

Where Are We Now? Effect-studies and the Rise of Diversity in Philosophy for Children

Pieter Mostert

“I like the scientific spirit—the holding off, the being sure but not too sure, the willingness to surrender ideas when the evidence is against them: this is ultimately fine—it always keeps the way beyond open—always gives life, thought, affection, the whole man, a chance to try over again after a mistake—after a wrong guess.”

-Walt Whitman

Let me start with a lengthy quotation from Lipman’s manifesto “Where we are now” (Lipman, 1986), claiming that enough research has been done to assert that Philosophy for Children¹ has a positive effect on ‘higher order thinking’. 35 Years later, I want to add a question mark and ask, “Where are we now?” In my attempt to answer this question, I will focus on three issues that have been raised over the years and that keep returning in conversations about whether, as an international community, we are doing the research we should do: (1) although commonly named ‘Philosophy for Children’, there is a wide variety in the actual ‘programmes’ that students are involved in; (2) in most effect-studies only a rudimentary description is given of the actual teaching and learning activities; (3) there is a wide variety in understandings of philosophy. A critical analysis of these issues leads me to a set of recommendations for effect-studies in the near future.²

Back to 1986

In 1986 Matthew Lipman opened his overview of “where we are now” with a manifesto:

The National Assessment of Educational Progress has urged that ‘higher order thinking skills’ of American students be systematically and significantly strengthened. Likewise, the Rockefeller Commission on the Humanities has emphasized that elementary and secondary education must focus upon the building of conceptual skills and must simultaneously intensify and enrich student awareness of our cultural heritage. The Philosophy for Children program helps meet both these needs by fostering the improvement of reasoning skills, starting with the early elementary grades, and by having children engage in logical dialogue with regard to such traditional philosophical concepts as ‘right’, ‘fair’, ‘good’ and ‘true’.

¹For Lipman, “Philosophy for Children” refers to the programme as developed at his Institute for the Advancement of Philosophy for Children (IAPC) at Montclair State University. In this article I use it as a generic term.

² In this article I focus on effect-studies of the cognitive domain, because the outcomes of these studies are commonly used as the evidence that Philosophy for Children brings clear benefits for the students involved.

The success of Philosophy for Children in meeting these objectives is reported upon here. Its effectiveness as a tool for improving various aspects of cognitive performance, including reasoning ability and ideational creativity, is evidenced on a variety of measures of skills. 'Higher order thinking' is confirmed as being not only a definite feature of early childhood cognition, but also as one that can be taught by this particular method. Improvements in academic performance, moreover, suggest that the program is indeed a vital component of humanities education. Furthermore, the improvements generated in reasoning by way of philosophical inquiry are shown to have their impact 'across the disciplines' at an early age. (p. 32)

The Success of the Philosophy for Children Programme

The international community has embraced Lipman's manifesto. There is a widely received view, that by now enough research has been done in order to claim that Philosophy for Children has a positive effect on the cognitive abilities across the disciplines. Trickey and Topping summarise this view in their systematic review of research: "Unlike many educational methods, P4C has relatively good quality and quantity of evidence for effectiveness" (Trickey and Topping, 2004, p. 377). The acronym P4C—currently replaced by P4wC, emphasising the 'for' and the 'with' equally—is understood to refer to the Philosophy for Children "programme". Effect-studies focus on the "effectiveness of Philosophy for Children programmes"; see for example Yan, Walters, Wang and Wang (2018). This way of speaking is not so much of a problem in a single study: "the programme" refers to the intervention as described in the research report. But what if a good number of reports show evidence of effectiveness, but refer to interventions which display a wide variety in the content and implementation of "the programme"? What can we conclude then?

The introduction of a Philosophy for Children programme in a school is a complex intervention. It consists of a whole 'package' of changes in the usual routines of a particular school: new study materials, new learning activities, new classroom arrangements, new kinds of interaction among students and between students and teachers, new teachers or newly trained teachers, presence of researchers and visitors from other schools, interviews, and questionnaires, pre- and post-tests, newsletters, and social media posts for sharing experiences. Philosophical dialogues don't end at the end of the class. The introduction of Philosophy for Children creates a web of changes, of which some will have a bigger impact than others, depending on the specific context of a particular school. In this multitude of changes, it is not more than reasonable to ask the question: What is it that makes Philosophy for Children 'effective'?

