

# Aptitude Testing



## Background

There has been a long and ongoing discussion about the validity and fairness of aptitude testing, particularly in relation to Māori, in job selection in New Zealand and in the Public Service in particular. The views formed, both for and against, have been based on anecdotal evidence or evidence from the United States or Britain because there has been a singular lack of research done in New Zealand on the topic. The debate has also sometimes been uninformed, with some people confusing Aptitude Tests with Intelligence Tests and basing their argument on that premise. The purpose of this paper is to provide some empirical evidence to the debate so that if nothing else, that debate can become more informed.

To do this, I have chosen to use a single occupational group, Policy/Senior Analysts, as the centre of the study for a number of reasons. The main reasons are: that a key competency for the role is analytical reasoning which relates to the ability of the individual to evaluate the logic of various kinds of arguments, which can be reliably measured by a verbal aptitude test: that we have sufficient data collected to be able to construct a norm group: and that there are external norms for the general population (which is predominantly non Māori) to compare the Māori norm group against.

The other reason is that there have been two pieces of research carried out using earlier subsets of the data I will use, which give me a benchmark to work from.

### Predicting performance

In relation to predicting job performance there is a seminal piece of research which is a meta analysis of 85 years of research findings looking at 19 different selection methodologies, published in the American Psychological Society Inc. Journal of Personal and Social Psychology in 1998, carried out by Frank. L. Schmidt of the University of Iowa, and John. E. Hunter of the Michigan State University.

This research indicates that for hiring employees without previous experience on the job, the single most valid predictor of future performance and learning is General Mental Ability (GMA). They also looked at what other selection methods could be used to improved the ability to predict when added to GMA.

In 2004 this research was expanded to provide evidence that the GMA predicts both occupational level attainment and performance within a chosen occupation better than any other (do you mean specific here?) ability, trait, or disposition and even better than job experience.

In essence this research indicates that on a scale of 0 to 1 (1 being a perfect predictor) the use of GMA in selection across a very wide variety of occupations will return a score of .51. Compare this to years of education which is .10 or “job tryout” which is .44 and it is clear that there are benefits to be accrued from using GMA test in a recruitment process.

The research also indicated that Work Sample tests had a slightly higher predictive validity .54, however these require work experience and are therefore not useful in the selection of people at an entry level. They were however valuable when the job required previous experience, such as Senior Analyst or Management positions.

The 2004 research expanded this to look at job complexity and indicates that the higher the level of complexity of the job, the better the GMA is likely to predict future job performance and, in addition, to predict the ability of the individual to acquire new skills through training. For example for a job with a level 1 complexity, the most complex jobs, the research indicates that the GMA has a predictive validity of .58 and will predict training success at .59, and level 2 had predictive validities of .56 and .65, level 3 figures were .51 and .57

### **Job Complexity**

Level 1 complexity jobs are described as professional, scientific and upper management, level 2 were complex technical jobs such as computer systems trouble shooter, or complex manufacturing jobs, and level 3 were skilled workers, technicians, and para-professionals.

I would argue that in general terms the role of Manager in this organisation could reasonable be described if not as “*professional, scientific and upper management*” then they could certainly be described as “*complex technical jobs such as computer systems trouble shooter, or complex manufacturing jobs*”. I would also argue that the role of analyst/senior analyst most certainly fits the latter description. I have attached the most recent competencies for these jobs as Appendix 1. This being the case it is reasonable to conclude that the use of GMA on its own in selection would produce at least a .55 predictive validity, and a .6 predictive validity for training for the vast majority of our positions.

### **Content Validity**

In relation to fairness it is important that what the tests purport to measure are important for the job the candidate is applying for, content validity.

Content validity is about showing that the tests used links to job content, that it is actually measuring things which are requirements of the job. This is generally validated from the data gathered from a job analysis exercise. A number of such exercises have been completed in the past and key job requirements identified. Each of these have identified that a key competency for the main business of The Organisation is that of verbal analysis as we employ in excess of 60 people who are described as policy analysts or senior policy analysts.

