

BAB III

METODE PENELITIAN

3.1 Desain Penelitian

Penelitian ini bertujuan untuk menganalisis dan mengevaluasi data statistik dari beberapa penelitian yang berupa studi primer terkait pengaruh model *flipped classroom* terhadap kemampuan berpikir tingkat tinggi siswa. Meta-analisis digunakan untuk mencapai tujuan tersebut. Dalam hal ini, meta-analisis termasuk ke dalam kategori penelitian tinjauan literatur, yaitu sintesis literatur sebelumnya mengenai topik tertentu. Meta-analisis termasuk ke dalam tinjauan literatur jenis metode *research synthesis* karena fokusnya adalah pada hasil penelitian dari literatur yang sudah ada.

Meta-analisis hanyalah salah satu dari banyak cara untuk merangkum, mengintegrasikan, dan menafsirkan kumpulan studi. Pendekatan kuantitatif digunakan untuk menilai secara sistematis hasil penelitian sebelumnya untuk ditarik sebuah kesimpulan (Juandi & Tamur, 2020). Meta-analisis merupakan metode sintesis dengan analisis statistik dari ukuran efek. Gene Glass dalam Rao et al. (2017) menuliskan, “*I use it (meta-analysis) to refer to the statistical analysis of a large collection of results from individual studies for the purpose of integrating findings. It connotes a rigorous alternative to the casual, narrative discussions of research studies which typify our attempts to make sense of the rapidly expanding literature.*” Berdasarkan hal tersebut, meta-analisis termasuk ke dalam teknik analisis data yang digunakan untuk mengkombinasikan dan menganalisis hasil-hasil dari banyak studi independen.

Meta-analisis menghitung ukuran efek dari banyak studi, sehingga kesimpulan yang diperoleh akan lebih kuat daripada pengujian hipotesis pada studi individual. Meta-analisis dipandang sebagai sebuah metode objektif kajian literatur karena penggunaan ukuran efek. Ukuran efek adalah cara untuk mengukur perbedaan antara dua kelompok yang diteliti. Metode meta-analisis mengabaikan interpretasi yang bersifat subjektif dari berbagai penelitian dengan topik yang sama.

Tahapan dalam metode meta-analisis antara lain (1) merumuskan masalah penelitian, (2) penelusuran literatur, (3) pengkodean, (4) analisis statistik, dan (5) representasi dan interpretasi hasil (Cooper et al., 2019; Hunter & Schmidt, 2004;

Juandi & Tamur, 2020). Selain analisis studi secara keseluruhan, beberapa karakteristik dijadikan tinjauan dalam penelitian ini untuk melihat apakah perlakuan akan mencapai tingkat efektivitas yang lebih tinggi jika diatur pada kondisi tertentu. Karakteristik atau variabel moderator tersebut antara lain jenjang pendidikan, ukuran sampel, tahun studi, jenis media pembelajaran, dan indikator berpikir tingkat tinggi.

3.2 Populasi dan Sampel

Sampel dalam sebuah penelitian adalah kelompok atau kumpulan di mana informasi diperoleh, sedangkan kelompok yang lebih besar di mana hasil penelitian diharapkan dapat diterapkan disebut populasi (Fraenkel & Wallen, 2009). Populasi dan sampel dari sebuah penelitian tidak selalu berupa individu, tetapi juga bisa berbentuk objek. Populasi pada penelitian ini adalah studi primer di Indonesia tentang model *flipped classroom* pada kemampuan berpikir tingkat tinggi yaitu kemampuan berpikir kritis, berpikir kreatif, penalaran, dan pemecahan masalah. Sedangkan, sampelnya adalah studi primer yang dilakukan oleh peneliti atau mahasiswa dengan metode eksperimen di Indonesia dan memenuhi kriteria inklusi yang ditentukan.

3.3 Kriteria Inklusi

Seperti halnya pada studi primer, membangun protokol menyeluruh yang mendefinisikan semua kriteria inklusi dan eksklusi untuk memperoleh data merupakan hal yang penting untuk dilakukan. Kriteria inklusi yang digunakan untuk menentukan sampel penelitian adalah elemen PICOS (*Populations, Interventions, Comparators, Outcomes, dan Study designs*) karena elemen PICOS merupakan bagian utama dari protokol yang menginformasikan seluruh desain studi. Lima elemen tersebut dibahas secara rinci untuk mengembangkan kriteria inklusi yaitu sebagai berikut.

1. *Populations*: populasi yang dianalisis dalam studi primer. Populasi yang dipilih untuk penelitian meta-analisis ini adalah siswa pada jenjang pendidikan Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), Sekolah Menengah Atas (SMA), dan perguruan tinggi di Indonesia.

2. *Interventions*: perlakuan yang diberikan dalam studi primer. Perlakuan yang dipilih untuk penelitian ini adalah penerapan model *flipped classroom* dalam pembelajaran matematika.
3. *Comparators*: adanya pembandingan untuk kelas dengan model *flipped classroom*. Dalam penelitian ini kelas pembandingan dari studi primer yang dipilih adalah model pembelajaran konvensional atau model pembelajaran lain.
4. *Outcomes*: hasil yang diperoleh pada studi primer. Hasil atau pencapaian yang dianalisis dalam penelitian ini adalah kemampuan berpikir tingkat tinggi berupa kemampuan berpikir kritis, kreatif, penalaran, dan pemecahan masalah siswa atau mahasiswa.
5. *Study designs*: jenis desain penelitian yang dipakai pada studi primer. Jenis penelitian yang dipilih untuk penelitian ini adalah penelitian kuantitatif dengan metode eksperimentasi dan kuasi-eksperimentasi. Secara khusus, studi primer menggunakan desain penelitian yang relatif ketat seperti *randomized control design* atau *controlled quasi-experimental design*.

