Challenging the “Law of Diminishing Returns”

Gerard J. Fogarty
University of Southern Queensland

and

Lazar Stankov
University of Sydney

Full reference: Fogarty,G., & Stankov, L. (1995). Challenging the Law of Diminishing Returns. *Intelligence*, 21 (2), 157-176.

Abstract

"The Law of Diminishing Returns" (Spearman, 1927) states that the size of the average correlation between cognitive tasks tends to be relatively small in high ability groups and relatively high in low ability groups. Studies supporting this finding have tended to contrast very low ability subjects (IQ < 78) with subjects from higher ability ranges and to use tests that have poor discriminatory power among the higher ability levels. In the first study described in this paper, tasks that provide good discrimination among the higher ability levels were used. A sample of High ability (N = 25) and of Low ability (N = 20) 15-years old boys completed four single tests, two with low and two with high *g* saturations, and two competing tasks formed from these single tests. The results indicated that, contrary to the predictions of the Law of Diminishing Returns, the amount of common variance was greater in the High ability group. It is suggested that the Law of Diminishing Returns does not take into account the factor of task difficulty and that there are situations where the exact reverse of this law holds. A second study again compared correlations obtained with extreme groups (N=28 & N=29), this time on measures of Perceptual Speed, which are easy for all ability levels. Results indicated that correlations among the Perceptual Speed measures were the same for both groups. In neither of these studies was there any support for the Law, which seems to be dependent on the very high correlations obtained from samples at the extreme lower end of the ability continuum.

Challenging the “Law of Diminishing Returns”

One of the very few "laws" in psychology has been identified in studies of individual differences in cognitive abilities. "The First Law of Intelligence" (see Guttman, 1992) has been known for a long time as the observation of 'positive manifold' - i.e., the tendency for cognitive tests to have positive correlations among themselves. As pointed out by Stankov (1983, p. 484), only half of this "law" is of real psychological interest. At the lower levels of intelligence there is a biological limit (i.e., non-living matter), so positive correlation is a necessity. Put simply, this means that if a person’s brain is hardly functioning, then all abilities are affected. Because there is no comparable limit at high-ability levels, the observation of positive manifold among high ability samples is of greater psychological interest.

Spearman (1927) also noted that if we divide subjects into groups with respect to their performance on a measure of general cognitive ability, the size of the average correlation tends to be relatively small in high ability groups and relatively high in low ability groups. This finding was a consequence of another law - "The Law of Diminishing Returns"[1]. Detterman & Daniel (1989) have rediscovered this "law" (cf Deary & Pagliari, 1991; Detterman, 1991). Neither of these authors, however, expressed surprise at the presence of positive manifold within the whole range of ability, including the above-average levels. Instead, they focused on the implications of this finding for the nature of the general factor. For Spearman, this "law" indicates that the more *g* (or mental energy) "... a person has available already, the less advantage accrues to his ability from further increments in it."(1927, p. 219). Detterman claims that Spearman's interpretation implies that *g* represents stupidity, not intelligence. According to him, "Each person can be thought of as having a set of independent abilities related to each other by a set of weights specifying each ability's relationship to other abilities in the performance of a particular task or test; *g* arises from this set of weights in combination with a person's independent abilities" (Detterman, 1991, p. 254).

While The First Law of Intelligence is widely accepted, the number of reported studies supporting The Law of Diminishing Returns is much smaller and therefore somewhat less convincing. It appears to have been forgotten until recently, although there have been debates over the years on issues not unrelated to the phenomena described by the Law. For example, there were suggestions by Hoffman (1962, ch.9) that correlations between IQ and academic achievement might be quite invalid for certain types of high ability student. Hoffman directed his criticism at the nature of multiple choice items that were subject to unintended interpretations at the upper end of the ability spectrum, but he might also have introduced the Law to support his argument of diminished correlations among cognitive variables in high ability groups. Similarly, the long-running debate on test bias looked at the question of differential validity for various groups in the population (see Cole & Moss, 1993). Initially, the groups studied were Blacks and Whites but many other groups have also been proposed and studied. The resulting literature shows that in most cases there are no differences between groups in terms of regression slopes for predicting job criteria and in terms of the factorial structure of test batteries within the groups (e.g., Jensen, 1980; Cunningham & Birren, 1980). With the group studies, the focus remained on structural concerns, especially the notion of “invariance”. The work of Linn (1982), who divided the selection bias literature into studies of differences in the predictor-criterion correlation and studies of differences in prediction systems, came closest to discussing issues central to the Law of Diminishing Returns when commenting on varying correlations between predictor and criterion variables for different groups.

Now that the Law has again been brought to the attention of the research community there are, in fact, good reasons to question its validity. One set of reasons derives from the fact that proponents of this law have relied upon somewhat obsolete methodology that focuses on correlation coefficients. The usual practice has involved the division of samples into ability groups on the basis of scores on one cognitive test variable and the comparison of correlation coefficients representing relations among other cognitive variables for each of the groups (e.g., Spearman, 1927). A similar logic is used in studies of "invariance". The literature on measurement invariance is vast (see, for example, Horn & McArdle, 1992; Meredith, 1964; Millsap, 1994). Among other things, the literature on this topic points out that invariance may be present in many parameters of structural equation models, not just in the correlations. Computer programs such as LISREL now provide a means of testing whether trends in correlations may be due to particular features of the structure and allow for the evaluation of these effects on the correlations. It is possible that observed trends in correlations may be the result of some particular statistical features of the data that have nothing in common with the substantive effects of the "Law of Diminishing Returns". In other words, statistical artefacts have not been eliminated as plausible accounts of trends in correlations.

