A meta-analysis of the effect of the Early Start Denver Model in children with autism spectrum disorder

Zhi Wang[1], Sau Cheong Loh[1], Jing Tian[2] and Qian Jiang Chen[3]

[1]Faculty of Education, University of Malaya, Kuala Lumpur, Malaysia; [2]School of Education, Guizhou Education University, Guiyang City, Guizhou Province, China; [3]School of Management, Guizhou University, Guiyang City, Guizhou Province, China

Objectives: To examine and analyse the intervention effects of the Early Start Denver Model (ESDM) on children with autism spectrum disorder (ASD).
Methods: This meta-analysis evaluated the effect sizes in four major domains of measurement (autism symptoms, language, cognition, and social communication). A total of 624 participants with ASD were included in 11 high-quality randomized controlled trial studies.
Results: The results indicated that the ESDM intervention resulted in significant improvement with moderate effect sizes in the cognition ($g = 0.28$), autism symptoms ($g = 0.27$), and language ($g = 0.29$) domains. The effect sizes of autism symptoms and language were moderated by country (Western versus Asian countries). However, there were no significant effects observed for the social communication domain.
Conclusion: The ESDM intervention significantly improved autism symptoms, language, and cognition. The effect sizes of autism symptoms and language were larger in Asian countries than in Western countries.

Keywords: autism spectrum disorder; Early Start Denver Model; meta-analysis; randomized controlled trials

Introduction

Autism spectrum disorder (ASD) is a developmental condition characterized by difficulties with social interaction and communication as well as restricted and repetitive behaviour (Lord *et al*. 2020). According to the 2020 statistics of the Centers for Disease Control and Prevention (CDC), 1 in 54 children younger than 8 years old in the United States is diagnosed with ASD (Knopf 2020). Early intensive behavioural intervention is recognized as an effective approach for improving development for young children with ASD. One of the main interventions is naturalistic developmental behavioural interventions (NDBIs), which incorporate applied behaviour analysis (ABA) and developmental principles and strategies for implementation (Vinen *et al*. 2018). The Early Start Denver Model (ESDM) is also an example of NDBI intervention and is designed especially for children aged 12–60 months (Rogers *et al*. 2019). The ESDM is a comprehensive early intervention program that facilitates social engagement and active learning and minimizes the impact of autism symptoms on children's learning by addressing deficits in attention, imitation, language, play skills, affect sharing, and social orientation (Vinen *et al*. 2018). This intervention program can be delivered by individuals, groups, primary professionals, or parents with low or high intensity over a brief or extended period of time (Cidav *et al*. 2017).

Correspondence to: Zhi Wang, Faculty of Education, University of Malaya, Jalan Universiti, Kuala Lumpur, 50603 Malaysia. Email: 17221470@siswa.um.edu.my

Previous studies have reported mixed outcomes in various domains. Some have reported that the participation of children with ASD in ESDM intervention resulted in significant improvements in autism symptoms or diagnosis (Gao *et al*. 2020, Xu et al. 2017, Yang 2015, Rogers *et al*. 2012, Vivanti *et al*. 2014), cognition (Eapen *et al*. 2013, Fulton *et al*. 2014, Rogers *et al*. 2019, Rogers *et al*. 2012, Vivanti *et al*. 2014), social communication (Rogers *et al*. 2019, Rogers *et al*. 2012), language (Rogers *et al*. 2012), adaptive behaviour performance (Colombi *et al*. 2018, Estes *et al*. 2015) and abatement of challenging behaviours (Fulton *et al*. 2014). However, other studies reported no significant changes in the severity of autism symptoms (Dawson *et al*. 2010, Fulton *et al*. 2014, Vivanti *et al*. 2016, Vivanti *et al*. 2014), language (Vivanti *et al*. 2017) and social communication (Li *et al*. 2018, Rogers *et al*. 2012, Xu et al. 2018). Therefore, it is necessary to conduct a meta-analysis to evaluate various targeting domains.

To address the mixed outcomes of the ESDM intervention, three meta-analyses reported effect sizes on different domains. Canoy and Boholano (2015) investigated the benefits of the ESDM intervention in the meta-analysis of five studies. The results indicated that the ESDM intervention improved language ability (including language reception and expression) and social communication and can be utilized in any type of setting. However, these results may not provide a reasonable conclusion owing to the low number of studies. Fuller *et al.* (2020) conducted a comprehensive meta-analysis to examine the effects of the ESDM intervention on six developmental outcomes (autism symptoms, language, social communication, cognition, adaptive functioning, and repetitive behaviours) of young children with ASD; they found significant improvements in cognition ($g = 0.41$) and language ($g = 0.41$) but not in autism symptoms, adaptive behaviour, social communication, or repetitive behaviours. Although this study supported the positive effects of the ESDM intervention on language and cognition, there were still two potential limitations that required further investigation, which included (1) the performance of the subgroup analysis and detailed assessment of the heterogeneity in the 12 studies and (2) establishment of the relationships between the length or intensity dosage and study outcomes. Yu et al. (2020) examined the 14 randomized controlled trials (RCTs) results of the ABA and other studies with related principles such as ESDM, picture exchange communication system (PECS), and discrete trial training (DTT), and found no significant effects on autism symptoms and social communication in the ESDM intervention compared with the ABA approach. Additionally, this study supported the substantial impacts of ESDM on receptive and expressive language. However, it was noted that only five studies were included in the ESDM compared to other interventions. Therefore, the low number of studies may not provide a reasonable conclusion. In summary, although the present meta-analysis results have supported the effects of ESDM, these results may be influenced by different factors (e.g. low number of empirical studies, experimental designs, and potential moderating variables).

Given the above research findings, we believe that there remain some gaps to fill. First, more rigorous experimental designs are needed. RCT experimental designs were adopted because they could be the gold standard for meta-analysis and more reliable than other experimental designs. Second, more potential moderating variables are required. Some potential moderating variables will be added in this study to explore intervention effects (e.g. intervention country, the length of intervention period, and the intensity of intervention). Third, literature searches in multiple languages (Chinese, English, and Malay) could reduce language bias in meta-analysis and support more comprehensive results and analysis. Hence, the first purpose of this study was to examine the effects of the ESDM intervention on targeting domains (autism symptoms, social communication, language, and cognition). Although some studies have reported the impacts of ESDM on other developmental outcomes (adaptive functioning or repetitive behaviours), they should be interpreted with caution due to the limited number of studies and possible low statistical power. Furthermore, this study also aimed to explore and analyse the effects of moderating variables on target domains. The last objective of this study was to compare the effects of ESDM on different languages.

Methods

Eligibility criteria

This study follows the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines to facilitate the transparent exposition of systematic reviews and meta-analysis. Studies were included in the meta-analysis if they met the following criteria: (1) studies published in Chinese, English, and Malay; (2) children diagnosed with ASD; (3) studies must report RCTs; (4) the intervention group only received the ESDM while the control group received either conventional or no intervention; and (5) reported outcomes for autism symptoms, cognition, language, and social communication.