Selain elemen PICOS tersebut, beberapa kriteria inklusi ditambahkan sebagai batasan dalam penelitian ini yaitu antara lain.

1. Studi primer berbentuk dokumen berupa artikel jurnal, artikel prosiding, skripsi, tesis, dan disertasi. Artikel ilmiah seperti skripsi, tesis, dan disertasi merupakan *grey literature* yang memberikan kontribusi penting dalam penelitian tinjauan sistematis seperti meta-analisis dengan menyajikan tinjauan yang komprehensif dan mengurangi bias publikasi untuk meningkatkan gambaran yang seimbang dari bukti-bukti yang tersedia (Paez, 2017). *Grey literatures* dan *conference proceedings* dalam penelitian tinjauan sistematis dan meta-analisis berguna untuk membantu meminimalisasi bias penelitian (Rethlefsen et al., 2021). Beberapa ahli menyatakan bahwa meta-analisis yang melibatkan *grey literature* mungkin memberikan data yang berkualitas rendah. Meski demikian, status publikasi tidak dapat menjadi jaminan kualitas data studi primer (Borenstein et al., 2009) dan karenanya peneliti tetap memasukkan data dari *grey literature*.

2. Studi primer memberikan informasi lengkap terkait data-data statistik seperti nilai rata-rata (*mean*), simpangan baku (*standard deviation*), ukuran sampel, *t-value* atau *p-value* pada kelas eksperimen yaitu kelas dengan model *flipped classroom* dan kelas kontrol yaitu kelas dengan model konvensional.
3. Studi primer yang diterbitkan melalui jurnal dan prosiding menerapkan proses *peer-reviewed* untuk meningkatkan kualitas data yang dipilih (Shi et al., 2020).

3.4 Teknik Pengumpulan Data

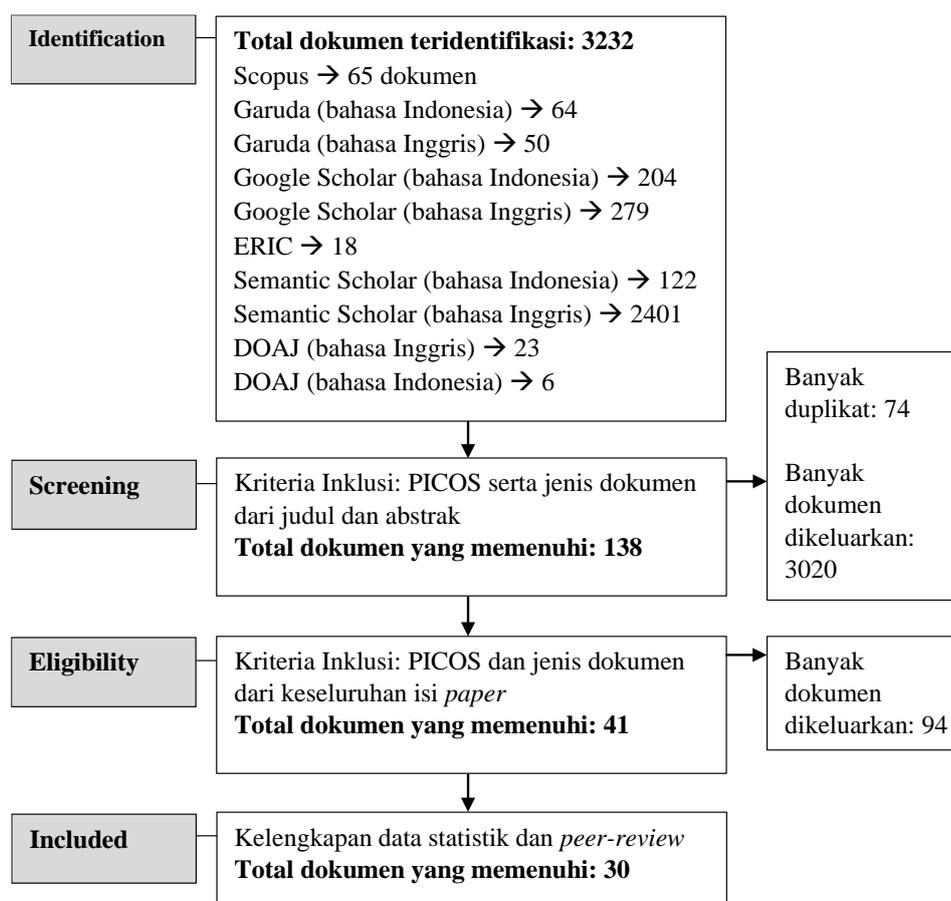
Tidak ada satupun database yang mampu memberikan daftar lengkap dan akurat dari semua penelitian yang memenuhi kriteria meta-analisis, sehingga dalam penelitian ini digunakan enam database yang umum dan sering digunakan di Indonesia. Penelusuran data dilakukan pada beberapa database yaitu Google Scholar (<https://scholar.google.com/>), Scopus (<https://www.scopus.com/>), Semantic Scholar (<https://www.semanticscholar.org/>), ERIC-Education Resources Information Center (<https://eric.ed.gov/>), DOAJ (<https://doaj.org/>) dan Portal Garuda (<https://garuda.kemdikbud.go.id/>). Kata kunci yang digunakan disajikan pada Tabel 3.1.

