Examining the Psychometric Properties of The McQuaig Occupational Test®

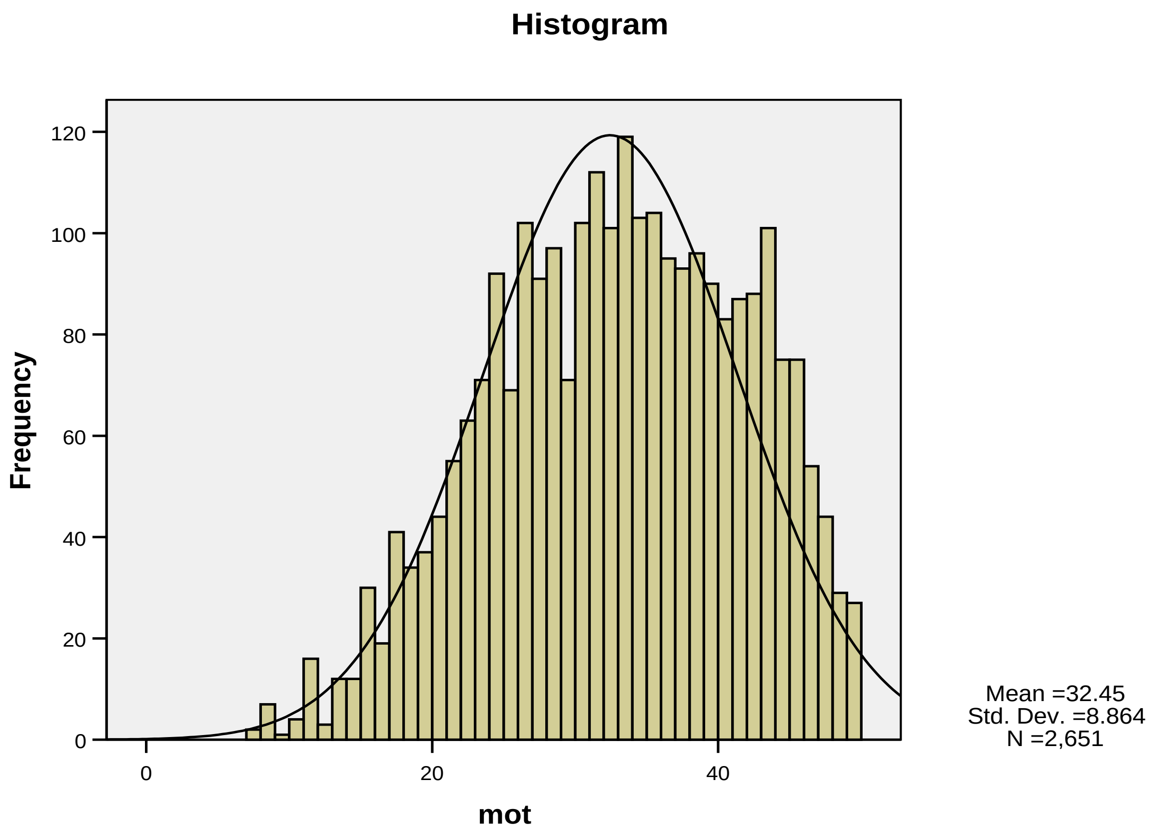
Prepared for: The McQuaig Institute of Executive Development Ltd., Toronto, Canada

Prepared by: Henryk Krajewski, Ph.D., Senior Consultant, MICA Consulting Partners

June 29, 2006

1. What are the percentile norms for The McQuaig Occupational Test® – The MOT (overall, male and female, country)

Using the available data the ages of the participant group range from 18-51 with an average of 30.98 years old (N=545; SD=7.9 years). The overall scores for The MOT are fairly normally distributed. This is consistent with what we know about intelligence test scores, and allows us to analyze the data using traditional statistical methods. As you can see, most people score near the average and less people get high and low scores. This is necessary to create "norms" (i.e., there has to be a middle where most people score). When the data is "non-normal," special statistics are required and percentile norms are difficult to interpret. You have no such problems with The MOT data.



a. Overall Sample Norms (Australian and UK data; N=2651):

MOT Score	Percentile Rank
0	0
21	10
25	20
28	30
30	40
32	50
35	60
37	70
40	80
44	90
47	95
49	97

b. Males (N=1934)

MOT Score	Percentile Rank
0	0
21	10
25	20
28	30
30	40
33	50
35	60
37	70
40	80
44	90
47	95
49	97

c. Females (N=716)

MOT Score	Percentile Rank
0	0
20	10
24	20
27	30
29	40
32	50
34	60
36	70
39	80
43	90
46	95
48	97-99

d. Norms by Country – United Kingdom – Overall (N=548)