Today's reality is that there is a wide variety of views and practices, and that only a minor part of these practices is covered by Matthew Lipman's description in 1986, quoted above, that we facilitate "logical dialogue(s) with regard to traditional philosophical concepts". Over the years, the emphasis on formal reasoning skills (as shown and practised in Lipman's Harry Stottlemeier's Discovery) has shifted towards a broader approach of communicating, questioning and (higher order) thinking skills. This development in the international community of Philosophy for Children can be characterised as a shift of focus from drawing inferences to constructing arguments. So far, this has not led to a new kind of homogeneity of practice, as I will demonstrate later. For effect-studies the absence of such a

homogeneity constitutes a problem in the case of comparing studies which measure a certain effect but refer to different practices ('interventions'), which have caused this effect.

Measuring Effectiveness

Overall, there were only a few high-quality studies that fulfil the indicators of quality in this study to support the effectiveness of P4C. Therefore, there is still much room for research to shed light on stronger evidence about the effectiveness of the programme. (Ventista, 2017, p. 459)

Two Obstacles

Ventista's "stronger evidence" is within reach when we are able to solve two issues. One has already been mentioned above: the variety of what students are actually doing under the umbrella of "the Philosophy for Children programme". The key-question here is: what constitutes a philosophical enquiry? Two different answers, each representing a widely shared view within the international community, can exemplify this variety. Tricky and Topping (2004) answer the question with "a routine classroom philosophical enquiry in terms of nine steps" (p. 369; for a more elaborate exposition of their view, see Topping, Tricky and Cleghorn, 2019). McCall and Weijers (2017), on the contrary, defend "the primacy of philosophy", as they call it:

The side benefits of practising philosophising, such as developing communications skills, listening skills, patience, tolerance of difference, and respect for others, could all be achieved in other ways. But we propose that the deep understanding of what underlies individual and collective human life can only be achieved by philosophising. It is our experience that internationally the emphasis in pwc [Philosophy with Children] has shifted from the core benefits of philosophising to attention to the side benefits. And we think that the loss of the primacy of philosophy in pwc is a huge loss for the children. (p. 91)

The lesson to be learned here for future effect-studies is, as Oyler (2019) recommends, to provide detailed information about, as he calls it, the "dosage" of activities: in which kinds of activities were the students (and teacher) involved and to which intensity? Such information would clarify what in a particular study is meant by "the programme."

The other obstacle to achieving stronger evidence is the variety in training the teachers have received before they start facilitating philosophical enquiries: "The amount of training and experience of the teacher who applied P4C could influence its effectiveness. Unfortunately, most of the papers did not include this information or, when it was included, they were always considered experts. Thus, this variable was not analysed as a moderator" (García-Moriyón, Rebollo and Colom, 2005, p. 16). So, we do not know how much of the measured effects can be attributed to training and experience, in comparison to the effect caused by the programme itself. We do know that when teachers have improved their own inquiry skills, they feel more competent and confident in facilitating the development of the inquiry skills of their students (see Nichols, Burgh and Fynes-Clinton (2017). Fair, Haas, Gardosik, Johnson, Price and Leipnik, (2015, p. 34) share the following observation:

Additionally, the teachers who included the sessions weekly were excited to see the students so engaged and eager to share their thoughts and listen to opinions of others. The teachers who did not include the sessions in their weekly routine shared that they did not feel comfortable with the open discussions and felt a lack of control in the classroom.

Similar outcomes have been recently found by Sindberg Jensen (2020). In 2004, Tricky and Topping discussed this phenomenon under the heading of implementation integrity: the degree to which an intervention is implemented as designed. After an analysis of ten different effect-studies one of their conclusions is that “implementation integrity may be highly variable” (p. 374). The lesson to be learned here is that effect-studies should provide specific information about the training (and mentoring) the teachers receive, not only about what kind of training and how much of it, but—more importantly—whether and how it has changed their teaching practice.