In addition, there would appear to be a weakness in the logic used to support the case for decreasing correlations with high ability groups. We suggest that the metaphor used to explain this law is somewhat limited and that an extension of the metaphor will highlight this deficiency. Thus, in terms of the original metaphor, a weak person will struggle to lift a moderately heavy weight and will benefit considerably from an increase in energy. A strong person will not show much benefit from a similar addition. We do not challenge this logic, nor do we challenge the supporting empirical data gathered from samples of low IQ and average subjects working on standard psychometric tasks. If we extend the metaphor, however, the picture changes. Consider the situation where the weights to be lifted are heavy. A weak person cannot lift the weights at all and does not benefit from a boost of additional energy. A strong person, however, will go close to lifting these weights under normal conditions and will probably benefit considerably from the same energy boost. The Law of Diminishing Returns will not apply under these conditions. In other words, typical intelligence tests are constructed for the "normal" range of abilities and their discriminatory power at the higher ranges may be severely limited because of ceiling effects. If more difficult tests were chosen, these effects would disappear and the lack of discrimination would apply to the low ability subjects. In practical terms, it means that we should be able to find tasks where positive manifold is greater among high ability subjects. One of the aims of the first study was to test this assumption using tasks that are known to impose a heavy cognitive load and which should therefore prove more suited to high ability groups. Competing tasks fall into this category.

Study 1: Competing Tasks And The Measurement Of Intelligence in High and Low Ability Groups

The term "competing task" is used to describe situations wherein two distinct tasks are competing for attentional resources. These tasks may come through the same or different sensory channels, the stimulus inputs may be simultaneous or interspersed; as long as the individual is forced to do more than one thing at a time, they are what we call competing tasks. Fogarty and Stankov (1982) and Stankov (1983) observed that tasks made a greater demand on general intelligence when measured under conditions where there was competition for limited resources. This meant that one could expect components of these tasks to have higher correlations with measures of general intelligence than their single test counterparts. These same authors later pointed out that despite the greater demand on general intelligence, psychometric indices might not always reflect this demand (Fogarty & Stankov, 1988). The crucial factor is the demand made on general intelligence by the single tests used to form the competing tasks. The metaphor of the human information processing system as a resource limited system (Norman & Bobrow, 1975) can be used to explain the dynamics of the situation. All competing tasks result in greater demand on central resources but if the single tasks themselves place a heavy load on central resources, as indicated by their *g* loadings, then the system itself has no way of coping with competition between two such tasks. The result may be a partial or complete breakdown in performance.

From a psychometric viewpoint, competing tasks can be seen as just another example of a complex task. In fact, they may have some practical utility because of this complexity. Stankov (1993) pointed out that typical tests of intelligence in use today have poor discriminatory power in the above-average range. In terms of the metaphor already employed, using these tests to rank high ability subjects is akin to using a set of moderate weights to determine the rankings in a weight-lifting competition. The true ranking is unlikely to emerge. Stankov suggested that the use of competing tasks with such samples may lead to better discrimination among high ability individuals. One implication of this suggestion is that positive manifold should increase for the high ability group if the tasks used are towards the upper limit of their potential. Under these same conditions, low ability people will fail to cope and positive manifold may not be observed. Such a situation could produce evidence contradicting the Law of Diminishing Returns.

In the present study, a group of High ability and a group of Low ability students were selected from a much larger sample and asked to perform four single tests and two competing tasks formed by the coupling of these single tests. The first of these competing tasks combined tests of perceptual processing, the second combined measures of fluid intelligence. The latter task was considerably more complex than the former. According to the Law of Diminishing Returns, the amount of common variance in the battery of tests (i.e., correlations among the tests) should be higher in the Low ability group than in the High ability group. This law has been reported with single tests of abilities, not competing tasks. On the basis of previous work with competing tasks, they were expected to impose processing demands that were likely to exceed the capacity among members of the Low ability group. With the High ability group, on the other hand, task demands were expected to be more in line with available capacity, especially for the easier of the competing tasks. Under these conditions, stronger evidence of positive manifold was expected in the data produced by the High ability group. This is quite at odds with what one would predict on the basis of the Law of Diminishing Returns.

Apart from testing the Law of Diminishing Returns, the conditions of the present study also enabled a test of claims arising from earlier studies (Fogarty & Stankov, 1982, 1988; Stankov, 1993) where it was observed that the phenomenon of increased *g* saturations did not apply when the two single tasks already imposed heavy demands on *g*, presumably because the demands exceeded available capacity. Under these conditions, *g* saturations actually decline when the tasks are combined. If this decline is indeed associated with the lack of available cognitive resources, then the decline should be much less marked in a group of high ability subjects. The use of two different ability groups and two different competing tasks - one formed from two perceptual tests (low *g* saturation), the other formed from two fluid intelligence markers (high *g* saturation) - provided an opportunity to explore further the hypothesis that competing tasks increase the demands placed on general intelligence. It was expected that for the Low ability group, *g* saturations would increase for the variables used in the “perceptual” competing task but decrease for variables in the “fluid intelligence” competing task. In the High ability group, *g* saturations were expected to increase in both tasks.

