

Ofsted is a very powerful and influential organisation in English education, which has a history of producing very useful subject reports that demonstrated what was happening in schools and gave indications of what they considered good practice. There has been a long gap since the 2012 Mathematics: Made to Measure Report and the return of subject reports was welcomed. In the Principles behind the reviews document (Ofsted, 2021) it says that, "*Educational research is contestable and contested, and so are documents such as these research reviews. Therefore, we are sharing our thinking with subject communities so that we can get input from the broader subject community.*" We hope that this submission will be welcomed as input from the broader mathematics education community.

In the Principles document Ofsted has stated:

*Our aim is that the reviews will support and inform those leading the thinking on subject education in our schools. Professionals from the education sector will also be able to see the research that is informing our conception of **a high-quality education** in a variety of subjects.* (Ofsted, 2021)

The Principles document goes on to say how this conception of a high-quality education will be used.

*When our inspectors carry out subject 'deep dives', they draw on a shared understanding of **a high-quality education**.* (Ofsted, 2021)

This demonstrates the importance of the Research Reviews to all involved in education and why it is vital that these reviews are of the highest quality.

AMET, the Association of Mathematics Education Teachers, has identified some serious problems with the scholarship in the recent Ofsted Research Review series: Mathematics document (published 25th May 2021). As many of our members are involved in research, as well as setting and marking academic assignments, we take scholarship very seriously. We teach students how to compile arguments, use sources with criticality and reference their work. When people read a paper they need to be able to trust that the references cited support the point that is being made. Unfortunately, our analysis of the Mathematics Research Review has shown that many of the references do not support the points that Ofsted has made. We emphasise that this complaint is not about the content of the report but on the poor match between the content and the references which are used to support it.

When we noticed some problems with certain references we decided to go through all 201 footnotes and 307 references thoroughly, reading each source and comparing it to the Ofsted statement. This has taken a team of people several weeks to complete. We felt that AMET is particularly well-placed to do this checking because of our familiarity with academic literature, our access to it and our understanding of educational research. A goodly number of the sources were open access, which means that classroom teachers would be able to download them. However, many of the sources were only accessible behind pay walls. Since many of our members work in universities we have access to a large range of university libraries and have had to use these to access many of the sources. We were also contacted by several authors who outlined how they felt their work had been misused.

We have included a table in this submission that goes through the sources in all 201 footnotes and 307 references and identifies where there are problems with them. We found 3 sources cited for information (1%), 139 matches (45%), 79 partial matches (26%) and 86 that did not match (28%).

The Principles document clearly sets out the approach that Ofsted is using:

The research reviews will set out the research that has informed our thinking on subject quality. When selecting literature for the reviews, we are drawing on research that aligns with the established principles for quality of education, as outlined in the [education inspection framework \(EIF\)](#) and summarised in our [‘Education inspection framework: overview of research’](#). (Ofsted, 2021)

This is quite different from the interpretation of a research or literature review that most teachers will have from their undergraduate and / or post-graduate degrees. A literature review starts with the literature and a thesis is then developed from this considering contrasting viewpoints, rather than starting with a thesis and seeking only supporting literature. Because of this we feel that it would be clearer to the education community to call these documents position papers rather than research reviews.

A key element of a research review is criticality, considering when, where, how, why and by whom the source was created. This helps to determine how generalisable the findings are and how relevant they are to the current, local situation. We welcome the fact that Ofsted have included a wide range of types of research, rather than looking exclusively at Randomised Control Trials. However, this means that the generalisability of the source must be considered. Footnote 166 consisted of a single source that involved research with four children in the United States. This is not a secure research base for making a general point for UK education. Small scale studies are often useful for looking at an issue in depth but their conclusions cannot be assumed to apply to all situations. A research review would normally highlight some details about the study being cited so that the reader would be able to judge the strength of the claim. That has not been done in this review and it is only by tracking down and reading the individual studies that this can be determined. The research presented also appears to be uncontested because only supporting sources have been used. The result is that the research base appears to be stronger than it is.

A wide range of countries was represented in the review, which is good because it is important to look beyond our borders to see how education is conducted elsewhere, as long as any findings are treated with caution since education has a complex relationship with culture. More than 40 of the citations related to East Asian countries. This seems reasonable since these countries have performed consistently well in mathematics in international studies such as TIMSS and PISA. However, it was surprising that more than three times this number of references came from the USA. The United States has not been a consistently high performing country in these international comparisons; in fact we outperformed them in the recent PISA (USA 478, UK 502) and TIMSS at Grade 4 (USA 535, England 556), although we had the same score at TIMSS Grade 8 (USA 515, England 515). Their education system is quite different to ours, with the curriculum being determined locally rather than controlled nationally. In fact, several of the American sources cited were calling for the creation of a standard curriculum in the USA so it was not clear why these were being used for a review in England that already has a National Curriculum. There were far more citations relating to the USA than there were to England / the UK. The rationale for this was not clear.

The age of the sources is sometimes a concern. The report includes research ranging from an article published this year to an article from 1980, with a book from 1939 used to illustrate a historical point. Approximately 50% of the references come from sources published within the last 10 years. However, that means that nearly half of the sources were published before the current National Curriculum. This can result in discussing practices that no longer occur. Choosing a study from 1988

for a statement related to computer use (footnote 161) seems ill-advised because of the fast rate of change in technology and the huge differences in children's familiarity with computers.

The original sources also need to be accessed to identify the age groups involved in the research studies. Most of the studies did involve school children but some were conducted with college or university students. Most of the studies with school age children were conducted in a specific phase or year group, e.g. secondary or Year 4, but they have often been used to support statements that generalise beyond these ages. This can be a particular problem with early years, especially since different countries start formal education at different ages. For example, footnote 46 relates to research with 1st and 2nd grade children (6 - 8 years old) in the USA but is erroneously equated to the "start of the academic journey" in England, which would be nursery or reception classes (3 - 5 years old).

Overall, we do not feel that the evidence base for this report, which sets itself out as a research review, is secure and call for its withdrawal until this can be corrected.

AMET

References

Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 International Results in Mathematics and Science*. Retrieved from Boston College, TIMSS & PIRLS International Study Center website: <https://timssandpirls.bc.edu/timss2019/international-results/>

OECD (2021). Mathematics performance (PISA) (indicator). doi: 10.1787/04711c74-en (Accessed on 10 July 2021)

Ofsted (2021). Principles behind Ofsted's research reviews and subject reports. <https://www.gov.uk/government/publications/principles-behind-ofsteds-research-reviews-and-subject-reports/principles-behind-ofsteds-research-reviews-and-subject-reports>