

The Tools of the Mind curriculum for improving self-regulation in early childhood: a systematic review

Alex Baron, Maria Evangelou, Lars-Erik Malmberg, G.J. Melendez-Torres

A Campbell Systematic Review
2017:10

Published: October 2017
Search executed: December 2016



The Campbell Library comprises:

- Systematic reviews (titles, protocols and reviews)
- Policies and Guidelines Series
- Methods Series

Go to the library to download these resources, at:

www.campbellcollaboration.org/library/

Better evidence for a better world



Colophon

Title	The Tools of the Mind curriculum for improving self-regulation in early childhood: a systematic review
Institution	The Campbell Collaboration
Authors	Baron, Alex Evangelou, Maria Malmberg, Lars-Erik Melendez-Torres, G.J.
DOI	10.4073/csr.2017.10
No. of pages	77
Last updated	12 October 2017

Citation	Baron A, Evangelou M, Malmberg L.E., Melendez-Torres, G.J. The Tools of the Mind curriculum for improving self-regulation in early childhood. Campbell Systematic Reviews 2017 DOI: 10.4073/csr.2017.10
ISSN	1891-1803
Copyright	© Baron et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Roles and responsibilities	Alex Baron and Maria Evangelou served as the early childhood education content area experts for this review. G.J. Melendez-Torres served as the systematic review methods expert; he and Alex Baron collectively created the data extraction form and performed coding for the included studies. Lars-Erik Malmberg served as the statistical analysis expert. Finally, Alex Baron will be responsible for updating this review as additional evidence accumulates and as funding becomes available.
Editors for this review	Editor: Sarah Miller Managing editor: Carlton Fong
Declarations of interest	The authors have no vested interest in the outcomes of this review, nor any incentive to represent findings in a biased manner.

Corresponding author	Alex Baron Department of Education/ Social Sciences Division University of Oxford 15 Norham Gardens Oxford, Oxfordshire OX2 6PY United Kingdom E-mail: alexander.baron@education.ox.ac.uk
-----------------------------	---

Full list of author information is available at the end of the article.

Campbell Systematic Reviews

Editor-in-Chief Vivian Welch, University of Ottawa, Canada

Editors

Crime and Justice David B. Wilson, George Mason University, USA
Charlotte Gill, George Mason University, USA
Angela Higginson, Queensland University of Technology, Australia

Disability Carlton J. Fong, Texas State University, USA

Education Sarah Miller, Queen's University Belfast, UK

International Birte Snilstveit, 3ie, UK

Development Hugh Waddington, 3ie, UK

Social Welfare Brandy Maynard, Saint Louis University, USA

*Knowledge Translation
and Implementation* Aron Shlonsky, University of Melbourne, Australia

Methods Therese Pigott, Loyola University, USA
Ryan Williams, AIR, USA

Managing Editor Chui Hsia Yong, The Campbell Collaboration

Co-Chairs

Crime and Justice David B. Wilson, George Mason University, USA
Peter Neyroud, Cambridge University, UK

Disability Oliver Wendt, Purdue University, USA
Joann Starks, AIR, USA

Education Sarah Miller, Queen's University Belfast, UK
Gary W. Ritter, University of Arkansas, USA

Social Welfare Brandy Maynard, Saint Louis University, USA

*Knowledge Translation
and Implementation* Robyn Mildon, CEI, Australia
Cindy Cai, AIR, USA

International Peter Tugwell, University of Ottawa, Canada
Development Hugh Waddington, 3ie, UK

Methods Ariel Aloe, University of Iowa, USA

The Campbell Collaboration was founded on the principle that systematic reviews on the effects of interventions will inform and help improve policy and services. Campbell offers editorial and methodological support to review authors throughout the process of producing a systematic review. A number of Campbell's editors, librarians, methodologists and external peer reviewers contribute.

The Campbell Collaboration
P.O. Box 4404 Nydalen
0403 Oslo, Norway
www.campbellcollaboration.org

Table of contents

TABLE OF CONTENTS	2
PLAIN LANGUAGE SUMMARY	4
EXECUTIVE SUMMARY/ABSTRACT	6
Background	6
Objectives	6
Search methods	6
Selection criteria	6
Data collection and analysis	6
Results	7
Authors' conclusions	7
1 BACKGROUND	8
1.1 The issue	8
1.2 The intervention	10
1.3 How the Tools program might improve children's self-regulation	11
1.4 Why it is important to do the review	13
2 OBJECTIVES	14
3 METHODS	15
3.1 Criteria for considering studies for this review	15
3.2 Search methods for identification of studies	17
3.3 Data collection and analysis	18
4 RESULTS	23
4.1 Description of studies	23
4.2 Risk of bias in included studies	24
4.3 Quantitative synthesis of results	27
5 DISCUSSION	29
5.1 Summary of main results	29
5.2 Quality of the evidence	30
5.3 Limitations and potential biases in the review process	31
5.4 Agreements and disagreements with other studies or reviews	31

6	AUTHORS' CONCLUSIONS	33
6.1	Implications for practice and policy	33
6.2	Implications for research	34
7	REFERENCES	36
7.1	References to included studies	36
7.2	References to excluded studies	36
7.3	Additional references	37
8	INFORMATION ABOUT THIS REVIEW	44
8.1	Review authors	44
8.2	Roles and responsibilities	45
8.3	Sources of support	46
8.4	Declarations of Interest	46
8.5	Plans for Updating the Review	46
8.6	Author Declaration	46
9	TABLES	48
	Characteristics of included studies	49
	risk of bias for included studies	54
	Characteristics of excluded studies	58
10	FIGURES	60
11	DATA AND ANALYSES	62
12	SPSS AND R CODE FOR ANALYSES	66
13	SEARCH NOTES FROM DATABASES	69
14	CODING MANUAL	71

Plain language summary

The Tools of the Mind curriculum improves self-regulation and academic skills in early childhood

The Tools of the Mind early childhood curriculum appear to improve children's self-regulation and academic skills. The assessment of the tools curriculum is hampered by a lack of rigorous evidence and more research is necessary to corroborate this finding.

What did the review study?

Tools of the Mind (Tools) is an early childhood education curriculum, which involves structured make-believe play scenarios and a series of other curricular activities.

Tools aims to promote and improve children's self-regulation and academic skills by having a dual focus on self-regulation and other social-emotional skills in educational contexts. This review examines the evidence on the effectiveness of Tools in promoting children's self-regulation and academic skills, in order to inform its implementation in schools.

What is the aim of this review?

This Campbell systematic review examines the evidence on the effectiveness of the Tools of the Mind curriculum in promoting children's self-regulation and academic skills, in order to inform its implementation in schools. The participants included students of all ages, gender, ethnicity, special education status, language-learning status, and socio-economic status. The review summarizes findings from 14 records across six studies conducted in the USA.

What studies are included?

Included studies had to have used randomized controlled trials or quasi-experimental studies and reported on one or more quantitative effect sizes regarding tools' effectiveness in self-regulatory or academic domains.

A total of 14 records across six studies were included in the review. The participants included students of all ages, gender, ethnicity, special education status, language learning status, and socio-economic status. The studies included measured at least one of four primary outcomes and did not measure any secondary outcome. Studies that compared Tools with a business-as-usual or another intervention were included in the review.

All included studies were conducted in the USA.

What are the main results of the review?

The Tools curriculum significantly improved children's math skills relative to comparison curricula, but the effect size was small. There are also shortcomings in the quality of evidence.

Although the average effect sizes for self-regulation and literacy favored tools compared to other approaches, the effect was not statistically significant. The evidence from the small number of included studies is mostly consistent with the evidence observed for other similar programs, but again the evidence is weak.

The results for the outcome measures were not statistically significant.

What do the findings of this review mean?

Generally, the Tools curriculum seems to improve children's self-regulation and academic skills. However, given the small number of included studies, as well as other methodological shortcomings, such as the high risk of bias in some of the included studies, this conclusion should be read with caution.

While there is doubt as to the validity of the findings, tools' educational approach seems to be consistent with many child developmental theories and as such, should not be ruled out. There is a need to conduct more high quality research, especially about studies focused on demonstrating tools' effectiveness in promoting children's self-regulation skills.

How up-to-date is this review?

The review authors searched for studies published up to December 2016. This Campbell Systematic Review was published in October 2017.

Executive Summary/Abstract

BACKGROUND

Tools of the Mind (Tools) is an early childhood education curriculum that aims to simultaneously promote children's self-regulation and academic skills. Given the increasing focus on self-regulation and other social-emotional skills in educational contexts, Tools has become increasingly implemented in classrooms around the United States, Canada, and Chile. Despite its growing popularity, Tools' evidence base remains mixed.

OBJECTIVES

The aim of this review is to synthesize the evidence on the effectiveness of the Tools program in promoting children's self-regulation and academic skills.

SEARCH METHODS

The systematic search was conducted from October 21 through December 3, 2016. The search yielded 176 titles and abstracts, 25 of them deemed potentially relevant. After full-text screening, 14 reports from six studies were eligible for inclusion.

SELECTION CRITERIA

In order to be included, a study must have had one or more quantitative effect sizes regarding Tools' effectiveness in the self-regulatory or academic domains. Moreover, the study must have employed statistical mechanisms to control for potential confounds.

Studies that compared Tools with a business-as-usual or another intervention were eligible for inclusion, whereas studies that did not pertain to the Tools curriculum were excluded. The reports, whether published or unpublished, could come from any national context, language, student population, or time period as long as the conditions outlined above were met.

DATA COLLECTION AND ANALYSIS

All included studies classified as randomized controlled trials, though, again, quasi-experimental studies had been eligible for inclusion. Each included study yielded effect sizes in the form of standardized mean differences. The outcomes of interest included assessor-reported self-regulation skills (e.g., teachers or parents rating children's self-regulation), task-based self-regulation skills (e.g., children performing a self-regulation task on a computer and receiving a score), literacy skills, and math skills. All effect sizes were interpreted as Tools' effect relative to other business-as-usual programs or other interventions.

RESULTS

The evidence indicated statistically significant benefits for Tools children on the math pooled effect size. The other pooled effect sizes for self-regulation and literacy favored Tools but did not reach statistical significance.

AUTHORS' CONCLUSIONS

The results indicate positive yet small effects for the Tools program. Three of the four pooled effect sizes did not reach statistical significance, but all four pooled effect sizes favored Tools. The small number of included studies reduced power, which could explain the lack of statistical significance across three of the four outcome measures. By contrast, it is also possible that Tools either does not substantially influence children's self-regulation or that the influence is too small to be detected with the current evidence base.

1 Background

1.1 THE ISSUE

1.1.1 Background on self-regulation

Self-regulation, defined as volitional control of attention, behavior, and executive functions for the purposes of goal-directed action (Blair & Ursache, 2011), is associated with multiple school-related outcomes (Calkins, S. D., Howse, R. B., & Philippot, 2004; A. Diamond & Lee, 2011; McClelland & Tominey, 2011). Children with robust self-regulation have been shown to more cooperatively participate in classroom activities (Fisher, Hirsh-Pasek, Newcombe, & Golinkoff, 2013; Ramani, 2012), sustain focus on tasks (K. L. Bierman, Nix, & Greenberg, 2008; Drake, Belsky, & Fearon, 2014), and exhibit reduced behavioral issues (Feng et al., 2008; Ponitz, McClelland, Matthews, & Morrison, 2009).

Conversely, lower levels of self-regulation skills are associated with externalizing behaviors (Flouri, Midouhas, & Joshi, 2014; Olson & Lunkenheimer, 2009), diminished attention (Raver et al., 2011; Tough, 2012), and lower academic achievement (Kim, Nordling, Yoon, & Kochanska, 2014; Nota, Soresi, & Zimmerman, 2004; Soares, Vannest, & Harrison, 2009). In addition to problems during the schooling years, children with poor self-regulatory competencies are more likely to have worse health and financial outcomes in adulthood (Moffitt, Arseneault, & Caspi, 2011; Schlam, Wilson, Shoda, & Mischel, 2013).

Previous studies demonstrate that self-regulation is amenable to improvement (Barnett et al., 2008; Diamond, Barnett, Thomas, & Munro, 2007; Nunes et al., 2007) as well as deterioration (Karreman, Van Tuijl, & Marcel, 2006; Raver, Blair, & Willoughby, 2013). Consequently, it is crucial to identify education practices that foster self-regulation growth, which emerges as the research rationale of this review.

1.1.2 Self-regulation development in educational contexts

Given the role of self-regulation in promoting both child and adult outcomes, early intervention in preschool contexts holds considerable promise for improving a child's development trajectory. As Heckman noted, early "skill begets skill; learning begets learning" (Heckman & Masterov, 2007, p. 3). Consequently, small self-regulatory differences in early childhood can be magnified to progressively larger differences over time (Alexander,

Entwistle, & Kabbani, 2001; O'Shaughnessy, Lane, Gresham, & Beebe-Frankenberger, 2003). Thus, early childhood emerges as an especially critical period in which to intervene.

Research about the challenges of self-regulation promotion further underscores the need for early interventions. A nationally representative survey indicated that 46% of American kindergarten teachers reported at least half of their students as routinely struggling with self-regulation (Rimm-Kaufman, Pianta, & Cox, 2000). In fact, American preschool students are three times more likely to be expelled for unmanageable behavior than primary and secondary students (Gilliam, 2005). Based on these statistics, it seems that many early childhood educational settings are neither meeting children's needs nor effectively promoting children's self-regulation.

Certain subpopulations of children face unique self-regulation challenges from a young age. Children growing up in poverty are more likely to experience self-regulatory problems (Raver et al., 2013; Raver, 2012), which make low-income children susceptible to disciplinary action (Alloway, Lawrence, & Rodger, 2013; Miller, Nevado-Montenegro, & Hinshaw, 2012). For example, a Washington DC report (Office of the State Superintendent of Education, 2013) revealed that students aged three and four received 181 suspensions during the 2012-2013 school year, most of which occurred for students in low-income schools.

Moreover, many children have been diagnosed with chronic regulatory deficits such as Attention Deficit Hyperactivity Disorder (ADHD) and conduct disorder. In 2013, 11% of American children between the ages of 4 and 17 had been diagnosed with ADHD, which reflects a 41% increase in diagnoses over a single decade (Center for Disease Control, 2013). In the UK, 7% of British boys and 3% of British girls aged 5 to 10 meet the diagnostic criteria for conduct disorder (National Institute for Health and Care Excellence, 2013), which presents challenges to the educators responsible for student learning (Webster-Stratton, Reid, & Stoolmiller, 2009).

Of course, the observed increase in children's issues in recent years likely does not mean that modern children have less self-regulation than did their parents; rather, the increase is likely a product of systemic changes in the way that self-regulation issues have been defined, measured, and diagnosed. Nevertheless, given the benefits of robust self-regulation skills for children and for the adults they will become, it is important to identify educational methods that cultivate all children's self-regulation.

In recent years, the number of self-regulation interventions has increased alongside the rising concerns regarding children's self-regulation issues (Harris, Friedlander, & Graham, 2005; Soares et al., 2009; Thompson, Ruhr, Maynard, Pelts, & Bowen, 2013), especially for children with special educational needs (Gulchak, 2008; K. Jones, Daley, Hutchings, Bywater, & Eames, 2007). Despite the growing number of interventions aiming to improve children's self-regulation, the body of evidence in respect of their effectiveness is sparse.

For example, the U.S. Department of Education's Institute of Educational Sciences (IES) funded a randomized controlled trial that assessed 14 preschool curricula; the results indicated that none of the curricula significantly improved children's self-regulation skills beyond traditional comparator curricula (Preschool Curriculum Evaluation Research Consortium, 2008). Moreover, none of the 14 programs identified self-regulation promotion as its primary curricular focus, despite abundant research indicating the benefits of self-regulation for children.

To the best of our knowledge, only one early childhood curriculum emphasizes self-regulation cultivation as its paramount aim: *Tools of the Mind* (Tools). Since its development in 1993, Tools has been adopted in parts of the United States, Canada, and South America. Twenty U.S. states now have at least one Tools school; in certain areas such as Washington DC, Tools has been implemented in the majority of local preschools (District of Columbia Public Schools, 2016).

In the face of the program's proliferation, it is important to establish evidence of Tools' effectiveness on hypothesized outcomes. That is, does Tools enhance children's self-regulation and academic outcomes as compared with traditional 'business-as-usual' or other interventions? This review aims to be the first to directly address this question.

1.2 THE INTERVENTION

1.2.1 Tools of the Mind (Tools)

Tools derives from the work of psychologist Lev Vygotsky. In his book *Thought and Language* (1962), Vygotsky develops the concept of 'mental tools,' which extend mental faculties in the way that physical tools extend physical faculties. For example, although young children typically struggle with task focus, they can be taught to use private speech (e.g., self-talk meant to guide one's actions as opposed to communicate with others) in order to maintain concentration amid distractions. In this case, private speech serves as a mental tool that enables children to focus beyond their baseline abilities (Vygotsky, 1962).