Method

Subjects

Subjects were 45 Year 10 boys drawn from a larger pool of 141 boys at a private school in Toowoomba, Australia, who agreed to take part in the study in return for the chance to win a raffle prize. All 141 boys had completed the Verbal Reasoning, Abstract Reasoning, Numerical Reasoning, Spelling, and Language Usage forms of the Differential Aptitude Test (Bennett, Seashore, & Wesman, 1989) some months earlier as part of a school assessment programme. Twenty five of the boys had combined Verbal Reasoning (VR) and Numerical Ability (NA) Differential Aptitude Test (DAT) results that placed them in the top three stanines (7,8,9) for their age group in Australia, twenty had scores that placed them in the bottom three (1,2,3) stanines. Combined VR and NA scores can be interpreted as an index of general cognitive capacity (Bennett et al., 1989), so these two groups represented populations that differed quite markedly in terms of scores on general intelligence. According to the information provided by the Australian Council for Educational Research (ACER), DAT stanine scores 1, 2, and 3 correspond to IQ levels that are below 88 points and stanine scores 7, 8, and 9 correspond to IQ levels that are above 112. The average age of the boys was 15 years.

Tests

Four single tests used in the Fogarty and Stankov (1988) study were chosen for inclusion in the present study. All tests were presented on MacIntosh Classic computers.

1. Hidden Figures. This test required the subject to study a complex figure and to indicate whether or not two figures which appeared below it were part of the larger figure. The legitimate answers were ‘0’, ‘1’, ‘2’, and ‘3’; corresponding to ‘neither’, ‘first figure only’, ‘second figure only’, or ‘both figures present’. (Source: Fogarty & Stankov, 1988 - adapted from Ekstrom, French, Harman, & Derman, 1976).

2. Tonal Memory. This test required the subject to indicate the position of a tone which changed pitch during the repetition of a tonal sequence. Tonal sequences varied from three tones to five tones in length. (Source: Fogarty & Stankov, 1988 - adapted from Seashore, Lewis, & Saetveit, 1960).

3. Sets. In this test, two sets of three letters were presented, the second set being a repetition of the first (following a 500 millisecond pause) with one letter replaced. Subjects were required to name both the letter that was missing in the second set and the letter that replaced it. (Source: Fogarty & Stankov, 1988 - adapted from Crawford & Stankov, 1983).

4. Tonal Reordering. This test required subjects to listen to a set of three tones which was then repeated, but usually in a different order. The task was to note the order of presentation on the second playing. (Source: Stankov & Horn, 1980).

All four single tests contained 20 items. The first two tests tapped lower order perceptual abilities in the Gf/Gc hierarchy: General Visualization (Gv) and General Auditory (Ga) function respectively. The second two tapped the higher order ability of Fluid Intelligence (Gf) (Fogarty & Stankov, 1988). Thus, these last two tests were more complex than the first two.

Tests one and two were combined to form a competing task of 30 items. Both components were presented simultaneously with subjects instructed to pay equal attention to both. Following the presentation of the items, subjects were asked to answer first one test, then the other, with a random post-cueing technique used so that over the 30 items each task received an equal number of first-prompts. Tests three and four formed a second competing task which was presented in an identical fashion. To form the competing tasks, only the easier items from the single tests were used. This is in accordance with our previous practice (Fogarty & Stankov, 1982, 1988). These tests yielded a total of eight measures: four single test scores and another set of four scores taken from these same tests administered under competing task conditions. A composite measure of general intelligence was added to the data set by combining DAT Verbal Reasoning, Numerical Ability, Abstract Reasoning, and Language Usage scores.

Results

Although the tests were modelled on previously-tried measures, some adaptations were made for the present study. Consequently, the first step in data analysis involved a check of reliabilities using Cronbach’s Alpha and the deletion of items which had a negative correlation with total test scores. Summary statistics for the adjusted scales are reported in Table 1.

Table 1

Descriptive Statistics for All Variables (N = 45)

Variable	Items	Mean	S.D.	Prop. Corr.	Alpha Reliab.	Corel. w/DAT[a]
1. Hidden Figures Single (hfs)	17	8.36	3.21	.49	.73	.58
2. Hidden Figures Competing (hfc)	27	16.98	6.02	.63	.87	.73
3. Tonal Memory Single (tms)	17	11.61	2.56	.68	.63	.43
4. Tonal Memory Competing (tmc)	29	19.13	5.75	.66	.85	.53
5. Tonal Reordering Single (trs)	20	10.25	5.84	.51	.90	.73
6. Tonal Reordering Competing (trc)	28	7.95	5.63	.28	.87	.61
7. Sets Single (sets)	20	15.59	3.76	.78	.80	.65
8. Sets Competing (setc)	30	14.30	6.23	.48	.85	.47

[a]All correlations higher than .372 are significant at the .01 level.

It can be seen that reliabilities ranged from moderate to high and there was reasonable variance on all measures. The fourth column in the table shows the mean proportion correct to facilitate comparisons across tasks. Proportion correct is used in the remainder of this study. As expected, decrement on components of the first of the competing tasks did not appear. This finding is in agreement with Wickens' (1980) multiple resources theory which assumes the existence of separate pools of resources corresponding to visual and auditory modalities (see Fogarty & Stankov, 1988b; Stankov, 1988; 1989). However, there is a pronounced decrement in scores for the components of the second competing task, even though it used the easier items from the component single tests.