Tabel 3. 1 Tabel Kata Kunci untuk Setiap Database

Database	Kata kunci
Google Scholar (pencarian lanjutan – dalam judul artikel)	("flipped classroom" ATAU "flipping classroom" ATAU "inverting classroom" ATAU "inverted classroom") DAN ("berpikir tingkat tinggi" ATAU "berpikir kritis" ATAU "berpikir kreatif" ATAU "penalaran" ATAU "pemecahan masalah")
Scopus (advanced search - title)	("flipped classroom" OR "flipping classroom" OR "inverting classroom" OR "inverted classroom") AND ("higher order thinking" OR "critical thinking" OR "reasoning" OR "problem solving" OR "creative thinking")
Semantic Scholar	("flipped classroom" OR "flipping classroom" OR "inverting classroom" OR "inverted classroom") AND ("higher order thinking" OR "critical thinking" OR "reasoning" OR "problem solving" OR "creative thinking")
ERIC (title: "__")	("flipped classroom" OR "flipping classroom" OR "inverting classroom" OR "inverted classroom") AND ("higher order thinking" OR "critical thinking" OR "reasoning" OR "problem solving" OR "creative thinking")
DOAJ	("flipped classroom" ATAU "flipping classroom" ATAU "inverting classroom" ATAU "inverted classroom") DAN ("berpikir tingkat tinggi" ATAU "berpikir kritis" ATAU "berpikir kreatif" ATAU "penalaran" ATAU "pemecahan masalah")

Database	Kata kunci
Portal Garuda	("flipped classroom" ATAU "flipping classroom" ATAU "inverting classroom" ATAU "inverted classroom") DAN ("berpikir tingkat tinggi" ATAU "berpikir kritis" ATAU "berpikir kreatif" ATAU "penalaran" ATAU "pemecahan masalah")

Kata kunci tersebut dimasukkan ke *search bar* secara bersamaan atau bergantian menyesuaikan sistem dari setiap database karena masing-masing database memiliki sistem penelusuran yang berbeda-beda. Pengumpulan data dibantu dengan menggunakan protokol PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*) karena konsep umum PRISMA relevan dengan jenis tinjauan literatur seperti meta-analisis. PRISMA *statement* bertujuan untuk membantu peneliti dalam meningkatkan pelaporan penelitian meta-analisis. PRISMA *statements* terdiri dari 4 fase yaitu *identification*, *screening*, *eligibility*, dan *included*. Alur hasil penelusuran data disajikan pada Gambar 3.1.



Gambar 3. 1 Hasil Penelusuran dengan *PRISMA Statement*

1. Identifikasi (*Identification*)

Banyak dokumen diidentifikasi berdasarkan kata kunci dari keenam database. Khusus database Semantic Scholar, untuk mengurangi dokumen yang tidak relevan, tahun publikasi dibatasi dari tahun 2000 hingga kini. Hasil penelusuran dari setiap database dan kata kunci yang digunakan disajikan pada Lampiran 5. Total dokumen teridentifikasi pada fase ini sebanyak 3232.

2. Penyaringan (*Screening*)

Semua hasil penelusuran dibaca secara manual pada bagian judul dan abstrak. Kriteria inklusi PICOS diterapkan berdasarkan apa yang dirumuskan sebelumnya. Dokumen yang menimbulkan keraguan untuk dikeluarkan tetap dilanjutkan ke fase selanjutnya untuk dibaca secara menyeluruh. Jenis dokumen juga disesuaikan berdasarkan kebutuhan yaitu artikel jurnal dan prosiding, serta dokumen skripsi/tesis/disertasi. Duplikat dari dokumen-dokumen yang telah diidentifikasi dibuang. Hasil proses penyaringan selengkapnya disajikan pada Lampiran 6. Dokumen yang memenuhi adalah sebanyak 138.

3. Kelayakan (*Eligibility*)

Artikel-artikel yang memenuhi dibaca secara menyeluruh untuk memastikan kesesuaian isi dengan kriteria yang ditentukan, khususnya pada paper yang belum bisa ditentukan apakah memenuhi kriteria PICOS hanya berdasarkan judul dan abstrak. Hasil pada fase kelayakan data disajikan pada Lampiran 7. Artikel yang mengandung informasi yang kurang jelas, ditelusuri lebih lanjut dengan menghubungi *corresponding author* melalui email dan sosial media *Instagram*. Terdapat empat artikel yang menyebabkan keraguan dan telah dihubungi melalui email. Tiga dari empat artikel telah memberikan balasan dan melengkapi kekurangan data, sehingga ketiga artikel dimasukkan pada tahap *Included*. Dengan demikian, banyak dokumen yang memenuhi adalah 41 dokumen.

4. Disertakan (*Included*)

Beberapa informasi seperti data statistik diidentifikasi untuk kebutuhan meta-analisis. Selain itu, sumber studi primer diselidiki apakah menerapkan proses *peer-review* untuk meningkatkan kualitas data penelitian. Hasil fase ini dapat

dilihat di Lampiran 8. Artikel yang sudah memenuhi kriteria inklusi kecuali beberapa data yang tidak tersedia pada artikel, peneliti berusaha menghubungi penulis terkait untuk memperoleh data yang dibutuhkan. Dari enam belas artikel yang tidak lengkap dan telah dihubungi, dua di antaranya memberikan balasan. Bukti penelusuran melalui email disajikan pada Lampiran 9. Banyak dokumen yang memenuhi dan disertakan dalam penelitian adalah tiga puluh studi.

