

SPACE FORTRESS AS AN IQ TEST? PREDICTIONS OF LEARNING AND OF PRACTISED PERFORMANCE IN A COMPLEX INTERACTIVE VIDEO-GAME *

Patrick RABBITT, Nicole BANERJI and Alex SZYMANSKI

University of Manchester, UK

Claims that scores on pencil and paper IQ tests predict performance in easy laboratory perceptual motor tasks are weakened by methodological inadequacies. With an experimental design avoiding these weaknesses the AH 4 IQ test predicted rate of learning and performance after 5 days practice on 'Space Fortress' better than did age, between 18 and 36 years, or amount of previous experience at video-games. However, AH 4 scores only weakly predicted initial performance. Correlations between AH 4 raw scores and practised 'Space Fortress' game scores were at $r = 0.69$, that is, as high as are normally attained between different IQ tests. These findings remained robust even when re-examined within a sub-group of 43 subjects scoring within the top 10th percentile of the population, both on the AH 4 and on a much more difficult IQ test, the AH 5, designed for selection within University populations. When combined with AH 4 raw scores initial performance scores on training sub-tests for 'Space Fortress' were very efficient predictors of practised performance (69% of variance between individuals) and rate of learning (38% of variance between individuals). Clearly interactive video games can be tailored to provide very efficient selection and training instruments for complex industrial and military information handling tasks.

Introduction

Psychologists have recently been excited by suggestions that scores on pencil and paper IQ tests may predict speed in easy Reaction Time (RT) tasks because they have taken this as evidence that the speed and accuracy with which people can discriminate between simple signals and solve complex verbal, arithmetical and spatial problems in IQ tests may, alike, reflect a common, inheritable biological property of the human Central Nervous System (Eysenck 1986; Jensen 1982; Vernon 1983). Indeed, some investigators have suggested that this property can

* Requests for reprints should be sent to P. Rabbitt, University of Manchester, Age and Cognitive Performance Research Centre, Oxford Road, Manchester, M13 9PL, UK.

directly be identified by measurable electrophysiological indices of 'neural signal to noise ratio' (A.E. Hendrickson 1982; D.E. Hendrickson 1982; Hendrickson and Hendrickson 1980; Blinkhorn and Hendrickson 1982).

In practical terms these studies imply that brief pencil and paper IQ tests, originally designed for academic selection and for the evaluation of high level cognitive skills, may also provide a very cheap, convenient and effective means for selecting people for training in complex perceptual motor tasks. In particular, to discover which individuals can best meet the complex perceptual-motor demands which arise from industrial and military applications of information technology.

Unfortunately the results so far published pose two problems: they may be unreliable, and we do not know what they mean. They may be unreliable because observed correlations between IQ test scores and RTs have usually been modest and sometimes absent (e.g. Sternberg (1984) quotes median values of $r = 0.2$, i.e. accounting for only about 4% of total variance). Even these weak associations have sometimes been obtained only when people with very low, or even with sub-normal, IQ test scores have been included in the populations sampled (Mackintosh 1981; Nettlebeck and Lally 1981). This last difficulty has been compounded by the fact that in nearly all published studies subjects have been given surprisingly little practice – often as few as 10 to 50 trials on each task investigated. Thus the correlations so far observed may only reflect differences in the times which more and less gifted individuals take to adjust to novel tasks in unfamiliar environments. So, while IQ test scores may predict from 4% to 10% of variance in efficiency between individuals on their first encounters with some simple tasks, we do not yet know whether they predict how rapidly people improve at these or other perceptual motor tasks or what maximum levels of performance they can attain with practice. In brief, it is premature to base a speculative neurology of individual differences on samples of only 10 to 50 trials on very easy reaction time experiments.

We do not know what published correlations between scores on pencil and paper IQ tests and on psychomotor tasks mean because the tasks investigated to date have been so unchallenging, dull and abstract that the observed individual differences in performance are quite as likely to reflect motivational factors as true individual differences in maximum information processing rate. No perceptual motor tasks yet

used in IQ comparisons have demanded capabilities such as the rapid interpretation of complex scenarios, the prediction of immanent events or the discovery and use of strategies to cope with a rapidly changing environment which both expert psychometricians and laymen associate with the concept of 'intelligence' (Sternberg et al. 1981). The Space Fortress game allowed tests of all these theoretical and methodological points.

