META-ANALYSIS OF THE EFFECTS OF PHILOSOPHY FOR CHILDREN PROGRAM ON STUDENTS' COGNITIVE OUTCOMES

A Thesis

by

SIJIN YAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Lynne Masel Walters
Committee Members,	Monica Neshyba
	Claire Elise Katz
Head of Department,	Michael Anthony de Miranda

May 2017

Major Subject: Curriculum and Instruction

ABSTRACT

Philosophy for Children (often abbreviated as P4C) is an educational program that aims at introducing philosophy into K-12 education with its distinguishable curriculum design and pedagogy. It begins with stimuli (children's literature, questions, or other media) that contain philosophical themes to inspire students' questions and discussions. The class proceeds through students' philosophical conversation in a community of inquiry, rather than through traditional lectures.

This meta-analysis examines the research on Philosophy for Children, published from 2002 to 2016 to show how the program affects students' cognitive outcomes and provide practical guidance to educators. A total of 10 studies (including 2 follow-up studies) between 2002 to 2016, representing 1509 students from second grade to twelfth grade, are included in this meta-analysis. Results show medium effects (d=0.58) on all outcomes except reasoning skills outcomes, which are significant (d= 1.06).

The results suggest that the philosophy for children program has overall moderate effects on students' cognitive and socio-psychological abilities, and has significant, positive effect on the reasoning abilities of students.

DEDICATION

To my mother, Ping Zhang, who will forever remain for me a teacher *par excellence*.

ACKNOWLEDGEMENTS

First of all, I would like to thank my committee chair, Dr. Lynne Masel Walters, who introduced me to pre-college philosophy education and encouraged me to explore beyond disciplinary boundaries.

I wish to thank Zhuoying for contributing the coding work of the selected literature.

To Dr. Chia-Chiang Wang, I owe a debt of gratitude. The project would have certainly been different without his mentorship in quantitative methodology.

I am deeply grateful to Dr. Claire Elise Katz for her appreciated support through the thesis process.

Thanks go to the education and philosophy department at Texas A&M University. I have been extraordinarily fortunate to receive such educative experiences.

I would like to thank my parents, without whom I could never have had the luxury of studying here at Texas A&M University.

Finally, to Gaofeng Fan I give my love. The enduring companionship he has offered me the last several years, especially during life's moments of uncertainty, has been invaluable.

To all who I have not given thanks personally but have enriched my life experiences in the "Aggieland," please accept my deepest appreciations.

CONTRIBUTORS AND FUNDING SOURCES

This work was supervised by a thesis committee consisting of Professor Lynne Masel Walters and Professor Monica Neshyba of the Department of Teaching, Learning and Culture and Professor Claire Elise Katz of the Department of Philosophy. All work for the dissertation was completed independently by the student.

The coding section for Chapter III was collaborated with Zhuoying Wang, a doctoral student in the Department of Teaching, Learning and Culture.

The graphs and tables in this thesis are created by the student.

There are no outside funding contributions to acknowledge related to the research and authorship of this thesis.

NOMENCLATURE

P4C	Philosophy for Children
SES	Socio-economic Status

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

It has been just over 100 years since the publication of John Dewey's *Democracy and Education* (Dewey, 1916), which aims at democratizing education as the instrument of progressive social transformation (Rescher et al., 2002). The Philosophy for Children program (P4C), which according to its founder Matthew Lipman (2004), built unapologetically on Deweyan foundations, has been implemented throughout the world and its materials have been translated into numerous languages (Gorard, Siddiqui, & Huat See, 2015).

Introduction of Philosophy for Children

Philosophy for Children is an educational program initiated by Matthew Lipman and his colleagues in the Institute for the Advancement of Philosophy for Children (IAPC) in the early 1970s (Brandt, 1988; Lam, 2012; Marashi, 2008; van der Straten Waillet, Roskam, & Possoz, 2015). Witnessing the deficit of college students' argumentative performance in public discourses and the tumultuous political environment of his time, Lipman argued that philosophy should no longer be confined to college and academic research. Children, he said, even in elementary grades, can begin a quest in philosophy so that they could learn how to think and reason (Brandt, 1988). Currently, with the help of numerous philosophers, educators and researchers, Philosophy for Children has become a global movement that has spread across 50

countries, and its material has been translated into 20 languages (Daniel & Auriac, 2011).

Curricular Design

The Philosophy for Children movement believes in the possibility of making the content of philosophy accessible to K-12 students by 1) removing the formidable terminology from philosophy and 2) using children's literature to bring the philosophical discussion to class (Lipman, 2009). Thus, educators and philosophers in P4C program have designed children's literature that entails philosophical questions and arguments in an approachable language. In class, students are encouraged to read the text aloud, normally each student read a part of the text and everyone has a turn, so that they could share meaning with each other, read aloud with expression and emotions, and learn to carefully listen to others (Lipman, 2009). Then, teachers collect students' questions, which they find puzzling, and write them down on the chalkboard/whiteboard for further discussion (Lipman, 2009).

The contents of the curriculum are not random. Rather, they are intentionally crafted in order to be consistent with students' lived experiences. Dewey proposed two principles that demonstrate the nature of *educative* experiences, as opposed to non-educative experiences (Dewey, 1986). The first one is continuity. It requires that all experiences are carried forward and influence future experiences (Dewey, 1986). The second one is interaction. It is the situation formed by the assigned activity and the internal condition of the person, which hugely influences the quality of experiences (Dewey, 1986).

To provide educative experiences, the educator should consider the background of students and provide appropriate new events that facilitate later experiences. Thus, the centrality of lived experiences for students in the P4C curriculum design cannot be overlooked.

Pedagogical Design

Community of Inquiry

The core method of implementing the Philosophy for Children program is that of a community of inquiry. This concepts was initially proposed by an American philosopher and scientist Charles Sanders Peirce in his article *The Some Consequences of Four Incapacities* (Peirce, 1868) as a rebuttal of Cartesian and traditional-modern epistemology. It roots on the view that it is pernicious to make single individuals absolute judges of truth (Peirce, 1868). We need a *community* of inquiry to "grind off the arbitrary and the individualistic character of thought" (Peirce & Houser, 1998).

Dewey fleshed out the Peircean theory of inquiry into his philosophy of education (Lipman, 2004), and further contended that the community of inquiry in the classroom can let students cultivate the habit to engage in a democratic life. The educators and philosophers in the Philosophy for Children movement are deeply influenced by both the Peircean and Deweyan community of inquiry (Sharp, Reed, & Lipman, 2010) and thus, regard it as the central pedagogy of this program.