Semua artikel yang memenuhi kriteria inklusi selanjutnya dikumpulkan dan diberi kode (*coding*). Artikel yang telah memenuhi kemudian diekstrak dengan mencatat semua informasi statistik yang dibutuhkan untuk perhitungan ukuran efek. Pengkodean dilakukan dengan bantuan Microsoft Excel berdasarkan lembar pengkodean. Pengkodean dilakukan oleh dua pengkoding yang kemudian dilakukan uji reliabilitas antar pengkoding untuk menghasilkan studi primer yang valid dan reliabel.

Data statistik yang digunakan adalah data kemampuan berpikir kritis, berpikir kreatif, pemecahan masalah, dan penalaran matematis pada *post-test* untuk kelas eksperimen dan kelas kontrol yang berupa (1) nilai-rata-rata, standar deviasi, dan ukuran sampel, (2) nilai *t-value* dan ukuran sampel, dan (3) nilai *p-value* dan ukuran sampel. Selain itu, untuk melakukan identifikasi artikel, dibutuhkan informasi lain seperti nama penulis, email penulis, tahun studi, tahun publikasi, jenis publikasi, indeks publikasi, nama jurnal atau prosiding atau penerbit, database, link unduh studi primer, jenjang pendidikan, indikator berpikir tingkat tinggi, dan jenis media pembelajaran yang digunakan.

3.5 Instrumen

Coding (pengkodean) adalah bagian penting untuk mengurangi informasi yang kompleks dan berantakan. Pengkodean dalam penelitian ini menggunakan instrumen pengkodean yang terdiri dari lembar koding dan protokol skema koding. Lembar koding atau pengkodean digunakan untuk mengubah data atau informasi dari setiap studi primer menjadi data numerik atau data kategorikal (Dagyar & Demirel, 2015). Lembar koding sangat membantu pengkode (coder) dalam

mengekstraksi data atau informasi dari setiap studi primer. Informasi ini tidak hanya digunakan untuk tinjauan internal saja, tetapi juga membantu menjawab tentang mengapa studi tertentu tidak dimasukkan dalam sintesis.

Prosedur pengkodean (coding) untuk meta-analisis ini menggunakan panduan protokol pengkodean untuk menentukan informasi yang akan diekstraksi dari setiap studi dan memenuhi syarat. Peneliti membuat protokol skema koding, lalu pengkoding membaca protokol skema koding yang telah dibuat oleh peneliti kemudian mengisi formulir yang peneliti berikan dengan tepat sesuai penelitian. Selanjutnya, hasil studi direpresentasikan dalam bentuk nilai ukuran efek.

Lembar skema koding divalidasi dan dievaluasi oleh dua ahli dalam bidang meta-analisis, yaitu Devita Anjarwati, M.Pd. dan Suparman, M.Pd. Validasi yang dilakukan merupakan validasi isi dengan menggunakan skala likert (sangat tidak baik, tidak baik, cukup, baik, dan sangat baik) untuk mengukur kesesuaian lembar skema koding dalam merekam data pada studi primer. Pengukuran yang digunakan untuk keputusan hasil validasi adalah nilai median semua item dalam aspek validasi karena data berbentuk skala ordinal. Hasil validasi oleh *expert judgements* ditampilkan pada Tabel 3.2.

Tabel 3. 2 Hasil Validasi Ahli

No	Aspek	Item	Validator 1	Validator 2
1	Bahasa	Setiap item protokol skema koding dideskripsikan dengan kalimat yang sederhana, jelas dan mudah dipahami oleh pengkoding.	4	4
		Setiap item protokol skema koding dideskripsikan dengan bahasa yang sesuai dengan kaidah penulisan Bahasa Indonesia.	5	5
2	Isi/ Konten	Protokol skema koding terdiri dari item dan deskripsi.	5	4
		Setiap item dalam protokol skema koding dideskripsikan masing-masing yang disertai dengan ilustrasi atau contoh.	5	5
Median			5	

Berdasarkan Tabel 3.2, nilai median untuk keseluruhan item adalah 5 yang artinya secara keseluruhan lembar protokol koding dinilai sangat baik (dapat digunakan tanpa revisi). Dua item yang memiliki nilai 4 yaitu dinilai baik (dapat digunakan dengan revisi sedikit), dievaluasi kembali dengan mempertimbangkan saran revisi dari validator untuk memperoleh instrumen yang benar-benar sesuai. Saran dan revisi secara lengkap disajikan pada Lampiran 4, sedangkan protokol lembar koding yang telah direvisi terlampir di Lampiran 1.

3.6 Teknik Analisis Data

3.6.1 Uji Reliabilitas

Reliabilitas koding dalam penelitian meta-analisis dinilai dari *intrarater reliability* dan *interrater reliability* (Cooper et al., 2019). Reliabilitas *intrarater* adalah kekonsistenan coder dari satu kesempatan ke kesempatan lain, sedangkan reliabilitas *interrater* adalah kekonsistenan koding antara pengkode yang berbeda. Jika reliabilitas *intrarater* bisa dilakukan dengan pengkode yang melakukan koding ulang, reliabilitas *interrater* dilakukan dengan membanding hasil koding dari setidaknya dua pengkode.

