

Psychology in the 21st C – Getting over our addiction to *p* so our research can be evidence for our practice

Neville M Blampied,
University of Canterbury
Christchurch, New Zealand
New Zealand Psychological Society
Annual Conference
30 Aug-1 Sept 2014
Neville.blampied@canterbury.ac.nz

Challenges for 21st C Psychology

... in the final analysis, the progress of ... psychology, as of every other science, will be determined by the value ... of its contributions to the advancement of the human race. (Witmer, 1907)

In the context of our disciplinary & professional
epistemic and ethical responsibilities -

**A commitment to
The Scientist-practitioner ideal/
Evidence-based practice**

Challenges for 21st C Psychology

... in the final analysis, the progress of ... psychology, as of every other science, will be determined by the value ... of its contributions to the advancement of the human race. (Witmer, 1907)

We have two problems

1. Dominant methods not ideal for yielding evidence relevant to practice
2. Researchers do not practice evidence-based practice

The French Connection

Conference dinner, 1st International Congress of Psychology, 1889

Delegates [including Bain, Galton, Helmholtz, James, Janet, Lombroso, Sechenov, & Wundt] admired *the wonderfully illuminated landscape of exhibition grounds, palaces and fountains spread out below, with all the lights and shadows of nocturnal Paris framing it in* (William James, 1889, p 615)



French contributions to method (1):

20th C

postmodernism ... rejects epistemological assumptions, refutes methodological conventions, resists knowledge claims, obscures all versions of truth ... [Rosenau, 1992, p 3]

1950's →

Post-modernism

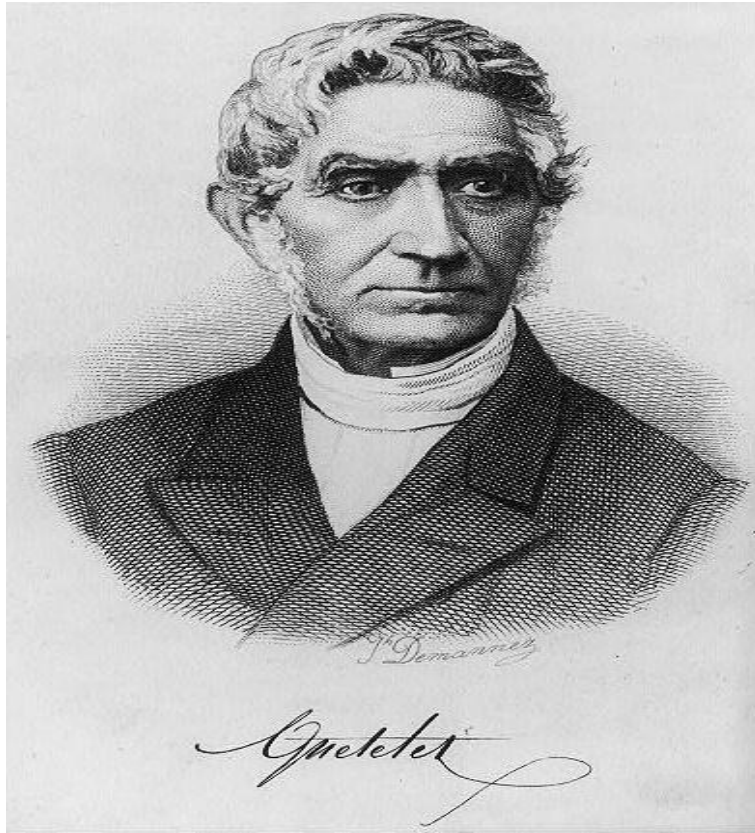
Post-structuralism →

Qualitative research

Derrida, Sartre, Levi-Strauss, Foucault, Lyotard, de Man

French contributions to method (2): 19th C

Adolphe Quetelet (1796- 1874)



Claude Bernard (1813 – 1878)



Divergent research traditions in Science

The particular was considered noise in the system. What one wanted was a display of the anatomy of nature, not its individual countenance. [Eisner, 2003, p 19]

nomothetic

- Concerned with general laws
- Concerned with the universal
- Abstract
- Timeless
- Objective/impersonal
- **Inter-individual research**
- **Legacy of Quetelet**

Divergent research traditions in Science

... *The true relations of phenomena disappear in the average.* [Bernard, 1865]

nomothetic

- Concerned with general laws
- Concerned with the universal
- Abstract
- Timeless
- Objective/impersonal
- **Inter-individual research**
- **Legacy of Quetelet**

idiographic

- Concerned with the individual case in context
- Concerned with the particular
- Concrete
- Historically situated
- Subjective/personal
- **Intra-individual research**
- **Legacy of Bernard**

Divergent research traditions in Science

... *The true relations of phenomena disappear in the average.* [Bernard, 1865]

SCIENCE

nomothetic

- Concerned with general laws
- Concerned with the universal
- Abstract
- Timeless
- Objective/impersonal
- **Inter-individual research**
- **Legacy of Quetelet**

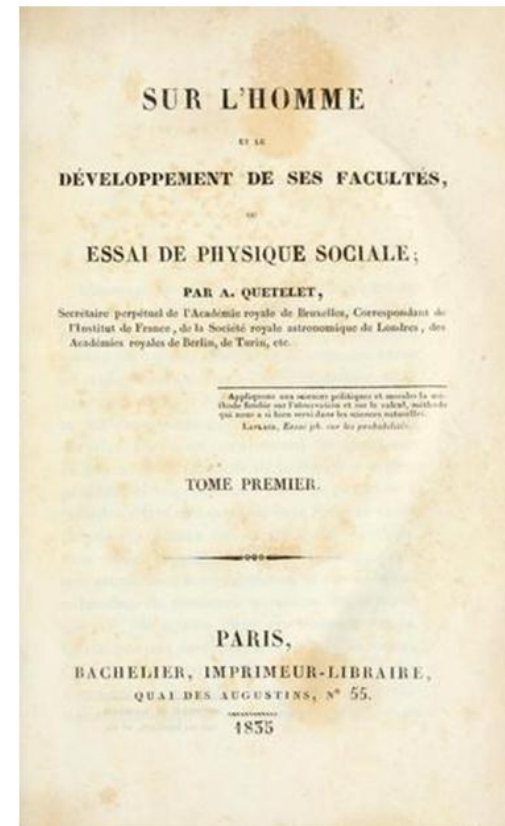
idiographic

- Concerned with the individual case in context
- Concerned with the particular
- Concrete
- Historically situated
- Subjective/personal
- **Intra-individual research**
- **Legacy of Bernard**

Quetelet's famous book & idea #1

A treatise on man and the development of his faculties
(English translation, 1842)