Process and Outcome

Reznitskaya (2005) takes us one step further. If we want to claim that a certain effect is the result of a specific process, it is not enough to offer a general description of this process. Part of the research should be the systematic examination of the classroom dialogues. Otherwise,

the relationship between the dialogic properties of interactions and the individual gains in reasoning is being assumed, rather than established”. <...> Also, outcome-type studies are typically designed to form an overall opinion regarding the success of a program, rather than to understand the underlying mechanisms of cognition. Such studies provide little information about the particular components of P4C practice and their relative contributions to the acquisition of intended skills. Thus, outcome-type research is ineffective for advancing our understanding of socio-cultural influences on learning and for providing P4C practitioners with specific, empirically-tested instructional strategies. (p. 10)

She advocates the use of the “Argument Schema Theory” [AST]:

Using AST as a theoretical framework, a P4C discussion can be assessed in terms of exhibiting various elements of an argument schema, including, reasons, counter-arguments, and rebuttals. This approach will help to make examinations of P4C discussions more systematic, methodologically sound, and grounded in an articulated theoretical orientation. (Reznitskaya, 2005, p. 8)

The Programme and the Test

Effect-studies rely on the use of one or more tests or assessments. Different studies choose different tests. Again, there is a variety within the international community of Philosophy for Children. Is that a problem? Not really, until Yan, Walters, Wang and Wang (2018, p. 27) observe: “Specifically, those studies conducted in non-Western countries have higher effect sizes than the Western ones.” The authors present several possible accounts of this phenomenon, but leave it open to the informed reader to select the most likely one. The most plausible account, to my understanding, is the relation between the programme and the test. In the three main non-western studies (Lam, 2012; Marashi, 2008; Othman and Hashim, 2006) ‘the programme’ is the IAPC

programme and ‘the test’ is the New Jersey Test of Reasoning Skills (see Shipman 1983), a test which was specifically designed to test the reasoning skills as taught in the IAPC-programme. In all other studies tests have been used which bear no close relation to ‘the programme’. In their meta-analysis, García-Moriyón, Rebollo and Colom (2005) address this close relation: “The New Jersey Test of Reasoning Skills was designed as one P4C proxy measure. Unsurprisingly, greater effect sizes were observed for that test. Such practices [an intimate connection between what is taught and what is tested, PM] must be avoided as far as possible.” (p. 21). On the other hand, in the same study the authors conclude that “a serious obstacle to comparing the accumulated evidence of the past 30 years (...) arises from the difficulties involved in reaching an accepted definition of reasoning skills” (p. 15).

Given this situation, a number of effect-studies have turned to tests which are not specifically tailored to measure philosophical thinking but are widely accepted and respected. The meta-analysis by Yan, Walters, Wang and Wang (2018) shows that CAT / CogAT [Cognitive Abilities Test] is the most commonly used test. One of the reasons for using such a test is that, if a Philosophy for Children programme makes a difference in the test score, and this is usually achieved within a limited period of time—in most studies the intervention is less than a year—one knows for sure that such a programme is powerful: it has the potential to raise the level of performance in cognitive abilities. Strategically, demonstrating such an effect may be necessary for Philosophy for Children in order to be taken seriously, as a valuable addition to the curriculum and as new perspective on teaching. Above that, using such a test will help to identify in which areas the performance has improved, for example: “All of the gain in the CAT scores comes from the verbal subscale. There was very little difference between treatment and control groups in terms of quantitative, non-verbal, and spatial elements of the CAT” (Gorard 2015, p. 4).

But we can’t deny that there is a tension between what’s asked and rewarded in such tests and the aims of Philosophy for Children. For example, in such a test one of the questions may be:

- “All questions have an answer.
Given this, which of the following statements is not possible?
(a) ..., (b) ..., (c)...,
(d) The professor discussed a number of questions which don’t have an answer”.

The only answer for which the test awards points is (d), but philosophically speaking, the word that turns this question into a philosophical one is the word ‘possible’: what does it mean when one says that statement (d) is ‘not possible’? And is ‘having an answer’ not an ambiguous statement? Do questions ‘have’ an answer? The question from the test could serve as a stimulus for a rich philosophical enquiry, but the students’ ability to analyse the different arguments for and against the ‘possibility’ of what the professor is doing is not an object of the test and is therefore not rewarded.

