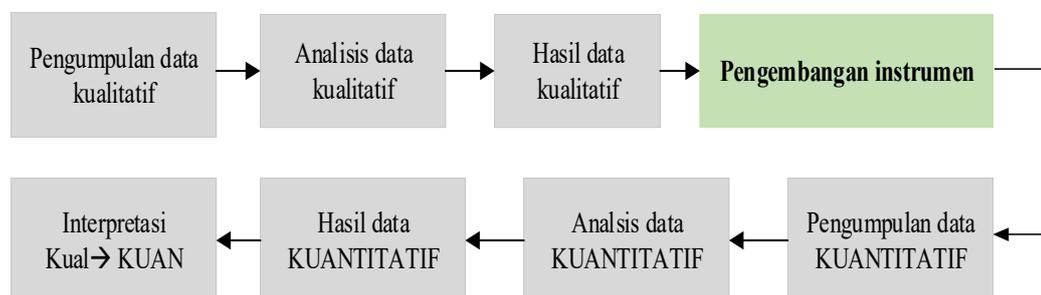


## BAB III METODE PENELITIAN

### A. Desain Penelitian

Penelitian ini merupakan penelitian *mixed methods* yang menekankan pada pengumpulan data kuantitatif dan kualitatif yang dilakukan secara simultan selama proses pengembangan. Dalam penelitian *mixed methods* terdapat empat jenis desain, yaitu: 1) *triangulation design*, 2) *embedded design*; 3) *explanatory design*; dan 4) *exploratory design*. Dalam penelitian ini digunakan *exploratory design*, dengan model *instrument development (QUAN emphasized)*. Model ini dilakukan dengan tiga tahap. Tahap pertama pengumpulan data secara kualitatif data hasil kualitatif digunakan sebagai dasar pengembangan item instrumen penelitian. Tahap kedua pengembangan instrumen. Tahap ketiga pengumpulan data secara kuantitatif, instrumen hasil pengembangan diimplementasikan dan diuji cobakan pada sampel sehingga diperoleh data kuantitatif (Cresswell & Clark, 2007). Tahapan pengembangan model *instrument development (QUAN emphasized)* dideskripsikan pada Gambar 3.1.



Gambar 3.1 Model Pengembangan *Instrument Development (QUAN Emphasized)*

Langkah-langkah dalam penelitian ini dijelaskan secara terperinci pada Gambar 3.2. Uraian tahapan-tahapan penelitian disajikan sebagai berikut:

#### 1. Pengumpulan Data Kualitatif

Tiga tahapan dalam pengumpulan data kualitatif, yaitu: a) pengumpulan data kualitatif, b) analisis data, dan hasil analisis data kualitatif. Penjelasan masing-masing tahapan dijelaskan lebih terperinci sebagai berikut:

##### a. Pengumpulan data kualitatif

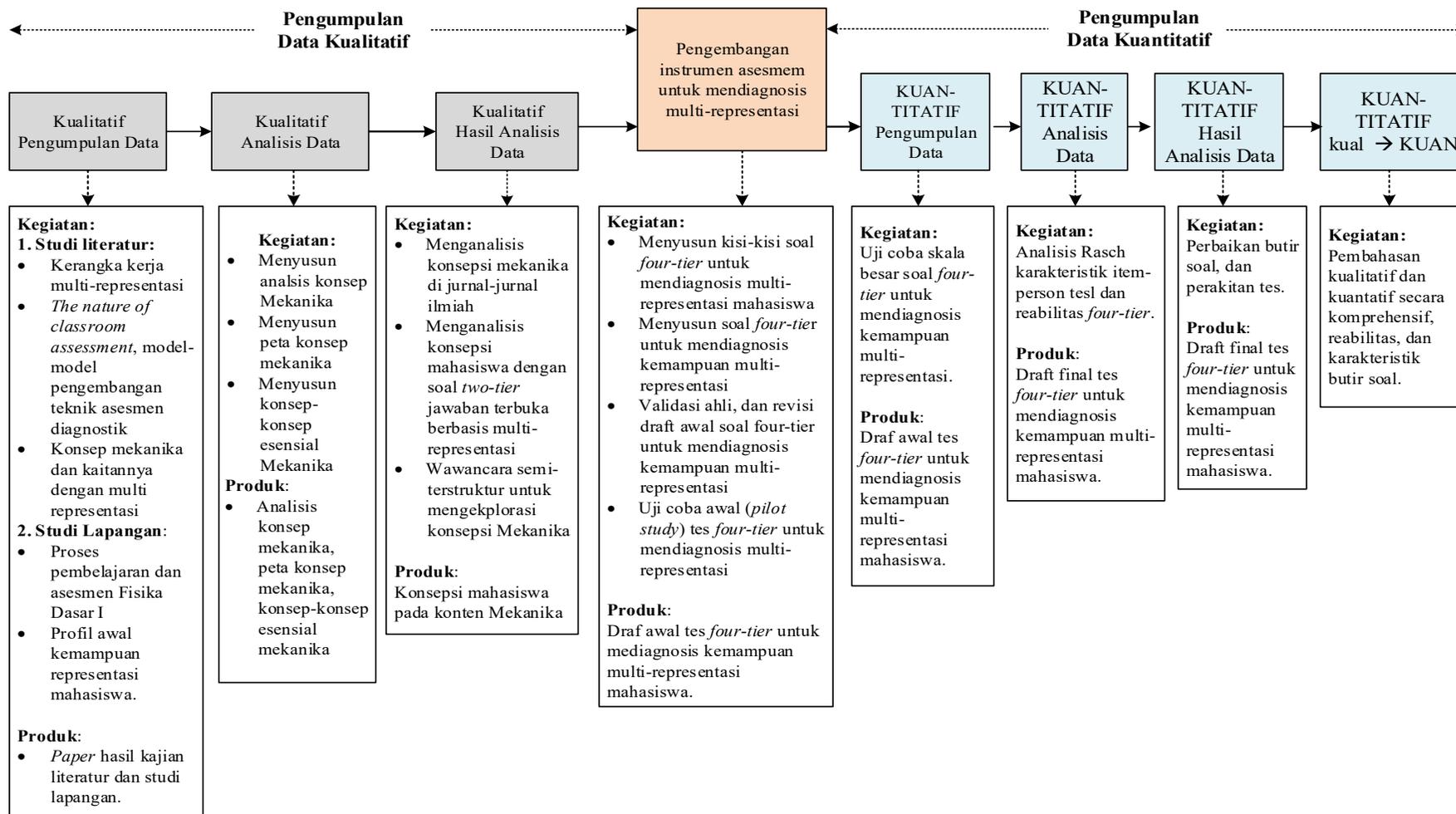
Tahap studi pendahuluan dilakukan untuk mencari kerangka kerja konseptual yang paling tepat dan memetakan penyelenggaraan perkuliahan dan

penerapan asesmen Fisika Dasar I di Jurusan Pendidikan Fisika. Tahapan ini fokus pada dua kegiatan yaitu studi literatur dan studi lapangan. Fokus kegiatan pertama adalah studi literatur mengkaji berbagai kerangka kerja konseptual multi-representasi, model-model asesmen pembelajaran, teknik asesmen diagnostik, tahapan pengembangan asesmen diagnostik, penggunaan asesmen diagnostik multi-representasi pada pendidikan fisika.

Pengembangan instrumen asesmen multi representasi merujuk pada beberapa kerangka kerja teoritik yang telah dikembangkan oleh para peneliti dalam menganalisis dan mengases kualitas multi-representasi. Kerangka kerja multi-representasi dibagi menjadi representasi eksternal dan internal (Cox, 1996; Goldin, 1998a, 2014a; Goldin & Karput, 1996; Goldin & Shteinglod, 2001; Schnotz & Lowe, 2003; Zhang & Norman, 1994; Zhang & Patel, 2006).

Representasi internal sering disebut representasi mental. Representasi internal dibangun dalam pikiran atau mental individu. Representasi eksternal merupakan wujud fisik yang dapat diamati seperti kata-kata, grafik, gambar, persamaan matematis (Goldin & Karput, 1996). Eksternal representasi pada berbagai jurnal dan tulisan ilmiah dikenal dengan istilah multi-representasi.

Kerangka kerja eksternal representasi dikembangkan dalam pendidikan sains (Prain & Waldrip, 2006; Wu & Puntambekar, 2012). Kerangka kerja multi-representasi dikembangkan lebih khusus untuk pendidikan kimia (Gilbert & Treagust, 2009). Lebih lanjut kerangka kerja multi-representasi dikembangkan untuk pendidikan fisika (Kohl & Finkelstein, 2006, 2008). Pengembangan instrumen asesmen diagnostik multi-representasi dalam penelitian ini mengadaptasi dari beberapa mode representasi dalam fisika yaitu: verbal, matematis, gambar, dan grafik (Klein dkk., 2017; Nieminen dkk., 2010; Van Heuvelen & Zou, 2001).



