving Pillai's $F(4,95) = 19.339$; $p < .01$. The significant interaction term

**TABLE 4**
**Testing Conjoint Measurement Assumptions With the Rate Measure (number correct/time in minutes).**

*a. Arithmetic Means and Standard Deviations Under Different Levels of W and M*

| | | The Number of Working Memory Placekeepers | | |
|---|---|---|---|---|
| | | $w_1$ | $w_2$ | $w_3$ |
| Practice | | 4.54(1.52) | 1.84(.90) | 1.18(.70) |
| 'Normal' Speed | $m_1$ | 8.55(2.84) | 4.12(1.53) | 2.88(1.46) |
| 75% Time | $m_2$ | 9.41(3.18) | 4.87(1.87) | 3.16(1.59) |
| 50% Time | $m_3$ | 11.21(3.63) | 6.21(2.62) | 3.84(1.89) |

*b. Tests of Single Cancellation (Independence)*

| Rows: | t-test | Columns: | t-test |
|---|---|---|---|
| $m_1w_1 > m_1w_2$ | 20.026** | $m_1w_1 < m_2w_1$ | 4.337** |
| $m_1w_2 > m_1w_3$ | 13.638** | $m_2w_1 < m_3w_1$ | 7.974** |
| $m_2w_1 > m_2w_2$ | 18.948** | $m_1w_2 < m_2w_2$ | 6.865** |
| $m_2w_2 > m_2w_3$ | 14.327** | $m_2w_2 < m_3w_2$ | 8.609** |
| $m_3w_1 > m_3w_2$ | 19.177** | $m_1w_3 < m_2w_3$ | 3.218** |
| $m_3w_2 > m_3w_3$ | 15.449** | $m_2w_3 < m_3w_3$ | 5.738** |

The ordering of cells is the same across each . . .
Row: $m_iw_1 > m_iw_2 > m_iw_3$ Column: $m_1w_j < m_2w_j < m_3w_j$

*c. Tests of Double Cancellation*[a]

| | t-test |
|---|---|
| $m_2w_1 > m_3w_2$ | 12.270** |
| $m_1w_2 > m_2w_3$ | 8.956** |
| $m_3w_3 > m_1w_1$ | 19.049** |

[a]This test is based on the assumption that single cancellation holds and that means in cells of section (a.) of this Table are ordered so that they increase as one moves from the top right-hand corner down and towards the left-hand side.

appears to derive from the finding that at the most difficult level—the $w_3$ level—performance does not vary under differing motivational manipulations as much as it does under the other two difficulty conditions. Again, in itself this finding does not threaten the non-independence assumption of the conjoint measurement.

Tests relevant to the conjoint measurement conditions are presented in the middle and bottom sections of Table 4. It is readily apparent that both single and double cancellation conditions are satisfied by the rate measure and therefore the conclusion is as it was with the time measure, that our data support a quantitative structure for the rate measure.

## CONCLUSION

In order to gain further insight into the significance of the obtained results, it is useful to consider the probability of obtaining the outcome satisfying single and double cancellation conditions within a 3 by 3 matrix. The 9 cells of such a matrix can be ordered in $9! = 362{,}880$ different ways. For single cancellation to be satisfied, rows and columns must be ordered in parallel. That can happen if order is fixed by the marginal orders. The number of marginal orders for both rows and columns is 3!. Thus, there are $3! \times 3! = 6 \times 6 = 36$ different ways that rows and columns are ordered in parallel. However, even if rows and columns are ordered in parallel, there are still 42 different rank orders for each marginal ordering of rows and columns. The number of orderings of 9 cells that satisfy single cancellation condition is $36 \times 42 = 1512$. Note, however, that only 36 out of the 42 orderings satisfy double cancellation condition. Thus, the number of orderings that satisfy both single and double cancellation conditions is $36 \times 36 = 1296$. The likelihood of any ordering satisfying both single and double cancellation conditions is $1296/362{,}880 = 1/280$.[6]

Of the three measures of the dependent variable of this study (i.e. Letter Series performance), two—'time' and 'rate'—show the pattern on means indicative of an underlying quantitative process. The results with the third measure—number correct scores—are somewhat equivocal, but we believe that they too support a quantitative structure of intelligence.