In previous work, changes in correlation between the components of the competing tasks has been taken as one index of a change in *g* saturations. An increase in correlation has been attributed to greater reliance on higher order processes necessitated by the complexity of the situation in which both tasks must now be solved. In the present study, the combined DAT score represented a further benchmark against which to judge the *g* saturations of the single and competing task measures. The last column in Table 1 contains correlations between test scores and DAT. These correlations, of course, are higher than the correlations from the whole population since they reflect the fact that two extreme groups, selected on the basis of DAT scores, were involved. The pattern of correlations of the various tasks with the overall DAT measure supports the claim that the changes in correlation reflect changes in *g* saturations. The components of the perceptual competing task had higher correlations with the DAT measure than did their single test counterparts. The reverse was true for the fluid intelligence competing task. As mentioned earlier, this finding is not at odds with what is known about the relationship between task complexity and *g*.

If complexity is indeed the critical factor, then one might expect to see some continuation of the trend for components of competing tasks to show higher *g* saturations if the analysis is restricted to the group of students with high scores on the DAT Verbal Reasoning and Numerical Ability subscales. To explore this possibility, both groups were analysed separately. The descriptive statistics (including re-calculated reliability estimates) are shown in Table 2.

Table 2
Descriptive Statistics for High and Low Ability Groups[a]

Vars.	High Ability Group (N = 25)			Low Ability Group (N = 20)			
	M	S.D.	Rel.	M	S.D.	Rel.	Diff.
hfs	.58	.17	.66	.43	.20	.72	.14*
hfc	.70	.16	.82	.44	.20	.76	.26**
tms	.70	.14	.65	.60	.21	.66	.10*
tmc	.72	.19	.85	.54	.21	.82	.18**
trs	.68	.24	.86	.32	.23	.82	.36**
trc	.36	.22	.87	.19	.18	.80	.17**
sets	.88	.13	.68	.64	.20	.74	.24**
setc	.58	.21	.84	.37	.19	.83	.21**

[a]Rel. = Alpha coefficient of reliability; Diff. = Difference between proportion correct in High and Low ability groups. Asterisks indicate the conventional levels of significance of the appropriate t-test: * $p < .05$; ** $p < .01$.

From this, it can be seen that the High ability group performed better on all tests. Furthermore most means of the High ability group were above the .50 level and most means of the Low ability group were below the .50 level. The difference in the proportion of correct answers was significant for all variables with standard deviations and reliabilities comparable in the two groups. This means that any differences in correlations between the groups cannot be attributed to differential restrictions in range or to reliabilities of the tests. The correlations among the tasks for the two groups are shown in Table 3[2].

Table 3
Correlations Among All Variables for High (Lower Triangle; N=25) and Low Ability (Upper Triangle; N=20) Groups

Vars.	DAT	hfs	hfc	tms	tmc	trs	trc	sets	setc
DAT		.09	.09	.10	-.04	.26	.11	.02	-.06
hfs	.05		.53	.10	.27	-.12	.09	-.19	-.37
hfc	.36	.62		.16	.04	-.17	.07	-.08	-.18
tms	.17	.19	.21		.67	.12	.15	.19	-.13
tmc	.36	.25	.19_	.64		.23	.35	.14	-.26
trs	.48	-.05	.04	.67	.59		.52	.23	-.01
trc	.59	.02	.08	.33	.63	.53		.30	.08
sets	.33	.11	.33	.04	-.13	.20	.04		.39
setc	.21	-.03	.43	-.22	-.22	-.05	-.13	.32	

[a]All correlations higher than .381 (lower triangle) and .423 (upper triangle) are significant at the .05 level.

Even a casual inspection of the matrices in Table 3 shows that the overall level of correlations displayed by the High ability group (lower triangle) is higher than the correlations within the Low ability group (upper triangle). This finding suggests that there is no support for the Law of Diminishing Returns in these data. In order to make this clearer, correlations were calculated among composite scores. Composite scores were used because they yielded an overall measure of performance in the competing task condition. Thus, the composite score for competing task one was formed by adding scores on the two components. Similar composites were formed by adding totals for the corresponding single tests. Four measures were obtained in this way. They were: a) Single 1 (perceptual: hfs + hfc); Single 2 (fluid intelligence: trs + sets); Comp 1 (perceptual: hfc + tmc); and Comp 2 (fluid intelligence: trc + setc). The descriptive statistics and correlations are presented in Table 4. Separate parts of the table are used to display single test composites and the competing task composites.

Table 4

a. Means, Standard Deviations and Correlations Between Combined Single Tasks and DAT Scores in High and Low Ability Groups

	Task	Mean	S.D.	Correlations		
				Sing. 1	Sing. 2	DAT
High	Single 1	.64	.13	1.00	.34	.14
(N=25)	Single 2	.78	.15		1.00	.53
	DAT	151.72	13.32			1.00
Low	Single 1	.40	.15	1.00	-.00	.19
(N=20)	Single 2	.51	.16		1.00	.16
	DAT	81.70	11.51			1.00

LISREL 7.13 (Jöreskog & Sörbom, 1988) test of the difference between the two correlational matrices produced a Chi-square value of 4.38 with df = 6 and p=.626.

b. Means, Standard Deviations and Correlations Between Combined Competing Tasks and DAT Scores in High and Low Ability Groups

	Task	Mean	S.D.	Correlations		
				Comp1	Comp2	DAT
High	Comp. 1	.71	.13	1.00	.44	.46
(N=25)	Comp. 2	.47	.14		1.00	.63
	DAT	151.72	13.32			1.00
Low	Comp1	.43	.13	1.00	-.02	.03
(N=20)	Comp2	.28	.13		1.00	.03
	DAT	81.70	11.51			1.00

LISREL 7.13 (Jöreskog & Sörbom, 1988) test of the difference between the two correlational matrices produced a Chi-square value of 9.28 with df = 6 and p=.159.

