

Formula *Effect size* yang digunakan adalah sebagai berikut:

$$Hedges' g = \frac{M_1 - M_2}{SD_{pooled}}$$

$$SD_{pooled} = \sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}}$$

Keterangan:

M_1 = mean eksperimen

M_2 = mean kontrol

SD_1^2 = standar deviasi eksperimen

SD_2^2 = standar deviasi kontrol

SD_{pooled} = standar deviasi gabungan

n_1 = jumlah sampel eksperimen

n_2 = jumlah sampel kontrol

Interpretasi ukuran efek, menggunakan klasifikasi (Tamur, 2020)

Tabel 3.2 Kategori Effect Size

Effect Size (ES)	Kategori
$ES \leq 0,15$	Efek yang diabaikan
$0,15 < ES \leq 0,40$	Efek kecil
$0,40 < ES \leq 0,75$	Efek sedang
$0,75 < ES \leq 1,10$	Efek tinggi
$1,10 < ES \leq 1,45$	Efek sangat tinggi
$1,45 < ES$	Pengaruh tinggi

3.9 Prosedur Langkah Pengujian

Penelitian meta-analisis akan memberikan hasil temuan yang valid dan reliabel jika memperhatikan langkah-langkah pengujian yang sesuai dengan konsep penelitian dengan metode meta-analisis itu sendiri. Adapun langkah-langkah pengujian yang harus dilakukan dalam meta-analisis yaitu:

1. Uji Reliabilitas Koding

Pengujian terhadap reliabilitas koding merupakan langkah awal yang perlu dilakukan untuk memastikan bahwa ekstraksi data yang dilakukan antar pengkoding akan menunjukkan variabilitas dari proses justifikasi terhadap data-data yang ada di dalam studi primer (Ustun & Eryilmaz, 2014). Pengujian ini juga bermanfaat untuk menjamin keketatan dari prosedur meta-analisis sehingga objektivitas penelitian dapat dicapai. Reliabilitas koding dalam penelitian ini dilakukan dengan mengukur konsistensi antar pengkoding (*inter-rater reliability/IRR*) menggunakan uji Cohen's Kappa dengan pertimbangan bahwa terdapat 2 orang koder yang melakukan ekstraksi data dan pengkodean. Pengujian IRR dengan Cohen's Kappa ini akan dibantu oleh program SPSS 22 dan kemudian hasilnya diinterpretasi berdasarkan kategori koefisien Kappa (Viera & Garret, 2005) yang disajikan pada Tabel 3.3 berikut.

Tabel 3.3 Kategori Koefisien Kappa

Kappa (κ)	Kategori <i>Strength of Agreement</i>
$\kappa \leq 0,02$	Rendah (<i>Poor</i>)
$0,20 < \kappa \leq 0,40$	Lumayan (<i>Fair</i>)
$0,40 < \kappa \leq 0,60$	Cukup (<i>Moderate</i>)
$0,60 < \kappa \leq 0,80$	Kuat (<i>Good</i>)
$0,80 < \kappa \leq 0,10$	Sangat Kuat (<i>Very Good</i>)

2. Uji Bias Publikasi

Pengujian bias publikasi dalam meta-analisis merupakan salah satu langkah yang tidak bisa dilewatkan dalam melakukan analisis. Bias publikasi sangat erat kaitannya dengan reliabilitas hasil meta-analisis (Tamur dkk., 2020). Uji bias publikasi ini dapat menjadi landasan argumen bahwa penelitian primer yang digunakan sudah valid dan dapat diteruskan untuk analisis lanjutan. Selain itu, uji bias publikasi ini juga dilakukan dengan tujuan mengantisipasi adanya kecenderungan bahwa seluruh studi yang dipublikasikan dan diterbitkan dalam jurnal merupakan studi yang

memberikan hasil yang signifikan saja sehingga *effect size* akan bernilai besar daripada nilai yang sebenarnya (Borenstein dkk., 2009). Uji bias publikasi ini dapat dilakukan melalui beberapa cara antara lain melalui diagram corong (*funnel plot*), nilai *fail-safe N* (FSN), dan nilai *Trim and Fill* (Kul, Celik & Aksu, 2018).

a. Diagram corong (*Funnel Plot*)

