International Journal of Contemporary Educational Research
Volume 12, Number 1, March 2025, Page 26-38 Article Type: Research Article

# Philosophy for/with Children: A Meta-Analysis

**Ben Kilby**[1] | | bkilby@vaps.vic.edu.au
Monash University, Faculty of Education, Melbourne, Australia

**Abstract**
This meta-analysis presents empirical research using a Philosophy for/with Children (P4wC) approach. P4wC involves a teacher engaging in philosophical problem posing and dialogue with students. There has been research focussed on the benefits derived from this practice for students. This meta-analysis calculates the effect size of these benefits that can be calculated using quantitative measures. This meta-analysis contains 62 data sets from 30 studies. P4wC is found to have a significant overall effect size of 0.65. Moreover, subgroup analyses show that when the research focusses on the development of critical thinking, the effect size is substantial at 0.89. This indicates that P4wC has a large impact on the capacity for children and young people to develop critical thinking. Other subgroup analysis also shows that this holds across age ranges. However, it does differ by region, with studies conducted in Eastern countries yielding a higher effect size than Western countries.

**Keywords:** Philosophy for/with Children; P4wC; Empirical Research; Critical Thinking; Meta-analysis

**Citation**
Kilby, B. (2025). Philosophy for/with Children: A Meta-Analysis. *International Journal of Contemporary Educational Research, 12*(1), 26-38. https://doi.org/10.52380/ijcer.2025.12.1.703

| | |
|---|---|
| Received | 21.08.2024 |
| Accepted | 05.03.2025 |
| Publication | 28.03.2025 |
| Peer-Review | Double anonymized - Double Blind |
| Plagiarism Checks | Yes - iThenticate |
| Conflicts of Interest | The author(s) has no conflict of interest to declare. |
| Complaints | editor@ijcer.net |
| Grant Support | The author(s) acknowledge that they received no external funding in support of this research. |


[1] Corresponding Author

## Introduction

This paper presents a meta-analysis of empirical research using a Philosophy for/with Children (P4wC) [2] approach. P4wC is an approach focussed on philosophical inquiry through dialogue, developing critical, creative, caring, and collaborative thinking in students from childhood through senior secondary and even tertiary level study (Lipman & Bynum, 1976). P4wC uses a dialogic approach to engage students in critical collaborative exploratory discussion (Kilby, 2021). This approach aims to develop a wide variety of philosophical outcomes, including in the areas of values, understanding, and intellectual humility – but for which quantitative empirical research has not been conducted. While there are variations in how P4wC is practices such as the role of the teacher or how discussions should be structured (e.g., Kilby, 2022a), there are generally accepted practices within P4wC (Murris, Haynes, & Gregory, 2016). There are a number of skills and dispositions present in all P4wC practice that are of educational value. Some of these are capable of being measured by quantitative empirical research and have been researched through quantitative empirical studies. This meta-analysis analyses data from quantitatively measured empirical studies in P4wC where the computation of an effect size is possible and creates an overarching effect size for interventions that utilise P4wC. This meta-analysis identifies 62 data sets from 30 studies (Table 7). The rationale for this study is to provide an updated meta-analysis of the effectiveness of P4wC. Two previous reliable meta-analyses were conducted in 2004 and 2005 (Trickey & Topping, 2004; Moriyón, Robello, & Colom, 2005). Therefore, given the time between previous meta-analyses and the additional empirical research that has taken place since that time, this updated meta-analysis of P4wC research is beneficial to conduct at this time.

## Philosophy for/with Children

What is Philosophy for/with Children? P4wC was initially a primary school curriculum and pedagogy but is now practised at all levels of education. It was created by American Philosopher Matthew Lipman in the 1970s, who criticised the existing education system for fostering children's lack of critical and creative thinking capabilities. Cam (2006, p. 2) states that part of the impetus for developing these capabilities is that "no developed society would tolerate unchecked endemic disease in the way that we suffer the consequences of widespread poor thinking in our society." Lipman began providing children a space to think philosophically and study complex ideas at a younger age through immersion in philosophical thinking and dialogue. He developed a set of purpose-written children's books with accompanying teacher manuals. The books were used as stimuli to encourage students to explore philosophical problems. The book's characters also modelled the kind of communal dialogue and reasoning skills that P4wC aimed for students to develop (Lipman, 1983). Through this, a wide array of thinking skills would be developed that would aid children in developing more rounded cognitive skills and a broader sense of questioning that would benefit them across all disciplines, as well as throughout their lifelong learning journey (Lipman, 1985). As Cam (2006, p. 1) explains,

> *There can be no doubt that the ability to think about the issues and problems that we face in our lives, to explore life's possibilities, to appreciate alternative points of view, to critically evaluate what we read and hear, to make appropriate distinctions and needful connections, and generally to make reasonable judgements are among the attributes of anyone who has learnt to think effectively in life.*

Lipman's philosophical view was not founded in *a priori* reason or rationalism, but in pragmatism. This pragmatism was developed out of learnings of others, including Charles Sanders Peirce and later by John Dewey; psychologist William James; sociologist George Herbert Mead; among others. Lipman stated that the goal of P4wC was to allow children to develop *reasonableness*, which he described as rationality tempered by judgment, involving the ability to provide reasons for one's opinions and claims and be moved by the reasons of others (Lipman, 2003, p. 10). His program had a reliance on critical thinking, which enabled children to make distinctions between stronger and weaker forms of reasoning. His program provided analytical and logical thinking and reasoning skills. These skills were seen to be best developed in what was called a Community of Inquiry (CoI), an idea derived from Peirce. Teachers would facilitate the acquisition of these skills by using philosophy as a learning

[2] Currently, there are a variety of terms used to talk about doing philosophy with children, these including Philosophy for Children (P4C) (Montclair State University, 2017), philosophy with children (SOPHIA, 2017), philosophical inquiry with children (ICPIC, 2017), philosophy in schools (FAPSA, 2017), philosophy alongside children (Murris, 2016), and the use of any particular term is still debated today (Kohan et al., 2017). This research does not take a terminological position on any of these, but will use a popular term, Philosophy for/with Children and "P4wC," to represent a range of approaches that have grown out of Lipman and Sharp's initial Philosophy for Children (P4C) approach.

