



**Some basic observations on conducting a systematic review: A brief reply to Goldstein, Vatalaro and Yair**

Journal:	<i>Journal of Children's Services</i>
Manuscript ID	JCS-11-2017-0051.R1
Manuscript Type:	Research Paper
Keywords:	Parental engagement, school attainment, systematic review, education, parental involvement, robust evaluations

SCHOLARONE™  
Manuscripts

1  
2  
3 Some basic observations on conducting a systematic review: A brief reply to Goldstein,  
4 Vatalaro and Yair  
5  
6

7 This paper is written in response to an attempted rebuttal of our prior paper in this journal  
8 (Author 2015), by Goldstein et al. (2017). Our paper reported a systematic review of  
9 interventions to engage parents more in their children's education, in order to raise school  
10 attainment. Goldstein et al. make a large number of unwarranted claims about the quality of  
11 our paper which reported a systematic review. They reproach us for using reports of  
12 unpublished evidence, for mis-labelling or mis-describing studies, and for denigrating studies  
13 by labelling them as 'bad'. We were very surprised when first alerted to this response and  
14 went back to look at all of the research reports that Goldstein et al. claimed we mis-  
15 represented in our assessment. We found that their claims are false and based on a very poor  
16 understanding of how evidence is reviewed. In our reply, we look first at why unpublished  
17 material *must* be included in a review like ours, and why the outlet for publication is not  
18 relevant, then at appropriate designs for causal questions, and at the confusion in Goldstein et  
19 al. between the quality of an evaluation and the impact of an intervention. We look at many  
20 examples where the confusion leads to Goldstein et al. making incorrect assertions about our  
21 paper, in order to make the point that their whole idea of how to conduct a systematic review  
22 is wrong.  
23  
24  
25

### 26 **The development of the review**

27  
28  
29 This brief paper is a response to an attempted rebuttal to a report of a systematic review we  
30 previously published in this journal (Author 2015). The paper has substantive implications  
31 for attempts to improve the school attainment of children and young people, and it is hoped  
32 that it will also help some readers to understand a little more about how to conduct a  
33 systematic review. As explained in our original paper (p.3), we had originally done a much  
34 larger review of parental and child attitudes, aspirations, and behaviours for the Joseph  
35 Rowntree Foundation:  
36

37 Author (2012) conducted a wide-ranging review of the evidence linking attainment to  
38 attitudes and aspirations and concluded that only parental involvement in education  
39 offered any promise as a causal contributor to attainment.  
40

41 The overall situation currently is that there is an established link between parental  
42 involvement and child performance at school. There are interventions that have been  
43 shown to enhance parental involvement and may improve wider outcomes such as child  
44 behaviour or attitudes. But we do not know whether intervening will improve  
45 attainment at school.  
46  
47

48 In response to this preliminary promise we were then funded by the Nuffield Foundation to  
49 do another review focused only on parental engagement interventions. This is a perfectly  
50 normal iteration in research, and we found that the promise from the first less focussed  
51 review was not realised on closer examination. For some reason, Goldstein et al. see this as  
52 being inconsistent and even dishonest – somehow “contradicting [our] own prior studies and  
53 reviews” (p.1).  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Our review was not about other aspects of parental involvement or its benefits, but solely  
4 about the evidence (or lack of it) that enhancing parental involvement can be shown to lead to  
5 improved child attainment at school. There could be many benefits from enhanced parental  
6 engagement for both parties but the evidence is that it does not improve school attainment.  
7 We could have made mistakes along the way or mis-transcribed from data extraction  
8 documents. **This is always possible in a large complex synthesis.** We are aware of one study  
9 in the original review where the results for reading and maths were transposed.  
10

## 11 12 **Conducting a systematic search**

13  
14 All guides, authorities and resources on the conduct of a systematic review make it plain that  
15 the search and synthesis involved must include as much research relevant to the research  
16 question as possible (e.g. Torgerson 2003, Author 2013). The EPPI Centre resources  
17 (<https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=67>) state that:  
18  
19

20 Systematic reviews aim to find as much as possible of the research relevant to  
21 the particular research question  
22

23 The international Campbell (and Cochrane) Collaboration  
24 ([https://www.campbellcollaboration.org/research-resources/writing-a-campbell-systematic-  
25 review/systemic-review.html](https://www.campbellcollaboration.org/research-resources/writing-a-campbell-systematic-review/systemic-review.html)) is equally clear:  
26