However, a community is not only a birthplace of respect, diversity and mutual learning, it can also generate conflicts, discontent feeling of divergences, bullying and

exclusion. The idea of building "an intellectually safe community" is a timely response to such concerns in P4C (Butnor, 2012).

Intellectually Safe Community

Intellectual safety is the idea that all participants in the classroom community (students and teachers) feel free to ask virtually any question (Jackson, 2001) or demonstrate any ideas, as long as respect for all persons is honored.

Intellectual safety doesn't mean complete relaxation, free of stress and doubt, since intellectual growth must involve some kind of discomfort (Butnor, 2012). Struggling through a difficult idea, rocked by new perspective, trying to defend one's point of view may also be intimidating but rewarding. Thus, an intellectually safe community for students is like a rich ground for plants where they can embrace the complexity of nature, learn to be stronger and healthy, instead of luxuriating in a warm house which over-protects them from any hardness. In the intellectually safe classroom, students as well as teachers can be challenged in their world-views, but at the same time feel supported and safe (Butnor, 2012).

As Butnor (2012) has said, showing vulnerability of teachers themselves (Butnor, 2012) is an effective way to let students feel that they are being cared for and respected. If teachers themselves would like to put their own ideas at risk (Butnor, 2012), students will be greatly motivated to do the same.

Conclusion

In conclusion, the P4C program is a series of educational experiments that put the interdependence of theory and practice at the center of its development. Thus, the evaluation of such a program is critical, not only because it examines the effectiveness of a unique mode of instruction, but because it also helps us to grasp a deeper understanding about the practical consequences of various classroom-based educational perspectives, such as community, theory of inquiry and learning, philosophy as a discipline, the cognitive abilities of K-12 students, children literature and philosophy.

CHAPTER II

LITERATURE REVIEW

The radical nature of P4C in transforming our vision of the function of philosophy from a mere sphere of intellectual minorities to a rich soil for every human being with diverse age and life experiences, generated a lot of skepticism and debate.

Questions Regarding P4C

The first question is whether children are intellectually mature enough for philosophy (Daniel & Auriac, 2011). According to Piaget, children are not equipped with the ability to do abstract thinking (Piaget, 1931). However, cognitive scientists recently have shown that children have surprisingly cognitive abilities (Gopnik, 2009). Furthermore, philosopher Gareth Matthews stressed the freshness of children's ideas through his philosophical discussion with the young people, and the deep-rooted condescending attitudes behind this "children are not capable of doing philosophy" argument (Matthews, 1980, 1994).

The second issue is the evaluation of the effectiveness of P4C and the various ways this program can be conducted to benefit more students, especially those who are challenged and disadvantaged, at an affordable cost (Gorard et al., 2015).

Since 1970s, the outcomes measured in P4C research can be divided into two categories: (1) thinking abilities and academic abilities (2) socio-psychological outcomes related to attitudes toward academics, prosocial attitudes and behavior and such. Even

though there are emerging studies that appraise the effectiveness of P4C in socio-psychological area (Abbasi & Ajam, 2016; Dasí, Quintanilla, & Daniel, 2013; Scholes et al., 2016), the extant literature is still limited.

Most studies focus on the goal of the Philosophy for Children program to provide a more formal training to develop students' thinking skills, especially for reasoning skills outside the realm of science and mathematics (García-Moriyón, Rebollo, & Colom, 2005; Gorard et al., 2015; Gregory, 2011; Säre, Luik, & Tulviste, 2016; Trickey & Topping, 2004). From Lipman's perspective, philosophy is the finest instrument yet devised for the perfection of the thinking skills (Lipman, Sharp, & Oscanyan, 2010).

This study tries to provide a quantitative synthesis and evaluation of the extant literature to get a whole picture of the cognitive effects of this program and what factors will influence the program.

Previous Review of the Evaluation of P4C

P4C may have a positive effect on students' cognitive abilities. In 2004 and 2005, two reviews (García-Moriyón et al., 2005; Trickey & Topping, 2004) were conducted to synthesize research on the effects of P4C.

The quantitative systematic analysis by Trickey and Topping (2004) investigated the influence of P4C on students in general, with the conclusion that P4C has a moderate positive effect on students' abilities with low variance. It collected 8 controlled experiments regarding Philosophy for Children from 1970s to 2002. But this meta-analysis combined the cognitive abilities and affective abilities without a theoretical

foundation, even though the relationship between the two has not been yet researched a consensus (Hidi, Renninger, & Krapp, 2004). The study did not include heterogeneity test, which could help the practitioners to understand if the variability in effect sizes is just from sampling error around the mean or have more to expect.

The meta-analysis conducted by García-Moriyón, Rebollo, and Colom (2005) examined the relationship between P4C and reasoning skills from 1976 to 2002, with the finding that P4C also has a positive moderate influence on students' reasoning abilities. Due to the various types of study designs in this field, the meta-analysis included posttest experiments, single group study with pre and posttest, and controlled experiments. The result showed significant differences among those types of studies, in which the more rigorous controlled experiments tended to show lower effect sizes.

Thus, despite the evident contributions of these two reviews (García-Moriyón et al., 2005; Trickey & Topping, 2004), they also have limitations which this current meta-analysis intends to dissolve:

First, to ensure the rigor of this meta-analysis, the studies included must be either random controlled experiments or quasi-experiments.

Second, since the P4C movement has spreaded worldwide and research was conducted on different continents (Lam, 2012; Marashi, 2008; Nia, 2014; Youssef, 2014), this meta-analysis aims to conduct an exhaustive search globally and include studies from both Western and non-western countries.

Third, this study will only combine effect sizes that relate to students' cognitive development, instead of synthesizing cognitive outcomes with socio-psychological outcomes.

After the publication of these two meta-analysis, a larger collection of literature on the effects of P4C on cognitive abilities has grown since these initial meta-analyses. Most importantly, these studies have included increasing rigor of study designs, a larger number of participants, and follow-up studies (Fair et al., 2015a; Fair et al., 2015b; Topping & Trickey, 2007a, 2007b). This suggests that we could have a more recent overview of the literature that addresses the limitations of previous two analyses and helps educators to have a clearer understanding of the overall effectiveness of P4C, the conditions under which it is more or less effective, or who is best served by this program.

The Present Study

The purpose of the current meta-analysis is to examine the reported effects of P4C from 2002 to 2016, immediately following the publication of the two articles that analyzed studies from 1970s to 2002. In addition, this meta-analysis examines which variables -- age, location, assessment measure, socioeconomic status, duration -- of the intervention might affect the degree of improvement in these effects.

