

# It's Time to Innovate

**Paul Barrett**

*email:* [paul.barrett@auckland.ac.nz](mailto:paul.barrett@auckland.ac.nz)

<http://www.liv.ac.uk/~pbarrett/paulhome.htm>

*Affiliations:* Universities of Auckland and Canterbury  
Departments of Psychology  
Dept. of Clinical Psychology, Univ. of Liverpool

# Why?

(Sternberg & Williams, 1998; p. 577)

"No technology of which we are aware -- computers, telecommunications, televisions, and so on -- has shown the kind of **ideational stagnation** that has characterized the testing industry. Why? Because in other industries, those who do not innovate do not survive...

# Why?

“In the testing industry, the opposite appears to be the case. Like Rocky I, Rocky II, Rocky III, and so on, the testing industry provides minor cosmetic successive variants of the same product where only the numbers after the names substantially change. These variants survive because psychologists buy the tests and then loyally defend them.”

# Why?

Michell, J. (1997), p. 406.

“No critic has explained why psychology, alone amongst the sciences, is entitled to its own definition of measurement....

Readers of this journal have been given no adequate reason, yet, to avoid the conclusion that methodological thought disorder is systemic in modern psychology.”

# Why?

Michell, J. (2000), p. 639.

“It is concluded that psychometrics is a pathology of science”

Michell, J. (2001), p. 211.

“the way in which psychometrics is currently, typically taught actually subverts the scientific method ”

# Measurement

*Michell (2001), p. 212 ...*

“Measurement, as a scientific method, is a way of finding out (more or less reliably) what level of an attribute is possessed by the object or objects under investigation.”

# Measurement

“However, because measurement is the assessment of a level of an attribute via its numerical relation (a ratio) to another level of the same attribute (the unit selected), and because only **quantitative attributes** sustain numerical ratios of this sort, measurement applies only to quantitative attributes”.

# Measurement

“Psychometrics concerns the measurement of psychological attributes using the range of procedures collectively known as psychological tests. As a precondition of psychometric measurement, these attributes must be quantitative”.



# Measurement

Measurement within a quantitative science is thus defined as the identification of a **magnitude** of a **quantitative variable** relative to some standard unit magnitude of that variable.

# Measurement

Joel Michell (1999)

“Because measurement involves a commitment to the existence of quantitative attributes, quantification entails an empirical issue: **is the attribute involved really quantitative or not?** If it is, then quantification can sensibly proceed. If it is not, then attempts at quantification are misguided. A science that aspires to be quantitative will ignore this fact at its peril”

# Measurement

Joel Michell (1999) ... p.75

“It is pointless to invest energies and resources in the enterprise of quantification if the attribute involved is not really quantitative. The logically prior task in this enterprise is that of addressing this empirical issue. I call it the *scientific task of quantification*.”

# Quantitative Structure

**From Michell (1990, p.52):**

Let  $X$ ,  $Y$ , and  $Z$  be any three values of a variable  $Q$ .  
Then  $Q$  is **ordinal** if and only if:

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

A relation possessing these three properties is called a simple order, so  $Q$  is ordinal if and only if  $\geq$  is a simple order on all its values.

# Quantitative Structure

All quantitative variables are simply ordered by  $\geq$ , but not every ordinal variable is quantitative, **for quantity involves more than order**, it involves **additivity**.

**Additivity** involves a ternary relation, symbolized as “ $X+Y=Z$ ”. Let  $Q$  be any ordinal variable such that for any of its values  $X$ ,  $Y$ , and  $Z$  ...

4.  $X+(Y+Z) = (X+Y)+Z$  (**associativity**)

5.  $X+Y = Y+X$  (**commutativity**)

6.  $X \geq Y$  if and only if  $X+Z \geq Y+Z$  (**monotonicity**)

7. If  $X > Y$  then there exists a value of  $Z$  such that  $X=Y+Z$  (**solvability**)

8.  $X+Y > X$  (**positivity**)

9. There exists a natural number  $n$  such that  $nX \geq Y$

(where  $1X = X$  and  $(n+1)X = nX + X$ ) (**Archimedean condition**)

## The questions I asked myself

So why am I continuing to treat psychometric test theory as critically important? From a quantitative measurement perspective it seems to be somewhat irrelevant to the axioms of quantity that are used within every other quantitative science. Could it be that psychology might actually be a **non-quantitative science** – with all that this term implies?

## The questions I asked myself

In which areas have the substantive innovations in major psychological tests and assessment methodologies taken place since the late 1950s?

Is it mainly in the technology of questionnaire test delivery allied to new measurement models for questionnaire data

*-or-*

the innovation of completely new kinds of psychological construct assessment?



## The questions I asked myself

Was Blinkhorn (1997) correct in his recent review of 50 years of Test Theory ...?

“Contemporary test theory, with its emphasis on statistical rather than psychological models, has become inaccessible to the majority of test users, and predominantly reflects educational rather than psychological concerns. **Real progress may depend on the emergence of a new and radical reconceptualization**”.



Was Paul Kline (1998) also wrong for concluding:

p. 196 ... The New Psychometrics...

“The original purpose of psychometrics was to provide the precise quantification necessary for science. As we can see, it has failed to do this, although it has provided us with promising results. From the arguments developed in this book, it can be concluded that it is pointless to attempt to develop further psychometric tests of the traditional kind”.