Gambar 3.2 Langkah-Langkah Penelitian

Teknik asesmen diterapkan dalam pembelajaran sains meliputi teknik tes dan non tes (Abruscato, 1996; Liu, 2010). Informasi tes berfungsi untuk akuntabilitas sekolah dan berpengaruh terhadap proses pembelajaran (Frederiksen, 1981). *Measurement-driven-instruction* adalah hasil pengukuran tes terkait erat dengan kinerja siswa dan proses pembelajaran (Madaus, 1988), proses pembelajaran mempersiapkan siswa untuk menghadapi tes dan hasil pengukuran tes sebagai umpan balik proses pembelajaran (Airasian, 1988; Popham, 1987). Keunggulan tes meliputi (Gipps, 1994): 1) interpretasi tes mempengaruhi individu, institusi, kurikulum, dan pembelajaran; 2) semakin banyak indikator kuantitatif sosial sebagai landasan pengambilan keputusan sosial, semakin kecil kemungkinan membelokkan pengambilan keputusan sosial; 3) Pengambilan keputusan penting terkait interpretasi tes, guru mengajar untuk persiapan tes.

Ahli-ahli asesmen menjelaskan tahapan-tahapan pengembangan tes dan manajemen pengelolaan tes (AERA dkk, 1999; Anderson & Morgan, 2008; Cohen & Swerdisk, 2010; Crocker & Algina, 2008; Haladyna, 2003; Haladyna & Rodriguez, 2013; Kusaeri & Suprananto, 2012; Mardapi, 2008, 2012; Tinkelman, 1971; Wiersma & Jurs, 1990). Beberapa ahli menjelaskan *framework* pengembangan tes diagnostik pilihan ganda dua-tingkat (Treagust, 1988, 1995), tes diagnostik tiga-tingkat (Arslan dkk., 2012; Caleon & Subramaniam, 2010a; Kirbulut & Geban, 2014; Romine dkk., 2015) dan tes diagnostik pilihan ganda empat tingkat (Caleon & Subramaniam, 2010b). Penelitian ini berfokus pada teknik tes diagnostik empat-tingkat.

Fokus kegiatan kedua adalah studi lapangan. Studi lapangan bertujuan mendapatkan informasi tentang: a) proses perkuliahan dan implementasi asesmen pada mata kuliah Fisika Dasar I, b) profil multi representasi awal mahasiswa. Fungsi tahap studi lapangan sebagai landasan rasionalitas empiris guna pengembangan instrumen tes asesmen diagnostik multi-representasi. Informasi mengenai pelaksanaan perkuliahan Fisika Dasar I diperoleh melalui wawancara dengan tim dosen, observasi langsung pelaksanaan perkuliahan, penyebaran angket, dan pemberian tes esai untuk mengetahui kemampuan awal multi-representasi mahasiswa.

b. Analisis dan hasil analisis data kualitatif

Fokus kegiatan pada tahap ini adalah pengembangan instrumen asesmen diagnostik-multi representasi. Tahapan-tahapan pengembangan instrumen tes diagnostik dimodifikasi dari Treagust (1995), antara lain: 1) mendefinisikan konten, 2) menganalisis konsepsi mahasiswa, 3) mengembangkan instrumen tes diagnostik.

1) Mendefinisikan konten

Tahapan mendefinisikan konten meliputi beberapa tahapan, yaitu: a) mengidentifikasi analisis konsep, dan b) mengembangkan peta konsep dan konsep esensial.

a) Mengidentifikasi analisis konsep

Pengembangan analisis konsep merupakan tahapan identifikasi dan analisis konten (materi subyek) (Hackling & Treagust, 1984). Analisis konten (materi subyek) sebagai landasan pengembangan pembelajaran, evaluasi siswa, dan perencanaan kurikulum (Stewart & van Kirk, 1981). Beberapa metode analisis konten antara lain: representasi struktur substantif (Finley & Stewart, 1982), teknik jaringan semantik (Stewart & van Kirk, 1981), jaringan struktur aktif (Lindsay & Norman, 1977), dan analisis konsep (Herron, Cantu, Ward, & Srinivasan, 1977). Karakteristik utama representasi struktur substantif, teknik jaringan semantik, dan jaringan struktur aktif adalah makna jaringan konsep dan hubungan sistematis diantara konsep.

Konsep adalah kelompok mental dengan sifat berbeda menjadi satu kategori berdasarkan indentifikasi kesamaan prinsip (Flavell, Miller, & Miller, 2002). Konsep merupakan kumpulan atribut atau karakteristik umum dari semua contoh (orang, obyek, kejadian, ide) menjadi kelompok (tipe, jenis, kategori) tertentu, atau karakteristik membuat tipe contoh tertentu dan membedakannya menjadi contoh dan noncontoh (Ehrenberg, 1985).

Analisis konsep pada penelitian ini diadaptasi dari Herron (1977). Analisis konsep dalam tujuh tahapan, yaitu: (1) menentukan nama atau label konsep, (2) definisi konsep, (3) jenis konsep, (4) atribut kritis, (5) atribut variabel, (6) posisi konsep (superordinat, ordinat, sub ordinat), (7) contoh dan noncontoh. Jenis-jenis konsep dalam analisis konsep: (1) konsep konkrit yaitu konsep yang contohnya, atribut kritis dan atribut variabel mudah dilihat dengan jelas, (2) konsep abstrak

yaitu konsep yang contohnya tidak dapat dilihat, (3) konsep dengan atribut kritis abstrak contohnya dapat dilihat, (4) konsep berdasarkan pengetahuan prinsip, yaitu konsep yang membutuhkan prinsip-prinsip pengetahuan untuk membedakan contoh dan noncontoh, (5) konsep melibatkan representasi simbol, 6) konsep yang menjabarkan suatu sifat atau atribut (Herron dkk., 1977). Atribut kritis menyatakan ciri-ciri utama suatu konsep yang menjabarkan definisi konsep. Atribut variabel menunjukkan ciri-ciri konsep yang nilainya dapat berubah tetapi besaran dan satuannya tetap (Liliasari, 1996).

b) Mengembangkan peta konsep dan konsep esensial

Manusia berkompetensi menggunakan simbol (lisan atau tulisan) guna menggambarkan makna konsep dan membangun hubungan antar konsep terhadap segala sesuatu obyek atau peristiwa (Novak, 1991). Peta konsep adalah perangkat skematis untuk menggambarkan keterkaitan makna konsep dalam kerangka kerja proposisi. Proposisi adalah makna hubungan dua atau lebih konsep (Novak & Gowin, 1984).

Peta konsep didasarkan teori pembelajaran bermakna Ausubel (Novak, 1990a; Novak & Cañas, 2006). Psikologi kognitif Ausubel menjelaskan pembelajaran merupakan proses asimilasi konsep dan proporsi baru ke dalam konsep dan proporsi dalam kognitif siswa (Novak & Cañas, 2007). Gagasan-gagasan utama teori belajar kognitif Ausubel sebagai landasan peta konsep, yaitu: 1) struktur kognitif tersusun secara hierarkis konsep dan proposisi lebih inklusif superordinat terhadap konsep dan proposisi kurang inklusif, 2) konsep-konsep dalam struktur kognitif mengalami diferensial progresif, artinya pengembangan konsep paling baik jika unsur-unsur paling umum (paling inklusif) diperkenalkan lebih dahulu, kemudian diberikan hal-hal yang lebih mendetail dan lebih khusus dari konsep, 3) penyesuaian integratif menjelaskan belajar bermakna meningkat bila siswa mengenal hubungan-hubungan baru sekumpulan konsep (Dahar, 2011).

Tahapan-tahapan pengembangan peta konsep, antara lain: 1) Meninjau domain topik atau konsep: membaca buku atau artikel, merefleksi konsep yang telah dikuasai; 2) Mengidentifikasi konsep utama: mendaftar beberapa konsep yang dianggap penting; 3) Menentukan hierarki konsep: menentukan konsep paling inklusif ke konsep yang kurang inklusif dari konsep utama; 4) Menghubungkan atau

memberi label hubungan antar konsep: memahami seluruh hubungan antar konsep dalam peta; 5) Memperdalam atau memperluas peta konsep: menambahkan konsep subordinat ke peta semaksimal mungkin untuk memperdalam dan memperluas pengetahuan pada konsep tersebut; 6) Membuat *cross-link*: mencoba menghubungkan dan membuat label hubungan antara konsep yang sudah ada pada peta (Mintzes, 2006).

Konsep-konsep esensial menunjukkan konsep-konsep penting dari konten Mekanika. Konsep-konsep esensial diturunkan dari tahapan sebelumnya yaitu pengembangan peta konsep. Tabel 3.1 menjelaskan adaptasi tahapan *framework* pengembangan instrumen asesmen pilihan ganda empat tingkat.