Through this meta-analysis, the researchers hope to find answers for the following question:

First, what does the cumulative research suggest regarding the impact of P4C on students' cognitive abilities?

Second, do characteristics of intervention, students, or outcome types influence the magnitudes or direction of the effect of the Philosophy for Children Program?

Because the two previous meta-analyses included studies from the 1970s to 2002, the research reported has dealt with information about studies on philosophy for children education published from 2002 to 2016 and estimated effect sizes of various components of the combined research studies.

CHAPTER III

METHODOLOGY

In this study, the effectiveness of P4C was tested through a meta-analysis, which is a method that merges the results of many independent researchers conducted on a particular topic and performs statistical analysis (Çoğaltay & Karadağ, 2016). The procedure and detailed descriptions of this quantitative synthesis was described in this chapter.

Study Search and Retrieval

This study included the online databases British Education Index, ERIC, Education Full Text (H.W. Willson), Education Source, Academic Search Complete, PsycINFO database from 2002 to 2016.

The keyword used was *philosophy n2 children*, which means it specified 2 maximum intervening words between philosophy and children, in any order. The researchers included both published journals and doctoral dissertations.

Second, the researcher conducted non-electronic journal search. The index of the journal *Thinking: Philosophy for Children* was consulted for articles. Then, potential relevant articles were retrieved from a library.

The third was a google scholar search engine, *Journal of Philosophy in Schools,* as well as references listed in collected studies.

Through the initial searches, 1180 articles were potentially relevant.

Inclusion Criteria

In order to be included in this meta-analysis, studies had to meet the following criteria:

1. Participants: The population of interest was pre-college students enrolled in a Philosophy for Children program and their control-group counterparts. College studies and teacher education research were excluded from the study.

2. Intervention/method: The study included a philosophy for/with children program as an intervention. Philosophy for Children programs have different names, from philosophy for children to philosophy with children, sometimes abbreviated to P4C or PwC. All of these studies were included.

3. Publication date: The study should be published between 2002 to 2016.

3. Research design:

a. The study must be either random controlled experiments or quasi-experiments.

b. A quantitative measure of outcomes was used in the study to calculate the magnitude of effects of the intervention.

c. The outcome variables contained a measurement of cognitive outcomes, such as reasoning ability, comprehension ability, general cognitive ability, academic development, pro-social ability and emotional intelligence.

d. This meta-analysis focuses on the cognitive outcomes of P4C. Thus, the control group must not implement any thinking skill intervention, which rendered the study as comparing the different critical thinking programs.

To ensure all studies included are well-designed and able to provide enough data for the computation of effect sizes, researchers left out studies that fail to conform any of those criteria.

A considerable number of studies were excluded in this stage particularly because many of them adopted qualitative methodology, which cannot provide enough effect size for meta-analysis. This is because of the nature and limitation of meta-analysis itself in which it applies only to research studies that produce quantitative findings, that is, studies using quantitative measurement of variables and reporting descriptive or inferential statistics to summarize the resulting data (Lipsey & Wilson, 2001). This rules out qualitative forms of research such as case studies, ethnography, and "naturalistic" inquiry (Lipsey & Wilson, 2001).

Because of the large amount of literature, there are two steps of screening during the selection of included studies. First, the author screened the abstract and results of each of the 1180 studies. There were 44 articles remained. Then, the author scrutinized the full text of each article and excluded 27 studies which failed to meet the requirement of the inclusion criteria. The included 17 studies were assigned to the coding process.

Coding Procedure

The coding process is a data extraction process, picking clear and appropriate data from the pile of complex information (Çoğaltay & Karadağ, 2016). The manual for coding studies was developed by the researcher before proceeding to the coding.

Content Validity and Interrater Reliability

Content validity for the coding sheet and coding manual was determined originally by submitting to two professors for feedbacks on the appropriateness of variables and categories created in this study. The first professor is working primarily on Curriculum & Instruction, and the second professor is in educational psychology.

There were two coders in this study. The first coder is the author of this meta-analysis, a Master student in Curriculum and Instruction; the second coder is a doctoral student in the same major. Both of the two coders have received basic statistical education for quantitative research.

After the coding manual was created, two coders met first to discuss and went over the coding manual until everything was clear. The coders scrutinized each of article and extracted the variables and outcomes from the studies and inputted them into an excel document.

To determine interrater reliability, the two researchers independently coded five studies to ensure that the inclusion/exclusion criteria were met. The researchers achieved an interrater reliability of 90.0 percent across those studies.

Analysis of coder disagreements resulted in the refinement of some definitions and decision rules for some codes. Then, each coder individually coded the remainder of the studies.

Missing Data

During the coding process, the first coder contacted the original authors from two different references for standardized deviations and means to calculate the effect sizes. One set of data was obtained.

Through these five means of search, the author attained a total of 17 articles, two of them were follow-up studies.

Data Analysis

Effect size computation, homogeneity test and moderator analysis were conducted in this study.

The effect size acquired in the meta-analysis study is a standard measure value used to determine the strength and direction (Çoğaltay & Karadağ, 2016) of the effectiveness of Philosophy for Children program on students' cognitive outcomes. Cohen's *d* was used to adjust and determine the effect sizes of each study. All the effect sizes in every study were aggregated. Each study has only one effect size for cognitive outcomes. This is because in meta-analysis, the unit of analysis is the individual research study and any two or more effect sizes that come from the same study are statistically dependent (Lipsey & Wilson, 2001). In meta-analysis, all data analysis involving effect sizes was weighted analysis (Lipsey & Wilson, 2001).

Two main models were used in the analysis of Heterogeneous distribution of Effect Sizes: the fixed effects model and the random effects model. Under a fixed effects model, according to Lipsey and Wilson (2001), an effect size observed in a study is

assumed to estimate the corresponding population effect with random error that stems only from the chance factors associated with subject-level sampling errors in that study. If it is believed that the research is not equal in terms of functionality, and if generalizations through the estimated effect size are to be made for greater populations, then the model that should be used is the *random effects model*(Çoğaltay & Karadağ, 2016).

Conclusion

After the analysis, the author evaluated the data and determined whether there P4C has a positive effect on students' cognitive abilities. In addition, the study also examined which study settings might relate to the effectiveness of P4C and which outcomes in cognitive abilities is the most significant.

This study will help educators understand the effectiveness of Philosophy for Children program especially in students' cognitive abilities and provide recommendations for future practices.

