How to Raise the Standards of Outdoor Learning and Its Research

Summary of 'The Existing Evidence-Base About the Effectiveness of Outdoor Learning', by Fiennes et al.

Rolf Jucker

In partnership with the Institute of Outdoor Learning, the Blagrave Trust commissioned Giving Evidence and The Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) at UCL Institute of Education to produce a systematic review of the existing literature about outdoor learning (www.giving-evidence.com/outdoor-learning, accessed 12/02/2021). The Institute of Outdoor Learning (a membership body of practitioners) and the Blagrave Trust (a funder) wanted to understand the following in order to improve their co-ordination of activities and their funding in this area:

1. Categorise the various outdoor learning (OL) activities being run in the UK, in order to provide a coherent sense of the sector as a whole;
2. Identify the various outcomes which organisations running outdoor learning activities are measuring, i.e., identify the outcomes which providers seem to be seeking to achieve; and
3. Assess the designs of individual evaluations (while aware that study designs vary in their openness to bias and hence inaccuracy) and the standard of evidence generally available for different types of outdoor learning. (Fiennes et al., 2015, 5)

From 3,536 titles and abstracts found, the authors finally included 4 UK surveys, 16 systematic reviews and 57 primary UK studies in their review (ibid., 48). I am attempting here a concise summary of their review, in particular the third part which

This summary has been submitted to the authors of the original systematic review of 2015 for correctness. It has been approved on 9th February 2021. The original review can be found here: www.giving-evidence.com/outdoor-learning. We are very grateful indeed for the support of the original authors and also for granting us the copyright to reproduce 3 boxes from the original review.

R. Jucker (✉)
Stiftung SILVIVA, Jenatschstrasse 1, CH-8002 Zürich, Switzerland
e-mail: rolf.jucker@silviva.ch

© The Author(s) 2022

R. Jucker and J. von Au (eds.), *High-Quality Outdoor Learning*,
https://doi.org/10.1007/978-3-031-04108-2_6

was concerned with assessing the quality of the research designs and the available evidence.

However, the authors point out in the previous two sections that there is "no comprehensive or regular (repeated) survey of the scale of outdoor learning in the UK". Disturbingly, they cite research which shows that, at least for the 30 years up to 2010, fieldwork and residential study have declined, not risen, in the UK (ibid., 11–12), at least in Biology. The factors cited for this decline are time and cost pressure, changes in curriculum and its assessment, as well as fears around health and safety and a decline in teachers' enthusiasm and expertise (ibid., 12).

With regard to the quality of research in the field, the authors found that the then-current research base (the research was done in 2015) in the UK raised issues of research ethics, the quality of systemic reviews available, and confusion between interventions and outcomes in studies. In addition, the primary studies they found in the UK are limited by the following factors:

- The studies are thinly spread across a wide variety of populations, age groups, interventions, settings and outcomes, so "few topics have been researched more than a handful of times." (ibid., 6)
- Types of activities and participants are limited mostly to adventure or residential activity; 11–14 year olds; and the general population.
- "The outcomes measured are mainly around 'character development-type' outcomes (communication skills, teamwork, self-confidence etc.). Very few studies addressed interventions with strong links to core curriculum subjects. (…) Looking internationally, only six of the 15 systematic reviews looked at educational attainment, and only one addressed employability." (ibid.)
- "Safety is little covered in the systematic reviews and was not measured as an outcome in any of the primary studies. Safety is obviously a major issue in outdoor learning since it can be dangerous." (ibid.)
- In terms of the methodological quality of the designs of the studies, the review, using a scale developed by Project Oracle,[1] found that many UK studies did not even reach Level One of this scale. This means that they did not have an explicit theory of change ("also known as a logic model: an articulation of the inputs, the intended outcomes, how the inputs are meant to produce those outcomes, and