MOT Score	Percentile Rank
0	0
25	10
29	20
32	30
34	40
36	50
38	60
40	70
42	80
46	90
49	95
50	97-99

e. Norms by country – Australia – Overall (N=2103)

MOT Score	Percentile Rank
0	0
20	10
24	20
27	30
29	40
32	50
34	60
36	70
39	80
43	90
46	95
48	97-99

Overall Norms for males and females differ only slightly and are very comparable. There was only a 1-point difference on overall MOT scores between males and females. This difference is statistically significant in large part due to the size of the sample. Practically speaking, the difference is insignificant. Therefore, for ease of use, overall norms should be used (by country – see below), and not by country and gender.

Overall means between the Australian and UK data are significantly different from one another, and are quite different in a practical sense (the mean scores are 4 points different for the Australian vs. UK participants: M=31.60 vs. 35.69, respectively). This gives rise to real differences in norms. It is suggested, therefore, that country specific norms be used, as is consistent with norms used in other intelligence tests (e.g., Watson Glaser).

2. Item Difficulty – What items do participants attempt? Is the test consistent in terms of the difficulty of the items?

Each item on the first 6/10ths of The MOT is attempted 45% of the time. This means approximately half the people attempt each question and half leave it blank. For example, half the people attempt Question 1, and half leave it blank. This is common for a speeded test and indicates that most of the items in the first 6/10ths are relatively equal in difficulty (i.e., the same number of people tend to skip the question as answer it). Because The MOT is speeded, less people attempt questions in the last 4/10ths (approximately 30%).

Most of the items on The MOT have a "base rate" of approximately 70%. This means that, of the people who attempt this question, 70% get it right. The perfect base rate is 50% on a non-speeded test. In other words, the most diagnostic questions – those that are maximally able to distinguish between candidates on non-speeded tests – are those that half the people get right, and half the people get wrong. Look at it this way, if everyone gets a question right (i.e., a base rate of 90%) or everyone gets a question wrong (i.e., a base rate of 10%), those questions are not going to allow you to distinguish between candidates.

However, because The MOT is speeded, it puts pressure on the participants to use both their intellectual *power* and *speed.* Speed is defined as the capacity to quickly work through complex information. Most people could get most of The MOT questions correct given unlimited time – which would cause difficulty in distinguishing the "smart" candidates from the "less smart" ones (because there would be little differentiation in scores). However, The MOT's time limit brings in *speed* capacity, inherently causing differentiation in overall scores. Those who score high vs. low are primarily different in speed capacity – though the average 70-75% base rate suggests there is some *power* capacity inherent in the test. It would not be uncommon for a "pure" speed test to have base rates in the 90-95% range.

The above has important implications for how MOT test scores are interpreted: The MOT is primarily (though not strictly) a speeded test. In practical terms, you could say that The MOT is a measure of speed and accuracy in intelligence. On the plus side, there is evidence that speeded tests are less susceptible to bias which would be of particular interest to HR practitioners who are concerned with racial or cultural bias.

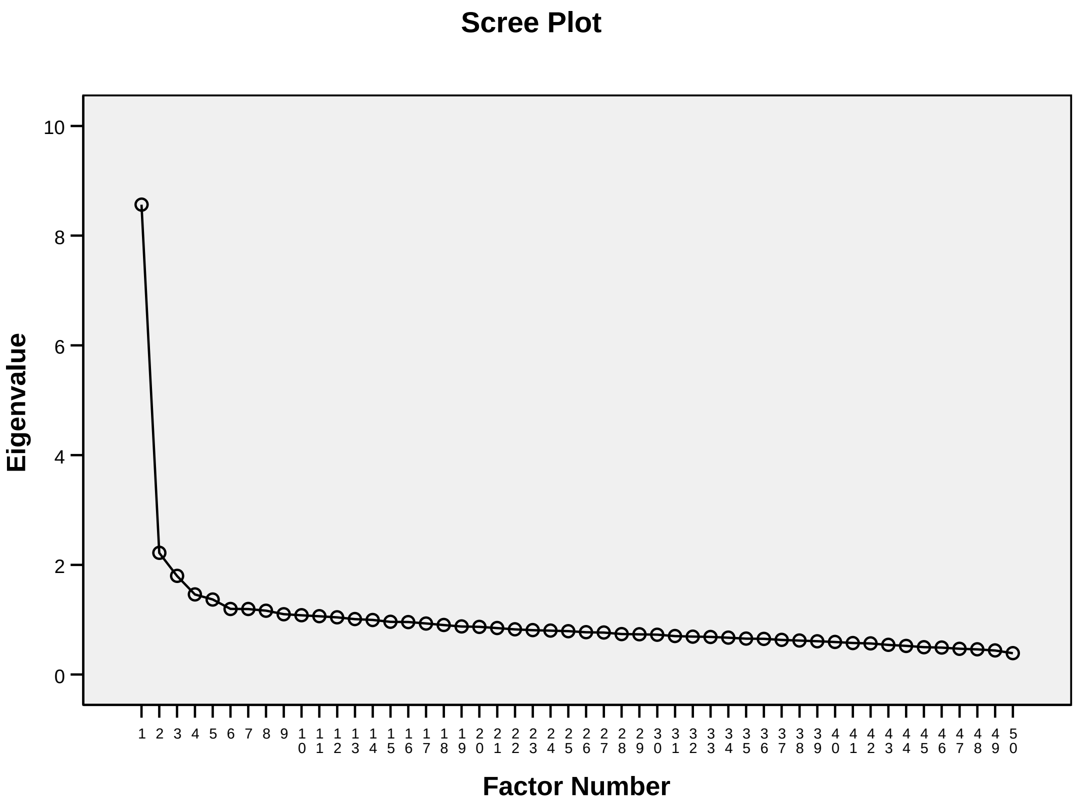
In terms of items that are either extremely difficult or easy: Items 28 and 50 are clearly the most "difficult." Of the people who attempt these items only about 35% get them correct. There are a variety of items with base rates over 95% (95% of the people who attempt these items, get these items correct). However, this is not necessarily a problem if you are comfortable with the notion of a *speed* test.