## QUALITATIVE PROPERTIES

### CORRELATIONS WITH RAVEN'S PROGRESSIVE MATRICES TEST

The data of this study also allow us to consider correlations between performance on the Letter Series test and measures of intelligence obtained with the Raven's Progressive Matrices test. These are presented in Table 5. For these data,

**TABLE 5**
**Correlations Between Raven's Progressive Matrices Test Scores and Dependent Variables Derived From The Letter Series Test.**

| *a. Sum of Scores Over All Treatment Conditions* | | *Scores From The Letter Series Test* | | |
|---|---|---|---|---|
| | | *Accuracy* | *Time* | *Rate* |
| Raven's Matrices Test | | .53 | −.29 | .54 |
| *b. Number Correct Scores* | | *Working Memory Placekeepers (W)* | | |
| | | $w_1$ | $w_2$ | $w_3$ |
| Practice | | .253 | .556 | .607 |
| Motivation (M): | | | | |
| Normal Speed | $m_1$ | .125 | .457 | .567 |
| 75% Time | $m_2$ | .224 | .423 | .559 |
| 50% Time | $m_3$ | .299 | .459 | .447 |
| Average: | | .235 | .474 | .555 |
| *c. Time Scores* | | *Working Memory Placekeepers (W)* | | |
| | | $w_1$ | $w_2$ | $w_3$ |
| Practice | | −.247 | −.269 | −.255 |
| Motivation (M): | | | | |
| Normal Speed | $m_1$ | −.396 | −.276 | −.240 |
| 75% Time | $m_2$ | −.388 | −.161 | −.015 |
| 50% Time | $m_3$ | −.369 | −.205 | .014 |
| *d. Rate Scores* | | *Working Memory Placekeepers (W)* | | |
| | | $w_1$ | $w_2$ | $w_3$ |
| Practice | | .308 | .505 | .550 |
| Motivation (M): | | | | |
| Normal Speed | $m_1$ | .409 | .440 | .527 |
| 75% Time | $m_2$ | .477 | .450 | .512 |
| 50% Time | $m_3$ | .472 | .476 | .394 |

difference between any two correlation coefficient within the same subset of variables that is greater than .17 is significant at the .05 level.

One reason for considering these correlations is in order to establish that, in the present sample, the Letter Series test behaves as a proper fluid intelligence test—a low correlation with the Raven's Progressive Matrices test would suggest that something unusual had happened in the study. Table 5a shows correlations between scores on the Raven's test and overall scores (i.e. summed over all nine experimental conditions) of the three measures used in this study. The highest correlations occur with the rate and number correct scores, and the size of these correlations ($r$ = .53 and .54) confirms the Letter Series test as a measure of fluid intelligence.

The time measure has the lowest correlation with the Raven's test ($r = -0.29$). This correlation is probably slightly inflated since the Raven's test was administered with a time limit 10 minutes shorter than that recommended in the manual. Nevertheless, it is clear that the time score did not measure the same trait as the number correct score (the correlation between the number correct and time scores of the Letter Series was −.31). The correlation of the rate scores with the Raven's is virtually the same as that of the number correct scores, so that in this study calculation of rate did not produce a measure superior to the number correct scores for measuring fluid intelligence.

Trends in the correlations across different treatment conditions provide additional information about the relationship between the different measures. These are apparent in the lower sections of Table 5—i.e. Table 5b, 5c, and 5d. It is convenient to consider the correlational trends for the three measures of the dependent variable separately.[7]

## CORRELATIONS BETWEEN THE RAVEN'S PROGRESSIVE MATRICES TEST AND THE LETTER SERIES TEST UNDER DIFFERENT WMP CONDITIONS

We shall first look at the changes occurring in WMP. It is apparent from Table 5b that correlations between the Raven's test scores and number correct scores increase as the number of working memory placekeepers increases.[8] This holds for the first two rows of the motivation variable and there is a very slight drop in correlation in the third row at the most difficult WMP condition. On the average, correlations show a clear increase across the three levels of WMP. Thus, WMP is an example of a 'complexity' manipulation (Stankov 1988).