# So now what?

① Think about **Psychometrics as Applied Numerics** (Barrett, 2003), quantitative science, and non-quantitative science

② Look again with clear, fresh eyes at Barrett and Paltiel (1996)... “*Can a single item replace an entire scale: POP vs the SHL OPQ 5.2 Questionnaire*”

## So now what?

③ Seriously examine why a hypothetical “True Score” was created by early psychometricians. How much thought actually went into the psychological “sense” of such a concept rather than the “statistical sense” which was required to develop the tenets of classical test theory?

# So now what?

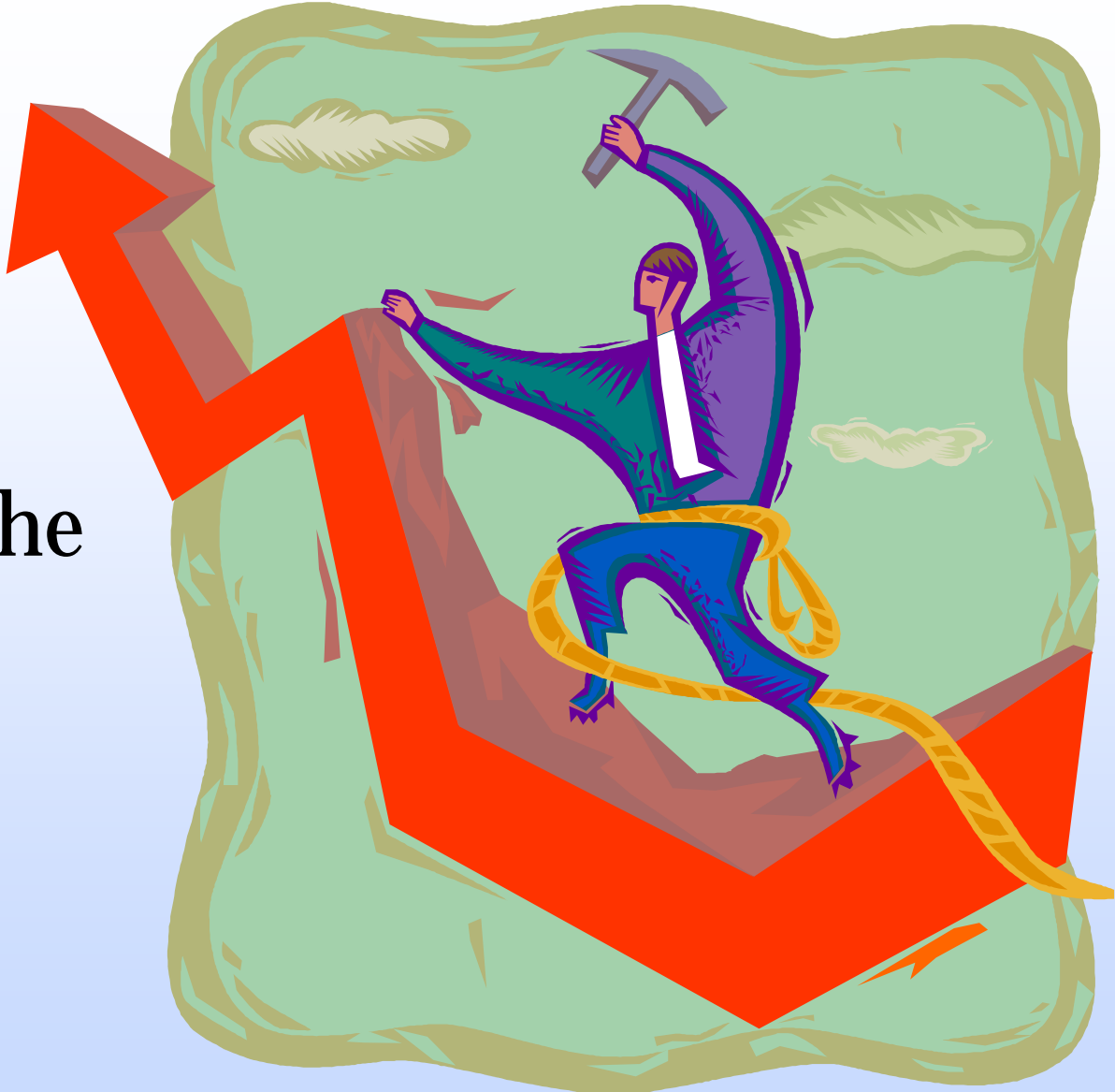
Think the unthinkable - “WYSIAYH”  
... and deal with it accordingly ...

**(What You See is All You Have)**

# So now what?

④ **BUT**, do not discard all the psychometric tests that have gone before – they possess an evidence-base for their validity and in some cases have very great utility.

So, there is the challenge ...



**What happens if you seek self-report construct magnitude measurement using a single composite questionnaire item with an 11-point response scale?**

The word "POP" is rendered in large, bold, red, three-dimensional block letters. The letters are set against a light purple rectangular background. The lighting creates a sense of depth, with shadows on the bottom and sides of the letters.

