

Reply to Tooley and Longfield (2014) by the Private Schools Rigorous Review Team

October 2014

SUMMARY

We thank Tooley and Longfield for their keen interest in our rigorous review of 'The role and impact of private schools in developing countries'. In response to issues raised in their document, the review clearly sets out the search strategy and criteria for the inclusion of studies, following the approach established by DFID's 'How to note'. This approach rigorously and objectively interrogates a number of hypotheses and assumptions that underpin the polarised debate about the potential and real contribution of private schools to improving education for children.

The rigorous review underwent external peer review at three stages: a) initial protocol detailing method and theory of change b) draft report and c) final report. Feedback from external peer reviewers in writing and in roundtable meetings with a mix of international experts in the field was fully taken into account at each stage. The issues raised by Tooley and Longfield were not raised by previous reviewers in their detailed feedback. We have thoroughly investigated each of the challenges proposed by Tooley and Longfield and have found that the stronger policy conclusions they reach in their document are not justified by the evidence.

Our review identified that arriving at general conclusions is difficult because of the diversity of the private school sector, the significant gaps in the evidence, and the fact that available research is rarely generalisable in itself. While the findings cannot be universally translated into policy regardless of context, they do merit policy-makers' attention. There is more contestation than there is consensus in the literature, with many findings inconsistent and some being outright contradictory. Further research in targeted areas could lead to a strengthening of this emerging evidence base.

**Reply to Tooley and Longfield (2014) by the Private Schools Rigorous Review Team
October 2014**

We thank James Tooley and David Longfield for their keen interest in the rigorous literature review of *The Role and Impact of Private Schools in Developing Countries* (Day Ashley, Mcloughlin, Aslam, Engel, Wales, Rawal, Batley, Kingdon, Nicolai and Rose, 2014) commissioned by the Department for International Development (DFID), as expressed in their document, *The Role and Impact of Private Schools in Developing Countries: A Response to DFID's 'Rigorous Literature Review'* of October 2014. We welcome the opportunity to check the validity and rigour of the review's findings on the basis of their response.

1. The challenge

- 1.1. Tooley and Longfield have not substantially challenged the rigorous review's overall search strategy, the overall findings from 11 out of the 17 assumptions presented, and the gaps in the evidence identified by the review.
- 1.2. Tooley and Longfield's challenge relates to three areas which are addressed in sections 3-6 below:
 - *Reading of the evidence*: based on their different interpretations of the evidence, Tooley and Longfield suggest a re-classification of a number of the studies under six of the assumptions included in the review. (This is addressed in section 3. Classifications of specific assumptions are discussed in section 6.)
 - *Framing of the assumptions*: Tooley and Longfield suggest a rewording of three assumptions¹. Each of these would give a more favourable view of the role of private schools.² (This is addressed in section 4.)
 - *Evidence missed or duplicated*: Tooley and Longfield suggest that evidence is missed or duplicated to such an extent that the overall findings would change. (This is addressed in section 5.)
- 1.3. To produce this rejoinder to those specific challenges, researchers from the review team have thoroughly re-checked classifications by referring back to original data extraction templates and to the original studies included in the review. Advisers within the team and the team leader have then carefully reviewed these responses, including thorough cross-checking against the original studies as appropriate.

2. General response

- 2.1 Following a competitive bidding process the team was appointed to carry out the review with the objective of assessing recent evidence on the role and impact of private schools in DFID-priority countries and identifying critical research gaps. It was one of a series of rigorous literature reviews commissioned by DFID in relation to their increased emphasis on

¹ Tooley and Longfield also propose removing five assumptions under the 'enabling environment' theme. See 4.4 below.

² Tooley and Longfield also critique the definition of private schools adopted in the review. See section 4.5 below.

evidence-based policy-making and programme design and was carried out according to the detailed terms of reference agreed with DFID.

2.2 The rigorous review underwent external peer review at three stages: (i) initial protocol detailing method and theory of change, (ii) draft report, and (iii) final report. The selection of external peer reviewers was at DFID's discretion. Feedback was fully taken into account at each stage from (i) external peer reviewers in writing and (ii) roundtable meetings with a mix of international experts in the field. The issues raised now by Tooley and Longfield (2014) were not raised by previous reviewers in their detailed feedback.

2.3 The resources committed to undertaking this review may not have been comparable to that of a systematic review covering a similar breadth of themes, nevertheless in our approach we strived to mimic several of the internal procedures used in a systematic review to ensure neutrality and rigour of analysis. This included, in brief:

- a multi-disciplinary team of six experienced researchers and four senior expert advisers, selected by DFID following a competitive bidding process;
- clear search protocols and criteria for the inclusion of studies in the review;
- the use of common criteria against which to assess the evidence³;
- a thorough review of selected studies to identify the rigour of the methodology used, and to assess their research findings;
- procedures for cross-checking assessments within the team;
- protocols to ensure that team members were not involved in assessing their own work⁴;
- a collaborative team approach involving feedback at each stage of synthesis;
- an internal expert panel of advisers to oversee the entire process.