Social physics



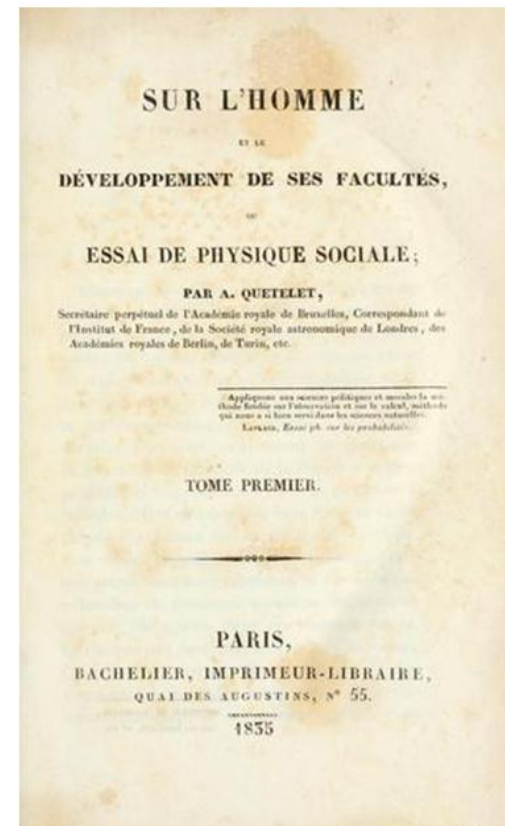
Quetelet's famous book & idea #1

L'homme moyen – the average man

If an individual ... possessed all the qualities of the average man [sic] he would represent all that is great, good, or beautiful.

[Quetelet, 1842, p 100]

A treatise on man and the development of his faculties
(English translation, 1842)



Quetelet's idea #2 : The normal distribution & human attributes

... all our qualities, in their greatest deviation from the mean, produce only vices.

[Quetelet, 1853, p 49]

Measures of human physical and 'moral' characteristics are normally distributed

Attributes = ideal (average) +/- error deviation (*vices*)

It was Quetelet's notion that every individual represented an attempt by nature to achieve perfection, but that, like a person shooting at a target, nature's aim was never exact.

[Johnson & Pennypacker, 1993]

Quetelet's idea # 3 : Development studied by cross-sectional research

*He suggested that instead of making numerous observations on an individual as he [sic] progressed through life, the changes from one age level to another might be studied by making observations on large numbers of people at different ages... [Quetelet] **established the precedent for drawing inferences concerning the nature of dynamic individual phenomena on the basis of statistical comparisons made between large groups of individuals.***

[Johnson & Pennypacker, 1993, p 91]

Quetelet's legacy: via R.A. Fisher

Every experiment may be said to exist only in order to give the facts a chance of disproving the null hypothesis. (Fisher, 1935).

Devised (or improved on)

- ❑ Control groups
- ❑ Randomization
- ❑ Factorial designs, & Analysis of Variance
- ❑ Null Hypothesis Significance Tests (NHST)

The methods ... of Fisher are beautifully suited and highly effective [for] ... the purposes for which they were developed – population genetics, agricultural research, and industrial quality control.

[Johnston & Pennypacker (1980), p 91]



The inference revolution in Psychology

Although their initial understanding of these techniques was often defective, psychologists embraced them with considerable enthusiasm. [Danziger, 1987].

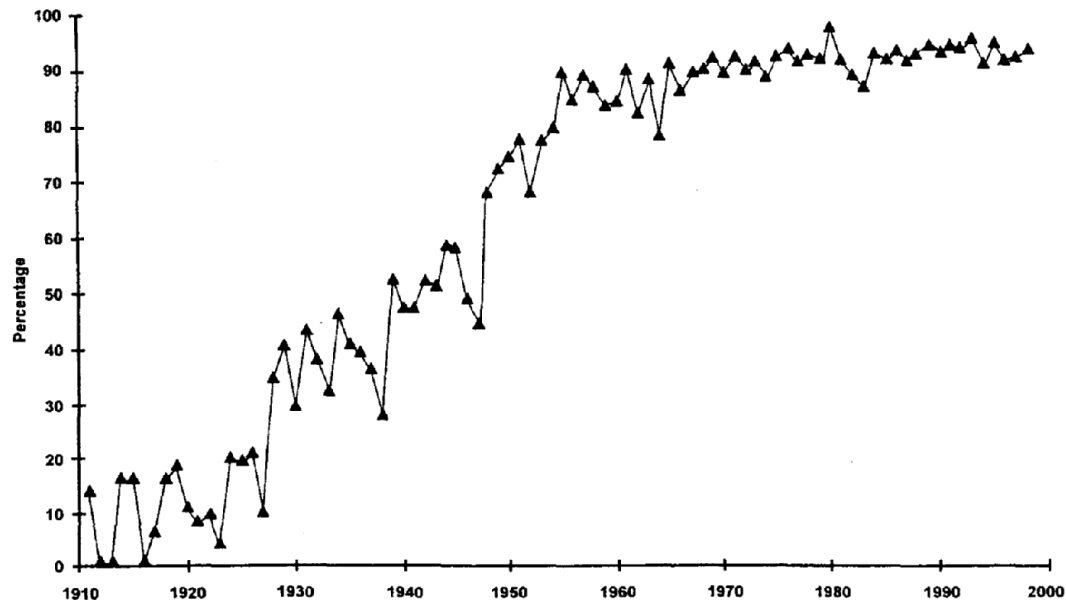


Figure 1.1. Percentage of articles reporting results of statistical tests in 12 journals of the American Psychological Association from 1911 to 1998. From “The Historical Growth of Statistical Significance Testing in Psychology—And Its Future Prospects,” by R. Hubbard and P. A. Ryan, 2000, *Educational and Psychological Measurement*, 60, p. 665. Copyright 2001 by Sage Publications. Reprinted with permission.

Science – the Standard Doctrine

...mainstream psychology ... engaged in 'methodolatry', worshipping a particular method – methodolatry of a remarkably monotheistic kind. [Malson, 2010]

- Postulate a [null] hypothesis about a population
 - Recruit as large a sample* as possible.
 - Randomly allocate participants to treatment conditions.
 - Aggregate individual data into **group averages** (sample statistics).
 - Draw inferences about the population from the sample statistics.
 - Use NHST to assess the data assuming H_0
 - reject H_0 if $p < \alpha$ (typically, α set at $p = 0.05$).
 - If H_0 rejected, accept H_A
 - Worry about Type 1 error (false rejection of H_0)
 - Don't worry about Type 2 error (false rejection of H_A)
- * From a known population

Criticisms of the Standard Doctrine

... one of the worst things that ever happened in the history of psychology [Meehl, 1978]

- Bernard criticised “statistics” in the 19thC
- Criticism in Psychology from 1942
- Criticism has increased substantially from the 1990’s
 - Harlow, L.L., et al. (1997) *What if there were no significance tests?*
 - APA Taskforce on Statistical Inference (1999)
 - Nickerson (2000). Null-hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, 5, 241-301
 - Ziliak & McCloskey (2008). *The cult of statistical significance*.
 - Lambdin, C. (2012). Significance tests as sorcery: Science is empirical – significance tests are not. *Theory & Psychology*, 22, 67-90.