Thus, Vygotsky's developmental theory is central in Tools' approach. According to the curricular developers, Tools is informed by "is inspired by the work of the Russian psychologist Lev Vygotsky and his students, and at the same time, is rooted in cutting edge neuropsychological research on the development of self-regulation/executive functions in children" (Bodrova & Leong, 2015, Tools website home page). Unlike several self-regulation interventions, which often involve individualized plans for specific children (Gulchak, 2008; Soares et al., 2009) or a set of exercises to supplement an existing curriculum (K. L. Bierman, Domitrovich, Blair, Nelson, & Gill, 2008; Domitrovich, Cortes, & Greenberg, 2007), Tools is intended to be a comprehensive curriculum delivered to all students.

Tools is centered around make-believe play as a mechanism to improve children's self-regulation. In the words of Vygotsky (1978), "in play the child is always behaving beyond his age, above his usual everyday behavior. In play he is, as it were, a head taller than himself" (p. 74). In addition to Vygotsky's assertions regarding play's potential for promoting child outcomes, his contemporaries such as Piaget remarked that "play is the answer to how anything new comes about" (Piaget, 1951, p. 72), whereas Elkonin and Zaporozhets vigorously argued (Elkonin & Zaporozhets, 1978) for the expansion of pretend play in early childhood contexts.

As for the connection between play and self-regulation, the Tools developers (Bodrova & Leong, 2007) assert that play scenarios require children to 1) remember their make-believe role and act it out (*working memory*), 2) inhibit the impulse to arbitrarily switch roles (*inhibitory control*), and 3) flexibly switch between their personalities as individuals versus the personalities of the role they have assumed (*cognitive flexibility*). The following section will now describe how Tools aims to integrate self-regulation promotion into all parts of curriculum.

1.3 HOW THE TOOLS PROGRAM MIGHT IMPROVE CHILDREN'S SELF-REGULATION

Tools' theory of change contains three elements: 1) the teacher regulates the students, 2) the students regulate one another, and 3) the students self-regulate (Bodrova & Leong, 2007). That is, a child's ability to regulate his or her internal thoughts and actions must begin with someone outside of the child (i.e., an adult or more competent peer) who first regulates the child's behavior. When the students first arrive in a classroom, Vygotsky wrote (1962) that they are "slaves to their environment," and education's aim must be to transform them into "masters of their own behavior" (p. 147).

Tools attempts to help children regulate their behavior by integrating self-regulation-oriented activities within academic instruction (Bodrova, Leong, & Akhutina, 2011, p. 18). That is, each Tools activity contains both a target academic skill (e.g., reading a book with a classmate) and a self-regulatory skill (e.g., waiting one's turn to read the book). Overall, Tools includes over 60 activities that simultaneously target students' self-regulation as well as their academic skills. Two such activities are now described below.

1.3.1 Two examples of Tools activities

Two activities, 'buddy reading' and make-believe play scenarios, are emblematic of Tools' approach to learning. Buddy reading involves two students who cooperatively read a book. One child receives a picture of a mouth, which designates him or her as the reader; the other child receives a picture of an ear, which designates him or her as the listener. The reader

then reads the story while the other child actively listens and checks for decoding errors. The children then switch roles after the first reader completes the story (Leong & Bodrova, 2011).

Given proper execution, buddy reading targets both literacy and executive function. Buddy reading hones literacy skills because children read a book, whereas the activity should theoretically hone executive function because children must 1) use *working memory* to remember and act out their roles, 2) demonstrate *cognitive flexibility* by switching across roles, and 3) exhibit *inhibitory control* to suppress desires to switch roles at inappropriate times (e.g., the listener should not attempt to become the reader before his or her turn).

The second activity emblematic of the Tools approach is make-believe play, which is meant to occur every day in Tools classrooms (Bodrova & Leong, 2013). In *Mind and society*, Vygotsky (1978) asserted that children achieve their “greatest self-control in play” (p. 99). This is because pretend play thus requires children to focus on a role (e.g., a grocer), enact that role (e.g., help a ‘customer’ bag groceries), and inhibit the impulse to switch roles (e.g., become the grocery store manager instead of the grocer) even when the child wishes to act spontaneously.

Vygotsky (1933) argued that effective play scenarios require three elements: children must 1) determine an imaginary scenario, 2) negotiate roles for themselves and one another, and 3) act out those roles with fidelity (i.e., not switch or cease a role simply because one has lost interest in it). In order to achieve such structured play scenarios, Tools teachers work with students to create play plans as depicted in figure 1.

As observed in figure 1, the play plan includes both textual and pictorial elements. According to the Tools manual (Leong & Bodrova, 2011), play planning involves multiple steps. First, the teacher convenes a group of students who collectively determine a play scenario. Second, students negotiate roles for each child to assume throughout the play block. For example, in figure 1, the children have decided to enact a scenario involving a princess and prince. Each child then creates a play plan that includes his or her name, a picture of the child acting out that role, and a textual description of the play plan. The plan from figure 1 indicates that the student will pretend to be Sleeping Beauty and marry a prince.

Thus, make-believe play planning simultaneously involves writing practice, drawing practice, and goal-oriented thinking to guide the child’s subsequent behavior. If students forget their roles in the play scenario, then the teacher and/or other students can reference the play plan (Leong & Bodrova, 2011). This play-planning process precedes the actual play scenario, which is where Vygotsky (1933) argues children’s self-control is directly taxed.

1.3.2 Tools summary

In sum, whether children are engaged in literacy, mathematics, or play scenarios, each Tools activity aims to target self-regulation. Tools is designed to be implemented by classroom

teachers throughout a full academic year (Leong & Bodrova, 2011). Moreover, in contrast to programs that target only children with self-regulatory deficits, the Tools curriculum developers argue that self-regulation instruction “should not be reserved only for ‘problem’ children” and that “all children benefit from practicing deliberate and purposive behaviors” (Bodrova & Leong, 2005, p. 35). Thus, Tools’ comprehensive nature emerges as a key mechanism of its purported efficacy in improving children’s self-regulation.

1.4 WHY IT IS IMPORTANT TO DO THE REVIEW

Given self-regulation’s role in promoting a multitude of desirable life outcomes, it is critical to identify educational practices that improve self-regulation skills. The Tools developers claim that the program effectively promotes children’s self-regulation, and Tools has already been implemented in the U.S., Canada, and parts of South America at a cost of \$3000 per classroom in the first year alone (United States Department of Education, 2008).

Although Tools’ proliferation has been consistent in recent years, the findings from Tools evaluation studies have been inconsistent (Blair & Raver, 2014). These mixed findings have thus far precluded any authoritative conclusion regarding the curriculum’s effectiveness. The present review aims to provide education policymakers and practitioners with useful information regarding whether to implement Tools.

2 Objectives

Our central objective was to identify, appraise, and synthesize the available evidence regarding Tools in order to evaluate Tools' effectiveness as compared with other curricula, including business-as-usual and other programs. Our ancillary objective was to examine study and student characteristics that explain observed heterogeneity in effect sizes across trials.

3 Methods

3.1 CRITERIA FOR CONSIDERING STUDIES FOR THIS REVIEW

3.1.1 Types of studies

We were prepared to include studies with experimental or quasi-experimental designs that adequately controlled for potential confounds. Thus, we sought to include studies of the following types:

- *Randomized controlled trial*: Random assignment of participants to treatment and control groups by the researcher, using a reliable method of randomization (e.g., random source allocation)
- *Regression discontinuity*: Researchers assign a threshold or cut-off point (e.g., a birthday cut-off for eligibility into an early childhood program) above or below which the intervention is delivered. Although formal randomization does not occur, comparison of observations lying close to either side of the threshold enables estimation of the treatment effect.
- *Matched control group studies*: Treatment group participants are compared against a matched group of controls who are similar on a set of pre-specified characteristics but do not receive the intervention.
- *Time-series*: Participants are observed before, during, and after the intervention to determine whether it had any effect differentiable from underlying trends over time.
- *Pre- and post-design*: The treatment and control groups, although not randomly assigned, are tested at the beginning and end of the intervention. The pre-test establishes whether significant group differences exist at the study's outset; the post-test reveals whether a significant effect manifests after the treatment has been administered.

Although the review would ideally restrict included studies to randomized trials, randomization of students in education research can be difficult given ethical concerns and school district policies. Thus, this review also aimed to include the quasi-experimental designs described above as long as those studies' designs enabled controlling for potential confounds.

3.1.2 Types of participants

We included data on students of any age, gender, ethnicity, special education status, language learning status, and socio-economic status in this review. This is because Tools is a comprehensive curriculum aimed at students of any background, so any student or classroom that experienced Tools was eligible for inclusion in this review.

3.1.3 Types of interventions

We included any study that analyzed Tools' effect in comparison to one or more business-as-usual curricula or another intervention program. Business-as-usual curricula are those the teacher had used before the intervention study began. Other intervention programs (i.e., newly implemented intervention programs that were not business-as-usual for the participating teacher) were also included in the review as long as they served as a comparison group for the Tools program.

Finally, we included studies where Tools was combined with another program or intervention that was new for the teacher (i.e., if Tools was combined with a math curriculum to create a composite intervention program). Studies that did not pertain to the Tools program were excluded.

3.1.4 Types of outcome measures

Primary outcomes

As indicated in section 1.2, Tools aims to simultaneously cultivate children's self-regulatory and academic skills. Thus, the primary outcome measures target both the self-regulatory and academic domains. To be eligible for inclusion, studies had to include at least one quantitative outcome pertaining to at least one of the four dimensions below:

- **Children's self-regulation as reported by teachers, school administrators, parents, and/or observers:** These subjective reports typically derive from observation periods during which a researcher or teacher rates the child's behavior. For example, parents, teachers, or researchers can fill out the Behavioral Rating Inventory of Executive Function – Preschool (BRIEF-P) rating form (Gioia, Espy, & Isquith, 2005), which has 63 items to assess children's inhibitory control, cognitive flexibility, working memory, and overall executive control.
- **Children's self-regulation as indicated by task-based measures:** These scores derive from children's task performance on an executive function exercise. For example, the "Heads-Toes-Knees-Shoulders" task involves touching the correct body part based on the teacher's instructions, which change after each round. This activity engages all aspects of executive function: 1) working memory (remembering the teacher's directions and acting upon them), 2) cognitive flexibility (switching among the rules as they change during each round, and 3) inhibitory control (not touching

the body part that you hear, but rather the body part that the teacher has previously specified through a rule).

- **Children’s academic skills as captured by various assessments:** Any literacy and math scores on preschool achievement tests were included. All recovered academic and self-regulation data from the included studies derived from standardized assessment instruments.

Secondary outcomes

This review did not include any secondary outcomes.

3.1.5 Duration of follow-up

We included data from any follow-up periods in the original studies. The follow-up data were classified into three categories: short-term (i.e., data taken between the end of the Tools intervention year to five months following the intervention), medium-term (i.e., data taken between six months and 11 months after the end of the Tools intervention), and long-term (i.e., data taken at 12 months or more after the end of the Tools intervention).

3.1.6 Types of settings

We included studies from any setting where Tools was implemented. Because Tools is a school-based curriculum, we expected that our search would yield only school-based studies, which is indeed what we found. Nonetheless, no a priori setting-based exclusion criteria were imposed.

3.2 SEARCH METHODS FOR IDENTIFICATION OF STUDIES

3.2.1 Electronic searches

We systematically queried the set of databases in the bulleted list below. For each database, we used some variant of “Tools of the Mind” as a search term. We aimed to capture every study that mentions Tools at any point in the title, abstract, or text body; thus, a simple search term that includes the program title seemed sensible. For example, in the ERIC database, we used the following search term: AB(“Tools of the Mind”) OR TI(“Tools of the Mind”). The search terms and results for each database are shown in Section 13. The full set of databases we searched is as follows:

- ERIC (ProQuest)
- ProQuest Dissertations and Theses (ProQuest)
- Applied Social Sciences Index and Abstracts (ProQuest)
- Sociological Abstracts (ProQuest)
- Social Sciences Citation Index (ProQuest)

- PsycINFO (Ovid)
- MEDLINE (Ovid)
- Embase (Ovid)
- CENTRAL (Cochrane Library)
- LILACS (<https://lilacs.bvsalud.org/en/>)
- OpenGrey (www.opengrey.eu/)

3.2.2 Searching other resources

In addition to the electronic database searches, we undertook five other strategies to maximize the comprehensiveness of our search:

- We examined the reference lists of relevant primary studies and reviews to identify additional articles.
- We conducted a forward citation search in *Web of Science* using the Tools’ developers’ curricular text (Bodrova & Leong, 2007) as the starting point.
- We hand-searched four journals including *Child Development*, *Early Childhood Research Quarterly*, *Early Childhood Education Journal*, and *Journal of School Psychology*.
- We reviewed four websites of education institutions and agencies including:
 - *Tools of the Mind* website (<https://www.toolsofthemind.org>)
 - What Works Clearinghouse at the Institute of Educational Sciences (<http://ies.ed.gov/ncee/wwc/>)
 - National Institute for Early Education Research (<http://nieer.org>)
 - Peabody Research Institute (<http://peabody.vanderbilt.edu/research/pri/>)
- We contacted experts in the field to inquire about ongoing studies, gray literature, and suggestions for additional contacts.

3.3 DATA COLLECTION AND ANALYSIS

3.3.1 Selection of studies

Two researchers (Baron and Melendez-Torres) independently conducted eligibility screening on all retrieved studies. Both researchers screened titles, abstracts, and (where appropriate) full texts in order to determine whether studies were suitable for inclusion in the review. All study inclusion disagreements were resolved through discussion and consensus.

3.3.2 Data extraction and management

Baron and Melendez-Torres also independently coded the studies selected for inclusion according to the data extraction form attached in Section 14. In instances of missing or unclear information, study authors were contacted for clarification. The level of agreement between the two coders was very high, and the only emergent disagreement was resolved

through discussion and consensus.

3.3.3 Assessment of risk of bias in included studies

Finally, Baron and Melendez-Torres independently coded each RCT for risk of bias using the Cochrane framework (Higgins & Green, 2011). We rated risk of bias for randomized trials as low-, high-, or unclear-risk across the following categories:

- *Random sequence generation:* How was random assignment executed? If the paper claims random assignment but does not explain the assignment mechanism, then this could be a source of bias.
- *Allocation concealment:* Did the person who conducted the assignment know which participants were being allocated to which group? If so, then the person might have assigned certain participants to an intervention in a non-random way (e.g., a teacher put a child in the Tools group because the child liked pretend play).
- *Blinding of participants and personnel:* Do the participants know they are receiving the treatment or control? In the present context, do teachers and students know whether they are receiving the Tools or comparison condition? If so, then their knowledge that they are in the treatment versus control group could bias their approach toward the study.
- *Blinding of outcome assessment:* Do the assessors know the condition assignment of the children they are assessing? If researchers know that the child is in a Tools classroom, then the researchers' evaluation of the child's self-regulation could, for example, be positively biased by an expectation that the child will be more self-regulated.
- *Incomplete outcome data:* Has there been substantial attrition from the study? If the missing data derives mostly from, for example, FRPL-eligible students who have moved homes or students with an IEP who get pulled out of the classroom for individualized instruction, then the results will not represent the true population of students.
- *Selective reporting:* Have all the outcome measures mentioned in the methodology section been reported in the results section? If the study collects data on certain outcome measures but does not report non-significant results, then the reported results could reflect the authors' biases regarding which outcomes were worthy to report.

3.3.4 Measures of treatment effect

As for effect size metrics, we used the standardized mean difference (Hedges' *g*) for continuous outcomes and planned to use the odds-ratio (OR) for binary outcomes. Overall, we conducted meta-analyses on the four outcomes noted in section 3.1.4:

- Task-based self-regulation measures (e.g., HTKS, peg tapping, etc.)
- Informant-based reports of children's self-regulation from teachers and researchers

(e.g., BRIEF – P, Child Social Behavior Questionnaire, etc.)

- Measures of children’s language and literacy skills
- Measures of children’s math skills

We computed effect sizes for each variant of the comparison condition (e.g., business-as-usual, other intervention, no treatment, etc.). For example, if a study compared Tools with another intervention as well as a business-as-usual curriculum, then we computed a separate effect size for each of the two comparison conditions.

3.3.5 Unit of analysis issues

In instances where one research program (i.e., a study) was associated with multiple manuscripts (i.e., reports), we treated the reports as deriving from a single study. Thus, our meta-analytic sample size equals the number of studies, not the number of reports associated with those studies.

For data extraction across multiple reports from the same study, we selected the report that yielded the most relevant information. The other reports were only used if they added unique information for data extraction purposes. All of the relevant recovered references for this review are listed in Section 7.1.

Because most studies yielded multiple effect sizes on the same outcome, data dependency among those nested effect sizes was treated through robust variance estimation (RVE) analysis with a multilevel meta-analysis robustness check (see section 3.3.9).