Uji reliabilitas dalam penelitian ini utamanya dilakukan untuk melihat konsistensi antar pengkode dalam menentukan studi primer yang akan digunakan dalam meta-analisis yaitu *interrater reliability*. Uji reliabilitas yang digunakan adalah Cohen's Kappa karena merupakan statistik yang kuat untuk menguji tingkat kesepakatan (Juandi & Tamur, 2020) dan memberikan hasil yang informatif (Cooper et al., 2019). Parameter k adalah proporsi kemungkinan peningkatan terbaik atas peluang yang sebenarnya diperoleh oleh penilai.

$$k = \frac{\text{Pr}(a) - \text{Pr}(e)}{1 - \text{Pr}(e)} \quad (3.1)$$

Keterangan:

k : koefisien Cohen's Kappa

$\text{Pr}(a)$: kesepakatan yang benar-benar diamati (*actual observed agreement*)

$\text{Pr}(e)$: kesepakatan kebetulan (*chance agreement*)

Perhitungan Cohen's Kappa dibantu dengan perangkat lunak SPSS. Interpretasi nilai Cohen's Kappa yang diperoleh disajikan pada Tabel 3.3 (McHugh, 2012; Sim & Wright, 2005).

Tabel 3. 3 Klasifikasi Reliabilitas Kappa

Nilai	Interpretasi
$k \leq 0$	Tidak ada kesepakatan
$0 < k \leq 0,20$	Kesepakatan sedikit
$0,20 < k \leq 0,40$	Kesepakatan cukup
$0,40 < k \leq 0,60$	Kesepakatan sedang
$0,60 < k \leq 0,80$	Kesepakatan kuat
$0,80 < k \leq 1$	Kesepakatan sempurna

Komponen yang diekstrak dari setiap studi primer meliputi kode; indikator berpikir tingkat tinggi; sitasi; data statistik yang berupa rata-rata, simpangan baku, *p-value*, *t-value*, dan ukuran sampel baik kelas *flipped classroom* (FC) maupun kelas pembelajaran konvensional (PK); kapasitas kelas eksperimen; jenjang pendidikan; jenis media; tempat penelitian; tipe dokumen; tahun publikasi; tahun studi; pengindeks; nama jurnal/prosiding/institusi; database; email; dan link penelusuran.

Lembar koding, lembar protokol koding, file PDF studi primer, dan beberapa file pendukung lain dikirimkan kepada dua pengkoding yang ahli di bidang meta-analisis atau sudah melakukan penelitian meta-analisis. Dua pengkoding yang terlibat adalah Chelsi Ariati, M.Pd. dan Khairunnisa, M.Pd. Ketidaksetujuan antara pengkoding diselesaikan melalui diskusi dan kesepakatan bersama (Sánchez-Meca et al., 2021). Hasil pengukuran reliabilitas data untuk masing-masing item yang dibantu dengan *software* SPSS versi 26 disajikan pada Tabel 3.4.

Tabel 3. 4 Hasil Reliabilitas Ekstraksi Data Studi Primer

No	Item	Nilai Reliabilitas	Interpretasi
1	Kode	1,00	Kesepakatan Sempurna
2	Indikator HOT	1,00	Kesepakatan Sempurna
3	Sitasi	1,00	Kesepakatan Sempurna
4	Rata-rata kelas FC	1,00	Kesepakatan Sempurna
5	Simpangan baku kelas FC	1,00	Kesepakatan Sempurna
6	Ukuran kelas FC	0,93	Kesepakatan Sempurna
7	Rata-rata kelas PK	1,00	Kesepakatan Sempurna

No	Item	Nilai Reliabilitas	Interpretasi
8	Simpangan baku kelas PK	1,00	Kesepakatan Sempurna
9	Ukuran kelas PK	0,93	Kesepakatan Sempurna
10	t-value	0,80	Kesepakatan Kuat
11	P-value	1,00	Kesepakatan Sempurna
12	Ukuran kelas FC	0,93	Kesepakatan Sempurna
13	Jenjang pendidikan	1,00	Kesepakatan Sempurna
14	Jenis media pembelajaran	1,00	Kesepakatan Sempurna
15	Tempat penelitian	1,00	Kesepakatan Sempurna
16	Tipe dokumen	1,00	Kesepakatan Sempurna
17	Tahun publikasi	1,00	Kesepakatan Sempurna
18	Tahun studi/penelitian	0,91	Kesepakatan Sempurna
19	Pengindeks	0,97	Kesepakatan Sempurna
20	Nama jurnal/prosiding/institusi	1,00	Kesepakatan Sempurna
21	Database	1,00	Kesepakatan Sempurna
22	Email	1,00	Kesepakatan Sempurna
23	Link penelusuran	1,00	Kesepakatan Sempurna

Berdasarkan Tabel 3.4, nilai reliabilitas untuk setiap item memiliki tingkat persetujuan kuat dan sempurna (McHugh, 2012; Sim & Wright, 2005). Dengan demikian, data yang diekstrak oleh peneliti dapat digunakan menyesuaikan hasil pengkodean 1 dan pengkodean 2 dengan melakukan peninjauan kembali (Güler et al., 2022). Data ekstraksi secara lengkap dapat dilihat pada Lampiran 10.

3.6.2 Ukuran efek (*Effect size*)

Ukuran efek (ES) adalah alat penting dalam penelitian meta-analisis. Ukuran efek menunjukkan besarnya pengaruh yang ditimbulkan oleh variabel yang diuji dalam pengujian hipotesis yaitu pengaruh model *flipped classroom* terhadap kemampuan berpikir tingkat tinggi matematis. Perhitungan ES menggunakan indikator perbedaan rata-rata terstandarisasi karena skala nilai pada studi primer tidak sama. Rumus *Hedges's g* digunakan dalam penelitian ini karena sampel yang diperoleh relatif kecil.

