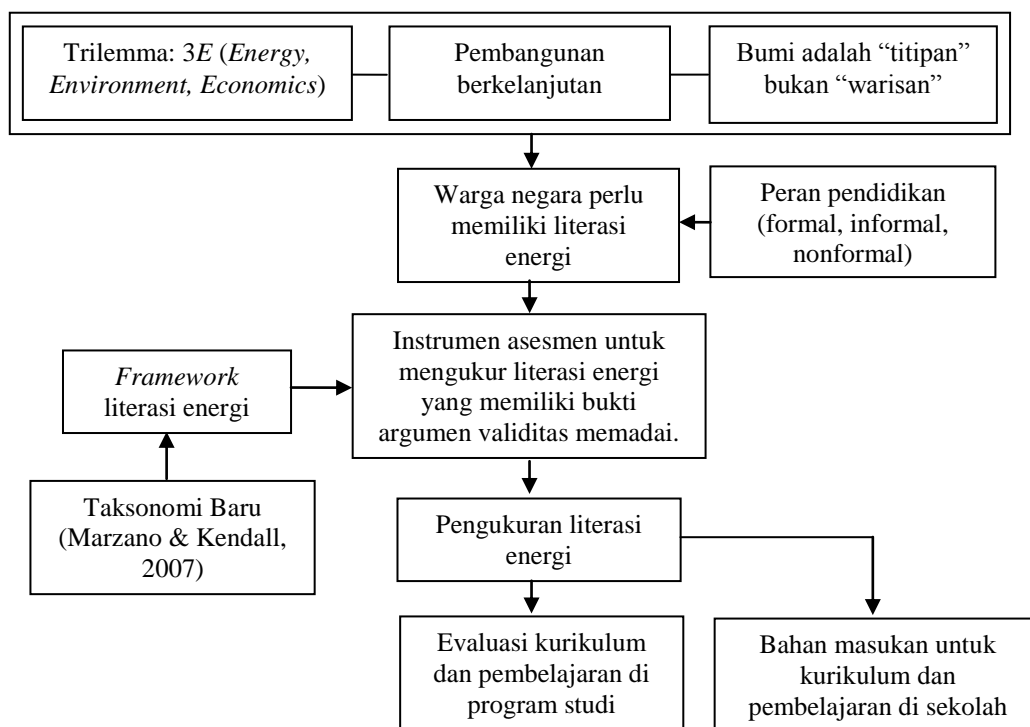


BAB III METODOLOGI PENELITIAN

3.1 Paradigma Penelitian

Pembangunan berkelanjutan akan terwujud jika warga negara memiliki literasi energi. Literasi energi mencakup domain pengetahuan, sikap, dan perilaku. Literasi energi dapat dibekalkan melalui pembelajaran terutama fisika yang terkait dengan konsep energi. Pembelajaran fisika yang dapat menjadikan mahasiswa memiliki literasi energi adalah pembelajaran yang berorientasi pada kompetensi literasi energi. Instrumen asesmen literasi energi disusun berdasarkan kompetensi literasi energi. Gambar 3.1 menampilkan bagan paradigma penelitian ini.



Gambar 3.1 Paradigma penelitian.

3.2 Metode dan Langkah-Langkah Penelitian

Penelitian ini adalah penelitian pengembangan dan validasi *framework* dan instrumen asesmen untuk mengukur literasi energi. Metode yang digunakan adalah penelitian berbasis desain (*design-based research*) (Brown, 1992; Collins,