Finally, we adjusted the effect sizes to account for the intra-cluster correlation (ICC) among students in the same classroom. This is necessary because students who are in the same classroom are likely to affect one another’s academic and self-regulatory outcomes. Thus, students’ outcome data are not statistically independent of one another, which shrinks standard errors below their actual values and increases type I error rates (i.e., finding a statistically significant result that does not exist in the actual population). By correcting the effect sizes with ICC values, we could achieve more accurate estimation of standard errors (Hedges & Hedberg, 2007). When ICC values were not reported for a study, then we substituted values commonly found in the literature. Specifically, we used values of .10 for literacy and .11 for math (L. Hedges & Hedberg, 2007) and .015 for self-regulation (Fuhs, Farran, & Nesbitt, 2013).

3.3.6 Dealing with missing data

The authors of each included study were contacted to obtain, where relevant, missing data. Any missing data that was not explained within the study report or through correspondence with the authors was considered as a source of bias.

3.3.7 Assessment of reporting biases

Had a sufficient number of studies (i.e., ten or more) been retrieved, then a funnel plot would have been used to assess publication bias. However, we retrieved an insufficient number of studies, which precluded formal publication bias assessment. Nonetheless, we retrieved multiple unpublished studies, as will be described in section 4.2.7.

3.3.8 Data synthesis

We used the robust variance estimation (RVE) SPSS (IBM) macro described in Tanner-Smith & Tipton (2014) to compute pooled effect sizes that controlled for data dependency issues. That is, some studies included multiple effect sizes for the same outcome (e.g., both peg tapping and Heads-Toes-Knees-Shoulders for task-based self-regulation skills). Those effect sizes cannot be considered statistically independent from one another because they arise from the same study sample.

Traditional meta-analyses often address data dependency issues by selecting one outcome per study or averaging the effect sizes, which lead to a loss of information and power. In an attempt to use all available data, we analyzed all relevant effect sizes from each study while correcting for dependency in effect sizes from the same study through RVE.

In accordance with Tanner-Smith & Tipton (2014), we specified a rho value of .80, which indicates the assumed inter-correlation among effect sizes nested within the same study. The high rho value provides more conservative standard error estimates, which reduces the likelihood of type I errors. As a robustness check for the high rho value of .80, we also specified models with low (.20) and medium (.50) rho values to assess whether the results changed based on varying levels of assumed inter-correlation.

In addition to the RVE package in SPSS, we also used the metafor package (Viechtbauer, 2010) in R to perform multilevel meta-analysis with random effects on effect size as a robustness check on the findings. As with RVE, multilevel meta-analytic methods control for effect sizes nested within studies and are thus appropriate for addressing data dependency issues (Van den Noortgate, López-López, Marín-Martínez, & Sánchez-Meca, 2014).

3.3.9 Heterogeneity analysis

The RVE approach used in this study does not estimate heterogeneity in the same way as traditional multivariate meta-analysis. Specifically, the Q-statistic and I² statistic reported in many meta-analyses are not relevant within the RVE context. Instead, RVE estimates overall between-study heterogeneity is reported as a tau-squared (τ^2) value, which does not include an attendant test statistic or significance test (Tanner-Smith & Tipton, 2014). The τ^2 values are shown in section 4.3.2.

3.3.10 Subgroup analysis

We had proposed to conduct moderation analyses to investigate heterogeneity in the event that more than ten studies had been retrieved (see Littell, Corcoran, & Pillai, 2008). Specifically, we had sought to investigate the following study-level moderators:

- *Study design*: Do experimental and quasi-experimental designs exhibit consistently different effect sizes and significance values?
- *Study location*: Since the intervention was developed in the U.S., then does Tools' effect change across national contexts?

We had also sought to investigate the following child-level moderators, which are aggregated at the study level:

- *Age* (pre-kindergarten versus kindergarten)
- *Gender* (percentage of boys in the study sample)
- *Special education status* (percentage of special education students)
- *Socio-economic background* (percentage of free and reduced-price lunch (FRPL) eligibility)

4 Results

4.1 DESCRIPTION OF STUDIES

4.1.1 Results of the search

The search of the 11 aforementioned electronic databases yielded 63 total records (see Section 13 for the search terms and results for each database). In addition to the electronic database search, 123 records were identified through other components of the search strategy outlined in section 3.2.2 (i.e., reviewing reference lists ($k = 2$), hand-searching ($k = 0$), a forward citation search ($k = 120$), contacting experts ($k = 1$), and screening relevant websites ($k = 0$).

Ten of the records were duplicates, so 176 records remained after de-duplication. After screening the 176 titles and abstracts, 151 records were excluded that did not pertain to the Tools curriculum. The remaining 25 full texts were screened, and 14 records across six studies were deemed eligible for inclusion in the present review (see Figure 2 for the systematic review flowchart). Those six research programs, each with its own study ID¹, were detailed in 14 separate papers (see table 1).

4.1.2 Description of included studies

The characteristics of included studies table (Section 9) provides descriptive information on the included studies. Although Tools has been implemented in the United States, Canada, and parts of South America, all included studies were conducted in the United States. Moreover, all of the included studies were independent evaluations of the program; that is, the Tools developers did not conduct any of the studies.

As for publication type and status, two of the included studies (Barnett et al., 2008; Blair & Raver, 2014) have been published in peer-reviewed journals, one has been published as a government report (Morris et al., 2014), and the other three are article-length manuscripts that are awaiting publication (Clements et al., 2014; Farran & Wilson, 2014; Lonigan & Phillips, 2012).

¹ The study ID was chosen by the report from which we gained the most information.

All six studies featured cluster randomized controlled trial (RCT) designs, thus meeting the methodological inclusion criteria. Quasi-experimental studies were also eligible for inclusion in this review, but no quasi-experimental studies were recovered in the search. For the six RCT studies, five studies (Blair & Raver, 2014; Clements et al., 2014; Farran & Wilson, 2014; Lonigan & Phillips, 2012; Morris et al., 2014) used schools as the unit of randomization, whereas the other study (Barnett et al., 2008) used classrooms as the unit of randomization.

As for data analysis strategy, five of the six included studies (Barnett et al., 2008; Blair & Raver, 2014; Clements et al., 2014; Farran & Wilson, 2014; Lonigan & Phillips, 2012) used multilevel regression models to analyze child outcomes. Morris et al. (2014) did not report their data analysis strategy. All outcomes were continuous and were thus converted to standardized mean differences (Hedges' *g*) for this meta-analysis. Moreover, effect sizes from the five studies featuring hierarchical models were adjusted for intraclass correlation coefficients as described in section 3.3.5.

Finally, with regard to the implementation approach, four of the six included studies implemented Tools as a stand-alone intervention against a business-as-usual or another program (Barnett et al., 2008; Blair & Raver, 2014; Farran & Wilson, 2014; Morris et al., 2014). By contrast, the other two studies implemented Tools as part of a combined intervention. Specifically, Clements et al. (2014) implemented Tools alongside the Building Blocks math curriculum, whereas Lonigan & Phillips (2012) included two Tools conditions: one with Tools as a stand-alone program and another where Tools supplemented the Literacy Express Comprehensive Preschool Curriculum (LECPC). In the Lonigan & Phillips (2012) study, separate effect sizes were reported for each of the two Tools conditions, but the authors did not release the requisite data to include their study in this meta-analysis.

4.1.3 Excluded studies

The 11 excluded studies and the reasons for their exclusion are outlined in Section 9.

4.2 RISK OF BIAS IN INCLUDED STUDIES

All included studies were assessed using the Cochrane Handbook's (Higgins & Green, 2011) risk of bias tool. Section 9 includes risk of bias tables for each study that provide textual evidence either from the relevant study report(s) or from our correspondence with the authors to substantiate our risk of bias rating. In addition to Section 9, the sections below assess all included studies across the six risk of bias dimensions outlined in Section 3.3.3.

4.2.1 Random sequence generation

Across the six included studies, four studies were considered low-risk for random sequence generation bias, whereas two studies were considered unclear risk. For the four low-risk studies, three used computer-generated randomization (Barnett et al., 2008; Blair & Raver,

2014; Farran & Wilson, 2014), whereas the fourth (Clements et al., 2014) used the circular sampling scheme, which has been shown (Lahiri, 1951) to ensure proper randomization.

The remaining two studies (Lonigan & Phillips, 2012; Morris et al., 2014) did not report their random sequence generation process, which explains their rating of unclear risk. Both studies were reported as randomized controlled trials, so it is likely that both studies either attempted or achieved effective randomization. However, without evidence from the study or authors, the studies' potential for bias arising from random sequence generation bias remains unclear.

4.2.2 Allocation concealment

Across the six included studies, three were considered low-risk for allocation concealment bias (Barnett et al., 2008; Blair & Raver, 2014; Farran & Wilson, 2014), whereas the remaining three studies (Clements et al., 2014; Lonigan & Phillips, 2012; Morris et al., 2014) were considered unclear risk.

Section 9 indicates the textual evidence from each study that indicates who conducted the randomization. Studies where researchers were not themselves controlling the assignment process received 'low-risk' ratings, whereas studies without specific information on the assignment process received 'unclear risk.'

4.2.3 Blinding of participants and personnel

As with most educational interventions, it was not possible to blind the students and teachers to their curricular assignment. Teachers must know what curriculum they are using in order to implement it, which precludes the possibility of true blinding. In instances where blinding of participants and personnel is impossible, the Cochrane Handbook (Higgins & Green, 2011) dictates that the studies should be considered to have an unclear risk of bias. Thus, all included studies were considered to exhibit an unclear risk of bias on this dimension.

4.2.4 Blinding of outcome assessment

All included studies implemented outcome assessment protocols that aimed to ensure blindness of the assessors to the children's condition. That is, assessors who filled out observational reports of children's self-regulation were meant to be blind to the child's group assignment during the assessment period.

However, researchers across studies indicated that assessors may have intuited children's group assignment based on student and classroom characteristics (e.g., researchers saw children engaging in a Tools activity). Thus, although the studies were designed to ensure blinding of outcome assessment, the studies could not guarantee that blindness occurred. Thus, the only study without a high risk of bias was Clements et al. (2014), which received an unclear rating because no information about potential bias in the outcome assessment was provided.

4.2.5 Incomplete outcome data

Across the six studies, three were considered low-risk for attrition-related issues (Barnett et al., 2008; Blair & Raver, 2014; Farran & Wilson, 2014), one was considered high-risk (Clements et al., 2014), and two were considered unclear risk (Lonigan & Phillips, 2012; Morris et al., 2014). The low-risk studies each reported the levels of missingness, their analyses to address the attrition, and the statistically non-significant differences between the attrited participants and the remaining participants.

By contrast, Clements et al. (2014) noted substantial attrition in their study but did not conduct analyses to assess the impacts of the attrition. Once again, Section 9 contains textual evidence to indicate why that study received a ‘high-risk’ rating. Finally, the remaining two studies provided no information regarding attrition, which explains the ‘unclear risk’ rating given to those studies.

4.2.6 Selective reporting

All six studies exhibited a low-risk of selective reporting bias. This is because all studies reported on all outcomes mentioned in their methodology sections.

4.2.7 Other sources of bias

No other source of bias was identified within the included studies. That said, across the entire set of studies, it is possible that the Tools literature base suffers from publication bias. Specifically, among the present set of included studies, the two studies to indicate significant positive results for Tools have both been published (Barnett et al., 2008; Blair & Raver, 2014), whereas three papers that show null or negative effects (Clements et al., 2014; Farran & Wilson, 2014; Lonigan & Phillips, 2012) have not.

In fact, the only ‘null effects’ study to have been published (Morris et al., 2014) was commissioned by the United States government and was thus published as a government report instead of as a research paper. Thus, no studies that indicate null effects for Tools have been published in peer-reviewed academic journals, even though these studies constitute the majority of the Tools evidence base.

With fewer than 10 studies eligible in the present review, however, a visual inspection of publication bias via funnel plot was not possible. It is possible that the imbalance between published and unpublished findings is simply due to chance. Given the relative nascency of the Tools evidence base, it is possible that the publication of various findings will naturally balance over time.

As for other common sources of bias, one strength of the current Tools research base is that the Tools curricular developers did not conduct any of the included Tools evaluation studies. A review by Gellis and Reid (2004) found that program developers’ financial and emotional investment in the research outcomes can sometimes bias results. This was not an issue for

the included Tools studies because the Tools developers did not conduct any of the included studies.

4.2.8 Risk of bias summary

The risk of bias results are visually summarized in Figure 3. Figure 3 indicates an unclear risk of bias across all studies for the blinding of personnel, which was not possible with the Tools program. Moreover, five of the six studies exhibited a high risk of bias for the blinding of outcome assessors, which was difficult to ensure across studies. Moreover, one of the six studies exhibited a high risk of attrition bias because of incomplete outcome data. Beyond that, none of the studies exhibited a high risk of bias across any of the other Cochrane handbook's (Higgins & Green, 2011) risk of bias dimensions.

4.3 QUANTITATIVE SYNTHESIS OF RESULTS

4.3.1 Overall findings

Whereas the previous sections included information across six studies, one of the studies (Lonigan & Phillips, 2012) did not report the necessary outcome data to be included in the quantitative synthesis. Despite numerous attempts to contact the study authors, the requisite data were not made available, thus precluding this study's inclusion in the meta-analysis.

Thus, this section presents the meta-analytic results across the five studies incorporated into the quantitative synthesis. As described in section 3.3.9, each study yielded multiple effect sizes on at least one of the relevant outcome measures. Because effect sizes from the same study are based on the same sample of children and the same study characteristics, those effect sizes cannot be considered statistically independent from one another. Thus, robust variance estimation (RVE) was used to account for shared variation among effect sizes from the same study.

The final results (see table 2) favored Tools for each pooled effect size, but the effect sizes were small and did not reach statistical significance for three of the four outcome measures. The only exception was math, where the pooled effect size was small yet statistically significant ($g = .061$, $p < .05$) in favor of the Tools condition. Forest plots with individual effect sizes across each of the four outcome measures are presented in Section 11.

4.3.2 Subgroup analysis to explore heterogeneity

The RVE analysis indicated low levels of between-study variability across the four outcome measures: Executive function ($\tau^2 = .02$), self-regulation ($\tau^2 = .03$), literacy ($\tau^2 = .00$), and math ($\tau^2 = .00$). Again, τ^2 values do not have a test statistic for significance testing, but the τ^2 values observed here are very small in magnitude. Given the small number of included studies ($k = 5$), we did not conduct subgroup analyses to explore heterogeneity.

4.3.3 Robustness check using multilevel meta-analysis methods

As indicated in section 3.3.9, we assessed the findings' robustness using multilevel meta-analysis in the metafor package (Viechtbauer, 2010) of R Studio. As with the robust variance estimation (RVE) method, multilevel meta-analysis addresses the issue of data clustering (i.e., effect sizes nested within the same study).

The robustness check (table 3) results largely mirror those observed in the RVE analysis. Specifically, all effect sizes favored the Tools condition, and the effect size magnitudes are mostly similar to those observed for the RVE analysis. One difference is that the math effect size went from being statistically significant ($g = .061$, $p < .05$) in the RVE analysis to marginally significant ($g = .061$, $p = .08$) in the multilevel meta-analysis approach. Beyond that, the results were highly robust across the two methods.

5 Discussion

5.1 SUMMARY OF MAIN RESULTS

This summary section is divided into two parts: 1) A discussion of the systematic review results and 2) a summary of the meta-analytic results.

5.1.1 Systematic review results

This review recovered 14 records across six studies. Through careful reading of the retrieved manuscripts, we ensured that overlapping samples reported across multiple records were counted as the same study instead of as different studies (Lipsey & Wilson, 2001). Each of the six recovered studies classifies as a randomized controlled trial (RCT); no quasi-experimental Tools studies were recovered in the search. Thus, the number of included studies was lower than the required ten for further analyses such as formal publication bias assessment and moderation analysis (Littell et al., 2008; Tanner-Smith & Tipton, 2014).

Nonetheless, the small number of recovered studies is consistent with the relatively sparse literature on other early childhood self-regulation programs. For example, the Chicago School Readiness Project, which, like Tools, is a new school-based self-regulation intervention, has only two RCT-based evaluation reports (S. Jones, Bub, & Raver, 2013; Raver et al., 2011), whereas Montessori, a well-established early childhood curriculum, has no RCT studies in its evidence base (Institute of Educational Sciences, 2016; Lillard, 2005).

Similarly, the Incredible Years has been implemented for over thirty years in more than twenty countries (K. Jones et al., 2007), yet the number of existing studies for any one of its multiple intervention arms is notably low. For example, Nye (2013) analyzed the evidence base for the Incredible Years Teacher Classroom Management program (TCM), which, like Tools, is a school-based intervention. Nye (2013) recovered only four studies to include in the systematic review, even though the Incredible Years has been in existence for longer than Tools.