This role requires the job holder to be able to be able to analyse and interpret high level written information. *Analyses data, identifying trends and developing creative solutions for Maori development; seeks empirical evidence to substantiate arguments; develops and tests alternative courses of action; thinks laterally to assess the implications for Maori and the Organisation’s position on issues; interprets policy, procedures and legislation and applies principles accurately* is the key competency being measured.

The test which we have consistently used to measure this has been the VMT3 verbal analysis test which is specifically designed to test the candidate’s ability to understand and interpret complex written reports and policy documents. This provides a high level of both face and content validity and it is able to tap into the level of General Mental Ability of the candidate, all of which is critical to the work of most positions in The Organisation.

## **Impact on Māori**

Researching this was quite problematic in that there are only three studies done which specifically look at the impact of “testing” on Māori. All of these rely on the Organisation data. There is however, a lot of research on the impact of testing on “ethnic minorities” which was also considered in our review as this has tended to have been the research which has informed debate on the subject within the Organisation in the past.

The issue which was pursued was: if General Mental Ability (GMA) was such a powerful predictor of future job performance for people with no work experience, would it unfairly discriminate against, or disadvantage Māori?

This has been a contentious issue and as a result, inevitably there has been some research done to try and provide the answer. The first New Zealand research looked at differences between Māori and non-Māori work style preferences rather than cognitive ability. The second looked at the same area but also added critical reasoning tests which measure cognitive ability which has been the most hotly debated part of good recruitment practices. The tests used were targeted at a management graduate level and did not represent the most complex available. The results of this study indicated that there were no statistically significant differences between Māori scores and non-Māori scores on either the verbal reasoning test or the numerical reasoning test. This study had a limited participation rate of 49 employees of the Organisation.

The third study had a much larger sample size to work with and focussed on three tests. Two of these were from the Advanced Managerial Test battery which was the most complex available. The other was from the same test bank as the previous study, the Management Graduate Test Bank.

The findings of this report were that there was a difference between Maori and non-Māori on two of the three tests measured. The differences were as follows.

	Verbal Reasoning Test Mean (worth indicating the test reference here)	Numerical Business Analysis Test Mean	General Numerical Reasoning Test Mean
Māori	19.28	8.21	14.72
Non-Maori	22.2	14.90	14.76

This does not suggest test bias, it simply indicates that for reasons unknown, Māori scored, on average, lower than non-Māori on two of the three tests. It is of interest to note here that both this study and the previous one found that the lower level tests (General Numerical Reasoning Test) did not produce different results for Māori. The question is therefore do the other two tests disadvantage Māori? The authors suggest that the difference between the results for the two numerical tests could be related to the fact that the ‘Business’ test requires business knowledge and experience.

They suggest that a way of minimising the impact that previous knowledge could have on assessing the candidates General Mental Ability, which as pointed out in international research still shows that lower test scores are accompanied by lower performance irrespective of ethnic origin, a test such as the Ravens Progressive

Matrices which does not rely on vocabulary or prior business knowledge may produce smaller ethnic group differences whilst still measuring cognitive ability. This does however create problems with demonstrating content validity as they are a form of diagrammatic reasoning test which bear no observable resemblance to the nature of the jobs. This is a key reason for their decline in popularity of recent years.

The paper concludes that there are limitations to the research which in some part may have been governed by the shortlisting process used by managers to select people who were to be tested, and the possibility that they were more likely to be prepared to shortlist a Māori candidate who may not have been the ideal match for the position than a non-Māori one on the basis that they could invest time and money in their eventual development if they were successful in getting the job.

It is worth noting here that the norm group which The Organisation uses in the selection process for the verbal reasoning test is a Composite norm group which has a mean of 19.84. Given that the above research indicates that the norm for the Māori sample is 19.28, there is no statistically significant difference between the means and therefore it is reasonable to conclude that the test does not unfairly disadvantage Māori applicants despite the fact that non-Māori applicants, on average, score higher. It is also worth noting that the standard deviation for the research above is 5.35 and for the composite norm group used it is 5.11, which is again a similar spread of scores, re-enforcing the idea that the test is not disadvantaging Māori unfairly.