Even though the correlational data form the bases of our discussion of the Law of Diminishing Returns, it is important to examine findings relating to group levels of performance (i.e., arithmetic means). The first point to note is that variablity within each condition is approximately equal, indicating that even the condition with the lowest mean - Comp 2 in the Low ability group - does not show any evidence of the effect of restriction in range that may be expected as a consequence of floor effects. Two previously noted features of the arithmetic means are also apparent in Table 4. Firstly, on the average, the Low ability group scored less than 50% on the composite tasks whilst the High ability group scored above 50% correct. Secondly, although Single 2 (a combination of fluid intelligence tests) proved to be easier than Single 1 (a combination of perceptual tests), competing versions of the two tasks displayed the opposite trend.

Evidence Relevant to the Law of Diminishing Returns

The correlations in both sections of this Table provided information that was used to test the generality of the Law of Diminishing Returns. Two procedures were used to test the differences between the amount of common variance within High and Low ability groups. Firstly, a z-test statistic of 1.30 was obtained for the difference between single composites for both groups (r = .35 versus r = -.00, Table 4a). For the corresponding difference between Comp 1 and Comp 2 composites (r = .44 and r = -.02, Table 4b), the z value was 1.70. Using a one-directional test, the difference in correlation coefficients between the competing task composites is significant at the .05 levels. This test therefore shows that the results contradict the Law of Diminishing Returns with competing tasks. Secondly, structural equation modelling procedures embodied in LISREL (Jöreskog & Sörbom, 1989) were employed to test the difference between correlational matrices for single tests (Table 4a) and competing tasks (Table 4b). For the single tests, the Chi-square was 4.38 which, with df = 6, produced a probability value of .626. For the competing tasks, the Chi-square value was 9.28 which, with df = 6, produced a probability value of .159. The difference between the groups in the amount of common variance in competing tasks was not significant with this test whereas it was with the z-test. Notice, however, that the Chi-square test was not a one-directional test and therefore the criterion probability value should be increased from the typical .05 level to something higher. How much higher we do not know.[3] Nevertheless, it is obvious that the difference between groups in the amount of common variance for the competing tasks comes close to being significant, as it was with the z-test. It is important to note here that the sign of the difference is the opposite to that predicted by the Law of Diminishing Returns. The overall results fail to support the Law and, in fact, go some way towards actually contradicting it.

Discussion

The main aim of this first study was to gather data which would test the generality of the Law of Diminishing Returns and to provide a balance to earlier studies that relied mostly on tests that did not discriminate well among the more intelligent groups used in those studies. The results indicate that the Law does not hold under the conditions of the present study. The only feature that is not a replication of our previous work with the same tests is a somewhat lower age of the subjects employed here. The subjects of this study were 15 year-old-boys while most of our work in the past was with adults. It is hard to see this age difference as the cause of lack of support. In fact, a parallel to the Law of Diminishing Returns exists within the developmental literature in the form of the "differentiation hypothesis" (Anderson, 1992; Garrett, 1946). According to this hypothesis, at the younger age levels only a general factor can be measured and, as a person grows older, primary abilities branch-out of it. If anything, this hypothesis would suggest that younger subjects would produce more favorable evidence for the "Law of Diminishing Returns" than older subjects.

Evidence That Competing Tasks Capture More Variance in the High Ability Group

The other feature of the first study that deserves comment is the use of competing tasks to follow a suggestion that these tasks may be good, perhaps better than single tests, measures of intelligence among high ability groups. A formal test of this hypothesis is not provided. Instead, it is pointed out that the highest correlations were obtained with the competing tasks within the High ability group (i.e., top part of Table 4b). The next highest correlations were displayed by the composites for single tests in the High ability group (i.e., correlations in the top part of Table 4a). On the other hand, correlations within the Low ability group are all virtually zero. The first latent root is sometimes taken as an indicant of the amount of common variance within a matrix. For the four matrices in Table 4, the first latent roots were as follows:

	1st Latent Root	% of Total Var.
Single - High ability group	1.723	58%
Single - Low ability group	1.176	39%
Competing - High ability group	2.011	67%
Competing - Low ability group	1.114	37%

The results provide support for the hypothesis that common variance among the competing tasks given to the High ability group was larger than common variance among the single tests given to the same group. It can be said that these tasks provide a challenge to the more able subjects; they motivate such subjects to mobilize processing resources that are at their disposal to solve items in complex intelligence tests. But difficulty may be a wrong word to use here. While the subjects may use the term "difficulty", the critical feature may be properly labeled as "complexity". Thus, the two perceptual tests when given concurrently did not become more difficult (the means remained much the same under the competing condition) but appeared more difficult to the subjects. In reality, they became more complex and that is why they tended to correlate more highly with the DAT measure of general intelligence (c.f., Stankov & Crawford, 1993; Stankov & Cregan, 1993). However, if the demands imposed by the competing tasks are excessively complex, increased *g* saturations will not be observed. This is similar to the situation wherein the *g* saturation of a task increases as additional bits of information are needed for task solution - e.g., tasks designed in accordance with Hick's law. Beyond a certain point, however, increasing the bits of information does not lead to further increases in *g* loadings; rather, decreases are observed (Jensen, 1982; Roberts, Beh & Stankov, 1988). This would suggest that tests which do not have high *g* saturations are good candidates for coupling together as competing tasks if one's aim is to construct better measures of intelligence. The two measures of fluid intelligence, on the other hand, had a lower correlation with the measure of general intelligence when they were combined into a competing task. This would suggest that such tests are not good candidates for competing presentation. The present study has confirmed a trend that has already been noted in earlier studies (cf., Fogarty & Stankov, 1988).

