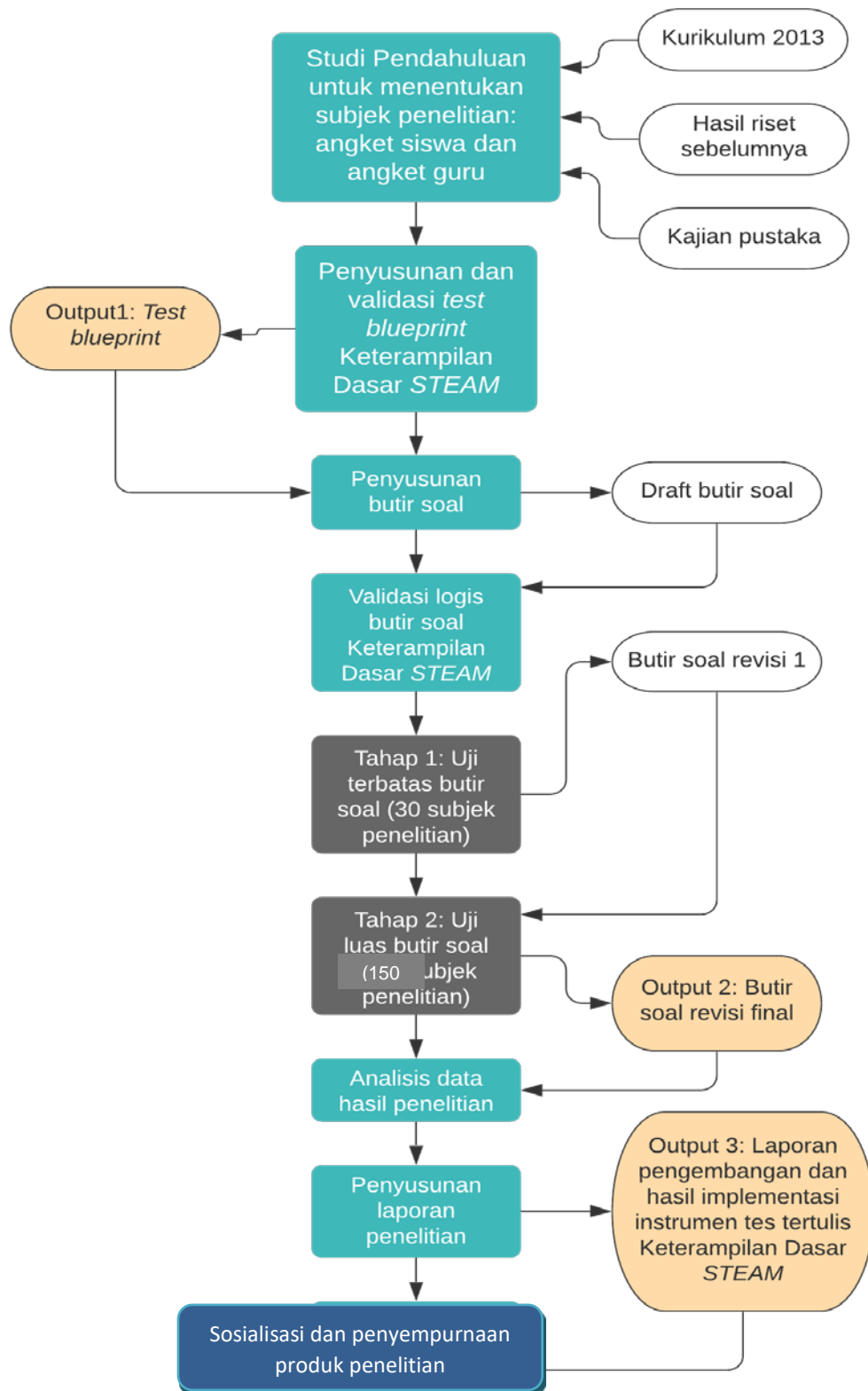


## **BAB III**

### **METODE DAN DESAIN PENELITIAN**

#### **3.1. Metode Penelitian**

Penelitian ini tidak melibatkan aktivitas pembelajaran berbasis *STEM* di dalam kelas, namun siswa akan diberikan tes tertulis yang harus dijawab untuk kemudian dianalisis keterampilan *STEM* yang dimiliki oleh siswa. Metode yang digunakan pada penelitian ini adalah Desain Penelitian Pengembangan yang mengacu pada Robert J. Beichner. Desain pengembangan terbatas adalah model pengembangan dimana bentuk penelitiannya berupa aktivitas merancang produk (instrumen tes tertulis) dan prosedur baru yang diuji di lapangan, dimodifikasi, dievaluasi, dilakukan penyesuaian, divalidasi dan disempurnakan (Beichner, Zavala, Tejada, & Barniol, 2017) . Instrumen tes harus memenuhi kriteria tertentu agar menghasilkan instrumen tes tertulis yang valid dan reliabel. Pengembangan instrumen tes tertulis untuk mengukur keterampilan *STEM* ini dilakukan sesuai dengan alur berikut ini.