[1] The five levels are: "Project Oracle's scale 'rates' what we know about interventions on whether there are: (1) detailed project descriptions and logic models; (2) before and after studies; (3) evaluations with a control group, which one would expect for interventions beyond the pilot stage; (4) replicated evaluations of impact; and (5) multiple independent evaluations in different settings, which may imply that further evaluations are less useful." (Fiennes et al., 2015, 9) More on the Project: "Project Oracle is a children and youth evidence hub that aims to improve outcomes for young people in London. We do this by building the capacity of providers and funders to develop and commission evidence-based projects, creating an ecosystem in which evidence is widely gathered, used and shared. We also work with specific 'cohorts' or sub-sets of the sector to embed good practice, and at a national and international level to promote the wider use of evaluation and evidence. Project Oracle is funded by the Greater London Authority (GLA), the Mayor's Office for Police and Crime (MOPAC) and the Economic and Social Research Council (ESRC)." (ibid., 9, FN 2).

assumptions about context, participants or other conditions", ibid., 7). This might mean that the practitioners had a poor understanding of their intervention, and more seriously, it impedes other practitioners in assessing whether the intervention might achieve the same outcomes in their context. "No UK study, or set of studies, featured the more demanding attributes of Levels Four or Five". This means that no intervention had been replicated and studied in multiple contexts (ibid., 7).

The authors make a very effective plea for research quality (see Box 3, ibid., 28–29; reprinted as Box 1 below):

Box 1: Why we evaluate research methods

(i) *Because different research methods give different answers*

"Two men say they're Jesus: One of them must be wrong" (Dire Straits lyric!)

Table 1 shows the effect of a reading programme in India measured using several research methods (Innovations for Poverty Action). These methods all used the same outcome measures, but the experimental designs were different.

The answers vary widely: some suggest that it works well, others show it to be detrimental. Clearly there can only be one correct answer! All the other answers are incorrect: and could mislead donors or practitioners to implement this programme at the expense of another which might be better.

Table 1 Different methods and impacts

Method	Impact estimate
(1) Pre-post	26.42[a]
(2) Simple difference	−5.05[a]
(3) Difference-in-difference	6.82[a]
(4) Regression	1.92
(5) Randomized experiment	5.87[a]

[a] Statistically significant at the 5% level

The answers vary because research methods vary in how open they are to biases (i.e., systematic errors). For instance, suppose that a medical trial involves giving patients a drug for two years. Suppose that that drug has horrible side-effects such that during the two years, some patients can't stand taking it so they drop out of the trial (or worse, perhaps the drug kills some of them). If the trial only collects data on patients who are still in the trial after two years, it

will systematically miss the important insights about those side-effects. This 'survivor bias' will make the drug look more effective than it really is.

Somebody reading the trial results without knowing that detail wouldn't be able to distinguish the actual effect of the drug from that of this survivor bias. Similarly, if a study only looks at the outcome (in the example above, it's reading level) before the programme and then afterwards (i.e., is a pre-post study), it won't be possible to distinguish whether any improvement in reading levels was due to the programme or just to the fact that children learn over time anyway.

{As an aside, contrary to popular myth, it is not invariably the case that robust research is more expensive than unreliable research, nor that randomised controlled trials (the most reliable design for a single primary study) are invariably terribly expensive: many are cheap or free. See Appendix 12, ibid., 73}.

(ii) *Because weaker research methods allow for more positive findings*

The UK National Audit Office searched for literally every published evaluation of a UK government programme (National Audit Office, 2013*: Evaluation in Government*). Of those, it chose a sample, and ranked on one hand, the quality of the research method ('robustness' on the x axis, i.e., how insulated the study is from bias), and on the other, the positive-ness of the programme ('claimed impact').

The trend line on the resulting graph below would slope diagonally downwards. It shows that more robust research only allows for modest impact claims whereas weak research allows much stronger claims.

Bad research can be persuaded to say almost anything, and won't allow researchers to distinguish the effects of a programme from other factors (e.g., the passage of time, the mindset of participants, other programmes) nor from chance.