Study 2: Testing the Law of Diminishing Returns With Perceptual Speed Tests in High and Low Ability Groups

The first study showed that it is possible to find situations where the reverse of the Law holds and, in this sense, provided a needed contrast to the other studies reported in the literature. A different test of the Law, however, involves the analysis of correlations among tasks that are of equal difficulty level across the ability continuum. If the differences in correlation are related to the suitability of tests for high and low ability populations then there should be no differences at all when tests of uniform difficulty are administered. Easy tests that are based on time-limited performance, such as Perceptual Speed tests, fall into this group. The spread of total scores on these tests is achieved through the imposition of severe time limits on test performance. These tests are important here because the items of these tests tend to show about the same discriminatory power at both high and low ability levels. The aim of the second study in this series was to examine the evidence for the Law of Diminishing Returns under these conditions. The data to be reported here were collected by R. Roberts as part of his Ph. D. research project (Roberts, 1995).

Method

Subjects

A total of 179 (110 females) subjects was employed in this study. The majority of these (146) were first year Psychology students at the University of Sydney, Australia - the remainder were recruited from the general population. The age of the subjects ranged from 17 to 50 years with a mean of 21.58 years and standard deviation of 6.18 years.

Tests

In this study, a battery of twenty one psychometric tests was administered. The tests measured fluid intelligence (Gf, eight tests), crystallized intelligence (Gc, four tests), short-term acquisition and retrieval function (SAR, two tests), broad visualization (Gv, four tests), and broad auditory function (Ga, three tests). A composite score combining all these psychometric tests - i.e., a measure of general ability - was calculated and used to divide the subject sample into High ability and Low ability groups. The High ability group consisted of all people whose total score was more than one standard deviation above the mean; the Low ability group were all those subjects whose total score was lower than one standard deviation below the mean. Because one of the tests of Gf was the Ravens Standard Progressive Matrices test, it was possible to obtain information about the positions of the two groups relative to the standardization sample. This test has a maximum score of 60, a mean of 43.34 and standard deviation of 6.64. The arithmetic mean of the High ability group in the second study was 56.83 whilst for the Low ability group it was 40.67. The IQ equivalents of these scores for the Low group ranged between 92 and 103. Those of the high group were all above 125.

In addition to the above twenty psychometric tests, subjects were given the four Perceptual Speed tests from the Stankov (1988) study.

1. Number Comparison. There were 48 items in this test and the task was to press the "Yes" key if two numbers displayed side by side in the center of the screen were exactly the same, and to press the "No" key when different. Score: time (in milliseconds) needed to provide answers to all items.
2. Search. The three parts of this test each consisted of 30 items and the order of presentation of the parts was randomized across subjects according to a Latin Square design. The task was to press the "Yes" key if the string of letters displayed across the center of the screen contained (i.) A or B or C (part 1); (ii.) A and B and C (part 2); and (iii.) A and either B or C (Part 3). In this test, the logical 'or' is inclusive. Score: time (in milliseconds) needed to provide answers to all 90 items.
3. Stroop task. This test consisted of 80 items where two words naming a color (BLUE, ORANGE, or GREEN) were presented one above the other. The colors used to print the words were also either blue, orange or green. The task was to press the "Yes" key if the color named by the bottom word was the same as the color used to print the top word. The fact that the bottom word was printed in one of the three colors used and the top word also named one of these three colors provided distraction in this task. The Stroop task is typically considered a measure of selective attention. Stankov (1988a) and Roberts (1995) found that these tasks define a Search factor when embedded into a larger battery of cognitive measures. Score: time (in milliseconds) needed to provide answers to all 80 items.
4. Digit Symbol. This was the subtest from WAIS. Score: number of correctly paired digits and symbols.

Results

Table 5a presents descriptive statistics for the whole sample and Tables 5b and 5c contain the same information for the High and Low ability groups respectively. The procedure used for the division into ability groups is described in the Tests section above.

Table 5
Means, Standard Deviations and Correlations Between Perceptual Speed Tests

a. Whole Sample (N=179)	Mean	S.D.				
1. Number Comparison	3000.97	903.87	1.00	.50	.51	-.51
2. Stroop	1724.92	604.45		1.00	.52	-.38
3. Search	1165.94	333.18			1.00	-.31
4. Digit Symbol	68.24	10.45				1.00
b. High Ability (N=28)	Mean	S.D.				
1. Number Comparison	2730.03	854.84	1.00	.56	.74	-.35
2. Stroop	1645.77	562.62		1.00	.46	-.25
3. Search	1086.026	254.09			1.00	-.25
4. Digit Symbol	71.50	10.02				1.00
c. Low Ability (N=29)	Mean	S.D.				
1. Number Comparison	3074.21	995.93	1.00	.57	.50	-.53
2. Stroop	1980.41	714.99		1.00	.59	-.15
3. Search	1350.62	427.33			1.00	-.27
4. Digit Symbol	66.00	10.55				1.00

As expected, there were differences between the three sets of data: means, standard deviations, and correlations all differed across the three samples considered here. The correlational matrix in Table 5a shows correlations that are more equal in magnitude while those of Tables 5b and 5c are somewhat more varied. Negative values in all these matrices are due to the fact that Digit Symbol is scored in terms of accuracy rather than time.