**Gambar 3.1. Diagram alur langkah penelitian**

Berdasarkan Gambar 2 diagram alur langkah penelitian, secara umum terdapat lima tahap penelitian, yaitu:

- a. **Tahap 1.** Studi pendahuluan untuk menentukan subjek penelitian.
  - 1) Analisis kajian pustaka, hasil riset sebelumnya, dan kajian terhadap kurikulum 2013 penting untuk dilakukan. Kajian terhadap kurikulum 2013 meliputi analisis Standar Kompetensi Lulusan, Kompetensi Inti, dan Kompetensi Dasar mata pelajaran IPA di Sekolah Menengah Pertama (SMP).
  - 2) Pengembangan *framework* yang terintegrasi antara saintifik literasi berdasarkan *OECD* dan *framework STEM* sebagai acuan pengembangan instrumen tes tertulis untuk mengukur keterampilan *STEM*
- b. **Tahap 2.** Merancang *test blueprint* berdasarkan indikator soal pada topik IPA yang digunakan dan juga descriptor keterampilan *STEM*. Tahap ini merupakan kajian validitas logis terhadap *test blueprint* yang telah dikembangkan’.
- c. **Tahap 3.** Pengembangan butir soal tes tertulis untuk mengukur keterampilan *STEM* pada konsep IPA.
  - 1) Penyusunan butir soal tes berdasarkan tabel spesifikasi (*test blueprint*) soal yang telah di validasi sebelumnya. Butir soal tes tertulis yang disusun terdiri dari 3 paket. Setiap paket soal terdapat sebanyak 30 soal. Pertimbangan dibuatnya paket soal sebanyak 3 paket dalam pengembangan tes tertulis ini adalah agar terciptanya instrumen asesmen dengan cakupan konsep IPA yang beragam, reliabel dalam mengukur keterampilan *STEM* dan juga diharapkan dapat dijadikan model asesmen pada pembelajaran *STEM* lainnya.
  - 2) Validasi logis (*expert judgement*) dilakukan untuk memberi pertimbangan terhadap validitas konstruksi, validitas isi butir soal, dan kejelasan bahasa pada butir soal. Validasi ini melibatkan ahli pedagogi IPA dan ahli pengukuran. Ahli tersebut dipilih yang memiliki latarbelakang penguasaan epi*STEM*ologi IPA. Berdasarkan hasil validasi ahli inilah diperoleh butir soal revisi.

- 3) Telaah keterbacaan soal dilakukan dengan melibatkan 2 orang guru SMP. Berdasarkan hasil keterbacaan tersebut, soal kemudian diperbaiki dari aspek tata bahasa dan penyajiannya sebelum digunakan dalam ujicoba.
- d. **Tahap 4.** Implementasi butir soal pada Uji Terbatas dan Uji Luas untuk validasi akhir perangkat butir soal.
  - 1) Demi terbentuknya kualitas desain butir soal yang valid dan reliabel maka perlu dilakukan dua tahap uji: (a) Tahap uji terbatas yang dilakukan pada 30 siswa sebagai subjek penelitian untuk kemudian di analisis keterbacaan dan kualitas butir soal, lalu menjalani revisi tahap 1, (b) Tahap uji luas yang melibatkan 150 siswa sebagai subjek penelitian, uji luas ini merupakan uji lapangan sesungguhnya. Uji lapangan dimaksudkan untuk memperoleh produk akhir berupa tes tertulis yang sudah teruji validitas dan reliabilitasnya untuk mengukur keterampilan *STEM*. Berdasarkan hasil uji coba lapangan utama ini selanjutnya adalah analisis karakter butir soal yang telah dikembangkan. Uji lapangan juga menyajikan hasil implementasi butir soal dalam memaparkan keterampilan *STEM* yang dimiliki siswa SMP pada konsep IPA.
  - 2) Analisis data menggunakan analisis pemodelan *Rasch* dengan software Winstep, kemudian revisi tahap akhir sebagai penyempurnaan perangkat instrumen tes tertulis berdasarkan hasil validasi uji lapangan. Soal – soal dengan kualitas yang baik, valid dan reliabel dipilih berdasarkan hasil analisis data pada tahap ini.
- e. **Tahap 5.** Penyempurnaan dan perbaikan kaidah tata tulis pada instrumen tes tertulis yang telah dikembangkan.
  - a. **Tahap 6.** Rekonstruksi instrumen tes tertulis dan penyempurnaan berdasarkan rekomendasi dan *review umum expert judgement*.