method, and students would participate in communal critical analyses of philosophically problematic concepts. Within philosophy, students might explore such problematic concepts like the good, freedom, justice, and fairness in sub-disciplines of ethics and political philosophy; truth, knowledge, and authority in the sub-discipline of epistemology; and reality, creation, and the mind in the sub-discipline of metaphysics; beauty, art, and the emotions in the sub-discipline of aesthetics. Moving through the 1990s and early 2000s, P4wC moved beyond the original Lipman analytical/logical model by including intersubjective experiences, pluralistic reasonableness, a connection to meaning making in children's lives, and the involvement of embodied experiences and emotions that affected children's reasoning (Vansieleghem, 2013, p. 1303). This moved away from the more analytical method that Lipman initially employed, which relied heavily on logic. Lipman himself embraced this change by adding "caring thinking" as one of three key types of thinking – the others being critical and creative thinking (Lipman, 2003). The influence of Lipman's long-term collaborator, Ann Sharp, played a key role in this, as she connected P4wC to ideas from feminist and continental philosophy (Gregory & Laverty, 2017).

Importantly, P4wC is not a matter of taking university level philosophy and implanting it in schools (Pritchard, 2014, s. 3; Lone, 2021; Kizel, 2024, p. 45). There is not a focus on philosophical knowledge in the form of learning about particular philosophers, theories, or ideas from the history of philosophy (Golding, 2014, p. 71). Instead, the focus has always been on *philosophical understanding*: insofar as that understanding supports children to make meaning of, respond to, and criticise philosophically problematic concepts in their own way (p. 70). The teacher provides dialogic skills, strategies, and procedures that may be used to enhance philosophical understanding (Kilby 2019a; Kilby 2022b; Kizel 2022). The aim of P4wC is to produce a philosophiser, not a historian of philosophy or a person who can recite the ideas of other philosophers, but a person who can think well, solve problems, and lead a meaningful, ethical life. This is an important caveat in relation to CCD. P4wC is not the teaching of philosophy as a discipline but a pedagogical approach. Therefore, the features of its pedagogical approach are not necessarily bound to a particular discipline area and can support diverse and inclusive thinking (Kilby 2023). The P4wC community is separate from the academic discipline of philosophy and focuses more on the ideas and thinking of students themselves, rather than the study of great philosophers and philosophies like you might find in an academic philosophy course. For this reason, you will find P4wC being applied across all disciplines, including maths (Kennedy, 2007; Kennedy & Marsal, 2023), science (Sprod, 1997; Dobashi, 2009), and art (Prior & Wilks, 2018; Wartenberg, 2019) to name a few.

## Search & Selection

A systematic search of the databases of ERIC, Web of Science, and Google Scholar was undertaken to identify studies that fit the profile for Philosophy for/with Children. The search terms used to identify studies were only those that are commonly used in P4wC. These are: "Philosophy for/with Children", "Philosophy for Children", "Philosophy with Children", "Philosophy in Schools", "Philosophical Inquiry". To ensure that as many empirical datasets on P4wC research were included, known Philosophy for/with Children spaces were also accessed to ensure that data was not missed. These included the *Institute for the Advancement of Philosophy for Children* at Montclair State University, which hosts a page dedicated to research in Philosophy for children (Montclair State University, 2017). It also included P4wC related academic journals were searched manually including the *Journal of Philosophy in Schools*, *Critical and Creative Thinking, Thinking: The Journal of Philosophy for Children*, *Childhood and Philosophy*, and *Analytic Teaching and Philosophical Praxis*. Additionally, further searching was conducted when new studies were found by looking through reference lists of papers to find other studies that were cited, and also using the Google Scholar citation features which provide information about other articles that have subsequently cited a specific article. 97 studies were initially identified. 30 of these studies met the inclusion criteria.

The inclusion criteria for this meta-analysis were studies that:

- Involved a Philosophy for/with Children intervention;
- Involved quantitative data;
- Involved a control group;
- Had the data required to perform a meta-analysis available or calculatable.

This meta-analysis requires the presentation of data in terms of effect size, calculated using Hedges G (Hedges & Olkin, 1985). To meet this requirement, data that needed to be presented or calculable from the original research papers included:

- Mean average of the gain in the intervention group

- Mean average of the gain in the control group
- Standard deviation of the gain in the intervention group
- Standard deviation of the gain in the control group
- Number of participants in the intervention group
- Number of participants in the control group

Some studies had already calculated effect sizes using this method themselves, therefore this effect size was used. Other studies had not calculated effect size at all. While other students had not calculated effect size using the same method (e.g., some studies used Cohen's D instead of Hedges G), therefore the effect size was recalculated using Hedges G for consistency. Research that lacked the above data sets were excluded on the basis that effect size was not able to be calculated.

At the conclusion of the selection process, 62 data sets were included for analysis from 30 different studies. Inclusive of the 62 data sets were some studies that were separated into multiple parts. For example, the study of Worley and Worley (2019) reported measures of improvement in reading, and also improvement the use and success critical thinking and metacognitive skills. Data is presented separately in this meta-analysis for reading and critical thinking, allowing a subgroup analysis to separate these different measures for greater insight into the benefits for reading as compared to the benefits in terms of critical thinking during a P4wC intervention. Similarly, Gür, Koçak, and Muharrem (2017) provided separate data for research on 5-year-olds compared to 6-year-olds. These data sets, and similar others, were treated separated.

## Method

This meta-analysis deliberately chose to include a wide variety of studies measuring different outcomes that resulted from a P4wC intervention. The purpose of which was to assess the overall educational effect of P4wC on students. However, subsequent subgroup analyses within this paper reveal more specificity about individual effects. Included in this meta-analysis are measures assessing students' reading, critical thinking, general academic achievement (including maths and writing), and socio-emotional skills. While these are disparate measures of achievement, they are all valuable measurements of educational achievement for schools, thus, the relevance of how P4wC impacts on all of these measures provides valuable insights. Further subgroup analyses within these categories detailed how P4wC impacts in those particular areas (e.g., how does P4wC impact reading achievement specifically). Some of the studies included in this paper were also included in previous meta-analyses such as Trickey and Topping (2004) and Moriyón, Robello, and Colom (2005). They were included again here if they fit the selection criteria to provide a cumulative update on the effect size of P4wC. Additionally, 22 of the 30 studies included in the current paper were published later than 2006, making them newer than the two previously cited meta-analyses.