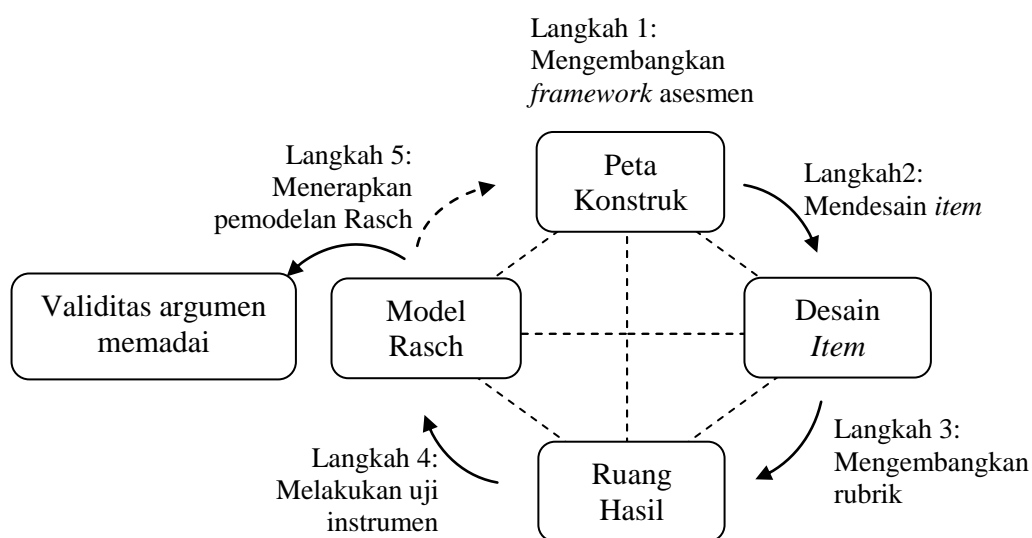
1992). Walaupun penelitian berbasis desain biasa digunakan dalam mengembangkan strategi atau model pembelajaran (misalnya Brown, 1992; Cobb, Confrey, DiSessa, Lehrer, & Schauble, 2003; The Design-Based Research Collective, 2003), metode ini juga dapat digunakan untuk mengembangkan asesmen (misalnya Mohan, Chen, & Anderson, 2009). Hal ini dikarenakan metode penelitian berbasis desain bersifat lebih fleksibel (Collins, Joseph, & Bielaczyc, 2004) dan menggunakan siklus iteratif dari perencanaan yang diikuti dengan implementasi, evaluasi dan modifikasi (The Design-Based Research Collective, 2003).

Dalam implementasinya, penelitian dilakukan melalui lima langkah yang diadaptasi dari Kuo, Wu, Jen, & Hsu (2015). Pertama, mengembangkan *framework* asesmen: menganalisis kompetensi dan komponen literasi energi serta pengorganisasiannya. Kedua, mendesain *item*: mendesain instrumen asesmen untuk mengukur literasi dari seluruh komponen dan tingkat kompleksitas dalam *framework*. Ketiga, mengembangkan rubrik penskoran: mengembangkan ruang hasil dengan basis *framework* asesmen untuk panduan penskoran tiap *item*. Keempat, melakukan uji coba: mengumpulkan bukti validasi untuk mendukung basis teoretis dari konstruk. Kelima, menerapkan model Rasch: menggunakan model pengukuran Rasch untuk menghubungkan data skor dengan komponen literasi energi yang telah ditentukan dalam *framework* asesmen. Langkah penelitian yang dilakukan secara skematis dapat dilihat pada Gambar 3.2.

3.2.1 Mengembangkan *framework* asesmen

Langkah penting dalam pengembangan instrumen pengukuran adalah menentukan dengan jelas apa yang akan diukur (Wu dkk., 2016), biasa disebut dengan konstruk atau *latent trait*. Konstruk sangat penting karena berpengaruh pada validitas konstruk dari instrumen (Baghaei, 2008). Konstruk disusun dengan memberikan deskripsi yang mendetail, disebut dengan *framework* asesmen (Wu dkk., 2016). *Framework* asesmen dikembangkan sebagai panduan untuk langkah-langkah berikutnya.

Framework asesmen literasi energi yang telah dikembangkan pada penelitian sebelumnya (K. Chen dkk., 2013; DeWaters & Powers, 2013) memiliki tiga kelemahan (Yusup dkk., 2017). Kelemahan pertama, *framework* tersebut tidak menggambarkan model atau teori berpikir. Kelemahan kedua, *framework* tersebut tidak memberikan informasi yang jelas mengenai level berpikir yang diperlukan untuk menjawab suatu *item*. Kelemahan ketiga, *framework* tersebut tidak menggambarkan dengan jelas bagaimana mengintegrasikan pengetahuan dan konteks ke dalam *item* asesmen.



Gambar 3.2. Langkah-langkah pengembangan dan validasi instrumen asesmen literasi energi untuk mahasiswa calon guru fisika. Diadaptasi dari “Development and validation of a multimedia-based assessment of scientific inquiry abilities,” oleh Kuo, C.-Y., Wu, H.-K., Jen, T.-H., & Hsu, Y.-S., 2015, *International Journal of Science Education*, 37(14), hlm. 2333.