The most recent systematic review, to the best of our knowledge, of early childhood executive function interventions (Jacob & Parkinson, 2015) also reported difficulties with identifying a large number of evaluation studies. In fact, the authors reported outcomes from only one

study on both the Chicago School Readiness Project as well as the Head Start REDI program. Thus, the small number of included studies in this review mostly mirrors the small evidence base observed for other similar programs.

5.1.2 Summary of the meta-analytic results

The meta-analytic results favor Tools across all four outcome measures. However, those effect sizes did not reach statistical significance across three outcomes: 1) assessor report-based ratings of children's self-regulation, 2) task-based self-regulation indicators, and 3) literacy skills. By contrast, small but statistically significant impacts were observed for the math pooled effect size.

The significant math effect for Tools is noteworthy given that the Tools developers consider math to be an area of weakness for the curriculum (Leong, Bodrova, & Hensen, 2008). One possible explanation for the observed effect is that few early childhood programs allocate time to math at all (Hirsh-Pasek, K., & Golinkof, 2016); thus, even though Tools may not have an especially strong math regimen, Tools students may still have been exposed to more math than were students in many of the comparator classrooms. Because included studies did not report on the time spent on math in comparator classrooms, this hypothesis could not be tested.

In summary, the effect sizes were all in the positive direction for Tools students, but the effect sizes were small (i.e., max = .12) and statistically insignificant for three of the four outcome measures. Consequently, despite potentially promising evidence from the positive effect sizes in favor of Tools, more research is necessary to demonstrate that those effects are statistically significant as opposed to arising from chance alone.

5.2 QUALITY OF THE EVIDENCE

Each of the recovered studies classifies as an RCT, which bolsters initial confidence in the quality of the evidence. However, the small number of included studies reduces power and increases the possibility of Type II errors, which could have occurred in the present review. Publication of additional Tools evaluations will improve power in future meta-analyses.

In addition to power issues, the literature also suffers from risk of bias issues. Specifically, one of the studies (Clements et al., 2014) exhibited a high risk of attrition bias. This is because the authors noted but did not analyze differences between the participants who remained and the participants who left the study, which introduces bias into those studies' results. In addition to the high risk of bias in that study, the other studies received an unclear bias rating across several dimensions because of inadequate reporting.

Thus, despite issues of power and potential risk of bias among included studies, the rigorous RCT designs across studies suggests that the quality of the Tools evidence base is relatively

high.

5.3 LIMITATIONS AND POTENTIAL BIASES IN THE REVIEW PROCESS

This review exhibits three main limitations. Firstly, we could not retrieve data from the Lonigan & Phillips (2012) study for this meta-analysis. That study's large sample size ($n = 2,564$) would have decreased the standard errors of the observed effect sizes, which would lend confidence to the robustness of the results.

The second limitation of this review is the small number of included studies, which yields uncertainty about the reliability of our RVE and multilevel estimates. For RVE, Tanner-Smith et al. (2014) recommends including at least 10 studies with at least one effect size per study. We had only five studies, though all of them yielded multiple effect sizes. Given that we had fewer than the recommended number of included studies, the reliability of the results remains in question. Nonetheless, the consistent results observed in the robustness check with the multilevel meta-analysis approach bolsters confidence in the findings.

Thirdly, the small number of included studies precluded formal publication bias assessment and moderation analysis. It is hoped that future updates to this review will include more studies to enable these analyses. Despite those three limitations, it is important to note that there were no deviations from the review protocol.

5.4 AGREEMENTS AND DISAGREEMENTS WITH OTHER STUDIES OR REVIEWS

No previous reviews have focused specifically on the Tools program. Nonetheless, a recent review by Jacob and Parkinson (2015) examined whether school-based executive function interventions improve academic achievement, and the authors included the Diamond et al. (2007) and Farran & Wilson (2014) Tools studies reviewed here. Jacob and Parkinson concluded that there is “no compelling evidence of a positive impact on EF and no evidence of positive impact on achievement for the Tools program” (p. 25).

However, that review did not meta-analyze the Tools data and included only two studies, so their results do not cover the full range of evidence considered here. Nonetheless, that study reached similar conclusions as the present review regarding self-regulation outcomes, whereas this review found math gains for Tools students; Jacob and Parkinson (2015) did not.

Beyond that review, two review studies (K. Bierman & Torres, 2016; Diamond & Lee, 2011) pertaining to the range of self-regulation interventions both mention Tools. However, neither study used systematic search or review methods, neither study included a meta-

analysis, and neither study included the full range of Tools studies included here. The Diamond & Lee (2011) review concludes that Tools is effective based on only one report (Diamond et al., 2007), which was conducted by the same lead author who conducted the review study. By contrast, Bierman & Torres (2016) mention four Tools studies but make no comment on the program's effectiveness.

6 Authors' conclusions

Tools' educational approach aligns with many child developmental theories as well as notions of best practice in the early childhood education field. The results presented here indicate small yet positive results for the Tools program. While these results are promising, they are based on a small evidence base; thus, more research is necessary to demonstrate Tools' effectiveness in promoting children's self-regulation skills.

6.1 IMPLICATIONS FOR PRACTICE AND POLICY

The Tools developers have repeatedly hypothesized (Bodrova & Leong, 2007, 2013; Leong & Bodrova, 2011) gains for Tools students, especially for self-regulation. In the results presented here, the effect sizes were all in the positive direction for Tools students, although the effect sizes were small (i.e., max = .12) and statistically non-significant for three of the four outcome measures.

That said, it is important to clarify that “no evidence of an effect is *not* the same as evidence of no effect; insufficient statistical power (too few studies, too much heterogeneity) is an alternative explanation for null results” (Littell et al., 2008, p. 135). In other words, we cannot conclude that Tools *does not work* in promoting children's self-regulation; rather, the evidence produced here simply does not conclusively demonstrate that Tools *does work* as hypothesized by the developers.

Although the null statistical effects did not align with the developers' expectations, the results are consistent with other evaluations of early childhood programs. For example, a national evaluation of 14 preschool curricula in the United States (Preschool Curriculum Evaluation Research Consortium, 2008) found that none of the curricula significantly improved children's self-regulation skills beyond comparator curricula.

This is not to say that children's self-regulation did not improve in any curriculum; rather, the studies showed no evidence that one curriculum promoted children's self-regulation significantly more than any of the other sampled curricula. Thus, the absence of an observed effect for Tools perhaps may not be surprising, even though Tools explicitly claims to hone children's self-regulation skills.

As for early childhood programs that have shown gains for participating children, the Perry Preschool project (Berrueta-Clement, 1984) and the Abecedarian project (Ramey & Campbell, 1984) rank among the most cited intervention programs with evidence bases suggesting effectiveness. In contrast to Tools, those two programs were implemented with a small number of students and highly trained teachers in a specific site (Hyson, Copple, & Jones, 2006). Those programs lack evidence of effective scaling among large numbers of students or proliferation across multiple geographic settings, but their effectiveness on a small scale remains compelling (Heckman, Moon, Pinto, Savelyev, & Yavitz, 2010).

Given the promising results from two of the Tools RCT studies (Barnett et al., 2008; Blair & Raver, 2014), it seems possible that, as with the Perry preschool and Abecedarian projects, the program can work well when implemented and tested on a smaller scale. Thus, looking forward, perhaps the Tools developers could lead a small-scale Tools study that identifies the core components of the program as well as the optimal training regimen for teachers.

6.2 IMPLICATIONS FOR RESEARCH

Multiple possibilities for future analyses would strengthen the existing Tools literature base. Three possibilities for future research include: 1) a multi-arm trial that directly compares Tools with other self-regulation programs, 2) a meta-analysis of several early childhood interventions and curricula, and 3) a study that accounts for measurement error in the self-regulation construct.

6.2.1 A multi-arm trial comparing Tools to other self-regulation programs

Four of the six Tools included Tools evaluation studies compared Tools against a single ‘business-as-usual’ condition. By contrast, the other two studies compared Tools against both a ‘business-as-usual’ group as well as another intervention group: Lonigan & Phillips (2012) used the Literacy Express curriculum and Clements et al. (2014) used the Building Blocks math curriculum. These latter two studies enable assessment of the relative effectiveness among Tools, another target early childhood curriculum, and a ‘business-as-usual’ program.

Unfortunately, in the existing research, the two multi-arm trials compared Tools against literacy and math curricula, respectively, instead of against other self-regulation interventions. In the future, a large-scale, multi-arm trial that compares Tools to other self-regulation interventions such as the Incredible Years, the Chicago School Readiness Project, the Promoting Alternative Thinking Strategies program, and other programs within one large-scale randomized trial may provide the most useful evidence base for understanding differential intervention effectiveness.

In the present meta-analysis, we observed that Tools did not predict significantly improved task-based or assessor-reported self-regulation relative to the set of comparator curricula.

However, the set of comparator curricula largely involved programs such as HighScope, Creative Curriculum, and others. Like Tools, those are all comprehensive curricula, but, in contrast to Tools, they lack a specific focus on self-regulation. Thus, a multi-arm trial comparing Tools with several other early childhood self-regulatory interventions could represent a significant contribution to the evidence base.

6.2.2 A network meta-analysis of several early childhood interventions

Whereas the research plan described in the previous section refers to a single, large-scale trial that concurrently analyzes multiple self-regulation interventions, another research program could conduct a network meta-analysis that aggregates data across multiple studies regarding multiple interventions. Many meta-analyses do analyze multiple interventions, but, to the best of my knowledge, no meta-analyses have investigated multiple self-regulation interventions' impact on children's self-regulation skills.

The aforementioned Jacob and Parkinson (2015) review did investigate the impacts of executive function interventions on children's academic achievement. However, that study only used only academic measures as outcomes, whereas students' self-regulation skills were not assessed. A future review could expand on the Jacob and Parkinson (2015) study by investigating multiple self-regulation interventions' effects on children's self-regulation outcomes instead of only academic skills.

6.2.3 Modeling measurement error in the self-regulation construct

Given that the meta-analysis revealed null effects across three of the four outcome measures, it may seem sensible to conclude that Tools has no effect on children's self-regulatory and literacy skills. Once again, however, no evidence of effect is not the same as evidence of no effect. Instead, it is possible that other factors masked Tools' impact on child outcomes.

One such factor could be measurement error in the assessment instruments. Although the included studies exclusively employed standardized testing instruments, it remains possible that those instruments, especially for self-regulation, had low construct validity, which has been noted in the self-regulation measurement literature (McClelland & Cameron, 2012).

Thus, it is possible that low construct validity across the measures may have contributed to the observed null results. As Kline (2015) writes, measurement error "generally reduces effect sizes below their true (population) values" (p. 92). Kline goes on to recommend the use of latent variable models, which account for measurement error in order to more accurately capture relationships among phenomena (e.g., between the Tools program and self-regulation).

It is hoped that future research can obtain raw data from all others in order to transform the observed measures into latent self-regulation constructs that account for measurement error. In so doing, the pooled effect sizes for Tools versus comparison group children would represent a more accurate estimation of Tools' true impact on children's self-regulatory skills.

7 References

7.1 REFERENCES TO INCLUDED STUDIES

- Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly*, 23(3), 299–313.
- Blair, C., & Raver, C. (2014). Closing the achievement gap through modification of neurocognitive and neuroendocrine function: Results from a cluster randomized controlled trial of an innovative approach to the education of children in kindergarten. *PloS One*, 9(11), 1–13.
- Clements, D., Sarama, J., Germeroth, C., & Fesler, L. (2014). Effects on executive function and mathematics learning of an early mathematics curriculum synthesized with scaffolded play designed to promote self-regulation versus the mathematics curriculum alone (in press). *Cognition and Instruction*, 1–67.
- Farran, D., & Wilson, S. (2014). Achievement and self-regulation in pre-kindergarten classrooms: Effects of the Tools of the Mind curriculum (in press). *Child Development*, 1–66.
- Lonigan, C. J., & Phillips, B. M. (2012). Comparing skills-focused and self-regulation focused preschool curricula: Impacts on academic and self-regulatory skills. In *Society for Research on Educational Effectiveness*. Washington DC.
- Morris, P., Mattera, S., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). Impact findings from the Head Start CARES: National evaluation of three approaches to improving preschoolers' social and emotional competence. *Office of Planning, Research and Evaluation Report*. Washington DC.

7.2 REFERENCES TO EXCLUDED STUDIES

- Bodrova, E., & Leong, D. (2001). Tools of the Mind: A case study of implementing the Vygotskian approach in American early childhood and primary classrooms. *Innodata Monographs*, 7, 1–44.
- Bodrova, E., Leong, D., & Akhutina, T. (2011). When everything new is well-forgotten old: Vygotsky/Luria insights in the development of executive functions. *New Directions for Child and Adolescent Development*, 133, 11–28.
- Copple, C. (2003). Fostering young children’s representation, planning, and reflection: A focus in three current early childhood models. *Journal of Applied Developmental Psychology*, 24(6), 763–771.
- Grigorenko, E. L. (1998). Mastering tools of the mind in school (trying out Vygotsky’s ideas in classrooms). *Intelligence, Instruction, and Assessment: Theory into Practice*. ELEC, Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.
- Hammer, C., Blair, C., Lopez, L., Leong, D., & Bodrova, E. (2012). Tools of the Mind: Promoting the school readiness of ELLs. In *Society for Research on Educational Effectiveness* (pp. 1–6). Washington DC.
- Hyson, M., Copple, C., & Jones, J. (2006). Early childhood development and education. In *Handbook of child psychology: Child psychology in practice* (6th ed., pp. 3–47). Hoboken, NJ, US: Wiley.
- Mackay, P. E. (2013). *The effects of tools of the mind on math and reading scores in kindergarten*. ProQuest Dissertations and Theses. University of Massachusetts - Amherst, Ann Arbor.
- Magalhaes, A. M. (2013). *Exploring effective strategies for developing executive functioning skills within an early childhood classroom*. ProQuest Dissertations and Thesis. Northeastern University, Boston.
- Millaway, S. A. (2015). *Impact of tools of the mind on middle school achievement*. ProQuest Dissertations and Theses. Saint Peter’s University, Ann Arbor.
- Rodgers, M. S. (2012). *Structured play and student learning in kindergarten: An outcome evaluation*. ProQuest Dissertations and Theses. Northeastern University, Ann Arbor.
- Shaheen, S. (2014). How child’s play impacts executive function-related behaviors. *Applied Neuropsychology: Child*, 3(3), 182–187.

7.3 ADDITIONAL REFERENCES

- Alexander, K., Entwisle, D., & Kabbani, N. (2001). The dropout process in life course perspective: Early risk factors at home and school. *The Teachers College Record*, 103(5), 760–822.
- Alloway, T., Lawrence, A., & Rodger, S. (2013). Antisocial behavior: Exploring behavioral, cognitive, and environmental influences on expulsion. *Applied Cognitive Psychology*, 27(4), 520–526.
- Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008a). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly*, 23(3), 299–313.
- Barnett, W. S., Jung, K., Yarosz, D. J., Thomas, J., Hornbeck, A., Stechuk, R., & Burns, S. (2008b). Educational effects of the Tools of the Mind curriculum: A randomized trial. *Early Childhood Research Quarterly*, 23(3), 299–313.
- Berrueta-Clement, J. (1984). Changed lives: The effects of the Perry preschool program on youths through age 19. *Monographs of the High/Scope Educational Research Foundation*, 393(6), 1–193.
- Bierman, K. L., Domitrovich, C., Blair, C., Nelson, K., & Gill, S. (2008). Promoting academic and social-emotional school readiness: The Head Start REDI program. *Child Development*, 79(6), 1802–1817.
- Bierman, K. L., Nix, R. L., & Greenberg, M. T. (2008). Executive functions and school readiness intervention: Impact, moderation, and mediation in the Head Start REDI program. *Development and Psychopathology*, 20(3), 821–843.
- Bierman, K., & Torres, M. (2016). Promoting the development of executive functions through early education and prevention programs. In J. Griffin, L. Freund, & P. McCardle (Eds.), *Executive function in preschool age children: Integrating measurement, neurodevelopmental and translational research* (Vol. 1, pp. 35–50). Washington DC: American Psychological Association.
- Blair, C., & Raver, C. (2014). Closing the achievement gap through modification of neurocognitive and neuroendocrine function: Results from a cluster randomized controlled trial of an innovative approach to the education of children in kindergarten. *PLoS One*, 9(11), 1–13.
- Blair, C., & Ursache, A. (2011). A bidirectional model of executive functions and self-regulation. In K. Vohs & R. Baumeister (Eds.), *Handbook of self-regulation: Research, theory, and applications* (2nd ed., pp. 300–320). New York: Guilford Press.
- Bodrova, E., & Leong, D. (2005). Self-regulation: A foundation for early learning. *Principal*, 85(1), 30–35.
- Bodrova, E., & Leong, D. (2007). *Tools of the Mind: The Vygotskian approach to early childhood education* (2nd ed.). Columbus: Pearson.
- Bodrova, E., & Leong, D. (2013). Play and self-regulation: Lessons from Vygotsky. *American Journal of Play*, 6(1), 111–123.
- Bodrova, E., & Leong, D. (2015). Locations and alignments.
- Bodrova, E., Leong, D., & Akhutina, T. (2011). When everything new is well-forgotten old: Vygotsky/Luria insights in the development of executive functions. *New Directions for Child and Adolescent Development*, 133, 11–28.