Hedges G (Hedges & Olkin, 1985) was used to measure all effect sizes. This was chosen as there was variability in the number of participants in the control compared to the experiment group in most studies. This meta-analysis computed effect sizes using the standardised mean difference for outcomes reported in the studies. For this reason, only studies with a pre-post control-intervention comparison were included. In many studies, the calculation was made using the data published. However, for some studies the standard deviations were missing from the publication. In these studies, efforts were made to contact the researchers to clarify the standard deviations. If this was unsuccessful, the standard deviations were calculated from the data provided and then used in the calculation of Hedges G. There is not a standardised way of reporting data in education (e.g., Cohen's D or Hedges G for calculating effect size). Therefore, as Bernard, Borokhovski, Schmid, and Abrami (2014) explain, it is sometimes required to calculate effect sizes "from test statistics (e.g., t-ratios), exact probabilities (e.g., $p = .023$) or even inexact hypothesis-test outcomes (e.g., $p < .05$) (cf. Glass, McGaw, & Smith, 1981; Hedges, Shymansky, & Woodworth, 1989)." An example of this in this paper was Fair et al. (2015b). This study provided the means and standard deviations of individual tests, as well as the number of participants, and t and p scores derived from parametric dependent sample t-tests. While standard deviations were provided for each test, standard deviations for the gains from pre-test to post-test were not provided. However, these were calculated by analysing the t-tests in the following way:

$$t = \frac{Me}{\mathrm{Sx}}$$

where

$$Sx = \frac{Se}{\sqrt{n}}$$

*t* = the published t score

*Me* = mean difference in gain scores

*Se* = standard deviation of the gain scores

*n* = number of participants

*t*, mean gain, and the number of participants were all provided, therefore we could solve this equation for *Se* to calculate the gain in standard deviation and then use it in the formula to calculate effect size using Hedges G.

Some effect sizes were slightly different because of the different tools used to calculate. For example, Topping and Trickey (2007b) used their data to calculate an effect size of 0.75. They used Cohen's D to calculate this. In this current meta-analysis, the study of Topping and Trickey (2007b) has an effect size of 0.83. The same data was used to calculate this, however this meta-analysis used Hedges G for all calculations. Hedges G was determined to be a more appropriate measure for two reasons. Firstly, Hedges G provides an effect size by weighting the participant groups. When the intervention and control groups have the same number of participants, Hedges G will be identical to Cohen's D. However, when there is a difference in participant numbers, Hedges G will provide a better analysis be weighting the effect size for participant numbers. This resulted in the difference in calculated effect size in the Topping and Trickey (2007b) paper compared to the effect size reported in this paper.

Where the standard deviation of the gain score was not reported, or the raw data was not available, the standard deviation of the gain score was derived from calculating the covariance with an estimated correlation of 0.5. The standard deviation of the gain score was then calculated using the following formula:

$Var(x) + Var(y) + 2Cov = Se(z)$

$Cov = Se(x) \times Se(y) \times Cor$

where

*Se(x)* = Standard deviation of pre-test

*Se(y)* = Standard deviation of post-test

*Cor* = Correlation (estimated at 0.5)

*Var(x)* = $\mathrm{Se(x)}^2$

*Var(y)* = $\mathrm{Se(x)}^2$

*Se(z)* = Standard deviation of the gain

Care was also taken to avoid double counting. For example, Youssef (2014) reported results from a pre-test and post-test, then also a follow up test. The results from the follow up test only were also reported in Youssef, Campbell, and Tangen (2016). Therefore, the data from Youssef (2014) only includes the unique pre-test and post-test results, as the follow up test results had already been included from Youssef et al. (2016).

In conducting subgroup analyses, there was some variability in how different studies were presented. Therefore, some categorisation was performed to fit each study into a specific category for analysis. For example, the subgroup related to 'time of intervention' was measured in semesters for this meta-analysis. However, some studies reported time in terms of weeks (Fair et al., 2015a), hours (Gür et al., 2017), or dates (Lam, 2012). These reported timeframes were roughly translated to semesters based on best guess. For example, Fair et al. (2015a) reported 10 weeks of intervention, this was translated to 1 Semester. Gür et al. (2017) reported 50 hours of intervention, this was translated to 1 Year. Lam (2012) reported October to March, and this was translated to 1 semester. Moreover, similar translations were made in terms of age of students, as some studies reported year level and others reported

age. Measures were also categorised as either Critical Thinking, General, SEL, or Reading. However, there were also different instruments used to measure within each of these categories. Studies that did not include any information were excluded from categorisation in that subgroup. For example, Marashi (2008) did not report the length of the intervention, so this study was excluded from the subgroup related to 'time of intervention.' Where the data allowed, studies were split between subgroups. For example, Worley and Worley (2019) conducted their intervention and used pre- and post-testing for both critical thinking and metacognitive skills, as well as reading comprehension. The data from this study was reported separately (i.e., critical thinking and reading comprehension data were not combined). Therefore, the data from this study pertaining to reading comprehension was used in the 'Measure' subgroup category of 'Reading', while the critical thinking and metacognitive skills data from this study was used in the subgroup category of 'Critical Thinking'. Some studies required subjective decisions to be made about the subgroup analysis. For example, Fair et al. (2015b) measured students when they were 16. However, this was a follow-up assessment. The intervention itself occurred 3 years prior and no further intervention had taken place. It was judged that this study fit into the 8–14-year-old age category because the intervention took place when students were 8-14, despite the assessment occurring at an age that would fit in the <14 category. Similarly, the participants in Walker, Wartenberg, and Winner (2013) were aged 7-8, splitting the subgroup categories of >8 and 8-14. It was decided to place this study in the 8-14 subgroup.

For some studies, only a part of the study were used. For example, Reznitskaya et al. (2012) were attempting to measure transfer of argumentation skills to different contexts. This goal is not relevant and was not considered. However, the study did include a measure of 'elaborated reasoning' among a number of other variables. Only the measure of elaborated reasoning was used from this study in the present meta-analysis.