# Examples of OPQ Concept 5.2 scale items (paraphrased for copyright purposes)

N=621 Barrett and Paltiel, 1996

**T8-Innovative**

**alpha=.85**

**Mean ITC=.59**

---

- 8 I do not find it easy to generate creative **ideas**
- 39 People approach me for creative **ideas**
- 70 I find it hard to be inventive
- 101 New **ideas** come easily to me
- 132 My **ideas** are rarely innovative
- 163 I enjoy coming up with lots of valuable **ideas**
- 194 I rarely have many original **ideas**
- 225 I generally have an original approach to problems



# Examples of OPQ Concept 5.2 scale items (paraphrased for copyright purposes)

R5-Affiliative

alpha=.78

Mean ITC=.51

---

- 29 I prefer my own **company** to that of others
- 60 I get much pleasure from other people's **company**
- 91 **Companionship** is not a major concern to me
- 122 I develop close attachments to people
- 153 I rarely long for the **company** of others
- 184 I have a large number of friends
- 215 I do not like making new friendships
- 246 I get enjoyment from the **companionship** of others

# The POP questionnaire

- ◆ R1 Persuasiveness I like selling, whether ideas or products
- ◆ R2 Controlling I like organising and taking charge of people.
- ◆ R3 Independent I speak my mind even if its unpopular
- ◆ R4 Outgoing I am an outgoing and sociable person
- ◆ R5 Affiliative I enjoy being in the company of others.
- ◆ R6 Soc Confident I am at ease in social settings.
- ◆ R7 Modest I am modest about my achievements.
- ◆ R8 Democratic I like the group to participate in decision-making.
- ◆ R9 Caring I am sensitive to other people's problems.
- ◆ T1 Practical I enjoy repairing objects or devices.
- ◆ T2 Data Rational I enjoy working with numbers and statistics.
- ◆ T3 Artistic I appreciate the performing and literary arts.
- ◆ T4 Behavioural I like analysing other people's behaviour.
- ◆ T5 Traditional I am described as something of a traditionalist.
- ◆ T6 Change Orient. I am usually critical of people's ideas.

## The response scale used for the POP items

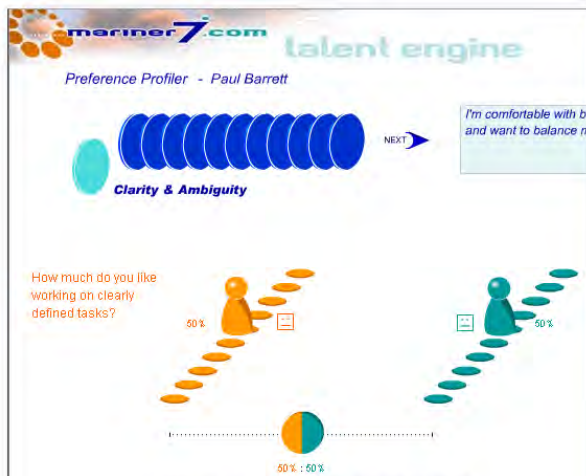
Strongly Agree	Agree	Uncertain	Disagree	Strongly Disagree
<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/>

# The POP questionnaire - Result

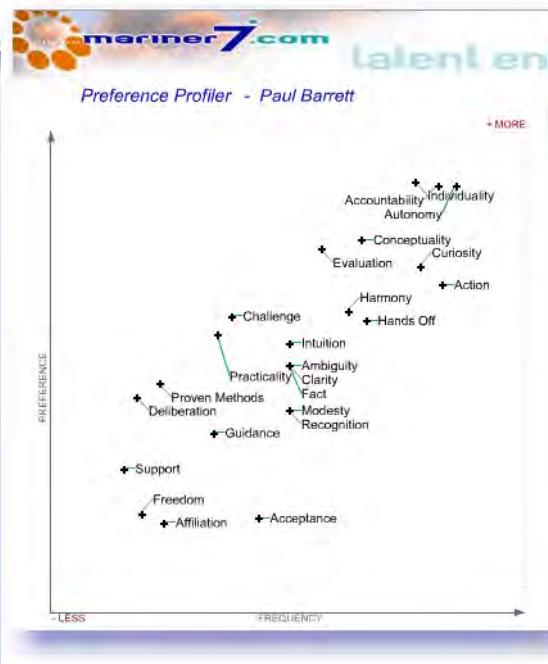
## Uncorrected Corrected

◆ R1 Persuasive	.64	.98	I like selling, whether ideas or products
◆ R2 Controlling	.73	>1.0	I like organising and taking charge of people.
◆ R3 Independent	.48	.79	I speak my mind even if its unpopular
◆ R4 Outgoing	.69	.98	I am an outgoing and sociable person
◆ R5 Affiliative	.58	.88	I enjoy being in the company of others.
◆ R6 Soc Confident	.64	.91	I am at ease in social settings.
◆ R7 Modest	.68	.99	I am modest about my achievements.
◆ R8 Democratic	.57	.93	I like the group to participate in decision-making.
◆ R9 Caring	.51	.74	I am sensitive to other people's problems.
◆ T1 Practical	.88	>1.0	I enjoy repairing objects or devices.
◆ T2 Data Rational	.84	>1.0	I enjoy working with numbers and statistics.
◆ T3 Artistic	.79	>1.0	I appreciate the performing and literary arts.
◆ T4 Behavioural	.64	.98	I like analysing other people's behaviour.
◆ T5 Traditional	.56	.85	I am described as something of a traditionalist.
◆ T6 Change Orient.	.48	.80	I am usually critical of people's ideas.
◆ T7 Conceptual	.68	>1.0	I enjoy the discussion of hypothetical ideas.
◆ T8 Innovative	.74	>1.0	I generate creative and innovative ideas.
<b>Median r</b>	<b>.64</b>	<b>.98</b>	

# Why bother with questionnaire items at all?



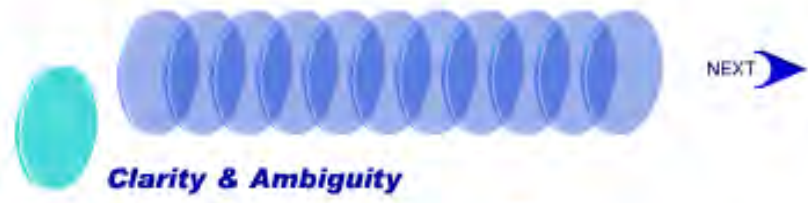
# The Graphical Profiler





## Preference Profiler - Paul Barrett

[Intro](#) [Help](#) [Finish](#)