Most social interventions have a small effect and a reliable research method will show what that is: bad research is likely to overstate it. The highest estimate for the reading programme above is from the pre-post study which is a weak study design (Fig. 1).

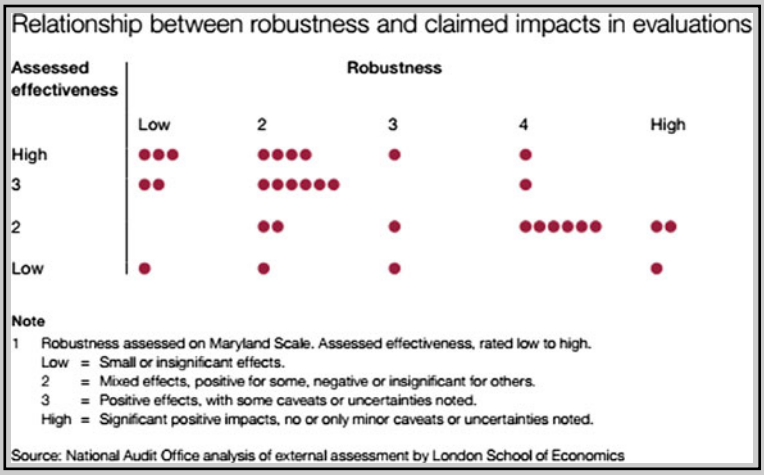

Fig. 1 Robust research allows for modest impact claims, weak research allows much stronger claims

This relationship between weak research methods and positive findings has been shown also in medical research. We found it in the studies of outdoor learning too.

It is therefore very important, and should certainly be a future aspiration both for practitioners and for researchers to adhere to robust and rigorous research designs. The authors note: "We were unable to find replicated studies that took into account differing contexts and that were sufficiently well documented for wider implementation." (Fiennes et al., 2015, 30)

Box 5 (reprinted as Box 2 below) shares guidelines for describing interventions from medical research which might help the outdoor learning sector to improve replicability of good practice (ibid., 33).

Box 2: Describing an intervention

Medical research has guidelines for describing interventions such that somebody else can replicate them accurately. They have a 12-point checklist for describing interventions, the Template for Intervention Description and Replication (TIDieR) (Hoffman et al., 2014), which is helpful and could easily be adapted for outdoor learning. It has been adapted elsewhere, e.g., by mental health charities (Kent County Council, 2014):

- The name of the intervention (brief name or phrase)
- The way it works (rationale, theory, or goal of the essential elements)
- What materials and procedures were used (physical or informational)
- What (each procedure, activity, and/or process)
- Who provided the intervention (e.g., nurse, psychologist, and give their expertise and background)
- How was it delivered (e.g., face to face, online, by phone, and whether it was provided individually or in a group)
- Where it took place
- When and how much (the number of sessions, schedule, dosage and duration)
- Tailoring (what if anything could be adapted to the individual, why and by how much)
- Modifications which happened after the study started
- How well was adherence to the plan assessed (i.e., the process for assessing adherence)
- The extent to which implementation adhered to the plan.

Given these limitations, it seems fair to suggest that the findings, implications and recommendations of the review about the effectiveness of interventions should be treated with caution. They might qualify as indications and trends, rather than established truth. The most solid findings were:

- "[The systematic reviews] almost all report that the various outdoor learning activities have positive effects on all their various outcomes, e.g., attitudes, beliefs, interpersonal and social skills, academic skills, positive behaviour, re-offending rates and self-image." (Fiennes et al., 2015, 17)
- "The effect attenuates over time: the effect as measured immediately after the intervention is stronger than in follow-up measures after a few months. This is common for social interventions. However, one meta-analysis found that effects relating to self-control were high and were normally maintained over time." (ibid.)
- "Longer programmes tend to be more effective than shorter ones. This fits with practice-based knowledge that length can allow for a more intensive and integrated experience and is obviously important given the pressure to cut length in order to reduce costs." (ibid.)
- "Strong benefits are also associated with well-designed preparatory work, and follow-up work." (ibid.)