- Calkins, S. D., Howse, R. B., & Philippot, P. (2004). Individual differences in self-regulation: Implications for childhood adjustment. In *The regulation of emotion* (pp. 307–332).
- Center for Disease Control. (2013). *Trends in the parent-report of health care provider-diagnosis and medication treatment for ADHD: United States, 2003-2011*. Atlanta, Georgia.
- Clements, D., & Sarama, J. (2012). The efficacy of an intervention synthesizing scaffolding designed to promote self-regulation with an early mathematics curriculum: Effects on executive function. In *Society for research on educational effectiveness* (pp. 1–7).
- Clements, D., Sarama, J., Germeroth, C., & Fesler, L. (2014). Effects on executive function and mathematics learning of an early mathematics curriculum synthesized with scaffolded play designed to promote self-regulation versus the mathematics curriculum alone (in press). *Cognition and Instruction*, 1–67.
- Diamond, A., Barnett, S., Thomas, J., & Munro, S. (2007a). Preschool program improves cognitive control. *Science*, 318, 1387–1388.
- Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007b). Preschool program improves cognitive control. *Science*, 318, 1387–1388.
- Diamond, A., Barnett, W. S., Thomas, J., & Munro, S. (2007). Preschool program improves cognitive control. *Science*. Diamond, Adele: Department of Psychiatry, University of British Columbia, Vancouver, BC, Canada, V6T 2A1, adele.diamond@ubc.ca: American Assn for the Advancement of Science.
- Diamond, A., & Lee, K. (2011). Interventions shown to aid executive function development in children 4-12 years old. *Science*, 333, 959–964.
- District of Columbia Public Schools. (2016). Early childhood academic programs.
- Domitrovich, C. E., Cortes, R. C., & Greenberg, M. T. (2007). Improving young children's social and emotional competence: A randomized trial of the preschool "PATHS" curriculum. *Journal of Primary Prevention*, 28(2), 67–91.
- Drake, K., Belsky, J., & Fearon, R. M. P. (2014). From early attachment to engagement with learning in school: The role of self-regulation and persistence. *Developmental Psychology*, 50(5), 1350–1361.
- Elkonin, D., & Zaporozhets, A. (1978). *The psychology of preschool children*. Cambridge, MA: MIT Press.
- Farran, D., & Wilson, S. (2014). Achievement and self-regulation in pre-kindergarten classrooms: Effects of the Tools of the Mind curriculum (in press). *Child Development*, 1–66.
- Feng, X., Shaw, D. S., Kovacs, M., Lane, T., O'Rourke, F. E., & Alarcon, J. H. (2008). Emotion regulation in preschoolers: The roles of behavioral inhibition, maternal affective behavior, and maternal depression. *Journal of Child Psychology and Psychiatry*, 49(2), 132–41.
- Fisher, K. R., Hirsh-Pasek, K., Newcombe, N., & Golinkoff, R. M. (2013). Taking shape: Supporting preschoolers' acquisition of geometric knowledge through guided play. *Child Development*, 84(6), 1872–1878.
- Flouri, E., Midouhas, E., & Joshi, H. (2014). Family poverty and trajectories of children's emotional and behavioural problems: The moderating roles of self-regulation and verbal

- cognitive ability. *Journal of Abnormal Child Psychology*, 42(1), 1–14.
- Fuhs, M., Farran, D., & Nesbitt, K. (2013). Preschool classroom processes as predictors of children's cognitive self-regulation skills development. *School Psychology Quarterly*, 28(4), 347–359.
- Gellis, Z., & Reid, W. J. (2004). Strengthening evidence-based practice. *Brief Treatment and Crisis Intervention*, 4(2), 155.
- Gilliam, W. S. (2005). *Prekindergarteners left behind: Expulsion rates in state prekindergarten systems*. New York: Foundation for Child Development.
- Gioia, G., Espy, K. A., & Isquith, P. K. (2005). Behavior Rating Inventory of Executive Function–Preschool Version (BRIEF-P). Odessa, Florida: Psychological Assessment Resources.
- Gulchak, D. J. (2008). Using a mobile handheld computer to teach a student with an emotional and behavioral disorder to self-monitor attention. *Education and Treatment of Children*, 31(4), 567–581.
- Harris, K. R., Friedlander, B. D., & Graham, S. (2005). Self-monitoring of attention versus self-monitoring of academic performance: Effects among students with ADHD in the general education classroom. *The Journal of Special Education*, 39(3), 145–156.
- Heckman, J., & Masterov, D. (2007). The productivity argument for investing in young children. *Review of Agricultural Economics*, 29(3), 446–493.
- Heckman, J., Moon, S. H., Pinto, R., Savelyev, P. A., & Yavitz, A. (2010). The rate of return to the HighScope Perry Preschool Program. *Journal of Public Economics*, 94(1), 114–128.
- Hedges, L., & Hedberg, E. (2007). Intraclass correlation values for planning group-randomized trials in education. *Educational Evaluation and Policy Analysis*, 29(1), 60–87.
- Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods*, 1(1), 39–65.
- Higgins, J., & Green, S. (Eds.). (2011). *Cochrane handbook for systematic reviews of interventions*. Cochrane Collaboration.
- Hirsh-Pasek, K., & Golinkof, R. M. (2016). The preschool paradox. *Science*, 351(6278), 1158–1158. *Science*, 361, 1158.
- Hyson, M., Copple, C., & Jones, J. (2006). Early childhood development and education. In *Handbook of child psychology: Child psychology in practice* (6th ed., pp. 3–47). Hoboken, NJ, US: Wiley.
- IBM. (2012). *SPSS Statistics for Windows*. Armonk, New York: IBM Corp.
- Institute of Educational Sciences. (2016). Montessori method: No report available - No studies meeting evidence standards.
- Jacob, R., & Parkinson, J. (2015). The potential for school-based interventions that target executive function to improve academic achievement: A review. *Review of Educational Research*, 10(1), 1–41.
- Jones, K., Daley, D., Hutchings, J., Bywater, T., & Eames, C. (2007). Efficacy of the Incredible

- Years basic parent training programme as an early intervention for children with conduct problems and ADHD. *Childcare, Health and Development*, 33(6), 749–756.
- Jones, S., Bub, K., & Raver, C. (2013). Unpacking the black box of the CSRP intervention: The mediating roles of teacher-child relationship quality and self-regulation. *Early Education and Development*, 24, 1043–1064.
- Karreman, A., Van Tuijl, C., & Marcel, A. G. (2006). Parenting and self-regulation in preschoolers: A meta-analysis. *Infant and Child Development*, 15, 561–579.
- Kim, S., Nordling, J., Yoon, J., & Kochanska, G. (2014). Effortful control in “hot” and “cool” tasks differentially predicts children’s behavior problems and academic performance. *Journal of Abnormal Child Psychology*, 41(1), 43–56.
- Lahiri, D. B. (1951). A method for sample selection providing unbiased ratio estimates. *International Statistical Institute Bulletin*, 33(2), 133–140.
- Leong, D., & Bodrova, E. (2011). *Tools of the Mind: Pre-k training manual*. Denver: Metropolitan State College of Denver.
- Leong, D., Bodrova, E., & Hensen, R. (2008). *Tools of the Mind: Kindergarten manual*. Metropolitan State College of Denver.
- Lillard, A. (2005). *Montessori: The science behind the genius*. New York: Oxford University Press.
- Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. London: Sage.
- Littell, J. H., Corcoran, J., & Pillai, V. (2008). *Systematic reviews and meta-analysis*. London: Sage.
- Lonigan, C. J., & Phillips, B. M. (2012). Comparing skills-focused and self-regulation focused preschool curricula: Impacts on academic and self-regulatory skills. In *Society for Research on Educational Effectiveness*. Washington DC.
- McClelland, M., & Cameron, C. E. (2012). Self-regulation in early childhood: Improving conceptual clarity and developing ecologically valid measures. *Child Development Perspectives*, 6(2), 136–142.
- McClelland, M., & Tominey, S. (2011). Introduction to the special issue on self-regulation in early childhood. *Early Education & Development*, 22(3), 355–359.
- Miller, M., Nevado-Montenegro, A. J., & Hinshaw, S. P. (2012). Childhood executive function continues to predict outcomes in young adult females with and without childhood-diagnosed ADHD. *Journal of Abnormal Child Psychology*, 40(5), 657–68.
- Moffitt, T. E., Arseneault, L., & Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences of the United States of America*, 108(7), 2693–2698.
- Morris, P., Mattera, S., Castells, N., Bangser, M., Bierman, K., & Raver, C. (2014). Impact findings from the Head Start CARES: National evaluation of three approaches to improving preschoolers’ social and emotional competence. *Office of Planning, Research and Evaluation Report*. Washington DC.
- Morris, P., Mattera, S. K., Castells, N., Bangser, M., Bierman, K., Raver, C., & MDRC. (2014). *Impact Findings from the Head Start CARES Demonstration: National Evaluation of Three Approaches to Improving Preschoolers’ Social and Emotional Competence*.

- National Institute for Health and Care Excellence. (2013). Antisocial behaviour and conduct disorders in children and young people: Recognition and management.
- Nota, L., Soresi, S., & Zimmerman, B. (2004). Self-regulation and academic achievement and resilience: A longitudinal study. *International Journal of Educational Research*, 41(3), 198–215.
- Nunes, T., Bryant, P., Evans, D., Bell, D., Gardner, S., Gardner, A., & Carraher, J. (2007). The contribution of logical reasoning to the learning of mathematics in primary school. *British Journal of Developmental Psychology*, 25(1), 147–166.
- Nye, E. (2013). *Effectiveness of the Incredible Years Teacher Classroom Management program on student and teacher outcomes: A systematic review and meta-analysis (Master's thesis)*. University of Oxford.
- O'Shaughnessy, T. E., Lane, K. L., Gresham, F. M., & Beebe-Frankenberger, M. E. (2003). Children placed at risk for learning and behavioral difficulties: Implementing a school-wide system of early identification and intervention. *Remedial and Special Education*, 24(1), 27–35.
- Office of the State Superintendent of Education. (2013). *Reducing out-of-school suspensions and expulsions in District of Columbia public and public charter schools*. Washington DC.
- Olson, S., & Lunkenheimer, E. (2009). Expanding concepts of self-regulation to social relationships: Transactional processes in the development of early behavioral adjustment. In A. Sameroff (Ed.), *The transactional model of development: How children and contexts shape each other* (pp. 55–76). Washington DC: American Psychological Association.
- Piaget, J. (1951). *Play, dreams, and imitation in childhood*. London: Heinemann.
- Ponitz, C. C., McClelland, M. M., Matthews, J. S., & Morrison, F. J. (2009). A structured observation of behavioral self-regulation and its contribution to kindergarten outcomes. *Developmental Psychology*, 45(3), 605–619.
- Preschool Curriculum Evaluation Research Consortium. (2008). Effects of preschool curriculum programs on school readiness. Washington DC: United States Department of Education Institute of Educational Sciences.
- Ramani, G. B. (2012). Influence of a playful, child-directed context on preschool children's peer cooperation. *Merrill-Palmer Quarterly*, 58(2), 159–190.
- Ramey, C. T., & Campbell, F. A. (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency*, 88, 515–523.
- Raver, C. (2012). Low-income children's self-regulation in the classroom: Scientific inquiry for social change. *American Psychologist*, 11, 681–689.
- Raver, C., Blair, C., & Willoughby, M. (2013). Poverty as a predictor of 4-year-olds' executive function: New perspectives on models of differential susceptibility. *Developmental Psychology*, 49(2), 292–304.
- Raver, C., Jones, S., Li-Grining, C., Zhai, F., Bub, K., & Pressler, E. (2011). CSRP's impact on low-income preschoolers' preacademic skills: Self-regulation as a mediating mechanism.

- Child Development*, 82(1), 362–378.
- Rimm-Kaufman, S., Pianta, R., & Cox, M. (2000). Teachers' judgments of problems in the transition to kindergarten. *Early Childhood Research Quarterly*, 15(2), 147–166.
- Schlam, T., Wilson, N., Shoda, Y., & Mischel, W. (2013). Preschoolers' delay of gratification predicts their body mass index 30 years later. *Journal of Pediatrics*, 162(1), 90–93.
- Soares, D. A., Vannest, K. J., & Harrison, J. (2009). Computer aided self-monitoring to increase academic production and reduce self-injurious behavior in a child with autism. *Behavior Interventions*, 183(4), 171–183.
- Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods*, 5(1), 13–30.
- Thompson, A. M., Ruhr, L. R., Maynard, B. R., Pelts, M. D., & Bowen, N. K. (2013). Self-management interventions for reducing challenging behaviors among school-age students: A systematic review. *Campbell Systematic Reviews*, 1–50.
- Tough, P. (2012). *How children succeed*. Boston: Houghton Mifflin Harcourt.
- United States Department of Education. (2008). *What Works Clearinghouse: Tools of the Mind*. Washington DC.
- Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2014). Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods*, (10), 1–22.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36(3), 1–48.
- Vygotsky, L. (1962). *Thought and language*. Cambridge: MIT Press.
- Vygotsky, L. (1978). *Mind in society: The development of higher psychological processes*. Cambridge: Harvard University Press.
- Webster-Stratton, C., Reid, M., & Stoolmiller, M. (2009). Preventing conduct problems and improving school readiness: Evaluation of the Incredible Years teacher and child training programs in high-risk schools. *Journal of Child Psychology and Psychiatry*, 49(5), 471–488.

8 Information about this review

8.1 REVIEW AUTHORS

Lead review author:

The lead author is the person who develops and co-ordinates the review team, discusses and assigns roles for individual members of the review team, liaises with the editorial base and takes responsibility for the on-going updates of the review.

Name: Alex Baron

Title: Mr.

Affiliation: Oxford University

Address: 15 Norham Gardens

City, State, Province or County:
Oxford

Postal Code: OX2 6PY

Country: United Kingdom

Phone: +447796322776

Mobile: +447796322776

Email:
alexander.baron@education.ox.ac.uk

Co-author(s):

Name: Maria Evangelou

Title: Associate Professor

Affiliation: Oxford University

Address: 15 Norham Gardens

City, State, Province or County: Oxford

Postal Code: OX2 6PY

Country: United Kingdom

Phone: +441865274012

Mobile:

Email: maria.evangelou@education.ox.ac.uk

Co-author(s):

Name: Lars-Erik Malmberg

Title: Associate Professor

Affiliation: Oxford University

Address: 15 Norham Gardenns

City, State, Province or County:
Oxford

Postal Code: OX2 6PY

Country: United Kingdom

Phone: +441865274012

Mobile:

Email:
maria.evangelou@education.ox.ac.uk

Name: G.J. Melendez-Torres

Title: Assistant Professor

Affiliation: Warwick University

Address: Warwick University,
Coventry CV4 7AL

City, State, Province or County: Coventry

Postal Code: CV4 7AL

Country: United Kingdom

Phone: 07847360259

Mobile:

Email:
Melendeztorres.gj@gmail.com

8.2 ROLES AND RESPONSIBILITIES

- Content: Alex Baron and Maria Evangelou
- Systematic review methods: G.J. Melendez-Torres
- Statistical analysis: Lars-Erik Malmberg

- Information retrieval: Alex Baron and G.J. Melendez-Torres

8.3 SOURCES OF SUPPORT

The first author, Alex Baron, has funding through the United Kingdom's Economic and Social Research Council.

8.4 DECLARATIONS OF INTEREST

The authors have no conflicts of interest.

8.5 PLANS FOR UPDATING THE REVIEW

Alex Baron will be primarily responsible for the review updates, which will occur every three years following the review's publication in Campbell.

8.6 AUTHOR DECLARATION

Authors' responsibilities

By completing this form, you accept responsibility for maintaining the review in light of new evidence, comments and criticisms, and other developments, and updating the review at least once every five years, or, if requested, transferring responsibility for maintaining the review to others as agreed with the Coordinating Group. If an update is not submitted according to agreed plans, or if we are unable to contact you for an extended period, the relevant Coordinating Group has the right to propose the update to alternative authors.

Publication in the Campbell Library

The Campbell Collaboration places no restrictions on publication of the findings of a Campbell systematic review in a more abbreviated form as a journal article either before or after the publication of the monograph version in *Campbell Systematic Reviews*. Some journals, however, have restrictions that preclude publication of findings that have been, or will be, reported elsewhere, and authors considering publication in such a journal should be aware of possible conflict with publication of the monograph version in *Campbell Systematic Reviews*. Publication in a journal after publication or in press status in *Campbell Systematic Reviews* should acknowledge the Campbell version and include a citation to it. Note that systematic reviews published in *Campbell Systematic Reviews* and co-registered with the Cochrane Collaboration may have additional requirements or restrictions for co-publication. Review authors accept responsibility for meeting any co-publication requirements.