The research does however raise issues which need to be carefully managed in the future and HR are constantly monitoring this and adding new data as it becomes available, to add further clarity to the issue, and to ensure that the use of the tools remains appropriate.

Does this research tell the full story? Possibly not. It for example indicates that there is an important difference between Māori and non-Māori on the Verbal reasoning test which suggests that it should not be used. The Organisation's problem is that these tests measure just the sorts of things we need to know about in terms of the capability of applicants. It was therefore considered appropriate to consider doing our own research and developing our own norm groups for the two most commonly used and most useful tests.

## **Why Test?**

Personal Psychology is about each person's individuality, including individual ability, which in this case will allow an employer to reasonably predict how well the individual is likely to behave (perform in the role) in the future. To have any use the tests used must provide two kinds of information. The first is evidence of validation. Validation is the establishment of an empirical relationship between the activities people perform on the job and the test. The validation study will demonstrate that high performance on the test is related to high performance on the job. For the purposes of this paper the validation studies done by the test's proprietor were considered to be adequate.

The second, and the focus of this paper, relates to how well others who have sat the same tests have performed. The tests must provide evidence about how, relatively, individuals compare. To have any value the comparison should be against a like group. For example it would be of little value to the employer for one individual's score on the VMT3 to be compared with the scores of a group of preschool children. All that this would tell the prospective employer is how well the individual compared

with a group which has little relevance to the work the individual is about to undertake. What needs to be done is for the score of the individual to be compared with either the scores of those people who actually do the job, or at least with all of those who have in the past applied to do the job. This is described as a norm group.

### **Developing norm groups**

The dictionary definition of norm is "a standard of development or achievement." In the evaluation of applicants for a position however, the word norm usually applies to an empirically derived distribution of scores on a test. (how well an individual scored in relation to all other individuals who have sat the same test.) In other words, a formal (norming) study is conducted in order to determine how a defined group of people performs on a particular test. In the case of The Organisation the population to be normed was all applicants for positions who sat the Saville and Holdsworth VMT3 ability test. The purpose of developing our own norm group was to be able to compare applicant's individual scores against the scores of all other applicants for the positions of analyst and senior analyst and identify any possible discrimination which was racially based in these tests as they are the most commonly used in the recruitment process as discussed above.

The importance of the development of our own norm is twofold. Firstly it is likely to reflect a more relevant set of results than a general composite group as it is likely to more closely reflect our own applicant group, possibly self evident in that it is in fact our own applicant group. This means that they will provide a relevant comparison group for the person being tested. The second which relates to the first, but has a different focus is that because it has a high and identifiable Māori population, we can do some analysis around its impact, or not, on Māori in terms of discrimination.

### **How does a norm group get formed?**

Ideally there should be at least 100 scores or more. In our case it was possible to proceed because we had the requisite number of more than one hundred (152 in total) The process of norming however is somewhat more complex. First we needed to calculate the mean of the group, then we needed to calculate the standard deviation, then put into a programme which linked the raw scores to positions on the standard deviation curve.

### **What did we do?**

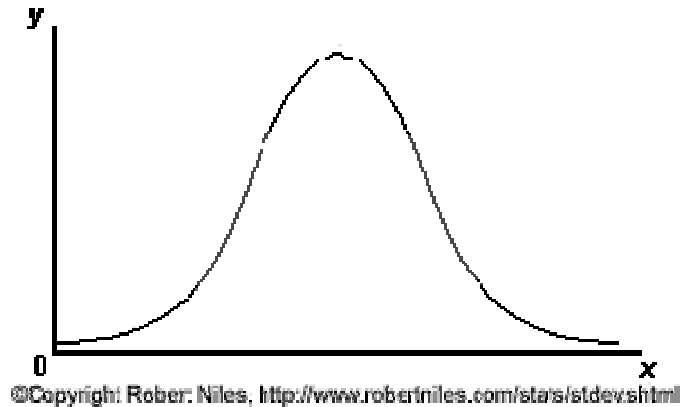
This was relatively straight forward. Take the results of the two aptitude tests used in the selection process by the employer where ethnic data had been collected. As there were over 100 samples available for each test, build a norm set for both tests. The fact that we are able to say that a significant proportion of those forming the groups had self identified as Māori was an additional bonus

### **What is a norm group?**

Norms are a way of converting raw scores into some form of relative measure. They are based on the assumption that in a given population (in the case of The Organisation all applicants for analysts and senior analyst positions) the results of any tests will be distributed in what statisticians call normal distribution of data.