Tabel 3.1 Adaptasi Tahapan Mendefinisikan Konten Mekanika Treagust (1995)

No	Tahapan analisis konten Treagust	Tahapan analisis konten penelitian ini	Alasan rasional
1	Identifikasi konten	Analisis konsep dengan <i>framework</i> Herron	Keunggulan <i>framework</i> Herron antara lain: a) mendefinisikan jenis dan definisi konsep secara spesifik; b) menjelaskan variabel fisis utama yang berpengaruh pada konsep; c) menentukan posisi konsep dengan tepat (superordinat, ordinat, sub ordinat) merupakan landasan penentuan hierarki konsep pada peta konsep; d) menjelaskan contoh dan non-contoh setiap konsep pada fenomena fisis sehari-hari
2	Pengembangan peta konsep	Peta konsep dengan <i>framework</i> Novak & Cañas (2007) dan Mintzes (2006)	Keunggulan peta konsep Novak, Cañas, dan Mintzes antara lain: a) mendefinisikan secara spesifik karakteristik konsep utama; b) mendefinisikan secara spesifik karakteristik hierarki konsep meliputi konsep kurang inklusif ke konsep inklusif, sebagai landasan penentuan hierarki konsep adalah posisi konsep (superordinat, ordinat, dan sub ordinat) pada analisis konsep; c) mendeskripsikan secara spesifik karakteristik <i>cross-link</i> dan label konsep dari <i>cross-link</i>
3	Identifikasi pertanyaan	Konsep-konsep esensial	Komponen konsep-konsep esensial diturunkan dari hasil analisis konsep dan pengembangan peta konsep.

Ratna Ekawati, 2021

**PENGEMBANGAN INSTRUMEN ASESMEN PILIHAN GANDA EMPAT-TINGKAT UNTUK MENDIAGNOSIS KEMAMPUAN MULTI-REPRESENTASI CALON GURU PADA KONSEP MEKANIKA**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

proporsi pengetahuan	Menghasilkan konsep esensial utama dan turunannya.
----------------------	--

2) Menganalisis konsepsi mahasiswa sebagai landasan menyusun distraktor soal

Tahap kedua pengembangan tes diagnostik adalah menganalisis konsepsi siswa meliputi: a) menganalisis literatur yang relevan terkait konsepsi mahasiswa pada konten mekanika, b) melakukan wawancara dengan siswa tentang pemahaman konten sains mereka, c) serta tanggapan pertanyaan *paper and pencil* yang bersifat terbuka.

a) Menganalisis literatur terkait konsepsi mahasiswa

Tahap ini mengidentifikasi konsepsi mahasiswa pada materi mekanika digunakan sebagai dasar informasi untuk mengembangkan distraktor pada pilihan jawaban soal *four-tier*. Identifikasi konsepsi mahasiswa pada materi mekanika berdasarkan hasil penelitian sebelumnya yang telah dimuat dalam jurnal-jurnal penelitian. Temuan informasi konsepsi mahasiswa dikelompokkan berdasarkan konsep-konsep esensial materi mekanika untuk mempermudah membuat distraktor pada tes pilihan ganda empat-tingkat.

b) Melakukan wawancara terkait konsepsi mahasiswa

Wawancara digunakan secara luas pada bidang psikologi kognitif untuk menyelidiki secara mendalam konten dan proses kognitif (Fraenkel & Wallen, 1993; Osborne & Gilbert, 1980b). Berbagai jenis teknik wawancara digunakan untuk mengungkap konsepsi mahasiswa, yaitu *Piagetian Clinical interview* (PCI) (Rice & Feher, 1987), *Interview-About-Instance* (IAI) (Osborne & Gilbert, 1979, 1980a), dan *Interview-About-Even* (IAE) (Kruger, 1990a).

PCI berfokus pada pengetahuan mahasiswa. Mahasiswa menjelaskan percobaan, pertanyaan, atau tugas selama wawancara (Bendall & Goldberg, 1993). Tujuan utama wawancara klinis adalah memahami pola penalaran siswa tanpa mengubah pola pemahaman dalam diri mahasiswa. Wawancara klinis memberikan rincian bagaimana mahasiswa memahami suatu konsep fisika tertentu (Engelhardt dkk., 2003).

Pengembangan tes tahap ini menggunakan wawancara semi terstruktur dengan teknik PCI. Data pada tahap ini adalah konsepsi mahasiswa pada materi

mekanika. Fungsi utama tahap ini mengidentifikasi konsepsi mahasiswa sebagai landasan pengembangan distraktor jawaban (Kaltakci-Gurel dkk., 2017).

c) Mengembangkan tes dua-tingkat dengan respon jawaban bebas

Pengembangan item tes dua-tingkat dengan respon jawaban bebas didasarkan analisis konsep, peta konsep, konsep esensial dan wawancara konsepsi mahasiswa. Tahap ini untuk melengkapi kajian konsepsi dari analisis literatur dan wawancara. Tes jawaban bebas terdiri dari dua tingkat. Tingkat pertama berisi pertanyaan pilihan ganda terkait konten mekanika, dan tingkat kedua berisi alasan terhadap tingkat pertama dengan respon jawaban terbuka. Tabel 3.2 menjelaskan adaptasi tahapan analisis konsepsi mekanika sebagai landasan pengembangan distraktor tes Treagust.

Tabel 3.2 Adaptasi Tahap Menganalisis Konsepsi Mahasiswa Treagust (1995)

No.	Tahap analisis konsepsi mahasiswa Treagust	Tahap analisis konsepsi penelitian ini	Keterangan
1	Menganalisis literatur terkait	Menganalisis literatur ilmiah	Analisis konsepsi mahasiswa pada konsep esensial mekanika
2	Menggali konsepsi mahasiswa dengan menerapkan item pilihan ganda jawaban terbuka	Menggali konsepsi dengan pilihan ganda dua-tingkat dengan jawaban bebas	Hasil respon jawaban bebas tingkat kedua sebagai landasan menyusun distraktor soal pada pilihan ganda empat-tingkat
3	Wawancara semi-terstruktur	<i>Piagetian Clinical Interview (PCI)</i>	Wawancara klinis konsepsi mekanika mahasiswa berdasarkan tes pilihan ganda dua-tingkat dengan jawaban bebas

## 2. Pengembangan Tes Diagnostik

Karakteristik tes diagnostik empat-tingkat meliputi tingkat-pertama pertanyaan pilihan ganda pada konten fisika, tingkat-kedua bertanya tingkat kepercayaan memilih jawaban pada tingkat-pertama, tingkat-ketiga bertanya alasan memilih jawaban pada tingkat-pertama dan tingkat-keempat bertanya tingkat kepercayaan memilih jawaban pada tingkat-ketiga. Tahapan pengembangan ini

adalah a) merancang kisi-kisi tes, b) mengembangkan tes pilihan ganda empat-tingkat, dan c) menelaah tes pilihan ganda empat-tingkat oleh ahli.

a. Merancang kisi-kisi tes

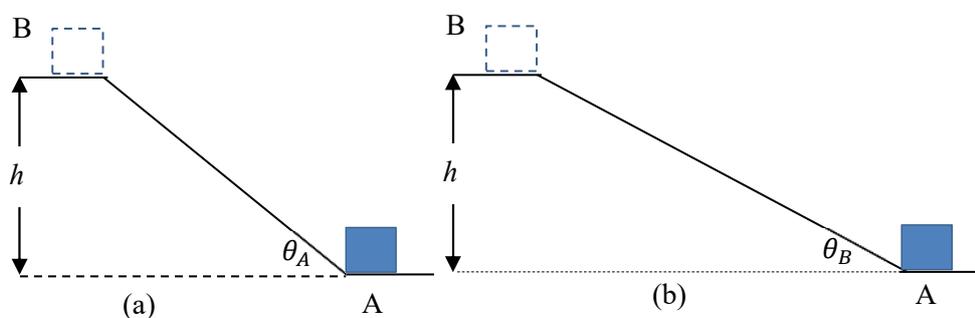
Tujuan mengembangkan kisi-kisi tes adalah menentukan konstruksi tes secara teliti, lingkup dan fokus dari tes diagnostik multi-representasi. Dua aspek komponen kisi-kisi tes subyek materi dan jenis kemampuan yang diuji (Tinkelman, 1971). Kisi-kisi tes mendefinisikan konten materi, jumlah item, proses kognitif yang diujikan, format item, dan pengelompokan item (AERA dkk., 1999). Kisi-kisi instrumen dikembangkan berdasarkan tahapan indentifikasi konsep.

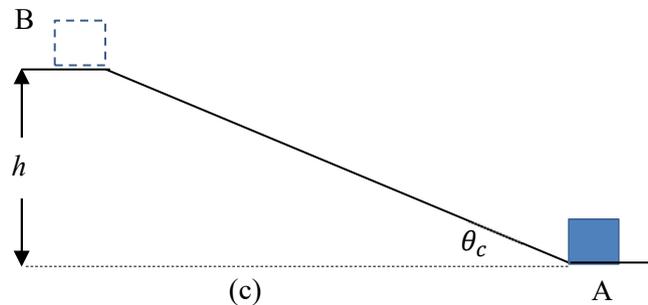
b. Mengembangkan tes diagnostik pilihan ganda empat-tingkat

Tingkat pertama adalah pertanyaan konsep Mekanika. Tingkat kedua merupakan tingkat keyakinan terhadap respon jawaban pada tingkat pertama. Tingkat ketiga alasan terhadap respon jawaban pada tingkat pertama, tingkat keempat adalah tingkat keyakinan terhadap respon alasan pada tingkat ketiga. Tingkat keyakinan terhadap respon jawaban, yaitu: 1) hanya menebak, 2) sangat tidak yakin, 3) tidak yakin, 4) yakin, 5) yakin sekali, 6) sangat yakin sekali (Caleon & Subramaniam, 2010b).

