

STEP 1

Do the findings and recommendations display the 3 characteristics of good reasoning?

1 Does it offer reasons for the conclusion?

2 Are the premises offered acceptable to the evaluation's audiences? Does the warrant for each conclusion hold? Think about what the warrant/s is of your particular evaluation approach (for examples, see the table below).

3 Does it contain enough relevant information required for the evaluation's audience to find the conclusion acceptable?

Table 2.1. Comparison of Different Definitions for the Same Phenomenon

<i>Evaluation Approach</i>	<i>Phenomenon of Interest</i>	<i>Source of Criteria (Locus of Values)</i>	<i>Evidence (Foundation for a Claim)</i>	<i>Warrant (authorizes Inference)</i>
Connoisseurial/critic approach to program evaluation	Program defined as set of qualities identifiable by an expert	Personally held values of an expert	Expert values	Expert is reliable and credible
Pluralistic approach to program evaluation	Program defined as set of values held by stakeholders	Stakeholder values	Stakeholder values and their connection to impact	Stakeholder values reflect what is desirable and important
Consumer approach to product evaluation	Functional product	Properties inherent in the product and consumer use	Properties and their connection to extent of performance	Accepted meaning of the word (such as car or watch)
Goal-free approach to program evaluation	Program defined as a means of meeting needs	Consumer needs	Needs and their connection to program effects	Needs accepted as necessary requirements for existence
Causal approach to program evaluation	Program defined as set of treatment-outcome relationships	Dependent variables in goals or research literature	Relationships among variables	Relationships were identified under reliable methods

Source: Fournier, D. (1995) Establishing Evaluative Conclusions: a distinction between general and working logic. *New Directions for Evaluation* (68)

STEP 2

Making the connections between the program theory, literature, claim and warrant explicit

What does the program theory say?



What is the warrant for the judgement?

(the principle, rule or assumption that justifies a judgement about value or underpins recommendations. This could include assumptions embedded within the evaluation approach)



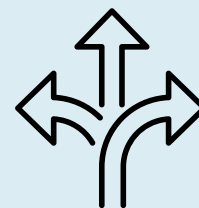
What does the data say? (the 'what')

- Qualitative
- Quantitative
- Literature sources



What is the judgement/recommendation?

(the 'now what')



What is the argument put in its most basic terms?

(the 'so what'. I.e. the premises leading to the conclusion? e.g. "Because of X, some Y experienced Z)

What is the warrant for the argument?

The warrant is the principle, rule or assumption that justifies the claims made about what the data means. This might be a principle of human behaviour drawn from literature, the general logic of evaluation, a working logic or another type of explicit or unspoken rule, principle or assumption or several working together)



STEP 3

Strengthen reasoning, and check for logical fallacies

a) Check if any of these are potentially occurring in any of the above columns (note: it is worthwhile checking for these while designing program theory/logic also!)

Logical Fallacy	Definition	Evaluation example
Composition fallacy	Assuming that what's true of the parts must be true of the whole	Drawing unwarranted conclusions about the generalised population from data about specific groups/contexts.
Ignoratio elenchi (irrelevant conclusion)	Making an argument which may be valid but which fails to address the issue at hand	Judging a program to be successful on the basis that it achieves benefits, despite a lack of connection between the outcomes and the original problem that prompted the need for the program.
Texas Sharpshooter	Drawing a target around an existing bullethole, i.e. cherry-picking data or overemphasising patterns to fit a particular claim	Selecting data to fit an existing theory, without sufficient warrant (e.g. overemphasising the views of small numbers of stakeholders).
Affirming the consequent	Assuming a particular cause because a particular outcome is present. 'If P leads to Q. Q. Therefore, P.'	Overlooking other possible causes for an outcome, e.g. If we improve children's confidence to play outdoors, they will be more physically active. Children are more physically active, therefore their confidence to play outdoors has improved.
Middle Ground/ Golden Mean Fallacy	Assuming that because something lies in the middle of extremes it must be true	Recommending a change to a program based on it being the middle ground between two groups of stakeholders opposing views, and assuming it is the best solution because of this.
False binary/ the Black and White fallacy	Assuming there are only two options when there are more or falsely presenting two options where more are possible	Providing a recommendation on the assumption that the options are only to continue to fund vs. stop funding, when other options are possible (e.g. scale down/ absorb elements into other programs)
Special Pleading	Moving the goalposts when the answer isn't the desirable one	Making a judgement that a program would 'usually' be valuable despite a lack of supporting evidence for any benefit because it was delivered in a 'special/unusual' set of circumstances (e.g. COVID). Changing the KEQs or outcomes under examination to better fit the findings without noting this was done or sufficiently explaining why.

