SYNTHESISING KIRKPATRICK'S FOUR LEVELS

Fran Demetriou

19 September 2018

Acknowledgement of country

I acknowledge the Palawa people who are the traditional custodians of the land on which we meet today and pay my respects to Elders past, present and emerging.
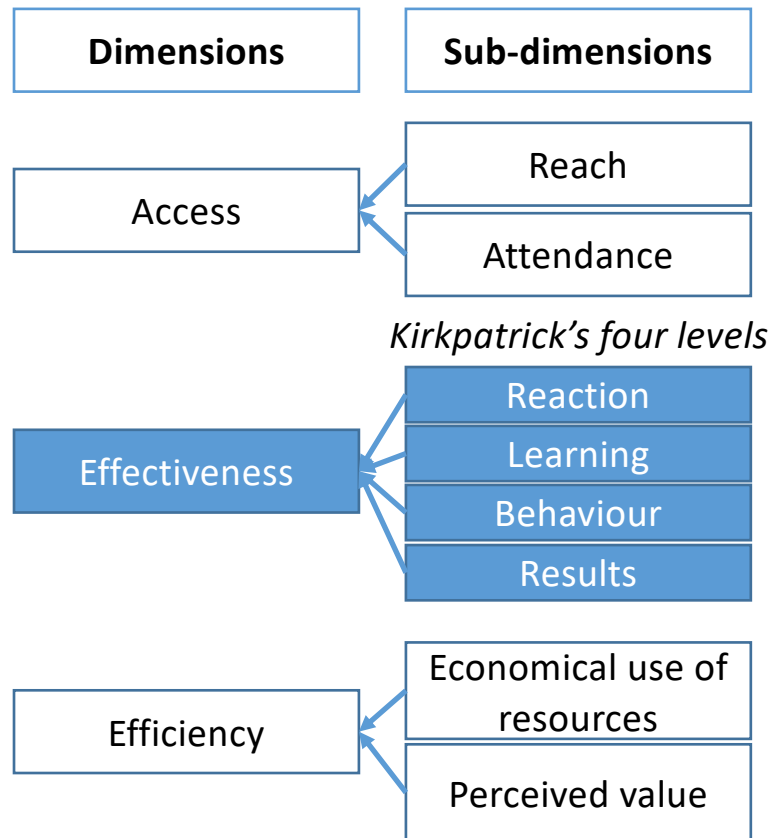
IN THIS SESSION

- The Emerging Community Leaders (ECL) program
- The evaluation design
- Evaluation synthesis – introduction/refresher
- Kirkpatrick's four levels and the considerations we made for determining importance for the ECL program
- Key takeaways and further thinking
- Questions

3

EMERGING COMMUNITY LEADERS

- A community sector leadership program

- Tasmanian Community Fund (TCF) identified a need for developing future community leaders, especially in rural and regional communities – no other leadership programs available in Tasmania that specifically target the community sector.

- Funded by TCF and delivered by SRA Corporate Change, in partnership with TCF

- Program offered each year to 24 participants, for five years.

- Program aim: *"to deliver an effective leadership program for employees and volunteers working in the community sector that will benefit participants, their employers or businesses, and ultimately the Tasmanian community at large."*

- Program vision: *"To create a highly valued and sought-after leadership program for the community sector"*

4

DIMENSIONS TO EVALUATE SUCCESS

Dimensions	Sub-dimensions

Access → Reach

Access → Attendance

Kirkpatrick's four levels

Effectiveness → Reaction / Learning / Behaviour / Results

Efficiency → Economical use of resources

Efficiency → Perceived value

- Evaluation purpose: Evaluate the success of the program in developing leadership capacity to its intended cohorts, and identify opportunities to improve.

- Kirkpatrick's four levels:
  - **Reaction**: The extent that participants find the program favourable, relevant and engaging.
  - **Learning**: The extent that participants acquire intended knowledge, skills and/or attitudes.
  - **Behaviour**: The extent that participants apply learnings back on the job.
  - **Results**: The extent tangible outcomes occur as a result of the training.

- Our challenge: Synthesising our findings for each of the four levels into a warranted conclusion about the **effectiveness** of the program within the constraints of a real-world evaluation.