### 3.2. Subjek Penelitian

Subjek penelitian yang terlibat pada penelitian ini adalah kelas IX dari setiap sekolah yang telah dipilih. Sampel dipilih dengan menggunakan cara *purposive sampling*. Menurut Arikunto (2006) *purposive sampling* adalah teknik mengambil sampel dengan tidak berdasarkan *random* (acak), daerah atau strata, melainkan berdasarkan atas adanya alasan pertimbangan yang berfokus pada tujuan penelitian tertentu. Adapun pertimbangan dari dipilihnya teknik *purposive sampling* ini adalah bahwa diperlukan sekolah yang memenuhi syarat sebagai berikut: (1) Telah atau pernah mengimplementasikan pembelajaran berbasis *STEM*, pembelajaran konstruktivisme atau pembelajaran dengan pendekatan saintifik pada konsep IPA, (2) Siswanya diharapkan memiliki keterampilan *STEM* yang disebutkan diatas, (3) Siswanya familiar dengan bentuk soal berbasis *STEM* yang memerlukan *higher order thinking skills (HOTS)* (4) Sekolah tersebut memiliki guru IPA dengan kualifikasi dan karakteristik yang kurang lebih sama dengan masa minimal mengajar selama 5 tahun. Keempat hal ini merupakan alasan yang krusial dipilihnya sekolah pada kluster yang baik agar dapat menghasilkan karakteristik butir soal yang representative dalam mengukur keterampilan *STEM*. Penelitian ini diimplementasikan pada tahap uji terbatas di SMP Al – Azhar Syifa Budi Parahyangan dengan 30 responden, kemudian pada tahap uji luas di 6 sekolah dengan kluster berbeda. Berikut ini adalah rincian sampel penelitian:

**Tabel 3.1. Rincian Sampel Penelitian**

No.	Nama Sekolah	Jumlah responden
1.	SMP Labschool UPI, Bandung	25
2.	SMP Daarul Hikam Integrated School, Bandung	25
3.	SMP Al Azhar Syifa Budi Parahyangan, Bandung	25
4.	SMP Istiqamah, Bandung	25
5.	SMPIT Nurul Imam, Bandung	25
6.	SMPIT Al-Azhar 44 Grand Wisata, Jakarta	25
<i>Total responden</i>		150 responden

### 3.3. Definisi Operasional

Definisi operasional dimaksudkan untuk menghindari kesalahan pemahaman dan perbedaan penafsiran yang berkaitan dengan istilah – istilah dalam judul penelitian tesis, diantaranya:

a. Instrumen Tes Tertulis

Instrumen tes tertulis yang dikembangkan dalam penelitian ini berupa pilihan ganda dengan empat opsi jawaban yang terdiri dari 3 paket soal dengan tema konsep IPA yang berbeda dengan 30 butir soal pada masing – masing paket soal. Instrumen tes tertulis ini telah melalui beberapa tahap pengembangan hingga akhirnya layak untuk diimplementasikan pada tahap uji terbatas (30 responden) dan juga tahap uji luas (150 responden). Instrumen tes tertulis ini dikembangkan untuk mengukur keterampilan STEM yang dimiliki oleh siswa SMP pada konsep IPA.

b. *Test blueprint*

*Test blueprint* disebut juga dengan *table specification* atau kisi – kisi soal yang disusun berdasarkan indikator soal dan deskriptor keterampilan STEM sebelum instrumen tes tersebut diimplementasikan pada tahap uji terbatas dan tahap uji luas. *Test blueprint* ini dapat memastikan validitas konten dan kesesuaian deskriptor penilaian yang harus dicapai oleh siswa (Sabbahi, 2020).

c. Keterampilan *STEM*

*STEM (Science, Technology, Engineering, and Mathematics)* merupakan sebuah pendekatan interdisiplin yang merupakan integrasi dari empat bidang ilmu yaitu ilmu sains, ilmu teknologi, ilmu teknik dan ilmu matematika. Keterampilan STEM yang diukur meliputi: keterampilan menganalisis fenomena saintifik, keterampilan memecahkan masalah, keterampilan matematis, keterampilan *engineering design* dan menggunakan teknologi, keterampilan berpikir kreatif dan keterampilan mengorganisasi data dan bukti saintifik. Keenam keterampilan STEM ini dijabarkan kedalam beberapa indikator dan deskriptor yang masing – masing diwakili oleh satu butir soal.

d. Konsep IPA

Konsep IPA yang dipilih adalah Energi dalam Sistem Kehidupan, Tekanan Zat Cair, Padat dan Gas dan Tekanan pada Tumbuhan, dan juga Gunung Api dan Gempa Bumi.