Correlations between the Raven's test scores and time measures of the Letter Series test tend to decrease as difficulty increases: people with high scores on fluid intelligence tend to work faster than average when test items are easy, but when the test becomes more difficult they perform at much the same speed as people who have lower fluid intelligence scores. The trend in correlations is as consistent (but opposite in direction and somewhat smaller in magnitude) with the time measure as it was with the number correct measure. From the opposing trends for the number correct and time measures, it is not surprising that the rate measure, which is a combination of these two scores, does not show a pronounced trend in correlations over the levels of the W variable especially under the high-pressured $m_2$ and $m_3$ conditions.[9]

The present results with the rate measure are different from those obtained by Spilsbury et al. (1990), in which both number correct *and* speed scores showed higher correlations with outside measures of intelligence under the more difficult conditions of the four-term series problem. The rate measure was recommended for future use on the basis of these findings. The present results also differ from those of Stankov and Crawford (1993) who found no change in correlation between fluid intelligence scores and time measures under conditions of increasing difficulty, although increases in correlations with number correct measures were reported. Stankov and Crawford (1993) therefore recommended use of the number correct measure, rather than the rate measure, in the study of complexity manipulations. On the basis of the present findings, we should recommend the use of both number correct and time measures and, with considerable caution, the rate measure. All studies, obviously, support the use of the traditional accuracy scores in the assessment of cognitive ability.

## CORRELATIONS BETWEEN THE RAVEN'S PROGRESSIVE MATRICES TEST AND THE LETTER SERIES TEST UNDER DIFFERENT LEVELS OF MOTIVATION

By and large, changes in motivation lead to small and less systematic changes in correlation between Letter Series and Raven's test scores than do changes in WMPs. With a couple of weak exceptions, correlations for the three measures either remain essentially the same or change unsystematically across the levels of motivation (M).

Since, as we have seen earlier, increased levels of the M variable lead to a decrement in the overall level of performance, we can follow Stankov (1988) in postulating that the M variable represents a "difficulty" manipulation in this study. This is in agreement Stankov and Crawford's (1993) finding that instructions to work as fast as possible, as opposed to working as accurately as possible, did not differentially affect correlations with intelligence.

Conclusions drawn from the analyses of correlations differ from those based on the overall level of performance (i.e. arithmetic means) in indicating that number correct scores should be the preferred measure in studies of intelligence. Of the three measures, these scores are the most sensitive to complexity manipulations and have the highest correlation with an outside measure of intelligence. They do not show systematic changes under different motivation conditions. Time taken on the test showed systematic changes with WMP and no change with M, as did number correct scores, but had a relatively low correlation with the outside measure of intelligence, while rate score was no better than number correct score in terms of size of correlation with the Raven's, and had a serious defect in being insensitive to both WMP and M manipulations.

From a correlational point of view, the usefulness of the time and rate scores is still open to question, since the present study shows correlational trends which differ from those of our previous work. It is obvious that further research into a variety of measures in studies of abilities is needed.

## STRUCTURAL EQUATION MODELING OF ACCURACY SCORES

In order to investigate qualitative structure in more detail we carry out further analyses of the number correct scores only. Our aims in this section are as follows. First, since it is apparent from the previous Tables that differences exist between variances of the Letter Series test under different treatment conditions, we wish to present covariances here. These covariances can be compared with the corresponding correlations from the top part of Table 5. Changes in correlations have been accompanied by changes in variances (see the diagonal in Table 6). As a result, changes in covariances across the treatment conditions are much more dramatic in appearance than changes in correlations. Since we can no longer see the trend as clearly as we did in the correlations, it is still true that increase in WMP levels lead to stronger relationship between the Raven's and Letter Series tests?

Second, we wish to establish if common parts of Raven's and Letter Series test scores display the same trend as raw correlations. This is important not only because of the interest in a factor-analytic model per se but also because of the close link between this model and the true-error score model which provides the basis of reliability measurement. For this purpose we analyse the covariance matrix of Table 6 with the COSAN—a covariance structure analysis program developed by McDonald (1978). A one-factor model was fitted using maximum likelihood procedure and the resulting solution is presented in the right-hand side of that Table. For comparison purposes, loadings on the common factor are arranged in the pattern analogous to that of Table 5 in the lower section of Table 6. Although the pattern on covariances differ from the patterns on correlations, it is nevertheless clear that WMP levels show an increase in relationship between the common parts of Raven's and Letter Series scores while, as before, changes in motivation lead to less systematic effects.

