

# What Works?

for children with speech, language and communication needs?

## What sort of evidence should we be looking for in interventions which are included in the What Works database?

**James Law and Jenna Charlton**

Whether an intervention is to be included in the *What Works* database depends on the quality of the empirical evidence underpinning it.

This document highlights the importance of specific aspects of that empirical evidence which have been considered when accepting interventions.

There are a number of different experimental designs but the key thing is that studies must have some way of demonstrating that the children included in the studies would not have improved anyway, just by virtue of getting older. Not all the points below will be relevant to those interested in providing evidence to *What Works*, as the relevance depends upon individual research methodology and design.

To determine a realistic and reliable effect of an intervention, research evidence must provide data for comparison between 'intervention' and 'no intervention'. This may be provided in a number of ways, the most common being with another comparable group of children, or comparison with the children's own earlier performance.

### Comparison between intervention and control groups

- Comparison groups may be provided with no intervention at all, with "treatment as usual" or a specific alternative intervention. The choice of comparison group clearly makes a difference to how one can generalise the results of the study.
- Randomised controlled trials are the most rigorous and unbiased way of determining whether a cause-effect relation exists between the treatment and outcome. The randomisation ensures there are no systematic differences between groups which may affect the outcome of the study.
- An alternative approach is a matched or "quasi-experimental" study in which two (or more) groups of children are matched on sample characteristics (such as age, gender, ability) which are considered likely to affect the outcomes of the study.

### Comparison "within" individuals

- It is also possible to compare whether a child's own progress is affected by the intervention.
- These studies are sometimes called "case studies" and can be "single cases studies" of a single child or "case series" studies where the results are from groups of individual children. Sometimes they are called "experimental single subject designs". Key is that there is an "experimental" aspect of the project. For example, we talk about control periods when no intervention is provided and treatment periods when it is.
- Some studies take a different approach by looking at treatment and control performance simultaneously. So, a practitioner could identify forty words that the child was not able to say or could not understand and would then teach half of these words but not the other half and the "effect" of the intervention would be the difference between the treated and the untreated words.
- A variety of different study designs for case studies are used. For example: No treatment (A)/Treatment (B) is an AB design or No Treatment/Treatment/No Treatment would be ABA. This depends on the order in which the intervention is provided and the data is collected. Sometimes people talk about "cross over designs" where one group of children would be allocated to follow an AB pattern and another a BA pattern to check to see whether there is an effect of the order in which the intervention is presented.

## What sort of evidence should we be looking for in interventions which are included in the *What Works* database?

### Measurement issues

- Intervention studies depend heavily on the quality of the measures used to assess change in the children.
- These measures should be both valid and reliable – that is other people would obtain the same result and the assessments do really measure what they set out to measure.
- These assessments should have been “standardised” on a large group of typically developing children. Informal tests developed by practitioners but not standardised are very difficult to interpret.
- Assessments can be based on the performance of the child, an observational schedule of the child (for example observations of the child interacting with others) or a parent or teacher (and sometimes child) report.
- Tests are often separated into norm referenced standard tests which tell you how close to average the child’s performance is, or criterion referenced tests which tell you whether a child has reached the same threshold as other children.
- It is always desirable that tests should have been recently standardised. This is because of the so-called ‘Flynn Effect’ where children’s performance on any test changes over time. Children tested in 2015 will do better than children who were tested on the same test in 1960s for example.

### Systematic reviews

- Systematic reviews of evidence are regarded as strong evidence for an intervention as they pool the results of studies of comparable interventions.
- They allow you to see how one study compares with another and what, on average, is the effect of a given intervention.
- They are considered to give a better overall picture of the value of a particular intervention avoiding overreliance on a single study which may or may not be representative of the other studies published for the same intervention.
- Systematic reviews have increased power and decreased bias as compared with the individual studies they include, and the pooling of outcomes can provide the most accurate overall assessment of an intervention.
- Although systematic reviews are generally considered to be very useful in providing an unbiased estimate of the real effect of an intervention, it is important to note that the quality of such reviews can vary and like the intervention studies themselves, they should be judged carefully against quality criteria.

### Summary

Understanding how a study was designed helps understand what weight to attach to its findings. There is a great deal to take into consideration when appraising evidence. The *What Works* database aims to make sound judgements based upon the quality of the experimental evidence. There are a great many ways of grading evidence but *What Works* uses a relatively straightforward three way criteria (Indicative, Moderate and Strong) to reflect that experimental evidence. It is still important to recognise that the quality of the experimental evidence is not the same as the quality of the intervention itself. So, it is possible to have good evidence that the intervention does not work particularly well or vice versa. Guidelines for meeting submission criteria based upon strength of research may be found in the ‘Support document 1- Guidelines for Submissions to the *What Works* Database’ as well as the ‘Criteria for Inclusion’ form.