I understand the commitment required to update a Campbell review, and agree to publish in the Campbell Library. Signed on behalf of the authors:

Form completed by: Alex Baron

Date: 26 November 2016

9 Tables

Table 1: List of included studies and records

Studies included for the systematic review	
Study ID	Records
Barnett (2008)	Diamond et al. (2007)
	Center on the Developing Child (2008)
	Barnett et al. (2008)
	Stechuk (2009)
Lonigan (2012)	Lonigan & Phillips (2012)
Blair (2014)	Blair & Raver (2014)
	United States Department of Education (2015)
Clements (2014)	Clements & Sarama (2012)
	Clements & Sarama (2014)
Farran (2014)	Wilson & Farran (2012)
	Farran & Wilson (2014)
Morris (2014)	Mattera et al. (2013)
	Hseuh et al. (2014)
	Morris et al. (2014)

Table 2: Robust variance estimation estimates across the four outcome measures

Outcome	n(k)	Effect size	SE	p-value	95% C.I.
Reported SR	12(3)	0.121	0.118	0.415	(-.387, .628)
Task-based SR	36(5)	0.072	0.079	0.418	(-.149, .293)
Literacy	43(5)	0.027	0.027	0.379	(-.049, .103)
Math	12(3)	0.061	0.019	0.035	(.007, .115)

(Note: 'n' signifies the number of effect sizes; k signifies the number of studies from which those effect sizes were drawn; 'effect size' signifies the pooled effect size across all studies)

Table 3: Robustness check with multilevel meta-analysis in the metafor R package

Outcome	n(k)	Effect size	SE	p-value	95% C.I.
Reported SR	12(3)	0.043	0.029	0.242	(-.023, .091)
Task-based SR	36(5)	0.119	0.079	0.418	(-.149, .293)
Literacy	43(5)	0.017	0.023	0.468	(-.029, .062)
Math	12(3)	0.053	0.03	0.084	(-.006, .111)

CHARACTERISTICS OF INCLUDED STUDIES

Barnett et al., 2008

Methods	Randomized controlled trial
Participants	Barnett et al. (2008) randomly assigned 210 preschool children (54% age four, 46% age three) to either the Tools group (n = 88) or the control group (n = 122).
Interventions	<p>The intervention for Barnett, 2008, and in all the sections below, is the Tools program. Since the nature of the Tools program has been extensively discussed, the sections here do not include additional detail about Tools. Instead, the ‘Intervention’ portions in these tables provide detail regarding the professional development regimen for Tools teachers, the timeline of the study, and information about Tools implementation in the particular study context (e.g., Tools could have been partially implemented in one of the studies, which would be noted in the structured summaries here).</p> <p>Tools teacher received four days of Tools curriculum training before the start of the school year. During the school year, certified Tools trainers also visited the classrooms once per week. Child-level data were collected during the first year of Tools implementation, so teachers had no training year. At the time of this study, Tools had approximately 40 activities (Barnett et al., 2008, p. 301) instead of the 61 activities the program currently includes.</p> <p>The control classrooms used a curriculum developed by the school district in the three years prior to the study. According to the authors (Barnett et al., 2008), “there was a greater emphasis on teacher-imposed control and less on children regulating each other and themselves” (p. 303). Information regarding the professional development training for control classroom teachers was not provided in the study.</p>
Outcomes	For academic measures, the authors administered the Woodcock Johnson Applied Problems and Letter-Word Identification subtests, the Peabody Picture Vocabulary Test (PPVT – III), the Expressive One Word Picture Vocabulary Test – Revised (EOWPVT – R), the Oral Language Proficiency Test, and the Weschler Preschool Primary Scale of Intelligence (WIPPSI). For self-regulation measures, the authors used the Social Skills Rating Scale (SSRS). Executive function data were collected in the form of both the Dots and

	Flanker tasks. Children took a pre-test and a post-test for each measure, but the timing of those assessment windows was not reported (in the Diamond et al., 2007 paper where the executive function data were included).
Notes	

Blair & Raver, 2014

Methods	Cluster randomized controlled trial
Participants	Blair & Raver (2014) randomly assigned 759 kindergarten children (age statistics not reported) in 79 classrooms in 29 schools to either a Tools group (n = 443) or control group (n = 316).
Interventions	<p>Teachers implemented the Tools curriculum in a two-year professional development cycle. In year one, teachers received five days of training. In year two, teachers received three days of training. Each school also had a Tools coach who provided feedback to teachers once per fortnight in year one and once per month in year two. The intervention was only delivered during children's kindergarten year of school.</p> <p>Control group teachers continued business-as-usual practice and professional development training during the two years of the study. According to the study authors, control group classrooms used "commercial literacy and mathematics curricula" that were aligned with state standards (Blair & Raver, 2014, p. 4).</p>
Outcomes	For academic measures, the authors administered the Woodcock Johnson Applied Problems and Letter-Word Identification subtests, the Peabody Picture Vocabulary Test (PPVT – III), the Expressive One Word Picture Vocabulary Test – Revised (EOWPVT – R), the Oral Language Proficiency Test, and the Weschler Preschool Primary Scale of Intelligence (WIPPSI). For self-regulation measures, the authors used the Social Skills Rating Scale (SSRS).
Notes	

Clements et al. (2014)

Methods	Cluster randomized controlled trial
Participants	Clements et al. (2014) randomly assigned 826 children in 84 "four-year-old classrooms" (Douglas Clements & Sarama, 2012, p. 2) to one of three conditions: Building Blocks math curriculum, Tools of the Mind plus Building Blocks combined curriculum, or business-as-usual. Since this meta-analysis pertains to the Tools curriculum, only the Tools students (n = 288) and business-as-usual students (n = 273) will be referred to hereafter.

Interventions	<p>The study spanned three years. In the first year, teachers began implementing the curriculum to which they had been randomly assigned, but no student data were collected. When the next cohort of pre-kindergarten students arrived in year two, those students were assessed in all measures during the fall and spring of pre-kindergarten. Finally, in year three, all children in all conditions reverted to their business-as-usual curricula (i.e., neither Tools nor Building Blocks were implemented), and follow-up data were collected for all children in all measures. For both years of Tools implementation, teachers received six days of professional development.</p> <p>Importantly, this study is the only trial in this meta-analysis for which Tools was not implemented on its own in any experimental condition. That is, in this study, Tools was implemented as part of a composite curriculum that included both Tools and the Building Blocks math curriculum. As such, the unique effects of the Tools curriculum vis-à-vis the business-as-usual curriculum cannot be ascertained in this study.</p> <p>The control classrooms continued the business-as-usual math curricula used by the three school districts in this study. Specifically, one district used <i>Everyday mathematics</i> (McGraw-Hill), one used <i>Developing math concepts in pre-kindergarten</i> (from Math Perspectives), and the third had no uniform math curriculum used throughout their schools (Clements et al., 2014, p. 18).</p>
Outcomes	<p>For self-regulation, the researchers used Heads-Toes-Knees-Shoulders (HTKS), Peg Tapping, Forward and Backward Digit Span, Self-Ordered Pointing, and the Item Selection tasks. For math skills, the researchers used the Tools for Early Assessment of Mathematics (TEAM) and the mathematics portion of the Early Childhood Longitudinal Study (ECLS) cognitive assessment. For literacy, the researchers used Alphabet Knowledge and Name Writing subtests of the Phonological Awareness Literacy Screening (PALS), the Expressive Vocabulary Test (EVT) for vocabulary, and the Refrenew Bus Story to measure oral language and narrative retell.</p>
Notes	

Farran & Wilson, 2014

Methods	Cluster randomized controlled trial
Participants	Farran and Wilson (2014) randomly assigned 877 preschool children (Mean age = 54 months) in 60 classrooms in 59 schools to the Tools condition (n = 646 children) or the control condition (n = 499 children).
Interventions	Teachers implemented the curriculum in a two-year cycle. In the first year, teachers received Tools professional development training (amount is unreported), but no outcome data were collected. In the second year, teachers received more professional development training (amount is unreported), and child outcome data was

	<p>collected. The Tools intervention was only implemented during children’s pre-kindergarten year of school.</p> <p>Control group teachers continued business-as-usual practice and professional development training during the study. The study took place in five school districts, so “the comparison classrooms used a variety of curricula, with the modal one being Creative Curriculum” (Farran & Wilson, 2014, p. 11).</p>
Outcomes	<p>For self-regulation, the researchers used the researcher-reported Self-Regulation Assessor Rating (SAR) and the teacher-reported Cooper-Farran Behavioral Rating Scale (CFBRS). In addition to those two informant-report measures, the researchers also used the Peg Tapping, Heads-Toes-Knees-Shoulders, Corsi Blocks, Copy Design, and DCCS tasks to measure executive function. For academic skills, the researchers used seven Woodcock Johnson III subtests: Letter Word, Applied Problems, Oral Comprehension, Spelling, Picture Vocabulary, Academic Knowledge, and Quantitative Concepts.</p>
Notes	

Lonigan & Phillips, 2012

Methods	Cluster randomized controlled trial
Participants	<p>Lonigan & Phillips (2012) randomly assigned 2,564 children (m age = 52.7 months, SD = 6.37) in 117 preschool centers to one of four conditions: Tools, Literacy Express Comprehensive Preschool Curriculum (LECPC), a combined curriculum with both Tools and LECPC, and ‘business-as-usual.’</p>
Interventions	<p>Teachers in the Tools-only condition implemented the entire Tools program, whereas teachers in the Tools-LECPC combined curriculum only implemented Tools’ make-believe play block activities (see section 1.3.1). Lonigan and Phillips (2012) state that teachers in both Tools conditions received professional development to support “sophisticated and self-regulated play by the children” (p. 3), but the study does not indicate how much training the teachers received. Each classroom maintained its condition assignment for two years, and data were collected across two sequential cohorts of students for each classroom. That is, each teacher delivered his or her target curriculum over two years with two different groups of students.</p> <p>Control classrooms continued their ‘business-as-usual’ practice throughout the two years of the study. Lonigan and Phillips (2012) indicated that ‘business-as-usual’ classrooms mostly used the HighScope or Creative Curriculum classrooms (see Chapter Seven for more information on these two curricula).</p>
Outcomes	<p>For self-regulation measures, the authors used the Heads-Toes-Knees-Shoulders task as well as the Behavioral Rating Inventory of Executive Function – Preschool (BRIEF-P) to rate children’s executive function. For academic measures, the authors used the</p>

	Bracken Basic Concept Scales – Revised (BBCS – R) as well as the Test of Preschool Early Literacy (TOPEL). The BBCS – R assesses children in six areas: colors, shapes, counting, letters, size, and comparisons. The TOPEL has four subscales: print knowledge, definitional vocabulary, blending sounds, and elisions (Lonigan & Phillips, 2012, p. 4).
Notes	Despite numerous attempts to contact the authors, the requisite data were not made available. Thus, this study was included in the systematic review but excluded from the meta-analysis.

Morris et al., 2014

Methods	Cluster randomized controlled trial
Participants	2,670 children in 307 classrooms in 104 preschool centers were randomly assigned to one of four conditions: Tools of the Mind, Incredible Years (IY), Promoting Alternative Thinking Strategies (PATHS), or business-as-usual. All reported comparisons were between an intervention group and business-as-usual; thus, no comparisons of Tools with the Incredible Years or PATHS program were reported. Thus, since this meta-analysis pertains to the Tools curriculum, only the Tools students (n = 678) and business-as-usual students (n = 676) will be referred to hereafter.
Interventions	<p>Tools training, implementation, and data collection took place in the course of one school year. Nonetheless, the researchers refer to the “comprehensive professional development system for teachers – including four to six training sessions, weekly coaching sessions in the classroom, a ‘real-time’ managing information system (MIS) to support monitoring, and technical assistance” (Morris et al., 2014, p. 2) to support robust implementation across all sites.</p> <p>The control classrooms continued business-as-usual practice and received no additional professional training above their usual schedule. Of business-as-usual classrooms, 88% used either Creative Curriculum or HighScope.</p>
Outcomes	For self-regulation, the researchers used pencil tapping, the Social Skills Rating Scale (SSRS), the Behavioral Problems Index (BPI), and the Cooper-Farran Behavioral Rating Scale (CFBRS). For academic skills, the researchers used 1) the Woodcock Johnson III Letter Word and Applied Problems subtests, 2) the Academic Rating Scale (ARS) Language and Literacy, Mathematical Knowledge, and General Knowledge subtests, and 3) the Expressive One Word Picture Vocabulary Test (EOWPVT).
Notes	

RISK OF BIAS FOR INCLUDED STUDIES

Barnett *et al.*, 2008

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The randomization was by computer generated sequence" (Author email).
Allocation concealment (selection bias)	Low risk	"The researcher who conducted the assignments was the project coordinator who was responsible for organizing data collection; that person was not involved in other aspects of the research including specification of hypotheses, design or analysis" (Author email).
Blinding of participants and personnel (performance bias)	Unclear risk	It is not possible to blind teachers and students to their condition assignment.
Blinding of outcome assessment (detection bias)	High risk	An email exchange with one of the authors indicated that "The intent was for testers to be blinded to condition, but the testers said they could tell which children were <i>Tools</i> because when it came to the most difficult test conditions, control children tended to give up, but <i>Tools</i> children kept saying, "I know I can do this" (author email). Moreover, teachers conducted ratings for the students' self-regulation. Thus, teachers were not blind to children's curricular assignment, nor could they have been.
Incomplete outcome data (attrition bias)	Low risk	"Among those who consented to the study, attrition was relatively minor. One child in each group moved out of the district prior to assessment. This left us with an initial sample of 218 children: 92 (42%) in <i>Tools</i> and 126 (58%) in the control group. Of these, four in each group were not tested in the Fall, due to the child's absence or discomfort with the testing situation. By Spring post-test, another six children in the <i>Tools</i> group and five children in the control group had moved. One child in each group was not tested due to absences so that 85 <i>Tools</i> (92%) and 120 control (95%) children were assessed in the Spring. It was not possible to conduct extensive analyses of attrition, because most attrition in this study was due to lack of active consent from parents prior to any data collection. However, we do know gender, ethnicity, and home language for most of the original sample children. Thus, it was possible to test for differences between those whose parents agreed to participate and those whose parents declined or did not respond. Analysis of Variance

		revealed no statistically significant main effects of attrition or interactions between attrition and treatment (curriculum assignment)" (303).
Selective reporting (reporting bias)	Low risk	All measured identified in the methodology section are reported in the results section.
Other bias	Low risk	No other sources of potential bias were identified.

Blair & Raver, 2014

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"The randomization was computer generated" (Author email).
Allocation concealment (selection bias)	Low risk	"The randomization was conducted independently by someone not associated with study" (Author email).
Blinding of participants and personnel (performance bias)	Unclear risk	It is not possible to blind teachers and students to their condition assignment.
Blinding of outcome assessment (detection bias)	High risk	"The outcome assessors may have been aware of the group assignment of the school. I can't say for sure, one way or the other, but I expect that some of them were" (Author email).
Incomplete outcome data (attrition bias)	Low risk	"I did [assess differences between attrited and non-attrited students] and differences were minimal" (Author email).
Selective reporting (reporting bias)	Low risk	All measured identified in the methodology section are reported in the results section.
Other bias	Low risk	No other sources of potential bias were identified.

Clements et al., 2014

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"Schools/centers were randomly assigned to the three conditions three at a time starting at a randomly chosen point in the sorted list and then moving to the top of the list. This is an application of the systematic circular sampling scheme (Lahiri, 1951), which was utilized to ensure three experimental groups that are balanced geographically and in terms of the length of the Pre-K program and key background characteristics of the schools/centers" (8).
Allocation concealment (selection bias)	Unclear risk	Not reported

Blinding of participants and personnel (performance bias)	Unclear risk	It is not possible to blind teachers and students to their condition assignment.
Blinding of outcome assessment (detection bias)	Unclear risk	Not enough information to judge.
Incomplete outcome data (attrition bias)	High risk	The authors mention substantial attrition from the study but do not analyze its impact on the results. "Table 5 presents the corresponding results. The first column in this table shows that the size of the analytic samples for these analyses is roughly ten percent smaller than those of the Spring 2011 measures, where the reduction in the sample size is mostly due to children's mobility between the two time points" (32-33)
Selective reporting (reporting bias)	Low risk	All measured identified in the methodology section are reported in the results section.
Other bias	Low risk	No other sources of potential bias were identified.