3. Reading of the evidence

3.1 Our review classifies bodies of evidence according to the number of studies in which evidence supported (positive), countered (negative) or was ambiguous (neutral⁵) in relation to the assumptions. Additionally, it classified the strength of the bodies of evidence (i.e. strong, moderate and weak) under each of the assumptions using the criteria of *quality, size, context and consistency*⁶. These assessments are transparently presented in the review so it is clear how the strength of evidence was arrived at.

3.2 On the basis of our thorough re-assessment, the review team has found no grounds to uphold any of the suggested reclassifications of overall bodies of evidence for four of the six assumptions (Assumptions 1, 3, 4 and 7) challenged by Tooley and Longfield. This is explained in more detail in sections 6.1, 6.3, 6.4 and 6.6, respectively.

3.3 There may be grounds to reclassify the overall evidence for two of the assumptions (Assumptions 6 and 8) if further evidence were included under these assumptions as identified by Tooley and Longfield from the studies included in the review. However, the

³ The assessment of quality of individual studies was carried out in accordance with DFID's *How to note* (DFID 2013).

⁴ To prevent any potential conflict of interest, team members who were also authors of relevant studies were not involved in decisions about the inclusion of studies in the review, nor in reviewing their own work – these tasks were carried out by other members of the team and this policy was strictly adhered to.

⁵ A neutral finding may be understood as 'ambiguous' and therefore in need of more investigation.

⁶ The assessment of the strength of bodies of evidence was based on DFID's *How to note* (2013).

inclusion of these studies would not result in the reclassifications that Tooley and Longfield propose i.e. *moderate positive*. This is explained in more detail in Sections 6.5 and 6.7 below.

4. Framing of assumptions

- 4.1 Tooley and Longfield suggest reframing Assumptions 3, 4 and 8. However, it would be methodologically unsound to reformulate any of the assumptions now *after the analysis has been done* as these are the testable assumptions against which the evidence was originally assessed. The review provides a clear and transparent trail from each testable assumption to the narrative detailing and discussing the nuances arising from the evidence, to the synthesis of evidence and key headline finding, to the overall assessment of the body of evidence.
- 4.2 The team is fully aware that there are various ways in which the assumptions may be precisely worded. The 17 assumptions in the review were carefully formulated and based on a Theory of Change and associated hypotheses developed by the team with the intent of maintaining balance and objectivity. The hypotheses that were identified, and used as the basis for the assumptions, were selected on the grounds that they had clear practical implications for policymakers and were testable. The Theory of Change, hypotheses and assumptions were reviewed by DFID and by external peer reviewers appointed by DFID, and changes were made on the basis of their feedback.
- 4.3 The assumptions were also used instrumentally to enable the review to capture relevant data on the key issues, recognising that authors of the studies may use different definitions of concepts. The team was careful not to impose definitions in these cases and to acknowledge the internal validity of the studies, particularly since the studies included in the review involved a range of different types of methodologies.
- 4.4 Tooley and Longfield (2014) propose reducing the 17 assumptions investigated in the review to 12⁷ assumptions, removing the assumptions under the 'enabling environment' theme. The Theory of Change sets out all the steps that are viewed as significant for achieving the desired outcomes covered in the review. It is, therefore, appropriate to address these steps through the assumptions. Assigning more or less weighting to one or another hypothesis in the theory of change was not appropriate for the review to pursue its course, which was to individually test each of the assumptions in turn.
- 4.5 Additionally, Tooley and Longfield (2014) question the definition of 'private school' used in the rigorous review. The definition that has been adopted is based on definitions established in the literature, developed through team discussion and debate, and refined in line with peer review comments. It fully acknowledges the blurred boundaries associated with any definition of private school. It is also explained in the review that the terms 'for profit' and 'low-fee' or 'low-cost' cannot always be used with certainty since studies do not always use these terms or provide information on the extent of fees charged or profits made. Following advice from DFID, the review only included studies that are identified as explicitly focusing on private schools. Other types of non-state providers, specifically, religious and philanthropic schools are the focus of a second rigorous review, *The Role and Impact of*

⁷ Tooley and Longfield (2014) justify this with reference to Mcloughlin (2013). It is important to note that Mcloughlin (2013) was a separate piece of work with a different purpose, a Topic Guide that was only partly based on work from Day Ashley *et al.* (2014).

Philanthropic and Religious Schools in Developing Countries: A Rigorous Review of the Evidence, Wales *et al.* (Forthcoming).

