Why learning 'What works' is often not enough - dealing with complication and complexity

Patricia J. Rogers

Professor of Public Sector Evaluation
Centre for Applied Social Research
Royal Melbourne Institute of Technology, Australia
Patricia.Rogers@rmit.edu.au

BetterEvaluation event

Rome, Italy 19 September 2011
A long history of evidence-based policy and practice

---

Blue-eyed people better off, say scientists

By staff writers | NEWS.com.au | August 20, 2007 12:23pm

- Blue-eyed people "better strategic thinkers"
- Are "likely to achieve more in life"
- Public quick to link study with racism

PEOPLE with blue eyes are likely to achieve more in life than those with brown, say US scientists.

Scientists who conducted the tests said brown-eyed people performed better at reaction time, but those with lighter eyes appeared to be better strategic thinkers, the Daily Mail reported.

Brown-eyed people succeeded in activities such
A common myth

EVIDENCE BASED POLICY

FIND OUT WHAT WORKS AND DO IT
Possible reasons why finding out what works and doing it might not produce better results

PROBLEMS WITH THE EVIDENCE

1. The evidence can be wrong
2. The evidence can be incomplete

PROBLEMS WITH GENERALIZING FROM THE EVIDENCE

Dealing with complication and complexity

3. What works here and now might not work in other places and times –
   - Causal mechanisms that work only in particular contexts
   - Limits of scaleability
1. The evidence can be wrong
Possible reasons for false negatives or false positives from RCTs

- **Internal validity**
  - Quality issues –
    - Poor measurement,
    - Poor adherence to randomisation,
    - Inadequate statistical power,
    - Ignored differential effects,
    - Inappropriate comparisons,
    - Fishing for statistical significance,
    - Differential attrition between control and treatment groups,
    - Treatment leakage,
    - Unplanned cross-over,
    - Unidentified poor quality implementation

  Other issues –
  - Random error,
  - Contamination from other sources,
  - Need for a complete causal package,
  - Lack of blinding,
Results can be manipulated

Strategies used to manipulate results in drug trials:

- choice of placebo as comparator
- selection of subjects (Bodenheimer, 2000)
- manipulation of doses (Angell, 2004).
- method of drug administration (Bodenheimer, 2000).
- manipulation of timescales (Pollack & Abelson, 2006).
- suspect statistical analysis
- deceptive publication
- suppression of negative results (Mathews, 2005)
- selective publishing (Mathews, 2005, Armstrong, 2006; Harris, 2006; Mathews, 2005; Zimmerman & Tomsho, 2005)
- opportunistic data analysis (Bodenheimer, 2000)
- control of authorship (Bodenheimer, 2000)

Evidence can be misrepresented

Blue-eyed people better off, say scientists

By staff writers | NEWS.com.au | August 20, 2007 12:23pm

- Blue-eyed people "better strategic thinkers"
- Are "likely to achieve more in life"
- Public quick to link study with racism

PEOPLE with blue eyes are likely to achieve more in life than those with brown, say US scientists.

Scientists who conducted the tests said brown-eyed people performed better at reaction time, but those with lighter eyes appeared to be better strategic thinkers, the Daily Mail reported.

Brown-eyed people succeeded in activities such as

Why blue-eyed boys (and girls) are so brilliant

By BEN CLERKIN
Last updated at 11:33 20 August 2007

Comments (49) | Add to My Stories
The colour of your eyes could determine your achievements in life, say scientists. They claim those with blue eyes are more likely to sparkle academically than those with brown.

They are more intelligent and gain more qualifications because they study more effectively and perform better in exams.
EXAMPLE of random error: RCTs of clinical interventions

It cannot be guaranteed that randomisation will automatically produce treatment and control groups that are the same on all observable and unobservable variables.

A study that compared results from RCTs to those from observational studies found that, while the overall average effect size was similar, results from RCTs were more varied.

Some of the RCTs produced paradoxical findings (that is, in some of trials the interventions produced negative effects and in other trials the same intervention had positive effects) which could be explained by random variation between treatment and control groups in terms of contributing factors (Concato et al, 2000).