CHAPTER IV

RESULTS

The present meta-analysis investigated the extent to which Philosophy for Children program effects students' cognitive outcomes, and conducted moderator analyses. This quantitative synthesis of empirical studies includes studies from 2002 to 2016. The year 2002 was chosen as the cutoff date because two previous meta-analyses have encompassed the literature in regard to P4C and its educational outcomes before 2002 (García-Moriyón et al., 2005; Trickey & Topping, 2004).

Studies Excluded

Seven studies were excluded from the seventeen research papers during the last stage of this analysis. Figure 1 provides an overview of the excluded papers. Two studies were excluded because their participants were younger than first grade (Dasí et al., 2013; Säre et al., 2016). This is because synthesizing and comparing children who are 3 to 5 years old and students are in K-12 education are inappropriate.

One study was excluded because the experiment's control group was still under another thinking skills intervention (Othman & Hashim, 2006). This study compared P4C to other thinking program (the Reader Response Program). Thus, the control group not a neutral group, still being affected by another intervention. The controlled groups in all the included studies were not under any thinking skills intervention.

Two studies were not included due to the lack of means and standard deviations to calculate the effect sizes (Colom, Moriyon, Magro, & Morilla, 2014; Walker, Wartenberg, & Winner, 2013).

One study was excluded because its outcome measure was spiritual development, which is not considered to be within the realm of cognitive outcomes (Abaspour, Nowrozi, & Latifi, 2015).

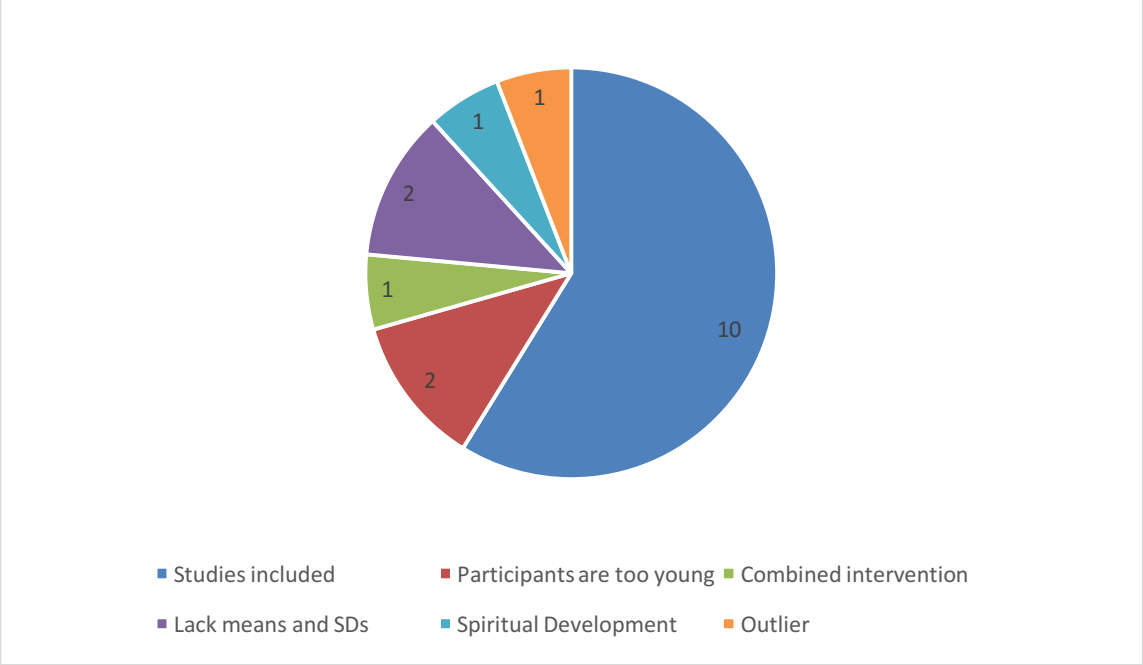


Figure 1 The Included and Excluded Studies

There was one study (Gorard et al., 2015) that was classified as an outlier, and then excluded from the study. This study had the largest sample size (16 times larger than the mean sample size) and the lowest effect size (0.13 times smaller than the mean

effect size). In meta-analysis, extreme effect sizes that are discrepant from the preponderance of those found in the research of interest are unrepresentative of the results of that research and may distort the result of meta-analysis (Lipsey & Wilson, 2001). To achieve at a reasonable summary of the quantitative findings of a body of research studies, the researcher eliminate them from the analysis.

Studies Included

A total of 10 controlled experiments were included in this analysis, which together report the findings of eight independent studies and two follow-up studies. Among the ten studies, nine of them are published journal articles, and one is dissertation (Youssef, 2014). Table 1 provides an overview of the characteristics of each citation included in the synthesis. The sample sizes in these studies ranged from 28 to 540, representing 1509 students from second grade in elementary to first grade in high schools. The sample sizes of studies were adjusted in following way:

First, if the sample size of one study in posttest is smaller than the pretest, then the whole sample size of this study is coded as the smaller one.

Second, in the case of one study with a follow-up study (Fair et al., 2015a; Fair et al., 2015b), the sample sizes of the independent studies were adjusted to the corresponding student groups with the follow-up studies.

Thus, the overall sample sizes in this meta-analysis is smaller than original literatures.

Table 1 Characteristics of Included Studies

Reference	*Study Type*	*Location*	*Sample Size*	*Grade/Age Level*	*Outcome Measure*
(Abbasi & Ajam, 2016)	Intervention	Iran	50	Second	Questionnaire of Educational Progress*
(Fair et al., 2015b)	Intervention	United States	177	Seventh	CogAT
(Fair et al., 2015a)	Follow-Up	United States	115	Seventh Grade- Two Years after	CogAT
(Lam, 2012)	Intervention	China	28	Secondary School First Grade	NJTRS
(Marashi, 2008)	*Intervention*	*Iran*	*60*	*Eighth*	*NJTRS*

Table 1 Continued

Reference	Study Type	Location	Sample Size	Grade/Age Level	Outcome Measure
(Nia, 2014)	Intervention	Iran	60	High School First Grade	Abedi's Test of Creativity
(Tok & Mazı, 2015)	Intervention	Turkey	74	Fifth Grade	Reading Comprehension Test* and Listening Comprehension Test*
(Topping & Trickey, 2007a)	Intervention	United Kingdom	540	Ten year old students	CAT

Table 1 Continued

Reference	Study Type	Location	Sample Size	Grade/Age Level	Outcome Measure
(Topping & Trickey, 2007b)	Follow-Up	United Kingdom	183	Ten Year Old Students – Two Years After	CAT
(Youssef, 2014)	*Intervention*	*Australia*	*222*	*Sixth Grade*	*Reading Comprehension Test*

*Note: CogAT: Cognitive Ability Test (American Version); CAT: Cognitive Ability Test (United Kingdom Version); NJTRS: New Jersey Test of Reasoning Skills; *: Tests developed by researchers*

Overall Effectiveness of P4C

There are 10 studies, representing 1509 students in this meta-analysis. The overall standardized mean effect size aggregated from the ten studies was 0.58 with a 95% confidence interval ranging from 0.33 to 0.53. According to Cohen's Rule of Thumb (VanVoorhis & Morgan, 2007), the mean effect size represents that P4C has a modertate, positive overall cognitive effect for students.