Funnel plot merupakan diagram hasil uji bias publikasi yang berbentuk segitiga seperti corong dimana *effect size* dari setiap studi independen tersebar di dalam segitiga dalam bentuk titik-titik. Garis simetri vertikal pada segitiga menunjukkan *effect size* gabungan sehingga dapat dilihat dan dibandingkan kedudukan dari masing-masing *effect size* studi terhadap *effect size* gabungan. Jika sebaran *effect size* pada diagram corong simetris terhadap *effect size* gabungan, maka tidak terdapat bias publikasi. Selain itu, jika terdapat titik (*effect size*) yang cukup jauh dari sebaran titik-titik lainnya dan dari nilai efek gabungan maka studi dengan *effect size* tersebut berpeluang menjadi studi dengan bias penelitian sehingga studi tersebut perlu dipangkas agar tidak mempengaruhi hasil temuan meta-analisis.

b. *Fail-Safe N* (FSN) Test

FSN merupakan cara lain yang dapat digunakan untuk menguji bias publikasi. FSN menggunakan nilai probabilitas untuk menghitung bias publikasi. Perhitungan FSN dilakukan dengan menggunakan rumus $\frac{N}{5k+10}$ dimana N merupakan nilai yang diperoleh dari hasil perhitungan dengan CMA 3.0 dan k merupakan banyaknya studi yang terlibat dalam analisis. Selanjutnya, jika diperoleh nilai $\frac{N}{5k+10} > 1$, maka dapat dikatakan bahwa seluruh studi yang digunakan tahan terhadap bias publikasi sehingga analisis lanjutan dapat dilakukan.

c. *Trim and Fill Test*

Uji bias publikasi lainnya yaitu uji *Trim and Fill*, dimana hasil uji ini memberikan informasi tentang banyaknya studi yang harus dipangkas untuk menghindari adanya bias dalam analisis. Hasil uji *Trim and Fill* ini erat kaitannya dengan tampilan hasil dalam diagram *funnel plot* yang menunjukkan posisi titik yang berada jauh dari sebaran titik-titik lainnya

dan nilai efek gabungan. Dengan adanya hasil dari uji *Trim and Fill* ini, maka posisi atau identitas studi yang harus dipangkas dari analisis menjadi lebih mudah untuk diketahui sehingga interpretasi nilai ukuran efek yang berlebihan dapat dihindari.

Bias dalam meta-analisis dapat disebabkan oleh beberapa faktor seperti adanya kelemahan dalam penentuan sampel, dan penggunaan instrumen penelitian atau prosedur penelitian yang kurang sesuai. Adanya bias dalam suatu penelitian dapat mempengaruhi reliabilitas hasil penelitian sehingga uji bias publikasi melalui tiga langkah di atas sangat penting untuk dilakukan. Jika dalam langkah pengujian bias publikasi ditemukan adanya bias dari hasil penelitian terdahulu yang sedang dianalisis maka perlu dilakukan analisis terhadap variabel moderator. Sebaliknya, jika tidak ditemukan adanya bias, maka analisis dapat dilanjutkan dengan menginterpretasi hasil meta-analisis menggunakan model estimasi efek yang sesuai (DeCoster, 2009).

3. Uji Heterogenitas

Uji heterogenitas dalam meta-analisis dilakukan dengan tujuan untuk mengetahui adanya keanekaragaman yang terdapat pada distribusi ukuran efek. Dengan mengetahui ada atau tidaknya heterogenitas antar studi ini, maka dapat ditentukan model estimasi statistik yang sesuai dengan kondisi alamiah data dan sesuai untuk analisis lanjutan. Model estimasi efek dalam meta-analisis terbagi menjadi dua yaitu model efek acak (*random effect model*) dan model efek tetap (*fixed effect model*) (Brockwell & Gordon, 2007).

Model efek tetap (*fixed effect model*) merupakan model efek yang mengestimasi bahwa ukuran efek secara keseluruhan memperoleh ukuran efek dalam populasi yang sama, atau dengan kata lain ukuran efek tunggal. Model efek tetap ini cenderung menampilkan bobot rata-rata hasil meta-analisis jika melibatkan studi atau sampel dalam jumlah yang besar. Oleh sebab itu, jika dalam suatu meta-analisis melibatkan studi dengan skala besar, maka studi dengan skala kecil tidak akan berdampak signifikan terhadap hasil analisis maupun interpretasi akhir dari meta-analisis.