From the practical point of view the developing technology of video-games raises the possibility that tailor-made complex interactive tasks such as Space Fortress can be used as safe, convenient and inexpensive training aids for the increasingly complex information handling tasks demanded by the growth of information technology. While insightful and comprehensive discussions of transfer of training between components of complex tasks suggest that optimism must be qualified (Weightman and Lintern 1985) there is the further interesting possibility that use of complex interactive tasks as selection rather than training aids may help to focus on those individuals who are most likely to achieve excellent performance; in particular:

(1) Whether existing pencil and paper IQ test can provide efficient selection criteria for training on complex tasks which demand very rapid evaluation of complex scenarios and the implementation of fast decisions about these scenarios using complex motor control procedures?

(2) Very complex interactive tasks can, logically, be broken down into sub-sets of component skills. The developing technology of video-games allows efficiency at each component skill, embedded in a specific 'training game' or 'sub-task' to be evaluated separately from efficiency at the overall game scenario. How do predictions from raw scores on pencil and paper IQ tests compare with predictions from initial performance, and from rate of learning on 'sub-tasks' which are either components, or 'stripped down' simplified versions of a very complex task? I.E. how do IQ test scores compare with the 'Space Fortress' 'sub-tasks', or 'component games' as predictors of initial performance, rate of improvement and maximum attainment at the complex Main Game scenario? Are tasks which simulate component skills in a complex scenario also the best selection tests as well as training devices for that scenario? Or do IQ scores add to the prediction – perhaps because they pick up 'Meta-components' of task performance (Sternberg 1984)

which allow individuals to integrate a repertoire of much simpler component skills in order to meet very complex demands? The time is now ripe for revision of excellent earlier discussions of predictions of initial performance, acquisition rates and terminal performance at very simple motor skills by authors such as Jones (1970) in terms of very complex, multi-component tasks.

(3) An existing literature encourages us to expect that individual differences in scores on IQ tests may correlate with initial performance, at least on very simple psychomotor tasks. Surprisingly, only very few studies (e.g. Jones 1970) have yet examined correlations between test scores and *rates of improvement* or *maximum levels of performance* attained with practice. The Jones (1970) study is a special case because the task (micrometer use) was simple enough to allow all trainees to converge on a common performance asymptote.

The 'Space Fortress' task offered unique opportunities to ask these questions in a task in which upper limits of performance did not approach ceiling. The main game scenario was assembled from a number of component sub-tasks, each derived from a particular experiment paradigm currently fashionable in Cognitive Psychology, and each independently testable as one of a set of 'training tasks'. This allowed us to compare how well scores on conventional pencil and paper IQ tests predicted initial performance, rate of learning and maximum skills attained on each sub-task as well as in a complex game scenario into which they were all integrated.

Method

Subjects

Fifty-six young men, undergraduates and post-graduates from the University of Manchester, and unemployed attending Manchester job-centres were each paid £15 to complete a 1 hour training session on each of 5 successive working days. During each training session each was tested once on each of the five individual sub-tasks (i.e. Ship Control, Aiming, Interval Judgement, Memory Search and Easy Game) and five times on the Main Space Fortress game. On recruitment, each volunteer rated his previous experience with video-games of any kind on a 10 point scale. He was then extensively interviewed about this experience and again rated by experimenters on a 7 point scale. All volunteers were given the AH 4 test of general intelligence (Heim and Batts 1948). Forty-three who scored at level 'A' on this test (i.e. whose scores fell into those attained

Table 1
Description of subjects tested on Space Fortress.

Number of subjects	56
Mean age	24.04
SD age	5.30
Youngest	18.2
Oldest	36.6
Mean AH 4 total score	97.0
SD total AH 4 scores	27.1
Lowest score	36.0
Highest score	117.0

by the top 10% of the UK population) were then given a further, much more difficult, IQ test, the 'AH 5', which had been developed to discriminate between members of a University population (Heim 1947). Details of subjects, and scores are given in table 1.