Taksonomi belajar dapat digunakan sebagai *framework* asesmen untuk mengatasi kelemahan dari *framework* asesmen literasi energi yang telah dikembangkan oleh para peneliti terdahulu. Beberapa taksonomi belajar yang sudah dikenal luas yakni Taksonomi Bloom (Bloom dkk., 1956) atau Revisi Taksonomi Bloom (Anderson dkk., 2001), Taksonomi Domain Afektif Krathwohl (Krathwohl dkk., 1964), Taksonomi SOLO (Biggs & Collis, 1982), dan Taksonomi Baru (Marzano & Kendall, 2007). Untuk mendapatkan *framework* yang tepat untuk IALE dilakukan pengkajian terhadap taksonomi belajar tersebut.

Tiga kriteria ditetapkan untuk menyeleksi taksonomi belajar yang sesuai (Yusup dkk., 2017). Kriteria pertama, taksonomi belajar tersebut mencakup domain kognitif, afektif, perilaku dan pengetahuan dalam satu model. Kriteria kedua, taksonomi belajar tersebut secara jelas membedakan antara proses berpikir dan pengetahuan. Kriteria ketiga, taksonomi belajar tersebut dapat memprediksi perilaku terkait energi.

3.2.2 Mendesain *item*

Hal terpenting pada tahap ini adalah *item* yang dikembangkan mengikuti *framework* asesmen yang telah disusun sebelumnya. *Item* dikembangkan dalam format asesmen tertulis (*paper-and-pen assessment*) berbentuk esai (*constructed response*) dan pilihan ganda. Aspek-aspek yang menjadi pertimbangan dalam mendesain *item* meliputi aspek konteks, sensitivitas, dan teknis.

Aspek konteks mempertimbangkan kriteria-kriteria bahwa *item* yang dikembangkan harus (a) relevan dengan kehidupan sehari-hari mahasiswa S-1 (usia 18-25 tahun); (b) konteksnya autentik; (c) mencerminkan aspek yang penting dari warga negara yang memiliki literasi energi; (d) mencakup konten yang seharusnya dikuasai mahasiswa calon guru fisika; (e) sesuai dengan kompetensi yang dirumuskan dalam *framework*.

Aspek sensitivitas mempertimbangkan kriteria-kriteria bahwa *item* yang dikembangkan harus (a) dapat digunakan secara nasional di Indonesia (misalnya bebas dari konteks budaya dan pengetahuan kelompok budaya tertentu); dan (b) tidak bias gender. Sementara, kriteria-kriteria yang dipertimbangkan pada aspek teknis adalah (a) asesmen dapat dilakukan di kelas; (b) penskorannya dan interpretasi hasilnya mudah; dan (c) tes dapat dijawab dalam waktu paling lama 120 menit, sehingga tidak menimbulkan kebosanan pada *testee* yang dapat berakibat bias pada hasil asesmen.

3.2.3 Mengembangkan rubrik penskoran

Setelah merancang *item*, langkah berikutnya adalah mengembangkan rubrik penskoran. Pengembangan rubrik penskoran terkait erat dengan dua pilar pendekatan pemodelan konstruk: desain *item* dan ruang hasil (Gambar 3.2).

Ruang hasil terdiri dari seperangkat kategori kualitatif yang berbeda untuk mengidentifikasi, mengevaluasi, dan menskor bukti kinerja jawaban mahasiswa (Wilson, 2005).

Analisis Rasch menginginkan agar data sesuai dengan model. Karena model yang dihasilkan dari penelitian ini akan digunakan sebagai basis untuk menarik inferensi, maka model tersebut tidak cukup dengan didasarkan pada kesesuaian data semata, tetapi juga pada bagaimana proses data itu dihasilkan (McCullagh, 1980), yakni model observasi (Wright & Linacre, 1992). Di sinilah pentingnya rubrik karena dalam penelitian ini rubrik menjadi acuan dalam mengklasifikasi kategori jawaban sebagai model observasi.