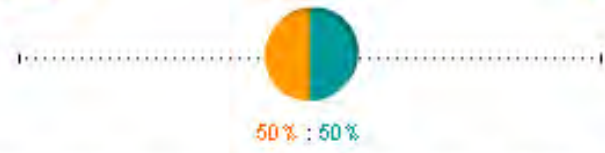
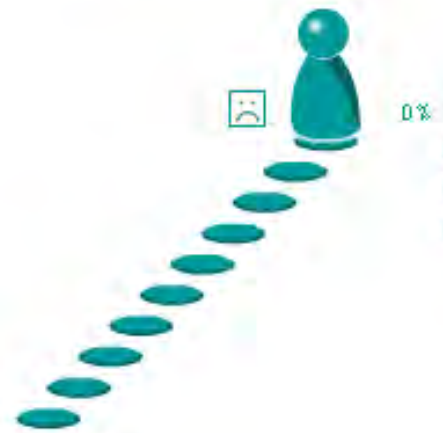
Click and drag the orange pawn  teal pawn  and click  along their axes to indicate your preferences and how you'd like to balance your time.

### Clarity & Ambiguity


How much do you like working on clearly defined tasks?



How much do you like having to make sense out of uncertainty?



How would you like to balance your time between them?

 SWITCH TO SUMMARY VIEW



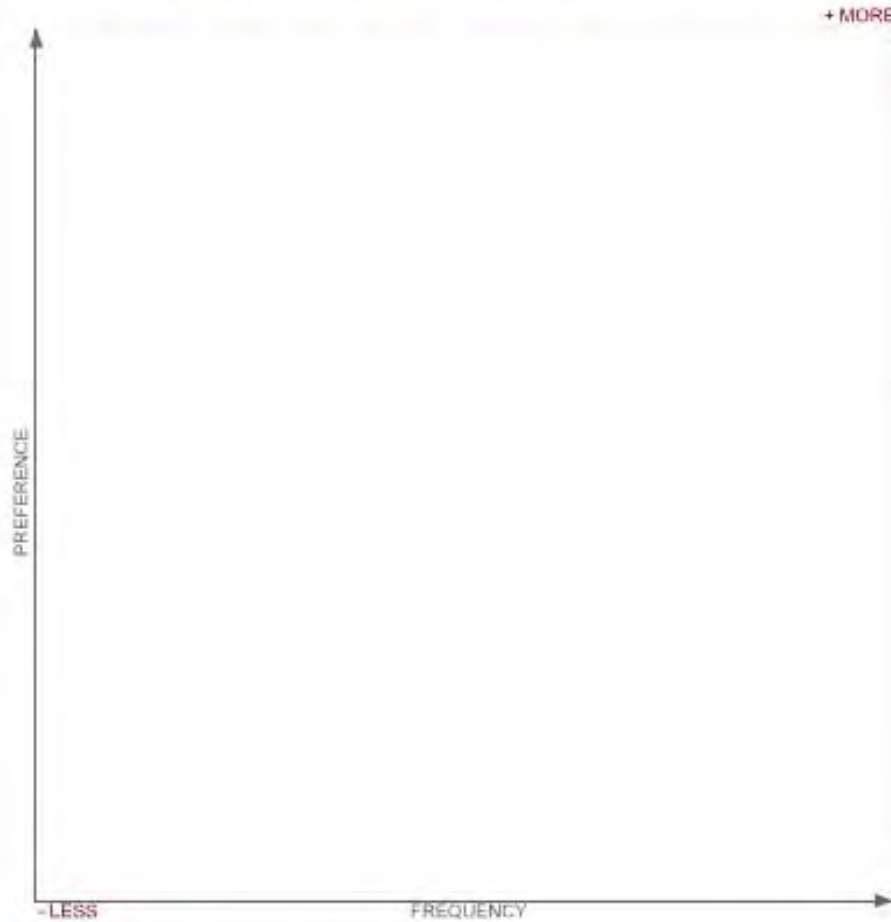


## Preference Profiler - Paul Barrett

Intro

Help

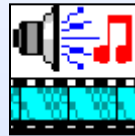
Finish



SWITCH VIEW



# The Complete Work-Preferences Profiler





# The cumulative, public, results for a user

## Preference Profile

- I like task clarity, and although I'm comfortable with uncertainty as well, I would rather spend the majority of my time working on clearly defined tasks.
- I'm comfortable with both confrontation and promoting harmony, and want to spend some time on both.
- I dislike team activities, and want to spend the majority of my time working by myself, which I'm super keen on.
- I'm put off by receiving guidance, and want to spend the majority of my time being independent, which I really like.
- I enjoy thinking conceptually, and although I'm comfortable with thinking practically as well, I would rather spend the majority of my time working with concepts.
- I'm comfortable with both going on intuition and fact based decision-making, and want to spend some time on both.
- I really dislike being accepting, and want to spend the majority of my time challenging current thinking, which I love.
- I dislike established processes, and want to spend the majority of my time exploring new ideas, which I love.
- I dislike providing guidance, and want to spend the majority of my time leaving others to work independently, which I'm keen on.
- I'm super keen on moving quickly, and although I'm okay with acting with complete information as well, I would rather spend the majority of my time prioritising outputs.
- I'm uncomfortable with both being publicly acknowledged and modesty, and want to spend some time on both.
- I love being accountable, and although I'm okay with being carefree as well, I would rather spend the majority of my time holding a lot of responsibility.