Sementara itu, model efek acak (*random effect model*) merupakan model efek yang mengestimasi bahwa populasi ukuran efek yang diperoleh serupa

namun tidak identik (Borenstein dkk., 2009). Model ini juga menampilkan bobot rata-rata dari dampak ukuran efek pada kelompok penelitian, tanpa memandang bobot dari masing-masing studi. Penetapan model estimasi ini didasarkan pada interpretasi nilai Q Cochran atau p -value. Jika pada *output* CMA 3.0 menunjukkan hasil terdapat heterogenitas yang signifikan, maka model estimasi yang sesuai dengan kondisi tersebut adalah model efek acak. Sebaliknya, jika tidak terdapat heterogenitas yang signifikan, maka model efek tetap menjadi model estimasi yang sesuai untuk digunakan dalam analisis lanjutan. Uji heterogenitas ini dilakukan dengan melihat hasil uji Q -value atau p -value. Kriteria penolakan H_0 yaitu jika $Q_{value} > \chi^2_{(df;0,05)}$ atau $p < 0,05$. Jika nilai $Q_{value} > \chi^2_{(df;0,05)}$ atau $p < 0,05$ artinya hipotesis pada uji heterogenitas *effect size* ditolak yang berarti bahwa terdapat heterogenitas ukuran efek antar studi sehingga model estimasi yang digunakan adalah model efek acak.

4. Analisis Karakteristik Studi

Berdasarkan penjelasan pada bagian sebelumnya yaitu analisis terhadap karakteristik studi dilakukan jika terdapat heterogenitas ukuran efek antar studi, atau dengan kata lain, model estimasi yang digunakan adalah model efek acak (Borenstein dkk., 2009). Merujuk kepada hipotesis yang disusun dalam penelitian ini, peneliti memperkirakan bahwa terdapat heterogenitas ukuran efek sehingga perlu dilakukan analisis lebih lanjut terhadap intervensi dari variabel moderator atau karakteristik studi yang mungkin mempengaruhi efektivitas dari penerapan model pembelajaran kontemporer terhadap keterampilan berpikir tingkat tinggi. Oleh sebab itu, adapun beberapa karakteristik studi yang diinvestigasi dan dianalisis dalam penelitian ini beserta pertimbangan terhadap data yang diperoleh dari pengkodean dan ekstraksi data studi primer antara lain sebagai berikut:

a. Tahun Penelitian

Karakteristik tahun penelitian dalam studi merupakan artikel penelitian dari tahun 2016-2021 yang kemudian dianalisis dengan membuat pengelompokan menjadi tahun 2016-2017 dan tahun 2018-2021.

Pembagian kelompok ini dipilih dengan pertimbangan bahwa pada tahun 2018 adalah masa-masa ujian nasional sekolah menengah (SMP dan SMA) sudah menerapkan HOTS.

b. Tingkat Pendidikan

Karakteristik tingkat pendidikan yang dianalisis dalam penelitian ini dikelompokkan berdasarkan tingkatan sekolah menengah yang ada di Indonesia berdasarkan Sistem Pendidikan Nasional yaitu tingkat Sekolah Menengah Pertama (SMP) / sederajat, dan Sekolah Menengah Atas (SMA) / sederajat.

c. Kelas Penelitian

Karakteristik kelas penelitian dikelompokkan menjadi kelas VII dan VIII untuk tingkat SMP, serta X dan XI untuk tingkat SMA. Pembagian kelompok ini sudah sangat jelas karena mengikuti aturan yang ada dalam Sistem Pendidikan Nasional.

d. Ukuran Sampel

Karakteristik ukuran sampel yang dianalisis diklasifikasikan menjadi ukuran sampel lebih dari 30 partisipan dan ukuran sampel kurang dari atau sama dengan 30. Pengklasifikasian kelompok ukuran sampel ini didasarkan pada teori statistik yaitu teori sampling yang menjelaskan mengenai syarat ukuran sampel kecil dan sampel besar.

e. Model-model pembelajaran kontemporer (berdasarkan model pembelajaran yang di gunakan)

Karakteristik model-model pembelajaran kontemporer yang di gunakan yaitu model *Discovery/Inquiry*, model pembelajaran *Problem-Based Learning* (PBL), model pembelajaran *Project-Based Learning* (PjBL), *creative problem solving* (CPS) dan *problem solving*. Penetapan karakteristik model pembelajaran kontemporer ini bertujuan untuk melihat bagaimana besaran pengaruh dari setiap model pembelajaran kontemporer yang di analisis tersebut.