27 Campbell reviews must include a systematic search for unpublished reports (to avoid  
28 publication bias).  
29  
30

31 Because of the file-drawer problem, whereby studies published on any topic are more often  
32 positive than those not subsequently published, relying only on the most easily-accessible or  
33 well-known reports in any review of evidence will lead towards bias. To ignore this, as  
34 Goldstein et al. **advocate and so presumably do in their** own literature reviews, is an  
35 elementary error. In the review reported in our earlier paper we made a deliberate effort to  
36 find 'grey' literature as well as papers published in journals and official reports to funders.  
37 This is the correct approach. Therefore, our synthesis of evidence correctly includes studies  
38 from a range of sources. This seems to confuse Goldstein et al., who say:  
39  
40

41 The examples that the authors chose to highlight should have been the ones formatted  
42 as academic papers. (p.14)  
43

44 This statement is clearly and totally false, and goes against the systematic element of a  
45 systematic review. Goldstein et al. repeatedly complain that we used "unpublished papers"  
46 that "have rarely passed a peer review process that would make them candidates for serious  
47 scientific review" (p.3). Again this shows **lack of knowledge** of how a review of evidence  
48 should be conducted (and presumably also on the part of the reviewers and editors of the  
49 journal **that published this purported** 'rebuttal'). We do not accept the source of **any**  
50 publication or the status of its author or funder as any guarantee of research quality, unlike  
51 Goldstein et al. who say things throughout like:  
52

53 ...received a high rating by the Home Visiting Evidence of Effectiveness... systematic  
54 review (p.10)  
55  
56  
57  
58  
59  
60

1  
2  
3 ...conducted according to very high standards as evidenced by its publication in  
4 Developmental Psychology, a very well respected journal (p.10).  
5

6 These **unwarranted** comments are the antithesis of a review approach that must treat an  
7 unpublished report in the same way as one in a highly cited **journal. Goldstein** et al. also try  
8 to draw a distinction between academic and non-academic work:  
9

10 Other untrustworthy unpublished research that Author cite in their paper included what  
11 we are labelling as non-academic research (p.14)  
12

13 An example they use to illustrate this point about what they are calling untrustworthy  
14 unpublished non-academic research is Nutbrown and Hannon (2011). Cathy Nutbrown and  
15 Peter Hannon are full professors in the School of Education, Sheffield University, UK. They  
16 are very far from being non-academic, and Nutbrown and Hannon (2011) is not a “one-page  
17 conference poster” as stated by Goldstein et al. (p.14) but a five-page referenced project  
18 summary [http://real-  
19 online.group.shef.ac.uk/docs/The%20aims,%20structure%20and%20framework%20of%20th  
20 e.pdf](http://real-online.group.shef.ac.uk/docs/The%20aims,%20structure%20and%20framework%20of%20the.pdf).  
21  
22

23 Our paper and the work it was based was right to include unpublished material, in order to try  
24 and reduce publication bias, despite the extra work that this involves. Goldstein et al. are  
25 wrong not to do the same in their own studies, and of course they are wrong to criticise us for  
26 doing the right thing.  
27

28 We searched for robust evaluations of parental engagement interventions with school  
29 attainment outcomes. The causal question of our review is best addressed by randomised  
30 control trials, or regression discontinuity designs, or similar (Author 2017). We started with  
31 an acceptance of an association between parental interest in their child’s education and school  
32 outcomes but wanted to know if **there** was more than an association. Therefore, it is clear that  
33 only such active designs as RCTs could provide strong evidence of impact. Research that  
34 used a passive approach such as *post hoc* statistical modelling of longitudinal data could not  
35 do so, however large it was, however complex its analysis, **and wherever it was published**.  
36 We ignored studies that had no impact data relevant to school outcomes, and we ignored the  
37 parts of **other** studies that reported no impact data relevant to school outcomes. All of this is  
38 explained **clearly in our** original paper.  
39  
40  
41  
42

### 43 **Separating quality and impact**

44  
45 Once we had assembled the studies for full review, we read and judged them in terms of their  
46 quality (graded 0 to 4 as per Author 2017) and their impact on attainment (none/harmful,  
47 mixed/unclear, and positive). We took no account of the report authors’ claims (unlike  
48 Goldstein et al.), instead using only the evidence they presented about quality and impact.  
49 Quality was judged in terms of design (as above), scale, attrition and so on. Impact was  
50 judged in terms of ‘effect’ size if reported, or an approximation **based on what was reported**,  
51 if not. And we said in our paper:  
52