[www.staffcv.com](http://www.staffcv.com)

# Reproducibility and Stability

Two small samples of data to date:

Adult Working Volunteers ( $N=61$  3-month long term sample and 10 x 5-day individuals)

Auckland University undergraduates ( $N=25$  x 5-day short-term retest, 23 x 1-month individuals)

# Reproducibility and Stability

	5 days	1 month	3 months
<b>N</b>	35 <b>29</b>	23 <b>17</b>	61 <b>47</b>
<b>Pearson r</b>	0.65 <b>0.85</b>	0.35 <b>0.77</b>	0.53 <b>0.83</b>
<b>ICC</b> (intraclass)	0.64 <b>0.84</b>	0.34 <b>0.76</b>	0.52 <b>0.82</b>
<b>MAD</b>	11.18 <b>7.34</b>	15.0 <b>8.26</b>	13.26 <b>7.85</b>

\*MAD = Mean Absolute Deviation (0-100 range)  
Figures in **RED** are for “clipped” data

# The Personality Profiler

Using the free 5-factor personality model item-bank at the International Personality Item Pool

<http://ipip.ori.org/ipip/>

Specifically using **10** facets taken from the AB5C 45-facet personality questionnaire as “typical” personality test scales, spanning about **110 questionnaire items** in total ...

<http://ipip.ori.org/ipip/newAB5CTable.htm>

\*Barrett and Ebbeling (in preparation)

# The Personality Profiler

single item rewords? ... facet = **Talkative**

H4	Do most of the talking.
H1138	Talk too much.
H527	Speak loudly.
H535	Make myself the centre of attention.
H1150	Like to attract attention.
H536	Never stop talking.
H531	Make a lot of noise.
H769	Demand to be the centre of interest.
H548	Speak softly.
X212	Dislike talking about myself.

# The Personality Profiler

single item rewords? ... facet = **Talkative**

## **The Rating Statement:**

I have no problem in talking about almost anything. In fact, I find it hard to stop sometimes, especially if I've become the centre of attention! Frankly, I just like talking with people.

# The Personality Profiler



The Personality Profiler	AB5C Scales	Profiler Scores
Friendliness	35.3	<b>39.3</b>
Leadership	34.1	<b>36.7</b>
Talkativeness	27.1	<b>33.3</b>
Efficiency	37.7	<b>41.5</b>
Purposefulness	41.5	<b>44.4</b>
Organization	<b>46.7</b>	42.5
Orderliness	32.9	32.5
Calmness	<b>36.4</b>	33.5
Impulse-Control	<b>37.2</b>	33.7
Happiness	35.7	<b>39.3</b>

Profiler  
-VS-  
questionnaire  
scale-score

**Means**

N=99 cases



The Personality Profiler	AB5C Scales	Profiler Scores
Friendliness	6.8	<b>8.1</b>
Leadership	6.9	<b>8.8</b>
Talkativeness	7.3	<b>9.9</b>
Efficiency	7.7	<b>9.6</b>
Purposefulness	7.7	<b>9.6</b>
Organization	7.3	<b>10.8</b>
Orderliness	8.3	<b>10.5</b>
Calmness	6.1	<b>9.3</b>
Impulse-Control	6.9	<b>11.3</b>
Happiness	6.9	6.7

Profiler  
-VS-  
questionnaire  
scale-score

**Std. Devns.**

N=99 cases

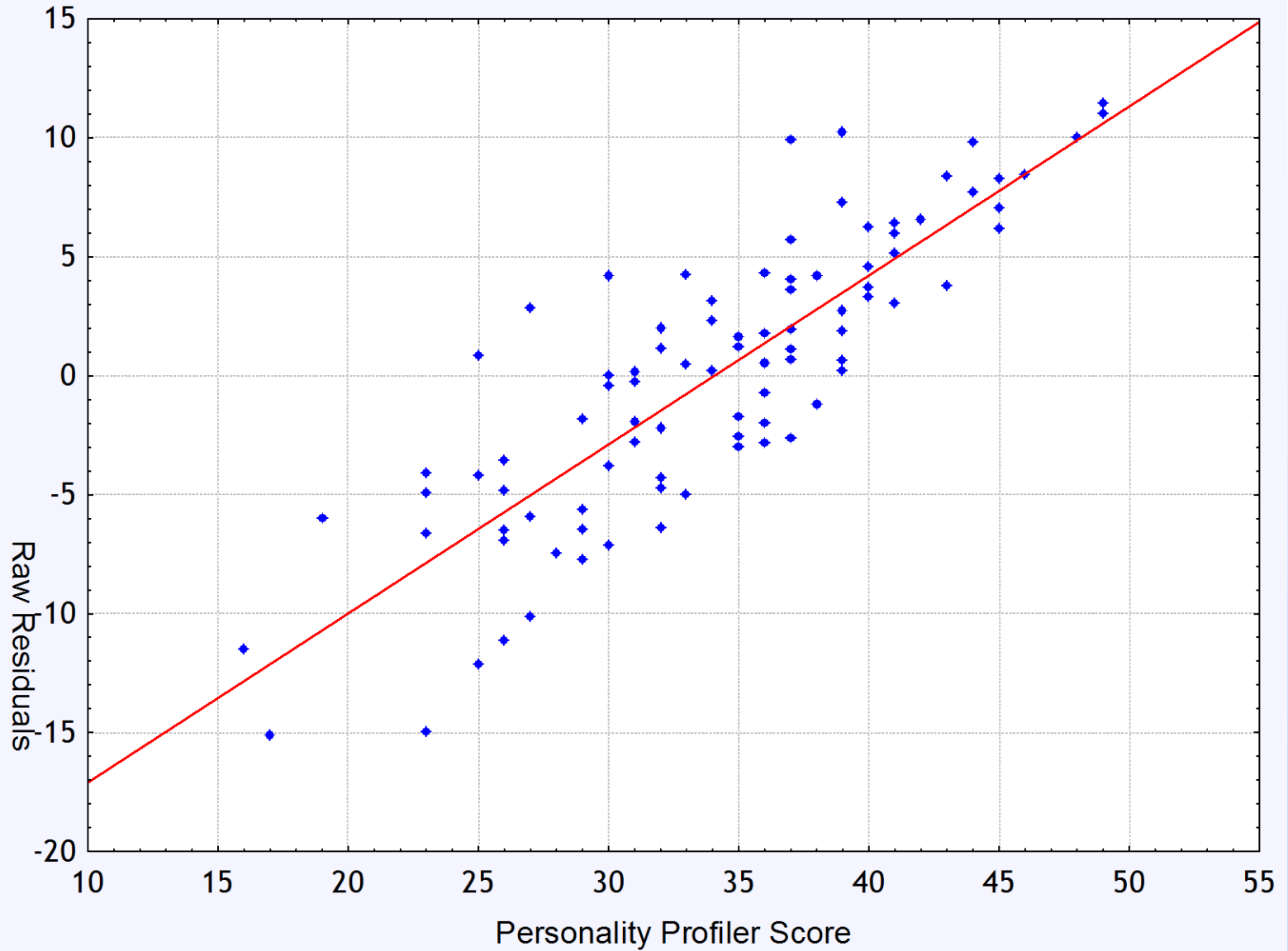
The Personality Profiler	Actual Correlation	Disattenuated
Friendliness	0.71	<i>0.77</i>
Leadership	0.54	<i>0.59</i>
Talkativeness	0.70	<i>0.75</i>
Efficiency	0.63	<i>0.68</i>
Purposefulness	0.56	<i>0.61</i>
Organization	0.60	<i>0.65</i>
Orderliness	0.75	<i>0.80</i>
Calmness	0.54	<i>0.62</i>
Impulse-Control	0.28	<i>0.32</i>
Happiness	0.64	<i>0.69</i>