For the following types of intervention, there was less or mixed evidence, considerable variation in effect sizes or only evidence for certain findings:

- Positive benefits on academic learning
- Creative development, emotional development and social skills. (ibid.)

For some interventions, such as mountaineering or rock climbing, evidence was weak, absent or there even was evidence of harmful impacts (ibid.).

In our context it is worth noting that the review found only "very few studies (…) of interventions with strong links to core curriculum subjects" (ibid., 21). In addition, there seemed far fewer studies looking at outdoor-based learning in a regular school day setting, compared to residential experiences (ibid., 22). In terms of age of the pupils researched, most concern 11–18 year olds.

"Strikingly few studies looked at educational attainment" (ibid., 23), whereas "non-educational outcomes", such as curiosity, relationship with nature, self-awareness, self-esteem, self-responsibility, communication or teamwork, health and well-being, healthy lifestyles, employability, youth leadership, community integration or community leadership, "have received much more research interest" (ibid., 24).[2] The authors sanguinely state: "We take no view here on whether non-educational outcomes are important, but rather notice the mismatch between research topics and the pressure schools face to achieve those educational outcomes." (ibid., 26)

Given that effect sizes of 0–0.2 are considered small, 0.5 is considered moderate, 0.8 or more is considered large, the average effect sizes in some of the systemic reviews of between 0.26 and 0.35 have to be considered small to moderate.

Recommendations

In terms of developing a coherent, robust agenda for practitioners and researchers of the outdoor-learning sector, I would translate the authors' recommendations into the following four strategies:

- On the level of practitioners of outdoor-based learning, they need to be enabled to create and use theories of change, i.e. they need to be clear about their operational models (see ibid., 32 and Box 4, reprinted as Box 3 below). Practitioner's organisations also need to have systems in place to collect relevant data but also to "support ethical practices for monitoring and research, particularly the storage and sharing of data from evaluations" (ibid., 8).
- On the level of researchers, they need to "create a system to regularly capture data on the types and volumes of activity". Only with a decent set of baseline data can the sector, funders or government agencies trace (positive or negative) developments.
- Researchers, practitioners, funders and government bodies need to reflect together on the important research topics and prioritise them deliberately. This includes

[2] "Other outcomes included: creativity, commitment to learning, respect for self / others, sense of social responsibility, sense of belonging, addressing fear, tenacity, confidence, social skills, motivation, concentration, physical skills, resilience, social behaviour, direction, mindset, enjoyment, inspiration, impact on schools, family and community, critical thinking, self-determination, competence, relatedness, task approach, task avoidance, ego approach, ego avoidance, Relative Autonomy Index (RAI), interest effort, value autonomy-support, metacognition, problem-solving skills, optimism, pedagogical skills." (Fiennes et al., 2015, 24).

the need for "creating a more shared language around the categories of activity" (ibid., 32).

- A new open-access culture needs to be developed which ensures that "both interventions and research are described clearly, fully and publicly" (ibid., 8).

Box 3: Theories of change

What is a theory of change?

A theory of change (or logic model: we use the terms interchangeably) is what is meant by Project Oracle's Level 1's 'we know what we want to achieve' and 'project model' (i.e., articulation of how the activities are supposed to create the intended impact). It lays out the assumptions behind an intervention, and links between activities and intended impacts (i.e., how the activities are supposed to produce those impacts, and what is assumed, e.g., parental engagement, weather…). They allow organisations to find and cite evidence suggesting that their activities are likely to produce their target outcomes.

A clear theory of change also helps other organisations considering running the intervention to see whether the assumptions are likely to hold in their contexts, i.e., whether they're likely to get similar results. It also helps other organisations make good decisions about what outcomes to try to achieve by showing what's involved in the interventions which ostensibly deliver them.

The diagram below shows the constituent pieces of a logic model (Fig. 2):

Fig. 2 Logic model

Why does having a clear logic model matter?