SYNTHESIS REFRESHER

- Put simply… it's bringing together multiple findings about a program in a systematic way to make a judgement about its performance.
- Involves two keys steps (Davidson, 2005):
  - 1) Setting performance standards – what do different levels of performance look like for the dimensions you are evaluating the program on?
  - 2) Importance determination – what is the relative importance of the dimensions you are evaluating the program on?
- Without a synthesis methodology you are either:
  - Presenting a conclusion without being explicit about the values you have used to reach it (Owen and Rogers, 1999; Nunns, Peace and Witten, 2015)
  - Not presenting any conclusion and leaving it to the reader to make their own determination – a "cop-out disguised as or rationalized as objectivity" (Scriven, 1971, p53)
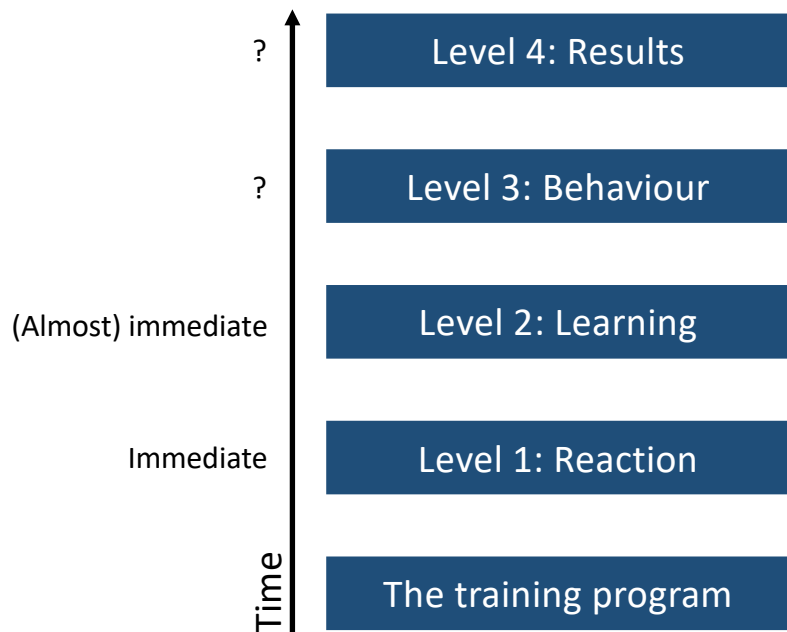
6

SYNTHESIS AND KIRKPATRICK LEVELS

Performance 1

	Poor	Adequate	Good	Excellent
Results				
Behaviour				
Learning				
Reaction				

Performance 2

	Poor	Adequate	Good	Excellent
Results				
Behaviour				
Learning				
Reaction				

Which performance can be considered better?

That depends on the relative importance of the four levels.

1. TIMEFRAME EXPECTATIONS

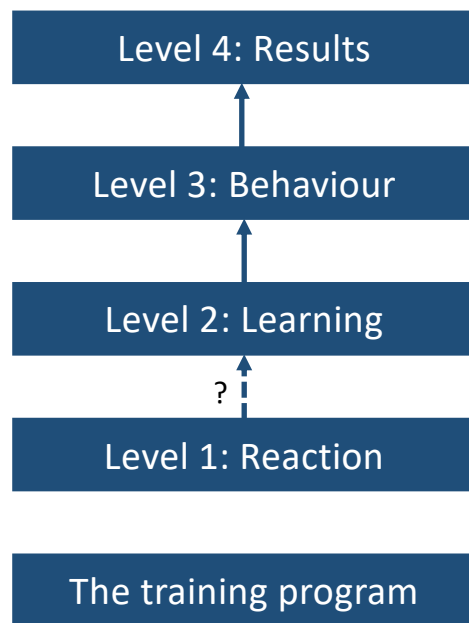| Time	
?	Level 4: Results
?	Level 3: Behaviour
(Almost) immediate	Level 2: Learning
Immediate	Level 1: Reaction
	The training program

- Issue: The time lapse between the training program and *behaviour/results* levels is unknown.

- Consideration: Client/stakeholders may have **expectations** about when behaviour/results should occur. These expectations may be **critical** or **non-critical** to program success.

- Implications:
  - A lack of findings at the *behaviour* and *results* levels may indicate lack of effectiveness OR a lack of sufficient data at this point in time.
  - Where clients/stakeholders have critical timeframe expectations for effects at these levels, a lack of positive findings will have a more detrimental effect on our assessment of effectiveness.

- Evaluation design: Evaluations should mitigate for this as best they can within budget and time constraints. If time constraints mean possible findings cannot be captured fully, this should be factored into the importance determination of the level. Limited evaluation designs will mean a lack of findings is less detrimental on our assessment of effectiveness.