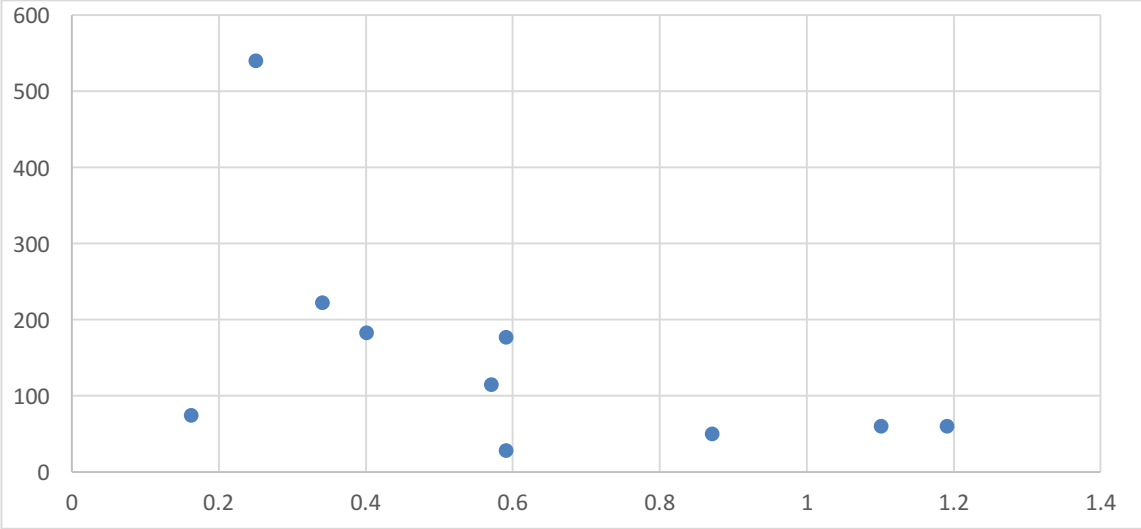
In this study, the homogeneity test was found to be statistically significant (Q = 26.59, p < 0.001), which means that there is more variability in effect sizes than would be expected from sampling error around the mean. Table 2 provides the overall results and omnibus test of this meta-analysis. Figure 2 is the scatter plot of mean effect sizes of each study included.

Table 2 Overall Results and Omnibus Test of P4C Studies

			Median	Fixed		Random		
	k	*N*	ES (*d*)	ES (*d*)	95% CI	ES (*d*)	95% CI	*Q*
P4C	10	1,509	.58	.43	[.33, .53]	.50	[.33, .66]	26.59**

Note. *k* = study size; *N* = total number of participants; CI = confidence interval; *Q* = omnibus test of homogeneity. *** $p < .001$

Figure 2 Scatter Plot



Note: x: Effect Size; y: Sample Size

Results of Moderator Analysis

Since the homogeneity test was found to be statistically significant, moderator analysis was used to find out the reasons of variance between effect sizes. In this meta-analysis, subgroup analysis was employed to detect moderating effects. Six moderator variables were tested: grade level, socio-economic status of students, location of studies, study design (random or quasi-experiments), total time of intervention and outcome measures. Table 3 provides a detailed statistical description of the result of moderator analysis.

Table 3 Moderator Testing of Study

Variable	*k*	*N*	*d*	95% CI	Q_B	ANOVA
Research Location						
Asia	5	272	.69	[.46, .91]		> W
Non-Asian Countries	5	1,237	.39	[.27, .51]	5.16*	
Grade at Intervention						
2-5	4	416	.51	[.34, .69]		
6-10	6	1,093	.42	[.29, .55]	0.75	
SES of Participants						
Disadvantaged	4	1,015	.40	[.27, .53]		
Others	6	494	.55	[.37, .72]	1.74	
Method of Group Assignment						
Random	4	811	.44	[.33, .54]		
Quasi Experiment	6	698	.52	[.38, .66]	0.83	
Total Time of Intervention						
5-20 Hours	4	445	.34	[.18, .51]		

Table 3 Contined

Variable	*k*	*N*	*d*	95% CI	Q_B	ANOVA
21-40 Hours	3	579	.28	[.13, .43]		
More than 40 Hours	3	427	.47	[.28, .66]	2.41	
Outcome Measure						
CAT or CogAT	4	1,015	.40	[.27, .53]		
Others	6	494	.55	[.37, .72]	1.74	
Type of Outcome						
General Cognitive Ability	4	1,015	.40	[.27, .53]		
Reasoning Skills	2	148	1.06	[.72, 1.40]		> C & R
Reading comprehension	2	296	.28	[.06, .50]	15.44***	

Note. k = study size; N = number of participants;

CI = confidence interval; Q_B = between-groups test of homogeneity;

ANOVA = significant result.

* $p < .05$, *** $p < .001$

In this study, two of the six moderators revealed statistically significant effects, including research location (two subgroups: western and non-western countries) and outcomes measures (three subgroups: general cognitive ability, reasoning skills and

academic achievement). The tests of homogeneity indicated no statistical differences by grade levels, socio-economic status of participants, methods of group assignments and duration of the intervention. The following is the detailed description of each subgroup analysis.

Grade Level

The included studies were divided into two categories in terms of the grade levels: 2 to 5 (k=4) and 6 to 10 (k=6). As seen in table 3, the average effect size (cohen'd) of studies which recruited grade 2 to 5 students was 0.51, and the average effect size of studies with grade 6 to 10 students was 0.42. Q_B was 0.75. From the results of this moderator analysis, no significant difference was found between effect sizes of studies according to the grade levels of their sample.

Socio-Economic Status of Students

In this sample of studies, two categories for the SES of Participants were present. The first group included students who received free lunch, or classified as "economically disadvantaged" by the local districts. The second group of students are not identified as part of the free-lunch program, or classified as from middle (or upper) class families. No significant heterogeneity in effect sizes was found between the two groups of students.