Responses to criticism

There is no comprehensive, definitive refutation of the criticisms of standard research doctrine!

Generally speaking, most defenders of NHST believe that it has been misinterpreted and badly used for decades.

[Balluerka, et al., 2005]

Result - Major problems for applied science

... the intimate relationship between psychological research and statistics verges on the indecent

[Howit, 2013]

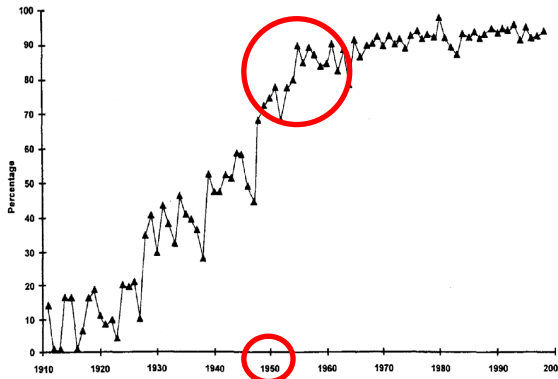


Figure 1.1. Percentage of articles reporting results of statistical tests in 12 journals of the American Psychological Association from 1911 to 1998. From "The Historical Growth of Statistical Significance Testing in Psychology—And Its Future Prospects," by R. Hubbard and P. A. Ryan, 2000, *Educational and Psychological Measurement*, 60, p. 665. Copyright 2001 by Sage Publications. Reprinted with permission.

1949 – year of the
Bolder Conference

Beginning of the
Scientist-Practitioner
ideal of applied
psychology

*Empirically supported psychological treatments ... [are those] that have been exposed to evaluation **using the accepted methods of psychological science.***

[Kendall (1998) p3.]

Result - Major problems for applied science

... the intimate relationship between psychological research and statistics verges on the indecent

[Howit, 2013]

Review of published research 1967 – 1988:

- Extreme misinterpretation and misuse of p values

We found that p values received an absurdly central position in many studies... often at the expense of descriptive statistics. In a study of weight loss, for example, authors chose to mention ... the imposing p of .001 rather than any measure of weight loss. ... Finally, one article ... contained only p values; no means, standard deviations; correlations coefficients; or any other statistics [i.e., actual data] were mentioned. ... In general, null hypothesis tests have been misinterpreted and misapplied all along.

- Abuse of alpha levels (*marginally significant*, etc)
- Belief that small p values mean that experimental hypotheses are confirmed

Things have improved? No

Reviewing 1993 – 2001:

- All the earlier problems found, plus
- Persistent confusion of statistical with clinical significance
- Failure to consider practical/clinical significance of findings

... only 40% of articles made any attempt to discuss clinical significance. This is serious. In a major journal dedicated to the research of psychotherapy and other interventions, clinical significance should be relevant to more than 40%... [Fidler, 2005]

Fidler, Cumming, Thompson, et al. (2005). Towards improved statistical reporting in the Journal of Consulting & Clinical Psychology. *JCCP*, 73, 136 -143

Problem: Nomothetic-idiographic mismatch

... by attempting to describe only the average, one runs the risk of describing nobody in particular.

(Molden & Dweck, 2006, 192-203).

- Nomothetic research not ideal for yielding evidence relevant to practice

...there is a growing disaffection from traditional experimental designs and statistical procedures which are held inappropriate to the subject matter ... [particularly] in the area of therapeutic change...

Bergin, A, & Strupp, H. (1972). *Changing frontiers in the science of psychotherapy*. Chicago, Il: Aldine [p 440].

And

The application of knowledge is always to the single case [Allport, 1942]

- Intervention is necessarily idiographic

*There are three fundamental questions [about any intervention]: (a) does it work under special, experimental conditions? [Efficacy](b) does it work in practice? [Effectiveness]and (c) **is it working for this patient?** [Progress]The crucial question posed by the practitioner is Question c.*

[Howard, et al.,1996, 1059-64]

Evidence-based practice

*Evidence-based practice is a transdisciplinary, **idiographic** approach* [Spring, 2007, p 611]

EBP is the contemporary expression of the scientist-practitioner ideal

***Evidence-based medicine** is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of **individual patients**.* [Sackett, et al., 1996, p 72]

***Evidence-based practice in psychology** is the integration of the best available research with clinical expertise in the context of patient characteristics, culture, and preferences. ... [It] **starts with the patient** and asks what research evidence ... will assist the psychologist in achieving the best outcome ... [and] articulates a decision-making process for integrating multiple streams of research evidence...into the intervention process.*

[APA Presidential Taskforce on Evidence-based Practice, 2006, p273]

Problem is intrinsic for Standard Research Doctrine

Consistent with nomothetic research

Hypotheses are about populations

[Hayes, 1963]

Not about

- The sample
- The individual cases in the sample

Yet, Evidence-based Practice

- *is about the care of individual patients*
- *starts with the patient*
- *is idiographic*

The “double standard” in Psychology

Throughout its history as a science, psychology has ben plagued by a double standard in its treatment of the individual subject...

*In psychological discourse (both scientific and applied) the individual ... is constantly given high relevance. In contrast, **the individual case is usually forgotten** in the practice of psychological research **because it is replaced by samples** of subjects that are assumed to represent some general population*

(Valsiner, 1986, p1).

Getting from the average to the individual – How?

Well – we can't, legitimately

So

- We attribute generic, prototypic, ideal status to the population average – Quetelet's Legacy!!
We present our significant result as if it were about all mankind and all time (Gigerenzer, 1987, p 13 – 33).
- We generalize (inductive generalization) from the average to all members of the putative population
...characteristics of that abstracted individual may easily become attributed to particular concrete individuals with whom psychologists ... work (Valsiner, 1986, p1-2).

One can always generalize; the question is whether one ought to.