Selanjutnya setiap karakteristik studi yang dianalisis dengan bantuan program CMA V3 diuji hipotesisnya dengan melihat nilai Q -value atau p -value. Kriteria penolakan H_0 yaitu jika $Q_{value} > Q_{tabel}$ atau $p < 0,05$. Jika

nilai $Q_{value} > Q_{tabel}$ atau $p < 0,05$ artinya hipotesis nol ditolak yang berarti bahwa terdapat perbedaan signifikansi pengaruh dari penerapan model pembelajaran kontemporer terhadap keterampilan berpikir tingkat tinggi (HOTS) ditinjau dari jenis karakteristik studi yang dianalisis. Sebaliknya jika hipotesis nol diterima, hal ini bermakna bahwa tidak terdapat perbedaan signifikansi pengaruh dari penerapan model pembelajaran kontemporer terhadap keterampilan berpikir tingkat tinggi (HOTS) ditinjau dari jenis karakteristik studi yang dianalisis.

BAB IV

TEMUAN DAN PEMBAHASAN

Bab ini berisi beberapa data temuan berupa hasil perhitungan akhir pada studi meta-analisis model pembelajaran kontemporer terhadap keterampilan berpikir tingkat tinggi (*Higher Order Thinking Skill*) dengan menentukan nilai besaran pengaruh (*Effect size*) pada pembelajaran matematika.

4.1 Temuan

Data pada penelitian ini berjumlah 17 artikel publikasi ilmiah yang sesuai dengan kriteria penelitian akan tetapi hanya 14 artikel penelitian yang dapat dianalisis dari 70 artikel yang telah dikumpulkan berdasarkan kriteria inklusi. Adapun data hasil penelitian tersebut sebagai berikut:

Tabel 4.1 rekapitulasi hasil data pengkodean

kode	Sitasi	Data statistic							t-value	p-value
		Pembelajaran kontemporer				Model pembanding				
		Model pembelajar ran	Rata- rata	Standar deviasi	Ukuran sampel	Rata- rata	Standar deviasi	Ukuran sampel		
A1	Yulianto, 2019	PBL	65,54 95	9,4413 6	91	54,6 81	9,941	94	0,032	0,000
A2	Nurfitriy anti, 2016	PjBL	85,19	7,24	35	77,9 3	7,17	35	3,87	0,000
A3	Riadi, 2016	PBL	78,28			68,6			2,564	0,014
A4	Riadi. 2016	PBL	82,96			73,9 3			3,309	0,003

A5	Hodiyono, 2018	PBL	73	9,51					11,64	0,000
A6	Juanengsih, 2020	PBL	74,02	12,56		66,79	12,18		2,331	0,000
A7	Wardani, 2019	PBL	77,11			71,15			2,311	0,025
A8	Setiasih, 2019	PjBL	91,74	4,56		74,08	8,76			
A9	Sasmi, 2020	CPS							8,737	0
A10	Herutomo, 2019	CPS	75,77	4,32		70,21	4,29			
A11	Boham, 2021	Discovery	76,54		26	55		26	7,18	0,000
A12	Suherman, 2019	PjBL	59,11	10,83		38,78	10,54			
A13	Sardin, 2018	Problem Solving	27,918	20,5	37	7,648	6,923	37	11,783	0,000
A14	Tambunan, 2019	Problem Solving	77,54	5,53	138	73,63	9,45	139		

Tabel 4.1 menyajikan data rekapitulasi dari hasil pengkodean yang telah diekstrak berdasarkan lembar pengkodean. Data yang digunakan dalam penelitian ini adalah data statistik dari nilai rata-rata, standar deviasi, ukuran sampel, nilai *t-value* dan *p-value*. Namun, tidak seluruh data statistik memiliki data yang lengkap sehingga studi tersebut dikelompokkan kedalam beberapa kelompok yaitu 1) kelompok dengan data statistik nilai rata-rata, standar deviasi, dan ukuran sampel sebanyak 10 artikel, 2) kelompok dengan data statistik ukuran sampel, *t-value*, dan *p-value* sebanyak 4 artikel. selanjutnya, data hasil pengkodean untuk karakteristik studi yang diamati disajikan pada tabel 4.2

Tabel 4.2 rekapitulasi hasil data pengkodean karakteristik studi

Ko de	Sitasi	Tingkat pendidi kan	Kelas peneliti an	Tahun peneliti an	Ukur an samp el	Pengind eks publikas i	Pering kat jurnal
A1	Yulianto,2 019	SMP	VIII	2018- 2021	>30	Google scholar	Non Sinta
A2	Nurfitriyan ti, 2016	SMA	XI	2016- 2017	>30	Sinta	S2
A3	Riadi, 2016	SMP	VIII	2016- 2017	≤30	Sinta	S3
A4	Riadi. 2016	SMP	VIII	2016- 2017	≤30	Sinta	S3
A5	Hodiyono, 2018	SMA	X	2018- 2021	>30	Sinta	S5
A6	Juanengsih, 2020	SMA	XI	2018- 2021	>30	Non index (Prosed ing scopus)	Non sinta
A7	Wardani,20 19	SMP	VII	2018- 2021	≤30	Non index	Non sinta
A8	Setiasih, 2019	SMP	VIII	2018- 2021	>30	Sinta	S5
A9	Sasmi, 2020	SMP	VII	2018- 2021	≤30	Sinta	S4
A1 0	Herutomo, 2019	SMA	XI	2018- 2021	>30	Sinta	S2