Third, by using the structural equation model one can constrain elements of the common factor vector to be equal. By comparing changes in goodness-of-fit indices under unconstrained and constrained conditions, it is possible to establish whether motivation or WMP have stronger effects on qualitative properties. For the unconstrained solution in Table 6 the chi-square value was 138.29 with 35 degrees of freedom.[10] The logic of the next two analyses is based on the assumption that constraints will produce a less satisfactory fit. The question is whether that fit will be worse if we constrain across the levels of WMP or across the levels of motivation. To obtain an answer to this question we first constrained values in each row (i.e. WMP levels) of the bottom section of Table 6 to be equal. This produced a chi-square value of 417.68 with 41 degrees of freedom. By constraining the values in each column (i.e. motivation levels) to be equal chi-square value becomes 184.54 ($df = 41$). It is obvious that the effects are both more systematic and much stronger for the WMP manipulation than they are for the motivational manipulation.

## DISCUSSION

The main finding of our study is that two measures of performance of the Letter Series test—rate scores and the amount of time spent—satisfy the single and double cancellation conditions of conjoint measurement. Although number cor-

**TABLE 6**
**Number Correct Scores: Covariances between the Raven's and Letter Series Tests under Different Treatment Conditions and Factor Loadings on the Common Factor. Right-hand Side Presents a One-factor Solution Based on Covariances.**

| | *Raven's* | $m_1w_1$ | $m_1w_2$ | $m_1w_3$ | $m_2w_1$ | $m_2w_2$ | $m_2w_3$ | $m_3w_1$ | $m_3w_2$ | $m_3w_3$ | *Common* | *Unique* |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Raven's | 33.180 | | | | | | | | | | 2.997 | 24.195 |
| L. S.: $m_1w_1$ | 1.041 | 2.074 | | | | | | | | | 0.715 | 1.562 |
| L. S.: $m_1w_2$ | 6.972 | 2.137 | 7.026 | | | | | | | | 2.378 | 1.373 |
| L. S.: $m_1w_3$ | 13.303 | 2.364 | 8.733 | 16.602 | | | | | | | 3.709 | 2.843 |
| L. S.: $m_2w_1$ | 2.560 | 1.855 | 3.546 | 4.991 | 3.924 | | | | | | 1.402 | 1.960 |
| L. S.: $m_2w_2$ | 7.397 | 1.892 | 6.559 | 9.872 | 4.035 | 9.210 | | | | | 2.711 | 1.861 |
| L. S.: $m_2w_3$ | 13.365 | 2.339 | 8.741 | 14.873 | 5.350 | 10.239 | 17.229 | | | | 3.795 | 2.824 |
| L. S.: $m_3w_1$ | 4.092 | 1.830 | 4.712 | 6.558 | 3.644 | 5.416 | 6.683 | 5.656 | | | 1.898 | 2.054 |
| L. S.: $m_3w_2$ | 9.278 | 1.958 | 7.511 | 11.177 | 4.310 | 8.310 | 11.121 | 6.623 | 12.321 | | 3.098 | 2.723 |
| L. S.: $m_3w_3$ | 10.405 | 2.365 | 8.573 | 13.867 | 4.770 | 9.988 | 14.512 | 6.853 | 11.903 | 16.304 | 3.708 | 2.555 |

*Common factor loadings within the cross of independent variables:*

| | | *Working Memory Placekeepers (W)* | | |
|---|---|---|---|---|
| | | $w_1$ | $w_2$ | $w_3$ |
| *Motivation (M):* | | | | |
| Normal Speed | $m_1$ | 0.715 | 2.378 | 3.709 |
| 75% Time | $m_2$ | 1.402 | 2.711 | 3.795 |
| 50% Time | $m_3$ | 1.898 | 3.098 | 3.708 |

Raven's loading on the common factor: 2.997

rect scores provide slightly poorer data, they are still supportive of these conditions. Overall, these findings support the view that fluid intelligence has a quantitative structure. It follows that we can legitimately employ all parametric statistical tests with the three measures of performance on the Letter Series test.