Penskoran yang digunakan dalam rubrik penelitian ini adalah *partial credit model* (PCM) (Masters, 1982, 2010). Rubrik yang dikembangkan mengikuti skema yang disarankan oleh Marzano dan Kendall (2008). Sebagai contoh, kompetensi yang diukur untuk pertanyaan 1 pada topik bohlam (*Item* BL1) adalah “Mengidentifikasi kesalahan logis dari informasi yang disajikan.” Kompetensi tersebut mencerminkan kemampuan pada level analisis. Skor maksimum (skor 4) diberikan untuk jawaban yang memenuhi indikator tersebut secara lengkap/sepurna. Skor 3 diberikan untuk jawaban pada level analisis namun masih ada kesalahan minor. Skor 2 diberikan untuk jawaban yang memenuhi satu level di bawah kemampuan pada skor 4, yakni mengintegrasikan yang berada pada level 2, pemahaman. Skor 1 diberikan untuk jawaban yang berada pada kemampuan satu level di bawah skor 2, yakni level 1, retrieval, atau jawaban pada skor 2 namun masih ada kesalahan. Dengan demikian, skor 0 diberikan untuk jawaban yang tidak relevan dengan kemampuan yang digambarkan pada skor di atas. Jawaban kosong (tidak ada jawaban) tidak diberi skor dan dianggap data hilang.

Skor maksimum untuk setiap *item* tidaklah sama. Besarnya skor maksimum untuk suatu *item* dipertimbangkan berdasarkan pada daya diskriminasi *item* sebagai hasil perkiraan. Memang, memperkirakan daya diskriminasi *item* lebih sulit daripada memperkirakan tingkat kesulitan *item*. Dua hal menjadi

pertimbangan untuk menentukan skor maksimum suatu *item* (Wu dkk., 2016). Pertama, seberapa relevan hubungan *item* dengan konstruk yang akan diukur. Semakin relevan *item* terhadap konstruk, semakin tinggi skor maksimum yang diberikan. Kedua, seberapa banyak "informasi" mengenai kemampuan *testee* yang dapat diperoleh dari *item*. Informasi itu dapat digunakan untuk memisahkan mahasiswa menjadi beberapa kelompok kemampuan. Jika respons mahasiswa pada suatu *item* dengan jelas dapat dikategorikan menjadi beberapa kelompok kemampuan, maka lebih banyak poin skor yang diberikan.

Setelah skor maksimum suatu *item* dalam PCM ditentukan, berarti kategori skor dalam suatu *item* mencerminkan peningkatan tingkat kemampuan *testee*. Skor maksimum *item* berkaitan dengan diskriminasi, sementara kategori skor dalam suatu *item* berhubungan dengan tingkat kesulitan (Wu dkk., 2016). Karena skor maksimum dari *item* yang disusun dalam rubrik merupakan taksiran, maka kesesuaian respons/jawaban *item* terhadap model perlu diuji melalui statistik kesesuaian. Kesesuaian skor maksimum suatu *item* tercermin dari terpenuhinya kriteria statistik kesesuaian *item* yang bersangkutan.

3.2.4 Uji coba

Sebelum dilakukan *pilot test*, *framework*, *item*, dan rubrik penskoran yang telah dirancang selanjutnya diajukan ke ahli untuk dinilai (*expert judgement*). Validasi oleh ahli terhadap rancangan awal IALE pada penelitian ini dilakukan untuk melihat kualitas *item* (bahasa sederhana/tidak rumit, jelas), kesesuaian isi konstruk yang diukur, dan keselarasan antara *item* yang dikembangkan dengan konstruk. Tiga orang ahli yang melakukan penilaian adalah dosen bergelar doktor yang berasal dari tiga perguruan tinggi negeri di Indonesia yakni dari Universitas Pendidikan Indonesia, Universitas Pendidikan Ganesha, dan Universitas Negeri Medan. Mereka memiliki pengalaman dalam riset di bidang asesmen pendidikan IPA.

Dua dokumen terpisah tapi saling terkait diberikan kepada para ahli. Dokumen pertama adalah rancangan *framework* asesmen, dan dokumen kedua adalah rancangan *item* asesmen dan rubrik penskorannya. Lembar validasi ahli

juga disertakan sebagai lembar penilaian terhadap rancangan yang diajukan. Lembar validasi ahli berisi empat bagian pokok. Bagian pertama adalah penjelasan tentang tujuan penelitian. Bagian kedua berisi petunjuk bagaimana mengisi lembar validasi. Bagian ketiga berisi pernyataan yang disediakan jawaban “Ya” dan “Tidak,” disertai komentar. Bagian “Ya” dan “Tidak” terdiri dari dua aspek, yakni *framework* (lima pernyataan), dan untuk penilaian *item* dan rubrik penskoran (12 pernyataan). Pada bagian akhir ahli diminta untuk memberikan komentar secara umum, yang dapat diisi berupa saran-saran perbaikan.