Scores for the Space Fortress game and associated training tasks are described elsewhere in this volume (see Mané and Donchin 1989, this vol.; Mané et al. 1989, this vol.).

Results

Predictions from task external measures of individual differences

The first analysis examined how each of three possible sources of individual differences, IQ test score, age and previous experience of video-games affected each of three different performance measures: (1) Initial performance, (2) Rate of Improvement with Practice and (3) Maximum Score attained. An exploratory overall correlation analysis guided subsequent specific multivariate regression analyses which tested how sources of individual difference individually and jointly predicted performance.

Initial scores

Ratings of previous experience of video-games correlated significantly with initial scores on the main game (self-ratings $r = 0.331$; experimenter's ratings $r = 0.404$ $p < 0.01$). Negative correlations between these ratings and chronological age ($r = -0.576$ and $r = -0.561$) must partly represent unsurprising generation effects in the availability of video-games during adolescence. However, predictions of initial game score from chronological Age remained significant even when variance associated with rated previous experience had been partialled out by regression analysis. Thus it seems likely that initial difficulties experienced by older volunteers were partly, but not solely, due to their relative lack of previous experience with video-games.

Over all 56 volunteers AH 4 scores correlated modestly but significantly with initial main game scores ($r = 0.283$). For the 43 volunteers with raw AH 4 scores within the top 10% of the UK population initial game scores also correlated significantly with AH

5 test scores ($r = 0.254$). A multiple regression analysis compared Age, experimenter's ratings of previous experience and AH 4 test scores in turn as predictors of initial scores after the variance predicted by both other factors had been partialled out. The three factors considered together gave a significant prediction ($F = 3.98$, df 3,39 accounting for 6.7% of the total variance between individuals), but no factor on its own gave a significant independent prediction. This was also true for the 43 volunteers with AH 4 scores within the top 10th percentile. For this group when AH 5 scores were compared as predictors with Age and Experience only Age gave any significant prediction of residual variance ($t = -2.174$, df 40, $p < 0.05$, 5.2%).

Rate of improvement with practice

Least squares regression functions were fitted to means for the 25 game scores obtained over 5 successive practice days. Both AH 4 ($r = 0.422$) and AH 5 ($r = 0.325$) test scores correlated significantly with the slopes of these learning functions but Chronological Age and rated previous experience did not. When contributions of Age, Experience and IQ test scores were jointly compared in a multiple regression analysis, only IQ test scores emerged as significant independent predictors (AH4, $t = 2.82$, df 53, $p < 0.01$, 11.8% of variance; AH 5, $t = 2.151$, $df = 37$, $p < 0.05$, 11% of variance).

Maximum scores

Unsurprisingly most volunteers attained their maximum scores on the fifth day of practice. Over all 56 volunteers the correlation between AH 4 scores and Maximum Game scores was $r = 0.68$ (significant on multiple regression analysis at $t = 4.29$, $p < 0.01$) accounting for 17.21% of variance residual to that associated with Age ($t = -2.35$, df 53, $p < 0.05$, 5.1% of residual variance) and rated Experience ($t = 0.93$ nsd). For the 43 volunteers in the top 10th percentile AH 5 test scores and maximum main game scores correlated at $r = 0.3$ (significant on multiple regression analysis after partialling out effects of Age and Experience ($t = 2.72$, df 37, $p < 0.01$; 13.3% of residual variance) when Age ($t = -2.447$, df 37, $p < 0.01$, 9% of residual variance) and Experience ($t = 0.907$ nsd) had been considered. Over all 56 volunteers age correlated significantly with maximum attained game score ($r = -0.28$, $p < 0.01$) but this correlation vanished when variance due to difference in AH 4 test scores was partialled out. These findings make two new points: first, scores on pencil and paper IQ tests predict practised levels of performance on a complex video-game better than does Age, or amount of previous Experience. The strength of this prediction is much higher than has often been reported for the first few trials on very easy reaction time tasks. This does not seem to be due to improvements in the strengths of correlations with increasing task difficulty (Jensen 1985) so much as to a second factor, unexamined in any previous experiments: IQ test scores predict rate of learning more powerfully than initial performance on a complex task. This factor, so far unexamined in easy choice RT and inspection time tasks, turns out to be crucial because, as a consequence, test scores predict maximum levels of performance achieved after extensive practice significantly better (Hotelling's test, $p < 0.01$) than they predict initial performance or rate of learning alone. All these associations appear even within a sub-group whose IQ test scores fall within the narrow band attained by the top 10% of the UK population.