Item soal dikembangkan berdasarkan kisi-kisi tes. Distraktor pilihan jawaban pada tingkat pertama dan kedua hasil merupakan hasil identifikasi konsepsi mahasiswa yaitu: tahap analisis literatur terkait konsepsi mahasiswa, wawancara terkait konsepsi mahasiswa, dan tes pilihan ganda dua-tingkat dengan respon jawaban bebas. Contoh format tes diagnostik pilihan ganda empat-tingkat dijelaskan pada Gambar 3.3.

10A Sebuah benda bermassa  $m$  akan dipindahkan dari A ke posisi lebih tinggi B melalui permukaan yang licin dengan tiga macam cara seperti pada gambar di bawah ini.





### A. Representasi Verbal

10.A.1 Penjelasan yang tepat tentang hubungan usaha dan perubahan posisi adalah....

- Semakin besar sudut kemiringan perpindahan benda maka semakin kecil usahanya
- Semakin besar sudut kemiringan perpindahan benda maka semakin besar usahanya
- Usaha ditentukan oleh posisi awal ke posisi akhir benda pada sistem gerak
- Usaha dipengaruhi oleh lintasan perpindahan benda dari posisi A ke posisi B

10.A.2 Tingkat keyakinan Anda terhadap jawaban 10.A.1 ....

- |                       |                        |
|-----------------------|------------------------|
| a. Hanya menebak      | d. Yakin               |
| b. Sangat tidak yakin | e. Yakin sekali        |
| c. Tidak yakin        | f. Sangat yakin sekali |

10.A.3 Alasan Anda memilih jawaban 10.A.1 ....

- Usaha total karena gaya gravitasi yaitu  $W = -mg \sin \theta \Delta h$ , maka semakin besar sudut kemiringan perpindahan maka semakin besar usahanya
- Usaha total berkaitan dengan gaya normal dan gravitasi yaitu  $W = -mg \cos \theta \Delta h$ , maka semakin besar sudut kemiringan perpindahan maka semakin kecil usahanya
- Usaha total karena gaya gravitasi  $W = mg \Delta h$ , maka variabel utamanya perpindahan dari posisi awal ke posisi akhir
- Lainnya:

.....  
 .....  
 .....

10.A.4 Tingkat keyakinan Anda terhadap alasan jawaban 10.A.3 ....

- |                       |                 |
|-----------------------|-----------------|
| a. Hanya menebak      | d. Yakin        |
| b. Sangat tidak yakin | e. Yakin sekali |

c. Tidak yakin

f. Sangat yakin sekali

Gambar 3.3 Contoh Item Tes Pilihan Ganda Empat-Tingkat

## c. Menelaah tes pilihan ganda empat-tingkat oleh ahli

Telaah soal diperlukan jika dalam pembuatannya masih ditemukan kekurangan atau kesalahan. Penelaahan terhadap instrumen diagnostik multi-representasi dilakukan oleh kelompok pakar (*expert*). Pertimbangan profesional oleh kelompok *expert* untuk menentukan validitas isi butir soal baik dari segi materi, konstruksi soal, maupun kejelasan bahasa agar butir soal yang diujikan representatif kompetensi dan konsep fisika yang harus dikuasai.

Kelompok *expert* harus dipilih yang berkompeten di bidang fisika dan/atau asesmen dengan latar belakang pendidikan S3 (doktor). Jumlah *expert* ganjil untuk memudahkan mengambil keputusan ketika menganalisis hasil penilaian dari *expert*. Indeks validitas Aiken digunakan untuk mengetahui kesepakatan validator *expert* (Aiken, 1980, 1985). Indeks validitas Aiken dirumuskan sebagai berikut:

$$V = \frac{\sum s}{n(c - 1)}$$

V adalah indeks kesepakatan rater mengenai validitas butir; s adalah skor yang ditetapkan setiap rater dikurangi skor terendah dalam kategori yang dipakai ( $s = r - l_o$ , r = skor kategori pilihan rater,  $l_o$  = skor terendah dalam kategori penyekoran); n adalah banyaknya rater; dan c banyaknya kategori yang dapat dipilih oleh rater. Hasil perhitungan indeks validitas Aiken (1980) diinterpretasi pada Tabel 3.3.

Tabel 3.3 Kategori Indeks Validitas Aiken

Koefisien Validitas Aiken	Kategori Validitas
0,00-0,39	Kurang
0,40-0,79	Sedang ( <i>mediocare</i> )
0,80-1,00	Tinggi

## d. Pengembangan Instrumen Diagnostik Berbasis Multi Representasi.

Kegiatan pada tahap ini melakukan uji coba (*pilot study*) dan menganalisis butir soal. Tujuan pelaksanaan uji coba (*pilot study*) adalah memperoleh data empirik. Data empirik diperlukan untuk menguji apakah produk yang dikembangkan memenuhi kriteria validitas konstruk, reabilitas, dan parameter butir yang baik. Validitas konstruk adalah validitas menunjukkan sejauhmana instrumen mengungkap kemampuan (*trait*) atau konstruk teori tertentu yang akan diukur

(Gregory, 2015). Uji coba dilakukan melalui dua tahap, yaitu: uji coba terbatas dan uji coba diperluas. Hasil administrasi uji coba terbatas diestimasi dengan model *Rasch*: 1) tingkat kesesuaian butir soal (*item fit*), 2) tingkat kesesuaian individu (*person fit*), 3) peta wright (*person-item maps*), 4) tingkat kesulitan butir, 5) reliabilitas *person-item*, 6) mendeteksi bias item (*item DIF*). Hasil administrasi data tahap ini adalah rancangan instrumen tes diagnostik multi-representasi telah direvisi sehingga diperoleh draf yang lebih baik.

Tahapan administrasi skor tes diagnostik multi-representasi adalah 1) menyusun kemungkinan pola jawaban, 2) menyusun rubrik penskoran pada setiap pola jawaban diadaptasi dari Gurcay dan Gulbas, 3) mengkategorisasi respon jawaban siswa dengan mengadopsi teknik analisis miskonsepsi *four-tier test* Kaltakci-Gurel (2015). Adapun teknik analisis miskonsepsi versi Kaltakci-Gurel (2015) dijelaskan pada Tabel 3.4.

Tabel 3.4  
*Kombinasi dan Kategori Jawaban pada Four-Tier Test*

Tingkat ke-1	Tingkat ke-2	Tingkat ke-3	Tingkat ke-4	Kategori
Benar	Yakin	Benar	Yakin	Konsepsi ilmiah
Benar	Yakin	Benar	Tidak yakin	Kurang pengetahuan
Benar	Tidak yakin	Benar	Yakin	Kurang pengetahuan
Benar	Tidak yakin	Benar	Tidak yakin	Kurang pengetahuan
Benar	Yakin	Salah	Yakin	<i>False-positive</i>
Benar	Yakin	Salah	Tidak yakin	Kurang pengetahuan
Benar	Tidak yakin	Salah	Yakin	Kurang pengetahuan
Benar	Tidak yakin	Salah	Tidak yakin	Kurang pengetahuan
Salah	Yakin	Benar	Yakin	<i>False negative</i>
Salah	Yakin	Benar	Tidak yakin	Kurang pengetahuan
Salah	Tidak yakin	Benar	Yakin	Kurang pengetahuan
Salah	Tidak yakin	Benar	Tidak yakin	Kurang pengetahuan
Salah	Yakin	Salah	Yakin	Miskonsepsi
Salah	Yakin	Salah	Tidak yakin	Kurang pengetahuan
Salah	Tidak yakin	Salah	Yakin	Kurang pengetahuan
Salah	Tidak yakin	Salah	Tidak yakin	Kurang pengetahuan

Tahapan kemungkinan pola jawaban pada instrumen tes diagnostik multi-representasi dijelaskan pada Tabel 3.7. Tahapan pengembangan penskoran dijelaskan pada Tabel 3.5.

Tabel 3.5  
*Kategori Skor Berdasarkan Respon Jawaban Mahasiswa*

No.	Kategori Jawaban	Respon Jawaban	Skor
1	Respon jawaban benar tingkat pertama, benar tingkat tiga, respon keyakinan tinggi pada tingkat dua dan empat	B-SYS-B-SYS B-YS-B-SYS B-Y-B-SYS B-SYS-B-Y B-YS-B-YS B-Y-B-YS B-SYS-B-Y B-YS-B-YS B-Y-B-Y	4
2	Respon jawaban benar tingkat pertama, benar tingkat tiga, respon keyakinan tinggi pada tingkat kedua, dan rendah pada tingkat empat	B-SYS-B-TY B-Y-B-TY B-Y-B-TY B-SYS-B-STY B-YS-B-STY B-Y-B-STY B-SYS-B-HM B-YS-B-HM B-Y-B-HM	3
3	Respon jawaban benar tingkat pertama, benar tingkat tiga, respon keyakinan rendah tingkat dua, dan tinggi tingkat empat	B-TY-B-SYS B-STY-B-SYS B-HM-B-SYS B-TY-BYS B-STY-B-YS B-HM-B-YS B-TY-B-Y B-STY-B-Y B-HM-B-Y	3
4	Respon jawaban benar tingkat pertama, benar tingkat ketiga, respon keyakinan rendah pada tingkat dua dan empat	B-TY-B-TY B-STY-B-TY B-HM-B-TY B-TY-B-STY B-TY-B-STY B-STY-B-STY B-HM-B-STY B-TY-B-HM B-STY-B-HM B-STY-B-HM B-HM-B-HM	3
5		B-SYS-S-SYS	2

