

# Assessing claims in education – the ACE concepts

**LISA-MARIA MÜLLER**  
CHARTERED COLLEGE OF TEACHING, UK

**ANDREW MORRIS**  
LEARNING AND SKILLS RESEARCH NETWORK, UK

**JONATHAN SHARPLES**  
EDUCATION ENDOWMENT FOUNDATION, UK

**JOE CHISLETT**  
FOUNDER, KRITIKOS LTD, UK

**NICK ROSE**  
AMBITION INSTITUTE, UK

**HAMISH CHALMERS**  
UNIVERSITY OF OXFORD, UK



“Is there a profession filled with more disagreement among its members than teaching?” This was a question posed on Twitter by lecturer in education at La Trobe University Emina McLean in October (<https://twitter.com/EminaMcLean/status/1178879273127403520?s=03>). Why, she mused, is there furious disagreement about such fundamental aspects of our profession as teaching reading, managing behaviour, assessment, inclusion, pedagogy and technology? And well might she ask. The EduTwitter community enthusiastically offered a multitude of responses to McLean’s question, leading to some more or less heated debates in the ensuing thread. Responses largely centred on: ideology; selective use of research evidence; a lack of understanding of research methods; design and its implications; the heterogeneity of classrooms; commercial interests; and the inherent complexity of the field. These factors, together with exaggerated claims in (social) media, press releases and academic articles themselves (Haber et al., 2018; Sumner et al., 2016), have

arguably led to the rapid rise and spread of educational claims, both on- and offline, about what is and isn’t effective, many of which you may have come across. For example, there are claims that inquiry-based learning is ineffective, that playing brain games can improve students’ executive functions, and that students learn less in larger classes. How can you know which of these claims are trustworthy? And how should you decide when to act on such claims?

A group of researchers from a variety of disciplines have come together to create a tool with key concepts that can help stakeholders in their fields to assess the basis of claims and support them in deciding whether to apply an intervention in their context. These concepts were originally developed for health (Oxman et al., 2018) and have since been adapted for an educational context, as well as a range of other disciplines such as policing or environmental studies (see Oxman et al., 2019, for further details). We, a group of educationalists under the banner of CEBE (Coalition for Evidence-Based Education), were involved in the development of key concepts to Assess Claims in Education (ACE), and it is the aim of this article to present these concepts and outline how they can help you to assess the validity of claims, determine the trustworthiness of comparisons and decide whether to adopt an intervention in your context. Hopefully this will be helpful in shedding some light on some of the underlying reasons for the high levels of fundamental disagreements in education.

The 37 key concepts are set out succinctly in a handy web-based tool, which has been designed for use in staff development sessions as well as by individuals. You can access the key concepts, further information and many related resources at [thatsclaim.org/](https://thatsclaim.org/) >

GETTY



educational, and this article is an extension of the summary and the introductory text that you can download on the website.

### Claims

A good place to start is at the level of the claim itself. Does it seem too good to be true? If so, this alone may give us reason to be sceptical. We should also consider what – if any – basis there is for the claim. Is it based on faulty logic? Or is it perhaps based on trust alone? These are the questions that the first category of concept cards helps you to answer. Let us assess some claims and why their basis may be shaky.

For example, a study, newspaper headline or company could claim that an intervention to train executive functions is 100 per cent effective or leads to ‘dramatic’ improvements in student outcomes. It is probably wise to stay cautious when encountering such a claim, as dramatic effects of interventions are very rare in education. Rigorous studies show that most education interventions have small or moderate effects. The claim may hence either exaggerate the effects of an intervention or have used an inappropriate methodology.

It is also important to keep in mind that we can rarely be 100 per cent certain about the size of the effect of an intervention, or what would happen if we used the intervention in a different context. This is because statistical results are average effects across groups, and may not adequately recognise the specific circumstances of a given individual.

Sometimes claims can also be based

on faulty logic, such as the conflation of correlation and causation. Just because an educational outcome is associated with an intervention, it does not necessarily mean that the intervention caused the outcome. The association could also be due to chance or some other factor that hadn’t been accounted for. Therefore, it is important to always check whether those who took part in an intervention were compared to a similar group who did not receive the same intervention. If they were not, we cannot be sure that the effects were actually due to the intervention.

Another shaky basis for claims is trust alone. Sometimes we can be inclined to trust people simply because they have a lot of experience in a field or because they are (self-declared) experts. Of course, professional expertise and experience are important, and should be taken into account, but they should also be combined with other sources of evidence to make sure that the recommendations are solid and trustworthy.

Overall, the first step in assessing claims in education is thus to check the basis of claims. Do they sound too good to be true?

**It is important that studies provide detailed information about the exact measure they used and how much participants have (or have not) improved**

Are they based on sound logic? And are they based on more than trust alone?

### Comparisons

As mentioned above, to know whether an

intervention (such as training executive functioning) causes an effect (such as improved attainment), the intervention has to be compared to something else (such as not specifically training executive functioning but continuing to use an existing teaching method, for example).