Profiler  
-VS-  
questionnaire  
scale-scores

**Correlations**

N=99 cases

Predicting AB5C Scale "Leadership" by Profiler Score  
Plotting Prediction Residuals by Profiler Score



# The Personality Profiler

**2. Did you like using the profiler – in that you felt you were able to describe yourself adequately on the measured attributes (e.g. talkative, orderly etc.)**

No, not at all – far too limiting - (1%)

Just like a questionnaire – limiting in what it conveys about you - (12%)

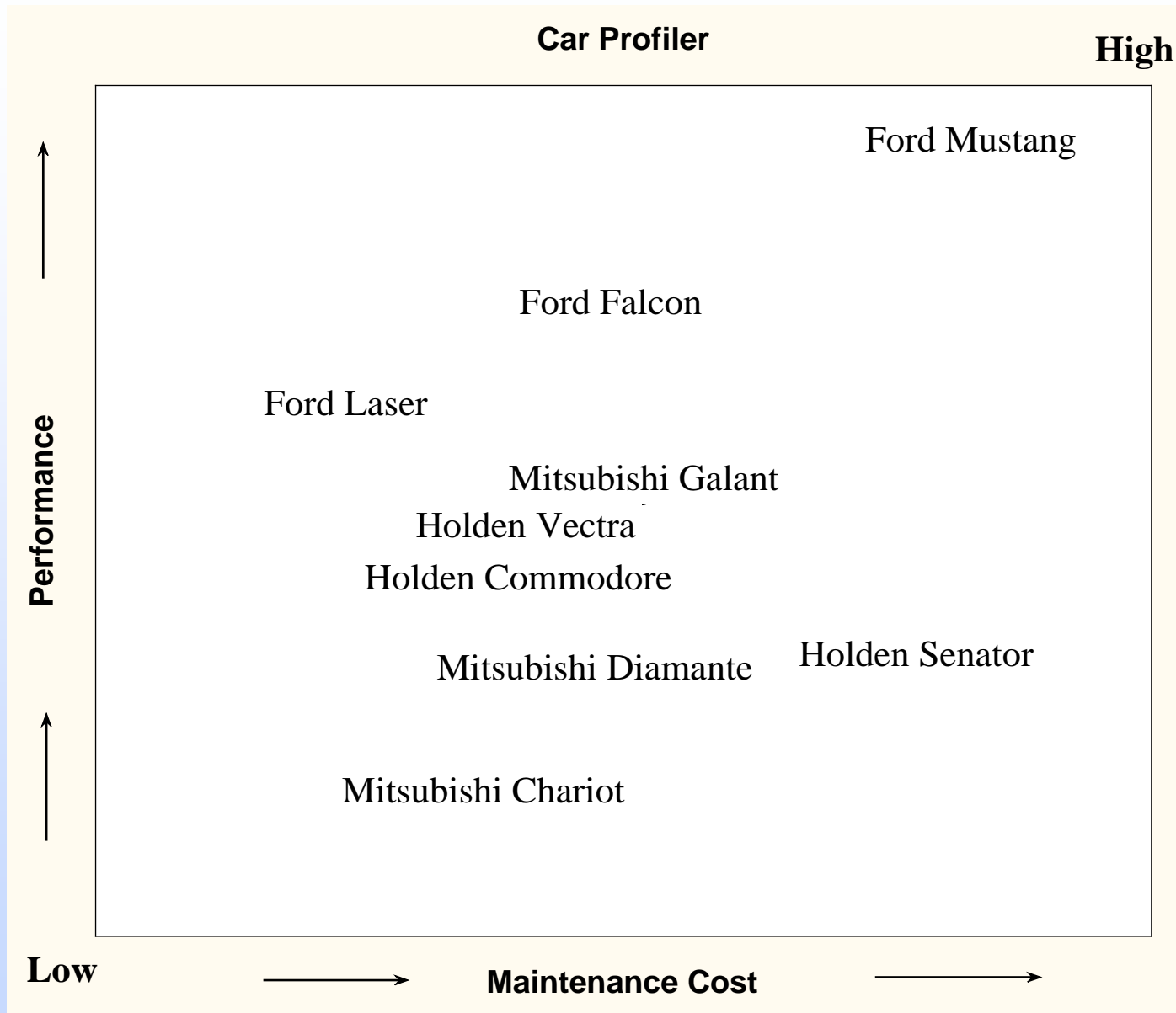