Comparisons of predictions from IQ test scores and from training tasks

Data from the 43 top 10th percentile subjects were analysed in order to make the most stringent, and so, for practical purposes, the most useful comparisons of predictive value of IQ test and sub-games. An initial problem was that scoring systems used for individual sub-games yielded very different ranges of scores (see Mané and Donchin (1989, this vol.) for details). Accordingly all scores on all training games and

Table 2
Predictions for Z transformed scores from all tasks from age and AH 4.

Task	Age (t)	AH 4 (t)	Total variance from both	Age component	AH 4 component
<i>Initial performance</i>					
Aiming	- 3.33	4.32	43.4%	15.9%	27.5%
Ship					
Control	- 2.90	1.68	19.9%	16.8%	3.1%
Interval					
Judgement	- 1.67	3.06	21.8%	5.0%	16.8%
Letters					
Memory Search	- 2.24	1.97	16.1%	10.5%	5.6%
Easy					
Game	- 3.07	2.270	29.2%	15.9%	13.3%
Main Game	- 2.53	1.64	16.6%	12.7%	3.9%
<i>Amount of improvement over practice</i>					
Aiming	- 0.88	0.656	3.30%	2.0%	1.3%
Ship					
Control	- 1.24	4.19	31.00%	2.6%	28.4%
Interval					
Judgement	0.68	- 1.04	0.50%	0.01%	49.0%
Letters					
Memory Search	1.02	1.42	1.76%	0.45%	1.31%
Easy Game	0.99	1.24	2.50%	1.3%	1.2%
Main Game	- 1.76	3.94	30.96%	5.2%	25.76%
<i>Performance on final practice session</i>					
Aiming	- 4.05	4.9	51.4%	20.66%	30.74%
Ship					
Control	- 2.78	3.80	35.4%	6.64%	28.76%
Interval					
Judgement	- 1.98	4.20	34.4%	3.27%	31.13%
Letters					
Memory Search	- 2.05	5.70	47.8%	5.20%	42.40%
Easy Game	- 2.3	4.65	40.3%	5.20%	35.1%
Main Game	- 3.5	4.14	43.0%	11.94%	31.06%

on the Main Game were normalised by z-transforms about a mean of 100; This was also done for AH 4 total raw scores. A series of separate multiple regression analyses were then run to compare Age with initial (first session) score on each of the training games, with score on first encounter with the Main Game, and with transformed AH 4

Table 3

(a) Multiple regression predicting practised performance on Space Fortress.

Variable	Mean	SD	B*	B	SE est.	t
AH 4	100.023	15.076	0.154	0.259	0.190	1.363
Aiming Task	100.069	14.968	0.350	0.591	0.276	2.139
Ship Control	100.209	14.897	0.184	0.313	0.253	1.235
Interval Task	100.093	15.077	0.247	0.414	0.184	2.247
Memory Search	100.023	15.015	0.51	0.086	0.186	0.465
Easy Game 1	100.093	15.001	0.018	0.183	0.208	0.878
Main Game 5	145.093	25.263				

Squared correlation coefficient = 0.730 (after adjustment for shrinkage 0.685)

Sample correlation coefficient = 0.854, Standard error of estimate = 14.163

Intercept = -39.978, t values to be evaluated with 36 degrees of freedom

Source	Sum of Sq	df	Mean Sq	F
Regression	19584.316	6	3264.052	16.272
Residual	7221.302	36	200.591	
Total	26805.619	42		

(b) Multiple regression predicting improvement in Main Game.