2. ESSENTIAL LINKS IN THE CHAIN

Level 4: Results

↑

Level 3: Behaviour

↑

Level 2: Learning

? ↑

Level 1: Reaction

The training program

- Issue: *Learning* is essential for *behaviour* to occur as a result of the program. *Behaviour* is essential for *results* to occur as a result of the program.

- Consideration: But is *reaction* essential for *learning*? Contested in the literature and may influence motivation to learn.

- Implication: Findings for the essential levels (*learning* and *behaviour*) are critical to effectiveness. *Reaction* needs to be considered on a case by case basis.
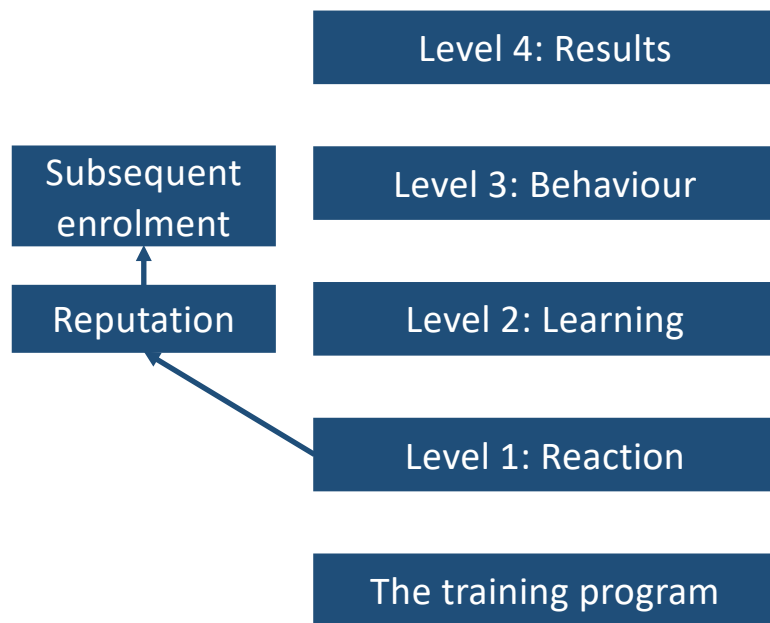
3. BARRIERS AND ENABLERS

Level 4: Results

Level 3: Behaviour

Level 2: Learning

Level 1: Reaction

The training program
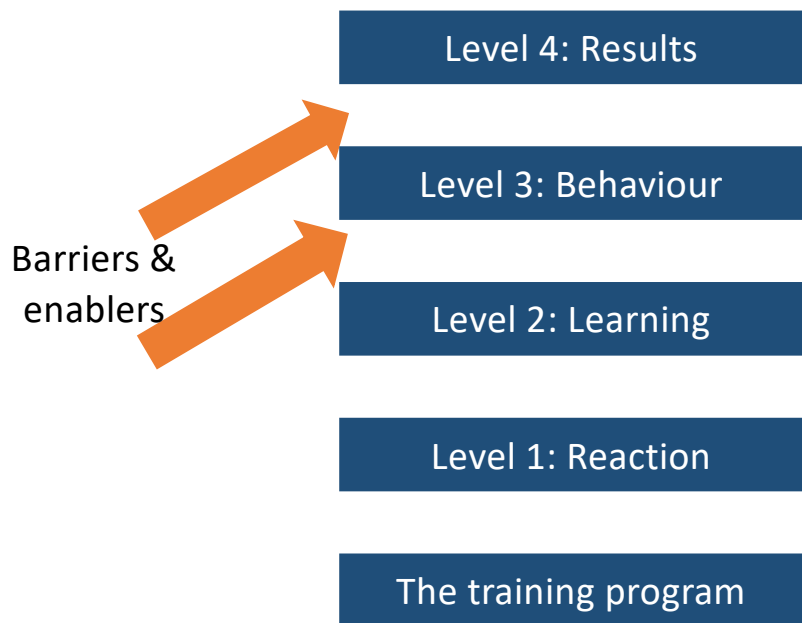
Barriers & enablers

- Issue: *Behaviour* and *results* operate externally to the program – and are influenced by barriers and enablers in participants' own contexts.

- Consideration: What is the role of the training program in influencing these factors?