A1 1	Boham, 2021	SMP	VIII	2018- 2021	≤ 30	Google Scholar	Non sinta
A1 2	Suherman, 2019	SMA	XI	2018- 2021	> 30	Sinta	S5
A1 3	Sardin, 2018	SMA	X	2018- 2021	> 30	Sinta	S5
A1 4	Tambunan, 2019	SMA	X	2018- 2021	> 30	Google Scholar	Non sinta

Berdasarkan 14 data ukuran efek yang diperoleh, dapat dilihat bahwa penelitian pada studi dilakukan dan dipublikasikan pada rentang tahun 2016–2021 dimana 3 (tiga) studi dipublikasi pada kelompok rentang tahun 2016–2017 dan 11 studi lainnya pada kelompok rentang tahun 2018–2021. Selain itu, penelitian yang dilakukan pada studi primer juga beragam pada jenjang sekolah menengah dengan persentase studi pada jenjang SMP dan SMA adalah 50% untuk keduanya. Jika mengacu pada kelas sesuai dengan sistem Pendidikan Nasional, terdapat 6 tingkat kelas pada sekolah menengah. Namun pada penelitian ini kelas yang menjadi pusat perhatian peneliti hanya terdapat pada kelas VII dan VIII untuk jenjang SMP, sedangkan untuk jenjang SMA/ sederajat hanya terdapat untuk kelas X dan XI. Selain itu, jika diamati dari ukuran sampel, maka terdapat dua kelompok ukuran sampel yaitu kelompok lebih dari 30 sampel sebanyak 10 studi dan kelompok kurang dari atau sama dengan 30 sampel sebanyak 4 studi. Sementara itu, pada data pengindeks publikasi, diperoleh informasi bahwa 9 studi merupakan artikel penelitian pada jurnal yang terindeks Sinta, sementara 3 studi terindeks oleh Google Scholar, dan 2 studi tidak terindeks karena merupakan proseding. Terakhir, pada informasi peringkat Sinta diperoleh bahwa terdapat empat peringkat Sinta dari studi yang diperoleh yaitu Sinta 2, Sinta 3, Sinta 4, dan Sinta 5.

Selain hasil rekapitulasi yang disajikan di atas peneliti juga melakukan uji reliabilitas koding yang di bantu oleh dua orang koder untuk menjamin reliabilitas koding yang digunakan. Setelah data dari kedua koder telah terkumpul, maka data

tersebut dikonversi ke dalam data kategorikal. Terdapat dua pilihan konversi yang digunakan peneliti, diantaranya:

- 1) Pada item Kode, Sitasi, Rata-rata, Deviasi Standar, Ukuran Sampel, *t-value*, *p-value*, Kelas Penelitian, Nama Jurnal/Prosiding, *Corresponding*, dan *Email*, peneliti menggunakan nilai 1 untuk menyatakan “setuju” terhadap hasil koding peneliti dan nilai 0 untuk menyatakan “tidak setuju”.
- 2) Pada item Tahun Penelitian, Tingkat Pendidikan, Ukuran Sampel, Pengindeks Publikasi, dan Peringkat Jurnal Sinta, peneliti menggunakan nilai sesuai dengan banyak kelompok pada setiap item karakteristik studi.

Hasil konversi koding dari kedua koder (*rater*) diuji dan dianalisis reliabilitas kodingnya dengan mengukur konsistensi antar pengkoding (*inter-rater reliability/IRR*) menggunakan statistik Cohen’s Kappa. Pengujian ini dibantu oleh program SPSS 22 dan hasil beserta interpretasi kategori persetujuan antar *rater* disajikan pada Tabel 4.3 berikut.