Variable	Mean	SD	B*	B	SE est.	t
AH 4	100.023	15.076	0.213	0.269	0.201	1.337
Aiming Task	100.069	14.968	0.493	0.628	0.292	2.149
Ship Control	100.209	14.897	-0.245	-0.314	0.268	-1.170
Interval Task	100.093	15.077	0.247	0.312	0.195	1.603
Memory Search	100.023	15.015	0.002	0.002	0.197	0.012
Easy Game 1	100.093	15.001	0.016	0.021	0.221	0.096
Main Game Imp	45.116	19.046				

Squared correlation coefficient = 0.469 (after adjustment for shrinkage = 0.380)

Sample correlation coefficient = 0.685, Standard error of estimate = 14.986

Intercept = -46.927, t values to be evaluated with 36 degrees of freedom

Source	Sum of Sq	df	Mean Sq	F
Regression	7150.808	6	1191.801	5.306
Residual	8085.610	36	224.600	
Total	15236.418	42		

scores as predictors for initial Main Game Score, final Main Game Score and differences between initial and final Main Game score (a better measure of improvement with practice than the slope of regressions fitted to non-linear learning curves). Regression analyses gave the total variance accounted for by both measures jointly, and the proportions of variance accounted for by each when all variance accounted for by the other had been partialled out. These, with associated t values, are given in table 2.

Table 2 shows that Age makes no significant predictions when effects of AH 4 raw scores, or of initial performance on training tasks has been partialled out. Only 3 tasks are significant predictors of initial Main Game performance when associated with Age as a joint predictor (Ship Control, 56.4% of total variance; Easy Game, 37.3% of total variance, and the Aiming task 33.7% of total variance). As we have seen, AH 4 scores do not predict initial performance where these other variables are considered. However, all tasks and AH 4 scores predict practised Main Game performance. First trial scores on the Aiming task are by far the best predictor (with Age, 60.6% of total variance). AH 4 scores predict terminal performance about as well as do first trial scores on all other training tasks (42.9% of variance as compared with 32% to 44.7% or variance) or, indeed, as well as or marginally better than first trial Main Game scores (45% of variance as against 31% of variance).

It is interesting that while subjects' scores on their first trials on the Main Game do predict the levels of performance that they will attain after 5 days practice (42.9% of variance) they do not predict the rate of improvement with practice ($p > 0.2$). Thus AH 4 test scores, and scores on two other simpler component tasks, predict rate of learning on a complex task better than do initial scores on the task itself.

Obvious practical questions were how well all these measures, taken together, predict rate of Main Game learning and performance after practice, and how far IQ test scores and task scores predict common, and how far independent, portions of the total variance in these performance measures. Multiple regressions used AH 4 scores and initial scores from the 5 training tasks as joint and independent predictors of final scores, and final-initial difference scores on the Main Game. Details are given in table 3.

It seems that AH 4 scores and initial scores on the various training tasks largely overlap in terms of the variance in Main Game performance which they predict. For predictions of final Main Game scores the squared correlation coefficient adjusted for shrinkage was 0.69; i.e. when used as joint predictors these measures account for 69% of variance in performance. However, we see from table 3 that only the Aiming task ($t = 2.14$; $p < 0.05$; 4.1% of variance) and the Interval judgement task ($t = 2.25$; $p < 0.05$; 4.4% of variance) predict rate of learning (terminal-initial Main Game scores); the 6 measures jointly predicted 38% of individual variance, but only Aiming task scores make any significant independent contribution ($t = 2.15$; $p < 0.05$; df 36; 7.9% of variance).

Discussion

Subjects' ratings of their previous experience with video-games predicted their initial performance at Space Fortress, but not their subse-

quent rate of learning nor the maximum scores which they achieved after playing the game 25 times over 5 successive days. Although self-ratings and interviews provide only very crude indices of differences in individuals' previous experience with video-games this pattern of correlations is intuitively plausible: people who discover that they are adept at learning new video-games will tend to play them more often; but not all potentially adept learners may discover, or necessarily enjoy, their latent ability. It is also plausible that with increasing specific practice on a new game the importance of transfer of training from other games should gradually become a less significant factor.