## Findings

This meta-analysis across 62 data sets from 30 studies using a random effects model to calculate Hedges G revealed an overall effect size for P4wC interventions of 0.65, statistically significant to the value of p = <0.0001. The Institute of Education Sciences (2017) suggests that effect sizes greater than 0.25 are of substantive interest, while Hattie (2020) suggests that the average effect size in education is 0.40. With an effect size of 0.65, this puts P4C in a high category of teaching strategies that improve educational achievement for students from early childhood to senior secondary schooling.

This meta-analysis shows a slightly higher effect size overall than previous meta-analyses conducted on P4wC. However, similar patterns are able to be seen across various meta-analyses. Moriyón et al. (2005) reported an overall effect size of 0.58 and Yan, Walters, Wang, and Wang (2018) reported an identical effect size (even though different studies were used). Trickey and Topping (2004) reported a lower effect size of 0.43.

**Table 1.Effect size of P4wC**

| k | Effect Size | 95% CI | t | p-value |
|---|---|---|---|---|
| 62 | 0.6511 | 0.4656 – 0.8366 | 7.02 | <0.0001 |

Subgroup analyses were conducted across five different subgroups:

1. Age of Participants: >8; 8-14; <15
2. Measure of Study: Critical Thinking; General Academic; Reading; Social-Emotional Learning
3. Time of Intervention: 1 Semester; 1 Year; 2 Years
4. Region: West; East
5. Decade of Study: 1980; 1990; 2000; 2010

Time of Intervention demonstrated higher effect sizes for studies of One Semester over studies of One Year. There was only one study that was in the Two Years category so this category was excluded. This was determined to be most likely a result of the reliability of the studies, with shorter term studies less likely to be as sound as longer-term studies, thereby appearing to generate more significant results.

**Table 2. Effect size of P4wC by 'Time of Intervention'**

| | k | Effect Size | 95% CI | p-value |
|---|---|---|---|---|
| 1 Semester | 36 | 0.8479873 | 0.605 – 1.091 | 0.000000000007474971 |
| 1 Year | 25 | 0.3976110 | 0.147 – 0.648 | 0.001872165 |
| 2 Years | 1 | 0.01 | -0.012 – 0.032 | 0.3809295 |

Similarly, Decade of Study also showed some non-significant variation but without any pattern. The highest effect size average came from the 2010 decade, following by 1990, 1980, then 2000. There was no substantive conclusion drawn regarding why this variation occurred.

**Table 3. Effect size of P4wC by 'Decade of Study'**

| | k | Effect Size | 95% CI | p-value |
|---|---|---|---|---|
| 1980 | 7 | 0.5885718 | -0.011 – 1.188 | 0.05418785 |
| 1990 | 3 | 0.6197084 | 0.376 – 0.863 | 0.000005968723 |
| 2000 | 14 | 0.2830657 | 0.084 – 0.482 | 0.005245003 |
| 2010 | 38 | 0.8020616 | 0.546 – 1.058 | 0.000000000819807 |

The Age of Participants subgroup category yielded almost identical results for the >8 group and the 8-14 group at 0.73 and 0.75 respectively. The <15 group was disregarded as the data from this age group all came from the same research.

**Table 4. Effect size of P4wC by 'Age of Participants'**

| | k | Effect Size | 95% CI | p-value |
|---|---|---|---|---|
| >8 | 12 | 0.726181921 | 0.279 – 1.173 | 0.001450949 |
| 8-14 | 42 | 0.753364629 | 0.533 – 0.974 | 0.00000000002275605 |
| <15 | 8 | 0.0090941040 | -0.071 – 0.089 | 0.8237544 |

Studies undertaken in Eastern countries yielded higher effects sizes than studies undertaken in Western countries, with effect sizes of 0.89 and 0.56 respectively with statistically significant p-values.

**Table 5. Effect size of P4wC by 'Location: East/West'**

| | k | Effect Size | 95% CI | p-value |
|---|---|---|---|---|
| East | 18 | 0.8854543 | 0.519 – 1.252 | 0.000002201270 |
| West | 44 | 0.5564449 | 0.353 – 0.760 | 0.00000008860444 |

Student data on reading and general academic achievement were significantly lower than data from measures of critical thinking. Data on socio-emotional skills was also high, but there were only four datasets that fit this category, all of which came from the same research and all of which were highly variable. For these reasons, the socio-emotional skills category was excluded from analysis here. Focussing on Critical Thinking exclusively, P4wC interventions generate an impressive effect size of 0.89. Reading skills and general academic achievement had much smaller effect sizes, with lower confidence intervals around 0 and, either non-statistically significant, or on the margins of statistically significant, p-values. It is clear that P4wC interventions in these studies had a greater effect on students' critical thinking skills than on their reading or general academic achievement.

**Table 6. Effect size of P4wC by 'Measure of Study'**

| | k | Effect Size | 95% CI | p-value |
|---|---|---|---|---|
| Critical Thinking | 32 | 0.8915116 | 0.623 – 1.130 | 0.000000000000218429 |
| General | 15 | 0.3581801 | 0.076 – 0.640 | 0.01281767 |
| Reading | 11 | 0.3566126 | -0.052 – 0.765 | 0.08707577 |
| Socio-Emotional | 4 | 0.6149649 | -0.621 – 1.851 | 0.3296109 |

## Discussion

This meta-analysis provides an update on the effectiveness of Philosophy for/with Children across a range of measures. Further, subgroup analyses enable a deeper exploration of the effectiveness of P4wC within specific subsets. Most interestingly, subgroup analyses for age, region, and measure of study provided significant insights. These will be unpacked in greater detail in this section.