Study Design

To warrant the rigor of this meta-analysis, the author set up stringent criteria for the inclusion of studies in which only random controlled trials and quasi-experiments were brought in the synthesis. From the moderator analysis, no significant effect size

difference was found between effect sizes of random controlled experiments and quasi-experiments which included in this meta-analysis.

Duration

In this study, the author divided the included literature into three subgroups in terms of the duration of interventions: 5 to 20 hours (k=4), 21 to 40 hours (k=3), and more than 40 hours (k=3). The result showed that none of the duration levels statistically varied from one another. Thus, there was no noteworthy difference between different levels of durations of intervention in the effect of P4C on students' cognitive outcomes.

Outcome Measure: CAT or Non-CAT

Studies included were examined according to their outcome measures. Four studies using Cognitive Ability Tests were accepted as CAT subgroup; six studies applying other types of outcome measures were accepted as Non-CAT subgroup. No significant heterogeneity was found between these two subgroups.

Types of Outcomes

A significant difference between different types of outcomes was found (Q_B = 15.44, $p < .001$). The studies which tested the improvement of reasoning skills through P4C yielded the largest estimations (d=1.06). While P4C used in improving general cognitive abilities (d=0.40) and reading comprehension ability (d=0.28) yielded positive but smaller estimations. This suggests that P4C has significant, positive influence on students' reasoning skills, and moderate effects on general cognitive ability and comprehension ability.

Research Location

This meta-analysis covers 5 Western studies and 5 non-Western studies. Five of the examined studies conducted in non-Western countries: Iran, Turkey, and China. The other five studies come from Western countries: United Kingdom, Australia and United States. A significant difference between the two groups was found in the moderator analysis ($Q = 5.16$, $p < .05$). The studies in non-Western countries had higher effect sizes ($d=0.69$) than studies conducted in Western countries ($d=0.39$).

Summary

The first research question in this meta-analysis concerned the direction and magnitude of the effectiveness of P4C on students' cognitive ability. The studies analyzed here showed a positive, moderator influence on cognitive outcomes in general.

The second question was whether and how the effectiveness of P4C differed significantly depending on the moderator variables. The moderator analysis found statistically significant results in regard to the location and outcome measures of these studies. No significant differences were found as to different grade level, socio-economic status of participants, different methods of group assignment, durations of intervention, and between cognitive ability tests and non-cognitive ability test.

The results suggest that Philosophy for Children has in general a positive moderate influence on students' cognitive outcomes, and significant positive impact on students' reasoning skills.

CHAPTER V

DISCUSSION

Ten studies have been included in this meta-analysis to determine the effects of Philosophy for Children program on students' cognitive abilities, and what characteristics of the intervention, students and outcomes measures could influence the magnitude and direction of such effect.

The Overall Effectiveness of P4C

According to the findings of this meta-analysis, in general, the Philosophy for Children program has shown a moderate, positive influence on students' cognitive outcomes. This result corroborates the previous literature on the program that states that P4C has a positive impact on students' various types of cognitive abilities (Fair et al., 2015a; Fair et al., 2015b; García-Moriyón et al., 2005; Topping & Trickey, 2007a, 2007b; Trickey & Topping, 2004).

The cognitive outcomes comprise general cognitive ability, reasoning skills, creative thinking abilities, educational progress in science, reading and listening comprehension abilities. Among all of these types of cognitive outcomes, the Philosophy for Children program has significant positive effect on students' reasoning skills, while moderate influences on other cognitive domains. The previous P4C meta-analysis that focused on reasoning abilities (García-Moriyón et al., 2005) also indicated the positive impact of P4C on students' reasoning skills.

Discussions about Findings between P4C and Students' Grade Levels

As stated in the results section, no statistically significant difference was found between the effectiveness of P4C on cognitive outcomes and grade levels of students. This result sheds lights on the question regarding P4C and students' age. Philosophy education is traditionally assumed as a discipline that is only appropriate for students no younger than secondary school age (Lipman & Sharp, 1978). But this moderate analysis of this meta-analysis indicates that both studies with grade 2 to 5 students and studies which included grade 6 to 10 students benefited from this program (grade 2-5: d=0.51; grade 6-10: d=0.42). There was no statistically meaningful difference between the effect sizes of the two subgroups.

In recent years, there are new studies (Dasí et al., 2013; Säre et al., 2016) that practiced P4C with very young children who are below the age of 6. For example, the study conducted by Dasi et al. (2013) showed a clear significant improvement in socio-psychological abilities among the 5-year-old children and a partial improvement in the 4-year-old children after participating a few sessions of P4C program. These studies provide information for educators and researchers to understand the unfamiliar area in which young children are involved in rather than excluded from philosophy.

Discussions about Findings between P4C and Duration of Interventions

The moderator analysis showed that the P4C's influence was not moderated by the duration of the intervention. This was not expected. Several studies (Fair et al., 2015a; Fair et al., 2015b; García-Moriyón et al., 2005; Topping & Trickey, 2007a) has

proposed that P4C should be implemented through a significant period of time before the program shows results.

In one study (Fair et al., 2015b), the researchers replicated a previous experiment conducted by Topping and Trickey (Topping & Trickey, 2007a), in which they shortened the duration of the P4C intervention to less than half of the former one: from 58 weeks to 22 to 26 weeks. The result showed that P4C still had a moderate effect on students' general cognitive ability. The phenomenon in this study overlaps with the results of moderator analysis in the current meta-analysis.

This suggests that a short time of exposure to P4C may also have a meaningful impact on students' cognitive outcomes. The practice of P4C should not be limited only within the realm of long-term implementation.

Discussions about Findings between P4C and Locations

Another significant finding of this study is the richness of literature from non-western countries. This suggests that in the recent years, the Philosophy for Children program has spread globally. Statistically significant difference was found between the effect sizes of studies in western (d = 0.39) and non-western (d=0.59) countries.

This is not expected. There are several possible accounts for this phenomenon.

First, the studies in Asia have smaller sample sizes. Because P4C is not familiar to educators and researchers in those countries (Lam, 2012; Marashi, 2008), including Iran, China and Turkey, the studies are often pilot studies with small sample sizes. On the countries, since P4C is initiated in the United States in the 1970s (Brandt, 1988), it is

a more relatively more well-known program. Thus, the studies in the United States, United Kingdom and Australia tended to evaluate P4C in a large school district. The mean sample size of Western studies is three times higher than the mean sample size of non-Western studies. Smaller sample sizes may contribute to the quality of teacher education and P4C implementation.