A clear logic model is important/essential to intelligent programme design because it enables predictions about whether a type of intervention is likely to work (for a specific population). An evaluation without a clear logic model simply shows whether a programme worked and the extent to which it worked: it gives no indication of why it worked (or not)—why it gets those results. That is, without a logic model, the intervention is like a black box: we gain no insight into whether it's likely to achieve those results again, nor elsewhere. It adds nothing to the 'science' (i.e., understanding) of these interventions. By contrast, if a provider starts with a clear logic model, they can use the existing research to see which parts are likely to be true, which are not evidenced, and therefore can:

(a) make an educated estimate of whether, when and for whom the intervention is likely to work,
(b) identify major risks and unsupported assumptions,
(c) change the design to make it more likely to succeed. It may transpire that the proposed logic model is totally fanciful and implausible, and hence this work will prevent them running a pointless intervention, or even a harmful intervention. And
(d) identify what needs testing. Maybe very little needs testing and so the practitioner is spared all the cost and hassle of evaluating.

In short, it enables practitioners to use existing research, rather than solely to produce research. Clearly this is more efficient. The focus on impact has led many organisations (particularly charities) to often produce research of bad quality), when (i) they are not set up nor incentivised to be researchers, and (ii) it might be more useful for them to leverage the (better quality) research which already exists.

Even though the review is a few years old, I think it is very useful indeed to sharpen our focus on what we need to do to improve the quality of outdoor-based learning provision as well as the quality of the research assessing its impact,[3] and thereby guiding future practice and policy development.

[3] Interestingly enough, this review reaches similar conclusions as the systematic review by Becker, C., Lauterbach, G., Spengler, S., Dettweiler, U., & Mess, F. (2017). Effects of regular classes in outdoor education settings: A systematic review on students' learning, social and health dimensions. *International Journal of Environmental Research and Public Health*, *14*(5), 1–20. http://doi.org/10.3390/ijerph14050485).

References

Fiennes, C., Oliver, E., Dickson, K., Escobar, D., Romans, A., & Oliver, S. (2015). *The existing evidence-base about the effectiveness of outdoor learning* (pp. 1–73). www.giving-evidence.com/outdoor-learning. Accessed 12 Feb 2021. The review can also be accessed on the sites of the respective organisations: https://www.blagravetrust.org/wp-content/uploads/2015/11/The-Existing-Evidence-base-about-the-Effectiveness-of-Outdoor-Learning-Executive-Summary-Nov-2015.pdf. Accessed 15 Feb 2021. https://www.outdoor-learning.org/Portals/0/IOL%20Documents/Research/outdoor-learning-giving-evidence-revised-final-report-nov-2015-etc-v21.pdf?ver=2017-03-16-110244-937. Accessed 15 Feb 2021.

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, R., Altman, D. G., Barbour, V., Macdonald, H., Johnston, M., Lamb, S. E., Dixon-Wood, M., McCulloch, P., Wyatt, J. C., Chan, A.-W., & Michie, S. (2014). Better reporting of interventions: Template for intervention description and replication (TIDieR) checklist and guide. *The British Medical Journal,* 348. https://doi.org/10.1136/bmj.g1687. Accessed 19 Apr 2021.

Kent County Council. (2014). HeadStart Kent. http://www.kelsi.org.uk/__data/assets/powerpoint_doc/0010/25588/Programme-Board-Main-Presentation-19-11-14-AF-UN.pptx. Accessed 19 Apr 2021.

Rolf Jucker is currently Director of the Swiss Foundation for Nature-based Learning (SILVIVA) and a learning for sustainability expert, having previously served as Director of the Swiss Foundation for Environmental Education from 2008 to 2012. He gained an MSc in Education for Sustainability (EfS) and worked extensively on education for a viable future, publishing widely on the subject. He is the author of *Do We Know What We Are Doing?* (2014) and *Can We Cope with the Complexity of Reality? Why Craving Easy Answers Is at the Root of our Problems* (2020).

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.