- Implication: Low effects on *behaviour* and *results* should not be detrimental to the training program's effectiveness is it has not been designed to influence these.
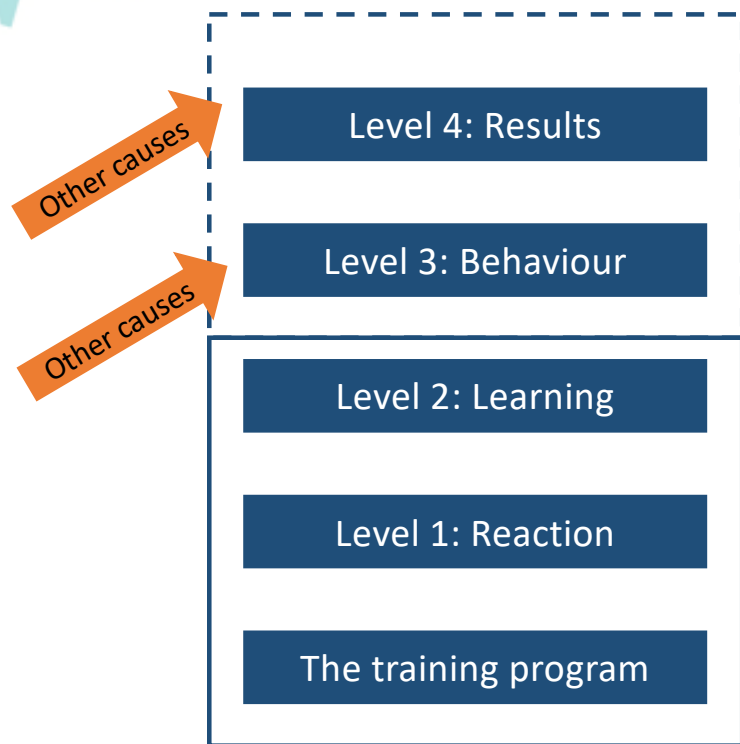
4. REACTION – OTHER IMPLICATIONS

Level 4: Results

Level 3: Behaviour

Level 2: Learning

Level 1: Reaction

The training program

Subsequent enrolment

Reputation

- Issue: Reaction may have reputational implications affecting subsequent enrolment. This is influenced by the program's context (such as whether it is delivered internally or externally, is mandatory or optional, and if it has competition)

- Implication: Reaction becomes more important when there are reputational implications.

5. ATTRIBUTION

Level 4: Results

Other causes

Level 3: Behaviour

Other causes

Level 2: Learning

Level 1: Reaction

The training program

- Attribution: Attributing *behaviour* and *results* findings to the training program are more challenging due to their existing externally to the program. Evaluations mitigate for this as best they can within time and budget constraints.

- Implication: Less robustly attributable findings are less beneficial to a conclusion about overall effectiveness of the training program.

12

CONSIDERATIONS FOR IMPORTANCE DETERMINATION

| Context considerations	Evaluation design reliability considerations
1. What is the program's level of control/influence over this level?	1. How sufficiently does the evaluation design factor in expected and unknown timeframes?
2. What are the implications of participant reaction on program reputation?	2. How strong is the evaluation design for making attribution claims?
3. What is the expected timeframe for effects at this level? How critical is this timeframe for program success?	
4. How essential is this level for effects on latter levels?	

A MATRIX TO DETERMINE IMPORTANCE

Taken from Davidson (2005)

		How detrimental… would it be overall if the program did very poorly on this dimension?		
		Not noticeably detrimental	Noticeably detrimental	Unacceptably detrimental
How beneficial… would it be overall if the program did very well on this dimension?	Somewhat beneficial	Somewhat important	Important	Important (and set a bar)
	Very beneficial	Important	Very important	Very important (and set a bar)
	Extremely beneficial	Very important	Extremely important	Extremely important (and set a bar)

EMERGING COMMUNITY LEADERS – YEAR 1 RESULTS
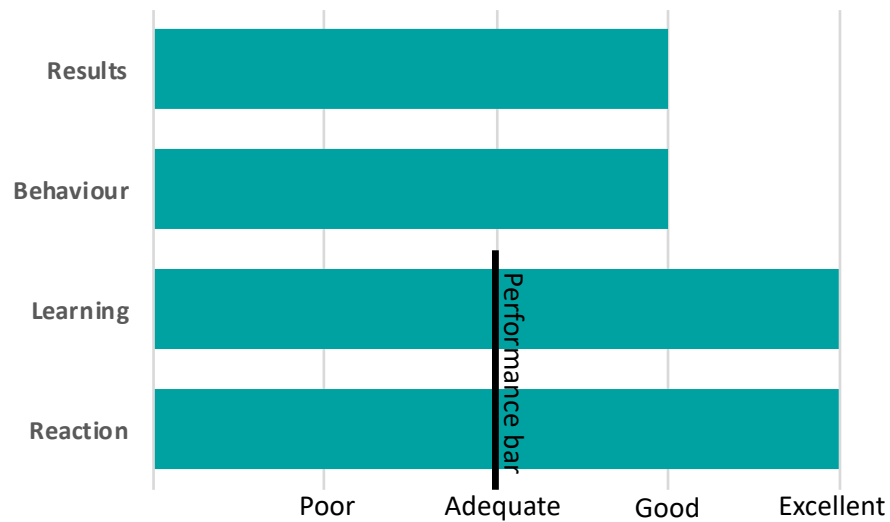
Level	Importance determination	Relative importance	Bar
Results	Important	Least	N/A
Behaviour	Very important	Middle	N/A
Learning	Extremely important	Most	Adequate
Reaction	Very important	Middle	Adequate