No.	Kategori Jawaban	Respon Jawaban	Skor
	Respon jawaban benar tingkat pertama, salah tingkat ketiga, respon keyakinan tinggi tingkat dua dan empat	B-YS-S-SYS B-Y-S-SYS B-SYS-S-YS B-YS-S-YS B-Y-S-YS B-SYS-S-Y B-YS-S-Y B-Y-S-Y	
6	Respon jawaban benar tingkat pertama, salah tingkat ketiga, respon keyakinan tingkat dua tinggi dan empat rendah	B-SYS-S-TY B-YS-S-TY B-Y-S-TY B-SYS-S-STY B-SY-S-STY B-Y-S-STY B-SYS-S-HM B-YS-S-HM B-Y-S-HM	2
7	Respon jawaban benar tingkat pertama, salah tingkat tiga, respon keyakinan rendah tingkat dua dan tinggi tingkat empat	B-TY-S-SYS B-STY-S-SYS B-HM-S-YS B-TY-S-YS B-STY-S-YS B-HM-S-YS B-TY-S-Y B-STY-S-Y B-HM-S-Y	2
8	Respon jawaban benar tingkat pertama, salah tingkat tiga, respon keyakinan rendah tingkat dua, dan rendah tingkat empat	B-TY-S-TY B-STY-S-TY B-HM-S-TY B-TY-S-STY B-STY-B-STY B-HM-S-STY B-TY-S-HM B-STY-S-HM B-HM-S-HM	2
9	Respon jawaban benar tingkat pertama, salah tingkat tiga, respon	S-SYS-B-SYS S-SY-B-SYS S-Y-B-SYS	2

No.	Kategori Jawaban	Respon Jawaban	Skor
	jawaban keyakinan tinggi tingkat dua, dan tinggi tingkat empat	S-SYS-B-YS S-YS-B-YS S-Y-B-YS S-SYS-B-Y S-YS-B-Y S-Y-B-Y	
10	Respon jawaban salah tingkat pertama, benar tingkat tiga, respon keyakinan tinggi tingkat dua, dan rendah tingkat empat	S-SYS-B-TY S-YS-B-TY S-Y-B-TY S-SYS-B-STY S-YS-B-STY S-Y-B-YS S-SYS-B-HM S-YS-B-HM S-Y-B-HM	2
11	Respon jawaban salah tingkat pertama, benar tingkat dua, respon keyakinan rendah tingkat dua dan respon keyakinan tinggi tingkat empat	S-TY-B-SYS S-STY-B-SYS S-HM-B-SYS S-TY-B-YS S-STY-B-YS S-HM-B-YS S-TY-B-Y S-STY-B-Y S-HM-B-Y	2
12	Respon jawaban salah tingkat pertama, benar tingkat dua, respon keyakinan rendah tingkat dua dan empat	S-TY-B-TY S-STY-B-TYS S-HM-B-TY S-TY-B-STY S-STY-B-STY S-HM-B-HM S-TY-B-HM S-STY-B-HM S-HM-B-HM	2
13	Respon jawaban salah tingkat pertama, salah tingkat dua, respon keyakinan tinggi tingkat dua dan empat	S-SYS-S-SYS S-YS-S-SYS S-Y-S-SYS S-SYS-S-YS S-YS-S-YS	1

No.	Kategori Jawaban	Respon Jawaban	Skor
		S-Y-S-YS	
		S-SYS-S-YS	
		S-YS-S-Y	
		S-Y-S-Y	
14	Respon jawaban salah tingkat pertama, salah tingkat dua, respon keyakinan tinggi tingkat tiga dan rendah tingkat empat	S-SYS-S-TY	1
		S-YS-S-TY	
		S-Y-S-TY	
		S-SYS-S-STY	
		S-YS-S-STY	
		S-Y-S-STY	
		S-SYS-S-HM	
		S-YS-S-HM	
		S-Y-S-HM	
15	Respon jawaban salah tingkat pertama, salah tingkat dua, respon keyakinan rendah tingkat tiga dan tinggi tingkat empat	S-TY-S-SYS	1
		S-STY-S-SYS	
		S-HM-S-SYS	
		S-TY-S-YS	
		S-STY-S-YS	
		S-HM-S-YS	
		S-TY-S-Y	
		S-STY-S-Y	
		S-HM-S-Y	
16	Respon jawaban salah tingkat pertama, salah tingkat dua, respon keyakinan rendah tingkat tiga dan empat	S-TY-S-TY	1
		S-STY-S-TYS	
		S-HM-S-TY	
		S-TY-S-STY	
		S-STY-S-TY	
		S-TY-S-HM	
		S-STY-S-HM	
		S-HM-S-HM	

Keterangan: B: benar, S: salah, SYS: sangat yakin sekali, YS: yakin sekali, Y: yakin, TY: tidak yakin, STY: sangat tidak yakin, HM: hanya menebak.

Setelah proses pengembangan pedoman penskoran, tahapan selanjutnya mengategorisasi kemampuan mahasiswa dalam konsepsi ilmiah atau miskonsepsi untuk format tes *four-tier* yang diadaptasi dari Kaltakci-Gurel (2017). Tabel 3.6 menjelaskan kategori pemahaman siswa untuk format tes *four-tier* berdasarkan pola kemungkinan respon jawaban mahasiswa.

Tabel 3.6

*Kategori Tingkat Pemahaman Berdasarkan Respon Jawaban Mahasiswa*

No.	Tier-1	Tier-2	Tier-3	Tier-4	Respon Jawaban	Kategori
1	Benar	Tinggi	Benar	Tinggi	B-SYS-B-SYS B-YS-B-SYS B-Y-B-SYS B-SYS-B-Y B-YS-B-YS B-Y-B-YS B-SYS-B-Y B-YS-B-YS B-Y-B-Y	Konsep ilmiah
2	Benar	Tinggi	Benar	Rendah	B-SYS-B-TY B-Y-B-TY B-Y-B-TY B-SYS-B-STY B-YS-B-STY B-Y-B-STY B-SYS-B-HM B-YS-B-HM B-Y-B-HM	Kurang paham
3	Benar	Rendah	Benar	Tinggi	B-TY-B-SYS B-STY-B-SYS B-HM-B-SYS B-TY-BYS B-STY-B-YS B-HM-B-YS B-TY-B-Y B-STY-B-Y B-HM-B-Y	Kurang paham
4	Benar	Rendah	Benar	Rendah	B-TY-B-TY B-STY-B-TY B-HM-B-TY B-TY-B-STY B-TY-B-STY B-STY-B-STY B-HM-B-STY B-TY-B-HM B-STY-B-HM	Kurang paham
5	Benar	Tinggi	Salah	Tinggi	B-SYS-S-SYS B-YS-S-SYS B-Y-S-SYS B-SYS-S-YS B-YS-S-YS B-Y-S-YS B-SYS-S-Y B-YS-S-Y B-Y-S-Y	<i>False-positive</i>

Ratna Ekawati, 2021

**PENGEMBANGAN INSTRUMEN ASESMEN PILIHAN GANDA EMPAT-TINGKAT UNTUK MENDIAGNOSIS KEMAMPUAN MULTI-REPRESENTASI CALON GURU PADA KONSEP MEKANIKA**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

No.	Tier-1	Tier-2	Tier-3	Tier-4	Respon Jawaban	Kategori
6	Benar	Tinggi	Salah	Rendah	B-SYS-S-TY B-YS-S-TY B-Y-S-TY B-SYS-S-STY B-SY-S-STY B-Y-S-STY B-SYS-S-HM B-YS-S-HM B-Y-S-HM	Kurang paham
7	Benar	Rendah	Salah	Tinggi	B-TY-S-SYS B-STY-S-SYS B-HM-S-YS B-TY-S-YS B-STY-S-YS B-HM-S-YS B-TY-S-Y B-STY-S-Y B-HM-S-Y	Kurang paham
8	Benar	Rendah	Salah	Rendah	B-TY-STY B-STY-S-TY B-HM-S-TY B-TY-S-STY B-STY-B-STY B-HM-S-STY B-TY-S-HM B-STY-S-HM B-HM-S-HM	Kurang paham
9	Salah	Tinggi	Benar	Tinggi	S-SYS-B-SYS S-SY-B-SYS S-Y-B-SYS S-SYS-B-YS S-YS-B-YS S-Y-B-YS S-SYS-B-Y S-YS-B-Y S-Y-B-Y	<i>False-negative</i>
10	Salah	Tinggi	Benar	Rendah	S-SYS-B-TY S-YS-B-TY S-Y-B-TY S-SYS-B-STY S-YS-B-STY S-Y-B-YS S-SYS-B-HM S-YS-B-HM S-Y-B-HM	Kurang paham
11	Salah	Rendah	Benar	Tinggi	S-TY-B-SYS S-STY-B-SYS	Kurang paham

Ratna Ekawati, 2021

**PENGEMBANGAN INSTRUMEN ASESMEN PILIHAN GANDA EMPAT-TINGKAT UNTUK MENDIAGNOSIS KEMAMPUAN MULTI-REPRESENTASI CALON GURU PADA KONSEP MEKANIKA**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