PROBLEMS WITH THE EVIDENCE

2. The evidence can be incomplete
“As with many interventions intended to prevent ill health, the effectiveness of parachutes has not been subjected to rigorous evaluation by using randomised controlled trials. Advocates of evidence based medicine have criticised the adoption of interventions evaluated by using only observational data. “

A systematic review was unable to find any randomised controlled trials of parachute intervention. (Smith and Pell, British Medical Journal, 2003)

RECOMMENDATION, FOLLOWING SYSTEMATIC REVIEW PROTOCOLS:

Parachute use should be discontinued until evidence from RCTs supports their use

“Only two options exist. The first is that we accept that, under exceptional circumstances, common sense might be applied when considering the potential risks and benefits of interventions. The second is that we continue our quest for the holy grail of exclusively evidence based interventions and preclude parachute use outside the context of a properly conducted trial.”

Impacts of After-School Programs on Student Outcomes: A Systematic Review for the Campbell Collaboration

- 110 separate potentially relevant studies were identified.
- Of these, only six were determined to be experimental design studies; three of the six met the criteria for the type of program to be included in this review
- Two others which met the criteria were added.

Main results
Based on 5 included studies, it does not appear that any one approach was more or less effective at contributing towards improved behavioral outcomes or other estimated effects.
- There is also no evidence that any one program model is more effective at changing students' context or improving academic outcomes.
EXAMPLE of synthesising a range of evidence: New Zealand Education Iterative Best Evidence Synthesis (Realist Synthesis)

- Among OECD countries, New Zealand has the biggest gap in achievement between high-achieving and low-achieving students.
- A best-evidence synthesis was undertaken of diverse evidence about how to effectively teach diverse students.
- A key piece of this was an example of one school whose students had similar demographics but were achieving significantly better outcomes.

The 2006 edition of the World Education Yearbook describes New Zealand's Iterative BES Programme "as the most comprehensive approach to evidence" and goes on to say: "What is distinctive about the New Zealand approach is its willingness to consider all forms of research evidence regardless of methodological paradigms and ideological rectitude, and its concern in finding...effective, appropriate and locally powerful examples of 'what works'".

PROBLEMS WITH GENERALISING FROM THE EVIDENCE

Addressing complication and complexity
Unhelpful ways in which the word ‘complex’ is used

Difficult
   “This evaluation is really complex”
   \textit{It is hard to collect good data}

Inadequately planned
   “This is a complex program”.
   \textit{We haven’t worked out what we’re doing yet}

Many components
   “This is a complex project”
   \textit{There are many different components/}
   \textit{This is a complicated project}
Sources for this thinking about evaluation and complexity

1997
Realistic Evaluation
Ray Pawson
Nick Tilley

2002
Commission of the Future of Health Care in Canada Discussion Paper No. 8
Complicated and Complex Systems – What would successful reform of Medicare look like?
Sholom Glouberman
Brenda Zimmerman

2003
IBM Systems Journal,
The new dynamics of strategy: Sense-making in a complex and complicated world
Cynthia Kurtz
Dave Snowden

2004
Evaluation
Theory-Based Evaluation and Types of Complexity
Nicoletta Stame
University of Rome ‘La Sapienza’, Italy

2006
Harvard Business Review
Leader’s Framework for Decision Making
Dave Snowden
Mary Boone

2007
Evaluation in Complex Adaptive Systems
Glenda Eoyang
Thomas Berkas

2007
Getting to Maybe
Frances Westley
Brenda Zimmerman
Michael Quinn Patton

2008
Evaluation Using Programme Theory to Evaluate Complicated and Complex Aspects of Interventions
Patricia Rogers
Sources for this thinking about evaluation and complexity

2008
Ben Ramalingan
Harry Jones

2009
NORAD conference
Evaluating the complex
Oslo, Norway

2010
Developmental Evaluation
Michael Quinn Patton

2011
Purposeful Program Theory
Sue Funnell
Patricia Rogers

Evaluating the complex
Kim Forss, Mita Marra, Robert Schwarz (eds)
Simple, complicated and complex — Glouberman and Zimmerman

Glouberman and Zimmerman 2002

**Simple**
- Tested ‘recipes’ assure replicability
- Expertise is not needed

**Complicated**
- Success requires high level of expertise in many specialized fields + coordination

**Complex**
- Every situation is unique – previous success does not guarantee success
- Expertise can help but is not sufficient; relationships are key

Simple, complicated and complex — Kurtz and Snowden

Kurtz and Snowden 2003

Simple
The domain of the ‘known’,
Cause and effect are well understood,
Best practices can be confidently recommended,

Complicated
The domain of the ‘knowable’
Expert knowledge is required,

Complex
The domain of the ‘unknowable’,
Patterns are only evident in retrospect.