A normal distribution of data means that most of the examples in a set of data are close to the "average," while relatively few examples tend to one extreme or the other. In the case of aptitude tests the number that people get right will tend to cluster around the mid-point and all of the results will be normally distributed. That is, for most people, the number they get right will be close to the mean, while fewer people score a lot higher or a lot lower than the mean.

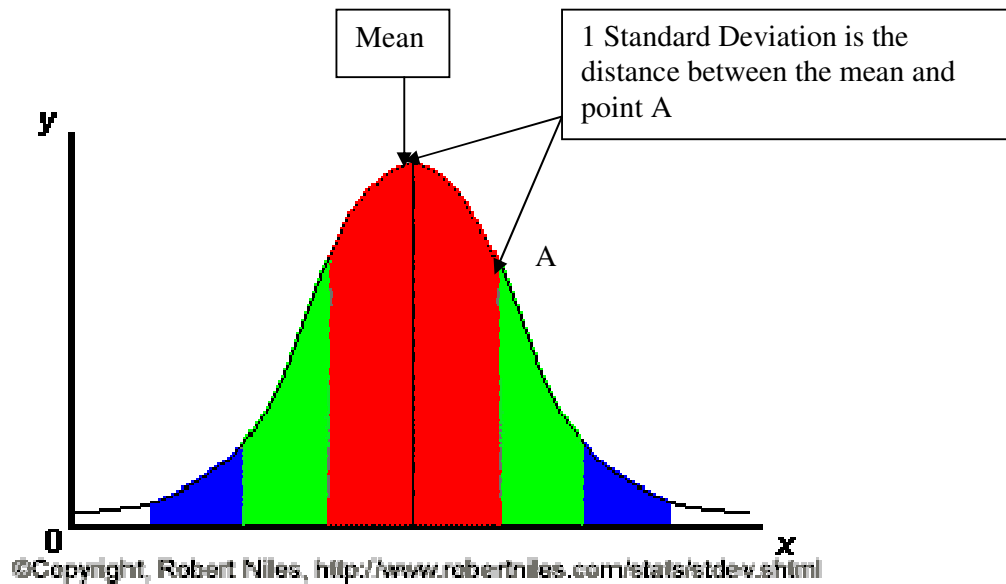
If you looked at normally distributed data on a graph, it would look something like this:



The x-axis (the horizontal one) is the value in question, the number each individual gets right. And the y-axis (the vertical one) is the number of data-points for each value on the x-axis, in other words, the number of people who got that number of answers right. Not all sets of data will have graphs that look this perfect. Some will have relatively flat curves, others will be pretty steep. Sometimes the mean will lean a little bit to one side or the other. But all normally distributed data will have something like this same "bell curve" shape.

Another important concept in the use of norms is that of standard deviation which is the average departure from the mean. The standard deviation is a statistical term that tells you how tightly all the various examples are clustered around the mean in a set of data. When the examples are tightly bunched together and the bell-shaped curve is steep, the standard deviation is small. When the examples are spread apart and the bell curve is relatively flat, that indicates that you have a relatively large standard deviation.

Computing the value of a standard deviation is complicated but graphically looks like the following.



One standard deviation away from the mean in either direction on the horizontal axis (the red area on the above graph) accounts for 68.26 percent of the people in this group. Two standard deviations away from the mean (the red and green areas) account for 95.44 percent of the people. And three standard deviations (the red, green and blue areas) account for 99.72 percent of the people.

If this curve were flatter and more spread out, the standard deviation would have to be larger in order to account for those 68.26 percent of the people. This tells us how spread out the examples in a set are from the mean.