There are two points that we wish to make in qualification of this conclusion. First, it is important to emphasize that support for the quantitative structure of intelligence relates only to the particular independent and dependent variables employed in the present study. Twenty or more levels of each independent variable instead of three, allowing examination of not only single and double, but also of triple, quadruple, etc. cancellation conditions, would certainly provide more convincing evidence. Nevertheless, if the structure of these variables were *not* quantitative, our crude procedure would detect it. Second, we acknowledge that our results cannot generalize to all tests of intelligence. But if they generalize to measures of fluid intelligence, this much is enough: it provides justification for using parametric statistics which have many important advantages over the alternatives.

Other important findings of our study derive from the correlations between the three dependent measures of the Letter Series test and the Raven's Progressive Matrices scores. Overall, number correct and rate scores had higher correlations with Raven's Matrices than did time scores. The correlations of number correct scores changed in a systematic way as the number of working memory placekeepers (WMPs) of the items increased, indicating that changes in the number of WMPs produce a complexity manipulation. On the other hand, the effects of manipulation of motivation on correlations was weak, suggesting that this is a difficulty manipulation.

The correlations for number correct and time scores showed opposing trends. Thus, whereas number correct scores showed higher correlations with the Raven's Progressive Matrices test as the number of WMPs increases, correlations of time scores tended to decrease. As a consequence, correlations of rate scores appeared insensitive to changes in the WMP levels. This contradicts our previous finding with rate measures, and cautions against indiscriminate use of scores derived from number correct and time measures in studies of human abilities.

The analysis based on covariances of the number correct scores and using the structural equation modeling method produced the result which is in agreement with the findings based on correlations. Thus, common parts of the variables show the same trends as those observed with correlations—i.e. systematic and strong increase in common processes of Raven's and Letter Series tests across the levels of WMP and weak unsystematic changes across the levels of motivation.

## SOME CONCEPTUAL ISSUES

Although we believe that there is much to be gained through a systematic application of the conjoint measurement procedure to the study of individual differences in cognitive abilities, we are aware that it will not be easy to design and execute such studies in a comprehensive manner. The difficulty does not lie in statistics, scaling, or in measurement theory, but rather in the problem of devising ways to systematically vary two task characteristics in order to obtain a particular structure on the performance measure. The appropriate conditions for the application of conjoint measurement require more than a superficial knowledge of the three variables involved. However, if a cognitive theory—like the limited processing capacity theory which provided a content for the manipulation of WMP and motivation—were able to furnish guiding principles, not only measurement properties but also basic cognitive processes of intelligence might be studied.

Even if task characteristics amenable to experimentation can be identified, the finding of appropriate levels to produce a sufficient range of scores for the measures or dependent variables might be a problem, and might be an impossible task in some cases. For example, if we had decided that the lowest WMP level should be WMP3 and had used Letter Series at the WMP4 and WMP5 levels in the study, manipulation of the motivation variable might have produced no observable effects in that there may be no change in performance between $m_2$ (i.e. working under the 75% of "normal" time) and $m_3$ (i.e. working under the 50% of "normal" time) levels at those levels of complexity.

In the present study, though a ceiling effect is present, the appropriateness of MANOVA statistical procedures and conclusions regarding quantity are not in question. However, given the sensitivity of correlations to restriction in range, our qualitative results may be criticized on the grounds that the lower correlations at the easy task level are statistical artefacts. While we cannot provide evidence at this stage to counter this argument, we wish to point out that rules relating to restriction in range apply to tests with a known distribution in the population given to a sample with a narrower range of abilities. In this study we have the same sample tested under different conditions. Although, from qualitative information, the different levels of the WMP factor measure things in common, it is also clear from the correlations that, at easier levels of WMP, specific variance is increased at the expense of common variance. Performance on the Letter Series test at each WMP level may involve a considerable and differing number of distinct elementary abilities. In other words, boundaries between both qualitative and quantitative categories are fuzzy. As performance at each level is somewhat different from the other levels, the restriction in range argument looses its force while, at the same time, the essential aspect of the argument for qualitative similarity remains valid. Further evidence that restriction in range may be of little importance derives from the structural equation modeling of the covariance matrix for the number correct scores.