Age between 18 and 36 years predicted initial levels of performance, even when ratings of previous experience had been taken into account. However, this prediction disappeared when effects of variance in AH 4 test scores was partialled out. Age, alone, did not predict rate of improvement, but did predict maximum attained scores. Older volunteers seem to start with an initial handicap but, thereafter, to learn as fast as the young. However, since they do not learn *faster* than the young their initial handicap persists throughout practice and they attain lower maximum scores. This is the first demonstration in the literature of such an early age-effect (i.e. 18 to 36 years) on performance of a complex task. It is especially interesting that Age, as a source of individual differences, should have effects discriminable from those of IQ test score (cf. Rabbitt and Goward 1986); i.e. within this age range IQ, but not Age, predicts rate of learning, but Age also predicts practised performance.

When effects of Age and previous Experience have been partialled out IQ test scores do not predict initial performance but do predict rate of subsequent learning and so, even more strongly, practised performance. It is surprising that the plausible finding that IQ test scores predict learning rate in a psychomotor task should, with rare exceptions (see Lintern 1976) be neglected in the literature. Perhaps this has been because some early investigations found no (e.g. Woodrow 1940) or only marginal (e.g. Zeaman and House 1967) correlations between IQ test scores and verbal learning scores. This new finding raises methodological difficulties for most previous studies of relationships between IQ test scores and psychomotor performance (e.g. Eysenck 1982, 1986; Jensen 1982, 1985; Vernon 1983; Nettlebeck and Lally 1981). If subjects are compared on only a very few trials of a task, predictions of task performance from IQ test scores may be absent or

negligible and may, in any case, be strongly moderated by quite different and dissociable factors such as Chronological Age and previous Experience. However, as practice continues it seems likely that any initial differences in attainment between High and Low test scores will be increasingly amplified by differences in their learning rates. Thus levels of correlations between test scores and performance indices will depend on the number of trials over which means have been computed, and are likely to increase with the amount of practice given. This also highlights a different methodological difficulty with studies in which many different tasks have been given during the same experimental session, and always in the same order, with difficult tasks experienced last (cf. Vernon et al., cited by Jensen 1985). Because people with High IQ test scores show more marked non-specific practice effects than Low test scorers, differences between groups are likely to be greatest on tasks administered late in the testing session. Models for IQ effects premised on suggestions that IQ test scores predict performance on difficult tasks better than on easy tasks must obviously be reevaluated to take account of this confounding factor in all the data on which they have been based.

A more important consequence of this correlation between IQ test scores and learning rates is that if correlations between test scores and average performance are computed from averages of data gathered over an entire testing session, they may disguise much more basic, and theoretically more informative, relationships between test scores and learning ability upon which they depend. Within the 43 most able subjects AH 4 test scores predicted rate of improvement on the Space Fortress game as well as any, and better than most, of a set of much simpler psychomotor tasks which were actually components of the Main Game scenario.

An equally important finding is that AH 4 test scores did not predict amount of improvement with practice promiscuously across all tasks, but rather selectively for some and not for others. Table 2 shows that, when effects of Age had been partialled out, AH 4 scores, on their own, predicted rate of improvement in Ship Control and in the main Space Fortress game but not in any other task. This cannot be interpreted as an increase in predictive value of IQ test scores with task difficulty (and so with 'g' loading) since AH 4 scores predicted improvement at Ship Control, but not at the much more difficult Easy Game which actually includes Ship Control as one among several diverse, component skills.