Search procedure

A total of nine databases were searched (EMBASE, PubMed, Cochrane Library, Web of Science, PsycINFO, ScienceDirect, ProQuest, CNKI, and WANFANG). The final search was completed in October 2020. To retrieve relevant papers, the following descriptors were used: (Autism OR ASD OR Autism Spectrum Disorder OR Developmental Disorder) AND (Early Start Denver Model OR ESDM OR Denver Model) AND (Randomized Controlled Trials OR RCTs). The first author searched unpublished or ‘grey’ literature using the online databases of dissertations and theses.

Data extraction and coded variables

All child outcome measures were recorded for each study. If a study reported the overall and subscale scores, the total/overall score was used. However, the subscale scores were appropriate for outcome-specific meta-analysis. Additionally, we recorded study-level characteristics, including average years, outcome measurement, country, primary implementer, intervention format, length of intervention period (in weeks), and intervention intensity (hours per week). Besides, two researchers (first and third authors) independently extracted data according to the inclusion criteria and performed an interrater agreement analysis using Kappa; consequently, any disagreements were resolved via discussion.

In addition, we used the Cochrane risk of bias assessment tool based on seven methodological quality indicators. Random sequence generation indicates selection bias (i.e. biased allocation to interventions) due to inadequate generation of a randomized sequence. Allocation concealment also signifies selection bias owing to the inadequate protection of allocations prior to the assignment. Blinding of participants and personnel refers to performance bias, which results from participants' and personnel's knowledge of the allocated interventions. Blinding of outcome assessment implies a detection bias as a result of the outcome assessors' knowledge of the allocated interventions. Incomplete outcome data denote attrition bias, which results from the amount, nature, or management of incomplete outcome data. Selective reporting indicates reporting bias due to the presentation of specific results. Other bias specifies includes bias associated with problems not covered elsewhere in the table. According to the Cochrane evaluation criteria, the fulfilment of the quality criteria by four or more items denotes the existence of low-risk bias. Additionally, the fulfilment attainment of the quality criteria by two or more items indicates medium-risk bias. However, if only one item meets the quality criteria, it implies the presence of high-risk bias (Higgins *et al*. 2019).

Analytic strategies

Calculation of effect sizes

The outcomes of each study were analysed using the meta-analysis software package Comprehensive Meta-Analysis version 2.0 (CMA2.0). Hedges' *g* was calculated using the standardized mean difference, as it constituted a conservative estimate. In CMA2.0, we input the sample size, postmean, and standard deviation of the intervention and control groups and subsequently calculated the *g*-value. A computer was used to convert the mean and standard deviation. For the interpretation of the *g*-value, Cohen provided rules of thumb as follows: values of 0.2, 0.5, and 0.8 indicate small, medium, and large effect sizes, respectively (Lipsey and Wilson 2001).

Model selection and heterogeneity test

A random-effects model was used in the meta-analysis owing to the lack of a common true effect among the involved studies. Furthermore, the existence of a significant relationship between the studies was also assumed. Given the possible moderating role of various variables (country, primary implementer, or length of period) on the intervention effect of ESDM, a random-effects model was adopted. In the following meta-analysis, a heterogeneity test was performed to further verify the model selection.

The purpose of the heterogeneity test is to determine whether the effects measured between studies are heterogeneous. Consequently, the Q statistic and I^2 indexes were utilized for the evaluation. The Q statistic is based on the test of total variation. It is assumed that the effect sizes are dependent on a chi-square distribution with $p < 0.05$ indicating significant heterogeneity; I^2 refers to the proportion of the variance between the studies in the population. Generally, I^2 indexes of 25%, 50%, and 75% are considered the limits of low, medium, and high heterogeneity, respectively (Lipsey and Wilson 2001).

Publication bias

The risk of publication bias was assessed by Egger's test and the linear regression method. Generally, we initially obtained the intercept of the linear regression equation and its 95% confidence interval and then conducted a hypothesis test to ascertain whether the intercept is 0. The lack of a significant value of intercept indicates the absence of publication bias. Additionally, we constructed a funnel plot to obtain a visual idea of potential publication bias. Following the convention, the precision of the studies was plotted along the y-axis with more precise studies (e.g. larger N) at the top and less precise studies (e.g. smaller N) at the bottom of the graph. Studies were plotted along the X-axis depending on the estimates of the effect sizes. The distribution of the effect size of the studies around the mean was assumed to be asymmetrical with the observation of less precise studies at the bottom and highly precise studies—closer to the estimated mean effect size—at the top. It is assumed that the presence of more studies on the right (high effect sizes) than on the left (small or null effects) side of the graph indicates the presence of publication bias. Finally, Duval and Tweedie's trim and fill method, which uses an iterative process, was applied to remove the extremely small studies, subsequent to the computation of the new effect size until a symmetrical funnel plot was achieved. In addition, the number of studies that might be missing from the funnel plot was examined to obtain the effect size, including the estimated missing studies.

Moderator variables analysis

There are two approaches to analysing the moderator variables. First, the meta-regression analysis examines the variation in effect sizes as a function of the following continuous variables: the length of the intervention period (in weeks) and the intensity of intervention (the average hours per week). Second, Q statistics determine whether the effect sizes vary as a function of the following categorical moderators: country (Western versus Asian), primary implementer (professional versus parent), and intervention format (individual versus group).

The authors grouped the studies according to the categorical variables being assessed. Then, the authors calculated the Q_{between} using the effect weights and pooled estimates to determine the existence of heterogeneity between the

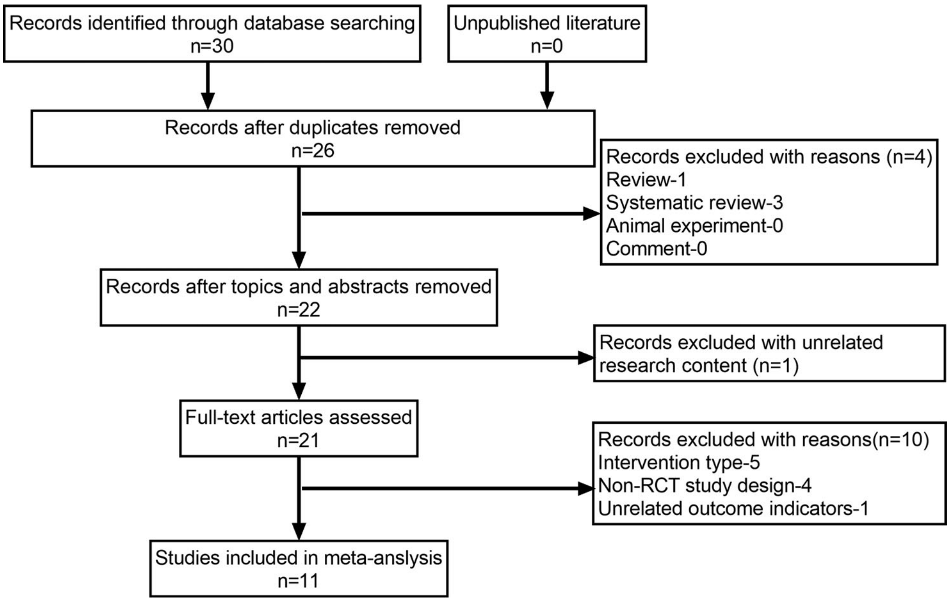


Figure 1. The PRISMA diagram of study inclusion.