Kesepakatan ahli dalam menilai rancangan *framework* dan perangkat IALE dinyatakan dalam *Fleiss kappa*. *Fleiss kappa* digunakan karena ahli yang memberikan penilaian sebanyak tiga orang. Secara matematik *Fleiss kappa* dirumuskan sebagai $\kappa = \frac{P - P_e}{1 - P_e}$. Jika semua ahli sepakat secara sempurna maka nilai $\kappa = 1$. Sebaliknya, jika tidak ada kesepakatan antar ahli maka nilai $\kappa = 0$.

Setelah *item* dan rubrik penskoran diperbaiki berdasarkan saran dari para ahli, langkah selanjutnya adalah melakukan *pilot test*. Tujuan dari *pilot test* adalah (1) mengetahui waktu yang diperlukan untuk mengerjakan seluruh *item* asesmen, (2) menguji kejelasan pertanyaan dalam *item*, (3) menguji apakah kata-kata atau kalimat-kalimat yang digunakan terlalu berlebihan (*window dressing/verbiage*), (4) menguji apakah pilihan kata negatif dinyatakan secara jelas. Jika terdapat kalimat atau kata negatif, seperti “Tidak” atau “Bukan”, apakah kata tersebut digunakan dengan tepat dan ditulis dengan menonjolkan kata tersebut, misalnya dengan huruf kapital dan dicetak tebal.

Tahap pertama *pilot test*, partisipan diberi tes dalam bentuk tertulis. Setelah partisipan menyelesaikan tes, pada tahap kedua *pilot test* dilakukan wawancara. Wawancara dilakukan terhadap lima orang partisipan setelah mereka selesai mengerjakan *item-item* IALE. Wawancara terhadap kelima orang tersebut dilakukan per individu. Pada saat wawancara, terwawancara diberi lembar jawaban yang telah mereka kerjakan. Pertanyaan diberikan untuk menggali bagaimana proses berpikir mereka ketika menjawab masing-masing *item* sehingga dapat menghasilkan jawaban seperti tertera dalam lembar jawaban tersebut.

Pada setiap pengujian instrumen (*pilot test* dan uji lapangan), *testee* mendapatkan satu set soal dalam bentuk buklet. Bentuk buklet ini membuat pertanyaan-pertanyaan yang merujuk pada tabel tertentu yang mungkin berada di halaman sebelumnya, akan tampak dalam satu muka, sehingga *testee* tidak perlu membolak-balikkan halaman. Gambar 3.3 menampilkan contoh halaman 1 dan 2 yang berisi pertanyaan 1 sampai pertanyaan 3 di mana ketiga pertanyaan tersebut merujuk pada satu tabel yang berada di halaman 1.

Lembar jawaban yang digunakan adalah kertas bergaris *double folio*. Kertas ini digunakan agar *testee* dapat memberikan jawaban yang panjangnya dapat ditentukan sendiri. Berbeda jika lembar jawaban khusus disediakan dengan menyediakan kolom/kotak untuk ruang menuliskan jawaban. Luas kolom/kotak tersebut seolah memberi kesan seberapa panjang jawaban yang diinginkan oleh pembuat tes.

BOHLAM

Gambar 1 menunjukkan tiga jenis bohlam yang sering kita temui. Gambar (a) adalah bohlam jenis incandescence light bulb (ILB) atau bohlam pijar. Gambar (b) adalah salah satu varian bohlam jenis compact fluorescent lamp (CFL). Gambar (c) adalah bohlam jenis light emitting diode (LED).

Untuk Pertanyaan 1 Sampai dengan Pertanyaan 5.
Tabel 1 menunjukkan efisiensi daya listrik untuk bohlam jenis ILB, CFL, dan LED. Luaran cahaya merupakan tingkat kecerahan (*brightness*) sebuah bohlam, semakin besar luaran cahayanya maka semakin cerah (terang) bohlam tersebut.