Tabel 4.3 Hasil Uji IRR Dengan Cohen’s Kappa

No.	Item data	Kappa (k)	p-value	Kategori
1	Kode	1	0,000	Sangat kuat
2	Sitasi	1	0,000	Sangat kuat
3	<i>Mean</i> model kontemporer	0,632	0,000	Kuat
4	Standar Deviasi model kontemporer	1		Sangat kuat
5	Ukuran Sampel model kontemporer	1		Sangat kuat
6	<i>Mean</i> pembanding	1		Sangat kuat

7	Deviasi Standar pembanding	1		Sangat kuat
8	Ukuran Sampel pembanding	1		Sangat kuat
9	<i>t-value</i>	1		Sangat kuat
10	<i>p-value</i>	1		Sangat kuat
11	Tahun Penelitian	1		Sangat kuat
12	Tingkat Pendidikan	1		Sangat kuat
13	Kelas Penelitian	1		Sangat kuat
14	Ukuran Sampel	1		Sangat kuat
16	Pengindeks Publikasi	1		Sangat kuat
17	Nama Jurnal/Prosiding	1		Sangat kuat
18	Peringkat Jurnal Sinta	1		Sangat kuat
19	<i>Corresponding email</i>	1		Sangat kuat

Merujuk kepada Tabel 4.3 tersebut, dapat diinterpretasi bahwa terdapat 18 item yang kategori atau tingkat persetujuan kedua penilainya sangat kuat, sementara 1 item berkategori kuat. Hasil uji IRR dengan Cohen's Kappa ini menunjukkan bahwa seluruh item yang digunakan dalam koding reliabel sehingga layak untuk digunakan dalam tahapan analisis selanjutnya. Oleh sebab itu, langkah berikutnya yaitu menguji *effect size* dari studi yang melibatkan dengan menggunakan data statistik yang telah diperoleh pada hasil koding dan kemudian melakukan analisis hasil ukuran efek tersebut.

4.1.1 Data Hasil *Effect Size* Berdasarkan Kategori

Data besar pengaruh (*effect size*) artikel publikasi ilmiah penggunaan model pembelajaran kontemporer terhadap keterampilan berpikir tingkat tinggi (HOTS). Berdasarkan kategori Hedges'g hasil perhitungan keseluruhan effect size dengan menggunakan perangkat lunak CMA dari masing-masing studi disajikan pada Tabel 4.4 output CMA sebagai berikut:

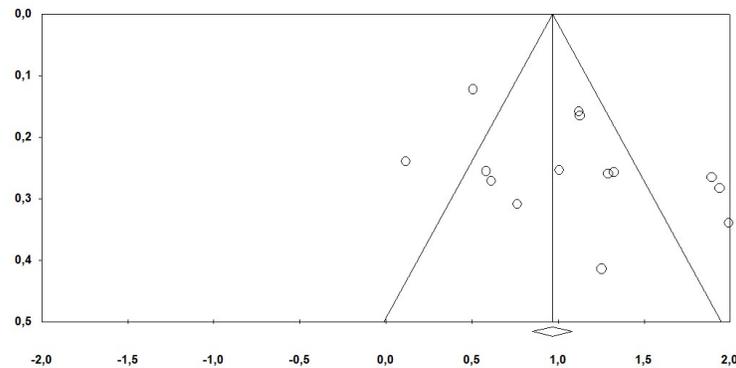
Tabel 4.4 *Effect size* secara keseluruhan

No	Kode Artikel	Hedges's	Kategori <i>effect size</i>
1	A1	1,121	Efek sangat tinggi
2	A2	1,008	Efek tinggi
3	A3	0,765	Efek tinggi
4	A4	1,254	Efek sangat tinggi
5	A5	1,940	Pengeruh tinggi
6	A6	0,583	Efek sedang
7	A7	0,612	Efek sedang
8	A8	0,161	Efek kecil
9	A9	1,128	Efek sangat tinggi
10	A10	1,291	Efek sangat tinggi
11	A11	1,991	Pengaruh tinggi
12	A12	1,878	Pengaruh tinggi
13	A13	0,506	Efek sedang
14	A14	1,325	Efek sangat tinggi
	Effect size gabungan	1,040	Efek Tinggi

Hasil data analisis pada Tabel 4.4 menunjukkan bahwa artikel berada pada kategori efek kecil 1 studi, kategori sedang 3 studi, kategori tinggi 2 studi, kategori efek sangat tinggi 5 studi, dan kategori pengaruh tinggi sebanyak 3 studi. Dari hasil tersebut akan dilakukan uji bias publikasi. Dalam studi meta analisis adalah melakukan uji bias publikasi untuk melihat apakah ada studi yang berbeda secara sistematis dari semua hasil studi yang akan dianalisis. Uji bias publikasi dapat

dilakukan dengan menggunakan hasil dari *funnel plot*, nilai *fail-safe N* (FSN), dan nilai *trim and fill*.

