Note 5.1 – On the role of meta-analysis in explaining variation in trial outcomes

Meta-analysis often takes the problem of variation in the results across trials to be a statistical matter: i.e., the view there is a meaningful average effect to be estimated for a general strategy (e.g., feedback) and the variation we see in the results across trials is due to a form of statistical variation. This variation is often assumed to be a form of random sampling variation. There is no basis for this assumption – which stems from a desire to treat an educational, scientific problem as a statistical one. Random sampling error is not a satisfactory explanation for why the results of trials of a general strategy such as feedback differ. A genuine explanation would consider a) the focus and design of the trials have been conducted (this will not be a random sample of all possible trials), b) how the educational factors vary, c) how the contexts vary and d) methodological aspects of the trials (e.g., measurements used, study limitations).

As a general rule, very little attention is paid to variation in results in a meta-analysis. The exception to this, and characteristic of high-quality meta-analyses, is when a form of correlational analysis called a 'meta-regression' is carried out. This is where the researcher looks for patterns in the effects relating to the study and intervention characteristics (such as the examples above). For example, we might collect data about the age of children in the trial and then see if there is an association between the effect of the intervention and pupil age; or we might examine whether an intervention contains certain features thought to be educationally important. In this way, we can 'code' the studies for features of the intervention and context and then do some analysis of the effect sizes, and thereby get a little beyond the average effect estimates and into questions of what works, for whom and in what circumstances.

While technical, meta-regression analyses we have seen are rarely sophisticated analyses in *educational* terms (e.g., in terms of the use and testing of educational theories of change) or in *causal* terms (i.e., going beyond simple additive, correlational models of causation). Typically they cannot be, because the requisite information is not collected in the original studies. Given the *status quo* around experimentation, IPE and reporting (as discussed in Chapter 5), the 'raw material' of the literature that systematic review and meta-analysis (SRMA) works with tends not to contain the detail required to do anything theoretically or practically sophisticated. It is also worth observing that the amount of information in a typical-length research paper can only ever be a brief and partial account. As a result, when it comes to variation in the effect size estimates, the 'explanatory power' of the data we have is often very limited.

Ultimately, we rarely know whether variation stems from differences in a) practice, b) implementation, c) context, d) the particular (non-random) 'sample' of studies that have been funded and tested, e) unexplained and/or unobserved confounding factors, or f) just random

variation and error. The statistician doesn't know but assumes it is all random sampling error from an imagined population which they assume that the studies are a sample from. The econometrician doesn't know either but hopefully has measured some variables that might be analysed in a meta-regression. There might be some big features that are e apparent; this is welcome but tends not to do more than scratch the surface when it comes to explaining variation in the results. Econometric analysis is not renowned for its ability to generate educational insights – although we note that statistical models are an important part of the EEI approach, as discussed in Part C, sometimes to good effect; there may be some lessons for meta-regression and trial design (in terms of IPE measures) for WW to be found there.