Researchers compare an intervention given to people in one group with something else given to people in another group. Those comparisons provide evidence – facts to support a conclusion about whether a claim is right or wrong. For those comparisons to be fair, the only important difference between the groups should be the interventions that they receive. For example, differences in participants’ socio-economic background, their age or their achievement levels prior to the intervention are all common factors that could influence the effect that an intervention has on their outcomes.

Once the interventions have been compared and results are available, these need to be described. There are several potential issues around the description of findings. For example, authors may jump to conclusions about the wider population based on a small sample, or results can be described verbally instead of numerically. The problem with the latter is that ‘big effect’ or ‘dramatic gains’ are highly subjective terms that depend heavily on people’s prior experience and reference points. We do not know whether ‘dramatic’ means one grade level or a couple of points on a test. This is why it is important that studies provide detailed information about the exact measure they used and how much participants have (or have not) improved. In education, it is also considered good practice to provide confidence intervals. They tell us a bit more about the range of scores that the true value is likely to be part of and the

range of scores that we can expect if we were to repeat the intervention.

Furthermore, when reading summaries of interventions, it is essential to check that these summaries are trustworthy. There are many different methods of summarising research evidence and some are more objective than others. Sometimes people can cherry-pick evidence to support their views instead of taking all evidence into account. Systematic reviews are designed to avoid this problem. They clearly outline the search terms that authors have used when compiling the review, and are transparent about the number of articles they found, which ones were included in the review and why. This process ensures that authors include all available evidence on a given topic and not just the evidence that supports their point of view. As such, systematic reviews (or meta-analyses) are generally more trustworthy than reviews that did not use a systematic methodology when looking for relevant research literature and summarising their results.

When looking at an intervention, it is thus important to always think ‘fair’. Is the evidence based on fair comparisons? Are the effects described appropriately? And are the summaries systematic?

Once you have considered the basis of claims and how results are reported, the final step is to consider whether the evidence is applicable to your context.

### Choices

A good choice is one that uses the best information available at the time. For education choices, this includes using the best available evidence of intervention effects. Good choices don’t guarantee good outcomes, but they make good outcomes more likely.

Firstly, it is important to consider what

the problem is that you would like to solve and what the available interventions are. For example, a student may be handing in work with lots of spelling mistakes because they’ve been distracted while doing their homework or because they are dyslexic. The appropriate intervention will depend on what the underlying issue is. This, in turn, will influence which sources of evidence are appropriate to consult and are applicable in your context.

Secondly, it is critical to check whether the evidence is relevant to your context and circumstances. If the intervention was carried out with students who are very different to yours – for example, in terms of their age, their socio-economic background or their prior achievement – the results may not be directly applicable to your context. This is not to say that they should not be considered at all, or are entirely irrelevant, but it is important to be mindful of the fact that a difference in student population may influence the ultimate results. Another important factor to consider is how and where a study was carried out. If the study was conducted in a highly controlled ‘lab’ setting, the results may not translate directly to the classroom. Moreover, if the intervention was administered by trained research staff following a strict protocol rather than by classroom practitioners, who also have to consider a myriad of other factors, such as student behaviour and administrative issues, this difference may influence the final outcomes. Again, this is not to say that studies that have been carried out in ‘lab’ settings do not translate to the classroom, but it is important to take these differences into account and manage expectations.

Finally, as with treatments in medicine, it is important to consider the advantages and disadvantages of a

specific intervention (e.g. time/cost) and whether the advantages are likely to outweigh the disadvantages. For example, if an intervention has led to moderate effects but is highly time-consuming and expensive, it may be worth considering an alternative.

In sum, when considering whether an intervention is appropriate for your context, take care! Consider what your problem is, what your options are, whether the evidence is relevant to your context and whether the advantages outweigh the disadvantages.

Next time you encounter a claim about an education intervention online or in the staff room:

**BEWARE** of claims that do not have a solid basis.

**THINK FAIR** and check the evidence from intervention comparisons.

**TAKE CARE** and make good choices.

Hopefully, this will help clear up some of the controversy surrounding claims in education. We hope you find the website ([thatsaclaim.org/educational](http://thatsaclaim.org/educational)) useful and would welcome your feedback.

### REFERENCES

- Haber N, Smith ER, Moscoe E et al. (2018) Causal language and strength of inference in academic and media articles shared in social media (CLAIMS): A systematic review. *PLoS ONE* 13(5): e0196346.
- Oxman AD, Aronson JK, Dahlgren A et al. (2019) Key concepts for making informed choices. *Nature* 572: 303–306.
- Oxman AD, Chalmers I, Austvoll-Dahlgren A et al. (2018) Key concepts for assessing claims about treatment effects and making well-informed treatment choices. *F1000Research* 7: 1784.
- Sumner P, Vivian-Griffiths S, Boivin J et al. (2016) Exaggerations and caveats in press releases and health-related science news. *PLoS ONE* 11(12): e0168217.