[Sechrest & Walsh, 1997]

Conclusion - for evidence-based practice we need a new way of doing science

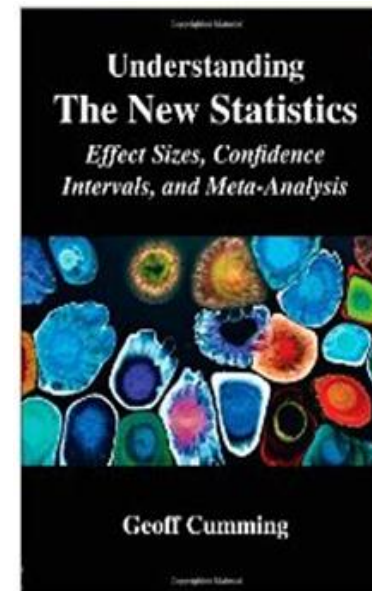
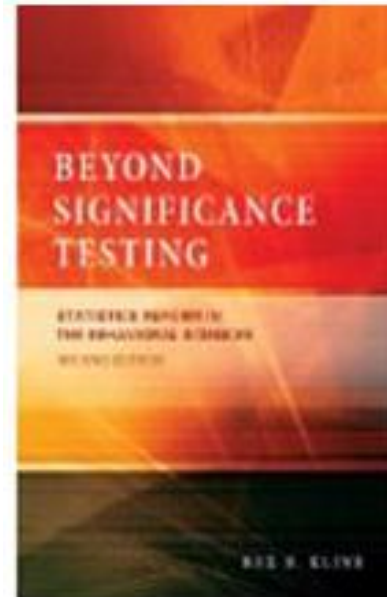
*Unanswered questions concerning individual[s]... will continue to puzzle the ... practitioner as he or she works with whomever happens to appear... **This is the one single factor that requires substantial alterations in ... the way we do science.** ... a methodology that highlights the *individual* and, at the same time, **maintains the integrity of an empirical and scientific approach** to the study of behavior.*

[Barlow, Hayes, & Nelson. (1984). *The scientist-practitioner: Research and accountability in clinical and educational settings*, p 52-53.]

New ways of doing science (1)

Nomothetic tradition: *The new statistics*

- Estimation }
- Precision – }*
 - Confidence intervals
- Effect sizes
- Meta-analysis
 - (best evidence synthesis)
- *Both lead to concern for
Measurement: validity/reliability/error
- Does not use NHST or $p < ?$



... *friends do not let friends compute p* [quoted in Klein, 2013).

I conclude from the arguments and evidence I have reviewed that

best research practice is not to use NHST at all [Cumming, 2012]

New ways of doing science (2)

Replication



Replication draws for its inductive force on
the principle of unlikely successive coincidences
[Barlow et al., 1984]

An ounce of replication is worth a ton of inferential statistics.
[Steiger, 1990]

... we must finally rely, as have the older sciences, on replication.
[Cohen, 1994]

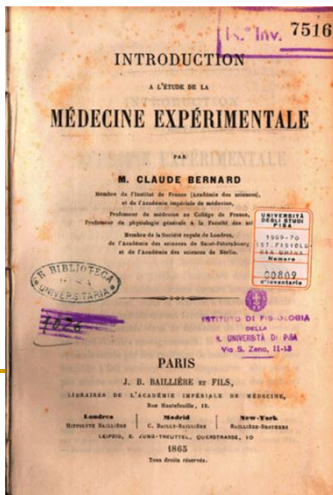
replication is at 'the heart of any science'...
[Schmidt, 2009]

New ways of doing science (3)

Re-discovering the legacy of Claude Bernard

A physician ... is by no means physician to living beings in general, nor even physician to the human race, but rather, physician to a human individual. ...The use of averages ... leads necessarily to error.

[Claude Bernard, 1865/1949, p 92, 134]



Idiographic science

Barlow & Nock (2009). Why cant we be more idiographic in our research? *Perspectives on Psychological Science*, 4, 19 – 21

Molenaar (2004). A manifesto on psychology as idiographic science: Bringing the person back into scientific psychology, this time forever. *Measurement: Interdisciplinary Research & Perspective*, 2, 201-218

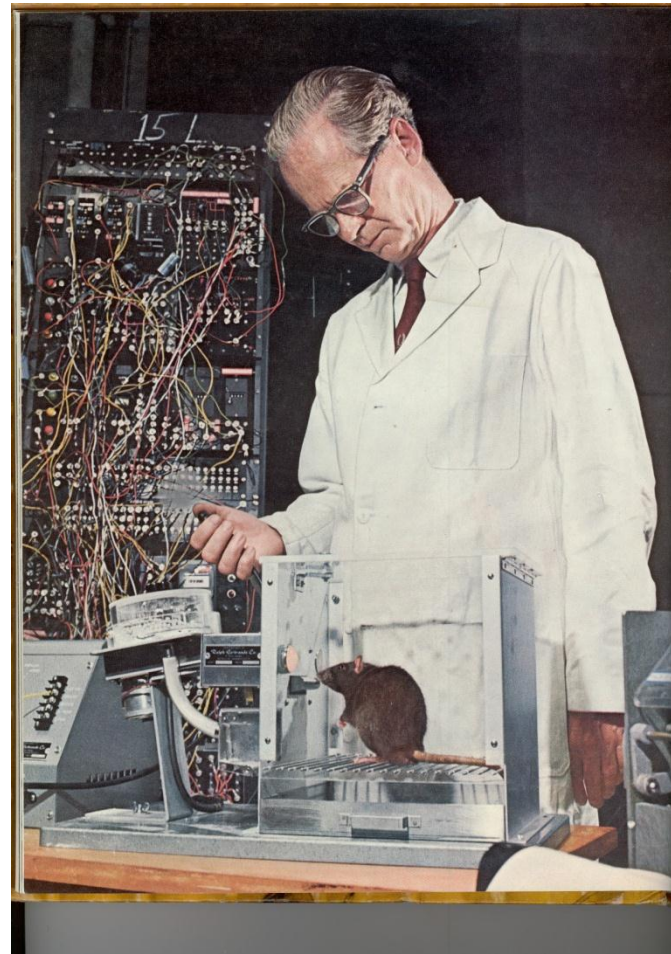
You cant treat a population [Grice, 2012]

New ways of doing science (4)

Single-case research = individuals + replication

the kind of science proposed here belongs on the non-statistical side. ... In placing itself in that position it gains the advantage of a kind of prediction concerning the individual that is necessarily lacking in a statistical science. ... Individual prediction is of tremendous importance, so long as the individual is to be treated scientifically.