### 3.4. Instrumen Penelitian

Penelitian pengembangan instrumen tes tertulis untuk mengukur keterampilan *STEM* dibagi dalam beberapa tahap, diantaranya adalah: Tahap 1. Studi pendahuluan untuk menentukan subjek penelitian; Tahap 2. Merancang tes *blueprint* dari instrumen tes tertulis untuk mengukur keterampilan *STEM* tersebut dan dilakukannya validasi tes *blueprint* oleh ahli; Tahap 3. Pengembangan butir soal tes tertulis untuk mengukur keterampilan *STEM* pada konsep IPA; dan Tahap 4. Implementasi Uji Lapangan (uji terbatas dan uji luas) untuk validasi akhir perangkat butir soal. Adapun untuk instrumen penelitian dibutuhkan lebih lanjut dijelaskan sebagai berikut.

a. Instrumen angket guru dan angket siswa

Peneliti telah menentukan target subjek penelitian berdasarkan informasi yang didapat dari kajian bahwa sekolah – sekolah ini telah dan pernah mengimplementasikan pembelajaran berbasis *STEM* dan atau pendekatan saintifik lainnya. Maka, instrumen angket guru dan angket siswa disusun sebagai studi pendahuluan untuk memastikan bahwa target subjek penelitian tersebut benar adanya. Hal ini bertujuan agar subjek penelitian menjadi tepat sasaran dan relevan dalam mengukur keterampilan *STEM* siswa. Angket guru berisi pertanyaan ya/tidak dan pernyataan *likert scale* mengenai pengimplementasian pembelajaran berbasis *STEM* dan atau pendekatan saintifik lainnya yang diberikan kepada guru. Sedangkan angket siswa diberikan kepada siswa untuk mengkonfirmasi pengalaman siswa mengenai pembelajaran berbasis *STEM* dan atau pendekatan saintifik.

b. Lembar Validasi Instrumen untuk tes *Blueprint*

Pengumpulan data menggunakan lembar validasi instrumen pada instrumen tes tertulis digunakan untuk melihat kesesuaian butir-butir soal dengan tujuan pembelajaran yang akan dinilai. Cara menilai validasi suatu alat ukur adalah

dengan mengundang judgment kelompok ahli dalam bidang yang diukur. Kelompok ahli yang menjadi validator instrumen penilaian otentik yang dikembangkan terdiri dari dua dosen yang ahli dibidangnya, satu guru IPA SMA yang pernah mengajar di SMP dan satu guru IPA jenjang SMP. Format lembar validasi instrumen yang disusun berisi daftar checklist kesesuaian, dengan kriteria nilai setuju dan tidak setuju. Selain itu, diberikan juga kolom catatan perbaikan yang bertujuan sebagai saran dan masukan untuk perbaikan instrumen yang dikembangkan.

c. Instrumen Tes Tertulis (*test blueprint* dan Butir Soal)

Instrumen tes tertulis digunakan untuk menilai pengetahuan siswa berdasarkan dimensi pengetahuan dan jenjang kognitif terhadap penilaian keterampilan *STEM*. Dimensi pengetahuan pada instrumen tes tertulis yang dikembangkan meliputi pengetahuan faktual, konseptual, dan prosedural mengenai konsep IPA. Instrumen tes tertulis ini disusun secara mandiri oleh penulis, sedangkan konten soal yang berisi gambar atau data tabel yang berkaitan dengan konsep IPA tertentu didapat berdasarkan data sekunder dari website yang valid atau video pada *scientific channel* pada *platform Youtube*. Beberapa butir soal yang mengandung pembahasan mengenai *scientific project*, kerja ilmiah atau *STEM based project* isi dari instrumen tes tertulis ini sedikit banyak pernah dialami secara langsung oleh penulis sehingga diharapkan penulis dapat membayangkan keterampilan – keterampilan *STEM* inilah yang sesungguhnya diukur pada siswa.

*Test blueprint* didefinisikan juga sebagai kisi – kisi soal atau rencana yang dibuat dan digunakan saat mengembangkan pengujian. *Test blueprint* adalah dokumen yang mencerminkan isi penilaian yang akan diberikan kepada responden *Test blueprint* berisi tujuan instruksional, keterampilan dan deskriptor *STEM*, indikator soal serta butir soal yang sesuai. Bentuk tes tertulis yang dikembangkan berupa pilihan ganda dan uraian terbatas. Tujuan pengembangan bentuk uraian terbatas adalah untuk menilai pengetahuan siswa pada jenjang kognitif yang lebih tinggi, yang sulit dinilai hanya oleh pilihan ganda. Tes



tertulis ini terdiri dari sembilan keterampilan *STEM* yang dikembangkan. Tes tertulis tersebut terdiri dari 30 soal pilihan ganda. Soal-soal tersebut sudah divalidasi dan diujicoba, sehingga instrumen tes yang disajikan bersifat valid dan reliabel. Selanjutnya soal tersebut digunakan pada tahap aplikasi instrumen untuk mengukur keterampilan *STEM* siswa.