No.	Tier-1	Tier-2	Tier-3	Tier-4	Respon Jawaban	Kategori
					S-HM-B-SYS	
					S-TY-B-YS	
					S-STY-B-YS	
					S-HM-B-YS	
					S-TY-B-Y	
					S-STY-B-Y	
					S-HM-B-Y	
12	Salah	Rendah	Benar	Rendah	S-TY-B-TY	Kurang paham
					S-STY-B-TYS	
					S-HM-B-TY	
					S-TY-B-STY	
					S-STY-B-STY	
					S-HM-B-HM	
					S-TY-B-HM	
					S-STY-B-HM	
					S-HM-B-HM	
13	Salah	Tinggi	Salah	Tinggi	S-SYS-S-SYS	Miskonsepsi
					S-YS-S-SYS	
					S-Y-S-SYS	
					S-SYS-S-YS	
					S-YS-S-YS	
					S-Y-S-YS	
					S-SYS-S-YS	
					S-YS-S-Y	
					S-Y-S-Y	
14	Salah	Tinggi	Salah	Rendah	S-SYS-S-TY	Kurang paham
					S-YS-S-TY	
					S-Y-S-TY	
					S-SYS-S-STY	
					S-YS-S-STY	
					S-Y-S-STY	
					S-SYS-S-HM	
					S-YS-S-HM	
					S-Y-S-HM	
15	Salah	Rendah	Salah	Tinggi	S-TY-S-SYS	Kurang paham
					S-STY-S-SYS	
					S-HM-S-SYS	
					S-TY-S-YS	
					S-STY-S-YS	
					S-HM-S-YS	
					S-TY-S-Y	
					S-STY-S-Y	
					S-HM-S-Y	
16	Salah	Rendah	Salah	Rendah	S-TY-S-TY	Kurang paham
					S-STY-S-TYS	
					S-HM-S-TY	
					S-TY-S-STY	

No.	Tier-1	Tier-2	Tier-3	Tier-4	Respon Jawaban	Kategori
					S-STY-S-TY	
					S-TY-S-HM	
					S-STY-S-HM	
					S-HM-S-HM	

Keterangan: B: benar, S: salah, SYS: sangat yakin sekali, YS: yakin sekali, Y: yakin, TY: tidak yakin, STY: sangat tidak yakin, HM: hanya menebak.

### 3. Pengumpulan Data Kuantitatif

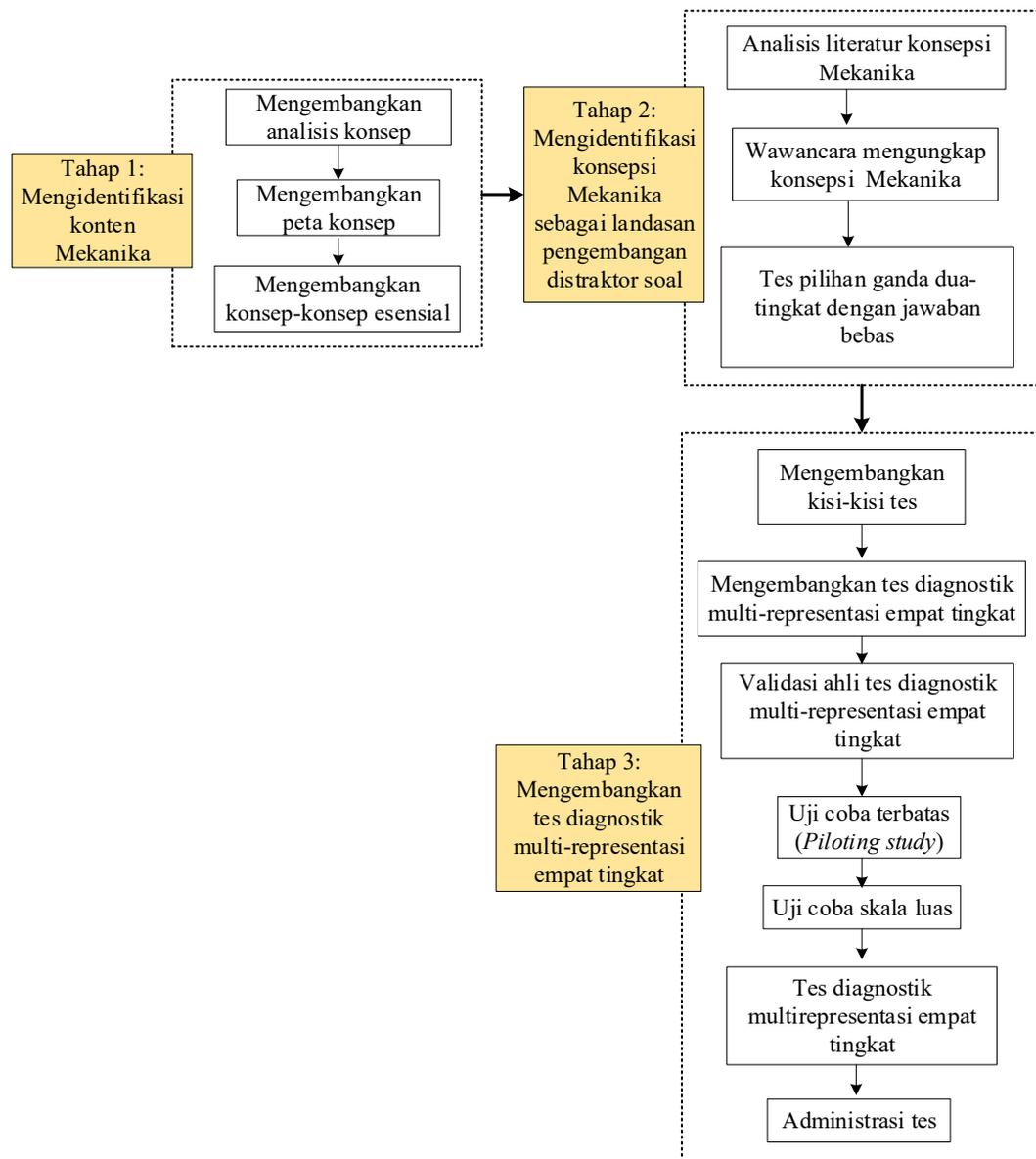
Dua tahapan dalam pengumpulan data kuantitatif: a) pengumpulan data dan analisis data kuantitatif, b) hasil analisis data kuantitatif. Adapun kegiatan dalam tiap tahapan dijelaskan sebagai berikut:

#### a. Pengumpulan dan analisis data kuantitatif

Setelah uji coba terbatas dilakukan analisis, langkah berikutnya uji coba skala luas. Administrasi uji coba diperluas bertujuan memperoleh estimasi empirik dengan model *Rasch*: 1) tingkat kesesuaian butir soal (*item fit*), 2) tingkat kesesuaian individu (*person fit*), 3) peta *wright* (*person-item maps*), 4) tingkat kesulitan butir, 5) reliabilitas *person-item*, 6) mendeteksi bias item (*item DIF*).

#### b. Hasil analisis data kuantitatif

Tahap ini dilakukan perbaikan bagian soal yang masih belum sesuai dengan yang diharapkan. Ada kemungkinan beberapa soal baik sehingga tidak perlu direvisi, dan beberapa lain harus dibuang karena tidak memenuhi standar kuliatas yang diharapkan. Setelah semua butir soal diperbaiki, langkah berikutnya adalah merakit butir-butir soal tersebut menjadi satu kesatuan. Keseluruhan butir perlu disusun secara hati-hati menjadi satu kesatuan soal yang terpadu. Dalam merakit soal, hal-hal yang dapat mempengaruhi validitas soal seperti nomor urut soal, dan *layout* harus diperhatikan. Berdasarkan hasil analisis data, perbaikan, dan perakitan tes diperoleh draf final instrumen diagnostik multi-representasi pada konsep mekanika. Tahapan pengembangan instrumen tes-diagnostik multi-representasi dijelaskan secara terperinci pada Gambar 3.4. Tahapan pengembangan instrumen diagnostik tersebut diintegrasikan pada desain penelitian *mixed-methods* dengan model *instrumen development*.