Farran & Wilson, 2014

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Low risk	"We used a computer random number generator (in excel) to perform the randomization" (author email).
Allocation concealment (selection bias)	Low risk	"All schools were recruited prior to assignment and all schools were randomized in a single randomization using the procedure described above. So, because knowledge of one assignment could not have affected recruitment or future assignments, allocation was effectively concealed – schools and the researchers were unaware of assignments or upcoming assignments because it was all done at once" (author email).
Blinding of participants and personnel (performance bias)	Unclear risk	It is not possible to blind teachers and students to their condition assignment.
Blinding of outcome assessment (detection bias)	High risk	"I think that for the most part assessors were blind to condition when completing the SAR. We did have some assessors who had also been observers in the pre-K classrooms, so if the assessor went to assess some children in the same classroom that they had observed previously, it would have been obvious to the assessor that those children were in a Tools or Control classroom. Also, I guess, the assessor could have noticed Tools materials, centers, etc. in the classroom when they went to pull the child for the assessment. But the assessment materials (roster of children's names and filemaker

		system for collecting assessment data) did not indicate if the classroom was Tools or Control. Also, this would have only occurred during the pre-K assessments, in kindergarten and first grade the children had moved into different classrooms and so assessors wouldn't have known if they were in Tools or Control during their pre-K year." (author email)
Incomplete outcome data (attrition bias)	Low risk	"Attrition during the study was minimal. No teachers dropped out during the test year. Attrition of students over the course of the study was low and similar across Tools and comparison classrooms" (11); "There were no statistically significant differences in attrition by condition" (11)
Selective reporting (reporting bias)	Low risk	All measured identified in the methodology section are reported in the results section.
Other bias	Low risk	No other sources of potential bias were identified.

Lonigan & Phillips, 2012

Bias	Authors' judgement	Support for judgement
Random sequence generation (selection bias)	Unclear risk	Not reported
Allocation concealment (selection bias)	Unclear risk	Not reported
Blinding of participants and personnel (performance bias)	Unclear risk	It is not possible to blind teachers and students to their condition assignment.
Blinding of outcome assessment (detection bias)	High risk	Teachers completed executive function ratings for children in their own classrooms. "In addition to measures of children's academic outcomes, children's classroom teachers completed the Behavioral Rating Inventory of Executive Function - Preschool version" (3).
Incomplete outcome data (attrition bias)	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	All measured identified in the methodology section are reported in the results section.
Other bias	Low risk	No other sources of potential bias were identified.

Morris et al., 2014

Bias	Authors' judgement	Support for judgement
-------------	---------------------------	------------------------------

Random sequence generation (selection bias)	Unclear risk	Not reported
Allocation concealment (selection bias)	Unclear risk	Not reported
Blinding of participants and personnel (performance bias)	Unclear risk	It is not possible to blind teachers and students to their condition assignment.
Blinding of outcome assessment (detection bias)	High risk	Teachers completed the Cooper-Farran Behavioral Ratings Scale (CFBRS) self-regulation assessments for children in their class
Incomplete outcome data (attrition bias)	Unclear risk	Not reported
Selective reporting (reporting bias)	Low risk	All measured identified in the methodology section are reported in the results section.
Other bias	Low risk	No other sources of potential bias were identified.

CHARACTERISTICS OF EXCLUDED STUDIES

Bodrova & Leong 2001

Reason for exclusion	Qualitative study without data on the target outcome measures
-----------------------------	---

Bodrova & Leong 2011

Reason for exclusion	Theoretical paper with no quantitative data
-----------------------------	---

Copple 2003

Reason for exclusion	Theoretical paper with no quantitative data
-----------------------------	---

Grigorenko 1998

Reason for exclusion	Not about the Tools of the Mind curriculum but rather Vygotsky's theoretical ideas
-----------------------------	--

Hammer 2012

Reason for exclusion	Study has not yet been conducted and thus has not produced results
-----------------------------	--

Hyson 2006

Reason for exclusion	A qualitative case study of Tools as part of a larger book chapter about child development
-----------------------------	--

Mackay 2013

Reason for exclusion	Doctoral dissertation with a non-experimental design that does not control for potential statistical confounds
-----------------------------	--

Magalhaes 2013

Reason for exclusion	Qualitative doctoral dissertation without quantitative data on the target outcome measures
-----------------------------	--

Millaway 2015

Reason for exclusion	Doctoral dissertation with a non-experimental design that did not control for potential statistical confounds
-----------------------------	---

Rodgers 2012

Reason for exclusion	Qualitative doctoral dissertation without quantitative data on the target outcome measures
-----------------------------	--

Shaheen 2014

Reason for exclusion	Review study with no original quantitative data
-----------------------------	---

10 Figures



Figure 1: Sample Tools of the Mind play plan

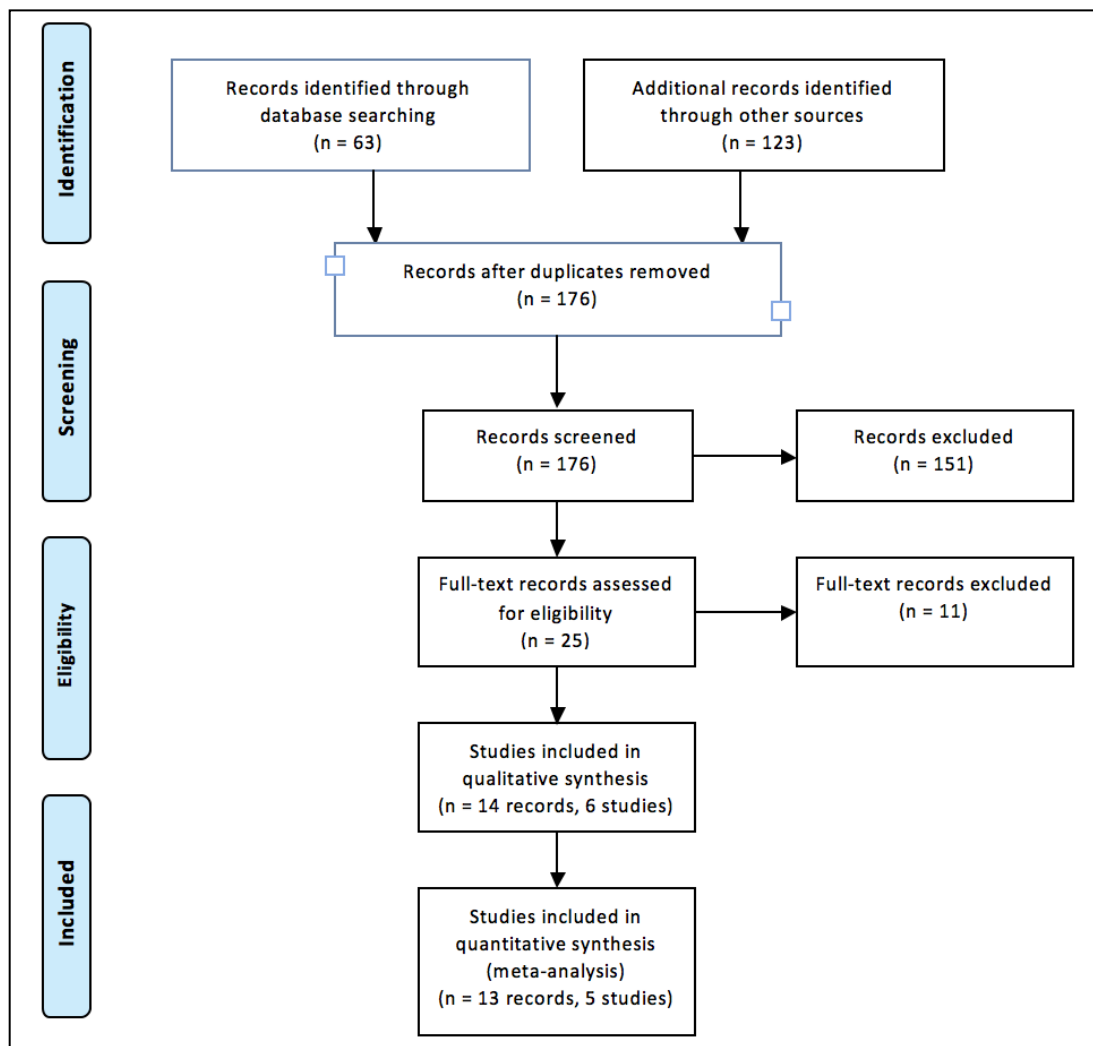


Figure 2: Systematic review flowchart

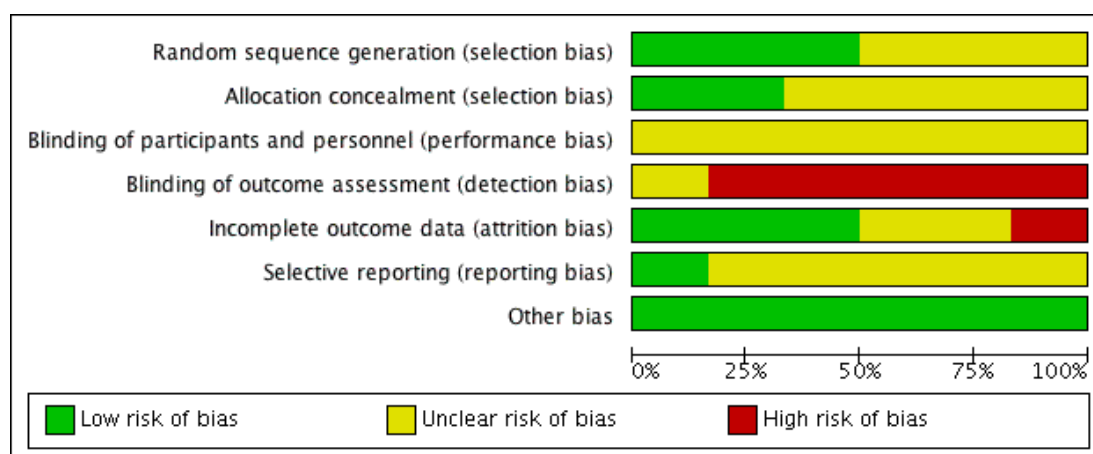
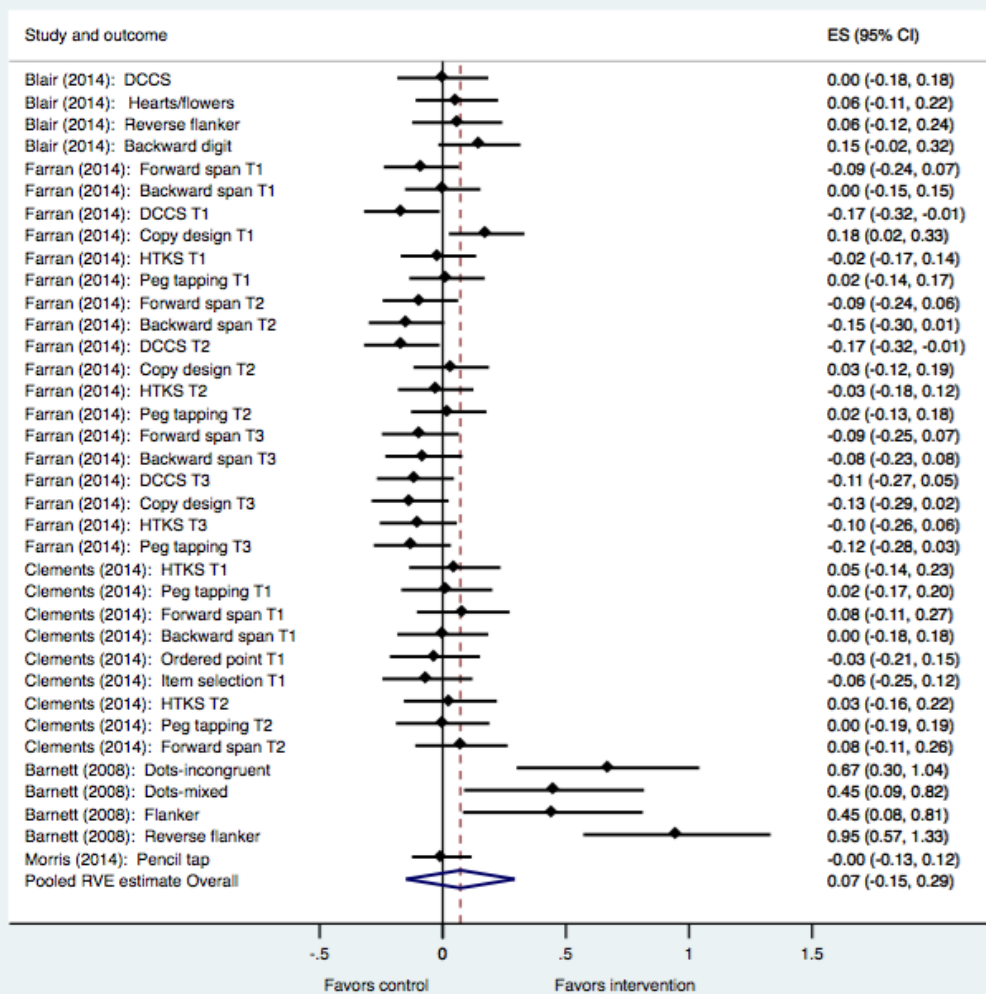


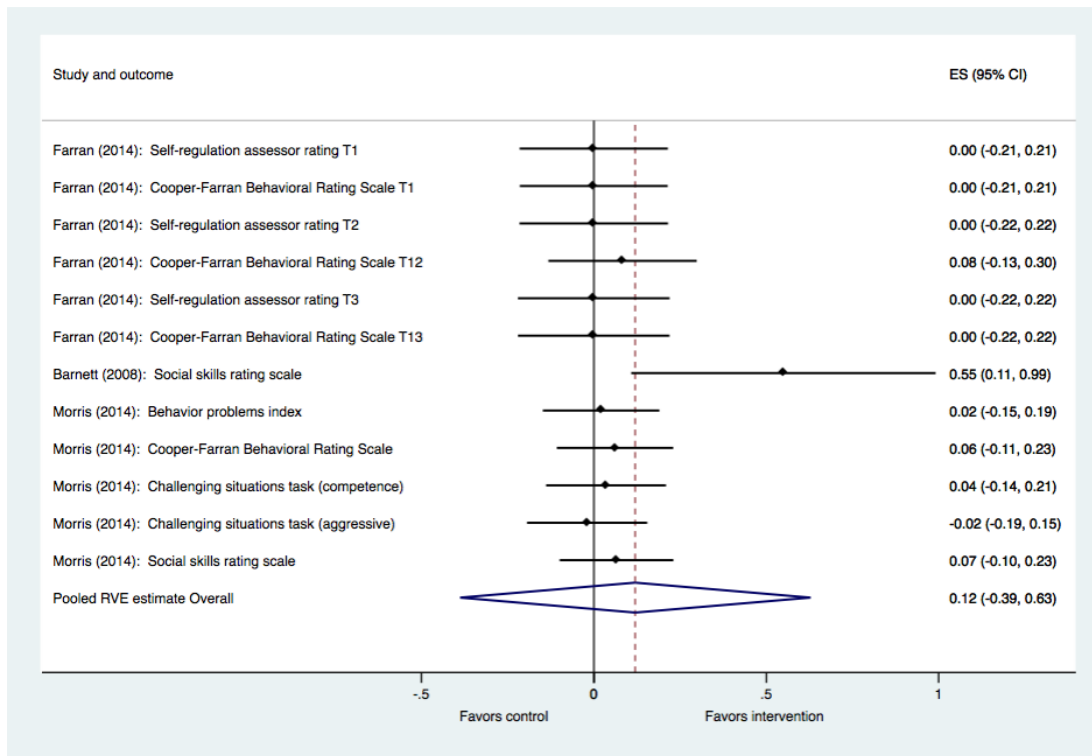
Figure 3: Risk of bias summary

11 Data and analyses

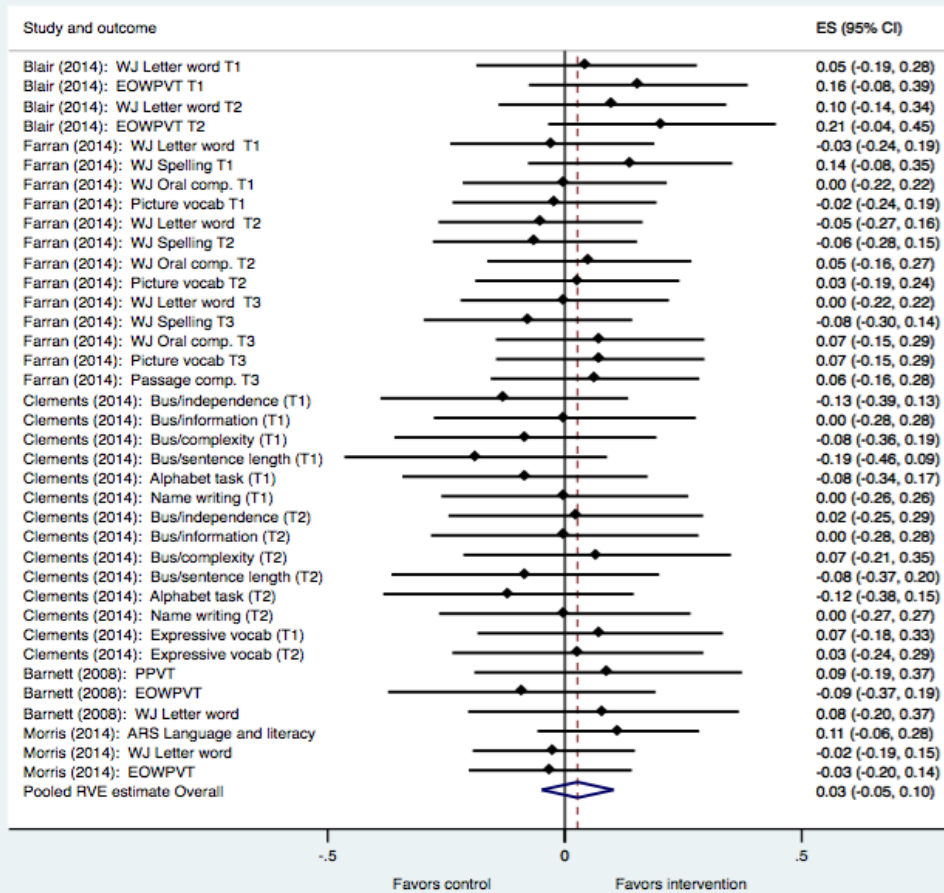
Analysis 1: Task-based self-regulation forest plot