53 Therefore, each study leads to two judgements – one on how effective the intervention  
54 has been, and one on how trustworthy the evidence for it is. It is important to realise  
55 that these are completely independent of each other. (p.254)  
56  
57  
58  
59  
60

1  
2  
3  
4 Goldstein et al. appear not to have understood this, and they collapsed our two dimensions  
5 into one (p.5), and so confused the impact results of any study and its quality of evidence  
6 throughout the rest of their paper. It is this kind of confusion – the idea that good studies also  
7 have positive results – that underlies publication bias. So their Table 1 has rows running  
8 across the actual results of any study and calling any studies with strong evidence ‘promising’  
9 for the purposes of parental engagement interventions - whether the impact results are  
10 negative, unclear or positive. This leads them (or their readers) to confusion throughout as to  
11 whether we had said a piece contained strong, medium or weak evidence and whether the  
12 research reported evidence that the intervention worked or not.  
13

14  
15 They create new names for their four resulting categories in Table 1 such as ‘bad’ and  
16 ‘untrustworthy’ research (p.5). And in the rest of the paper they repeatedly criticise us for  
17 labelling pieces of research as bad or untrustworthy. However, we do not use either term in  
18 the paper (as a simple electronic search will attest). Their ‘bad’ category conflates any  
19 research we found that was based on weak impact evidence, whether it reported no benefit  
20 from the intervention or were mixed or unclear as to the impact. But this misguided  
21 conflation gives them an unfounded reason to suggest that it is we who have used these  
22 inappropriate academic terms. For example, they say (p.8) “Our qualm with interventions that  
23 were categorised as bad is twofold”. This is absurd for so many reasons. Goldstein et al. are  
24 confusing our discussion of the quality of evaluation of the impact of interventions with the  
25 quality of the *interventions* themselves. Anyway, we do not have a category labelled ‘bad’,  
26 and throughout are only concerned with the quality of evidence of impact, and the level of  
27 impact, as portrayed by research reports.  
28

29  
30 As another example, Goldstein et al. state (p.12) “Author categorized a total of 55 studies as  
31 untrustworthy”. This is also just plain wrong. The term untrustworthy does not appear  
32 anywhere in our paper. Goldstein et al. have created their own, invalid, categories and then  
33 criticised us for using them (even though we did not). It is, presumably, much easier to make  
34 up something and criticise us for something we did not say than to engage with what we  
35 actually reported.  
36  
37

### 38 **Checking our claims about studies**

39  
40  
41 Oddly, given their claim that we set out to ‘denigrate’ the work of others, in their first  
42 example they claim we have actually been too generous. Chavkin et al. (2000) was based on  
43 five schools with around 1,600 students. The impact evidence comes from the official test  
44 results in four schools. It is true that Chavkin et al. also report some in-depth work with a  
45 smaller number of students, but we ignored this because, as explained above, we were only  
46 concerned with the impact evaluation at this stage. We rated the impact as being medium  
47 quality because of its large scale, use of a (weak) comparator, and the standardised official  
48 test results. If we had rated it of lower quality as Goldstein et al. wish us to do this would  
49 actually strengthen our conclusion that there is no solid evidence yet that enhancing parental  
50 engagement works to raise child attainment.  
51

52  
53 The second example is even odder. We rated Reynolds et al. (2011) as also having medium  
54 quality evidence but no better than this because it is not an RCT or RDD (i.e. it did not have a  
55 control group). The study had comparison “groups matched on age, eligibility for  
56 intervention, and family poverty”. There is therefore no randomisation to treatment, or  
57

1  
2  
3 random selection of cases, and so the use of significance testing in the Reynolds et al. paper  
4 is meaningless. This is so despite being peer-reviewed and published in the journal Science.  
5 We cannot do what Goldstein et al. do, which is to accept the outlet as any guarantee of  
6 quality. Our approach to judging trustworthiness therefore puts Reynolds et al. (2011) and  
7 Chavkin et al. (2000) in the same broad category of medium quality evidence. They both  
8 represent studies that need to be taken seriously in any review without being definitive or  
9 conclusive at all – hence ‘inconclusive’.

