



# CEM

Centre for Evaluation & Monitoring

“I’d like to see the evidence for  
that...”

*A talk to Future Leaders on the need for evaluation in education*

Stuart Kime

# What I will talk about

- What robust evaluation of interventions in education can offer you
- Why I think evaluation is a key component of systemic school improvement
- Why I think Future Leaders can help set the agenda for empirical educational research
- Why the agenda is a troublesome one

# What's my angle?

- 10 years as a teacher
- 5 years doing research in education
- Policy Fellow at DfE
- CEM research consultant
- Methods
- Pragmatic learner
- Optimist

# Why do we need evaluation?

- *Approximately, how many police officers are there in France (to the nearest 1,000)?*
- Two treatment conditions, 'randomly' assigned
  - A
  - B

# Treatment A

- The USA has 794,300 police officers.
- How many are there in France?

# Treatment B

- Finland has 7,800 police officers.
- How many are there in France?

# Treatment group averages?

- Treatment A
- Treatment B
- And the actual number of police officers in France?

# 220,000 police officers in France

- We need RCTs
- Bias is everywhere
- Anchoring effects
  - These effects occur when we try to think of a value for an unknown quantity before estimating that quantity
    - Kahneman, D. (2011) *Thinking, Fast and Slow*

# ‘Is Brain Gym an Effective Intervention?’

- “If we are unwilling to subject children and adults to experimental medical treatments that have not been validated through high quality, empirical research over a long period of time, we see no reason to be any more willing to subject students in our schools and classrooms...to educational interventions that remain similarly untested and not validated.” (Spaulding et al., 2010)

# What robust evaluation of interventions in education can offer you

- Large-scale, randomised trials
  - What works *there*
- Small-scale (“DIY”) trials (preferably randomised)
  - What works *here*
- Causal claims require randomised trials

# Why I think evaluation is a key component of systemic improvement

- Give me an example of a strategy / technique which improves children's maths attainment scores (primary or secondary school).
- How well does it work?
- What can evaluation tell us?

# DIY Evaluation

- Small-scale
- Low-cost
- Designed and implemented by teachers and school leaders
- Begins to address the ‘intention implementation gap’

# Why I think the Future Leaders can help set the agenda

- Small-scale DIY evaluations done well *could* be ‘knitted together’
- Cyclical process of piloting, scaled-up trialling and implementation driven by teachers and school leaders
- Embedding evaluation at the beginning rather than adding as an afterthought
- Leadership is crucial

# Marginal gains

- Acknowledge that we often don't know what works / the degree to which something works
- Recognise that evaluation offers tools to answer causal inference questions, but it's not a panacea
- Accept responsibility for the development of better evaluation practices in our own work contexts

# Threshold concepts

- 'Threshold concepts' lead to new and previously inaccessible ways of thinking about something (Meyer and Land, 2003)

# Threshold concepts

- “As a consequence of comprehending a threshold concept there may thus be a transformed internal view of subject matter, subject landscape, or even world view. Such a transformed view or landscape may represent how people ‘think’ in a particular discipline, or how they perceive, apprehend, or experience particular phenomena within that discipline, or more generally.” (Land, 2008)

# Liminality (thanks to Ray Land)

- Existing certainties transform and are problematised
- A state of growing understanding characterised by imitation, trial & error
- Liminality is unsettling – sense of loss of the familiar, without a firm replacement

# Troublesome knowledge

- “Knowledge is troublesome for a variety of reasons (Perkins 2006). It might be alien, inert, tacit, conceptually difficult, counter-intuitive, characterised by an inaccessible ‘underlying game’, or characterised by supercomplexity.
- Such troublesomeness and disquietude is purposeful, as it is the provoker of change that cannot be assimilated, and hence is the instigator of new learning and new ontological possibility.” (Land, 2008)

# This is not easy...

- Evaluation creates troublesome knowledge
- Evaluation leads you into liminal spaces
- Evaluation requires certain threshold concepts

# But...

- You are ahead of an interesting curve
- You need to be ready to lead others through the liminal spaces
- You need to help others navigate troublesome knowledge
  
- But that's what Future Leaders do

# Stuart's final 3 points

- Recognise the need for evaluation and the crucial role Future Leaders have to play
- Recognise that evaluation is not a silver bullet
- Recognise that evaluation is challenging, troublesome and unsettling for a start, but that it offers a promising, rewarding way forward for education



# CEM

Centre for Evaluation & Monitoring

## Thank you.

[stuart.kime@cem.dur.ac.uk](mailto:stuart.kime@cem.dur.ac.uk)

@StuartKime

@CematDurham