“Not Under any Thinking Skills Intervention”

So far, effect-studies have been understood to be studies which assess the effect of a specific intervention in comparison to a classroom situation without this intervention. This is understandable, as the common context is that of a school in which a Philosophy for Children programme is newly introduced. The effect of this programme is assessed by comparing students who took part in the

programme with students who did not. Briefly stated: the effect = ‘treatment A versus no treatment.’ There is, however, another type of effect-studies. In these studies, two approaches are compared, in order to find out about differences in effects and effectivity. This is the common format of research in, for example, medical treatments: is protocol or medicine A more effective than protocol or medicine B? In the realm of Philosophy for Children there are only a few of these effect-studies. A recent example is Worley and Worley (2019) on teaching critical thinking and metacognitive skills. An earlier example is Othman and Hashim (2006). Interestingly, Yan, Walters, Wang and Wang (2018) have excluded this study from their meta-analysis with the following reason: “The study conducted by Othman and Hashim (2006) was excluded because the experiment’s control group was still participating in another thinking skills intervention. This study compared P4C to another thinking programme (the Reader Response Program). Thus, the control group is not neutral. The control groups in all the included studies were not under any thinking skills intervention” (p. 20). In a strict sense this may have been the case: the students in the control group were not involved in another programme of which one of its aims is to improve thinking skills. But in a wider sense, isn’t education by definition a “thinking skills intervention”? Reading, writing, listening, memorising, repeating, answering—for a student it is hard to practise them without applying thinking skills. Schools organise all kinds of learning activities which involve thinking skills, like improving reading skills, developing learning strategies, and practicing memory training techniques. All in all, it is hard to imagine a control group being ‘neutral.’

The lesson we learn here is that for an effect-study it would not be enough to provide a detailed description of ‘the intervention’. It should also contain a similar description of ‘the programme’ of the control group.

Reflecting on the use of the word ‘intervention’, we can learn another lesson, namely that an effect-study should make its arguments explicit for the duration of the programme / intervention as object of the study. It may be that the effects of a longer delivery of the programme are different from the effects of a shorter version. Studying long-term practices of having philosophical enquiries may also reveal specific conditions which are favourable or even necessary for a solid implementation and establishment of a Philosophy for Children programme. See for example Lord, Dirie, Kettlewell and Styles (2021), who report:

The trial started in October 2016 with programme delivery from September 2017 to July 2019. <...> Of the 75 intervention schools, after two years from commencement, a substantial minority (35 of 75 schools) were not implementing P4C at the expected level. Of these, six did not implement P4C at all due to other priorities and/or senior leader turnover. The evaluation suggests that it takes time for teachers to become confident with, use and embed the P4C approach and this could have impacted the outcomes. (p. 6)

Understandings of Philosophy

“An Accepted Definition of Reasoning Skills”

In 2005 García-Moriyón, Rebollo and Colom envisaged that the international community would be “reaching an accepted definition of reasoning skills” (p. 15). We have not reached that point

yet, and it may remain a wish for quite a few more years. Why? Basically, because there is (and will remain) criticism from two opposing sides: any list of reasoning skills will be both too narrow and too wide at the same time.

It will be too narrow because such a list will focus on individual reasoning activities. Hence the popularity of “21st century skills”, with much emphasis on communication and collaboration. But, as these critics say, as long as we restrict ourselves to ‘reasoning’ and ‘skills’, our scope will be too narrow. Isn’t philosophy about wondering, exploring, reflecting, and finding out? David Velleman reflects on his life as an academic philosopher and addresses his colleagues in the introduction to his new book, “On being me”:

Academic philosophers who read this book will say, ‘But there are no arguments!’. Let me be the first to say it: There are no arguments. There are only observations [...]. For many years I thought the arguments were meant to convince the readers, but I found myself oddly unconcerned when few if any readers were convinced. I finally realized that I have all along been reporting on personal explorations. (Velleman, 2020, p. xii–xiii)

On the other side the criticism stands that a list of reasoning skills (and a fortiori critical thinking skills and 21st century skills) will be too wide because the listed skills are only for a part specifically philosophical. Many are general, in the sense that they are practised in all kinds of other learning activities, from historical reasoning to coding for computers, from mathematical reasoning to chess. Participating in philosophical enquiries may contribute to the development of these ‘general’ skills, but many other learning activities come with similar claims.