3. Factor Analyses – Does The MOT measure one factor or dimension (i.e., general intelligence) or does it measure specific facets of intelligence (i.e., verbal, analytic, mathematics)?

The results of a principle axis factor analysis on the data reveal that The MOT has one large, underlying factor that accounts for approximately 18% of the total variance in the test – 18 times as much variance accounted for by any other extracted factor. As can be seen from the diagram below, a scree plot also suggests that there is one general intelligence factor being measured in the data.

Because there are problems associated with factor analyzing categorical data (i.e., data scored as "right/wrong"), and particularly on timed tests (i.e., items could cluster together because of where they fall on the test vs. because they tap into a similar type of intelligence), a second-order factor analysis was conducted. This involved first, rationally deriving scales for the math, verbal, and reasoning questions of The MOT, summing respective items together to create scale scores for each dimension, and then treating the scale scores as continuous "items" (i.e., there could be a range of possible scores beyond just 1=right, 0=wrong). Finally, the "items" were factor analyzed to see if they would all load on a single intelligence factor, or whether they were better considered to represent separate dimensions. Results showed that a single factor accounted for 63% of the variance in the "items" as created through summed scale scores. Consistent with a scree plot that clearly indicated the presence of a single factor, these results suggest The MOT is indeed tapping into a single intelligence factor.

This common intelligence factor, commonly known as "*g*", has been found to underlie many popular intelligence tests including the Stanford-Binet, Weschler Adult Intelligence Scale, etc. Although there are clearly different "types" of intelligence questions found on The MOT, the suggestion that The MOT may measure verbal, critical reasoning, and mathematical ability as separate, but related, factors of intelligence, is not supported by the data. The MOT should therefore be used as a measure of work-related general intelligence which provides one total score vs. a series of subscale scores (i.e., verbal, reasoning, math).


Scree Plot
Eigenvalue
Factor Number
0
2
4
6
8
10

4. Reliability – Does the test consistently measure intelligence and do the items measure the same thing?

    a. Overall MOT

The overall MOT internal consistency reliability (Cronbach's Alpha) is .71, based on a sample of 237 individuals with complete item data (a requirement for reliability analyses). This value meets Nunnally's (1970) .70 criterion for internal consistency reliability. The value suggests that items are reliable, and are consistently measuring the same thing.

Reliability Statistics

Cronbach's Alpha	N of Items
.708	50

    b. Verbal items

The internal consistency reliability of the verbal subscale is .60 based on 468 observations. This level would be termed a low, but acceptable estimate.

Reliability Statistics

Cronbach's Alpha	N of Items
.603	18

    c. Reasoning items

The internal consistency reliability of the verbal subscale is .52 based on 456 observations. A value of this magnitude would suggest not all the items are consistently measuring the same thing.

Reliability Statistics

Cronbach's Alpha	N of Items
.520	18

d. Math subscale

The internal consistency reliability of the verbal subscale is .51 based on 303 observations. A value of this magnitude would suggest not all the items are consistently measuring the same thing.