The question of main interest in this work concerned the difference in an overall measure of association across the Low and High ability groups. As in Study 1, the first latent roots of the three matrices provided a useful measure of the degree of overall association. These values were 2.367 (Table 5a), 2.415 for the High ability group (Table 5b) and 2.338 for the Low ability group (Table 5c). Clearly, these values are very close to each other indicating that a rather similar amount of common variance is present in all three matrices. To examine the difference further, correlational matrices of High and Low ability groups were compared using LISREL 7.3 (Jöreskog and Sörbom, 1988). The resulting test of the equality of these matrices ($X^2 = 13.41$, df = 10, $p = .201$) indicated that correlations between the four Perceptual Speed measures were essentially the same in two groups.

Discussion.

The findings of Study 2 indicate that the Law of Diminishing Returns is not supported with measures of Perceptual Speed for subjects who are divided into High and Low ability groups on the basis of their general ability scores. This finding with tests whose items have about the same discriminatory power across all ability levels casts further doubts on the validity of the Law. The findings from Study 2 thus complement those obtained in Study 1. The basic argument presented here is that earlier studies used tests that failed to discriminate among high ability groups, with a consequent reduction in the size of the correlation coefficients for this group. Study 1 showed that by using tests that favoured the high ability group, the trend can be reversed with the reduced correlations now appearing at the lower end of the distribution. Study 2, employing tests with a more or less uniform difficulty level for different abilty groups (as measured by some external criterion), showed that under these conditions, the correlations are the same for the various groups.

General Discussion and Conclusion

In neither of the studies reported here was evidence found for the Law of Diminishing Returns. A possible reason was alluded to earlier. The available literature suggests that the most convincing evidence for the Law obtains when one compares a very low ability group (say, below an IQ of 78) with groups immediately above that range. This is certainly consistent with the findings of Detterman & Daniel (1989)[4]. The Low ability subjects in the first study, although at the lower levels with respect to their scores on DAT, did not in fact belong to the "IQ below 78" group. It is doubtful that even a single subject had an IQ below 80. This was also true of subjects in the second study. In the context of the present studies, however, there is an even more compelling reason for the failure to find evidence in support of the Law. It has to do with the fact that the Law, in its present form, predicts “diminished returns” as tasks are performed by increasingly more able people. This happens because there is no scope for them to exercise their superior capacity and, consequently, the tasks fail to provide a true ranking of those individuals who find the task within their capabilities. Stankov (1993) suggested that many current psychometric tests, including those used by Detterman and Daniel (1989), are of this nature. The Law will hold under these conditions. Discrimination will be present among lower ability subjects and correlations will be more robust in that group. The Law breaks down, however, when more difficult tasks, such as the competing tasks used in Study 1, are introduced. Now the more robust correlations are observed with the higher ability group because discrimination is occurring. The lower ability group, on the other hand, is struggling to cope with the task and it becomes increasingly difficult to obtain reliable rankings of these individuals on difficult tasks. Positive manifold can completely disappear when lower ability groups work on complex tasks, just as it can when high ability groups work on simple tasks. When tests with good discrimination across the full range are used, positive manifold is observed in all groups. The Law of Diminishing Returns, like the original Yerkes-Dobson law describing the relationship between performance and arousal level, should be modified to accommodate the factor of task complexity (see Matthews & Westerman, 1994).

In conclusion, the present data challenge the empirical basis of the “Law of Diminishing Returns” while leaving the “First Law of Intelligence” untouched. Indeed, we would go a bit further and suggest that the Law of Diminishing Returns is really just half of the First Law of Intelligence. Viewed in this light, it refers to the presence of positive manifold where the test has discriminatory power and its absence otherwise. There is no need to invoke a separate law to account for this phenomenon.

Footnotes

[1]Spearman was considerably more lax than most people today in his use of the the word "law".

[2] The only somewhat unexpected and hard to explain finding of Table 3 is a relatively large number of negative correlation coefficients involving the competing version of the Sets test. Not one of these correlations is significantly different from zero.

[3] It is possible to calculate the one-directional probability value from the Chi-square that has df = 1, but we are not aware of the procedure involving df > 1.