The subgroup analysis for the variable of age provides important findings for the application of P4wC. P4wC may often be viewed from the outside as an educational initiative that most strongly supports older and/or highly-able students (cf. Haynes, 2007). This usually implies that older students – those who already have a strong grasp on the basics of literacy and numeracy – can access P4wC to extend their thinking capabilities. The data from this meta-analysis challenges this narrow view of P4wC as (1) only beneficial as an extension program for high ability students, or (2) only beneficial to older students who have already grasped basic understandings in literacy and numeracy. This data suggests that there is an equal benefit for P4C in early childhood (>8) as there is in middle

schooling (8-14). For studies conducted with students under the age of 8, the effect size was 0.73. For students aged 8-14 the effect size was 0.75. This demonstrates that there is a clear benefit to students under 8 years of age as well as students aged 8-14. This is an important facet of the research conducted for this paper. It dispels assumptions about P4wC as a practice for older or high ability students. This finding should provide an impetus towards further research and practice in P4wC for the early childhood years of schooling, as the data demonstrates a significant benefit to that age range in addition to the 8-14 age range.

Another subgroup analysed the distinction between regions, separated into Eastern and Western regions of the world. While both Eastern and Western subgroups had an impactful effect size, the effect size in Eastern Regions was significantly more pronounced. This has been explored theoretically in the past, with the idea that Western countries are already adopting teaching practices related to open-ended inquiry, whereas Eastern countries are less likely to be doing so (Park, 1997; Ee & Seng, 2008; Lam & Park, 2016). Therefore, the exposure of P4wC in educational institutions that are more familiar with traditional teaching methods may yield greater outcomes. The subgroup analysis conducted here supports this conclusion.

The most interesting results from subgroup analyses came from the Measure of Study subgroup category. The effect size for critical thinking was 0.89, which reading and general academic achievement were 0.36. Patterns in critical thinking compared with other measures of achievement were similar in previous meta-analyses compared to this one. Moriyón et al. (2005) reported an effect size of 0.63 when the measure was the New Jersey Test of Reasoning Skills (a Critical Thinking measure), but only 0.31 when other measures were used. Similarly, Yan et al. (2018) reported an effect size of 1.06 from measures of "reasoning skill", a lower effect of 0.40 on general cognitive ability, and a lower effect again on reading of 0.28. Trickey and Topping (2004) did not present a subgroup analysis as part of their paper. The results of the current meta-analysis align with these previous ones. A much higher effect size in Critical Thinking measures was found, and much lower effects on measures of general academic and reading ability. Additionally, both reading and general academic achievement had p-values indicating results that are only marginally - or not at all - statistically significant at just below or above 0.05. In contrast, critical thinking had a p-value well below the 0.05 threshold for what is generally considered statistically significant. This indicates that P4wC has a strong significant effect on students' critical thinking, and a potentially small effect size on reading and general academic achievement.

There has been debate and criticism within the P4wC community about the use of P4wC to further reading and other general academic achievement measures (such as maths or writing; Anderson, 2020). This is a point of contention in P4wC due to concerns about the practice being used instrumentally to try and advance standardised test scores in literacy and numeracy, rather than advocating for the intrinsic benefits of philosophical thinking being these aspects of education. This subgroup analysis in this study shows a clear strength of P4wC in fostering critical thinking skills, but not in reading or general academic achievement. It is therefore evidenced that P4wC is more strongly associated with improving standardised quantifiable measures of critical thinking than of standardised quantifiable measures of reading or general academic achievement. This does not discount the possibility that P4wC may improve reading or general academic achievement in non-standardised measures or in non-quantifiable ways, such as through oracy or creativity which are important in both areas. However, given the results of this subgroup analysis demonstrating that critical thinking measures are improving to a much more significant degree than measures of reading or general academic achievement, it is recommended that those interested in implementing P4wC understand that one of the primary positive effects of its implementation is in the improvement of critical thinking for students. Reading, maths, and writing skills may be slightly improved through P4wC, but this improvement is likely to be in a much smaller range than that of critical thinking. The focus of P4wC is not improving standardised test scores in reading, maths, or writing, and the data analysed here shows that P4wC is not a very effective way to improve these facets of education. Therefore, the P4wC community should aim at fostering an understanding about the more significant positive effects of implementing P4wC, or focussing on measures that are understudied in this area such as the impacts on oracy or socio-emotional learning.

## Conclusion

This meta-analysis provides an update on the effectiveness of Philosophy for/with Children across a number of educational measures. This is a reinforcement of the value of P4wC as a valuable educational endeavour that is evidenced-based in improving outcomes for children and young people. There is further evidence that suggests the adoption of P4wC in Eastern countries may be of increased benefit. This is perhaps related to the traditional teaching styles more often practiced in Eastern countries compared to Western countries. Moreover, the value of P4wC is not restricted to certain age groups. This study finds similar results for children in early childhood as it does for children in middle years schooling. P4wC is not something that needs to wait until children are 'old

enough', there is value to be found in P4wC in early childhood education. The most valuable measure that is calculable based on quantitative empirical data is the development of critical thinking in children and young people. Using P4wC as a means to improve general academic achievement or reading comprehension appears to be less effective. However, as a means to improve the critical thinking capacity of students, this meta-analysis demonstrates that P4wC is a powerful tool to achieve this outcome. Future studies in P4wC should not focus on the notion that P4wC can be used to improve outcomes in reading, writing, and maths in a way measured by standardised testing. Research in P4wC should focus on what is particularly important and valuable, which this study shows is the development of critical thinking. Further research in P4wC should also be balanced with the many areas of educational value that P4wC contributes to that are not capable of being measured through quantitative empirical research, and are thus not able to be included in a meta-analysis such as this one (Kilby, 2019b). P4wC is a powerful educational tool, evidenced by the effect sizes calculated in this paper, and should be further promoted through schools across the world.