Gambar 3.4 *Framework* Pengembangan Tes Diagnostik Multirepresentasi Empat Tingkat

Tabel 3.7 Respon Jawaban Mahasiswa Tes Pilihan Ganda Empat-Tingkat

	Respon jawaban	Tier-1 dan Tier-2											
		B-SYS	B-YS	B-Y	B-TY	B-STY	B-HM	S-SYS	S-YS	S-Y	S-TY	S-STY	S-HM
Tier-3 dan Tier-4	B-SYS	B-SYS- B-SYS	B-YS- B-SYS	B-Y- B-SYS	B-TY- B-SYS	B-STY- B-SYS	B-HM- B-SYS	S-SYS- B-SYS	S-YS- B-SYS	SY-B- B-SYS	S-TY- B-SYS	S-STY- B-SYS	S-HM- B-SYS
	B-YS	B-YS- BSYS	B-YS- B-YS	B-Y- B-YS	B-TY- B-YS	B-STY- B-YS	B-HM- B-YS	S-SYS- B-YS	B-YS- B-YS	B-Y- B-YS	S-TY- B-YS	S-STY- B-YS	S-HM- B-YS
	B-Y	B-SYS- B-Y	B-YS- B-Y	B-Y- B-Y	B-TY- B-Y	B-STY- B-Y	B-HM- B-Y	S-SYS- B-Y	S-YS- B-Y	S-Y- B-Y	S-TY- B-Y	S-STY- B-Y	S-HM- B-Y
	B-TY	B-SYS- B-TY	B-YS- B-TY	B-Y- B-TY	B-TY- B-TY	B-STY- B-TY	B-HM- B-TY	S-SYS- B-TY	S-YS- B-TY	S-Y- B-TY	S-TY- B-TY	S-STY- B-TY	S-HM- B-TY
	B-STY	B-SYS- B-STY	B-YS- B-STY	B-Y- B-STY	B-TY- B-STY	B-STY- B-STY	B-HM- B-STY	B-SYS- B-STY	S-YS- B-STY	S-Y- B-STY	S-TY- B-STY	S-STY- B-STY	S-HM- B-STY
	B-HM	B-HM- B-SYS	B-SY- B-HM	B-Y- B-HM	B-TY- B-HM	B-STY- B-HM	B-HM- B-HM	S-SYS- B-HM	S-YS- B-HM	S-Y- B-HM	S-TY- B-HM	S-STY- B-HM	S-HM- B-HM
	S-SYS	B-SYS- S-SYS	B-YS- S-SYS	B-Y- S-SYS	B-TY- S-SYS	B-STY- S-SYS	B-HM- S-SYS	S-SYS- S-SYS	S-YS- S-SYS	S-Y- S-SYS	S-TY- S-SYS	S-STY- S-SYS	S-HM- S-SYS
	S-YS	B-SYS- S-YS	B-YS- S-YS	B-Y- B-YS	B-TY- S-YS	B-STY- S-YS	B-HM- S-YS	S-SYS- S-YS	S-YS- S-YS	S-Y- S-YS	S-TY- S-YS	S-STY- S-YS	S-HM- S-YS
	S-Y	B-SYS- S-Y	B-YS- S-Y	B-Y- S-Y	B-TY- S-Y	B-STY- S-Y	B-HM- S-Y	S-SYS- S-Y	S-YS- S-Y	S-Y- S-Y	S-TY- S-Y	S-STY- S-Y	S-HM- S-Y
	S-TY	B-SYS- S-TY	B-YS- S-TY	B-Y- S-TY	B-TY- S-TY	B-STY- S-TY	B-HM- S-TY	S-SYS- S-TY	S-YS- S-TY	S-Y- S-TY	S-TY- S-TY	S-STY- S-TY	S-HM- S-TY
	S-STY	B-SYS- B-STY	B-YS- B-STY	B-Y- S-STY	B-TY- S-STY	B-STY- S-STY	B-HM- S-STY	S-SYS- S-STYS	S-YS- S-STY	S-Y- S-STY	S-TY- S-STY	S-STY- S-STY	S-HM- S-STY
	S-HM	B-SYS- B-HM	B-YS- S-HM	B-Y- S-HM	B-TY- S-HM	B-STY- S-HM	B-HM- S-HM	S-SYS- S-HM	S-YS- S-HM	S-Y- S-HM	S-TY- S-HM	S-STY- S-HM	S-HM- SHM

Keterangan: B: benar, S: salah, SYS: sangat yakin sekali, YS: yakin sekali, Y: yakin, TY: tidak yakin, STY: sangat tidak yakin, HM: hanya menebak.

### C. Subyek Penelitian

Subyek penelitian ini adalah mahasiswa program studi pendidikan fisika yang sedang menempuh mata kuliah Fisika Dasar I pada tahun akademik 2017/2018. Subyek penelitian ditentukan dengan teknik *purposive sampling*. Teknik *purposive sampling* adalah pengambilan sampel didasarkan informasi awal untuk menghasilkan data yang dibutuhkan (Fraenkel & Wallen, 1993; Gay, Millis, & Airasian, 2012; Patton, 2002).

Uji coba utama melibatkan mahasiswa program studi Pendidikan Fisika dari tiga Lembaga Pendidikan Tenaga Pendidikan (LPTK) di Serang, Yogyakarta, dan Bengkulu. Uji coba skala kecil pada mahasiswa program studi Pendidikan Fisika pada salah satu LPTK negeri di Serang. Jumlah subyek penelitian pada uji coba skala kecil adalah mahasiswa pendidikan Fisika semester 3 yang sudah mengontrak mata kuliah Fisika Dasar 1 sejumlah 33 mahasiswa. Subyek uji coba skala besar dideskripsikan pada Tabel 3.8. Pemilihan subyek penelitian pada uji coba skala besar berdasarkan abilitas mahasiswa tinggi, sedang, dan rendah untuk mendapatkan sebaran peta analisis person yang lengkap pada model Rasch.

Tabel 3.8

*Jumlah Subyek Penelitian Uji Coba Skala Besar*

No.	Nama LPTK	Subyek Penelitian
1	LPTK di Yogyakarta	124
2	LPTK di Bengkulu	28
3	LPTK di Serang	22
Jumlah		174

### D. Instrumen Penelitian

Instrumen yang digunakan adalah instrumen penelitian yang dirancang dan dikembangkan berdasarkan permasalahan dan tahapan penelitian. Instrumen yang digunakan dalam penelitian terdiri dari:

1. Tes esai pada studi pendahuluan untuk mendapatkan informasi profil kemampuan multi-representasi awal mahasiswa. Format tes berupa tes esai dengan respon jawaban bebas.
2. Lembar angket untuk mahasiswa pada studi pendahuluan untuk mendapatkan informasi pelaksanaan pembelajaran dan implementasi asesmen pada mata kuliah Fisika Dasar I.

3. Lembar wawancara untuk dosen pada studi pendahuluan untuk mendapatkan informasi pelaksanaan pembelajaran dan implementasi asesmen pada mata kuliah Fisika Dasar I.
4. Lembar pedoman wawancara untuk mengetahui konsepsi Mekanika mahasiswa sebagai landasan pengembangan distraktor pilihan jawaban pada tes tingkat pertama dan ketiga.
5. Tes diagnostik multirepresentasi dua-tingkat dengan jawaban bebas untuk menjangring konsepsi mahasiswa pada konsep Mekanika sebagai landasan pengembangan tes multi-representasi empat-tingkat.
6. Tes diagnostik multi-representasi empat tingkat. Sebuah konsep mekanika direpresentasikan menjadi empat format representasi. Dimana setiap satu format representasi dikembangkan menjadi tes empat tingkat. Tes tingkat pertama bertanya konten fisika, tes tingkat kedua merupakan tingkat keyakinan terhadap respon jawaban pada tingkat pertama, tingkat ketiga adalah alasan terhadap respon jawaban pada tingkat pertama, tingkat keempat merupakan tingkat keyakinan terhadap respon jawaban pada tingkat ketiga.
7. Lembar validasi ahli untuk tes diagnostik multi-representasi empat-tingkat. Lembar validasi ahli meliputi: perbaikan rumusan/ stem soal, perbaikan pilihan jawaban item soal tingkat pertama, perbaikan alasan soal, perbaikan pilihan jawaban, perbaikan tingkat keyakinan pada item tes tingkat ketiga dan keempat, perbaikan indikator soal, perbaikan gambar.

#### **E. Teknik Analisis Data**

Analisis data dalam penelitian ini adalah menganalisis karakteristik instrumen tes. *Item Respon Theory* (IRT) digunakan untuk mengetahui karakteristik instrumen tes. IRT adalah kerangka khusus fungsi matematis yang menjelaskan interaksi antara orang (*persons*) dan butir soal (*test items*) (Sumintono & Widhiarso, 2013). IRT menjelaskan hubungan antara performa tes individu bersifat *observable* dan trait atau kemampuan bersifat *unobservable* (Hambleton & Swaminathan, 1985). IRT membangun model mengenai bagaimana konstruk psikologis yang bersifat laten (*unobservable*) dapat dinyatakan dalam bentuk respons terhadap butir

yang teramati (*observable*) (Azwar, 2015). Dua postulat sebagai dasar IRT: 1) Performa individu pada item tes dapat diprediksi dari kumpulan faktor trait, laten trait, atau kemampuan (*ability*); 2) Kaitan antara item performa individu dan item performa trait (*ability*) dijelaskan dalam karakteristik fungsi item atau *Item Characteristic Curve* (ICC) (Hambleton, Swaminathan, & Rogers, 1991). Berdasarkan penjelasan ahli psikometri tersebut dapat disimpulkan, IRT adalah hubungan probabilitas jawaban benar siswa pada suatu butir soal dan kemampuan siswa (*trait/latent trait*).