Reliability Statistics

Cronbach's Alpha	N of Items
.511	14

The low reliabilities for the individual subscales are a likely product of two factors, (a) reduced scale length (shorter scales have lower reliability) and (b) the nature of the abilities measured by the items. Scales with shorter item lengths naturally have lower internal consistency reliability. As referenced in the factor analysis section above, it is unlikely that separate subscales can be reliably measured in The MOT, so it is not surprising that reliabilities for the respective verbal, reasoning, and math items is low.

Although reporting the above is industry standard, there are issues with interpreting Cronbach's alpha reliability with categorical data on speeded tests. Therefore, we recommend a Test-Retest reliability study. Test-Retest reliability estimates would give a better indication of whether The MOT is consistently measuring intelligence. Another recommendation would be to correlate separately timed sections of The MOT as a better, more accurate, measure of internal consistency reliability.

Overall, the reliability analyses suggest it is most favorable to treat The MOT as a single test with no subscales. Interpretation of the subscales becomes difficult due to the fact that the subscales do not appear to measure distinct factors in a consistent fashion.

5. Relationships between MOT and WS:

   a. Are there any significant relationships between The MOT and real scores on The McQuaig Word Survey®?

The most straightforward way of assessing relations between variables is to interpret the correlation coefficient. The table below shows Pearson Product Moment correlations for the entire sample. The corresponding sample sizes appear in each cell.

Correlations

		Mot	rd	rs	rr	rc
Mot	Pearson Correlation	1	.115(**)	.078(**)	-.036	-.209(**)
	Sig. (2-tailed)		.000	.000	.070	.000
	N	2651	2594	2594	2594	2594
Rd	Pearson Correlation	.115(**)	1	-.134(**)	-.601(**)	-.661(**)
	Sig. (2-tailed)	.000		.000	.000	.000
	N	2594	2595	2595	2595	2595
Rs	Pearson Correlation	.078(**)	-.134(**)	1	-.484(**)	-.432(**)
	Sig. (2-tailed)	.000	.000		.000	.000
	N	2594	2595	2595	2595	2595
Rr	Pearson Correlation	-.036	-.601(**)	-.484(**)	1	.394(**)
	Sig. (2-tailed)	.070	.000	.000		.000
	N	2594	2595	2595	2595	2595
Rc	Pearson Correlation	-.209(**)	-.661(**)	-.432(**)	.394(**)	1
	Sig. (2-tailed)	.000	.000	.000	.000	
	N	2594	2595	2595	2595	2595

** Correlation is significant at the 0.01 level (2-tailed).

As can be seen in the table above, some of the correlations – like The MOT/Dominance correlation – appear significant (see **). However, practically speaking, The MOT/Dominance correlation (r = .11) is very small. More interesting is the correlation between The MOT and Compliance (r = -.21). This number suggests that higher scores on The MOT are associated with greater Independence. Though the correlation is moderate (.21), it is still *somewhat* meaningful. As research has shown, correlations of this magnitude can result in very high levels of utility (i.e., return on investment).

Descriptively, we could say that people who are smarter tend to describe themselves more as risk takers and individualists, or perhaps, those who are risk takers and individualists tend to be smarter. The correlation does not dictate which causes which. The sample size is too large for this to be a “random” finding. There is likely some degree of real relation here.

b. Is there a profile of WS scores that predicts higher MOT scores?

In order to assess this, we can use a statistical method called Discriminant Function Analysis. This is a “classification” program that optimally “weights” The WS scores in such a way as to optimally predict scores on The MOT.

According to the results, we can correctly classify individuals into MOT quartiles (bottom quarter, mid lower quarter, mid upper quarter, upper quarter) 31% of the time based on weighting Independence very heavily, Dominance very heavily, Sociability moderately and Driving moderately. This weighting would suggest that a “Hi Dominance, Low Compliance” type of profile would predict MOT performance significantly. The same “Hi Dominance, Low Compliance” profile predicts upper and lower half MOT scores at a 60% success rate.

The bottom line: You can predict who would fall into the top and bottom halves of performers on The MOT using a “Hi Dominance, Low Compliance” profile with 61% accuracy. You can predict who would fall into quartile groups on The MOT using a “Hi Dominance, Low Compliance” profile with 31% accuracy. Because you are trying to be more specific in terms of identifying quartile groups, it gets harder to predict with accuracy, given the moderate relationship between this profile and intelligence.

Structure Matrix

	Function
	1
Rc	-.971
Rd	.665
Rs	.212
Rr	-.197

Above is the actual weighting of The WS scores that would predict MOT scores.

The “industry standard” for using this weighting to predict outcomes is approximately 70-80% accuracy. Thus, it is not recommended to use The WS to predict MOT scores; however, there is some evidence that individuals with this profile will be smarter, on average, than people without this profile.