Diagram funnel plot yang menunjukkan bentuk simetris disekitar effect size gabungan, dapat disimpulkan bahwa tidak ada pengaruh bias publikasi pada meta analisis. Diagram funnel plot disajikan pada gambar 4.1 berikut ini:



Gambar 4.1 funnel plot of standard error by hedges's g

Berdasarkan gambar 4.1 tampak bahwa nilai effect size tersebar antara 0 – 2 dan terlihat asimetris di tengah funnel plot, dan disisi kiri dan kanan garis vertikal. Garis vertikal tersebut menunjukkan effect size gabungan. Kemudian, untuk menentukan apakah ada data studi yang perlu untuk dihilangkan, sehingga perlu melakukan uji Trim dan Fill. Nilai trim and fill berkaitan erat dengan funnel plot untuk mengidentifikasi jumlah studi yang harus dihilangkan dari analisis untuk menghindari bias publikasi dan adanya interpretasi effect size yang berlebihan. Hasil uji Trim dan Fill dapat dilihat pada Tabel 4.5 di bawah ini.

Tabel 4.5 Publikasi Bias dengan Uji *Trim and fill*

	studi yang dihilangkan	Poin estimasi	Limit terendah	Limit tertinggi	Nilai Q
Nilai yang diamati		1,07192	0,79366	1,35047	69,73089
Nilai yang disesuaikan	4	0,80890	0,51015	1,10766	127,01408

Pada tabel 4.4 diatas dapat diperhatikan bahwa terdapat 7 studi yang harus di potong atau di keluarkan dari analisis. Berdasarkan funnel plot pada

gambar 4.1 peneliti mencoba mengeluarkan satu studi yaitu A12 dengan alasan bahwasanya pada funnel plot terlihat berada jauh dari funnel plot. Setelah artikel dengan kode A12 tersebut dikeluarkan dari analisis, selanjutnya dilakukan analisis ulang terhadap uji *Trim and Fill* sehingga diperoleh hasil yang disajikan pada Tabel 4.6

Tabel 4.6 Publikasi Bias dengan Uji *Trim and fill* -1 studi

	studi yang dihilangkan	Poin estimasi	Limit terendah	Limit tertinggi	Nilai Q
Nilai yang diamati		1,02156	0,74654	1,29659	57,58879
Nilai yang disesuaikan	0	1,02156	0,74654	1,29659	57,58879

Analisis ulang yang dilakukan melalui uji *Trim and Fill* menunjukkan bahwa tidak ada studi yang harus dipangkas lagi, sehingga 13 studi lainnya layak untuk digunakan dalam tahapan analisis selanjutnya. Dengan demikian analisis bias publikasi dengan nilai FSN dan Trim and Fill bahwa terdapat 13 studi tidak rentan terhadap bias publikasi. Dapat diperhatikan bahwa penyebaran distribusi effect size asimetris, maka nilai *fail-safe* N (FSN) diidentifikasi untuk menghitung probabilitas bias publikasi. Hasil nilai *fail-safe* N (FSN) diperoleh dengan bantuan perangkat lunak CMA nilai N (FSN) adalah 1261. Dengan rumus $\frac{N}{(5k+10)} = \frac{787}{(5(13)+10)} = \frac{787}{75} = 10,493$, karena hasilnya adalah 10,493 lebih besar dari 1. Hasil penelitian ini menunjukkan bahwa semua studi yang terlibat dalam meta analisis tidak rentan terhadap bias publikasi.

Selain ketiga hasil uji bias publikasi tersebut, analisis sensitivitas yang dapat dilihat pada Lampiran juga dilakukan dengan menggunakan menu tab “*One Study Removed*” pada program CMA 3.0 untuk menginvestigasi apakah terdapat himpunan data effect size yang nilainya abnormal (berbeda signifikan dari yang lainnya). Tampilan diagram yang diperoleh dari tab “*One Study Removed*” disajikan pada Gambar 4.2. Berdasarkan model efek acak,