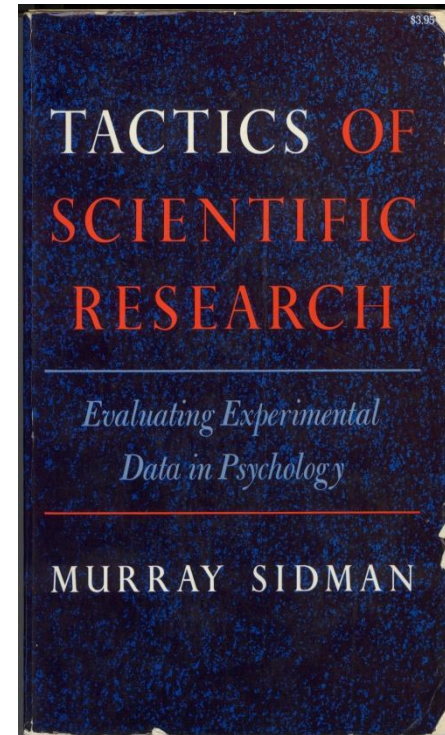
[Skinner, 1938]



Single-case research ...

The soundest empirical test of the reliability of data is provided by replication ... intersubject replication is a more powerful tool than intergroup replication. ... With intersubject replication ... each additional experiment increases the representativeness of the findings.

[Sidman, 1960]



Single-case research...

What it does not do:

- Rejects sampling theory.
- Does not average over participants.
- Does not use NHST-based inference.

What it does do:

- Maintains a commitment to quantification
- Uses rigorous experimental designs that permit causal inferences
- Uses Visual Analysis as its key data analysis technique
- Uses Replication as its key justificatory procedure

New ways of doing Science (5)

Visual Analysis

Recent advocacy -

Loftus, G.R. (1993). A picture is worth a thousand p values: On the irrelevance of hypothesis testing in the microcomputer age. *Behavior Research Methods, Instruments, and Computers*, 25, 250 – 256.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594 – 604.

*As soon as you have collected your data, before you compute any statistics, **look at your data.***

[Wilkinson & Task Force, 1999, emphasis in original]

Visual Analysis ...

you can learn a lot by looking. Yogi Berra

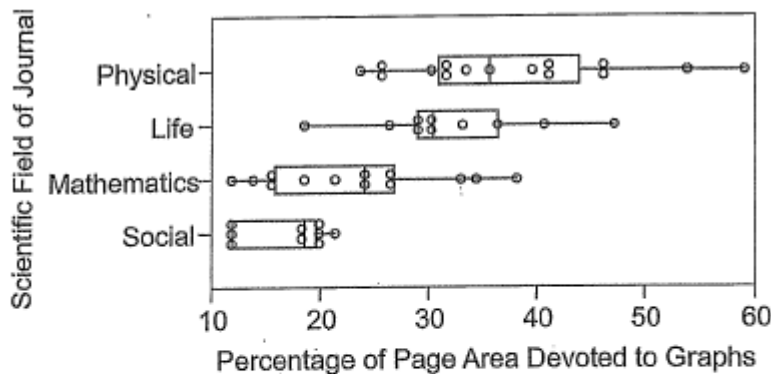


FIGURE 4.1. Dot-boxplot of results in Cleveland (1984) on the use of graphics in scientific journals. The horizontal axis represents the percentage of total page area devoted to graphics in 47 articles sampled from the 1980–1981 volumes of selected science journals.

Visual analysis ... can simultaneously detect curvilinear trends, repeating patterns or cycles in data, delayed or lagged responses following intervention onset, and within-phase changes in variability. In addition it can detect changes in mean level and trend slope across phases ... That breath is not equalled by any other analytic technique.

[Parker & Hagen-Burke, 2007]

[Wilkinson, I. (2012). Graphic displays of data. In H. Copper [Editor-in-chief]. *APA Handbook of Research Methods in Psychology: Vol 3* (pp73 – 100). Washington, DC:APA]

A new methodological synthesis for nomothetic & idiographic research?

Humans, not the gods, created all forms of enquiry, and we can and should modify them.

[Camic, Rhodes, & Yardley, 2003, p4]

The new statistics + replication + single-case research + visual analysis

Above all, abandon our addiction to p

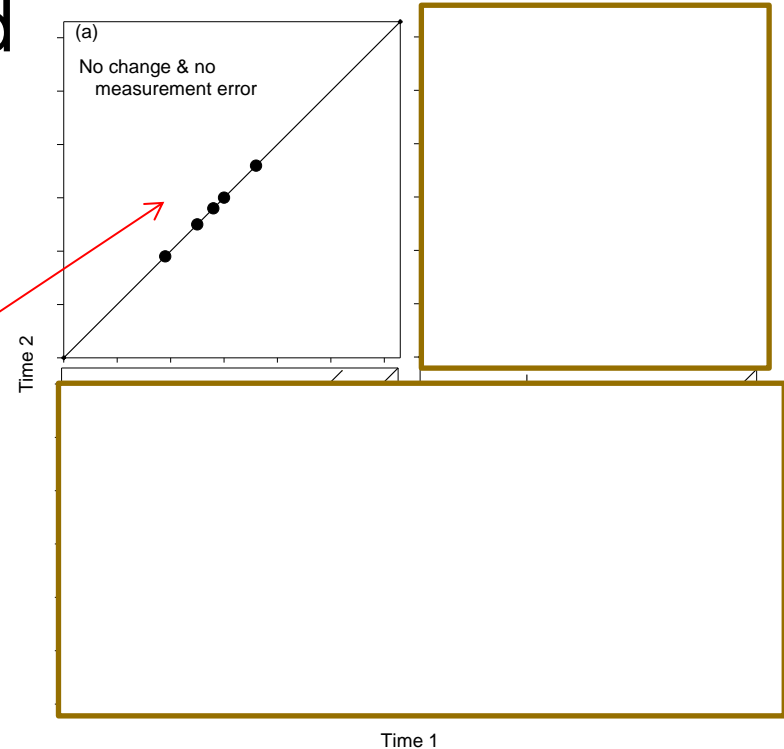
And add

- Proper respect for measurement quality (& error - SEM)
- Use a Reliable Change Index (Jacobson & Truax, 1991)
- Attend to clinical/practical significance