5. Evidence included

- 5.1 Tooley and Longfield question the choice of selection criteria, including the cut-off year for the inclusion of literature. The review clearly sets out the search strategy and criteria for the inclusion of studies, following the approach established by DFID's (2013) 'How to note'. It explains that the 'first sift' resulted in a large literature base beyond the scope and size of project and how the set of inclusion criteria were further narrowed in the 'second sift' resulting in the final set of 59 studies included in the review. The inclusion criteria were carefully employed since they provided objective and transparent measures to reduce the studies included in the review to a manageable number.
- 5.2 The narrowing of the publication date criterion to the past five years (from 2008) was purposefully employed as a practical, transparent and objective criterion precisely to prevent the possibility of subjective selection. This was not an arbitrary decision, but rather was employed in response to peer reviewer comments from the Initial Report suggesting that we focus on more recent evidence. Some relevant research predating 2008 is referred to for context and background in the Introduction and for cross referencing in footnotes, but they were not included in the studies reviewed given the set criteria. 'Published' is defined broadly in the review to include peer reviewed journal articles, books and book chapters and research and working papers published online, so long as work was completed and fulfilled all other inclusion criteria. We acknowledge that, as is inevitable in any literature review, not every research finding relevant to all 17 assumptions was reported or recorded through the synthesis process, and this was not the intention of the rigorous literature review. At no point does the review claim to be an exhaustive account of all the literature. Crucially for policymakers, from our broader scoping of the literature in the first sift, we do not consider any omissions to be sufficiently significant so as to substantially alter the overall findings of the review.
- 5.3 We are confident that the sample of 59 papers reported – more than would be typically included in a systematic review - is of a sufficient size and mix (of journal articles, books and working papers) and that the findings can be read by a policy audience as a reliable representation of the overall balance of findings in the body of evidence reviewed.
- 5.4 Policymakers can be confident in the quality and rigour of the materials included in the review. Aside from the fact that their quality was assessed against DFID's criteria, of the 59 studies, 27 were journal articles, 11 were published in academic books, and 21 were published online by reputable institutions⁸. DFID and the review team were keen to ensure the review captured up-to-date materials produced by active research institutes both in the West and also the Global South (where opportunities to publish in peer-reviewed, western-oriented journals can be limited).

⁸ Online publications came from the World Bank, International Growth Centre, Oxford Young Lives Centre, CREATE consortium, Bonn Institute for Labour, National Bureau for Economic Research, ASER, RECoup consortium and the Lahore Institute of Economics.

6. Assumption Classifications

We outline briefly here our responses to re-classification of assumptions.

- 6.1 Assumption 1 (learning outcomes – 21 studies in original review): On the basis of different interpretations, Tooley and Longfield challenge the classification of three studies and propose changing the overall classification from moderate positive to *strong positive*. Following our re-review we do not accept this challenge and are confident in our original assessment of the body of evidence as *moderate positive*.
- 6.2 Assumption 2 (teaching – 14 studies in original review): Tooley and Longfield do not challenge the original overall assessment of findings of this assumption which is *strong positive*. However, they do challenge the classification of one of the 14 studies, Ohba (2012, cited in Day Ashley *et al.*, 2014). We accept that the following reference to Ohba's (*Ibid.*) study on p. 21 is incorrect 'PTRs [pupil-teacher ratios] in private schools were often higher than in government schools'. Rather, it should be explained that only two out of the 12 private schools in the sample had higher PTRs than the two government schools in the sample. Since these schools were purposively selected for the study and involve an incomparable number of private and government schools, this evidence should be reclassified as *neutral* and not as positive as suggested. Tooley and Longfield point out an inadvertent factual inaccuracy on p. 20 of the review in the reference to Hartwig (2013, cited in Day Ashley *et al.*, 2014) that does not change the classification of the study⁹. Taking these corrections into account, the review's original classification of findings remains unchanged as *strong positive*.
- 6.3 Assumption 3 (geographically reach the poor – 8 studies in original review): Tooley and Longfield propose a reinterpretation of this assumption, a reclassification of three individual studies, and a reclassification of the overall findings from weak neutral to *moderate positive*. It would be methodologically inappropriate to reinterpret this assumption retrospectively. The team does not accept the reclassification and stands by the *review's original* overall assessment of the evidence as *weak neutral*.
- 6.4 Assumption 4 (gender equity – 12 studies in original review): Tooley and Longfield propose a reformulation of the assumption, a reclassification of eight individual studies, the removal of one study due to duplication, and a reclassification of the overall findings from moderate negative to *moderate positive*. Again, it would be methodologically inappropriate to reinterpret or reframe the assumption after the analysis has been completed. It is important to note that the assumption was not interpreted literally to mean precisely 50% boys and 50% girls attending private schools at any one time. In the review the findings drawn from the evidence clearly acknowledge the nuances around this assumption. The team does accept the removal of Härmä and Rose (2012, cited in Day Ashley *et al.* 2014) due to duplication. However, the team does not accept any of the challenges to the classifications of individual studies and is confident in the overall assessment of the evidence as *moderate negative*.