Sebagaimana yang dijelaskan pada Bab II, setelah menghitung ukuran efek dan berat masing-masing studi, ukuran efek gabungan dihitung dengan cara mengalikan ukuran efek dan berat masing-masing studi kemudian dibagi jumlah berat keseluruhan (Borenstein et al., 2009). Rumus perhitungan ukuran efek

Hedges's g jika diketahui rata-rata dan simpangan baku kelas eksperimen dan kelas kontrol adalah sebagai berikut.

$$g = \left(\frac{\bar{X}_1 - \bar{X}_2}{S_{pooled}} \right) \times \left(1 - \frac{3}{4df-1} \right), \text{ dengan } df = n_1 + n_2 - 2 \quad (3.2)$$

Apabila data statistik yang diketahui adalah nilai uji t , maka rumus perhitungan ukuran efek *Hedges's g* adalah sebagai berikut.

$$g = t \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad (3.3)$$

Rata-rata ukuran efek semua studi yang terlibat dihitung sebagai ukuran efek gabungan (*pooled*) dengan rumus perhitungan sebagai berikut.

$$Pooled = \frac{\sum g_i \times W_i}{\sum W_i} \quad (3.4)$$

Keterangan:

\bar{X}_1	: rata-rata kelompok eksperimen
\bar{X}_2	: rata-rata kelompok kontrol
S_{pooled}	: standar deviasi gabungan
n_1	: jumlah sampel kelompok eksperimen
S_1	: standar deviasi kelompok eksperimen
n_2	: jumlah sampel kelompok kontrol
S_2	: standar deviasi kelompok kontrol
t	: nilai uji statistik t
W	: Bobot studi
i	: banyak ukuran efek

Ukuran efek masing-masing studi dan ukuran efek gabungan diperoleh dengan bantuan *software Comprehensive Meta Analysis (CMA)* versi 4. Program ini dikembangkan khusus untuk perhitungan meta-analisis yang mencakup fungsi perhitungan ukuran efek secara otomatis, melakukan meta-analisis dasar dan lanjutan, serta membuat grafik kualitas publikasi (Juandi & Tamur, 2020). Ukuran efek baik masing-masing studi dan gabungan yang diperoleh kemudian

diinterpretasikan menggunakan klasifikasi (Juandi & Tamur, 2020; Thalheimer & Cook, 2002) yang disajikan pada Tabel 3.4.

Tabel 3. 5 Klasifikasi Ukuran Efek

Ukuran efek (ES)	Interpretasi/Kategori
$0 \leq ES < 0,20$	Efek diabaikan
$0,20 \leq ES < 0,50$	Efek rendah
$0,50 \leq ES < 0,80$	Efek sedang
$0,80 \leq ES < 1,30$	Efek tinggi
$ES \geq 1,30$	Efek sangat tinggi

Uji hipotesis dilakukan untuk menginvestigasi signifikansi pengaruh model *flipped classroom* dibandingkan dengan pembelajaran konvensional. Nilai *p-value* atau nilai *Z* digunakan untuk melihat signifikansi pengaruhnya. Jika $Z_{hitung} > Z_{tabel}$ dengan $p < 0,05$ maka hipotesis ditolak.

Ukuran efek *Hedges's g* memang memberikan gambaran besar pengaruh model *flipped classroom* terhadap kemampuan berpikir tingkat tinggi matematis, tetapi seberapa lebih besar pengaruhnya dibandingkan model konvensional dapat diketahui dengan ukuran *U*, yaitu persentase *nonoverlap* (Cohen, 1988). Jika asumsi kedua populasi yang dibandingkan normal dan homogen, maka memungkinkan untuk mendefinisikan ukuran *nonoverlap* (*U*) terkait ukuran efek yang diperoleh secara intuitif berkekuatan dan bermakna. Untuk memperoleh ukuran *U*, ukuran efek *Hedges's g* diubah ke ukuran efek *Cohen's d* terlebih dahulu dengan menggunakan rumus versi Borenstein et al. (2009) yaitu $g = d \times J = \frac{\bar{X}_1 - \bar{X}_2}{S_{pooled}} \times \left(1 - \frac{3}{4df-1}\right)$. Dengan demikian, ukuran efek *Cohen's d* diperoleh dengan rumus $d = \frac{g}{J} = \frac{g}{\left(1 - \frac{3}{4df-1}\right)}$. Ukuran efek *Cohen's d* hasil konversi tersebut kemudian diinterpretasikan dengan tabel ekuivalen *d* yang dilampirkan pada Lampiran 12.

3.6.3 Uji Bias Publikasi dan Sensitivitas

Meta-analisis dapat menghasilkan sintesis yang akurat secara matematis dari penelitian-penelitian yang dimasukkan dalam analisis. Namun, Borenstein et al. (2009) menjelaskan jika penelitian-penelitian yang dilibatkan merupakan sampel yang bias dari semua penelitian yang relevan, maka efek gabungan yang dihitung

akan mencerminkan bias ini. Padahal setiap penelitian tidak akan lepas dari bias, meskipun peneliti telah berusaha meminimalkan bias. Dalam penelitian meta-analisis terdapat ancaman bias publikasi yaitu peneliti lebih banyak melibatkan studi yang dipublikasikan dibanding yang tidak dipublikasikan. Dengan demikian penting untuk melakukan uji sensitivitas dan bias publikasi dari studi yang dilibatkan.