Understandings of Philosophy and Its Impact

In a thorough literature review of articles about the teaching and learning of philosophy in primary and secondary schools, Bowyer, Amos and Stevens describe how they were able to identify ten different understandings of what philosophy is: (1) a foundational concept, (2) thinking—a skill, a disposition, a practice, (3) a method or process, (4) a tool or instrument, (5) a creative task, (6) inquiry, (7) search for truth, (8) non-dogmatic teaching and hence the emancipation of thought, (9) communal activity, (10) a way of life (Bowyer, Amos and Stevens, 2020, p. 41). The authors demonstrate how these views are embedded in specific educational practices. In a second literature review the authors apply the same method to all the answers of the question “What does philosophy do?”. They focus on the claims in the literature about the impact of philosophy in the classroom and show how these claims can be categorised in two distinct types:

The first claim is that philosophy improves academic and cognitive abilities, where the idea of ‘cognition’ is captured by forms of reasoning that can be tested and measured. The evidence for improvement in academic and cognitive abilities takes the form of IQ scores, Cognitive Abilities Test (CAT) and school academic assessments, including norm-referenced tests of reading, reasoning, and other curriculum-related assessments. The second claim is that engaging with the world philosophically promotes the art of living well together. It is argued that philosophical engagement is a collaborative endeavour, aimed at cultivating understanding through respectful interactions that are open to exploring, questioning and challenging aspects of the world. (Bowyer, Amos and Stevens, 2021, p. 71)

The authors mention a number of tests which can be used to substantiate claims of the first type. Some of them have been discussed here. Regarding the claims of the second type, the authors' view is that "the evidence for this work of philosophy is based on testimony from teachers, principals, parents and students themselves." (Bowyer, Amos and Stevens, 2021, p. 74)

Few attempts have been made to create a theoretical model for 'philosophical engagement' *per se*. One of them is the PhD thesis by Rondhuis (2005), which is devoted to the design and empirical validation of a general model of what she calls "philosophical talent". This study is one of the very few studies which combine a thorough analysis of the philosophical activity, the development of a conceptual model and an extensive effect-study of philosophical discourses among children and adolescents. In her research project *Philosophical Talent*, the key-question "Which features signify philosophical quality?" is answered as follows:

six indicators for philosophical quality were distinguished, correlating with six thinking patterns or attitudes representing specific philosophical qualities or aspects of them. Each indicator covers a group of linguistic expressions. These indicators are:
1. Indecisive thinking, 2. Openness, 3. Tentative behaviour, 4. Epistemic position, 5. Reasoning quality, and 6. Anecdotal quality. (Rondhuis, 2005, p. 58)

For a concise presentation of her research, see Rondhuis (2006), also Rondhuis and Van der Leeuw (2000).

Effect-studies in the Near Future

In his presentation on "Philosophy and the cultivation of reasoning", Lipman clearly stated:

There are many ways in which philosophy can be taught to school children from kindergarten through high school. It does not have to be done the way we at the Institute for the Advancement of Philosophy for Children have done it, but we can only report our own experience." (Lipman, 1985, p. 37)

The international community of Philosophy for Children demonstrates how philosophical dialogues can be done in different ways, with a wide variety of aims and topics. There is no reason to assume that this diversity will lessen in the future. We will have to live with it, but it presents a number of specific challenges to future effect-studies, as we have seen. The following recommendations are offered for further consideration. They should be read in addition to, not in competition with other recent proposals for future research, like Reed-Sandoval and Sykes (2017, p. 223- 224).

(1) Effect-studies should contain a detailed description of both the intervention and of the programme of the control group. This should include detailed information about the 'dosage' of activities: in which kinds of activities were the students (and teacher) involved and to which intensity. Under the umbrella of ICPIIC, a special interest group on research should be created in order to draft a format for such a description. Researchers complete it for their specific studies and add it to their reports.