Example – basic elements

Assume change measured
from t_1 to t_2

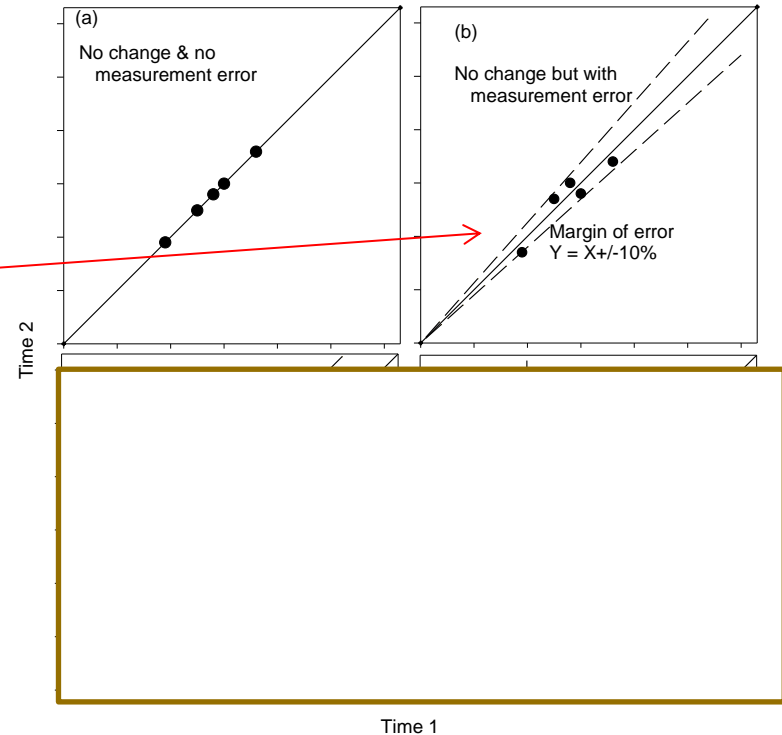
On a scatter-plot
The diagonal is
the line of no effect



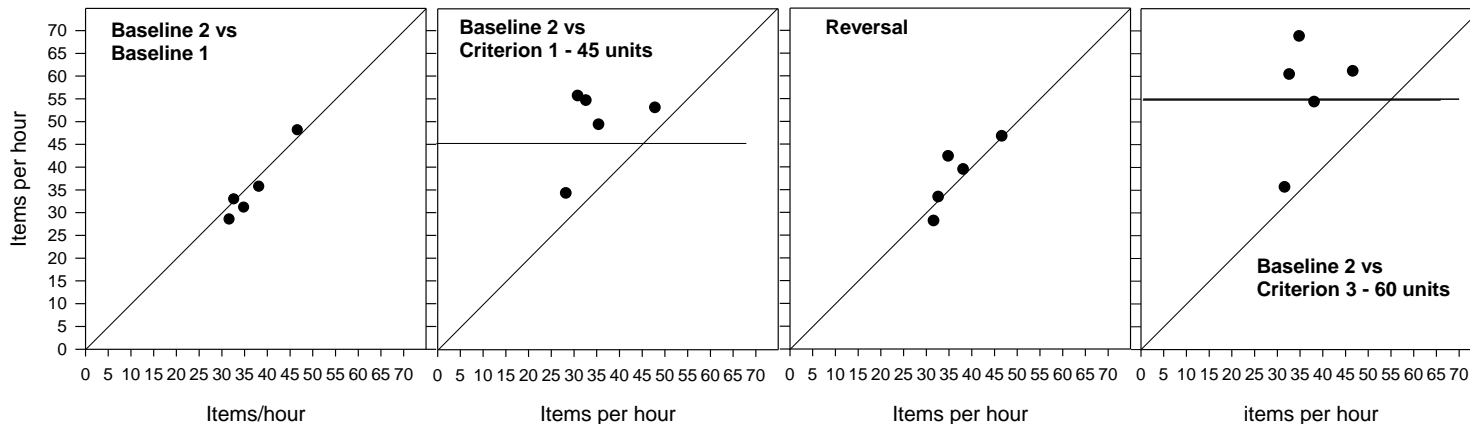
Example – basic elements

Measurement error/
unsystematic variability
shows as variability round
the line

Systematic change shows
as deviations above/below
diagonal



Example - Single-case research – Changing-criterion & reversal design

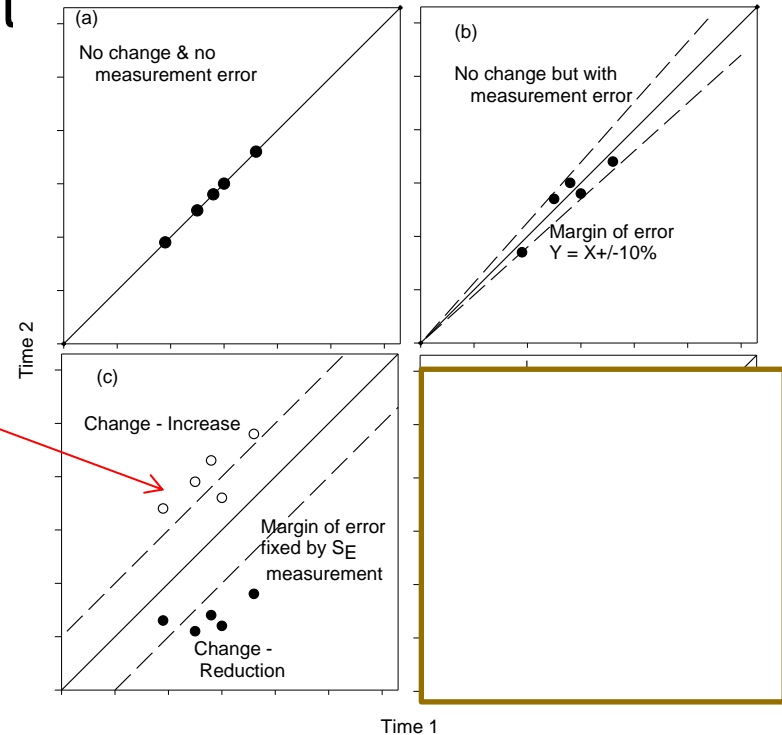


Changes in worker productivity with the introduction of incentives for meeting target criteria.

[Reworked from Straight & Blampied (1998). *NZ Journal of Disability Studies*, 5].