Tabel 1. Efisiensi daya listrik bohlam jenis ILB, CFL, dan LED

Luaran cahaya (lumens)	Konsumsi daya listrik (watt)	ILB	CFL	LED
200	25	3-5	3	
450	40	8-11	5-8	
800	60	13-15	8-12	
1.100	75	18-20	12-16	
1.600	100	24-28	18-22	
2.400	150	30-52	20	
3.100	200	40-75	22	
4.000	300	70-100	40, 1	

Sumber: https://www.wiki.jedta.org/wiki/111/111.com_jenis%2F%20bohlam

Pertanyaan 1. Bohlam
Berdasarkan Tabel 1, Siswa A mengatakan bohlam pada luaran cahaya yang sama, bohlam ILB menghasilkan cahaya yang paling cerah (terang) diantara ketiga jenis bohlam. Sementara Siswa B mengatakan bohlam bohlam LED yang menghasilkan cahaya paling cerah (terang).
Identifikasi pendapat mana (jika keduanya) yang salah, dan jelaskan kesalahan (jika ada) dari pendapat tersebut. Jika tidak ada pendapat yang salah, jelaskan mengapa kedua pendapat tersebut benar.

Pertanyaan 2. Bohlam
Berdasarkan Tabel 1, buktikan kesimpulan terkait efisiensi energi dari ketiga jenis bohlam tersebut.

Pertanyaan 3. Bohlam
Kerapatan daya, masa hidup, dan harga bohlam ILB, CFL, dan LED ditunjukkan pada Tabel 2.

Tabel 2. Masa hidup dan harga bohlam ILB, CFL, dan LED.

Jenis Bohlam	Kerapatan Daya (watt)	Masa hidup (jam)	Harga (Rp)
ILB	60	1.000	8.000
CFL	14	8.000	28.000
LED	30	25.000	88.000

Jika Anda memiliki uang Rp 90.000 untuk membeli bohlam, bohlam jenis mana yang akan Anda beli?
Seberapa dalam yang menjadi pertimbangan pilihan Anda tersebut?

Gambar 3.3 Tampilan buklet IALE. Pertanyaan 1 sampai 3 berada di halaman 1 dan 2 dapat dilihat dalam satu muka.

Pada saat pelaksanaan tes, *testee* diposisikan dalam tempat duduk yang diatur sedemikian rupa sehingga mengurangi kemungkinan *testee* untuk dapat berkomunikasi dengan teman yang ada di sebelah kanan-kiri ataupun depan-

belakangnya. Pengawasan di ruang tes hanya dilakukan oleh satu orang dan tidak dilakukan secara ketat. Hal ini dilakukan agar *testee* merasa nyaman dan tidak cemas.

3.2.5 Menerapkan model Rasch

Penskoran hasil tes dilakukan oleh satu orang yakni peneliti sendiri. Skor mentah yang diperoleh oleh masing-masing *testee* pada setiap *item* dimasukkan ke program pengolah data *MS Excel* untuk kemudian diimpor oleh program *Winsteps*. Hasil analisis secara kualitatif dan kuantitatif data *pilot test* menjadi dasar untuk memperbaiki *item-item* asesmen yang diuji pada tahap berikutnya.

Data yang diperoleh dari pengujian instrumen dianalisis menggunakan model Rasch. Semua analisis Rasch dalam penelitian ini dilakukan menggunakan perangkat lunak *Winsteps* versi 3.92.1 (Linacre, 2016). Karena skor jawaban *item* politomi dan juga skor maksimum di antara *item* tidak sama maka analisis Rasch yang digunakan adalah PCM dengan persamaan:

$$\ln\left(\frac{P_{nik}}{1-P_{nik}}\right) = \theta_n - \delta_{ik} \quad (5)$$

dimana P_{nik} adalah probabilitas *testee* n dengan kemampuan θ_n menjawab pada level k dari *item* i secara benar, dan δ_{ik} adalah kesulitan pada level k dari *item* i .

Pengembangan instrumen menggunakan analisis Rasch merupakan proses iterasi. Revisi terhadap *item* ataupun rubrik dilakukan berdasarkan hasil analisis Rasch, meliputi statistik kesesuaian *item*, struktur kategori *item*, *differential item functioning* (DIF), peta person-*item* (peta Wright), dan dimensionalitas. Kesesuaian *item* dengan model ditentukan dengan melihat *mean square residual* (*Mnsq*) dan *standardized mean square residual* (*Zstd*). *Mnsq* dan *Zstd* didasarkan pada perbedaan antara apa yang teramati dan apa yang diharapkan oleh model Rasch. *Mnsq* merupakan kuadrat residual, sementara *Zstd* adalah skor t ternormalisasi dari residual. Terdapat dua *Mnsq* dan *Zstd* atas semua person untuk masing-masing *item*. Dengan demikian akan dihasilkan empat statistik kesesuaian, statistik *infit* (*infit mnsq* dan *infit zstd*) dan statistik *outfit* (*outfit mnsq* dan *outfit Zstd*).