Table 1 Description of studies

Authors (year)	N	Average age (Years)	Outcome measure	Country	Primary implementer	Intervention format	Intervention length (weeks)	Intervention intensity (hours/week)
Dawson et al. (2010)	48	1.95	ADOS/MSEL	USA	Professional	Individual	104	20
Rogers et al. (2012)	98	1.75	ADOS/MSEL/MCDI	USA	Parent	Individual	12	1
Vivanti et al. (2014)	57	3.4	MSEL/MCDI	Australia	Professional	Group	52	15
Yang (2015)	36	3.96	CARS/PEP-3	China	Professional	Individual	8	5
Xu et al. (2017)	36	3.58	CARS	China	Professional	Individual	8	5
Vismara et al. (2018)	30	2.46	ADOS	USA	Parent	Individual	12	1.5
Vinen et al. (2018)	59	3.11	ADOS	USA	Professional	Group	156	17
Li et al. (2018)	35	3.25	CARS	China	Professional	Group	12	12
Rogers et al. (2019)	118	1.72	ADOS/MSEL	USA	Professional	Individual	116	16
Wang et al. (2019)	40	3.50	CARS/PEP-3	China	Professional	Group	24	12
Gao et al. (2020)	70	2.08	CARS/PEP-3	China	Parent	Individual	12	15.5

Note. ADOS: Autism Diagnostic Observation Schedule, CARS: Childhood Autism Rating Scale, MESL: Mullen Scales of Early Learning, MCDI: MacArthur Bates Communicative Development Inventory, PEP-3: Psycho Educational Profile-Third Edition.

groups. If the $Q_{between}$ value is significant ($p < 0.05$), moderator variables will be the grouping variable.

Results

Study selection

Figure 1 illustrates the study selection process. Of the 30 studies initially identified, 26 were selected after the duplication check. A review article and three systematic reviews were further excluded, resulting in the remainder of 22 studies. However, only 11 studies were included after removing the irrelevant content ($n = 1$) and unrelated outcome indicators ($n = 10$). The interrater agreement on the task classification was high, with a kappa of 0.84 ($p < 0.001$), *95% CI* [0.71, 0.97]. There was no interrater disagreement between the two independent raters (the first author and the third authors).

Study characteristics

Table 1 represents the characteristics of the studies included in the present analysis. The studies were conducted in three countries (America, Australia, and China) and included a total of 624 participants. The age range was from 1.72 to 3.96 years. The studies utilized five outcome measures to assess autism symptoms, social communication, cognition, and language domains. The primary implementer included professionals ($n = 8$) and parents ($n = 3$). The main intervention format was individuals ($n = 7$), followed by groups ($n = 4$). The length of the intervention period ranged between 8 and 156 weeks (mean = 46.91, SD = 50.80), with an intensity of 1–20 h/week (mean = 10.91, SD = 6.34).

Methodological quality of included studies

Figures 2 and 3 indicate that all the studies included had low-risk bias. Among the studies, seven had a quality score of 5 points, while the remaining four attained 4 points, indicating their high-quality level.

Meta-analysis results

Autism symptoms

Figure 4 shows the forest plots of nine autism symptom outcomes reported across eight studies. The effect sizes were

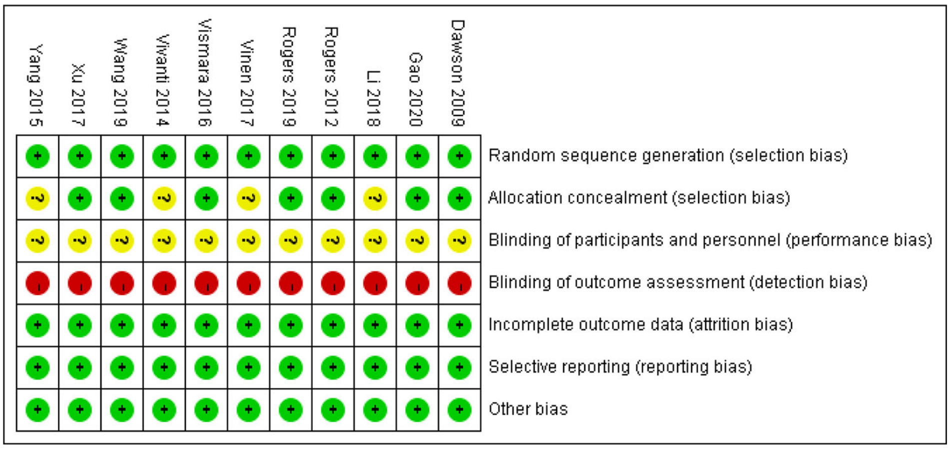


Figure 2. The methodological quality of included studies. Note: '+' indicates compliance with the standard; '-' implies non-compliance with the standard; and '?' means unclear content.

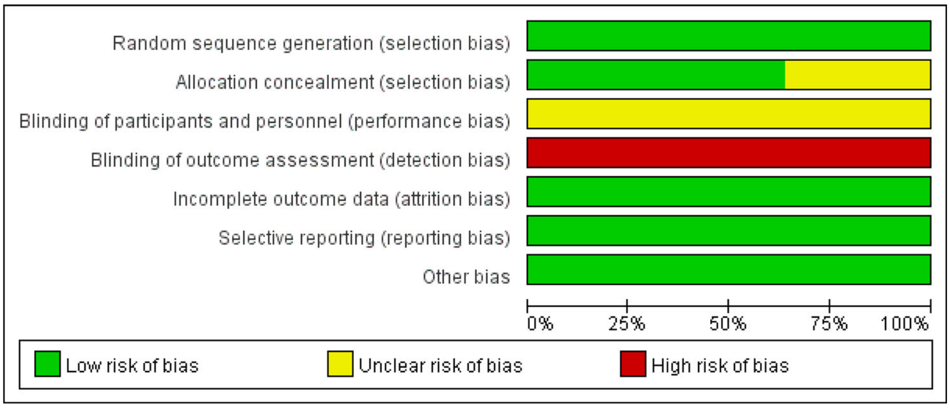


Figure 3. The distribution of the methodological quality of the included studies.