Permasalahan yang ditangani pada uji sensitivitas untuk meta-analisis serupa dengan permasalahan untuk studi primer, yaitu fokusnya adalah pada sejauh mana hasil yang diperoleh kuat atau tidak terhadap asumsi dan keputusan yang dibuat saat melakukan sintesis (Borenstein et al., 2009). Analisis sensitivitas digunakan untuk memastikan hasil yang diperoleh bersifat *robust* atau kuat dalam arti kesimpulan perhitungan tidak bergantung pada studi tertentu, misalnya adanya pengaruh dari *outlier*. Uji bias publikasi dilakukan untuk memeriksa apakah data yang digunakan memiliki kecenderungan bias dan bagaimana dampak yang dihasilkan bila memang terdapat bias. Analisis sensitivitas dalam penelitian ini adalah *one study removed*, diagram corong, dan uji *Rosenthal FSN*, sedangkan uji bias publikasi adalah uji *trim and fill*.

1. *One study removed*

Prosedur ini digunakan untuk memperkiraan hasil ukuran efek yang akan diperoleh apabila studi tertentu dikeluarkan dari analisis. Dengan demikian, peneliti dapat melihat bagaimana suatu studi dapat mengubah ukuran efek gabungan dan apakah kesimpulan dari ukuran efek yang diperoleh akan berbeda apabila studi tertentu dikeluarkan dari analisis.

2. *Funnel plot* (Diagram corong)

Diagram corong adalah diagram representasi untuk memeriksa potensi adanya bias publikasi di mana diperkirakan akan menyebabkan asimetri. Model pemikiran yang dibangun dari *Funnel plot* adalah sebagai berikut.

- a. Penelitian dengan sampel besar memiliki kemungkinan lebih besar pula untuk dipublikasikan, meskipun mendapatkan hasil yang tidak signifikan secara statistik. Penelitian-penelitian ini akan muncul sebagai bulatan-bulatan pada diagram corong di bagian atas dan memiliki kecenderungan mendekati garis *mean effect size* atau ukuran efek gabungan.

- b. Penelitian dengan sampel sedang, hanya efek menengah ke atas yang memiliki kemungkinan lebih besar untuk dipublikasikan, sehingga beberapa studi akan diperkirakan menghilang atau tersembunyi.
- c. Penelitian dengan sampel kecil hanya memiliki kemungkinan untuk dipublikasikan bila efek yang dihasilkan besar. Penelitian dengan sampel kecil akan muncul sebagai bulatan atau poin-poin di bagian bawah yang cenderung lebih sedikit dibanding bulatan-bulatan di bagian tengah dan atas karena penelitian ini memiliki kemungkinan kecil untuk dipublikasikan. Selain itu, bulatan-bulatan cenderung lebih banyak di bagian kanan karena penelitian yang memiliki ukuran efek besar lebih cenderung dipublikasikan dibanding penelitian dengan ukuran efek kecil.

Jika penyebaran *effect size* dari studi primer menunjukkan distribusi yang simetris, maka dikatakan tidak ada bias publikasi terdeteksi. Interpretasi dari diagram corong bersifat subjektif, sehingga tidak cukup digunakan sebagai satu-satunya alat untuk melihat bias publikasi.

3. Uji *Rosenthal Fail-safe N* (FSN)

Fail-safe N menghitung berapa banyak studi yang hilang yang perlu diambil dan dimasukkan dalam analisis agar nilai p menjadi tidak signifikan (Borenstein et al., 2009). Dalam uji ini, asumsi yang digunakan adalah efek gabungan pada studi yang hilang adalah nol. Jika studi tersembunyi yang dibutuhkan untuk membuat nilai p tidak signifikan hanya sedikit, terdapat kekhawatiran bahwa efek sesungguhnya memang nol. Namun, apabila dibutuhkan studi tersembunyi yang sangat banyak untuk membuat nilai p tidak signifikan, maka tidak perlu ada kekhawatiran berarti. Untuk menentukan jumlah studi tersembunyi yang dibutuhkan tergolong sedikit atau banyak, dihitung probabilitas bias publikasi menggunakan rumus $\frac{N}{(5k+10)}$, dengan N adalah hasil perhitungan dari program CMA yaitu banyak studi tersembunyi yang dibutuhkan untuk membuat nilai p tidak signifikan dan k adalah jumlah studi yang diikuti. Jika nilai *fail-safe N* yaitu $\frac{N}{(5k+10)} > 1$, maka semua studi yang terlihat tahan terhadap bias publikasi (Guler et al., 2021).

4. Uji *trim and fill*

Uji *trim and fill* didasarkan pada metode *Funnel Plot* di mana studi primer yang menyebabkan diagram tidak simetris akan dipangkas serta studi yang tersisa dan memenuhi bentuk simetris digunakan untuk menghitung pusat corong yang sesungguhnya.

3.6.4 Uji Heterogenitas dan Pemilihan Model Estimasi

Uji heterogenitas dalam penelitian ini dilakukan untuk melihat keberagaman yang terdapat pada setiap ukuran efek. Selain itu, uji ini juga digunakan untuk memperkuat pemilihan model estimasi yang digunakan dalam meta-analisis. Apabila yang diperoleh dari *software Comprehensive Meta Analysis* (CMA) untuk nilai statistik Q yaitu $p < 0,05$ maka homogenitas ditolak. Artinya ukuran efek antar studi dikatakan heterogen yaitu ukuran efek antar kelompok studi tidak mengukur parameter populasi yang sama (Borenstein et al., 2009). Selain nilai statistik Q , digunakan nilai I^2 untuk melihat proporsi variansi yang teramati yang mencerminkan perbedaan nyata dalam ukuran efek. Interpretasi nilai I^2 adalah nilai I^2 lebih dari 30% tergolong heterogenitas rendah, nilai I^2 lebih dari 50% tergolong sebagai heterogenitas sedang, nilai I^2 lebih dari 75% tergolong sebagai heterogenitas tinggi (Higgins & Thompson, 2002). Nilai statistik Q dan nilai I^2 tersebut mendukung asumsi peneliti bahwa jumlah heterogenitas lebih dari apa yang diharapkan dari *sampling error* (kesalahan pengambilan sampel) dan memberikan bukti bahwa model *random-effect* sesuai untuk digunakan.