Perspektif statistik Rasch memandang person dan *item* merupakan parameter yang sama. Dengan demikian kriteria *fit* keduanya juga sama. Dari perspektif substantif, person dan *item* berbeda. Untuk menentukan statistik kesesuaian *item*, dilakukan langkah-langkah berikut.

- 1) Memeriksa apakah terdapat *point-measure correlations* yang nilainya negatif. Jika ada yang negatif, beberapa kemungkinan menjadi penyebabnya.
 - a. Kesalahan pada kunci jawaban (rubrik), hal ini bisa ditelusuri dari tabel 10.3 keluaran *Winsteps*. Jika kesalahan ini terjadi maka rubrik perlu diperbaiki.
 - b. Kesalahan saat memasukkan data ke dalam *MS Excel*.
- 2) Memeriksa *outfit* lebih dulu sebelum melihat *infit*.
- 3) Memeriksa *mean square (mnsq)* lebih dulu sebelum melihat *z standardized (Zstd)*.
- 4) Memeriksa *underfit* lebih dulu sebelum melihat *overfit* atau negatif, karena *underfit* lebih mengancam validitas daripada *overfit*.

Dalam penelitian ini, rentang nilai 0,70 – 1,30 untuk *mnsq outfit* dan *mnsq infit* ditetapkan sebagai kriteria sebuah *item* dikatakan sesuai model Rasch (*item fit*). Jika suatu *item* tidak sesuai model (*misfit*), maka *item* tersebut dipertimbangkan untuk diperbaiki atau dibuang. Keputusan akhir apakah suatu *item* diperbaiki atau dibuang diambil dengan mempertimbangkan keseluruhan hasil analisis Rasch sebagaimana telah disebutkan.

3.3 Partisipan

Selama pengembangan instrumen, penelitian ini melibatkan mahasiswa calon guru fisika dari tingkat (semester) dan perguruan tinggi yang berbeda. Mahasiswa calon guru fisika yang menjadi partisipan berasal dari tiga perguruan tinggi negeri (PTN): satu dari Sumatera Selatan, dan dua dari Jawa Barat; dari tahun pertama hingga tahun ketiga kuliah. Pemilihan subjek didasari pertimbangan perbedaan geografi dan kelompok budaya. Partisipan yang terlibat dalam setiap tahap adalah subjek yang berbeda. Partisipan dalam penelitian ini diringkas dalam Tabel 3.1. Keterlibatan partisipan dalam penelitian ini bersifat

sukarela, artinya mereka tidak memperoleh kompensasi dalam bentuk apapun atas keterlibatan mereka dalam penelitian ini.

Tabel 3.1
Partisipan pada Setiap Tahapan Uji Coba Instrumen

Tahapan	Jumlah Awal	Jumlah akhir	Institusi
<i>Pilot test</i>	35	30 (5) ^a	A
Uji lapangan 1	65	62	A
Uji lapangan 2	185	123	A, B, C

Catatan. Jumlah awal adalah jumlah partisipan yang mengikuti tes. Jumlah akhir adalah jumlah partisipan yang dianalisis datanya. Jumlah ini berkurang dari jumlah partisipan yang mengikuti tes karena jawaban partisipan kurang dari 50%. Institusi A : PTN di Sumatera Selatan. Institusi B dan C : PTN di Jawa Barat.

^aAngka di dalam tanda kurung menunjukkan jumlah partisipan yang diwawancarai.