10  
11 In the actual classification in our paper that Goldstein et al. claim to be responding to we have  
12 a section labelled ‘unpromising approaches’. Work could be in this section either because  
13 there was no real evidence or because the studies suggested that the approach had no  
14 beneficial impact. Goldstein et al. (p.9) take exception to Necoechea (2007) being in this  
15 section and claim that it is reasonable evidence, based on a comparator and reporting ‘effect’  
16 sizes. This is true, and in no way contradicts what we said. However, it cannot be rated any  
17 higher because it only involved 52 cases. Anyone actually reading our paper will see that all  
18 we say about the study is that it found no good evidence of benefit from HIPPI. Again this is  
19 true. Incidentally, this unpublished study rated highly by Goldstein et al. should have alerted  
20 them to the fact that not all reasonable studies are published (and that not all published  
21 studies are high quality). They are being inconsistent here.

22  
23  
24 Our review was looking for evaluations of interventions, and so the most appropriate designs  
25 would be RCTs, RDD or similar (Author 2017). Post hoc modelling of existing longitudinal  
26 data is useful in identifying risk factors or similar, but provides no robust test of impact  
27 (Author 2013). The latter is true of Chang et al. (2009, labelled 2009b in Goldstein et al.).  
28 Chang et al. (2009b) has a passive design, and no randomisation, again making their use of  
29 significance testing incorrect and misleading. Goldstein et al. suggest that another study by  
30 Chang et al. (2009a) is better. It is not. It is also based on passive modelling of results, and  
31 has more than 30% of cases missing through attrition.

32  
33  
34 Goldstein et al. state (p.12) that the Chang et al. “studies certainly did not show mixed,  
35 unclear, or no benefit as evaluated by Author”. Again actually reading our paper shows that  
36 the only reference to Chang et al is “The situation for Head Start and Early Head Start is  
37 similar with relatively weak studies, including those with no comparators or unfair ones,  
38 showing both benefits and no benefits (Chang et al., 2009; Starkey and Klein, 2000; Hughes,  
39 2003; St Pierre et al., 2005).” The heading of our section was “Studies with inconclusive  
40 evidence” and included those claiming the intervention had been beneficial and others. Our  
41 concern here is with quality not whether the paper portrayed benefit or not, as our sentence  
42 makes clear to any reader. Goldstein et al. have fallen foul of their confused conflation of our  
43 clear separate consideration of quality and impact.

44  
45  
46 St Pierre et al. (2005) is described by Goldstein et al. as a key study of a “very high standard”  
47 but had no pre-test and so did not establish baseline equivalence as any high standard trial  
48 should. It anyway showed, as Goldstein et al. agree, that early parental intervention was not  
49 effective. It is hard to see how this can be used as an example of us downplaying the  
50 effectiveness of parental engagement.

51  
52  
53 Starkey and Klein (2000) is described by Goldstein et al. as “a superbly conducted study”. But  
54 it had only 28 cases (and so was not rated as high or medium quality) and found no difference  
55 in literacy outcomes between Headstart children and the comparison group. As we reported  
56  
57  
58  
59  
60

1  
2  
3 accurately it showed “both benefits and no benefits” (p.7). Again it is hard to see what  
4 Goldstein et al. are objecting to in our categorisation and reporting of this.  
5

6 In discussing Sylva et al. (2008), Goldstein et al. (p.17) again confuse the 100 or so children  
7 in the trial or impact evaluation, with the larger numbers of teachers and parents completing a  
8 survey. They criticise us for saying the trial was small with only around 50 children in each  
9 arm of the trial. They claim this is incorrect because the trial arms were 60 and 52  
10 respectively. But this ignores attrition and missing values. In the analysis, as reported by  
11 Sylva et al., there *were* around 50 children in each arm of the trial.  
12

13 We are also criticised for stating that Moore (2011) had 322 students in the HIPPY group  
14 when there were actually 1,194 according to Goldstein et al. (p.17). For the removal of doubt  
15 here we quote Moore (2011, p.13) exactly:  
16

17         The participants of the study were 322 students in the HIPPY program and 3,577  
18 nonHIPPY students.  
19

20  
21 As with all of these examples from their supposed ‘rebuttal’, Goldstein et al. are just plain  
22 wrong.  
23

### 24 25 **Checking suggested studies**

26  
27 Goldstein et al. complain that we did not quote some of the studies referenced by all of the  
28 studies that we do cite. This of course is a matter of paper length and timing, and some of  
29 these suggested studies have already been covered (above). We did not include very old  
30 studies or repetitions of the same study. We did not claim we reported every study – merely  
31 that new studies were unlikely to alter our conclusions by much. A study we could have  
32 included was Zigler et al. (2008) suggested by Goldstein et al. (p.14) as being one of  
33 “numerous esteemed publications about PAT that included RCTs”. However, Zigler et al.’s  
34 (2008) study was *not* an RCT. It was another passive design, modelling non-randomised data  
35 with the inappropriate use of significance testing - “The study used path analysis to test  
36 hypothesized models of how the Parents as Teachers (PAT) program affects children's school  
37 readiness and subsequent third-grade achievement” (p.103). Including it or any other of the  
38 suggested papers would have made no substantive difference to our conclusions. Goldstein et  
39 al. do not seem to know what an RCT is.  
40  
41  
42  
43

