EARLY INTERVENTION FOUNDATION

Technical summary: Impact on the EIF Guidebook

As of April 2021, the Guidebook includes information on the *impact* of programmes: that is, on the size of the improvements they have generated for children and young people. This information includes: (i) the size of improvements as they were originally measured, (ii) a transformation of effect sizes, called the improvement index, and (iii) the time point at which the outcome was achieved. This technical summary provides more detail on this new information.

Inclusions and exclusions

We have decided to only publish information on the size of improvements for programmes that receive our level 3 and 4 strength of evidence ratings (and for studies that have been assessed as robust). This is because we can be confident in these cases that there is a causal relationship between participation in the programme and improvements in outcomes, and that the evidence provides unbiased and trustworthy estimates of improvement in outcomes.

Effects as they were originally measured in evaluations investigating the impact of the programme

- This number describes the difference between the average outcomes of those who have received the programme, and the average outcomes for those who did not receive it – the difference between these outcomes is the improvement that we can attribute to the programme. An example of this information might be 'A 5-point improvement on the Problem Behaviour Scale', 'A 20% reduction in smoking', or 'A 15-percentage point reduction in the proportion of participants who have developed a major depressive disorder'.
- In terms of calculation:
    - This is typically calculated by subtracting the post-test mean for the treatment group from the post-test mean for the comparison group (as reported in the underlying evaluation studies), and expressing this as an absolute value.
    - Sometimes researchers will report the group mean difference itself, or a regression coefficient that is equivalent to this, and we will use these figures when they are available.

- When possible, we will report the intervention effect as the difference between covariate-adjusted means, as these estimates are more precise (owing to adjusting for differences that may exist between the treatment and control groups).

The improvement index

What is it?

- The **improvement index score** is a number between 0 and 50 that captures the magnitude of an effect and facilitates comparisons across effects measured on different scales (much like an effect size, such as Cohen's d), and does so in a way intended to be more interpretable to non-researchers than alternative ways of describing effects. This metric is sometimes called 'percentile growth', or 'percentile rank improvement'. This approach is described as a useful way of describing effect sizes by a number of methodologists in the field (Coe, R., 2002; Baird, M.D., & Pane, J.F., 2019), and is also used by colleagues at the What Works Clearinghouse for Education in the US, and the Best Evidence Encyclopaedia.
- The improvement index score is the difference between the percentile rank corresponding to the mean value of the outcome for the intervention group, and the percentile rank corresponding to the mean value of the outcome for the comparison group distribution (What Works Clearinghouse., 2020). It estimates the percentage of the control group who would be below the average person in the experimental group on a given outcome (or, conversely, it tells us that the mean score of the intervention group exceeds a certain percentage of those in the control group) (Coe, R., 2002).
- More simply, it can be interpreted as an estimate of how much we'd expect the average participant in the control group to improve if they had received the intervention, relative to other participants. For example, if the improvement index score is 25, that means that the average control group participant (who has better outcomes than 50% of participants, and worse outcomes than 50% of participants), would improve their percentile rank from 50 to 75. This means that if they had received the intervention, they would now have better outcomes than 75% of participants, and worse outcomes than only 25% of participants.
- In terms of **calculation**, this fundamentally involves converting Cohen's d or another similar metric into percentiles.
- This is possible owing to the fact an effect size is equivalent to a z-score of a standard normal distribution. This property permits us to calculate what percentage of the area in a normal distribution would fall below the z-score (Coe, R., 2002).
- For example, a Cohen's d of 0.6 indicates that the mean score of participants in the treatment group is 0.6 SDs above the mean score of participants in the control group, and that the mean score of the treatment group exceeds the scores of 73 percent of those in the control group (Coe, R., 2002). Here, the improvement index score is 23. This means that we would expect that the average participant in the control group's percentile score would improve by 23 percentile ranks (from the 50th percentile to the 73th percentile) if they had received the programme, and so belong to the group containing the top 27% of participants with the most favourable outcomes.

How is it calculated?

Calculating the improvement index score is a two-stage process: (i) calculate an effect size, and (ii) translate this into the improvement index score. This process varies slightly depending on whether outcomes are continuous variables or binary variables.

Broadly speaking, effect sizes (Cohen's d, Hedge's g, or Glass' delta for continuous variables, and odds ratios for binary outcomes) are taken directly from evaluation studies. If an effect size is not available, reviewers at EIF will calculate a Hedge's g[1] effect size for continuous variables (using post-test means, standard deviations, and analysis sample sizes reported in the evaluation study, for the treatment and control groups), or calculate an odds ratio for binary variables (using the number of participants belonging to each outcome group).

Effect sizes, whether directly reported, or calculated by EIF reviewers, should:

- Be a between-group effect size (i.e. compare the treatment and control group, rather than pre-post change within the treatment group).
- Be describing a main effect (i.e. for all of the available sample, subject to missing data), rather than a subgroup effect (analyses of outcomes for specific subsets of the sample).
- Be adjusted for covariates included in the statistical analysis of outcomes. However, EIF reviewers will accept unadjusted effect sizes provided that there are no serious baseline imbalances (either due to issues with the design, or due to attrition) likely to introduce bias to the effect size.
- Use post-test means and not mean gain scores.
- Use unadjusted standard deviations.
- Use standard deviations based on outcome scores at post-test or follow-up. Standard deviations based on gain scores are not acceptable.