⁹ PTR figures cited from Hartwig (2013) should read 23.5:1 in private and 61:1 in government schools.

- 6.5 Assumption 6 (financial stability – 2 studies in original review): Tooley and Longfield propose a reclassification of one study, the introduction of other evidence from seven further studies included in the review, and the reclassification of the overall findings from weak negative to *moderate positive*. The team acknowledges the error made in misattributing a quote to Tooley *et al.* (2008, cited in Day Ashley *et al.*, 2014) which will result in a reclassification of this study. However, a review of the proposed additional evidence from seven further studies raises a number of concerns which does not justify a change in the strength and direction of the evidence suggested by Tooley and Longfield. First, the evidence identified does not relate to research findings that focus directly on financial stability, but rather relates to contextual data about the private schools included in study samples, specifically the number of years these private schools had been established at the time of study. Second, some studies use purposive sampling of the private schools which can lead to a selection bias in relation to their length of establishment. Third, most of the studies do not provide data on the lifespan or survival rates of schools which is more accurately reflective of the concerns related to their financial stability. If we were to include this additional evidence under the assumption, the final classification would at best change from weak negative to *weak neutral*, as the findings are ambiguous.
- 6.6 Assumption 7 (ability of poor and poorest to pay fees – 13 studies in original review): On the basis of different interpretations of the evidence, Tooley and Longfield propose to change the classification of 10 individual studies, remove two studies, and change the overall classification of this assumption from weak neutral to *moderate positive*. As a rigorous review, we are limited to what the literature tells us – there was no clear consensus on definitions of the poor and poorest in the literature. These were often context specific and used different language and indicators (e.g. lowest economic quintile or lower two quintiles, or lower income households). We were careful, therefore, not to impose our own prescriptions but rather use the text to provide contextual information which also indicates how evidence was assessed as neutral and negative. The team accept the suggestion that Härmä and Rose (2012, cited in Day Ashley *et al.*, 2014) should be removed from this assumption on the grounds of duplication. We also accept that Tooley *et al.* (2011, cited in Day Ashley *et al.*, 2014) should be removed since findings are not relevant to this assumption. In relation to all other reclassifications of studies, the team stand by their original assessments. The removal of the two studies identified does not change the review's original overall assessment of the evidence as *neutral* i.e. ambiguous and therefore *weak*.
- 6.7 Assumption 8 (affordability of private schools compared with government schools – 5 studies in original review): Tooley and Longfield propose a reformulation of the assumption, a reclassification of two studies, the exclusion of one study, the addition of evidence from two further studies included in the review, and the reclassification of overall evidence from weak negative to *moderate positive*. Again, it would be methodologically inappropriate to reframe the assumption after the review has been conducted. The team stands by their original classifications of the two studies challenged and does not accept the exclusion of the study suggested. The team accepts that evidence from Ohba (2012, cited in Day Ashley *et al.*, 2014) and Heyneman and Stern (2013, cited in Day Ashley *et al.*, 2014) could be included. However, we do not agree with the classification of the evidence within these studies as positive and would rather classify both studies as providing neutral (ambiguous)

evidence. The inclusion of this additional evidence would change the review's original overall assessment of the finding from weak negative to *moderate negative*.

Closing comments

The team values comments and debate on the issues in the review, and thank Tooley and Longfield for their detailed response. However, after considerable re-checking we have found that the majority of challenges proposed by Tooley and Longfield that lead to them reaching stronger policy conclusions are not justified by the evidence. The small number of inadvertent factual inaccuracies and pieces of evidence missed or duplicated in specific assumptions can be easily rectified. Importantly, they do not substantially change the conclusions of the review: many of the assumptions remain weakly evidenced.

References

Day Ashley, L., Mcloughlin, C., Aslam, M., Engel, J., Wales, J., Rawal, S., Batley, R., Kingdon, G., Nicolai, S., Rose, P. (2014) *The role and impact of private schools in developing countries: a rigorous review of the evidence*. Education Rigorous Review. London: DFID

DFID (2013) *How to note: Assessing the strength of evidence*. DFID Practice Paper. London: DFID
Mcloughlin, C. (2013) *Low-cost private schools: Evidence, approaches and emerging issues*. EPS-Peaks

Tooley, J. and Longfield, D. *The Role and Impact of Private Schools in Developing Countries: A Response to DFID's 'Rigorous Literature Review'*, October 2014.

Wales, J., Wild, L., Aslam, M., Hine, S., Rawal, S., Batley, R., Day Ashley, L., Mcloughlin, C., Nicolai, S., Rose, P. (Forthcoming) *The role and impact of philanthropic and religious schools in developing countries: a rigorous review of the evidence*. Education Rigorous Review. London: DFID