Guidance of Effect Sizes

Displaying Outcomes using Effect Sizes on What Works

What Works is a database of evidenced interventions that aim to support children and young people's speech, language and communication hosted by The Communication Trust.

Decisions about which interventions are listed on the database are made by a moderating group of highly experienced academics and practitioners. The group check each intervention against criteria for standards of evidence; indicative, moderate or strong. The intervention's evidence rating is then displayed on the site to inform users of the strength of the evidence base behind it.

User feedback

Over recent years, users have told us that the emphasis on the evidence rating on What Works can be confusing. It is important to know that a 'strong' evidence rating doesn't necessarily mean that the intervention is likely to work well.

In fact, in some cases there may be strong evidence that the intervention has no positive effect at all, or even that there is a negative effect on children and young people's speech, language and communication skills; whereas indicative evidence (although meeting a lower standard of evidence) may suggest that it is likely to have the effect you want.

How can What Works support users to interpret the outcomes of an intervention?

To understand what the evidence actually says about an intervention, it is important to look at the strength of the intervention's outcomes for children and young people. The best way to do this is to read the journal articles and consider:

- The strength of the effects that researchers report.
- How well designed the studies are.

Many practitioners may not have time to read or have access to the required journal article. The What Works database enables practitioners to work from summaries of the relevant papers to the intervention.

The evidence ratings (strong, moderate, indicative) on What Works already support users to decide how well designed studies are. The **display of effect sizes** on What Works will support users with interpreting children and young people's outcomes for interventions based on the strength of effects.

What is an effect size?

An effect size is a number representing the size of a research result (the difference between the treatment and the control group on a given outcome in a particular study). It shows us not just whether or not an effect is statistically significant, but also the strength or size of that effect. These have been calculated from one of the research studies included for each intervention.

Results from studies have been converted into 'Hedge's g', which is a standardised measure of effect size. 'Hedges G' is used to calculated the size of the difference between the outcomes for the treatment and the control (not treatment, placebo or alternative treatment) group.

A negative Hedges' g indicates that an intervention results in poorer scores for children receiving it than for a control group. Positive Hedges' g values indicate that an intervention has "worked" to some extent and quantify the benefit produced by an intervention.

A number of different sites offer effect size calculators but this one has them all https://www.psychometrica.de/effect_size.html.

How will this information be shown?

For each intervention, there will be a graphic displaying:

- the largest/strongest effect size
- the smallest/weakest effect size
- the effect size for the primary outcome as stated by the researchers (in other words what the researchers were primarily looking at in their study)
- The reference for the study from which the effect sizes have been calculated

How can I interpret effect sizes?

In targeted "health" interventions an effect size can be roughly classified as small (0.2), medium (0.4) or large (bigger than 0.8) (see box below). In education and in universal interventions the threshold may be rather different. The UK's Education Endowment Foundation suggest that in education research it is appropriate to speak of effect sizes as being low (0.01 to .18) moderate (0.26 to 0.44) and high (0.56-0.69) (Coe et al.2013). According to the What Works Clearinghouse (US Department of Education, 2014) an effect size of 0.25 standard deviation units or larger can be considered "substantially important".

But it is important to note that these interpretations are just a guide, so although they can give you a good starting point for considering how effective an intervention is according to the studies, you do still need to consider not just the size of the effect but parameters such as cost. So a short, low-cost intervention which reaches, for example, all children in a school district but produces a small positive Hedges' g might be of practical importance whereas a very expensive intervention which is relevant to a very small group of children might not be.

One of the most commonly employed "rules of thumb" about the interpretation of effect sizes comes from Cohen (1969, p23). Cohen describes an effect size of 0.2 as 'small' and illustrates with the example that the difference between the heights of 15 year old and 16 year old girls in the US corresponds to an effect of this size.

An effect size of 0.5 is described as 'medium' and is 'large enough to be visible to the naked eye'. A 0.5 effect size corresponds to the difference between the heights of 14 year old and 18 year old girls.

Cohen describes an effect size of 0.8 as 'grossly perceptible and therefore large' and equates it to the difference between the heights of 13 year old and 18 year old girls.

Are effect sizes calculated for all interventions?

For some interventions there isn't enough information available to calculate the effect size. Other times, there may be sufficient information for an effect size to be calculated for some outcomes, but

not all of them. Whenever sufficient information is presented, an effect size will be calculated and will be displayed as easy-to-read 'pop-up' diagram on the intervention's What Works page.

References

Coe, R, Kime, S. Neville, C. and Coleman, R. (2013) *The DIY evaluation guide* London: Education Endowment Foundation.
Cohen, J. (1969). Statistical power for the behavioural sciences. London: Academic Press
What Works Clearinghouse Procedures Handbook Version 4.0, 7/14/2017
https://ies.ed.gov/ncee/wwc/Docs/ReferenceResources/wwc_procedures_handbook_v4_draft.pdf