Study name	Statistics for each study						
	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Dawson (2009)	-0.159	0.294	0.086	-0.735	0.417	-0.541	0.589
Rogers (2012a)	-0.067	0.201	0.040	-0.460	0.326	-0.335	0.738
Rogers (2012b)	0.102	0.201	0.040	-0.291	0.495	0.507	0.612
Vivanti (2014)	0.402	0.264	0.070	-0.116	0.920	1.521	0.128
Yang (2015)	1.090	0.352	0.124	0.399	1.780	3.094	0.002
Xu (2017)	1.088	0.352	0.124	0.398	1.778	3.090	0.002
Vinen (2017)	0.083	0.257	0.066	-0.421	0.588	0.323	0.747
Li (2018)	0.265	0.332	0.110	-0.386	0.915	0.797	0.425
Gao (2020)	0.179	0.237	0.056	-0.285	0.643	0.756	0.450
	0.272	0.130	0.017	0.018	0.526	2.096	0.036

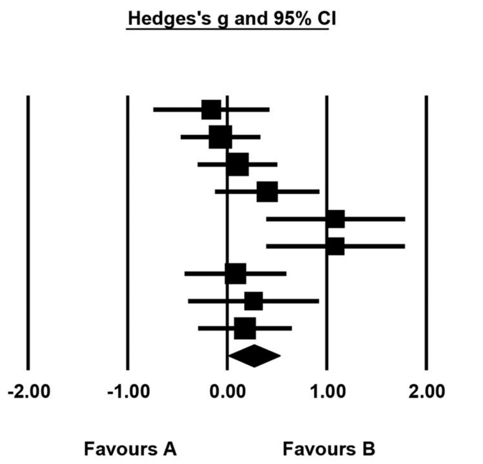


Figure 4. Forest plot for ASD symptoms outcomes.

represented such that positive values indicate a reduction in autism symptoms. The effect size was $g = 0.27$ ($p = 0.04$), which means that exposure of children with ASD to ESDM resulted in significant improvements in autism symptoms. A moderate level of heterogeneity was observed ($I^2 = 53.06\%, p = 0.03 < 0.05$).

Publication bias test

A subjective judgement was made to determine the existence of publication bias. Figure 5 shows that asymmetrical distribution was dispersed on both sides of the total effect sizes, with more on the right side, indicating the possibility of publication bias. Egger's test of the

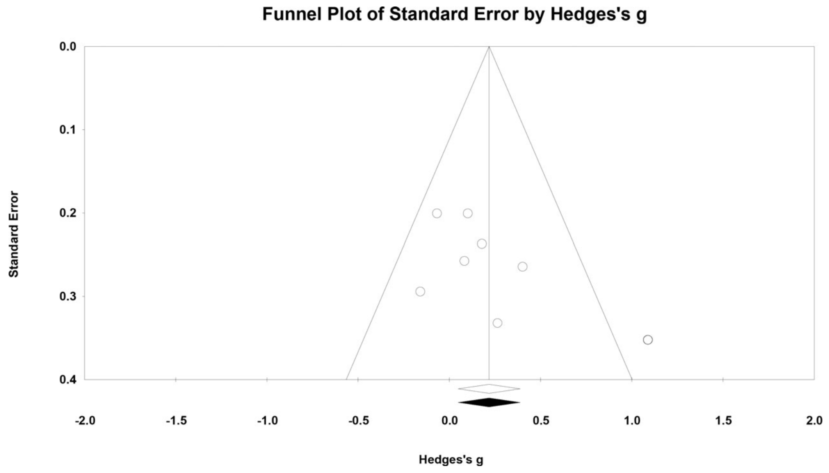


Figure 5. Funnel plot of autism symptoms. Note: solid diamond indicates revised effect sizes.

Table 2 Analysis of the effect of moderator variables on autism symptoms

Moderator variables	Group	*k*	*G*	*LL*	*UL*	$Q_{between}$	*p*
Country	Western country	5	0.07	−0.14	0.27	3.99	0.046
	Asian country	4	0.62	0.12	1.12		
Primary Implementer	Parent	3	0.06	−0.18	0.30	2.46	0.12
	Professional	6	0.43	0.04	0.82		
Intervention Format	Group	3	0.25	−0.07	0.56	0.07	0.80
	Individual	6	0.31	−0.07	0.69		

Table 3 Meta-regression of continuous variables on autism symptoms

Moderator variables	Standard Error	*LL*	*UL*	*Tau-squared*	*p*
Length of intervention	0.002	−0.005	0.001	0.09	0.26
Intensity of intervention	0.14	−0.03	0.02	0.10	0.74

small study bias ($p = 0.03 < 0.05$) indicated that there is a risk of publication bias in the sample. Given the possibility of publication bias, we applied Duval and Tweedie's trim and fill method and observed that there were no missing data in the study, and the revised effect size remained unchanged ($g = 0.27$, *95% CI* [0.02, 0.53]).

Moderator variables results

The heterogeneity test revealed the presence of heterogeneity in the effect sizes of the included studies, indicating the possible involvement of moderator variables. Hence, the moderating role of the country, the primary implementer, and the intervention format was investigated. It can be concluded from Table 2 that the effect size was significantly moderated by the country ($Q_{between} = 3.99$, $p = 0.046 < 0.05$), and the effect size of the Asian country subgroup was significantly larger than that of the Western country subgroup. The primary implementer and the intervention format were independent of the heterogeneity in the effect sizes.

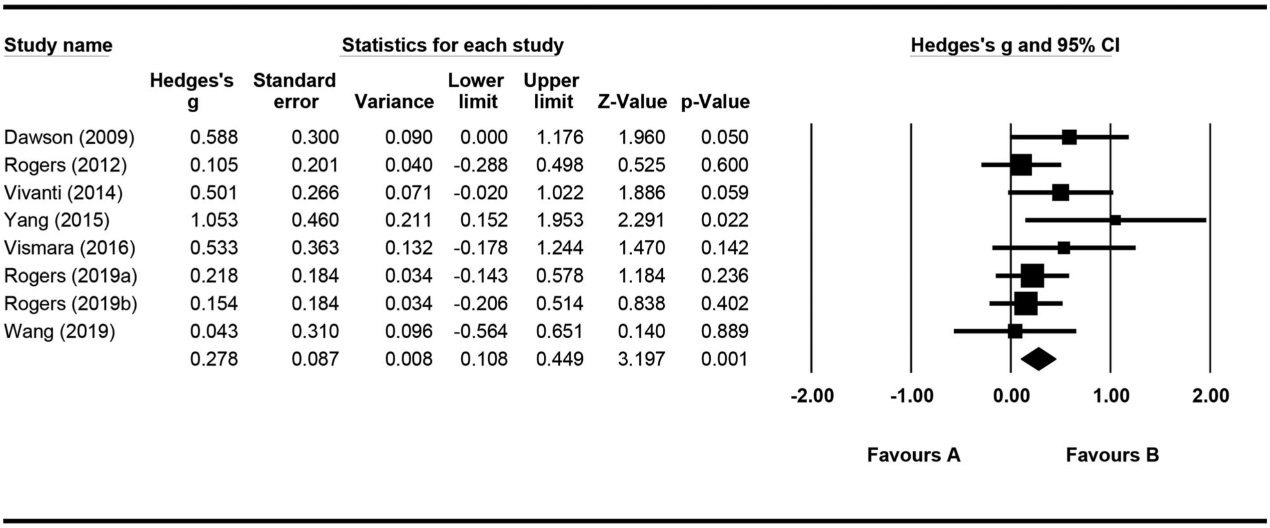
The continuous variables (the length and intensity of intervention) were also analysed in the meta-regression. Table 3 shows that the results of the regression model were not significant, and the length and intensity of the intervention could not predict the results. In other words, increasing the length and intensity of the intervention does not produce significant effects on autism symptoms.