### 3.5. Analisis Data

Data hasil judgement ahli dan hasil telaah keterbacaan soal oleh guru dianalisis secara kualitatif dan kuantitatif. Pengujian empirik butir soal dilakukan dengan analisis menggunakan aplikasi pemodelan *RASCH* yang baik digunakan pada *assessment* pendidikan. Pemodelan *RASCH* dapat mentransformasi data, dari data skor mentah menjadi data interval yang sama. Proses transformasi data ini mampu menghasilkan skala pengukuran yang linear, presisi dan mempunyai satuan (Sumintono, 2016). Selain itu, pemodelan *RASCH* juga bisa digunakan untuk menganalisis kualitas instrumen tes, menentukan tingkat abilitas siswa dan menentukan tingkat kesulitan test sampai pada mendeteksi miskonsepsi, atau bias pada instrumen tes. Analisis pemodelan *RASCH* ini menggunakan software Winstep dalam pelaksanaannya. Pengolahan data dan analisis angket skala likert menggunakan software Microsoft Excel.

### 3.6. Prosedur Penelitian

Tahap 2 (merancang tes *blueprint* dari instrumen tes tertulis untuk mengukur keterampilan *STEM* dan dilakukannya validasi tes *blueprint* oleh ahli) dari desain pengembangan mengenai pengembangan instrumen tes tertulis yang terdiri dari sembilan tahapan, yaitu: (1) menyusun spesifikasi tes, (2) menulis soal tes, (3) menelaah soal tes, (4) melakukan ujicoba tes, (5) menganalisis butir soal, (6) memperbaiki tes, (7) merakit tes, (8) melaksanakan tes, dan (9) menafsirkan hasil tes.

- a. Menyusun kajian pustaka terkait dengan instrumen tes tertulis untuk mengukur keterampilan *STEM* dan menyusun kajian kurikulum terkait dengan konsep IPA yang dipilih

b. Memodifikasi dan mengembangkan *framework* keterampilan *STEM*

c. Menyusun dan mengembangkan *test blueprint*.

Tahap penyusunan *test blueprint* ini merupakan tahap penyusunan kisi – kisi soal yang telah dikaji sebelumnya pada tahap kajian kurikulum yang berisi konsep IPA yang dipilih yaitu Energi dalam sistem Kehidupan, Tekanan Zat dan Tekanan pada Tumbuhan, dan Gunung Api dan Gempa Bumi.

d. Menyusun Spesifikasi Tes

Pada tahapan spesifikasi ini tujuannya adalah untuk mendeskripsikan seluruh karakteristik yang harus ada dari sebuah instrumen tes. Berikut adalah yang harus diperhatikan dalam proses spesifikatis tes: (1) menentukan tujuan tes, (2) menyesuaikan antara kisi-kisi tes yang telah di susun dengan kompetensi atau keterampilan *STEM* dan (3) menentukan bentuk tes.

e. Menulis Butir Soal

Pada tahapan ini, penulis menyusun soal tes berupa pertanyaan yang dibuat berdasarkan pada indikator dan silabus yang sudah ditentukan sebelumnya. Soal dibuat dalam bentuk tes objektif berupa soal pilihan ganda dan soal uraian terbatas. Selain itu dalam penyusunan soal juga mengacu pada kisi-kisi yang sudah relevan dengan keterampilan *STEM* yang telah dikemukakan sebelumnya. Artinya, setiap aspek dan indikator pada kisi-kisi yang sudah dibuat kemudian di implementasikan dalam bentuk soal.

f. Menelaah Butir Soal

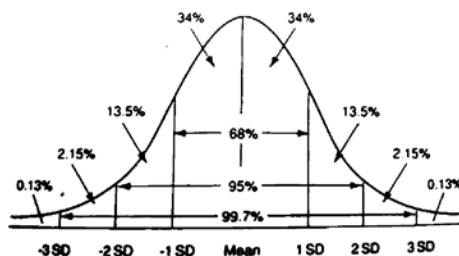
Tahap telaah butir soal dilakukan oleh dosen ahli sebagai pengujian validitas dan reliabilitas yang kemudian menjalani tahap revisi sebelum instrumen tes disebarkan pada responden di langkah uji terbatas.