This experiment gives little support to the idea that AH 4 scores pick up a single factor common to all tasks. Indeed it brings into question the idea that correlations between IQ test scores and reaction times are a royal road to understanding the biological bases of intelligence (Brand and Deary 1982; Eysenck 1986; Jensen 1985). These authors argue that such correlations are best explained on the assumption that a single psychometric factor 'g', common to all IQ test, can be identified with a single performance characteristic, 'information processing rate' which is picked up by all perceptual motor tasks, to a greater or lesser degree depending on their relative difficulty. It has been further suggested that this performance characteristic can, in turn, be reified in terms of a single hypothetical neurophysiological index such as 'neural noise level' (Blinkhorn and Hendrickson 1982; A. Hendrickson 1982; D. Hendrickson 1982). The present results favour an opposing, common sense, view that IQ tests may be regarded as collections of very disparate cognitive tasks, so diverse that at least some of them are likely to make demands on some of the disparate cognitive sub-systems that are involved in each of most everyday activities. On this model it is likely that some of the problems in any good IQ test will test at least some of the skills required in any given everyday task. This overlap between task demands will result in significant, but generally quite weak, correlations between most IQ tests and most laboratory tasks which are designed to test a very specific one of many possible cognitive skills. This is precisely what the majority of studies to date have found. More complex skills, which make demands on much wider ranges of cognitive sub-systems, are correspondingly likely to have greater overlap with any good IQ test and to show correspondingly higher correlations with test scores. Thus in general it is likely, but not inevitable, that correlations between IQ test scores and everyday tasks will increase with task difficulty.

The present data also suggest that high correlations between test scores and game performance may occur because people who can master most of the wide range of problems included in IQ tests can also more rapidly learn to master complex systems of rules, to attend selectively to the critical portions of complex scenarios, to make rapid and correct predictions of immanent events and to prioritise and update information in working memory. There is no evidence that any of these higher level skills depends *solely* on information processing rate; indeed there is increasing evidence that such skills as working

memory efficiency, recognition memory and selective management of information are well predicted by performance on the complex, and very diverse, problems posed by paper and pencil tests and not at all by simple measures of decision speed (Goward 1987).

While test data raise some severe methodological problems for earlier experiments they also resolve other uncertainties about the reliability of correlations between IQ test scores and perceptual motor performance. Evidently the correlations obtained in the present experiment do not solely reflect individual differences in motivation at dull laboratory tasks, or inclusion of low IQ people within groups compared, or transient individual differences in speed of adaptation during early trials on unfamiliar tasks in unfamiliar situations. AH 4 scores predicted performance in a complex, fast interactive video-game which was greatly enjoyed, and very competitively played, by all our volunteers; correlations remained significant even within the top 10% of the population and the advantage in performance for high over low test scores significantly increased with practice. Thus, although these limited results cannot, of course, tell us whether IQ test scores predict differences in the maximum levels of performance which individuals can attain after indefinitely long practice, they do weigh against the possibility that low IQ test scorers simply need longer practice to converge upon the same performance asymptotes attained by high test scorers.

In practical terms it is interesting that in a group of 56 young subjects with a very wide range of IQs levels of correlations between a relatively crude summary index of efficiency after training at a complex video-game (total game score) and scores on a well standardised and validated test of general intelligence (the AH 4) are as high as $r = 0.68$, thus accounting for up to 46% of individual variance in performance. This association is comparable to those between the AH 4 and other, well standardised, and reliable pencil and paper tests (i.e. $r = 0.58$ to 0.85 , or 34% to 72% of individual variance, Heim 1968). One interesting implication is that a relatively unsophisticated video-game, on which performance may reasonably be expected to be independent of native language or acquired literacy, and which is greatly enjoyed by young people who play it, rank orders individual differences in 'intelligence' nearly as well as pencil and paper psychometric tests which have been specially developed for this purpose over the last 80 years. An equally important implication is that structure of the Space Fortress game and its sub-tasks now provides us with an exemplar which can help us to

design, for any complex interactive information processing task, a 'battery' of performance tests which will jointly predict rate of improvement and maximum attained performance far better than initial performance at the task itself. It seems to be easily possible to design very efficient selection batteries to detect precisely those individuals who are most capable of attaining excellent performance at complex, interactive, decision-making tasks in which rapidly changing information is symbolically displayed on a video screen. Pencil and paper IQ tests are useful components of such selection batteries in that they predict rate of learning of complex tasks as well as, or better than, sub-tasks which train and evaluate component skills.