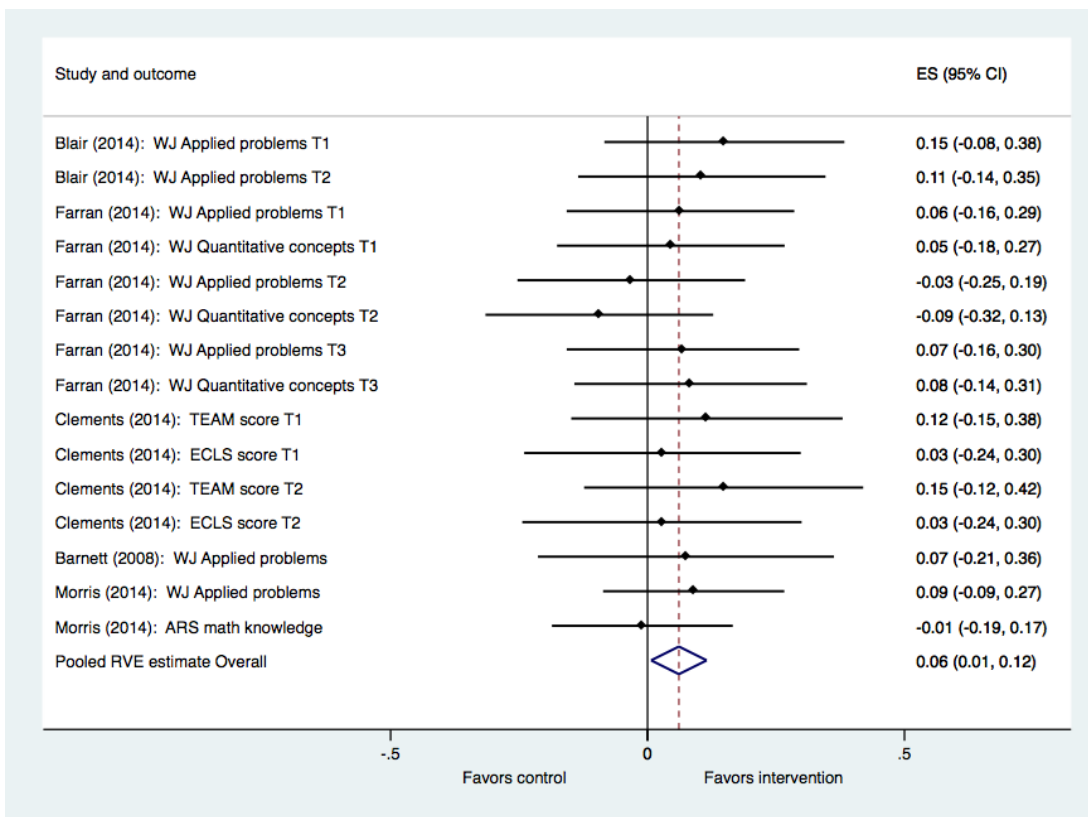
Analysis 2: Reported self-regulation forest plot



Analysis 3: Literacy forest plot



Analysis 4: Math forest plot



12 SPSS and R code for analyses

We used the robust variance estimation (RVE) macro in SPSS (IBM, 2012) with the following syntax.

```
DEFINE ROBUST ( STUDYID !CHAREND ("/")
               / EFFSIZE !CHAREND ("/")
               / VAREFFS !CHAREND ("/")
               / RHO    !CHAREND ("/") !DEFAULT ("")
               / DESIGN !CHAREND ("/") !DEFAULT ("")
               / WEIGHTS !CHAREND ("/") !DEFAULT ("")
               / RESID  !CHAREND ("/") !DEFAULT ("")
               / HWEIGHT !CHAREND ("/") !DEFAULT ("")
               / PRINT  !CHAREND ("/") !DEFAULT (DEF) ).
PRESERVE.
SET MPRINT OFF .
SET PRINTBACK OFF.
```

```
ROBUST STUDYID = studyid / EFFSIZE = es / VAREFFS = var / RHO = .8 .
```

Sample output from the RVE syntax with the math pooled effect size is below seen below.

```
Parameter Estimates and Robust Standard Errors
      Coef      SE      T  Pr > |T|  95% Conf  Interval
INTERCEP  .060563  .019468  3.110938  .035839  .006512  .114614

N Level 1
  15

N Level 2
  5

Average Level 1 N
  3.00

T-Test DF
  4

Tau-squared estimate
 .000000

Assumed Rho
  .80
```

Weighted Residual Sum of Squares Qe 3.234

Specifically, the output shows the parameter estimates in the first row. The estimates include the pooled effect size estimate, its standard error, the t- and p-values, and the 95% confidence interval.

The tau-squared (τ^2) value of zero in the RVE indicates the proportion of shared variance across effect sizes above and beyond that which would be expected by sampling error. Instead, the shared variation exists among effect sizes nested within the same cluster. Thus, we specified a rho value of .80 in the last line of the syntax, which signifies a very high dependency among effect sizes from the same study. This high inter-correlation value imposes a conservative estimation process on the analysis, which, in turn, reduces the likelihood of a Type I error. The rho value of .80 is the recommendation of the SPSS coders who created the RVE macro (Tanner-Smith & Tipton, 2014) and has been also recommended in other RVE literature (Larry V. Hedges, Tipton, & Johnson, 2010).

Nonetheless, we also performed a robustness check with Cohen's (1988) recommendations of .2, .5, and .8. Thus, we tested the model with low, medium, and high levels of assumed inter-correlation. Neither the beta coefficients nor the significance values changed across models, so the results from the RVE can be said to be robust

In addition, the RVE analysis, the meta-analysis was conducted through a multilevel framework as an additional robustness check. The multilevel meta-analysis was conducted using the R packaged called metafor (Viechtbauer, 2010). The syntax is depicted below.

```
> library(metafor)
> obj <- read.csv("/Users/abaro2/Documents/DPhil/DPhil Writing/Meta-
analysis/MLM_ES_table_Obj.csv")
> View(obj)
> MLM <-
rma.mv(yi=effectsize,V=var,data=obj,random=list(~1|esid,~1|studyid))
> summary(MLM)
```

Sample output from R Studio for the multilevel meta-analysis model with the assessor-reported self-regulation data is as follows:

```
Multivariate Meta-Analysis Model (k = 12; method: REML)
logLik Deviance AIC BIC AICc
13.7438 -27.4876 -21.4876 -20.2939 -18.0590
Variance Components:
estim sqrt nlvs fixed factor
sigma^2.1 0.0000 0.0000 12 no esid
sigma^2.2 0.0710 0.2664 3 no studyid

Test for Heterogeneity:
```

```

Q(df = 11) = 15.8510, p-val = 0.1468
Model Results:
estimate    se    zval    pval    ci.lb    ci.ub
0.1722  0.1607  1.0716  0.2839 -0.1427  0.4871

```

In the output above, the log-likelihood (loglik), Deviance, AIC, BIC, and AICc are all fit indices to compare the relative appropriateness of nested model specifications. This meta-analysis does not compare any nested models with various predictors because meta-regression was not conducted; thus, the model fit indices do not provide useful information for these meta-analytic results.

Below the model fit indices section in the output, the variance components row indicates the amount of variance observed at different levels of analysis. Those components are denoted as σ^2 in R Studio, whereas some other texts and software packages refer to those components as tau-squared (τ^2) values. In R Studio, the σ^2 at level one indicates the amount of shared variance among effect sizes from all studies, whereas σ^2 at level two indicates the amount of shared variance among effect sizes from the same study.

Thus, as we would expect, there are small but observable values of shared variation among effect sizes from the same studies (σ^2 at level two) because those effect sizes are based on information from the same participants. However, they are capturing different pieces of information about the participants, so we would not expect their shared variation to be extremely high. In the output above, 7.1% is the amount of shared variation among effect sizes from the same cluster (i.e., study).

By contrast, we would not expect any additional shared variation among all effect sizes from all studies at level one. Thus, σ^2 at level one is, as expected, zero. Once again, this number quantifies the amount of shared variation across all effect sizes that is observed above and beyond the prediction of sampling error. Since there is no reason to expect shared variation among the twelve effect sizes in the analysis above and beyond that among effect sizes clustered within the same study, the σ^2 value is 0.

Beneath the variance components analysis, we observe the Q-statistic value. The Q-statistic in the output above is relatively small and not statistically significant, which indicates that different studies did not reach significantly different conclusions regarding Tools' effectiveness on children's assessor-based self-regulation scores.

Finally, in the results above, the final row entitled 'Model Results' indicates the pooled effect size, its standard error, the Z- and p-values, and the 95% confidence interval. The output indicates a small to moderate effect size ($g = .17$) for assessor-reported self-regulation with a confidence interval that crosses zero, which indicates a lack of statistical significance (the p-value is .28).

13 Search notes from databases

The systematic electronic database search was conducted on October 21, 2016 beginning at 9 am British Standard Time in the United Kingdom. Whereas the search results may have changed since then (i.e., newer papers could have been added or removed from databases), the results presented below represent exactly what was recovered on the day of the search.

Applied Social Sciences Index and Abstracts (ProQuest)

Search term: AB("Tools of the Mind" OR TI("Tools of the Mind"))

Results: 2 hits

CENTRAL (Cochrane Library)

Search term: "Tools of the Mind"

Results: 0 hits

Embase (Ovid: 1947 to October week 2 2016)

Search term: "Tools of the Mind"

Results: 0 hits

ERIC (ProQuest)

Search term: AB("Tools of the Mind" OR TI("Tools of the Mind"))

Results: 22 hits

LILACS (<http://lilacs.bvsalud.org/en/>)

Search term: "Tools of the Mind"

Results: 0 hits

MEDLINE (Ovid: 1946 to 20 October 2016)

Search term: "Tools of the Mind"

Results: 4 hits

OpenGrey (www.opengrey.eu/)

Search term: "Tools of the Mind"

Results: 0 hits

PsycINFO (Ovid: 1967 to October week 2 2016)

Search term: “Tools of the Mind”

Results: 22 hits

ProQuest Dissertations and Theses (ProQuest)

Search term: AB(“Tools of the Mind” OR TI(“Tools of the Mind”)

Results: 7 hits

Social Sciences Citation Index (ProQuest)

Search term: AB(“Tools of the Mind” OR TI(“Tools of the Mind”)

Results: 6 hits

Sociological Abstracts (ProQuest)

Search term: AB(“Tools of the Mind” OR TI(“Tools of the Mind”)

Results: 0 hits

TOTAL HITS = 63

14 Coding manual

Item	Description	Notes
Section 1: Study Identification		
1	Study ID:	
2	Author(s) and year: e.g., Bodrova & Leong, 2007	
3	Type of report (select one) 1) Journal article 2) Book/book chapter 3) Government report (e.g., federal, state, local) 4) Thesis or dissertation 5) Conference proceedings 6) Unpublished past report (e.g., non-government technical report) 7) Unpublished in press/in progress manuscript 8) Other (specify)	
Section 2: Study Context		

1	Country in which the study was conducted: 1) USA 2) Canada 3) Chile 4) Other country (specify) 5) Cannot tell	
2	Regional location of the research site: 1) Suburban 2) Urban 3) Rural 4) Mixed 5) Cannot tell	
Section 3: Sample Description		
1	Number of students (for treatment group, comparison group, and total)	
2	Child gender (0 = female, 1 = male)	
3	Child age (0 = pre-kindergarten, 1 = kindergarten)	
4	Special education status (0 = no, 1 = yes)	
5	Ethnicity information (as described in the study)	
6	Socio-economic status (as described in the study)	
7	English language learners (as described in the study)	

8	Participant attrition rate (treatment group, comparison group, or two groups combined)	
9	Reason for attrition (as described in the study)	
Section 4: Description of intervention and comparison condition		
1	Comparison condition (as described in the study)	
2	<p>Were efforts made to monitor and measure fidelity of implementation?</p> <ol style="list-style-type: none"> 1) Yes (how) <ul style="list-style-type: none"> • Observations • Interviews of participants • Surveys of participants • Participant logs • Administrative records • Checklists • Other 2) No 	
3	Duration/frequency of Tools implementation (as described in the study)	
Section 5: Research Design		
1	<p>Research design type:</p> <ol style="list-style-type: none"> 1) Experimental design (included randomized controlled trials or cluster-randomized trials) 2) Quasi-experimental design– Regression discontinuity, differences-in-differences, instrumental variables 3) Quasi-experimental design– two groups, pre-and post-test design 	

	4) Quasi-experimental design— two groups, post-test only (no pre-test) 5) Longitudinal study—outcomes were measured at least twice after intervention	
2	Unit of assignment to conditions: 1) Individual 2) Group/cluster/sites (specify)	
3	Unit of analysis: 1) Individual 2) Group/cluster/sites (specify)	
4	Method of assignment to conditions: 1) Completely random 2) Random after matching, stratification, blocking, etc. 3) Quasi-random-assigned by some naturally existing situations 4) Nonrandom, but matched or statistically controlled on major characteristics or pretest measures	
5	If matching was used, how were the groups matched? (select all that apply) 1) Matched on pretest measures 2) Matched on demographics or other major features 3) Propensity score matching	
6	Were the participants (i.e., teachers and children) blinded to their conditions? 1) Yes 2) No	
7	Was the data collector blind to the group assignment? 1) Yes 2) No	

8	Results of statistical comparisons of pre-intervention group differences <ol style="list-style-type: none"> 1) No statistically significant differences 2) Statistically significant differences 3) No comparisons were made 	
9	Upon what kind of the statistical analyses were the major findings of the original study based? <ol style="list-style-type: none"> 1) Descriptive analysis 2) <i>t</i>-tests 3) ANOVA/MANOVA 4) ANCOVA/MANCOVA 5) Regression/multiple regression 6) Factor analysis 7) Path analysis 8) Multilevel modeling 9) Structural equation modeling (SEM) 10) Other (specify) 	
Section 6: Outcome Measures		
1	Outcome measures (select all that apply) <ol style="list-style-type: none"> 1) Achievement/learning outcome measures (e.g., standardized test scores, course grades) 2) Performance-based executive function tests (e.g., inhibitory control, working memory, cognitive flexibility) 3) Rating scales, survey, questionnaire, and checklist 4) Behavioral observation 	
2	Source of outcome data: <ol style="list-style-type: none"> 1) Child 2) Parent report 3) Teacher report/caregiver report 4) Other 	

4	<p>Were the reliability and validity of the outcome measures reported in the study?</p> <p>1) Yes (specify)</p> <p>2) No</p>	
5	<p>When did the post-test measure(s) take place?</p> <p>1) Immediately following the intervention</p> <p>2) Follow-up/delayed (specify)</p>	
6	<p>Quantitative information on outcomes of interests (e.g., means, standard deviations, <i>t</i>-values)</p> <p>(Note: all related outcomes will be extracted from the study and will be recorded in an Excel file for effect size calculations)</p>	
7	<p>Effect size calculation</p> <p>(e.g., Hedges' <i>g</i>, odd ratio, page number where the related original outcome data located, corresponding to each calculated effect sizes)</p>	
Section 7: Coding Information		
1	Coder	
2	Coding time: How much time (minutes) does it take to complete the coding?	
3	Date of coding	
4	Coding agreement rate with another independent coder (%)	

5	Areas/reasons of coding discrepancies (specify)	
6	How coding discrepancies were resolved (specify)	

About this review

Tools of the Mind (Tools) is an early childhood education curriculum, which involves structured make-believe play scenarios and a series of other curricular activities. Tools aims to promote and improve children's self-regulation and academic skills by having a dual focus on self-regulation and other social-emotional skills in educational contexts.

This review examines the evidence on the effectiveness of Tools in promoting children's self-regulation and academic skills, to inform its implementation in schools.