Cognition

Figure 6 displays the forest plot of the eight cognitive outcomes that were reported across seven studies. The effect size was $g = 0.28$ ($p = 0.001$), which indicated that children who received the ESDM experienced significant improvement in cognitive development compared to the control groups. A small level of heterogeneity was observed ($I^2 = 1.38\%$, $p = 0.42$).

Publication bias test

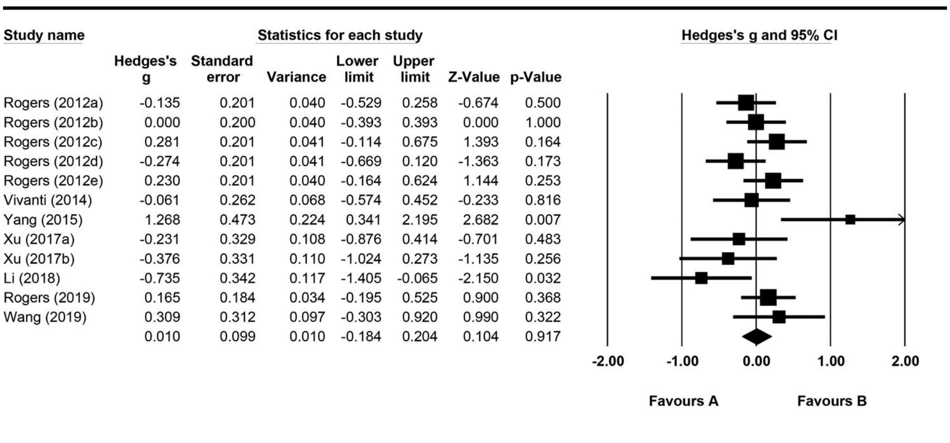
Figure 7 shows that asymmetrical distribution was scattered on both sides of the total effect sizes, with more on the right side, indicating the possibility of publication bias. Egger's test of publication bias ($p = 0.03 < 0.05$) indicated a possible risk of small publication bias in the sample. Consequently, Duval and Tweedie's trim and fill method revealed that two effect sizes fell outside the highlighted area, suggesting a small bias. Notably, the inclusion of additional data would reduce the average effect size. Accordingly,



Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Dawson (2009)	0.588	0.300	0.090	0.000	1.176	1.960	0.050
Rogers (2012)	0.105	0.201	0.040	-0.288	0.498	0.525	0.600
Vivanti (2014)	0.501	0.266	0.071	-0.020	1.022	1.886	0.059
Yang (2015)	1.053	0.460	0.211	0.152	1.953	2.291	0.022
Vismara (2016)	0.533	0.363	0.132	-0.178	1.244	1.470	0.142
Rogers (2019a)	0.218	0.184	0.034	-0.143	0.578	1.184	0.236
Rogers (2019b)	0.154	0.184	0.034	-0.206	0.514	0.838	0.402
Wang (2019)	0.043	0.310	0.096	-0.564	0.651	0.140	0.889
	0.278	0.087	0.008	0.108	0.449	3.197	0.001

Figure 6. Forest plot for ASD cognition outcomes.

Figure 7. Funnel plot of cognition. Note: solid diamonds indicate revised effect sizes, and solid dots denote additional data by the trim and fill method.



Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Rogers (2012a)	-0.135	0.201	0.040	-0.529	0.258	-0.674	0.500
Rogers (2012b)	0.000	0.200	0.040	-0.393	0.393	0.000	1.000
Rogers (2012c)	0.281	0.201	0.041	-0.114	0.675	1.393	0.164
Rogers (2012d)	-0.274	0.201	0.041	-0.669	0.120	-1.363	0.173
Rogers (2012e)	0.230	0.201	0.040	-0.164	0.624	1.144	0.253
Vivanti (2014)	-0.061	0.262	0.068	-0.574	0.452	-0.233	0.816
Yang (2015)	1.268	0.473	0.224	0.341	2.195	2.682	0.007
Xu (2017a)	-0.231	0.329	0.108	-0.876	0.414	-0.701	0.483
Xu (2017b)	-0.376	0.331	0.110	-1.024	0.273	-1.135	0.256
Li (2018)	-0.735	0.342	0.117	-1.405	-0.065	-2.150	0.032
Rogers (2019)	0.165	0.184	0.034	-0.195	0.525	0.900	0.368
Wang (2019)	0.309	0.312	0.097	-0.303	0.920	0.990	0.322
	0.010	0.099	0.010	-0.184	0.204	0.104	0.917

Figure 8. Forest plot of ASD social communication outcomes.

$g = 0.23$, *95% CI* [0.03, 0.42] after correction. Based on the above analysis, although publication bias may exist in the study, the conclusion is still valid.