Example – conventional group data

Magnitude of systematic effect compared with measurement error

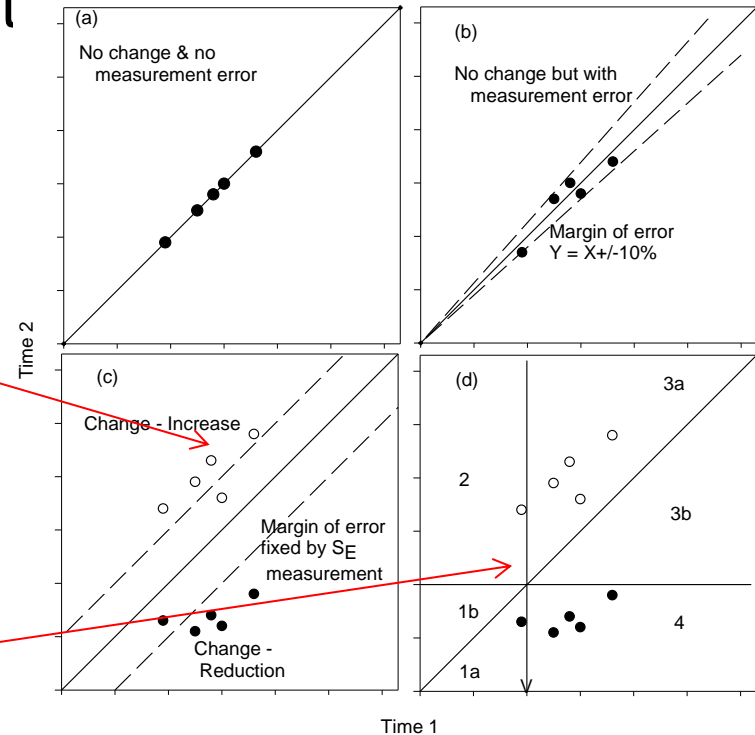


Example – conventional group data

Magnitude of systematic effect compared with measurement error

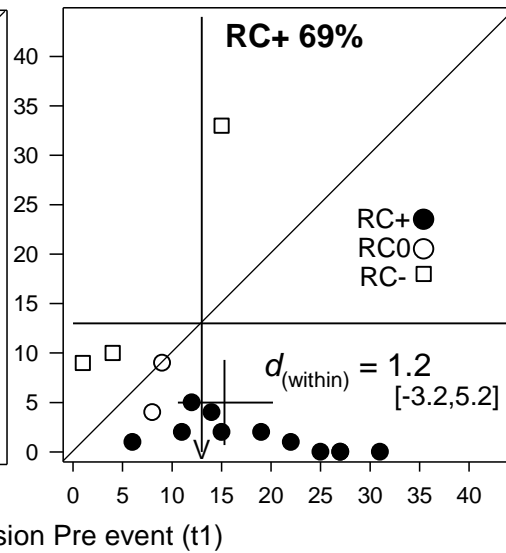
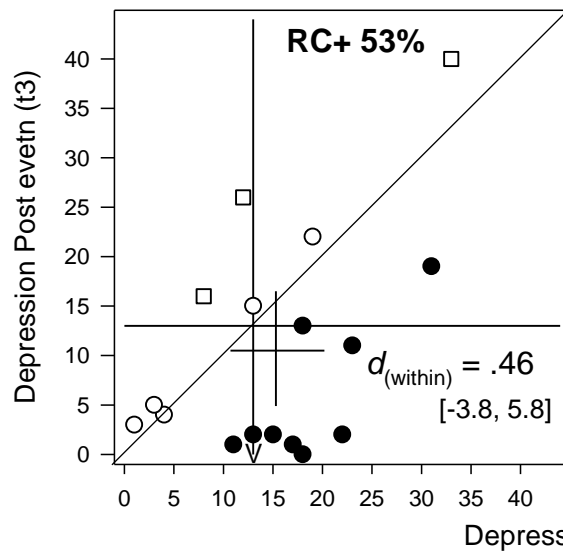
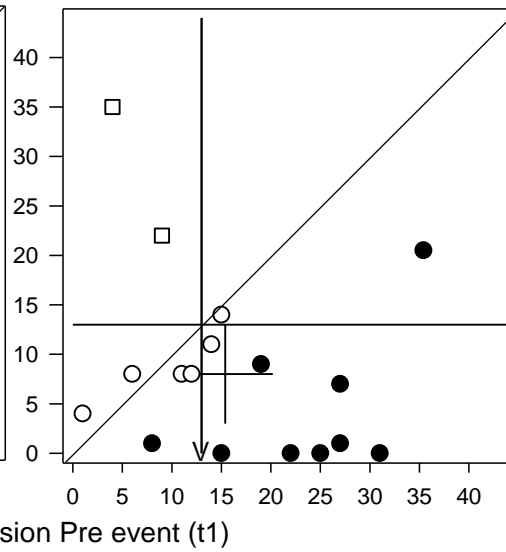
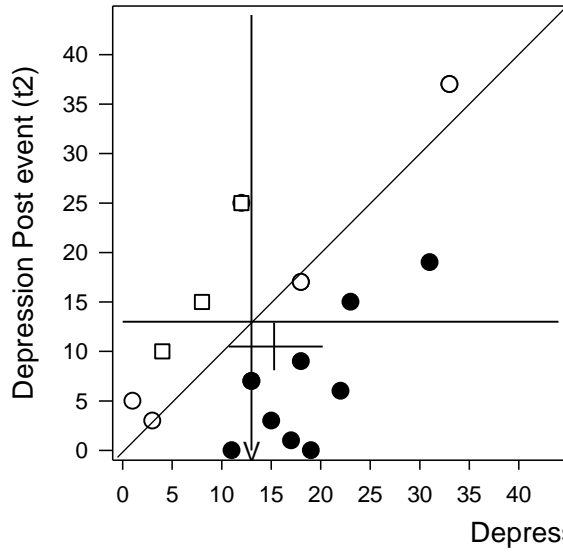
Determines Reliable Change