# The Personality Profiler

**2. Did you like using the profiler – in that you felt you were able to describe yourself adequately on the measured attributes (e.g. talkative, orderly etc.)**

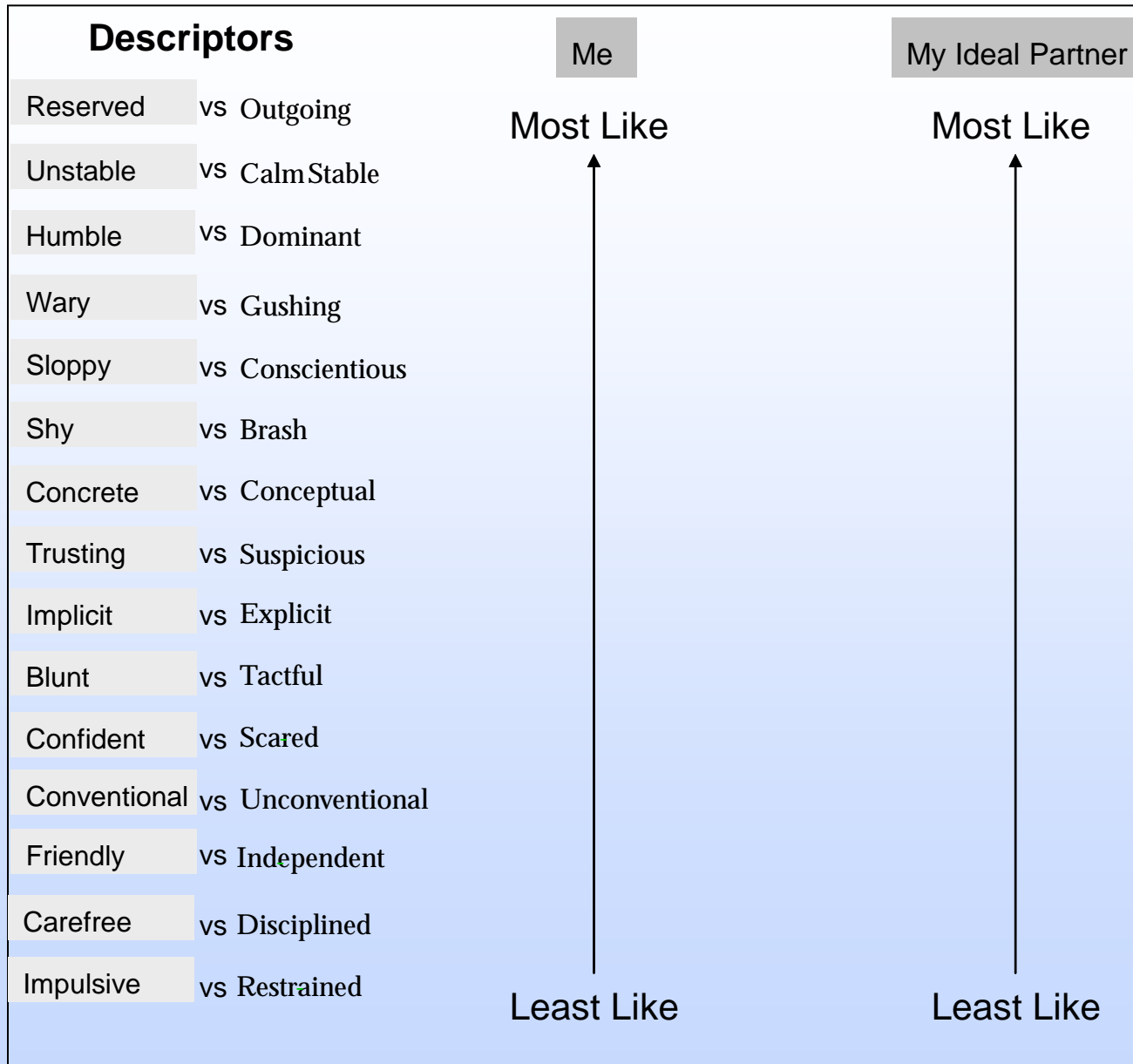
Well, not bad actually – better than a questionnaire, and probably about as good as this kind of assessment can get (in that it is still a simple assessment of something very detailed i.e. personality)

**- (87%)**

# The Market Research Profiler



# The Dating Preference Profiler



# Dynamic Testing

Grigorenko, E.L., & Sternberg, R.J.  
(1998) *Dynamic Testing*. *Psychological  
Bulletin*, 124, 1, 75-111.