Social communication

Figure 8 displays the forest plot of the 12 social communication outcomes that were reported across seven

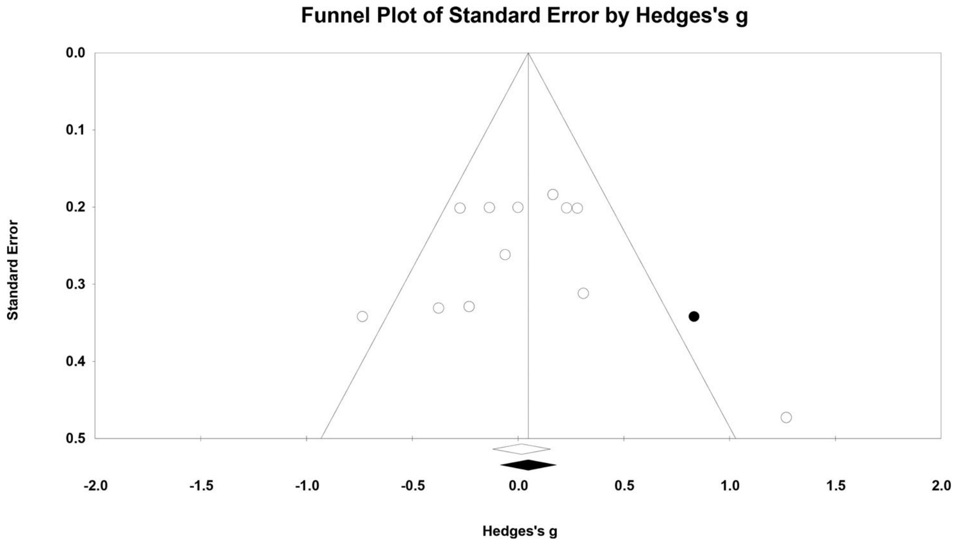


Figure 9. Funnel plot of social communication. Note: solid diamonds indicate revised effect sizes, while solid dots signify the additional data by the trim and fill method.

Study name	Hedges's g	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
Rogers (2012a)	-0.242	0.201	0.040	-0.636	0.153	-1.201	0.230
Rogers (2012b)	-0.187	0.201	0.040	-0.581	0.206	-0.932	0.351
Rogers (2012c)	0.050	0.200	0.040	-0.343	0.442	0.247	0.805
Rogers (2012d)	0.281	0.201	0.041	-0.114	0.675	1.393	0.164
Yang (2015a)	1.104	0.463	0.214	0.198	2.011	2.388	0.017
Yang (2015b)	0.960	0.454	0.206	0.069	1.850	2.112	0.035
Xu (2017)	1.082	0.352	0.124	0.393	1.772	3.076	0.002
Vinen (2017)	-0.335	0.259	0.067	-0.843	0.173	-1.292	0.196
Rogers (2019)	0.231	0.184	0.034	-0.129	0.592	1.257	0.209
Wang (2019a)	0.313	0.312	0.097	-0.299	0.924	1.002	0.316
Wang (2019b)	0.356	0.312	0.098	-0.257	0.968	1.139	0.255
Wang (2019c)	1.358	0.345	0.119	0.682	2.035	3.935	0.000
Gao (2020)	-0.758	0.538	0.290	-1.813	0.297	-1.408	0.159
	0.278	0.141	0.020	0.002	0.555	1.973	0.048

(Statistics for each study)

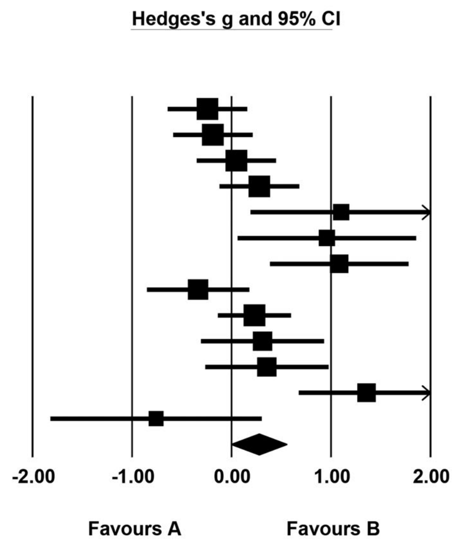


Figure 10. Forest plot of ASD language outcomes.

studies. The effect size was $g = 0.01$ ($p = 0.92$), which indicated that ESDM does not significantly improve social communication in children with ASD. A medium level of heterogeneity was observed ($I^2 = 47.48\%$, $p = 0.34$).

Publication bias test

Figure 9 reveals that the asymmetrical distribution was scattered on both sides of the total effect sizes but more to the left side, indicating the possibility of publication bias. However, Egger's test of publication bias ($p = 0.97 > 0.05$) indicated the absence of publication bias in the sample. Given the inconsistency of the results, the trim and fill method was applied and detected an effect size falling outside the highlighted area, suggesting a small bias. It is notable that adding more data would change the average effect size. Consequently, $g = 0.06$, *95% CI* [−0.15, 0.26] after correction. This result should be interpreted with caution, as the confidence intervals overlapped with zero.

Language

Figure 10 displays the forest plot of the 13 language outcomes that were reported across seven pieces of literature. The effect size was $g = 0.29$ ($p = 0.048$), indicating that children who received ESDM made significant progress in language compared to those in the control groups. A medium level of heterogeneity was observed ($I^2 = 70.82\%$, $p < 0.05$).

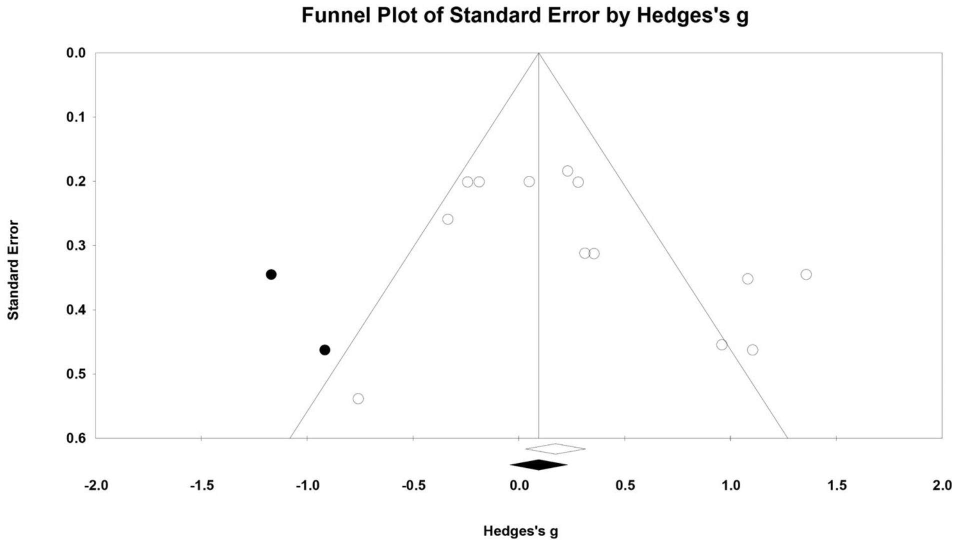


Figure 11. Funnel plot of language. Note: solid diamonds indicate revised effect sizes, while solid dots signify additional data by the trim and fill method.

Table 4 Analysis of the moderator variables on language

Moderator variables	Group	*k*	*g*	*LL*	*UL*	$Q_{between}$	*p*
Country	Western country	6	−0.01	−0.22	0.19	7.12	0.008
	Asian country	7	0.69	0.22	1.17		
Primary Implementer	Parent	5	−0.06	−0.31	0.19	6.82	0.12
	Professional	8	0.57	0.17	0.97		
Intervention Format	Group	4	0.40	−0.28	1.08	0.22	0.64
	Individual	9	0.23	−0.08	0.53		

Table 5 A meta-regression of continuous variables on language.