Another possible explanation is that several studies in non-Western countries tested the improvement of reasoning skills among students (Lam, 2012; Marashi, 2008; Othman & Hashim, 2006), while no included studies in Western countries specifically examined the reasoning abilities of students. According to the moderator analysis in this study regarding the effect sizes of studies with different types of outcomes, there is a statistically significant difference between reasoning skills and other types of outcomes. If P4C is more effective to the improvement of reasoning skills, then the discrepancy between the effect sizes in Western and non-Western studies is understandable.

Discussions about the Excluded Literature

A huge amount of studies were excluded in the process of the analysis. The first reason is because the majority of studies in the field of P4C are qualitative and theoretical, whereas the methodology employed in this study is a quantitative meta-analysis that needs to extract the data from many independent studies conducted on a particular topic and performs statistical analysis (Çoğaltay & Karadağ, 2016). For example, numerous insightful articles regarding P4C have been published in Africa (Di

Masi & Santi, 2016; Ndofirepi & Cross, 2015), but none of them was quantitative and could be used in this study.

The second reason is that even if some studies utilized a quantitative methodology, they were often lack of sufficient information especially for the means and standard deviations for the researcher to compute effect sizes.

During the exclusion of studies, the author was surprised by the studies in P4C has a huge age range. Two studies gave a novel practice and detailed observation of children who are below the age of five. To narrow down the age range to k-12 education, these two studies were excluded but definitely show the potential of teaching and introducing philosophy to very young children.

There is one study that compares the P4C with other thinking programs. The quantitative studies on the comparison between P4C and other critical thinking programs are too limited to synthesize.

Third, the result of exhaustive literature search and process of study inclusion/exclusion showed that more rigorous quantitative studies regarding P4C program are still in need. The researchers gathered over 1180 studies at first, after coding procedure, there were only 17 studies remained. Throughout the data analysis process, seven more articles were excluded from the study. The main reason of this phenomenon is that the majority of the literature regarding P4C are qualitative and theoretical studies. While due to the nature of meta-analysis, which is a quantitative synthesis study, it cannot process and analyze qualitative and non-empirical literature (Lipsey & Wilson, 2001). The second reason is that there is not only a paucity for

quantitative experiments in P4C, but also the rigor of such studies are in need to be improved. It is constant that the data are not sufficient enough for computing an effect size. Thus, this suggests that this field needs more studies to form a larger cluster of rigorous research.

CHAPTER VI

CONCLUSION

Since P4C appears to have a moderate, positive effect on students' cognitive outcomes. In particular, P4C has a significant, positive effect on reasoning skills. The author suggests that P4C may be considered as an effective thinking program for teachers in K-12 education.

Based on the findings of this meta-analysis, several recommendations and suggestions for future research are advanced:

First, addition to long-term implementation of P4C in classroom, a short time of exposure to P4C may also have meaningful impact on students' cognitive outcomes. The author suggested that the practice of P4C should not be limited only within the realm of long-term applications.

Second, this study suggests that grade level is not a moderator of the effectiveness P4C on students' cognitive abilities. Moreover, a small number of studies (Dasí et al., 2013; Säre et al., 2016) have practiced P4C with very young children who are below the age of five. Thus, age should not be the sole reason for the excluding students from philosophy education, and more studies are needed in terms of the impacts of P4C on very young children. As Lone and Burroughs have said (2016), at one time or another we all ask philosophical questions of some kind, consider our values and reflect on the rightness and wrongness of our actions. It is possible that all children, regardless

of age and grade level, have the capacity and interests in engaging philosophical activities (Lipman, 2009).

Third, there is a presupposition of P4C that assumes philosophy education should be assigned to "brighter" students or students who are from a particular advantageous background (Lipman & Sharp, 1978). However, this study indicates that students from different social background have no statistically meaningful difference in the extent to which they are bettered through P4C. It is suggested that educators in P4C program should strive to build a community of inquiry that encourages students to share not only divergent social backgrounds and life experiences (Lipman, 2009) but also different styles of thinking (Lipman & Sharp, 1978) so as to involve them in the classroom discussion.

Fourth, one study in this meta-analysis was excluded because it measured the outcomes based on spiritual development(Abaspour et al., 2015). It is recommended that further discussion on the relationship between philosophy and religion could be initiated in the future.

Fifth, relevant parties need to be aware of the limitations in this meta-analysis where a significant amount of literature was excluded from this study. This is because this quantitative meta-analysis by nature was in need of extracting and analyzing data from literature. Thus, only quantitative studies can be included. This leads to the overlook of the rich qualitative and theoretical literature within the realm of P4C. For example, there are many insightful studies regarding the practice and theoretical

interrogation of P4C in Africa (Di Masi & Santi, 2016; Ndofirepi & Cross, 2015), which could not be added in this study.

Sixth, different from previous meta-analyses, this study emphasized the exhaustive search for P4C studies around the world. The results show that during recent years, a considerable amount of practices have been taken in various continents. This study calls for more research that consider the nuances and details of P4C practices in different cultural, social, educational and philosophical contexts.

Seventh, limited research on socio-psychological outcomes have been generated in this field. P4C as a famous thinking skill program has been relatively examined (Daniel & Auriac, 2011; Fair et al., 2015a; Fair et al., 2015b; García-Moriyón et al., 2005; Säre et al., 2016; Topping & Trickey, 2007a, 2007b). However, the affective outcomes are happened to be more philosophically informed ones (Love, 2016). Thus, more studies on the socio-psychological dimensions in this field are strongly suggested.

In the end, considering the number of studies included in this meta-analysis, more quantitative research regarding P4C is needed. Especially, researchers should report the means, standard deviations, durations of intervention, and detailed information regarding students' background so that their studies will not stand alone as single research finding (Çoğaltay & Karadağ, 2016).

REFERENCES

Abaspour, N., Nowrozi, R. A., & Latifi, Z. (2015). Investigating the Effect of Educating Philosophy in the Children on the Spiritual Development of Female Students with 12-14 Years Old in the City of Isfahan. *Journal of Education and Practice, 6*(11), 162-166.

Abbasi, Z., & Ajam, A. A. (2016). The Effects of Philosophical Stories on Emotional Intelligence and Educational Progress of Students in Science Lessons. *Mediterranean Journal of Social Sciences, 7*(2), 282.

Brandt, R. (1988). On Philosophy in the Curriculum: A Conversation with Matthew Lipman. *Educational Leadership, 46*(1), 34-37.

Butnor, A. (2012). Critical Communities: Intellectual Safety and the Power of Disagreement. *Educational Perspectives, 44*, 29-31.

Çoğaltay, N., & Karadağ, E. (2016). The Effect of Educational Leadership on Organizational Variables: A Meta–Analysis Study in the Sample of Turkey. *Educational Sciences: Theory & Practice, 16*(2).