### **What did we find?**

The norm tables (see note at start of Appendix 2) for the tests are appended to this document as Appendix 2 for the VMT3

### **VMT3**

The differences are;

	<b>NZ Composite Group</b>	<b>Public Sector Group</b>	<b>the Organisation General</b>	<b>the Organisation Non-Māori</b>	<b>the Organisation Māori</b>
Mean	19.84	19.68	19.70	21.63	19.25
Std Deviation	5.11	6.04	6.03	4.97	6.25
Numbers	774	328	152	63	69

These results suggest that The Organisation could be confident that there are no meaningful difference between the Organisation Maori means and Standard Deviations and the means and SD's for the public sector or even the New Zealand composite group norms with the following provisos. The first is that a “top down” approach is not adopted as a rigid “cut off point” and that the results of the test are used in conjunction with some other exercise which also measures the same competencies. In the case of Te Puni Kokiri the test results are used in conjunction with a written exercise in which the applicant is required to analyse a document and write recommendation on possible options.

We still however need to be mindful that the Organisation non-Māori group still has a higher mean score (approx 2 raw scores 21.63 against 19.25), the norm groups themselves are appropriate for benchmarking all (including Māori) applicants against.

### **Implications for The Organisation**

The Organisation has an imperative to recruit high performing staff to ensure that it is able to fulfil its role of making the external environment such that Māori are able to succeed as Maori. The most reliable predictor of both future job performance and the ability to learn new information is an individual’s cognitive ability. There are a variety of ways which cognitive ability can be measured. Some of these ways may disadvantage Māori and some may even discriminate on the basis of ethnicity.

Logically it would be unwise to cease using the most effective tool for predicting on the job performance, and likewise it would be unwise to continue to use tools which may disadvantage, or discriminate against, our main stakeholder group.

the Organisation could take a lead on looking at ways in which the disadvantage, if it in fact exists, can be mitigated against so that Māori are not disadvantaged in both their applications to work here, or in any other organisation which uses cognitive ability testing as part of their selection process, given that this type of testing is becoming more prevalent because of its predictive ability. A recent study of 100 randomly selected organisations and 30 recruitment firms showed that almost one half of them used cognitive tests for selecting managerial personnel. (Taylor, Keetly, McDonnell 2002) which is an increase of fifty percent in the last ten years.

### **Conclusion**

Overall it would be reasonable to conclude that the use of the VMT3 (Verbal analysis test) is appropriate in our recruitment process it has a high level of both face and content validity (it looks like it is measuring a key component of our jobs and does), and measures GMA and is therefore the single most valid predictor of future job performance available to us.

## Bibliography

- Debra Negus, (2001) *Equal Opportunities and Validity Research*. Unpublished
- Frank L. Schmidt, John E. Hunter (1998) *The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings*. *Psychological Bulletin* 124 (No2) 262 – 274.
- Frank L. Schmidt, John E. Hunter (2004) *General Mental Ability in the World of Work: Occupational Attainment and Job Performance*. *Journal of Personality and Social Psychology* Vol. 86 No1 162 - 173
- Ivan T. Robertson, Helen Baron, Patrick Gibbons, Rab McIver, Gill Nyfield (2004) *Integrity and Managers*, Unpublished
- Jenson, A. (1980). *Bias in Mental Testing*. New York: Free Press.
- Kathryn D Martin-Gray (2002). *Pre-Employment Screening Tools: Literature Review Of Testing Issues, Benefit, Legal Issues And Bias Concerns*. Unpublished research paper.
- Nigel Guenole, Paul Englert, Paul J. Taylor (2003) *Ethnic Group Differences in Cognitive Ability Test Scores within a New Zealand Applicant Sample*. *New Zealand Journal of Psychology* Vol. 32 No 1, 49 – 54.
- Phillip L. Roth, Philip Bobko, Fred S. Switzer III, Michelle A. Dean. (2001) *Prior Selection Causes Biased Estimates of Standardised Ethnic Group Differences: Simulation and Analysis*. *Personnel Psychology* 54, 591 – 616.
- Pulakos, E, & Schmitt, N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance*, 9, 241-258.
- Schmidt, F., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schmitt, N., Clause, C., & Pulakos, E. (1996). Subgroup differences associates with different measures of some common job relevant constructs. In C. L. Cooper & I. T. Robertson (Eds.), *International Review of Industrial and Organizational Psychology* (pp. 115-139). New York, Wiley
- Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology*, 82, 719-730.