### 44 **Conclusion**

45  
46 Goldstein et al. make the elementary mistake, when attempting to review evidence, of  
47 accepting what researchers claim their evidence shows in reports rather than the reported  
48 evidence itself. It is only the latter that was being systematically reviewed by us. They  
49 confuse our different studies and searches over several years, with **appropriately** different  
50 objectives and research questions, in what is claimed to be a ‘rebuttal’ of our paper in this  
51 journal. They ignore the genesis of our paper as explained carefully in its introduction (and  
52 summarised above). And they choose to conflate the quality and the results of study in their  
53 Table 1 and throughout. They then use **their own confusion** to attribute to us motives for  
54 which they have no evidence at all, and to call us dishonest in an academic article, again on  
55 the basis of no evidence. **They were permitted by the journal editors to say:**  
56  
57  
58  
59  
60

This method appears to be motivated by a search for errors and irregularities, rather than an honest evaluation of evidence (p.15).

They have purposely focused their analysis on specific programs and unpublished or lesser known publications” (p.21).

This claim that we were behaving dishonestly should not have been allowed in the journal (and we should have been alerted to this attempt at rebuttal immediately, rather than having to find out about it after publication). We hope that readers will realise that we were trying to encapsulate a huge body of work, and we did so as scrupulously as possible. Mistakes may have been made, as in any large enterprise, but inadvertently so and not the kinds of elementary errors portrayed in this purported rebuttal. We recognise that in every review some studies will have been missed, but the key is whether including these studies would have altered the overall results. As shown from the examples given here – they would not. The Goldstein et al. ‘rebuttal’ adds nothing but confusion to progress in this area.

## References

- Chang, M., Park, B., Singh, K. and Sung, Y. (2009b) Parental involvement, parenting behaviors, and children’s cognitive development in low-income and minority families, *Journal of Research in Childhood Education*, 23, 3, 309-24
- Chang, M., Park, B. and Kim, S. (2009a) Parenting classes, parenting behavior, and child cognitive development in Early Head Start: A longitudinal model, *The School Community Journal*, 19, 1, 155
- Chavkin, N., Gonzalez, J. and Rader, R. (2000) A home-school program in a Texas-Mexico border school: voices from parents, students, and school staff, *The School Community Journal*, 10, 2, 127-37
- Moore, O. (2011) *The Home Instruction for Parents of Preschool Youngsters (HIPPO) program’s effect on academic achievement of TAKS tests*, unpublished dissertation, University of North Texas, Denton, TX
- Necoechea, D. (2007) *Children at-risk for poor school readiness: the effect of an early intervention home visiting program on children and parents*, unpublished PhD dissertation, University of California, Riverside, CA
- Nutbrown, C. and Hannon, P. (2011) *Effects of the REAL Project: Raising Early Achievement in Literacy*, available at [www.sheffield.ac.uk/research/impact/stories/fcs/22](http://www.sheffield.ac.uk/research/impact/stories/fcs/22)
- Reynolds, A., Temple, J., Ou, S., Arteaga, I. and White, B. (2011) School-based early childhood education and age-28 well-being: effects by timing, dosage and subgroups, *Science*, 333, 6040, 360-64
- St Pierre, R., Ricciuti, A. and Rimdzius, T. (2005) Effects of a family literacy program on low-literate children and their parents: findings from an evaluation of the even start family literacy program, *Developmental Psychology*, 41, 6, 953-70
- Starkey, P. and Klein, A. (2000) Fostering parental support for children’s mathematical development: an intervention with head start families, *Early Education and Development*, 11, 5, 659-80
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I. and Taggart, B. (2008) *Final report from the primary phase: pre-school, school and family influences on children’s development during Key Stage 2 (7-11) DfE RR 061*. Nottingham: The Department for Education
- Torgerson, C. (2003) *Systematic reviews*, London: Continuum



Zigler, E., Pfannenstiel, J., and Seitz, V. (2008) The Parents as Teachers program and school success: A replication and extension, *The Journal of Primary Prevention*, 29, 2, 103-120

Journal of Children's Services