Professor John Jerrim (UCL Institute of Education) and Dr Alex Sutherland (BIT) peer-reviewed the updated Magic Breakfast report. They agreed that, since there was an error in the random assignment, it would be most appropriate to treat the study as a matched comparison group study. The main method of analysis now used to judge the effectiveness of the Magic Breakfast intervention is regression. Using the EEF criteria, the Magic Breakfast study could receive a maximum of three padlocks using this approach, but only if the treatment and control groups are "well-matched".

Jerrim and Sutherland agreed that the Key Stage 2, but not Key Stage 1, analysis can be considered "well-matched". The critical difference is in in the quality of the measure of prior achievement. In particular, the Key Stage 1 analysis only controls for a binary measure of prior "achievement" based upon the Foundation Stage Profile (whether the child had achieved a "good level" of development or not). This is a coarse measure of children's pre-intervention abilities, and captures quite different skills to the outcome measures. As prior achievement is potentially a key confounder, Jerrim and Sutherland felt that this measure was not of sufficient quality to verify the claim that the intervention and control group were "well-matched".

It was therefore agreed that a two-padlock rating was most appropriate for the Key Stage 1 result. As the Key Stage 2 analysis included a stronger measure of prior achievement, Jerrim and Sutherland agreed that three-padlocks could be awarded to this part of the study.

Jerrim and Sutherland also discussed the issue of whether the IFS could continue to be considered *independent* evaluators, mainly due to the potential for what is known as confirmation bias[1]. It was agreed that this was unlikely to represent a major issue, particularly given that the results now reported for Key Stage 2 are somewhat different to the initial report. It was nevertheless recommended that the EEF commission an independent third party to "stress test" the result, using a range of different methods and model specifications. Jerrim and Sutherland felt that this should not delay the publication of the revised report, but that the padlock ratings awarded be considered 'provisional' until this has been completed.

Professor John Jerrim, IoE

Dr. Alex Sutherland, BIT

---

[1] Confirmation bias is where one (either consciously or sub-consciously) favours information that confirms previously held beliefs or results. In this case, given that a positive effect was initially reported by the IFS, there might be a concern that further analysis that confirms this finding might be favoured over analysis that would reject it.