Model	nama	metode	KELAS	TINGKAT	Statistics with study removed							Std diff in means (95% CI) with study removed					
					Point	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value	-2,00	-1,00	0,00	1,00	2,0	
A1	FBL	SMP	VIII		1,015	0,156	0,024	0,709	1,321	6,495	0,000						
A2	FBL	SMA	XI		1,025	0,151	0,023	0,728	1,321	6,774	0,000						
A3	FBL	SMP	VIII		1,042	0,149	0,022	0,750	1,334	6,989	0,000						
A4	FBL	SMP	VIII		1,009	0,146	0,021	0,722	1,295	6,898	0,000						
A5	PBL	SMA	X		0,945	0,132	0,017	0,686	1,204	7,156	0,000						
A6	PBL	SMA	XI		1,059	0,149	0,022	0,767	1,351	7,109	0,000						
A7	PBL	SMP	VII		1,056	0,149	0,022	0,764	1,348	7,086	0,000						
A8	PBL	SMP	VIII		1,096	0,136	0,018	0,829	1,362	8,059	0,000						
A9	CPS	SMP	VII		1,014	0,195	0,024	0,709	1,319	6,523	0,000						
A10	CPS	SMA	XI		1,000	0,148	0,022	0,709	1,291	6,738	0,000						
A11	D/I	SMP	VIII		0,953	0,135	0,018	0,688	1,217	7,048	0,000						
A13	PS	SMA	X		1,075	0,142	0,020	0,797	1,353	7,573	0,000						
A14	PS	SMA	X		0,997	0,148	0,022	0,707	1,287	6,740	0,000						
Random					1,022	0,140	0,020	0,747	1,297	7,280	0,000						

Gambar 4.2 model Efek acak

Tahap selanjutnya yaitu menguji atau mengidentifikasi heterogenitas dalam distribusi ukuran efek. Analisis heterogenitas ini dilakukan dengan mengidentifikasi hasil keluaran (*output*) dari program CMA 3.0 yang disajikan dalam tabel berikut.

Tabel 4.7 Heterogenitas dari Pengaruh Ukuran Distribusi

Heterogenitas				
Q hitung	Df	P hitung	I-squared	Q table
57,589	12	0	91,273	21,026

Berdasarkan Tabel 4.7 menunjukkan bahwa nilai Q-hitung adalah 57,589 lebih besar dari nilai Q-tabel ($\alpha = 5\%$) yaitu 21,026. Hal ini berarti memiliki heterogenitas dari pengaruh distribusi effect size. Dengan kata lain, ukuran efek dari masing-masing studi berbeda dan memberikan pengaruh yang diperhitungkan. sehingga model estimasi yang digunakan dalam analisis ini adalah effect random model. Model efek acak (*random effect model*) dimana model estimasi ini mampu mengestimasi bahwa populasi ukuran efek yang diperoleh serupa namun tidak identik. Dengan kata lain, melalui model estimasi ini diasumsikan bahwa populasi ukuran efek berdistribusi normal di sekitar nilai rata-rata dan deviasi standar. Informasi *effect size* studi primer dengan model estimasi *random effect model* disajikan pada tabel berikut.

Tabel 4.8 Effect Size Menurut Model Efek Acak

Model	N	Z	P	<i>Effect size dan 95% interval kepercayaan</i>				
				Hedges	Standar Error	Variance	Lower Limit	Upper Limit
Random	13	7,280	0,000	1,022	0,140	0,020	0,747	1,297

Berdasarkan Tabel 4.8 hasil dari 13 studi yang dianalisis berdasarkan effect random model dengan interval kepercayaan 95% memiliki batas bawah 0,747 dan batas atas 1,297 dengan rata-rata nilai effect size yaitu 1,022 dengan kategori efek tinggi. Sebagai hasil perhitungan uji Z untuk menentukan signifikansi statistik, skor Z diperoleh 7,280 dan nilai $p < 0,05$.

Analisis selanjutnya dilakukan untuk melihat efektivitas penggunaan model pembelajaran terhadap keterampilan berpikir tingkat tinggi (HOTS) ditinjau dari karakteristiknya yaitu jenjang pendidikan, kelas penelitian, jumlah sampel, tahun penelitian dan model-model pembelajaran kontemporer yang digunakan.

Tabel 4.9 Rekapitulasi Hasil Analisis ditinjau dari Karakteristik

Karakteristik	Grup	N	Hedges'g	Test of null		Heterogenitas		
				Z	P	Q	Df	P
Jenjang Pendidikan	SMP	7	0,971	4,985	0,000	0,141	1	0,707
	SMA	6	1,084	4,756	0,000			
Tingkatan Kelas	VII	2	0,915	3,604	0,000	0,401	4	0,940
	VIII	5	1,021	3,467	0,001			
	X	3	1,230	2,724	0,000			
	XI	3	0,959	4,672	0,000			