**Atomistic Theory Revisited.** An atomistic theory of the structure of intelligence provides a theoretical background for our work. The quantitative structure of intelligence can be understood in terms of elementary abilities providing convenient units of measurement for intelligence. Consistent with this theory are findings that decrements in performance on the Letter Series test that parallel changes in levels of the motivation (M) independent variable result from a call for the investment of more abilities of the *same kind*. Manipulation of WMP, leading to a systematic increase in correlation with other measures of intelligence reflected a call on elementary abilities drawn from a wider grouping—i.e. that of fluid intelligence.

Several of our recent studies provide clues to the possible nature of these elements and our findings with the manipulation of WMPs are in agreement with these studies. We shall briefly introduce this additional work here. Following the early work of Wittenborn (1943) and Stankov (1983), Crawford (1991) investigated elementary abilities involved in measures of sustained attention. Although simple in nature, these tests showed correlations with fluid intelligence that were of the same order of magnitude as the correlations shown by the Letter Series test. As with the present work involving the WMP manipulation, it was possible to decompose the sustained attention tasks into smaller subtasks and study the ingredients that lead to the increase in correlation. Stankov and Crawford's (1993) study suggests that the basic cognitive processes that lead to an increase in correlation have to do with the number of elements and relations one has to keep in mind while solving a putative test of sustained attention. The number of working memory placekeepers (WMP) in the present study has the same interpretation. Each level of WMP differs from the immediately preceding or following levels in terms of the number of additional relations (operators) that have to be taken into account. The elements and relations in all these studies are relatively simple; the number of these elements and relations is what seems to be operative. The term "working memory" is a good description of what is involved at the locus of mental process, while "complexity" is the term we have chosen to describe the matching test characteristic.

The atomistic view sketched above is psychological in the sense that elementary cognitive operations (or "bonds" in G. Thomson's sense) are not linked to physical systems. On the evidence available now, both the number of neurons (i.e. capacity) and the efficiency of their functioning determine psychological processing. Neural efficiency hypothesis has been supported recently by the finding that oxidative metabolism of glucose appears to be affected in people with poor performance on fluid intelligence tests (Stankov & Dunn 1993).

**ACKNOWLEDGMENTS:** We are grateful to Joel Michell for his generous help and advice during all stages of work on this project. Discussions by the members of the Individual Differences Seminar at the University of Sydney influenced many aspects of our thinking about the subject matter of this article. We are also thankful to David Grayson for his suggestions regarding some statistical aspects of this study.

## APPENDIX: QUANTITATIVE VARIABLES

A variable Q (with values X, Y, and Z) is said to be quantitative if its values are ordered and possess additive structure. Values of a variable are ordered if and only if:

1. If $X \geq Y$ and $Y \geq Z$ then $X \geq Z$ (transitivity);
2. If $X \geq Y$ and $Y \geq X$ then $X = Y$ (antisymmetry);
3. Either $X \geq Y$ or $Y \geq X$ (strong connexity).

If these three conditions are satisfied then it can be said that $\geq$ is a simple order. In the applications of conjoint measurement it is assumed that the variable being tested is ordinal.

In order to establish whether a variable also possessed additivity it is necessary to apply empirical tests. The present article is an example of such a test. Conditions for additivity are as follows:

4. $X + (Y + Z) = (X + Y) + Z$ (associativity);
5. $X + Y = Y + X$ (commutativity);
6. $X \geq Y$ if $X + Z \geq Y + Z$ (monotonicity);
7. If $X \geq Y$ then there exists a value $Z$ such that $X = Y + Z$ (solvability);
8. $X + Y \geq X$ (positivity);
9. There exists a natural number n such that $nX \geq Y$ [where $1X = X$ and $(n + 1)X = nX + X$] (Archimedean condition).

In this notation, $X + Y = Z$ does not mean that $X$ added to $Y$ equal $Z$, but rather that this is a relation of $Z$ being entirely composed of discrete parts $X$ and $Y$ (cf. Michell, 1990, pp. 52–53).