## Table 7. Summary of Studies Used in Meta-Analysis

| Study | Effect Size | Number of Participants |
|---|---|---|
| Cooke (2015) | 1.473035982 | 26 |
| Fair et al. (2015a) | 0.203146349 | 540 |
| Fair et al. (2015b) | 0.34071392 | 183 |
| Gasparatou and Kampeza (2012) | 0.420065749 | 30 |
| Giménez-Dasí, Quintanilla, and Daniel (2013) (4-year-olds emotional comprehension) | -0.81313661 | 27 |
| Giménez-Dasí et al. (2013) (4-year-olds strategies for interaction) | 0.478567233 | 27 |
| Giménez-Dasí et al. (2013) (5-year-olds emotional comprehension) | 0.527299689 | 33 |
| Giménez-Dasí et al. (2013) (5-year-olds strategies for interaction) | 2.260482902 | 33 |
| Jenkins (1986) | 0.63321872 | 60 |
| Jo (2001) | 0.76747721 | 54 |
| Lam (2012) | 0.61570622 | 42 |
| Marashi (2008) | 1.026746898 | 60 |
| Othman and Hashim (2006) (critical thinking) | 0.361586106 | 48 |
| Othman and Hashim (2006) (reading) | 0.358824478 | 48 |
| Pourtaghi, Hosseini, and Hejazi (2014) (creative elaboration) | 1.570134134 | 32 |
| Pourtaghi et al. (2014) (creative flexibility) | 1.051608837 | 32 |
| Pourtaghi et al. (2014) (creative fluency) | 1.037555227 | 32 |
| Pourtaghi et al. (2014) (creative innovation) | 1.280935499 | 32 |
| Säre, Luik, and Tulviste (2016) (making connections) | 1.317195914 | 125 |
| Säre et al. (2016) (sense-making) | 1.287162424 | 125 |
| Säre et al. (2016) (reasoning) | 1.501159397 | 125 |
| Säre et al. (2016)(talkativeness) | 0.672080214 | 125 |
| Sprod (1997) | 0.488786883 | 54 |
| Tok and Mazi (2015) (Listening comprehension) | 0.298236912 | 74 |
| Tok and Mazi (2015) (Reading comprehension) | -0.085209845 | 74 |
| Topping and Trickey (2007a) | 0.650602188 | 148 |
| Topping and Trickey (2007b) | 0.800601534 | 177 |
| Walker et al. (2013) | 1.33 | 23 |
| Worley and Worley (2019) (Reading) | 0.119565712 | 213 |
| Worley and Worley (2019) (Critical Thinking and Metacognitive Strategies - Success) | 1.410820787 | 220 |

| | | |
|---|---|---|
| Worley and Worley (2019) (Critical Thinking and Metacognitive Strategies - Use) | 2.005494575 | 220 |
| Youssef (2014) | 0.239308792 | 246 |
| Youssef et al. (2016) (Reading Comprehension) | 0.323925408 | 222 |
| Zulkifli and Hashim (2020) | 0.088040065 | 61 |
| Rahdar, Pourghaz, and Marziyeh (2018) (critical thinking) | 1.934146056 | 54 |
| Rahdar et al. (2018) (critical openness) | 1.64672722 | 54 |
| Rahdar et al. (2018) (reflective scepticism) | 3.014506862 | 54 |
| Rahdar et al. (2018) (self-efficacy) | 0.779460014 | 54 |
| Gür et al. (2017) (5-year-olds) | 0.16675672 | 63 |
| Gür et al. (2017) (6-year-olds) | 0.067988472 | 63 |
| Slade (1989) (high achieving) | 0.4946893 | 30 |
| Slade (1989) (low achieving) | 0.242532067 | 20 |
| Palsson (1996) | 0.839639068 | 126 |
| Moriyón, Colom, Lora, Rivas, and Traver (2000) (spatial) | -0.090438402 | 175 |
| Moriyón et al. (2000) (abstract reasoning) | -0.075824404 | 175 |
| Moriyón et al. (2000) (verbal) | -0.04243683 | 175 |
| Moriyón et al. (2000) (numerical) | 0.157829608 | 175 |
| Moriyón et al. (2000) (verbal intelligence) | 0.055296144 | 175 |
| Moriyón et al. (2000) (non-verbal intelligence) | -0.093623681 | 175 |
| Moriyón et al. (2000) (general intelligence) | -0.020117044 | 175 |
| Moriyón et al. (2000) (cognitive) | 0.217043723 | 115 |
| Reznitskaya et al. (2012) | 1.693649189 | 260 |
| Gorard, Siddiqui, and Huat See (2015) (Reading) | 0.134081456 | 1529 |
| Gorard et al. (2015) (Writing) | 0.035847035 | 1529 |
| Gorard et al. (2015) (Maths) | 0.102481007 | 1529 |
| Gorard et al. (2015) (Cognitive) | 0.068740783 | 1511 |
| Allen (1988) (low-ability students - reading) | 2.292065685 | 52 |
| Allen (1988) (high-ability students - reading) | -0.282583441 | 52 |
| Allen (1988) (low-ability students - reasoning) | 0.376777921 | 52 |
| Allen (1988) (high-ability students - reasoning) | 0.345379761 | 52 |
| Lord, Dirie, Kettlewell, and Styles (2021) (Reading) | 0.01 | 7677 |
| | | Total[3] Participants: 12066 |

## Ethical Approval

This meta-analysis deliberately chose to include a wide variety of studies measuring different outcomes that resulted from a P4wC intervention. The papers are open-access publications. Ethical principles and rules have been meticulously observed. In addition, ethics committee approval is not required to conduct this study.

[3] This total includes participants who engaged in a follow up study (e.g., Fair et al. (2015a) and Fair et al. (2015b)), but does not double count the splitting of studies that was done for this meta-analysis (e.g., Gorard et al. (2015) was split in reading, writing, maths, and cognitive assessment categories but used the same students. These students were not double counted for this total.

## References


Allen, T. (1988). I Think, Therefore I Can: Attribution and Philosophy for Children. *Thinking: The Journal of Philosophy for Children, 8*, 14-18.

Anderson, A. (2020). Categories of Goals in Philosophy for Children. *Studies in Philosophy and Education, 39*, 607-623.

Bernard, R. M., Borokhovski, E., Schmid, R. F., & Abrami, P. C. (2014). A Meta-Analysis of Blended Learning and Technology Use in Higher Education: From the General to the Applied. *Journal of Computing in Higher Education, 26*, 87-122. doi:https://doi.org/10.1007/s12528-013-9077-3

Cam, P. (2006). *20 Thinking Tools: Collaborative Inquiry for the Classroom*. Camberwell: ACER Press.

Cooke, P. A. (2015). *The Impact of Engaging in Philosophy with Middle School Children on the Development of Critical Thinking.* (Doctor of Philosophy). University of Rochester, New York.