Our judgement: Nearing excellence!

KEY TAKEAWAYS AND FURTHER THOUGHTS

- This is a work-in-progress! Further testing and development is required!
- Useful for considering… in the context of this training program, and within the constraints of this program, how effective can we say the training program is?
- Helpful for keeping the Kirkpatrick four-levels a simple four-level framework while still factoring in some of the underlying assumptions and influences to those levels.
- Issues still to consider:
  - The interplay between the different considerations (e.g. barriers and enablers and time lapse)
  - When time and budget enables further data capture of assumptions, how then to incorporate into the synthesis?
  - Methodological pureness of mixing values claims in with evaluation design reliability issues – potential to separate these out?

THANK YOU

Licata Ltd is an independent not-for-profit organisation based in Melbourne, Australia. We support the work of individuals and organisations who are responding to those in need and developing constructive solutions to social issues.

Our key business areas include:

- **Monitoring and evaluation**
- **Strategic and operational planning**
- **Organisational assessment and review**
- **Organisational development**
- **Research**
- **Data systems development**

Fran Demetriou - Evaluator

fran.d@lirata.com

www.lirata.com

lirata

CONSULTING | CAPABILITY | CREATIVE

REFERENCES

Alliger G M. and Janak E. A. (1989), Kirkpatrick's levels of training criteria: thirty years later, Personnel Psychology, Vol. 42 (2) pp. 331-342

Davidson, E. J. (2005b). Evaluation Methodology Basics: The Nuts and Bolts of Sound Evaluation. Sage Publications. ISSN: 0761929290

Holton, E. F. (1996). The flawed four-level evaluation model. Human Resource Development Quarterly, 7(1), 5-21. http://dx.doi.org/10.1002/hrdq.3920070103 K

Kirkpatrick, D.L. (1994). Evaluating Training Programs, Berrett-Koehler Publishers

Kirkpatrick, D.L., & Kirkpatrick, J.D. (2006). Evaluating Training Programs, Berrett-Koehler Publishers

Kirkpatrick, D.L., & Kirkpatrick, J.D. (2007). Implementing the Four Levels, Berrett-Koehler Publishers

Nunns, H., Peace, R., & Witten, K. (2015). Evaluative reasoning in public-sector evaluation in Aotearoa New Zealand: How are we doing? Evaluation Matters—He Take Tō Te Aromatawai, (1), 137–163. https://doi.org/10.18296/em.0007

Owen, J. M., & Rogers, P. J. (1999e). Evaluation fundamentals. In Program Evaluation: Forms and Approaches (International ed., pp. 2–22). Sage

Schneier, C. E. (Ed.). (1994). The training and development sourcebook. Human Resource Development.

Scriven, M. (1980). The Logic of Evaluation. Edgepress.

Scriven, M. (1986). New Frontiers of Evaluation, American Journal of Evaluation, Vol 7 (1), pp. 7 -44

Scriven, M. (1994) The final synthesis, Evaluation Practice, vol. 15, no. 3, pp 367-382

Shadish, W. R., Cook, T. D., & Leviton, L. C. (1991e). Michael S. Scriven: The Science of Valuing. In *Foundations of Program Evaluation: Theories of Practice* (pp. 73–118). SAGE. ISBN: 080393551X

Sloman, M. (1994). A Handbook for Training Strategy.

Tamkin, P., Yarnall, J., & Kerrin, M. (2002). Kirkpatrick and Beyond: A Review of Models of Training Evaluation. IES Report.

Warr P. B., Allan, C., Birdi, K. (1999) Predicting three levels of training outcome, *Journal of Occupational and Organizational Psychology,* Vol 72, pp. 351-375