(2) Effect-studies should be specific about the training (and mentoring) the teachers receive, not only what kind of training and how much of it, but—more importantly—whether and how it has changed their teaching practice.

(3) An international initiative should be launched for the development of a common framework for the description of philosophical activities within Philosophy for Children. Confidence that this can be done may be derived from other, much larger but nevertheless successful, international efforts to develop a common framework, like the CEFR standard, the Common European Framework of Reference for Languages, and for levels of education the NQF, the National Qualifications Framework.

(4) More effect-studies should be done outside the context of schools. Part of the diversity within the Philosophy for Children is that it is no longer exclusively done within the setting of a school but is seen as a community-based activity; see Lockrobin (2019).

(5) More effect-studies should focus on non-cognitive effects, as they are particularly relevant within the understandings of philosophy as a creative task, as communal activity, and as a way of life (as described by Bowyer, Amos and Stevens, 2020).

(6) There should be a better balance between effect-studies focusing on effects in terms of general cognitive skills and studies which focus on progress in philosophical thinking specifically.

(7) It's time to move towards studies which assess the effects of long-term practices of having philosophical enquiries. The arguments for the relevance of a Philosophy for Children programme have always positioned such a programme as a permanent part of the curriculum. Effect-studies which show the effects of short-term interventions do not provide much support to these arguments. It is time to correct this imbalance.

Acknowledgements

I want to thank Pablo Lamberti, Griet Galle and the journal's reviewer for their valuable suggestions. In my role as a 'veteran' I cannot write about Philosophy for Children without paying tribute to my late colleague Karel van der Leeuw. With him I attended the IAPC summer school in 1983. There the first outline of our joint PhD thesis took shape, devoted to the teaching of 'doing philosophy'. In 1988 we presented it in its final form, the product of years of intense dialogues. A year later, we founded the Dutch centre for Philosophy for Children. In 2015 Karel died; I write this article in commemoration of his contribution to Philosophy for Children.

References

Bowyer, L., Amos, C. and Stevens, D. (2020). What is 'philosophy'? Understandings of philosophy circulating in the literature on the teaching and learning of philosophy in schools. *Journal of Philosophy in Schools*, 7(1), 38–67.

- Bowyer, L., Amos, C. and Stevens, D. (2021). What does philosophy do? Understanding the work that philosophy does: A review of the literature on the teaching and learning of philosophy in schools. *Journal of Philosophy in Schools*, 8(1), 71–103.
- Colom, R., García Moriyón, F., Magro, C., and Morilla, E. (2014). The long-term impact of Philosophy for Children: A longitudinal study (Preliminary results). *Analytic Teaching and Philosophical Praxis*, 35(1), 50–56.
- Fair, F., Haas, L., Gardosik, C., Johnson, D., Price, D. and Leipnik, O. (2015). Socrates in the schools from Scotland to Texas: Replicating a study on the effects of a Philosophy for Children program. *Journal of Philosophy in Schools*, 2(1), 18–37.
- García-Moriyón, F., Rebollo, I and Colom, R. (2005). Evaluating Philosophy for Children: A meta-analysis. *Thinking*, 17(4), 14–22.
- Gorard, S., Siddiqui, N. and Huat See, B. (2015). *Philosophy for Children: Evaluation report and executive summary*. Millbank, England: Educational Endowment Foundation, 45 p.
- Lam, C.-M. (2012). Continuing Lipman’s and Sharp’s pioneering Work on Philosophy for Children: Using Harry to foster critical thinking in Hong Kong students. *Educational Research and Education*, 18(2), 187–203.
- Lipman, M. (1985). Philosophy and the cultivation of reasoning. *Thinking*, 5(4), 33–41.
- Lipman, M. (1986). Where we are now. *Thinking*, 6(4), 39–50.
- Lockrobin, G. (2019) Relocation and repopulation: Why community philosophy matters. In: Fulford, A., Lockrobin, G and Smith, R. (Eds), *Philosophy and community: Theories, practices and possibilities*. London, England: Bloomsbury, 15–37.
- Lord, P., Dirie, A, Kettlewell, K. and Styles, B. (2021). *Evaluation of philosophy for children: an effectiveness trial. Evaluation report*. Millbank, England: Education Endowment Foundation, 125p.
- McCall, C.C., and Weijers, E. (2017). Back to basics: A Philosophical analysis of philosophy in Philosophy with Children. In M. Rollins Gregory, J. Haynes, and K. Murriss (Eds.), *The Routledge International Handbook of Philosophy for Children* (83–92). London: Routledge.
- Nichols, B., Burgh, G. and Fynes-Clinton, L. (2017). Reconstruction of thinking across the curriculum through the community of inquiry. In M. Rollins Gregory, J. Haynes, and K. Murriss (Eds.), *The Routledge international handbook of Philosophy for Children* (245–252). London, England: Routledge.