Dobashi, T. (2009). Philosophising as Archetype Science: Takeji Hayashi's Clinical Pedagogy (H. Hague, Trans.). In E. Marsal, T. Dobashi, & B. Weber (Eds.), *Children Philosophise Worldwide: Theoretical and Practical Concepts* (pp. 185-200). Frankfurt: Peter Lang.

Ee, J., & Seng, T. O. (2008). Cultural Influences of the East & West: Implications on the Educational Reforms in the Singapore Context. *KEDI Journal of Educational Policy, 5*(1).

Fair, F., Hass, L. E., Gardosik, C., Johnson, D. D., Price, D. P., & Leipnik, O. (2015a). 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).

Fair, F., Hass, L. E., Gardosik, C., Johnson, D. D., Price, D. P., & Leipnik, O. (2015b). Socrates in the Schools: Gains at Three-Year Follow-Up. *Journal of Philosophy in Schools, 2*(2).

FAPSA. (2017). Federation of Australasian Philosophy in Schools Associations. Retrieved from http://fapsa.org.au/

Gasparatou, R., & Kampeza, M. (2012). Introducing Philosophy for Children (P4C) in Kindergarten in Greece. *Analytic Teaching and Philosophical Praxis, 33*(1), 72-82.

Giménez-Dasí, M., Quintanilla, L., & Daniel, M. F. (2013). Improving Emotion Comprehension and Social Skills in Early Childhood through Philosophy for Children. *Childhood & Philosophy, 9*(17), 63-89.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-Analysis in Social Research*. Beverly Hills: Sage.

Golding, C. (2014). A Handy Account of Philosophy in Schools. *Journal of Philosophy in Schools, 1*(1), 68-88.

Gorard, S., Siddiqui, N., & Huat See, B. (2015). *Philosophy for Children: Evaluation Report and Executive Summary*. Retrieved from Durham:

Gregory, M., & Laverty, M. (2017). *In Community of Inquiry with Ann Margaret Sharp*: Routledge.

Gür, C., Koçak, N., & Muharrem, S. (2017). The Study of the Effect of Thinking from Multidimensional Perspectives Training Program on the School Readiness of 5 and 6 Year-old Children. *Education and Science, 42*(189), 153-165.

Hattie, J. (2020). Visible Learning Meta: Global Research Database. http://www.visiblelearningmetax.com/Influences: Corwin.

Haynes, J. (2007). Freedom and the Urge to Think in Philosophy with Children. *Gifted Education International, 22*(2-3), 229-237.

Hedges, L. V., & Olkin, L. (1985). *Statistical Methods for Meta-Analysis*. Orlalndo: Academic Press.

Hedges, L. V., Shymansky, J. A., & Woodworth, G. (1989). *A Practical Guide to Modern Methods of Meta-Analysis*. Washington: National Science Teachers Association.

ICPIC. (2017). The International Council of Philosophical Inquiry with Children. Retrieved from http://icpic.org/

Institute of Education Sciences. (2017). *What Works Clearinghouse: Standards Handbook* (4 ed.). Washinton: U.S. Department of Education.

Jenkins, J. (1986). Philosophy for Children Programme at a Gloucestershire Comprehensive School in Great Britain. *Thinking: The Journal of Philosophy for Children, 6*(3), 33-37.

Jo, S. (2001). Literacy: Constructing Meaning through Philosophical Inquiry. *Analytic Teaching, 21*(1), 44-52.

Kennedy, N. (2007). Philosophy for Children and Mathematics Teaching and Learning: Insights, Models, Possibilities. In E. Marsal, T. Dobashi, & B. Weber (Eds.), *Children Philosophise Worldwide: Theoretical and Practical Concepts*. Frankfurt: Peter Lang.

Kennedy, N., & Marsal, E. (2023). *Dialogical Inquiry in Mathematics Teaching and Learning: A Philosophical Approach*. Berlin: LIT Verlag.

Kilby, B. (2019a). Why teachers' beliefs and values are important in p4c research: an australian perspective. *Childhood & philosophy*, *15*.

Kilby, B. (2019b). Who am I? Why am I here? Why children should be taught philosophy (beyond better test scores). *The Conversation*. Retrieved from https://theconversation.com/who-am-i-why-am-i-here-why-children-should-be-taught-philosophy-beyond-better-test-scores-120160

Kilby, B. (2021). Dialogic Pedagogies: Defining and Analyzing Four Types of Dialogue in Education. *Analytic Teaching and Philosophical Praxis, 41*(2).

Kilby, B. (2022a). Competition in Philosophy is a Feminist Issue. *Journal of Philosophy in Schools, 9*(2), 90-113.

Kilby, B. (2022b). Philosophy for/with Children and Teachers' Perspectives: The Need to Understand Teachers' Beliefs and Values. In *Philosophy with Children and Teacher Education* (pp. 42-50). Routledge.

Kilby, B. (2023). Gender and communication in children and school: aligning theory and evidence. *SN Social Sciences*, *3*(2), 36.

Kizel, A. (Ed.). (2022). *Philosophy with children and teacher education: global perspectives on critical, creative and caring thinking*. Taylor & Francis.

Kizel, A. (2024). *Enabling Students' Voices and Identities: Philosophical Inquiry in a Time of Discord*. Lanham: Lexington Books/Fortress Academic.

Kohan, W., Murris, K., Mostert, P., Sutcliffe, R., D'Olimpio, L., & Worley, P. (2017, 30 June). *What Does P4C Capture.* Paper presented at the 18th International Council of Philosophical Inquiry with Children, Madrid.

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 Evaluation, 18*(2), 187-203.

Lam, C. M., & Park, J. W. (2016). Introduction: A Sociological and Philosophical Approach to Education. In C. M. Lam & J. W. Park (Eds.), *Sociological and Philosophical Perspectives on Education in the Asia-Pacific Region. Education in the Asia-Pacific Region: Issues, Concerns and Prospects* (Vol. 29). Singapore: Springer.

Lipman, M. (1983). *Lisa*. Montclair: IAPC.

Lipman, M. (1985). Thinking Skills Fostered by Philosophy for Children. *Thinking and Learning Skills, 1*, 83-108.

Lipman, M. (2003). *Thinking in Education* (2 ed.). Cambridge: Cambridge University Press.

Lipman, M., & Bynum, T. (1976). *Philosophy for Children*. Oxford: Blackwell.