Colom, R., Moriyon, F. G., Magro, C., & Morilla, E. (2014). The Long-term Impact of Philosophy for Children: A Longitudinal Study (Preliminary Results). *Analytic Teaching and Philosophical Praxis, 35*(1).

Daniel, M., & Auriac, E. (2011). Philosophy, Critical Thinking and Philosophy for Children. *Educational Philosophy and Theory, 43*(5), 415-435.

Dasí, M. G., 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.

Dewey, J. (1916). *Democracy and Education: An Introduction to Philisophy of Education.* Macmillan.

Dewey, J. (1986). *Experience and Education.* Paper Presented at the The Educational Forum.

Di Masi, D., & Santi, M. (2016). Learning Democratic Thinking: A Curriculum to Philosophy for Children as Citizens. *Journal of Curriculum Studies, 48*(1), 136-150.

Fair, F., Haas, L. E., Gardosik, C., Johnson, D., Price, D., & Leipnik, O. (2015a). Socrates in the Schools: Gains at Three-year Follow-up. *Journal of Philosophy in Schools, 2*(2).

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

García-Moriyón, F., Rebollo, I., & Colom, R. (2005). Evaluating Philosophy for Children. *Thinking: The journal of philosophy for children, 17*(4), 14-22.

Gopnik, A. (2009). *The Philosophical Baby: What Children's Minds Tell us About Truth, Love & the Meaning of Life*. Random House.

Gorard, S., Siddiqui, N., & Huat See, B. (2015). Philosophy for Children: Evaluation Report and Executive Summary. *Education Endowment Foundation, Millbank, UK*.

Gregory, M. (2011). Philosophy for Children and Its Critics: A Mendham Dialogue. *Journal of Philosophy of Education, 45*(2), 199-219.

Hidi, S., Renninger, K. A., & Krapp, A. (2004). Interest, A Motivational Variable That Combines Affective and Cognitive Functioning. *Motivation, Emotion, and Cognition: Integrative Perspectives on Intellectual Functioning and Development*, 89-115.

Jackson, T. (2001). The Art and Craft of "Gently Socratic" Inquiry. *Developing Minds: A Resource Book for Teaching Thinking, 3*.

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.

Lipman, M. (2004). Philosophy for Children's Debt to Dewey. *Critical and Creative Thinking, 12*(1), 1-8.

Lipman, M. (2009). (USA) Philosophy for Children: Some Assumptions and Implications. *Children Philosophize Worldwide: Theoretical and Practical Concepts, 9*, 23.

Lipman, M., & Sharp, A. M. (1978). Some Educational Presuppositions of Philosophy for Children. *Oxford Review of Education, 4*(1), 85-90.

Lipman, M., Sharp, A. M., & Oscanyan, F. S. (2010). *Philosophy in the Classroom*. Temple University Press.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-analysis* (Vol. 49). Sage Publications Thousand Oaks, CA.

Lone, J. M., & Burroughs, M. D. (2016). *Philosophy in Education: Questioning and Dialogue in Schools*. Rowman & Littlefield.

Love, R. (2016). The Case for Philosophy For Children in the English Primary Curriculum. *Analytic Teaching and Philosophical Praxis, 36*(1).

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

Matthews, G. B. (1980). *Philosophy and the Young Child*. Harvard University Press.

Matthews, G. B. (1994). *The Philosophy of Childhood*. Harvard University Press.

Ndofirepi, A., & Cross, M. (2015). Child's Voice, Child's Right: Is Philosophy for Children in Africa the Answer? *Interchange, 46*(3), 225-238.

Nia, A. T. (2014). Investigate the Effect the Philosophy for Children Program (p4c) on Reducing Trait Anger in Teens. *Stud, 4*(2), 449-455.

Othman, M., & Hashim, R. (2006). Critical Thinking & Reading Skills. *Thinking: The Journal of Philosophy for Children, 18*(2), 26-34.

Peirce, C. S. (1868). Some Consequences of Four Incapacities. *The Journal of Speculative Philosophy, 2*(3), 140-157.

Peirce, C. S., & Houser, N. (1998). *The Essential Peirce: Selected Philosophical Writings* (Vol. 2). Indiana University Press.

Piaget, J. (1931). Children's Philosophies. *A Handbook of Child Psychology,* 377-391.

Rescher, N., Shusterman, R., Alcoff, L. M., Code, L., Harding, S., On, B.-A. B., . . . Wartenberg, T. E. (2002). *The Philosophical I: Personal Reflections on Life in Philosophy*: Rowman & Littlefield Publishers.

Säre, E., Luik, P., & Tulviste, T. (2016). Improving Preschoolers' Reasoning Skills Using the Philosophy for Children Programme. *Trames: A Journal of the Humanities and Social Sciences, 20*(3), 273.

Scholes, L., Lunn Brownlee, J., Walker, S., Johansson, E., Lawson, V., & Mascadri, J. (2016). Promoting Social Inclusion in the Early Years of Elementary School: A Focus on Children's Epistemic Beliefs for Moral Reasoning. *International Journal of Inclusive Education*, 1-14.

Sharp, A. M., Reed, R. F., & Lipman, M. (2010). *Studies in Philosophy for Children: Harry Stottlemeier's Discovery*. Temple University Press.

Tok, Ş., & Mazı, A. (2015). The Effect of Stories for Thinking on Reading and Listening Comprehension: A Case Study in Turkey. *Research in Education, 93*(1), 1-18.

Topping, K. J., & Trickey, S. (2007a). Collaborative Philosophical Enquiry for School Children: Cognitive Effects at 10–12 Years. *British Journal of Educational Psychology, 77*(2), 271-288.

Topping, K. J., & Trickey, S. (2007b). Collaborative Philosophical Inquiry for Schoolchildren: Cognitive Gains at 2-year Follow-up. *British Journal of Educational Psychology, 77*(4), 787-796.

Trickey, S., & Topping, K. J. (2004). 'Philosophy for Children': A Systematic Review. *Research Papers in Education, 19*(3), 365-380.

van der Straten Waillet, N., Roskam, I., & Possoz, C. (2015). On the Epistemological Features Promoted by 'Philosophy for Children'and Their Psychological Advantages When Incorporated into RE. *British Journal of Religious Education, 37*(3), 273-292.

VanVoorhis, C. W., & Morgan, B. L. (2007). Understanding Power and Rules of Thumb for Determining Sample Sizes. *Tutorials in Quantitative Methods for Psychology, 3*(2), 43-50.

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

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* (Doctoral Dissertation). Queensland University of Technology, Australia.