Chaotic
Patterns cannot be seen even in retrospect

Disordered
Cannot determine the nature of the system

# Implications of simple, complicated and complex for evaluation

<table>
<thead>
<tr>
<th>What do interventions look like?</th>
<th>Simple</th>
<th>Complicated</th>
<th>Complex</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Discrete, standardized intervention</td>
<td>Interventions that have multiple components, or that only work in conjunction with other interventions or favorable</td>
<td>Nonstandardized and changing, adaptive, and emergent in response to changing needs, opportunities and understandings of what is working</td>
</tr>
<tr>
<td>Who implements them?</td>
<td>Single organization</td>
<td>Multiple identifiable organizations in predictable ways</td>
<td>Multiple organizations with emergent and unpredictable roles</td>
</tr>
<tr>
<td>How do interventions work?</td>
<td>Pretty much the same everywhere</td>
<td>Differently in different situations (different people or different implementation environments), which can be clearly identified</td>
<td>Generalizations rapidly decay, results are sensitive to initial conditions as well as to context</td>
</tr>
</tbody>
</table>
Using the framework

Can be used to refer to a situation or to an intervention
Not useful as a way of classifying the whole situation or intervention
  - most useful to consider aspects of interventions
Not normative
  - complex is not better than simple
  - simple interventions can still be difficult to do well, or to get good data about

“Everything should be made as simple as possible, but no simpler.”
Possible reasons why finding out what works and doing it might not produce better results

3. What works here and now might not work in other places and times
# NECESSARINESS (STATE OF BEING NECESSARY)

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simple</td>
<td>Only way to achieve the intended impacts</td>
</tr>
<tr>
<td>Complicated</td>
<td>One of several ways to achieve the intended impacts – which can be identified in advance</td>
</tr>
<tr>
<td>Complex</td>
<td>One of several ways to achieve the intended impacts – which are only evident in retrospect</td>
</tr>
</tbody>
</table>
Necessariness – with/without comparisons for complicated interventions

A US program to assist poor families through linking them to services found that families receiving the program experienced improvements in welfare — but so did the families that were randomly assigned to a control group that did not receive the visits (St. Pierre and Layzer 1999).

[As this case shows], a good study helps avoid spending funds on ineffective programs and redirects attention to improving designs or to more promising alternatives.’ (When Will We Ever Learn?)

But families in the control group had also accessed services..

The appropriate comparison would have been to compare the costs incurred in the different groups

SUFFICIENCY

Simple  Sufficient to produce the intended impacts. Works the same for everyone

Complicated  Only works in conjunction with other interventions (previously, concurrently, or subsequently) and/or only works for some people and/or only works in some circumstances – which can be identified in advance

Complex  Only works in conjunction with other interventions (previously, concurrently, or subsequently) and/or only works for some people and/or only works in some circumstances – which is only evident in retrospect
Sufficiency – the potted plant thought experiment

If 200 potted plants are randomly assigned to either a treatment group that receives daily water, or to a control that receives none,

and both groups are placed in a dark cupboard,

the treatment group does not have better outcomes than the control.

Possible conclusions: Watering plants is ineffective in making them grow.

Better conclusion: Water is not sufficient.
Failure to address combined attribution – 2nd potted plant thought experiment

If 200 potted plants are randomly assigned to either a treatment group that receives daily water, or to a control that receives none,

and both groups receive light,

the treatment group has better outcomes than the control.

Inappropriate question: What proportion of survival and growth is due to the water?
Ways in which an intervention can work with other interventions

Figure 8: Five ways in which Strategy projects worked together with other interventions

1. Building on a previous activity
   - Something before
   - Strategy project

2. Benefiting from a concurrent project
   - Strategy project
   - Concurrent project

3. Jointly funded through another program
   - Jointly funded project
   - Other program funding

4. Strategy project part of a larger project
   - Larger project
   - Strategy project

5. Laying foundation for subsequent activity
   - Strategy project
   - Something after
Complicated necessariness

Intervention that works differently for different types of participants

- **INTERVENTION**
  - PARTICULAR PARTICIPANT CHARACTERISTICS
- **DIFFERENT INTERVENTION**
  - DIFFERENT PARTICIPANT CHARACTERISTICS