# The Psytech Programming Potential Test

## The Test Rationale

- ★ The test is designed to measure variables associated with learning.
- ★ It is measuring how a person interacts with a relatively complex interface.
- ★ The testee is initially presented with a complete “one-screen” interface in which all problem solutions take place.

# The Psytech Programming Potential Test

- ★ This interface possesses a problem area, a tool area, and a message area.
- ★ Problems are introduced in a structured way, to exercise basic tool use prior to “stretching” the testee with more complex problems.
- ★ All problems are answered correctly by a testee – because the system will provide the answer if necessary.

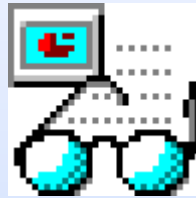
# The Psytech Programming Potential Test

★ What we are concerned with is the manner and speed with which a person achieves a resolution to a problem. There can only ever be one answer, which will be exposed to the testee either as a result of the correct use of a set of computational tools, or via the interactive progressively structured context-sensitive help which is presented at key junctures by the program itself.

# The Psytech Programming Potential Test

★ The only requirement for testing an individual is that they can read in the language in which the test is administered. No other knowledge is required.

# The Psytech Programming Potential Test



# The Psytech Programming Potential Test

Some of the variables acquired per item:

- ★ Item completion time
- ★ Prompt message required at start
- ★ Wrong tool error count
- ★ How many prompts required
- ★ Maximum level of prompts required
- ★ How many times main help activated
- ★ How many times interactive help activated
- ★ How many times tutorial help activated
- ★ Durations for various help accesses

# Sex-Offender Visual Tracking Time

Consider answering the following question asked within Forensic Psychology:

**How do you determine a person's sexual preferences without relying in any way upon their self-reports or interview responses...**

# Sex-Offender Visual Tracking Time

- ★ Observed Behaviour
- ★ Penile Plethysmography
- ★ Abel Screen (visual inspection time)
- ★ VTT (Visual Tracking Time)



# Sex-Offender Visual Tracking Time

Using precise eye-tracking and head-movement hardware and software ...  
pictorial stimuli ... and normative data on  
the following meta-variables ...

- ★ Stimulus Regions of Interest
- ★ Initial Stimulus Trajectory Tracking
- ★ Trajectory-Vector Maps
- ★ Time Off-Picture (sabotage)
- ★ Head and Eye-Movement Geometries

# Making better use of Assessment Data?

- ★ GeneSys Intelligent Psychometrics (*Psytech International*)
- ★ Smart Profiling

Some thoughts ...

# Who Can Afford to Innovate?

The M7 2D-Profiler ... ~ US\$ 100,000

The finished and validated Talent ~ US\$ 300,000

Engine asp model application

-----  
~ US\$ 400,000

\*The actual product in its  
entirety cost nearly **US\$ 1,000,000 (NZ\$ 2m)**

The Psytech PPT ... ~ US\$ 17,000 ↗

The VTT ... ~ US\$ 150,000

# Who Can Afford to Innovate?

What is the ROI on both the client and developer side for these new applications?

# Who Can Afford to Innovate?

Innovation requires failures to occur, albeit in a financially favourable ratio to the number of successes. But, failures will occur – and sometimes never make it to market. So, who in the Psychological Test “Industry” can afford to sustain such failures as part of a “Drive for Innovation”?

# Who Can Afford to Innovate?

Perhaps the most fertile and obvious domain for innovation is at the Universities. Certainly, some individuals/teams in the US (**Susan Embretson** and colleagues at Kansas, **Fritz Drasgow** and colleagues at Illinois, **Neal Schmitt** and colleagues at Michigan, and **Julie Olson-Buchanan** at California) are forging ahead - but ...

# References

Barrett, P.T. (August 2003) **Beyond Psychometrics: Measurement, non-quantitative structure, and applied numerics.** *Journal of Managerial Psychology.*

Sternberg, R. J., & Williams, W. M. (1998). **You proved our point better than we did: A reply to our critics.** *American Psychologist, 53, 576-577.*



# References

Goldberg, L.R. (in press). **The Comparative Validity of Adult Personality Inventories: Applications of a Consumer-Testing Framework.** In S. R. Briggs, J. M. Cheek, & E. M. Donahue (Eds.). *Handbook of Adult Personality Inventories.* (In press). New York: Plenum Publishing Corp.

# References

Brian Haig (2002) *Towards an abductive theory of scientific method*. In N. Stephenson et al (eds), *Theoretical Issues in Contemporary Psychology*. Boston: Kluwer.

Michael Maraun (1998) *Measurement as a Normative Practice*. *Theory and Psychology*, 8, 4, 435-461

Michell, J. (1990) *An Introduction to the Logic of Psychological Measurement*. Lawrence Erlbaum.

# References

Michell, J. (1997) *Quantitative science and the definition of measurement in Psychology. British Journal of Psychology*, 88, 3, 355-383.

Joel Michell (1999) *Measurement in Psychology: a critical history of a methodological concept*. London: Cambridge University Press

# References

Michell. J. (2000) *Normal science, pathological Science, and psychometrics. Theory and Psychology*, 10, 5, 639-667

Michell, J. (2001) *Teaching and misteaching measurement in psychology. Australian Psychologist*, 36, 3, 211-217

# References

Michell, J. (2002) *Stevens's Theory of Scales of Measurement and its place in Modern Psychology*. *Australian Journal of Psychology*, 54, 2, 99-104.

Peter Schönemann (1994) *Measurement: The Reasonable Ineffectiveness of Mathematics in the social sciences*. In I. Borg, and P. Mohler (eds), *Trends and Perspectives in Empirical Social Research*. Berlin: Walter De Gruyter