g. Melakukan Tahap Uji Terbatas Tes

Langkah uji terbatas ini dilaksanakan pada 30 responden dari 1 sekolah yang familiar dengan pembelajaran berbasis inkuiri atau pembelajaran berbasis *STEM*.

h. Menganalisis Butir Soal

Pada tahap analisis butir soal ditunjukkan untuk memperoleh instrumen tes yang layak digunakan dalam penelitian. Soal yang sudah melewati tahap uji coba kemudian dilakukan analisis soal guna menentukan tingkat kesukaran dan daya beda soal. Pada tahap ini penulis bisa mengetahui tingkat kelemahan dan kekuatan instrumen tes yang telah di susun, sehingga penulis bisa menentukan butir soal yang bisa digunakan dan butir soal yang dinyatakan gugur (*rejected*). Software yang digunakan pada tahap analisis butir soal menggunakan *software* Winstep dengan Pemodelan *Rasch* sebagai output dari analisis. Hasil pada langkah ini adalah tersedianya informasi perihal karakteristik soal, sehingga bisa membantu pada proses penyusunan perangkat soal. Pemaparan pengukuran model *Rasch* yang mengacu pada (Sumintono, 2016) dijelaskan seperti berikut ini.



**Gambar 3.2. Kurva normal**

(Sumber: (Sumintono, 2016))

### 1) Person – Item Wright Map

Bagian ini menyajikan peta konsep yang menjelaskan sebaran kompetensi responden dan sebaran tingkat kesukaran soal pada skala yang sama secara komprehensif. Peta konsep sebaran tersebut ditentukan dari nilai logit yang dapat bernilai positif (+) maupun negatif (-). Data sebaran logit person maupun logit butir bisa diterangkan melalui penjelasan melalui kurva normal biasa, dalam hal ini jarak antara nilai rata – rata di tengah (mean), dengan 1 deviasi standar (1SD) adalah 34% jumlah data. Jika didapati nilai logit di luar dari jarak +2SD sampai ke -2SD, kondisi tersebut bisa disebut dengan *outlier*.

### 2) Tingkat Kesulitan Butir Soal (*Item Measure*)

*Item measure* merupakan tabel yang menjelaskan secara rinci terkait informasi logit pada setiap butir soal. Data logit butir soal ini memiliki skala dengan jarak yang sama.

### 3) Tingkat Kesesuaian Butir Soal (*Item Fit*)

Pada tahap analisis *item fit* digunakan untuk mengetahui kemampuan butir soal dalam melakukan pengukuran, apakah berfungsi normal atau tidak.

Apabila ditemukan instrumen tes yang tidak fit, maka hal tersebut bisa menjadi justifikasi bahwa siswa mengalami miskonsepsi terkait materi pada instrumen soal tersebut. Kriteria pada contoh tabel *item fit* tiga kolom di bawah ini; *outfit MNSQ*, *outfit ZSTD* dan *Pt. Measure Correlation* (poin biserial) merupakan standar yang digunakan untuk menilai tingkat kesesuaian butir soal.

**Tabel 3.2. Contoh Tabel Misfit Order**

Item STATISTICS: MISFIT ORDER

ENTRY NUMBER	TOTAL SCORE	TOTAL COUNT	MEASURE	MODEL S.E.		INFIT		OUTFIT		PTMEASUR-AL		EXACT MATCH		Item
				MNSQ	ZSTD	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.	DBS%	EXP%	
12	27	30	-.96	.62	1.12	.41	2.13	1.47	A	-.02	.17	89.3	89.2	S12
11	27	30	-.96	.62	1.08	.32	2.04	1.40	B	.02	.17	89.3	89.2	S11
6	6	30	3.34	.57	1.22	.65	1.67	1.25	C	.43	.59	82.1	86.2	S6
14	25	30	-.34	.51	1.14	.54	1.23	.60	D	-.10	.22	82.1	82.0	S14
13	25	30	-.34	.51	1.03	.19	.93	.01	E	.21	.22	82.1	82.0	S13
1	28	30	-1.42	.74	1.00	.21	.81	.07	F	.15	.14	92.9	92.8	S1
5	29	30	-2.16	1.02	1.00	.31	.81	.27	G	.10	.10	96.4	96.4	S5
9	29	30	-2.16	1.02	1.00	.31	.81	.27	H	.10	.10	96.4	96.4	S9
10	15	30	1.47	.40	.98	-.09	.97	-.14	I	.42	.41	64.3	63.2	S10
7	27	30	-.96	.62	.94	.02	.82	-.05	J	.22	.17	89.3	89.2	S7
4	17	30	1.15	.40	.91	-.83	.85	-.83	K	.45	.38	64.3	62.5	S4
8	18	30	.99	.40	.89	-.97	.83	-.87	L	.45	.36	67.9	62.5	S8
3	8	30	2.78	.49	.87	-.37	.87	-.24	M	.61	.54	82.1	80.0	S3
15	25	30	-.34	.51	.86	-.40	.71	-.52	N	.34	.22	82.1	82.0	S15
2	24	30	-.10	.47	.85	-.52	.72	-.61	O	.37	.25	78.6	78.5	S2