[4] Detterman & Daniel (1989) report that the "Average correlations for lowest to highest ability groups assembled by Information subscale scores, uncorrected for attenuation, for the WISC-R were .42, .29, .26, .21, and .22 and for the WAIS-R were .56, .37, .30, .25, and .26. Even these uncorrected differences were substantial and systematic." (p. 356)

References

Anderson, M. (1992). Intelligence and Development: A cognitive theory. Oxford: Blackwell.
Bennett, G.K., Seashore, H.G., & Wesman, A.G. (1989). Differential Aptitude Tests: Australian Manual. NSW, Australia: The Psychological Corporation.
Cole, N.S., & Moss, P. A. (1993). Bias in test use. In R. Linn (Ed.), Educational measurement (3rd edn.), Phoenix: Oryx Press.
Crawford, J., & Stankov, L. (1983). Fluid and crystallized intelligence and primacy/recency measures of short-term memory. Intelligence, 7, 227-252.
Cunningham, W., & Birren, J. (1980). Age changes in the factor structure of intellectual abilities in adulthood and old age. Educational and Psychological Measurement, 40, 271-290.
Deary, I. J., & Pagliari, C. (1991). The strength of *g* at different levels of ability: Have Detterman and Daniel rediscovered Spearman's "Law of Diminishing Returns? Intelligence, 15, 247-250.
Detterman, D. K. (1991). Reply to Deary and Pagliari: Is *g* intelligence or stupidity? Intelligence, 15, 251-255.
Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low-IQ groups. Intelligence, 13, 349-359.
Ekstrom, R. B., French, J. E., Harman, H. H., & Derman, D. (1976). Manual for kit of factor referenced cognitive tests. Princeton, N.J.: Educational Testing Service.
Fogarty, G. (1987). Timesharing in relation to broad ability domains. Intelligence, 11, 207-231.
Fogarty, G., & Stankov, L. (1982). Competing tasks as an index of intelligence. Personality and Individual Differences, 3, 407-422.
Fogarty, G., & Stankov, L. (1988). Abilities involved in performance on competing tasks. Personality and Individual Differences, 9, 1, 35-49.
Garrett, H. E. (1946). A developmental theory of intelligence. American Psychologist, 1, 372-378.
Guttman, L. (1992). The irrelevance of factor analysis for the study of group differences. Multivariate Behavioral Research, 27, 175-204.
Hoffman, B. (1962). The tyranny of testing. New York: Crowell-Collier Press.
Jensen, A.R. (1980). Bias in mental testing. New York: The Free Press.
Jensen, A.R. (1982). Psychometric *g* and reaction time. In H.J. Eysenck (Ed.), A model for intelligence. N.Y.: Springer-Verlag.
Jöreskog, K. G., & Sörbom, D. (1988). LISREL 7: A Guide to the Program and Its Applications. Mooresville, IN: Scientific Software.
Linn, R.L. (1982). Ability testing: Individual differences, prediction, and differential prediction. In A.K. Wigdor & W.R. Garner (Eds.), Ability testing: Uses, consequences, and controversies: Part II (pp. 335-388). Washington DC: National Academy Press.
Linn, R.L. (1993). Educational Measurement (3rd Edn). Phenix: Oryx Press.
Matthews, G., & Westerman, S.J. (1994). Energy and tension as predictors of controlled visual and memory search. Personality and Individual Differences, 17, 617-626.
Meredith, W. (1964). Notes on factorial invariance. Psychometrika, 29, 177-185.
Millsap, R.E. (1994). Recent work in measurement bias, factorial invariance, and predictive bias. Cattell Address given at the annual meeting of the Society of Multivariate Experimental Psychology, Princeton, NJ.
Navon, D., & Gopher, D. (1979). On the economy of the human information processing system. Psychological Review, 86(3), 214-255.

Norman, D.A., & Bobrow, D.G. (1975). On data-limited and resource-limited processes. Cognitive Psychology, 7, 44-64.
Roberts, R. D. (1995). Speed of processing within the structure of human cognitive abilities. Unpublished PhD thesis, The University of Sydney.
Roberts, R. D., Beh, H. C., & Stankov, L. (1988). Hick's law, competing tasks, and intelligence. Intelligence, 12, 2, 111-131.
Seashore, C.E., Lewis, D., & Saetveit, J.C. (1960). Manual of instruction and interpretation for the Seashore Measures of Musical Talent (2nd Revision). N.Y.: The Psychological Corporation.
Spearman, C. (1927). The abilities of man. MacMillan: New York.
Stankov, L. (1983). Attention and intelligence. Journal of Educational Psychology, 75, 4, 471-490.
Stankov, L. (1988a). Aging, intelligence and attention. Psychology and Aging, 3(2), 59-74.
Stankov, L. (1988b). Single tests, competing tasks, and their relationship to the broad factors of intelligence. Personality and Individual Differences, 9, 1, 25-33.
Stankov, L. (1989). Attentional resources and intelligence: A disappearing link. Personality and Individual Differences, 10, 9, 957-968.
Stankov, L. (1993). Theory-based cognitive tasks for the identification of intellectually gifted students. Australian Educational and Developmental Psychologist, 10, 2, 16-23.
Stankov, L., & Horn, J.L. (1980). Human abilities revealed through auditory tests. Journal of Educational Psychology, 72(1), 21-44.
Stankov, L., & Crawford, J. D. (1993). Ingredients of complexity in fluid intelligence. Learning and Individual Differences, 5, 2, 73-111.
Stankov, L., & Cregan, A. (1993). Quantitative and qualitatitve properties of an intelligence test: Series completion. Learning and Individual Differences, 5, 2, 137-169.
Wickens, C. D. (1980). The structure of attentional resources. In R. Nickerson (Ed.), Attention and Performance VIII (pp. 239 - 257). Hillsdale, NJ: Lawrence Erlbaum.

Authors' Note
This research was supported by a grant from the Faculty of Sciences at the University of Southern Queenland. We wish to thank the boys and staff of Toowoomba Grammar School for their cooperation.
We are grateful to P. Ackerman who suggested the use of Perceptual Speed measures in the second study. We are also grateful to R. Roberts for permission to use his data in this second study.