## NOTES

1. From the statistical point of view, of course, correlations and variances are "metric" or quantitative. The meaning of "qualitative" in the present context has substantive psychological rather than statistical interpretation.
2. The meaning of "interaction" here is different from the meaning attached to it in the literature on experimental design. We elaborate on this distinction in a later part of this paper.
3. Since it can be assumed that personality traits mediate motivational effects, we collected data on self-concept (Marsh 1987) and Eysenck's Personality Inventory (EPI). This would have allowed us to partial-out personality traits from the ANOVA model and perhaps reduce the undesirable interaction effects. As our experimental manipulation in this study proved effective with two measures and effective with some reservation for a third, there is no need to explore the role of these moderating variables here. In order to save space, we do not present any information regarding the EPI or self-concept in this article.
4. Hunt (1980) and Stankov (1983) have suggested that intelligence and attentional

resources (i.e. processing capacity) are related constructs. The theoretical argument of the present article is easily derived from this link. We avoid use of the term "attentional resources" here since several studies in our laboratory (Myors et al. 1989; Roberts, Beh & Stankov 1988; Stankov 1987, 1989; Sullivan & Stankov 1990) have shown that cognitive tasks used by cognitive psychologists for the purpose of demonstrating the usefulness of the construct of processing capacity, do not always correlate with intelligence in the expected way. Also, there are cognitive tasks that violate the assumptions of attentional resources theory but nevertheless show correlations with intelligence.

5. There are 15 t-tests in Tables 2, 3, and 4. By keeping the Type I error rate (i.e. alpha error rate) at .01 for the individual t-tests, the experiment-wise error rate is .15. In order to have the experiment-wise error rate of .01, the individual t-test's significance level should set at .001—corresponding to the criterion t-test of 3.37. We can easily see that this Bonferroni correction would lead to a single change in our decision regarding the significance of the t-tests in Tables 3 and 4, and to four changes in our decisions regarding Table 2.

6. All calculations in this paragraph are based on the assumption that there are no ties. If ties are to be allowed, the prior probability value would be lower.

7. In evaluating the trends in correlations, it is necessary to take into account the reliabilities of the variables involved—in this case, the Letter Series test under different treatment conditions. We do not have estimates of reliabilities under all nine conditions of the W by M cross. However, the three levels of WMP used here had satisfactory reliabilities in the Myors et al. (1988) study. For the number correct score of the Letter Series of this study, estimates of Cronbach's coefficient alpha are: a. for WMP1 = .64; for WMP2 = .72; and c. for WMP3 = .70.

8. We should note that Myors et al. (1989) report similar increases in correlation up to the most difficult (i.e. WMP3 or W3) condition and a drop in correlation at that level. This was interpreted by Stankov (1989) as an indication that under the most difficult conditions a breakdown in performance takes place and the trend towards a systematic increase in correlation disappears as a consequence. The present data indicate that this breakdown does not always take place at the WMP3 level.

9. It is conceivable also that these opposing trends in correlations of number correct and time scores may be responsible for the fact that the overall measure of rate and the overall measure of accuracy have about the same correlation with Raven's scores.

10. It is obvious that this is a significant chi-square value. A better fit could be obtained by postulating the existence of additional factors. Since our aim here is to compare solutions rather than to obtain a completely satisfactory fit, we report only the simplest one-factor solution.

## REFERENCES

Butterfield, E.C., D. Nielsen, K.L. Tangen, & M.B. Richardson. (1985). "Theoretically based psychometric measures of inductive reasoning." Pp. 77–147 in *Test Design: Developments in psychology and psychometrics,* edited by S.E. Embertson. Orlando: Academic Press.

Carroll, J.B. (1980). Individual difference relations in psychometric and experimental cognitive tasks. Final Report, April 1980. Personnel and Training Research Programs, Psychological Sciences Division, Office of Naval Research, Arlington, Virginia.

Cattell, R.B. (1971). *Abilities: Their structure, growth, and action.* Boston: Houghton-Mifflin.

Crawford, J.D. (1991). "The relationship between tests of sustained attention and fluid intelligence." *Personality and Individual Differences, 12.*

Eysenck, M. (1982). *Attention and arousal.* Berlin: Springer Verlag.

Holzman, T.G., J.W. Pellegrino, & R. Glaser. (1983). "Cognitive variables in series completion." *Journal of Educational Psychology, 75,* 603–618.

Horn, J.L. (1988). "Thinking about human abilities." Pp 645–685 in *Handbook of multivariate psychology,* edited by J.R. Nesselroade. New York: Academic Press.