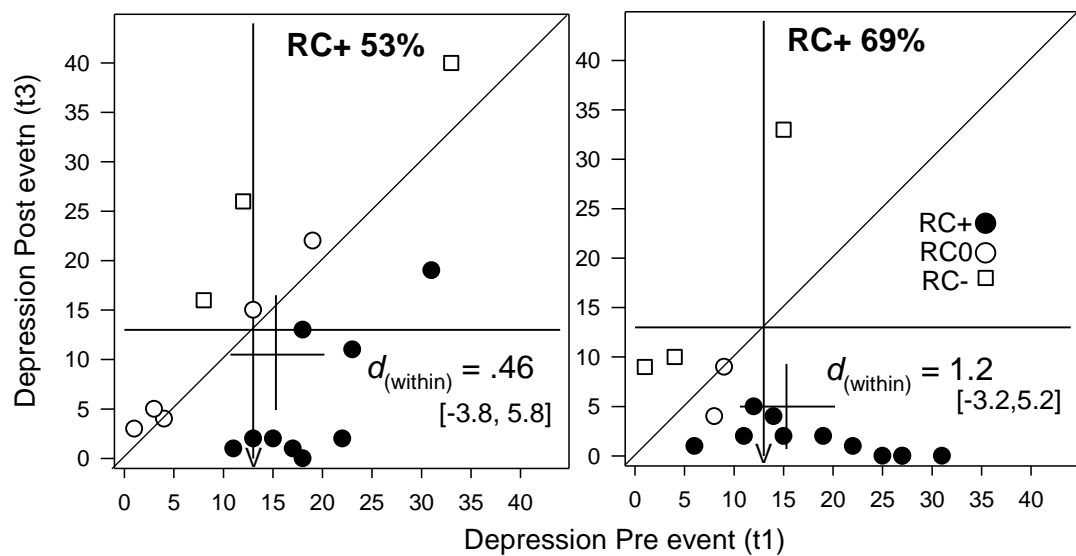
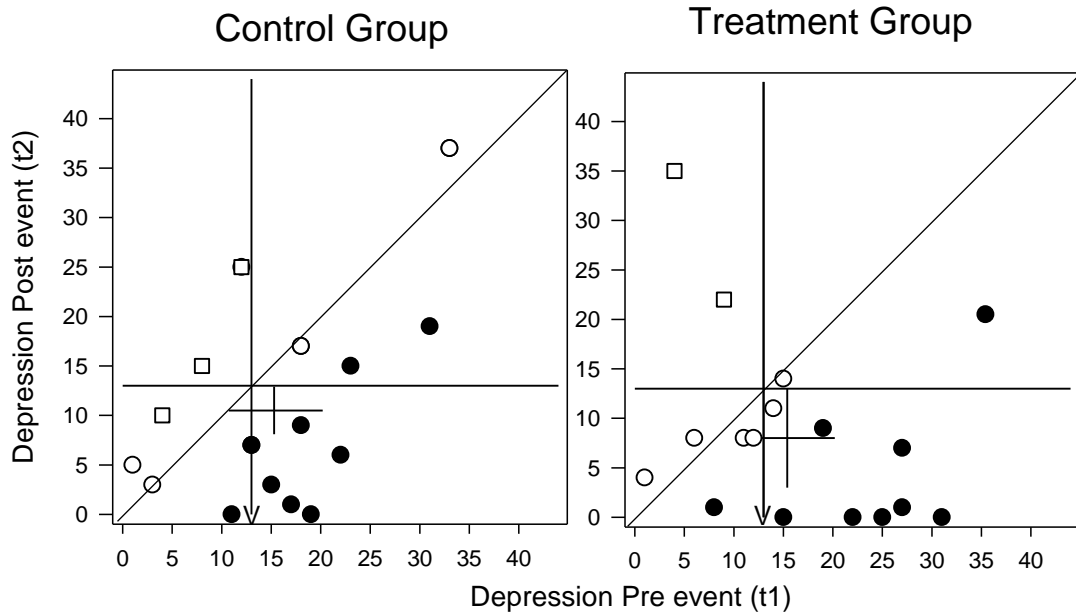
Clinical cut-offs can be added to help interpretation



Control Group

Treatment Group





- Visual analysis ✓
- Groups ✓
- Individuals ✓
- Means ✓
 - 95% Confidence intervals ✓
- Reliable Change ✓
- Effect size ✓
 - 95% CI on ES ✓
- % with reliable change ✓
- Clinical/practical significance ✓

The Challenge

... in the final analysis, the progress of ... psychology, as of every other science, will be determined by the value ... of its contributions to the advancement of the human race. (Witmer, 1907)

We need to take responsibility for disciplinary change to
achieve

Evidence-based research practice,

Producing evidence fit for evidence-based professional
practice

And then do

Professional practice using that evidence

- Personal
- Collaborative
- Institutional
- Organizational

References

- Blampied, N.M. (1999) A legacy neglected: Restating the case for single-case research in cognitive-behaviour therapy. *Behaviour Change*, 16, 89 - 104.
- Blampied, N.M (1999). Murray Sidman – An appreciative introduction. *Behaviour Change*, 16, 75 - 78.
- Blampied, N.M. (2000). Single-case research designs - A neglected alternative. *American Psychologist*, 55, 960.
- Blampied, N.M. (2001). The third way - Single-case research, training and practice in clinical psychology. *Australian Psychologist*. 36, 157 – 163.
- Blampied, N.M. (2013). Single-case research and the scientist-practitioner ideal in applied psychology. In G. Madden, et al. *Handbook of Behavior Analysis*, Vol 1. Washington, DC: APA.
- Blampied, N.M. (2014). Using modified Brinley plots to analyse behaviour change in individuals within groups. Under review.
- Rucklidge, J.J. and Blampied, N.M. (2011). Post-earthquake psychological functioning in adults with Attention-Deficit/Hyperactivity Disorder: Positive effects of micronutrients on resilience. *New Zealand Journal of Psychology* 40(4): 51-57.

Abstract

In the middle years of the 20th C two things happened that had far-reaching impacts on psychology. The first was the invention by R.A Fisher and other statisticians of modern factorial research designs, requiring random assignment of participants to conditions and statistical inference based on null-hypothesis statistical tests (NHST) of group averages. By the mid 1950's researchers in psychology were 'addicted to p ' and the use of NHST became essential for research to be published. The second development occurring at almost the same time, was the development by the American Psychological Association of the scientist-practitioner model of clinical practice. This ideal rapidly became the dominant model for university training of clinical psychologists in the USA and has been generalised to the training of applied psychologists in general and across the world. Not surprisingly, the 'scientist' part of the scientist-practitioner ideal became closely associated with NHST-based research. Clinical and applied research has for nearly 50 years thus also been 'addicted to p ', dominated by the search for statistical significance among group mean differences rather than clinical or practical significance and unable legitimately to make inferences about individual clients. The contemporary rise of the evidence-based practice movement, which can be considered a reformulation of the scientist-practitioner model, has brought sharply into focus again what has also been known for most of those 50 years: Our research methods, and especially our data analysis methods, are poorly adapted to the needs of practice. Research is about ideal, abstract, average types; practice is about individuals in all their diversity and variability. Furthermore, there is now an emergent 'crisis' in psychology due to the recognition that much of our research fails to replicate. I will review this lamentable history, and then consider some of the ways that we can adapt our research practices to make them much better adapted to evidence-based practice. These include the use of single—case research designs and novel methods of visual analysis of data.

Criticisms of NHST

NHST has been called:-

- a *religion* (Cohen, 1990)
 - an *addiction* (Schmidt & Hunter, 1997)
 - an *enduring tyranny* (Loftus, 1991)
 - a *justification for betraying the evidence of the raw data* (Faverau, 1993)
 - *An empty and damaging ceremony* (McCloskey & Ziliack, 2008)
 - a ... *kind of sorcery, a form of statistical shamanism* (Lambdin, 2012)
 - *is among the most questionable things we do* (Trerweiler & Stricker, 1998)
 - *Is the most bone-headedly misguided procedure ever institutionalized in the rote training of science students* (Rozenbaum, 1997)
- one of the worst things that ever happened in the history of psychology.*** (Meehl, 1978).