Or are you making the fallacy fallacy? (the assumption that because the argument contains a fallacy or is lacking some of the elements of a good argument, the argument itself is not just poorly made, but wrong)

[illegible]

Can you identify the 3 characteristics of good reasoning?	The problem, the program, the evaluation Parents of children within several local schools identified concerns about their children’s time on devices, lack of interest in physical activity, and lack of confidence in conversing with adults outside their immediate families. In response, a program was put together by a local mental health organisation TogetherWell (TW), designed to support improvements in the confidence and mental health of children through a 12 week roster of nature play and outdoor challenges. The program is delivered through a peer-to-peer facilitation model, with a trained staff member delivering training to members of 5 school communities on how to run the program. This allows it to achieve larger reach, and also to create positive connections between children and adults in their school community. The purpose of the evaluation of the program was to understand lessons, outcomes and implications for scaling up. The evaluation overview and findings The evaluation approach The evaluation took a participatory, developmental approach, in line with the community-development and community-led approach of the Leaping Lizards (LL) program. A program logic was initially developed with input from the TW staff. This was later revisited during implementation, with TW staff, and the trained program volunteers. Findings Analysis shows the LL program delivered positive outcomes, including the three key outcomes that were contained in the final program logic, after its review by program volunteers. There was consensus among the TW staff members that the program had a positive effect on adults and children involved, and that the program was ready for further roll out. There were challenges with implementation, including retaining enough trained parents and school staff to deliver the 1:5 children ratio - which impacted on the number of children who were able to attend. Nevertheless the program had a strong impact on those who did attend.	Identify any logical fallacies
---	---	--

Attrition of trained parents and school staff from the program likely reflects external factors outside the control of TW, including a local outbreak of measles that occurred just prior to the first planned outdoor play session, and a disruptive weather event half-way through the program period (which also impacted on several of the intended locations for nature challenges).

The program had a positive impact on children's physical activity levels and confidence to play outdoors. The program logic identified that improving children's confidence to play in and interact with their local environment would increase their time spent being active outdoors, thereby reducing time spent on devices. Pre- and post-surveys of all children showed that on average, the participants enrolled spent 1.5 hours less per week on devices compared to baseline.

88% (7) of program volunteers said that they thought participants' confidence in engaging with other adult members of the school community had improved 'somewhat' or 'very much'. Parents of the participants were less certain of whether or not this was the case, however the program volunteers had more opportunity to observe this outcome. It is likely that there was some change in confidence in engaging with adult members of the school, as observed by the program volunteers, but that this may have been less visible in the home.

While the program target for the number of participants (n=100, 20 per school) was not reached, due to attrition of the program volunteers, the program had a deeper impact than anticipated on the 45 children who did participate in the program. The program volunteers who did commit to the full length of the program were highly dedicated, and instead of 1 hour per week, extended the sessions to 90 minutes. This meant that physical activity targets for the participants were well exceeded.

Recommendation

As the peer-to-peer model of the program is low cost for the benefits it delivers (improvements in connection to nature, confidence in engaging with adults, and interest in outdoor play), it is recommended that TW resource the program model to be rolled out in additional schools, and delivered longer-term, with regular training sessions to onboard and build the volunteer base.

Fallacies found?

Report section	Potential fallacy
Evaluation approach – revision of program logic without explanation of what changed	Special pleading
Findings – There was consensus among the TW staff members that the program had a positive effect on adults and children involved	Texas sharpshooter
Attrition for trained parents and school staff from the program likely reflects external pressures outside the control of TW...	Special pleading
The program had a positive impact on children’s physical activity levels and confidence to play outdoors. The program logic identified that improving children’s confidence to play in and interact with their local environment would increase their time spent being active outdoors, thereby reducing time spent on devices. Pre- and post-surveys of all children showed that on average, the participants enrolled spent 1.5 hours less per week on devices compared to baseline.	Affirming the consequent (if P>Q. Q, therefore P)
88% (7) of program volunteers said that they thought participants’ confidence in engaging with other adult members of the school community had improved ‘somewhat’ or ‘very much’. Parents of the participants were less certain of whether or not this was the case, however the program volunteers had more opportunity to observe this outcome.	Affirming the consequent (if P>Q. Q, therefore P)
It is likely that there was some change in confidence in engaging with adult members of the school, as observed by the program volunteers, but that this may have been less visible in the home.	Golden mean (i.e. truth lies somewhere between program volunteers and parents observations)
While the program target for the number of participants (n=100, 20 per school) was not reached, due to attrition of the program volunteers, the program had a deeper impact than anticipated on the 45 children who did participate in the program. The program volunteers who did commit to the full length of the program were highly dedicated, and instead of 1 hour per week, extended the sessions to 90 minutes. This meant that physical activity targets for the participants were well exceeded.	Golden mean (i.e. truth lies somewhere between program volunteers and parents observations)
Recommendation As the peer-to-peer model of the program is low cost for the benefits it delivers (improvements in connection to nature, confidence in engaging with adults, and interest in outdoor play), it is recommended that TW resource the program model to be rolled out in additional schools, and delivered longer-term, with regular training sessions to onboard and build the volunteer base.	Composition fallacy – assumes the issue identified by parents originally is the same at other schools False binary – there are different and perhaps more appropriate scale options available than scaling out.

Others?