Humphreys, L. (1979). "The construct of general intelligence." *Intelligence, 3,* 105–120.

Hunt, E. (1980). "Intelligence as an information-processing concept." *British Journal of Psychology, 71,* 449–474.

Kotovsky, K. & H.A. Simon. (1973). "Empirical tests of a theory of human acquisition of concepts for sequential patterns." *Cognitive Psychology, 4,* 399–424.

Krantz, D.H., R.D. Luce, P. Suppes, & A. Tversky. (1971). *Foundations of measurement* (Vol. 1). New York: Academic Press.

Lord, F.M. (1953). "On the statistical treatment of football numbers." *American Psychologist, 8,* 750–751.

Luce, R.D., & J.W. Tukey. (1964). "Simultaneous conjoint measurement: A new type of fundamental measurement." *Journal of Mathematical Psychology, 1,* 1–27.

Marsh, H.W. (1987). "The hierarchical structure of self-concept and the application of hierarchical confirmatory factor analysis." *Journal of Educational Measurement, 24,* (1), 27–39.

McDonald, R.P. (1978). "A simple comprehensive model for the analysis of covariance structures." *British Journal of Mathematical and Statistical Psychology, 31,* 161–183.

Michell, J. (1986). "Measurement scales and statistics: A clash of paradigms." *Psychological Bulletin, 100,* 398–407.

———. (1988). "Some problems in testing the double cancellation condition in conjoint measurement." *Journal of Mathematical Psychology, 32,* 466–473.

———. (1990). *An introduction to the logic of psychological measurement.* Hillsdale, NJ: Lawrence Erlbaum.

Myors, B., L. Stankov, & G.W. Oliphant. (1989). "Competing tasks, working memory, and intelligence." *Australian Journal of Psychology, 41,* (1), 1–16.

Norman, D.A. & D.G. Bobrow. (1985). "On data-limited and resource-limited processes." *Cognitive Psychology, 7,* 44–64.

Roberts, R.D., H.C. Beh, & L. Stankov. (1988). "Hick's Law, competing tasks, and intelligence." *Intelligence, 12,* (2), 111–131.

Simon, H.A. & K. Kotovsky. (1963). "Human acquisition of concepts for sequential patterns." *Psychological Review, 70,* 534–546.

Spilsbury, G., L. Stankov, & R. Roberts. (1990). "The effect of a test's difficulty on its correlation with intelligence." *Personality and Individual Differences, 11,* (10), 1069–1077.

Stankov, L. (1983). "The role of competition in human abilities revealed through auditory tests." *Multivariate Behavioral Research Monographs,* 83-1.

———. (1987). "Competing tasks and attentional resources: Exploring the limits of primary-secondary paradigm." *Australian Journal of Psychology, 39,* (2), 123–137.

———. (1988). Complexity and intelligence. Paper presented at the XXIV International Congress of Psychology, Sydney, Australia.

———. (1989). "Attentional resources and intelligence: A disappearing link." *Personality and Individual Differences, 10,* 957–968.

Stankov, L. & J.D. Crawford. (1993). "Ingredients of complexity in fluid intelligence." *Learning and Individual Differences, 5,* (2), 73–111.

Stankov, L. & S. Dunn. (in press). "Physical substrate of mental energy: The number of neurons and efficient cerebral metabolic processes." *Learning and Individual Differences.*

Stankov, L. & B. Myors. (1990). "The relationship between working memory and intelligence: Regression and COSAN analysis." *Personality and Individual Differences, 11,* (10), 1059–1068.

Stevens, S.S. (1946). "On the theory of scales of measurement." *Science, 103,* 677–680.

———. (1951). "Mathematics, measurement and psychophysics." In *Handbook of experimental psychology,* edited by S.S. Stevens. New York: Wiley.

Sullivan, L. & L. Stankov. (1990). "Shadowing and target detection as a function of age: Implications for attentional resources theory and intelligence." *Australian Journal of Psychology, 42,* (2), 173–185.

Thurstone, L.L. (1938). "Primary mental abilities." *Psychometric Monographs, 1.*

Wittenborn, T.R. (1943). "Factorial equations for tests of attention." *Psychometrika, 8,* 19–35.