Moderator variables	Standard Error	*LL*	*UL*	*Tau-squared*	*p*
Length of intervention	0.001	−0.005	0.001	0.19	0.24
Intensity of intervention	0.10	−0.07	0.33	0.20	0.54

Publication bias test

Figure 11 shows that asymmetrical distribution was dispersed on both sides of the total effect sizes, with more on the right side, indicating the possibility of publication bias. Egger's test of a small study bias ($p = 0.10 < 0.05$) indicated a lack of publication bias in the sample. However, the trim and fill method observed two effect sizes falling outside of the highlighted area, suggesting the presence of a small bias. It is notable that additional data would minimize the average effect size; thus, it was corrected to $g = 0.14$, *95% CI* [−0.16, 0.44]. However, the result should be interpreted with caution due to the overlap of confidence intervals with zero.

Moderator variables results

The heterogeneity test revealed the possible moderating role of some variables in the study. Consequently, the moderating role of the country, primary implementer, and intervention format was analysed. From Table 4, it can be concluded that the effect size was significantly moderated by the country ($Q_{between} = 7.12$, $p < 0.05$)—the effect size of the Asian country subgroup was significantly higher than that of the Western country subgroup. The primary implementer and intervention format were independent of the heterogeneity of effect sizes.

The continuous variables (the length and intensity of the intervention) were also analysed in the meta-regression. Table 5 shows that the results of the regression model were not significant, suggesting that the length and intensity of intervention cannot predict the results.

Discussion

This meta-analysis examined the effects of the ESDM intervention on children with ASD in the following four major measurement domains: autism symptoms, cognition, social communication, and language. Among them, the ESDM intervention significantly improved autism symptoms and language and achieved medium effect sizes. The result of the language domain was consistent with the study of Fuller *et al.*, suggesting that children with ASD who received the ESDM intervention made significant progress in their language (Fuller *et al*. 2020). However, the result of autism symptoms contradicts their findings, which suggested that the ESDM intervention does not significantly enhance autism symptoms in children with ASD (Fuller *et al*. 2020). There are three possible reasons to explain these contrasting observations. First, different study designs were included. The meta-analysis from Fuller *et al.* included RCTs and quasi-experimental designs. However, the RCTs included in this study were fitted with a more rigorous than quasi-experimental design. Notably, one of the main concerns of the quasi-experiment is internal validity, since the treatment and control groups may not be comparable at baseline. Any change in post-intervention characteristics may be attributed to the different interventions, and the quasi-experiment may not be possible to convincingly demonstrate a causal link between the treatment condition and observed outcomes. Second, the number of included studies in language and autism symptoms was smaller than in the study of Fuller *et al.* (10 for autism symptoms, 19 for language). This lack of data may result in a reduction in the effect size. Third, it should be noted that although the result confirmed the positive effect of ESDM intervention on the language of children with ASD, this conclusion should be taken cautiously due to the inclusion of zero in the 95% confidence intervals.

This study discovered the existence of a medium amount of heterogeneity in autism symptoms and language. In addition, the heterogeneity analysis showed that the overall effect sizes in these two domains were significantly moderated by country (Western versus Asian). The effect sizes of the Asian countries were larger than those of the Western countries, which may be partly due to the following challenges associated with the adoption of various measurements. First, all the measurements of autism symptoms in Asian countries follow the CARS measurement, which is a rating scale that is more subjective than the ADOS and may cause an increase in the effect sizes. Therefore, researchers should pay more attention to problems relating to the measurement of scales. For example, the ADOS is intended to capture relatively stable characteristics of ASD symptoms, including social communication for diagnostic purposes and is not created to measure a treatment-related change. A more recent measure, known as the Brief Observation of Social Communication Change (BOSC), was created for this purpose and may be a more useful tool for capturing changes in the outcomes (Grzadzinski *et al*. 2016). Additionally, the majority of studies in Asian countries adopted the combination of PEP-3 and CARS to evaluate the intervention effects on language. It is expected that parental scores in the PEP-3 assessment result will be relatively high, as they are optimistic that the ESDM intervention can effectively improve their children's language ability. As a result, most parents' reports are positive. For instance, some studies (Wang et al. 2019, Yang 2015) revealed that children with ASD who received the ESDM intervention made significant progress in language development according to their parents' reports. In contrast, the MESL scale, which is widely used in Western countries, has a detailed scoring method and includes the evaluation of expressive and receptive language, thereby reducing subjective bias to a certain extent. Another reason for the variation in the effect sizes between Asian and Western countries is the influence of the ESDM reference standard of the norm on the experimental results. The norms of the Denver development assessment are derived from American children, which may cause varying intervention effects among children of the same age in different countries due to national and cultural backgrounds. The third reason is the impact of the low number of outcomes in autism symptoms and language on intervention effects. For example, there were only 13 language and 9 autism symptom outcomes that met the requirement and may cause instability in the analysis of the results.

Children with ASD who received ESDM also made significant progress in cognition. Notably, the effect sizes of cognition were smaller than those in Fuller *et al.*'s findings ($g = 0.41$). There are two possible explanations for this observation. First, Fuller's study included gross and fine motor data in cognition, which may cause bias in the effect size. Second, the revised effect size ($g = 0.23$) was tested using the trim and fill method and led to a small effect compared to Fuller *et al.*'s study. On the other hand, no significant difference was observed in social communication compared to the control groups. This finding is consistent with the previous observation; however, further studies are needed to accumulate evidence of the effect of ESDM on social communication.

Limitations

The most prominent limitation is the failure to thoroughly analyse the relationship between moderator variables, owing to the low number of studies. Notably, although the relationship between the length and intensity of intervention and intervention results was not established in this study, the effects of the intensity and intervention period on intervention results could be further explored in future studies. For example, Rogers *et al*. (2012) found that the low intensity and short intervention period (one hour per week and three-month intervention period, respectively) may hinder the impact of ESDM intervention. Another limitation is the inclusion of only Chinese literature in this study to represent Asian countries, causing a potential study bias. Furthermore, multilingual researchers should enrich the research findings. For example, a Japanese study showed that ESDM intervention in different countries may have varying influences due to cross-cultural backgrounds (Yoshimura et al. 2019). Therefore, more data from various Asian countries should be considered in the overall analysis.

Acknowledgements

We would like to thank those who assisted in this study, including special education teachers, Chinese scholars and Malaysian supervisors.

Conflict of interest

The authors declare no conflict of interest.

Funding

This work was supported by the Youth Science and Technology Talent Growth Project from Guizhou Provincial Education Department, China (2018-259).