Menurut (Boone *et al.*, 2014), kriteria yang dipakai dalam memeriksa kesesuaian butir soal yang tidak cocok (*outliers*) yaitu:

- Nilai *outfit MNSQ* (*means-square*) yang diterima:  $0.5 < MNSQ < 1.5$
- Nilai *outfit ZSTD* (*z-standard*) yang diterima:  $0.2 < ZSTD < +2.0$
- Nilai *Pt. Measure Correlation* (*poin biserial*) yang diterima:  $0.3 < Pt. Measure Correlation < 0.85$

Nilai *outfit MNSQ* (*means-square*), *outfit ZSTD* (*z-standard*) dan *Pt. Measure Correlation* (*poin biserial*) merupakan kriteria yang dipakai dalam melihat tingkat kesesuaian butir (*item fit*). Apabila butir soal yang disusun belum

memenuhi ketiga kriteria yang telah ditetapkan, maka bisa disimpulkan bahwa butir soal tersebut harus diperbaiki (*revised*) atau ditolak/diganti (*rejected*)

#### 4) Deteksi Adanya Butir Soal yang Bias

Salah satu indikasi dari suatu pengukuran yang valid adalah ditunjukkan dengan instrumen soal yang digunakan dalam penelitian tidak mengandung unsur bias. Suatu instrumen tes dikategorikan sebagai instrumen yang bias jika ditemukan karakteristik individu dari responden tertentu memperoleh keuntungan dibandingkan dengan karakteristik individu yang lain. Untuk mengetahui tingkat bias dalam instrumen tes, dibutuhkan analisis deteksi DIF (*differential item functioning*) dalam pemodelan *RASCH*. Deteksi tersebut menunjukkan tingkat keberfungsian butir diferensial. Suatu butir soal bisa dijustifikasi bias jika memiliki nilai probabilitas lebih kecil dari 5% atau 0.05) seperti contoh tabel berikut ini.

**Tabel 3.3. Contoh Tabel Deteksi Butir Soal Bias**

Person CLASSES	SUMMARY DIF		PROB.	BETWEEN-CLASS/GROUP			Item Number Name
	CHI-SQUARED	D.F.		UNWTD	MNSQ	ZSTD	
2	.1344	1	.7139	.1368	-.56	1	S1
2	.9098	1	.3402	.9336	.42	2	S2
2	1.7195	1	.1898	1.7843	.92	3	S3
2	.1478	1	.7007	.1507	-.52	4	S4
2	.5345	1	.4647	.5476	.09	5	S5
2	1.4037	1	.2361	1.4506	.75	6	S6
2	3.7014	1	.0544	3.9501	1.70	7	S7
2	1.2123	1	.2709	1.2495	.63	8	S8
2	.1587	1	.6903	.1606	-.50	9	S9
2	.2061	1	.6499	.2104	-.39	10	S10
2	.0320	1	.8580	.0325	-.97	11	S11
2	.4116	1	.5212	.4182	-.06	12	S12
2	2.0941	1	.1479	2.2483	1.13	13	S13
2	.0800	1	.7773	.0836	-.72	14	S14
2	5.4361	1	.0197	5.9815	2.20	15	S15
2	.4212	1	.5164	.4304	-.05	16	S16
2	3.7513	1	.0528	4.0148	1.72	17	S17
2	.0270	1	.8695	.0349	-.96	18	S18
2	.0047	1	.9451	.0061	-1.26	19	S19
2	.0062	1	.9371	.0081	-1.22	20	S20
2	.8371	1	.3602	.8595	.37	21	S21
2	.0800	1	.7773	.0836	-.72	22	S22
2	5.7542	1	.0164	6.6164	2.33	23	S23
2	.4180	1	.5179	.4260	-.05	24	S24
2	.0122	1	.9119	.0158	-1.12	25	S25
2	4.8318	1	.0279	5.1663	2.02	26	S26
2	.1574	1	.6916	.1599	-.50	27	S27
2	.0343	1	.8530	.0357	-.95	28	S28

#### 5) Tingkat Abilitas Individu (*Person Measure*)

Selain mengukur tingkat kesulitan butir soal, pemodelan *Rasch* juga dapat mengukur tingkat abilitas individu dan tingkan kesesuaian individu.