Lone, J. M. (2021). *Seen and Not Heard: Why Children's Voices Matter*. Landham: Rowman & Littlefield.

Lord, P., Dirie, A., Kettlewell, K., & Styles, B. (2021). *Evaluation of Philosophy for Children: An Effectiveness Trial*. Retrieved from London:

Marashi, S. M. (2008). Teaching Philosophy to Children: A New Experience in Iran. *Analytic Teaching, 27*(1), 12-15.

Montclair State University. (2017). Institute for the Advancement of Philosophy for Children. Retrieved from http://www.montclair.edu/cehs/academics/centers-and-institutes/iapc/

Moriyón, F. G., Colom, R., Lora, S., Rivas, M., & Traver, V. (2000). Valoración de «Filosofía para Niños»: un programa de enseñar a pensar. *Psicothema, 12*(2), 207-211.

Moriyón, F. G., Robello, I., & Colom, R. (2005). Evaluating Philosophy for Children: A Meta-Analysis. *Thinking: The Journal of Philosophy for Children, 17*(4).

Murris, K. (2016). *The Posthuman Child: Educational Transformation Through Philosophy with Picturebooks*. Abingdon: Routledge.

Murris, K., Haynes, J., & Gregory, M. J. (2016). *The Routledge International Handbook of Philosophy for Children*. London: Routledge.

Othman, M., & Hashim, R. (2006). Critical Thinking & Reading Skills: A Comparative Study of Reader Response & the Philosophy for Children Approaches. *Thinking: The Journal of Philosophy for Children, 18*(2), 26-34.

Palsson, H. (1996). We Think More than Before about Others and their Opinions (An Evaluation Report from Ireland). *Thinking: The Journal of Philosophy for Children, 12*(4), 24-29.

Park, J. W. (1997). Education for Democratic Citizenship and the Community of Inquiry: Some Implications for Educational Reform in Korea. *Thinking: The Journal of Philosophy for Children, 13*(3), 2-5.

Pourtaghi, V., Hosseini, A., & Hejazi, E. (2014). Effectiveness of Implementing Philosophy for Children Program on Students' Creativity. *Scientific Journal of Pure and Applied Sciences, 3*(6), 375-.

Prior, S., & Wilks, S. (2018). Philosophy in Public and Other Educational Spaces. *Philosophical Inquiry with Children*, 185-198.

Pritchard, M. (2014). Philosophy for Children. In Edward N Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 ed.). Stanford: Metaphysics Research Lab, Stanford University.

Rahdar, A., Pourghaz, A., & Marziyeh, A. (2018). The Impact of Teaching Philosophy for Children on Critical Openness and Reflective Skepticism in Developing Critical Thinking and Self-Efficacy. *International Journal of Instruction, 11*(3), 539-556.

Reznitskaya, A., Glina, M., Carolan, B., Michaud, O., Rogers, J., & Sequeira, L. (2012). Examining Transfer Effects from Dialogic Discussions to New Tasks and Contexts. *Contemporary Educational Psychology, 37*, 288-306.

Säre, E., Luik, P., & Tulviste, T. (2016). Improving Pre-Schoolers' Reasoning Skills using the Philosophy for Children Programme. *Trames, 20*(3), 273-295.

Slade, C. (1989). Logic in the Classroom. *Thinking: The Journal of Philosophy for Children, 8*(2), 14-20.

SOPHIA. (2017). SOPHIA: European Foundation for the Advancement of Doing Philosophy with Children. Retrieved from http://www.sophianetwork.eu/

Sprod, T. (1997). Improving Scientific Reasoning through Philosophy for Children. *Thinking: The Journal of Philosophy for Children, 13*, 11-16.

Tok, S., & Mazi, A. (2015). The Effect of Stories for Thinking on Reading and Listening Comprehension: A Case Study in Turkey. *Research in Education, 93*. doi:http://dx.doi.org/10.7227/RIE.0006

Topping, K. J., & Trickey, S. (2007a). Collaborative Philosophical Enquiry for School Children: Cognitive Gains at Two-Year Follow-Up. *British Journal of Educational Psychology, 77*(4).

Topping, K. J., & Trickey, S. (2007b). Collaborative Philosophy Enquiry for School Children: Cognitive Effects at 10-12 Years. *British Journal of Educational Psychology, 77*(2).

Trickey, S., & Topping, K. J. (2004). Philosophy for Children: A Systematic Review. *Research Papers in Education, 19*(3).

Vansieleghem, N. (2013). What is Philosophy for Children? From an educational experiment to experimental education. *Educational Philosophy and Theory, 46*(11), 1300-1310. doi:10.1080/00131857.2013.771442

Walker, C. M., Wartenberg, T. E., & Winner, E. (2013). Engagement in Philosophical Dialogue Facilitates Children's Reasoning about Subjectivity. *Developmental Psychology, 49*(7), 1338-1347.

Wartenberg, T. (2019). *Philosophy in Classrooms and Beyond: New Approaches to Picture-Book Philosophy*. Lanham: Rowman & Littlefield.

Worley, P., & Worley, E. (2019). Teaching Critical Thinking and Metacognitive Skills through Philosophical Enquiry: A Practitioner's Report on Experiments in the Classroom. *Childhood & Philosophy, 15*, 1-34.

Yan, S., Walters, L. M., Wang, Z., & Wang, C. 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.

Youssef, C. (2014). *A Multilevel Investigation into the Effects of the Philosophical Community of Inquiry on 6th Grade Students' Reading Comprehension, Interest in Maths, Self-Esteem, Pro-Social Behaviours and Emotional Well-Being.* (Doctor of Philosophy). Queensland University of Technology, Brisbane.

Youssef, C., Campbell, M., & Tangen, D. (2016). The Effects of Participation in a P4C Program on Australian Elementary School Students. *Analytic Teaching and Philosophical Praxis, 37*(1), 1-19.

Zulkifli, H., & Hashim, R. (2020). Philosophy for Children (P4C) in Improving Critical Thinking in a Secondary Moral Education Class. *International Journal of Learning, Teaching and Educational Research, 19*(2), 29-45. doi:https://doi.org/10.26803/ijlter.19.2.3