**SHORT-TERM OUTCOMES**

**LONG-TERM OUTCOMES**
Sufficiency – Early Head Start

- Early Head Start program - on average effective. Listed as an ‘evidence-based program’


## Change trajectory

<table>
<thead>
<tr>
<th>Simple</th>
<th>Constant, linear relationship between effort and results (eg twice the investment produces twice the results)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complicated</td>
<td>Well understood but not linear relationship between effort and results (eg curvilinear dose-response relationship such as diminishing returns or too much of a good thing)</td>
</tr>
<tr>
<td>Complex</td>
<td>Emergent relationship between effort and results (eg unknown tipping points)</td>
</tr>
<tr>
<td>Type</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Simple</td>
<td>Single set of intended outcomes/impacts</td>
</tr>
<tr>
<td>Complicated</td>
<td>Different intended outcomes/impacts intended by different partners/stakeholders</td>
</tr>
<tr>
<td>Complex</td>
<td>Emergent intended outcomes/impacts</td>
</tr>
</tbody>
</table>

Are there agreed objectives? Different objectives? Emergent objectives?
What are the implications for evaluating the intervention?
Complicated focus

Intervention that produces different outcomes valued by different stakeholders
What works here and now might not work in other times and places

• Sufficiency – what other parts of the causal packaged are needed?
• Change trajectory – does the causal relationship hold at different levels of intensity?
• What is valued – are the same evaluative criteria relevant?
• Scaleability – does the causal relationship hold when the intervention is scaled up?
Knowledge transfer – seeing evidence to policy as a leaky pipeline

Figure. Glasziou’s Pipeline Model (2004).
Knowledge translation (Nature June 2008)

Translational research

Basic biomedical research is booming. But its impact, in terms of new therapies and diagnostics, is growing far more modestly. 'Translational research' is seen as a solution to this disparity, ensuring that the bounty of discoveries is effectively 'translated' into benefits in the everyday world of medicine. *Nature* puts this new discipline in the spotlight, with a special package of features and comment.
### Need to translate ‘what works’

<table>
<thead>
<tr>
<th>Simple</th>
<th>Implement what has been identified as ‘best practice’ or ‘evidence-based practice’ – what works</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complicated</td>
<td>Classify the situation and implement what has been adapted for that context – what works for whom in what situation</td>
</tr>
<tr>
<td>Complex</td>
<td>Ongoing adaptation to emerging conditions – what is working here</td>
</tr>
</tbody>
</table>
Are different theories of change needed for different contexts or just different theories of action?

<table>
<thead>
<tr>
<th>Situation analysis</th>
<th>Theory of action</th>
<th>Theory of change</th>
<th>Intended result</th>
</tr>
</thead>
<tbody>
<tr>
<td>A tidy room</td>
<td>Deterrence – avoiding a negative interaction</td>
<td>Positive incentive</td>
<td>A tidy room</td>
</tr>
<tr>
<td></td>
<td>Withhold pocket money</td>
<td>Offer a reward (money, movie tickets)</td>
<td>A tidy room</td>
</tr>
<tr>
<td></td>
<td>Shouting</td>
<td>Praise efforts</td>
<td>A tidy room</td>
</tr>
</tbody>
</table>
Examples of adapting practice to context

Simple – best practice – “what works”
Teachers select a reading program which has been shown in RCTs to be effective (e.g., USA Reading First program - $1b p.a.)

Complicated – adapted – “what works for whom”
Teachers identify children’s learning stage and provide exercises to match this (e.g., Victorian Catholic Education Systems Literacy Assessment Project)

Complex – adaptive – “what is working?”
Teachers monitor students’ progress and respond to emerging problems and opportunities

Griffin, P. 2009 ‘Ambitious new project to raise literacy and numeracy levels in Victorian Schools.’ http://newsroom.melbourne.edu/studio/ep-29
EVIDENCE BASED POLICY = LEARNING FROM DIFFERENT SOURCES WHAT IS NEEDED AND POSSIBLE, WHAT WORKS FOR WHOM IN WHAT SITUATIONS, AND WHAT’S WORKING AND SUPPORTING TRANSLATION, ADAPTATION AND ONGOING LEARNING