Pengambilan sampel untuk *pilot test* tidak harus menggunakan prosedur yang ketat (Wu dkk., 2016). Pada tahap *pilot test*, partisipan yang terlibat sebanyak 35 orang mahasiswa dari salah satu perguruan tinggi negeri (PTN) di Sumatera Selatan. Dalam pemodelan Rasch, *pilot test* dengan jumlah sampel sebanyak 35 orang (berada dalam kisaran 16-36) sudah layak untuk mendapatkan hasil estimasi yang stabil dengan kisaran pada ± 1 *logit* dan tingkat kepercayaan 95% (Linacre, 1994). Namun, dari 35 partisipan yang mengerjakan *item* IALE, hanya jawaban dari 30 partisipan digunakan sebagai data. Sebanyak lima orang partisipan tidak dapat digunakan datanya karena *item* yang dijawab kurang dari 50%. Kriteria minimal 50% jawaban dari keseluruhan *item* yang digunakan sebagai data untuk dianalisis juga diterapkan pada uji lapangan 1 dan uji lapangan 2.

3.4 Analisis Data

Untuk menjawab pertanyaan penelitian kedua, analisis data dilakukan dengan menggunakan pemodelan Rasch *partial credit model*. Analisis data dilakukan dengan bantuan perangkat lunak *Winsteps version 3.92.1* (Linacre, 2016). Hasil pengoperasian *Winsteps* digunakan untuk memberikan bukti-bukti validitas yang sesuai. Bukti-bukti tersebut diantaranya kesesuaian kategori skor (*category fit*), dimensionalitas, kesesuaian *item* (*item fit*) dengan model, reliabilitas dan separasi untuk *item* dan person, peta Wright, dan *differential item functioning* (DIF).

Analisis data *pilot test* dan uji lapangan dilakukan dengan mengevaluasi bukti empiris dan teoretis untuk mengambil tindakan berikutnya, apakah *item* itu dibuang atau diperbaiki. Hasil analisis data *pilot test* menjadi dasar tindakan terhadap *item-item* yang problematik untuk selanjutnya diujikan pada uji lapangan 1. Demikian pula, hasil analisis data uji lapangan 1 menjadi dasar tindakan terhadap *item-item* problematik untuk uji lapangan 2.

Tabel 3.2
Inferensi dan Asumsi sebagai Argumen Interpretif

Inferensi	Asumsi
1. Inferensi penskoran	1.1. Kategori skor pada <i>item</i> bersifat monotonik.
2. Inferensi generalisasi	2.1. Indeks reliabilitas dan separasi person berada pada rentang yang dapat diterima. 2.2. Indeks reliabilitas dan separasi <i>item</i> berada pada rentang yang dapat diterima.
3. Inferensi eksplanasi	3.1. <i>Item</i> sesuai model Rasch. 3.2. Hierarki tingkat kesulitan <i>item</i> pada peta Wright sesuai. 3.3. <i>Principal component analysis residual</i> (PCAR)
4. Inferensi ekstrapolasi	4.1. <i>Item</i> dapat menarget kemampuan person dengan baik. 4.2. <i>Item</i> tidak mengalami DIF.

Untuk memberikan argumen validitas maka data hasil uji coba IALE dianalisis untuk memberikan bukti inferensi penskoran, generalisasi, eksplanasi, dan ekstrapolasi. Secara operasional, argumen interpretif yang dievaluasi dalam argumen validitas terdiri dari empat inferensi yang masing-masing memiliki asumsi. Keempat inferensi dan asumsi dalam argumen interpretif ditunjukkan pada Tabel 3.2.

3.5 Menentukan Level Kompetensi Literasi Energi

Data yang dikumpulkan menggunakan sejumlah *item* IALE berupa *partial credit*. *Testee* mendapat skor mentah dengan rentang tertinggi dari 0 sampai 3 (0,1,2,3) pada satu *item*. Skor yang diperoleh menggambarkan “level” kompetensi mereka. Pengkategorian kompetensi ke dalam level tertentu dilakukan dengan menentukan batas antar kategori. Batas kategori yang dapat ditentukan dari data yang dikumpulkan menggunakan IALE adalah batas antara kategori 1 dan

kategori 2, serta antara kategori 2 dan kategori 3. Dengan kata lain, kompetensi yang diukur menggunakan IALE dikategorikan dalam tiga level literasi energi.

Nilai batas antar kategori diperoleh dengan menghitung median dari probabilitas kumulatif 50% (Rasch-Thurstone *threshold*). Rasch-Thurstone *threshold* memisahkan skala pengukuran pada masing-masing batas kategori menjadi 50% kemungkinan teramati di bawah kategori dan probabilitas 50% teramati pada atau di atas kategori. Nilai batas antar kategori kemudian diplot dalam peta Wright, sehingga diperoleh gambaran kompetensi pada masing-masing level/kategori.