[1] When an effect size for continuous variables is not directly reported in a study, EIF reviewers will calculate a Hedge's g, rather than a Cohen's d, as it is robust to a broader range of circumstances (i.e. smaller sample studies). However, practically speaking, Hedge's g and Cohen's d will be identical in the majority of the studies that EIF reviews.

When effect sizes are not directly reported, EIF uses the following calculations.

Hedge's g calculation	$$Hedge's\ g = \frac{\omega(M_1 - M_{2)}}{SD\ pooled}$$ Where M_1 is the post-test mean for the treatment group, M_2 is the post-test mean for the control group, and SD pooled is given by: $$\sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}}$$ Where SD_1 is the standard deviation for the treatment group, SD_2 is the standard deviation for the control group, n_1 is the sample size for the treatment group, and n_2 is the sample size for the control group. ω is a small sample correction which is calculated as follows: $$\omega = 1 - \frac{3}{4(n_1 + n_2) - 9}$$
Odds ratio calculation	$$Odds\ ratio = \frac{(a/b)}{(c/d)}$$ Where a is the number of participants in outcome group 1 from the treatment group, b is the number of participants in outcome group 1 from the control group, c is the number of participants in outcome group 2 from the treatment group, and d is the number of participants in outcome group 2 from the control group.

This effect size is then converted into a percentile rank:

Percentile rank calculation for continuous variables	$$\Phi(\beta) - 0.5$$ Where Φ is the standard normal cumulative distribution function (or CDF), and β is the relevant effect size – indicating the proportion of the area under the standard normal curve for the effect size value (Baird M.D., & Pane, J.F., 2019). Calculating $\Phi(\beta)$ is equivalent to using a statistical z table (or standard normal z table) to ascertain what percentage of the area in a normal distribution falls beneath a given z-score. This is then subtracted by .5 to give an indication of percentile growth. For example, a Cohen's d of .6 indicates that 73% of the area in a standard normal distribution falls beneath it (or in other words, the mean score of the experimental group exceeds the scores of 73% of those in the control group). We then subtract this by 50 to

	provide an estimate of expected growth for the median (50th percentile ranked) control group participant if they had received the intervention – estimating that the average control group participant would improve their percentile rank by 23 (moving from 50th percentile to 73rd percentile).
Percentile rank calculation for binary variables	Step 1: Convert odds ratio into Cox Index. This is an effect size that can be broadly interpreted similarly to Cohen's d/Hedge's g type effect sizes (What Works Clearinghouse., 2020). $$D_{Cox} = \omega \frac{LOR}{1.65}$$ Where LOR is the natural logarithm of the adjusted odds ratio, and ω is a small sample correction. Step 2: The Cox Index is then used as the effect size in the Improvement Index calculation described above.

In summary, EIF reviewers use one of five methods for producing an improvement index score:

- **Method 1** – When an effect size for a continuous variable is directly reported (Cohen's d, Hedge's g, or Glass's delta), EIF reviewers will convert this directly into an improvement index score.
- **Method 2** – When an effect size for a continuous variable is not directly reported, EIF reviewers will calculate a Hedge's g from the available data, and then convert this into an improvement index score.
- **Method 3** – In cases where EIF reviewers have determined an estimate of effect is likely to be biased in the absence of adjustment for key covariates, and an adjusted effect size (or a set of adjusted means) is not available, EIF reviewers will calculate a difference-in-differences Hedge's g effect size (the g describing the magnitude of the difference between treatment and control at post-test, subtracted from the g describing the magnitude of the difference between treatment and control at pre-test), and then convert this into an improvement index score.
- **Method 4** – When an effect size for a binary variable is directly reported (odds ratio), EIF reviewers will convert this directly into an improvement index score.
- **Method 5** – When an effect size for a binary variable is not directly reported, EIF reviewers will calculate an odds ratio from the available data, and then convert this into an improvement index score.

In some cases it will not be possible to use the methods described above, when the underlying data is not reported in evaluation studies. In these cases, EIF reviewers will attempt to collect this information from programme providers and evaluators. If the data cannot be acquired, an improvement index score will not be reported.

References

Baguley, T., (2009). Standardized or simple effect size: What should be reported? *British journal of psychology*, *100*(3), 603–617. https://bpspsychub.onlinelibrary.wiley.com/doi/abs/10.1348/000712608X377117

Baird, M. D., & Pane, J. F. (2019). Translating standardized effects of education programs into more interpretable metrics. *Educational Researcher*, *48*(4), 217–228. https://journals.sagepub.com/doi/abs/10.3102/0013189X19848729

Coe, R. (2002). It's the effect size, stupid: What effect size is and why it is important. http://www.leeds.ac.uk/educol/documents/00002182.htm

Cox, D. R. (1970). Analysis of binary data. New York: Chapman & Hall/CRC

Lipsey, M. W., Puzio, K., Yun, C., Hebert, M. A., Steinka-Fry, K., Cole, M. W., ... & Busick, M. D. (2012). Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms. *National Center for Special Education Research*. https://eric.ed.gov/?id=ED537446

What Works Clearinghouse (2020). Procedures Handbook, Version 4.1. https://ies.ed.gov/ncee/wwc/Handbooks