- Othman, O. and Hashim, R. (2006). Critical thinking and reading skills: A comparative study of the Reader Response & the Philosophy for Children approaches. *Thinking*, 18(2), 26–34.
- Oyler, J. (2019). *Methodological integrity: The role of philosophy in educating for complex aims*. Paper presented at the Dialoguing Democracy Conference, Galway, Ireland
- Reed-Sandoval, A. and Sykes, A. (2017). Who talks? Who listens? Taking ‘positionality’ seriously in Philosophy for Children. In M. Rollins Gregory, J. Haynes, and K. Murriss (Eds.), *The Routledge International Handbook of Philosophy for Children* (219–226). London, England: Routledge.
- Reznitskaya, A. (2005). Empirical research in Philosophy for Children: Limitations and new directions. *Thinking*, 17(4), 4–13.
- Rondhuis, T. (2005). *Philosophical talent: Empirical investigations into philosophical features of adolescents’ discourse*. Rotterdam, The Netherlands: Veenman, 213 p.
- Rondhuis, T. (2006). Philosophical quality of children’s thinking patterns. *Thinking*, 18(3), 18–24.
- Rondhuis, T. and Van der Leeuw, K. (2000). Performance and progress in philosophy: An attempt at operationalisation of criteria. *Teaching Philosophy* 23(1), p. 23–42.
- Shipman, V. (1983). *The New Jersey Test of Reasoning Skills*. Totowa NJ, USA: Totowa Board of Education.
- Sindberg Jensen, S. (2020). The art of facilitating philosophical dialogues from the perspective of teachers. *Educational Studies*, DOI: [10.1080/03055698.2020.1835612](https://doi.org/10.1080/03055698.2020.1835612)
- Trickey, S. and Topping, K. (2004). ‘Philosophy for Children’: A systematic review. *Research Papers in Education*, 19(3), 365–380.
- Topping, K., Trickey, S. and Cleghorn, P. (2019). *A Teacher’s guide to philosophy for children*. London, England: Routledge, 179 p.
- Velleman, J. (2020). *On being me: A personal invitation to philosophy*. Princeton, USA: Princeton University Press, 91 p.
- Ventista, O. (2018). A literature review of empirical evidence on the effectiveness of Philosophy for Children. In: Duthie, E., García Moriyón, F. and Robles Loro, R. (Eds), *Family resemblances: Current trends in Philosophy for Children* (pp. 450–471). Madrid, Spain: Anaya.
- Whitman, W. (2019). *Walt Whitman speaks: His final thoughts on life, writing, spirituality and the promise of America*. Ed. B Wineapple. New York, Library of America. Prepublication in: *The New York Review*

of Books, 66(7), April 18. Retrieved from: <https://www.nybooks.com/articles/2019/04/18/walt-whitman-alone/>

Worley, E. and Worley P. (2019). Teaching critical thinking and metacognitive skills through philosophical enquiry: A practitioner's report on experiments in the classroom. *Childhood and Philosophy*, 15(32), p. 1-34.

Yan, S., Walters, L, Wang, Z and Wang, C. (2018). Meta-analysis of the effectiveness of Philosophy for Children programs on students' cognitive outcomes. *Analytic Teaching and Philosophical Praxis*, 39(1), 13-33.

Address Correspondences to:

Pieter Mostert,

Dorchester, England.

The Philosophy Foundation, London.

Email: pmostert@xs4all.nl