#### 6) Reliabilitas soal pada Ringkasan statistik

Ringkasan statistik menyajikan informasi secara lengkap terkait mutu reaksi *feed back* siswa, mutu instrumen tes, serta hubungan antara person dan butir soal. Reliabilitas merupakan derajat ketetapan suatu instrumen tes dinyatakan bersifat konsisten, sehingga penenliti yakin instrumen tes atau butir soal yang disusun bisa dipercaya dijadikan sebagai instrumen penelitian. Penjelasan ringkasan statistik disajikan pada tabel 3.4.

**Tabel. 3.4. Contoh Tabel Ringkasan Statistik**

Person	150 INPUT		150 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	19.2	31.0	.88	.57	1.00	.0	1.00	.0
P.SD	6.4	.0	1.58	.40	.15	1.0	.30	1.0
REAL RMSE	.69	TRUE SD	1.42	SEPARATION	2.06	Person	RELIABILITY	.81

Item	31 INPUT		31 MEASURED		INFIT		OUTFIT	
	TOTAL	COUNT	MEASURE	REALSE	IMNSQ	ZSTD	OMNSQ	ZSTD
MEAN	93.0	150.0	.00	.20	1.00	.0	1.00	.0
P.SD	20.7	.0	.80	.02	.08	1.2	.16	1.1
REAL RMSE	.20	TRUE SD	.78	SEPARATION	3.82	Item	RELIABILITY	.94

CRONBACH ALPHA TEST RELIABILITY = 0.86

Keterangan:

- a) Nilai alpha Cronbach (mengukur reliabilitas, yaitu interaksi antara *person* dan butir – butir soal secara keseluruhan):

<0.5: Buruk                                      0.5 – 0.6 : Jelek                      0.6 – 0.7 : Cukup  
 0.7 – 0.8 : Bagus                                      > 0.8 Bagus sekali

- b) Nilai *person reliability* dan *item reliability*:

<0.67 : Lemah                                      0.67 – 0.80 : Cukup                      0.80 – 0.90 :  
 Bagus  
 0.91 – 0.94 : Bagus sekali                      >0.94 : Istimewa

- c) Pengelompokan *person* dan butir dapat diketahui dari nilai *separation*. Makin besar nilai *separation*, maka kualitas instrumen dalam hal keseluruhan respond dan butir soal juga makin bagus karena bisa mengidentifikasi kelompok responden dan kelompok butir. Analsis lain yang digunakan untuk mengetahui pengelompokkan yang lebih akurat adalah persamaan pemisahan strata, dengan rumus sebagai berikut:

$$H = \frac{[(4 \times SEPARATION) + 1]}{3}$$

7) *Anchor Linking Items* (Soal Penghubung Antar Paket)

Proses *equating* (penyetaraan) adalah hal yang penting dilakukan dalam pengolahan hasil uji instrumen tes tertulis agar diperoleh pemetaan mutu instrumen soal yang akurat dan valid, tanpa distorsi perbedaan tingkat kesulitan meskipun dikembangkan 3 paket soal yang berbeda (Herkusumo, 2011). Jumlah minimal *anchor items* yang baik adalah 20% dari keseluruhan butir soal.

i. Merevisi Tes

Instrumen soal hasil uji terbatas ini kemudian menjalani tahap revisi kembali berdasarkan hasil analisis pemodelan *Rasch* yang telah dilaksanakan sebelumnya.

j. Melaksanakan Tes pada Tahap Uji Luas

Tahap ini adalah tahap dimana instrumen tes hasil revisi dituangkan ke dalam platform *Google Form* agar mudah diakses oleh 150 responden dari 6 sekolah. *Output* dari responden yang mengerjakan adalah dalam bentuk *sheets* pada Microsoft Excel yang kemudian dituangkan menjadi formula yang dapat dibaca oleh *software* Winstep pada Pemodelan *Rasch*.

k. Menafsirkan Hasil Tes

Setelah tahap uji luas dilaksanakan, file Microsoft Excel yang berisi data kemudian kembali dimasukkan ke dalam *software* Winstep untuk kemudian dianalisis hasilnya menggunakan kriteria pengukuran yang telah dijelaskan di atas. Hasil analisis tahap akhir lalu kembali direvisi tahap akhir, sehingga menghasilkan instrumen tes untuk mengukur keterampilan *STEM* yang valid dan reliabel berdasarkan perhitungan Pemodelan *Rasch* juga menghasilkan karakteristik instrumen tes untuk mengukur keterampilan *STEM* yang berbeda dari yang lain.

l. Rekonstruksi instrumen tes tertulis dan penyempurnaan berdasarkan rekomendasi dan *review* umum *expert judgement*.