## Senior / Analysts Competencies

### *Manages Time*

Is aware of the work of others which may impact on your priorities, and prepares in advance where possible; willingly reprioritises work to meet new or additional demands on time; can be relied on to meet deadlines, putting in extra effort to accommodate changes; clearly defines what is required by asking questions; realistically identifies where new work will have an impact on other deadlines and renegotiates changes in advance.

### *Networking*

Develops a comprehensive set of networks and uses these to gather and distribute information; brokers relationships between agencies or other units; establishes a positive working relationship with the Minister and Associate Minister; develops and uses personal contacts to build knowledge base and challenge thinking; builds knowledge of who key community or industry contacts are; manages relationships with public in a positive way.

### *Influencing*

Represents the Organisation's case through presenting logical factual arguments in a way that influences the decision makers; develops strategies to change people's positions and see the Organisation point of view; builds an understanding of audience and what drives their position or stance; presents information in a way that will be commonly understood; recognises where the other party is not likely to change their stance and agrees to compromise or disagree to progress an outcome; identifies the people who can influence, and provides them with arguments and strategies.

### *Teamwork*

Actively works within teams internally and externally; sees what needs to be done and fills in gaps without having to be asked; takes a leadership role to push the project along if there is a need; will say if there are problems or issues to deal with; knows when it is appropriate to have some fun with colleagues; recognises when it is more appropriate for other team members to take a lead role.

### *Analysing*

Analyses data, identifying trends and developing creative solutions for Maori development; seeks empirical evidence to substantiate arguments; develops and tests alternative courses of action; thinks laterally to assess the implications for Maori and the Organisation's position on issues; interprets policy, procedures and legislation and applies principles accurately.

### ***Meets Challenges***

Has a clear sense of where their career is headed; takes responsibility for their own development; identifies challenges as an opportunity to learn; is willing to try new things; puts in extra effort when needed; takes ownership of issues, getting on and doing the unpleasant things that need to be done.

### ***Judgement***

Sees how issues fit into the bigger picture; makes recommendations which take into account the political context; ensures that all the implications and risks for the Maori and the Organisation have been covered off; use others to assess the implications of recommendations, obtaining signoff from appropriate managers before committing it to action.

### ***Written Communication***

Writes clearly and succinctly, only includes relevant information; written documents are short and precise; identifies main messages and sticks to these; tailors style to purpose and audience; proofs own work before seeking others input; accepts and incorporates the input of others.

### The Organisations VMT3 2004

Grade	%ile	VMT3		Sten
			T-Score	
A	99	35	75	10
	99	34	74	
	99		73	
	99	33	72	
	98		71	
	98	32	70	
	97	31	69	9
	96		68	
	96	30	67	
	95		66	
	93	29	65	
	92	28	64	
B	90		63	8
	88	27	62	
	86		61	
	84	26	60	
	82	25	59	7
	79		58	
	76	24	57	
	73		56	
C	69	23	55	6
	66	22	54	
	62		53	
	58	21	52	
	54		51	
	50	20	50	
	46	19	49	5
	42		48	
	38	18	47	
	34	17	46	
31		45	4	
27	16	44		
24		43		
21	15	42		
18	14	41		
16		40		
14	13	39		
12		38		
10	12	37	3	
8	11	36		
E	7		35	2
	5	10	34	
	4		33	
	4	9	32	
	3	8	31	
	2		30	
	2	7	29	1
	1		28	
	1	6	27	
	1	5	26	
	1		25	
	1	0 - 4		

**VMT3**      Number of Cases: 152 Mean: 19.70 SD : 6.03

**Ethnic composition:** Approx 45.5% Maori, 41.5% NZ European, 13% Other

**Gender:** 46% Female, 41% Male, 13% Not known

**Applicants:** Mainly Policy Analysts, Senior Policy Analysts or 1<sup>st</sup> Level Managers