ORCID

Sau Cheong Loh http://orcid.org/0000-0001-5962-6815

References

Canoy, J. P. and Boholano, H. B. 2015. Early start Denver model: A meta-analysis. *Journal of Education and Learning*, 4, 314–327.

Cidav, Z., Munson, J., Estes, A., Dawson, G., Rogers, S. and Mandell, D. 2017. Cost offset associated with Early Start Denver Model for children with autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, 9, 777–783.

Colombi, C., Narzisi, A., Ruta, L., Cigala, V., Gagliano, A., Pioggia, G., Siracusano, R., Rogers, S. J., Muratori, F. and Team, P. P. 2018. Implementation of the Early Start Denver Model in an Italian community. *Autism*, 2, 126–133.

Dawson, G., Rogers, S., Munson, J., Smith, M., Winter, J., Greenson, J., Donaldson, A. and Varley, J. 2010. Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. *Pediatrics*, 125, e17–e23.

Eapen, V., Črnčec, R. and Walter, A. 2013. Clinical outcomes of an early intervention program for preschool children with autism spectrum disorder in a community group setting. *BMC Pediatrics*, 1, 3.

Estes, A., Munson, J., Rogers, S. J., Greenson, J., Winter, J. and Dawson, G. 2015. Long-term outcomes of early intervention in 6-year-old children with autism spectrum disorder. *Journal of the American Academy of Child & Adolescent Psychiatry*, 7, 580–587.

Fuller, E. A., Oliver, K., Vejnoska, S. F. and Rogers, S. J. 2020. The effects of the Early Start Denver Model for children with autism spectrum disorder: A meta-analysis. *Brain Sciences*, 10, 368.

Fulton, E., Eapen, V., Črnčec, R., Walter, A. and Rogers, S. 2014. Reducing maladaptive behaviors in preschool-aged children with autism spectrum disorder using the Early Start Denver Model. *Frontiers in Pediatrics*, 40, 6–14.

Gao, D., Ting, Y., Chun-Li, L., Fei-Yong, J. and Hong-Hua, L. 2020. Effect of parental training based on Early Start Denver Model combined with intensive training on children with autism spectrum disorder and its impact on parenting stress (in Chinese). *Chinese Journal of Contemporary Pediatrics*, 2, 158–163.

Grzadzinski, R., Carr, T., Colombi, C., McGuire, K., Dufek, S., Pickles, A. and Lord, C. 2016. Measuring changes in social communication behaviors: Preliminary development of the Brief Observation of Social Communication Change (BOSCC). *Journal of Autism and Developmental Disorders*, 46, 2464–2479.

Higgins, J. P., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M. J. and Welch, V. A. 2019. *Cochrane handbook for systematic reviews of interventions*. New York: John Wiley & Sons.

Knopf, A. 2020. Autism prevalence increases from 1 in 60 to 1 in 54: CDC. *The Brown University Child and Adolescent Behavior Letter*, 6, 4.

Li, H. H., Li, C. L., Gao, D., Pan, X. Y., Du, L. and Jia, F. Y. 2018. Preliminary application of Early Start Denver Model in children with autism spectrum disorder (in Chinese). *Chinese Journal of Contemporary Pediatrics*, 10, 793–798.

Lipsey, M. W. and Wilson, D. B. 2001. *Practical meta-analysis*. London: SAGE Publications, Inc.

Lord, C., Brugha, T. S., Charman, T., Cusack, J., Dumas, G., Frazier, T., Jones, E. J., Jones, R. M., Pickles, A. and State, M. W. 2020. Autism spectrum disorder. *Nature Reviews Disease Primers*, 1, 1–23.

Rogers, S. J., Estes, A., Lord, C., Munson, J., Rocha, M., Winter, J., Greenson, J., Colombi, C., Dawson, G. and Vismara, L. A. 2019. A multisite randomized controlled two-phase trial of the Early Start Denver Model compared to treatment as usual. *Journal of the American Academy of Child & Adolescent Psychiatry*, 9, 853–865.

Rogers, S. J., Estes, A., Lord, C., Vismara, L., Winter, J., Fitzpatrick, A., Guo, M. and Dawson, G. 2012. Effects of a brief Early Start Denver Model (ESDM)-based parent intervention on toddlers at risk for autism spectrum disorders: A randomized controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry*, 10, 1052–1065.

Vinen, Z., Clark, M., Paynter, J. and Dissanayake, C. 2018. School age outcomes of children with autism spectrum disorder who received community-based early interventions. *Journal of Autism and Developmental Disorders*, 5, 1673–1683.

Vismara, L. A., McCormick, C. E., Wagner, A. L., Monlux, K., Nadhan, A. and Young, G. S. 2018. Telehealth parent training in the Early Start Denver Model: Results from a randomized controlled study. *Focus on Autism and Other Developmental Disabilities*, 2, 67–79.

Vivanti, G., Dissanayake, C. and Team, V. A. 2016. Outcome for children receiving the Early Start Denver Model before and after 48 months. *Journal of Autism and Developmental Disorders*, 7, 2441–2449.

Vivanti, G., Duncan, E., Dawson, G. and Rogers, S. J. 2017. *Implementing the group-based Early Start Denver Model for preschoolers with autism*. New York: Springer.

Vivanti, G., Paynter, J., Duncan, E., Fothergill, H., Dissanayake, C., Rogers, S. J. and Team, V. A. 2014. Effectiveness and feasibility of the Early Start Denver Model implemented in a group-based community childcare setting. *Journal of Autism and Developmental Disorders*, 12, 3140–3153.

Wang, J. H., Gu, D. D., Sun, Y. Y., Jia, F. Y. and Li, H. H. 2019. Efficacy analysis of Early Start Denver model in children with autism spectrum disorder (in Chinese). *Chinese Journal of Behavioral Medicine and Brain Science*, 8, 684–688.

Xu, Y., Yang, J., Yao, J., Chen, J., Zhuang, X., Wang, W., Zhang, X. and Lee, G. T. 2018. A pilot study of a culturally adapted early intervention for young children with autism spectrum disorders in China. *Journal of Early Intervention*, 1, 52–68.

Xu, Y., Yang, J. and Yang, J. 2017. Application of Early Start Denver Model for early intervention on autistic children (in Chinese). *Chinese Journal of Clinical Psychology*, 1, 188–191.

Yang, J. 2015. *Trial of the Early Start Denver Model for early intervention on autistic children*. Master's Dissertation, Zhejiang University of Technology (in Chinese).

Yoshimura, Y., Tanaka, S. and Haramaki, T. R. 2019. Early intervention and perspectives for children with autism spectrum disorder in Japan. *Pediatric Medicine*, 12, 1–6.

Yu, Q., Li, E., Li, L. and Liang, W. 2020. Efficacy of interventions based on applied behavior analysis for autism spectrum disorder: A meta-analysis. *Psychiatry Investigation*, 5, 432.