References

- Blinkhorn, S.F. and D.E. Hendrickson 1982. Averaged evoked responses and psychometric intelligence. *Nature* 295, 596–597.
- C.R. Brand, and I.J. Deary 1982. 'Intelligence and inspection time'. In: H.J. Eysenck (ed.), *A model for intelligence*. Heidelberg: Springer-Verlag. pp. 133–148.
- Eysenck, H.J., 1982. What's new in the measure of intelligence? *Bulletin of the British Psychological Society* 35, A7, abstracts.
- Eysenck, H.J., 1986. 'The theory of intelligence and the psychophysiology of cognition'. In: R.J. Sternberg (ed.), *Advances in the psychology of human intelligence*, Vol 3. Hillsdale, NJ: Erlbaum. pp. 1–34.
- Goward, Louise, 1987. An investigation of the factors contributing to scores on 'Intelligence Tests'. Unpublished PhD thesis, University of Manchester.
- Heim, A.W., 1947. An attempt to test high grade intelligence. *British Journal of Psychology* 37, 70–81.
- Heim, A.W., 1968. AH 4 test. Windsor: NFER-Nelson
- Heim, A.W. and V. Batts, 1948. Upward and downward selection in intelligence testing. *British Journal of Psychology* 39, 22–29.
- Hendrickson, A.E., 1982. 'The biological basis of intelligence, Part 1: Theory'. In: H.J. Eysenck (ed.), *A model for intelligence*. Heidelberg: Springer-Verlag.
- Hendrickson, D.E., 1982. 'The biological basis of intelligence, Part II: Measurement'. In: H.J. Eysenck (ed), *A model for intelligence*. Heidelberg: Springer-Verlag.
- Hendrickson, A.E. and D.E. Hendrickson, 1980. The biological basis for individual differences in intelligence. *Personality and Individual Differences* 1, 3–33.
- Jensen, A.R., 1982. 'Reaction time and psychometric g'. In: H.J. Eysenck (ed), *A model for intelligence*. New York: Plenum. pp. 93–132.
- Jensen, A.R., 1985. the nature of the black–white difference on various psychometric tests: Spearman's hypothesis. *The Behavioural and Brain Sciences* 8, 193–219.
- Jones, M.B., 1970. A two-process theory of individual differences in motor learning. *Psychological Review* 77 353–360.
- Lintern, G., 1976. Field independence, intelligence, and target detection. *Human Factors* 18, 293–298.
- Mackintosh, N.J., 1981. A new measure of intelligence. *Nature* 289, 529–530.

- Mané, A. and E. Donchin, 1989. The Space Fortress game. *Acta Psychologica* 71, 17–22 (this volume).
- Mané, A.M., J.A. Adams and E. Donchin, 1989. Adaptive and part–whole training in the acquisition of a complex perceptual-motor skill. *Acta Psychologica* 71, 179–196 (this volume).
- Nettlebeck, T. and M. Lally, 1981. IQ put to the test. *Nature* 290, 240.
- Rabbitt, P. and L. Goward, 1986. Effects of age and raw IQ test scores on mean correct and mean error reaction time in serial choice tasks. *British Journal of Psychology* 77, 69–73.
- Sternberg, R.J., 1984. *Beyond IQ: A triarchic theory of intelligence*. Cambridge: Cambridge University Press.
- Sternberg, R.J., B.E. Conway, J.L. Ketron and M. Bernstein, 1981. Peoples conceptions of intelligence. *Journal of Personality and Social Psychology* 41, 37–55.
- Vernon, P.A., 1983. Speed of information processing and general intelligence. *Intelligence* 7, 53–70.
- Weightman, D.C. and G. Lintern, 1985. Part task training for tracking and manual control. *Human Factors* 27, 267–283.
- Woodrow, H., 1940. Interrelations of measures of learning. *Journal of Psychology* 10, 49–73.
- Zeaman D. and B.J. House, 1967. In: R.M. Gagne (ed.), *Learning and individual differences*. Columbus: OH: Merrill.