Hambleton dan Swaminathan (1985) menjelaskan karakteristik IRT, yaitu: 1) model IRT beranggapan bahwa performa individu dalam tes dapat diprediksi dengan satu atau lebih karakteristik trait, 2) model IRT merinci hubungan item performa individu dan trait atau kemampuan disumsikan sebagai dasar performa dari tes, 3) keberhasilan model IRT menghasilkan metode estimasi skor untuk individu pada trait, 4) trait diestimasi (disimpulkan) dari performa individu dalam tes yang bersifat *observable*. Keuntungan penerapan IRT: 1) Asumsi sejumlah besar item soal mengukur trait sama, estimasi kemampuan individu independen dari sampel tertentu item tes; 2) Asumsi individu populasi besar, deskripsi tes (misalnya, tingkat kesukaran dan daya pembeda) independen bagi sampel individu tertentu untuk tujuan kalibrasi soal; 3) Statistik menunjukkan secara presisi kemampuan masing-masing individu yang diestimasi.

Ada empat model IRT (DeMars, 2010; Depdiknas, 2008; Hambleton & Swaminathan, 1985; Retnawati, 2016) . 1) Model satu parameter (Model Rasch), yaitu proses menganalisis data menitikberatkan parameter tingkat kesukaran soal. 2) Model dua parameter, yaitu proses menganalisis data menitikberatkan pada parameter tingkat kesukaran soal dan daya pembeda soal. 3) Model tiga parameter, yaitu proses menganalisis data menitikberatkan pada parameter tingkat kesukaran soal, daya pembeda soal, dan menebak (*guessing*). 4) Model empat parameter, yaitu proses menganalisis data menitikberatkan pada parameter tingkat kesukaran soal, daya pembeda soal, menebak (*guessing*), dan penyebab lain.

Model IRT yang diterapkan pada penelitian ini adalah model satu parameter (model *Rasch*). Tujuan utama model Rasch adalah membuat skala pengukuran dengan interval yang sama. Karena skor mentah (*raw score*) tidak memiliki sifat

keintervalan, maka skor mentah tidak digunakan langsung untuk menafsirkan kemampuan orang (Sumintono & Widhiarso, 2015). Skor mentah (*raw score*) dikonversi pada mistar *logarithm odd unit* (logit) untuk person atau butir soal.

Kelebihan model Rasch adalah alat analisis yang sangat berguna untuk menguji validitas dan reabilitas instrumen, peserta tes/ individu/ *person* dan item secara sekaligus. Kelebihan lain dari model ini, memenuhi lima prinsip model pengukuran (Sumintono & Widhiarso, 2013). Kelima prinsip tersebut yaitu: 1) memberikan skala linier dengan interval yang sama; 2) dapat melakukan prediksi terhadap data yang hilang (omit); 3) memberikan estimasi (perkiraan) yang lebih tepat; 4) mampu mendeteksi ketidaktepatan model; 5) pengukuran bersifat *replicable*.

Keterbatasan model Rasch, yaitu: 1) aplikasi model memerlukan pengetahuan/ kemampuan pengguna tentang dasar matematika dan membutuhkan tingkat pemahaman lebih tinggi tentang perangkat lunak (Bond & Fox, 2015). 2) Model Rasch memerlukan sejumlah besar pengamatan/ulangan/replikasi untuk memperkirakan parameter model. 3) Model Rasch tidak memiliki parameter menebak (*guessing*).

Pendekatan IRT diterapkan untuk mengetahui karakteristik instrumen tes. Uji asumsi yang mendasari IRT (Hambleton & Swaminathan, 1985; Hambleton dkk., 1991) unidimensi. Butir tes bersifat unidimensi berarti butir tersebut mengukur satu kemampuan. Asumsi unidimensi terpenuhi apabila ada komponen tunggal yang dianggap paling dominan dalam menentukan performa individu. Uji asumsi tersebut dapat dianalisis dengan analisis faktor eksploratori.

Tahap selanjutnya, pengujian kesesuaian butir soal dengan model Rasch (*item infit*) dan kesesuaian person/individu dengan model Rasch (*person infit*). Adapun tahapan menguji *item fit* dan *person fit* antara lain: 1) menginterpretasi nilai *outfit means-square* (OMSQ), 2) *outfit z-standard* (OZTD), 3) *point measure correlation* (*Pt Mean Corr*) (Bond & Fox, 2015; Boone, Yale, & Staver, 2014). Jika butir soal tidak memenuhi ketiga kriteria tersebut maka butir soal diperbaiki atau diganti (Sumintono & Widhiarso, 2015). Untuk mengetahui apakah model Rasch dapat memprediksi respons setiap responden maka digunakan statistik *infit mean-square* (IMSQ) dan *outfit mean-square* (OMSQ) (Bond & Fox, 2015). Statistik IMSQ dan

OMSQ merupakan ukuran tingkat kesesuaian antara data observasi dengan nilai yang diprediksi oleh model Rasch. Butir tes dikatakan *fit* model Rasch jika nilai IMSQ dan OMSQ berkisar dari 0,5 sampai 1,5 (Lincare, 2002). Interpretasi nilai IMSQ dan OMSQ dijelaskan pada Tabel 3.9.

Tabel 3.9  
*Rentang Nilai IMSQ dan OMSQ (Lincare, 2002)*

Nilai	Implikasi Pengukuran
>2,0	Merusak sistem pengukuran
1,5 – 2,0	Tidak mempunyai makna bagi pengukuran
0,5 – 1,5	Bermanfaat bagi pengukuran
< 0,5	Tidak bermanfaat bagi pengukuran, tetapi tidak merusak

Interpretasi nilai *outfit z-standard* (OZSTD) yang diterima adalah  $-2,0 < OZSTD < 2,0$ . Interpretasi nilai *point measure correlation* (*Pt Mean Corr*) pada rentang  $\geq 0$ .

Kriteria acuan untuk menentukan kesulitan butir soal dengan model Rash diadaptasi dari Sumintono & Widhiarso (2015). Interpretasi tingkat kesulitan butir soal dideskripsikan pada Tabel 3.10.

Tabel 3.10  
*Kriteria Kesulitan Butir Soal*

No.	Kriteria	Keputusan
1	<i>Measure</i> > +1SD	Sangat sulit
2	0,0-(0,0 logit + 1 SD)	Sulit
3	0,0-(0,0 logit – 1 SD)	Mudah
4	<i>Measure</i> < –1 SD	Sangat mudah

Beberapa tahapan menentukan reliabilitas butiir soal meliputi nilai *person reliability*, nilai *item reliability*, dan nilai *alpha Cronbach*. Interpretasi data untuk nilai *person reliability* dan *item reliability* dijelaskan pada Tabel 3.11 (Sumintono & Widhiarso, 2015).

Tabel 3.11  
*Kategori Nilai Person Reliability dan Item Reliability*

No.	Nilai Person dan Item Reliability	Keterangan
1	> 0,94	Istimewa
2	0,91 - 0,94	Bagus sekali
3	0,8 – 0,90	Bagus
4	0,67 – 0,80	Cukup
5	< 0,67	Buruk

Nilai *alpha Cronbach* untuk mengukur reliabilitas meliputi interaksi antara *person* dan butir-butir soal secara keseluruhan. Interpretasi nilai *alpha Cronbach* lebih spesifik dijelaskan pada Tabel 3.12 (Sumintono & Widhiarso, 2013, 2015).

Ratna Ekawati, 2021

PENGEMBANGAN INSTRUMEN ASESMEN PILIHAN GANDA EMPAT-TINGKAT UNTUK MENDIAGNOSIS KEMAMPUAN MULTI-REPRESENTASI CALON GURU PADA KONSEP MEKANIKA

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

Tabel 3.12  
*Kategori Reliabilitas Alpha Cronbach*

No.	Nilai Alpha Cronbach	Keterangan
1	> 0,8	Bagus sekali
2	0,7-0,8	Bagus
3	0,6-0,7	Cukup
4	0,5-0,6	Jelek
5	< 0,5	Buruk

Apabila salah individu dengan karakteristik tertentu lebih diuntungkan dalam mengerjakan item soal dibanding individu dengan karakteristik lain disebut item bias. Pemodelan Rasch untuk mendeteksi item bias yaitu DIF (*Differential Item Functioning*). Suatu butir soal dikatakan bias jika nilai probabilitas butirnya di bawah 5% (0,05).

Peta konstruk merupakan representasi visual untuk mengetahui secara tepat dimana posisi butir soal dan responden terkait dimensi/pengetahuan yang diukur. Keunggulan peta konstruk antara lain: 1) mendefinisikan secara menyeluruh dan subtanatif pada isi konstruk yang diukur; 2) menggambarkan secara kontinum data-data pengukuran dengan dua bagian penting yaitu: a) bagian peta konstruk responden (*person*) dimana abilitas responden dikelompokkan secara kualitatif dari tinggi ke rendah, b) bagian konstruk item dimana item soal dikelompokkan secara kualitatif dari tingkat kesulitan tertinggi ke terendah (Wilson, 2005).

Peta konstruk pada pemodelan Rasch dikenal dengan istilah Peta *Wright* (*Wright Maps*). Karakteristik peta Wright yaitu menggambarkan person dan item secara komprehensif. Data sebaran logit mahasiswa (*person*) dan logit butir soal dengan sebaran kurva normal. Konstruksi kurval normal yaitu nilai rata-rata (*mean*) ditengah-tengah, nilai satu deviasi standar (1SD) adalah 34% dari jumlah data, rentang nilai dari +1SD dan -1SD sebesar 68%, rentang nilai +2SD dan -2SD sebesar 95%. Jika nilai logit *person* dan item butir di luar dari +2SD sampai -2SD disebut *outline*.