3.6.5 Analisis Karakteristik Studi

Analisis karakteristik studi dilakukan dengan tujuan untuk menguji apakah ada perbedaan ukuran efek antar kelompok studi (Juandi & Tamur, 2020). Berdasarkan hipotesis yang dibuat, peneliti memprediksi adanya ukuran efek yang berbeda, sehingga analisis karakteristik studi perlu dilakukan untuk melihat lebih dalam terkait variabel-variabel yang menyebabkan heterogenitas. Dalam penelitian ini, karakteristik studi yang diprediksi akan menyebabkan heterogenitas ukuran efek dibagi menjadi dua yaitu karakteristik bawaan (*gifted variable*) dan karakteristik yang ditentukan (*chosen variable*). Karakteristik studi bawaan yang dimaksud adalah karakteristik yang melekat pada sampel dan pelaksanaan studi eksperimentasi serta tidak dapat diubah atau “dikendalikan” begitu saja menyesuaikan keinginan peneliti pada studi ekperimentasi, sedangkan karakteristik

yang ditentukan adalah karakteristik yang disesuaikan dengan keinginan dan kepentingan peneliti pada studi eksperimentasi.

1. Ukuran sampel

Karakteristik ukuran sampel dibagi menjadi ukuran sampel yang kurang dari sama dengan 32 siswa dan ukuran sampel yang lebih dari 32 siswa (Permendikbud No. 17 Tahun 2017; Permendikbud No. 22 Tahun 2016). Selain itu, pengelompokan ukuran sampel juga didasarkan pada teori statistik yang menyatakan syarat ukuran sampel kecil dan besar.

2. Jenjang pendidikan

Karakteristik jenjang pendidikan dalam penelitian ini dibagi berdasarkan Sistem Pendidikan Nasional dalam Undang-Undang Nomor 20 tahun 2003 yang mengatur kategori jenjang pendidikan yaitu Sekolah Dasar (SD), Madrasah Ibtidaiyah (MI) atau bentuk lain yang sederajat; Sekolah Menengah Pertama (SMP) dan Madrasah Tsanawiyah (MTs) atau bentuk lain yang sederajat; Sekolah Menengah Atas (SMA), Madrasah Aliyah (MA), Sekolah Menengah Kejuruan (SMK), Madrasah Aliyah Kejuruan (MAK), dan bentuk lain yang sederajat; serta perguruan tinggi yang meliputi pendidikan diploma, sarjana, magister, dan doktor yang diselenggarakan oleh perguruan tinggi.

3. Tahun studi

Karakteristik tahun studi dalam penelitian ini adalah tahun akademik ketika penelitian pada studi primer dilaksanakan, seperti tahun 2019/2020, 2020/2021, dan seterusnya.

4. Jenis media pembelajaran

Jenis media yang digunakan dalam pembelajaran yang menerapkan model *flipped classroom* bervariasi antara lain video; sosial media; dan penggunaan *platform* tertentu seperti *Google Classroom*, *Edmodo*, *Moodle*, atau sejenisnya.

5. Indikator berpikir tingkat tinggi

Berdasarkan pemaparan kajian literatur pada Bab 2, karakteristik indikator berpikir tingkat tinggi dalam penelitian ini dikategorikan menjadi empat jenis kemampuan yaitu kemampuan berpikir kritis, berpikir kreatif, pemecahan masalah, dan penalaran matematis.

3.7 Tahap Penelitian

Tahapan prosedur penelitian meta-analisis yang dilakukan antara lain adalah sebagai berikut. Tahapan dalam metode meta-analisis antara lain (1) merumuskan masalah penelitian, (2) penelusuran literatur, (3) pengkodean, (4) analisis statistik, dan (5) representasi dan interpretasi hasil (Cooper et al., 2019; Hunter & Schmidt, 2004; Juandi & Tamur, 2020).

1. Merumuskan masalah penelitian yaitu pengaruh pengaruh model *flipped classroom* terhadap kemampuan berpikir tingkat tinggi matematis.
2. Melakukan penelusuran literatur melalui database Google Scholar, Scopus, Portal Garuda, ERIC, DOAJ dan Semantic Scholar menggunakan kata kunci yang telah ditentukan sesuai dengan permasalahan penelitian. Pada tahap ini, peneliti menyortir artikel-artikel yang memenuhi kriteria.
3. Melakukan ekstraksi data dan pengkodean data pada semua artikel yang memenuhi.
4. Melakukan perhitungan ukuran efek untuk masing-masing studi, ukuran efek gabungan serta ukuran efeknya berdasarkan karakteristik studi bawaan dan yang ditentukan antara lain jenjang pendidikan, ukuran sampel, tahun studi, jenis media pembelajaran, dan indikator berpikir tingkat tinggi. Perhitungan ukuran efek menggunakan bantuan *software Comprehensive Meta-Analysis (CMA)*. Selain itu, dilakukan uji bias publikasi dengan melihat *one study removed*, diagram corong, uji *trim and fill*, dan uji *Rosenthal FSN*.
5. Melakukan interpretasi terhadap ukuran efek yang diperoleh secara keseluruhan.