

Aplikasi *Rasch Model*: Pengembangan *Fluids Assessment (FAss)* Berdasarkan *Taxonomy of Introductory Physics Problems (TIPP)*

D N Angraeni, A Suherman, Y Guntara*

Program Studi Pendidikan Fisika, Universitas Sultan Ageng Tirtayasa

*E-mail: guntaray@untirta.ac.id

Received: 20 April 2020, Accepted: 14 Mei 2020, Published: 30 September 2020

Abstrak. Tujuan dari penelitian ini untuk menghasilkan produk *Fluids Assessment (FAss)* berdasarkan *Taxonomy of Introductory Physics Problems (TIPP)* yang layak pada materi fluida dinamis. Analisis validitas menggunakan pendekatan *Item Response Theory (IRT)* dengan satu Parameter Logistik (PL) yakni parameter tingkat kesukaran atau lebih dikenal dengan *Rasch Model*. Selain itu, produk yang dihasilkan dianalisis dengan pendekatan teori klasik untuk melihat daya beda dan efektivitas distraktor. Penelitian ini merupakan *Research and Development* dengan mengadaptasi model pengembangan Borg and Gall. Hasil penelitian menunjukkan bahwa 34 butir soal dinyatakan valid menurut para ahli dengan analisis *Content Validity Ratio (CVR)*. Sedangkan hasil analisis berdasarkan *Rasch Model* menunjukkan 17 butir soal valid dan reliabilitas soal diperoleh sebesar 0,90 dengan kategori baik. Dapat disimpulkan bahwa FAss memiliki tingkat kesukaran relatif sedang, daya beda yang baik, dan pengecoh yang berfungsi. Hasil implementasi menunjukkan bahwa sebagian besar siswa menguasai setiap level TIPP kecuali level 3a, 4d dan 4c yang berkaitan dengan *matching*, *investigating*, dan *experimenting*.

Kata kunci: *Fluids assessment, TIPP, rasch model*

Abstract. The purpose of this study was to produce a *Fluids Assessment (FAss)* product based on an appropriate *Taxonomy of Introductory Physics Problems (TIPP)* on dynamic fluid material. Validity analysis uses the *Item Response Theory (IRT)* approach with a *Parameter Logistic (PL)*, which is a difficulty level parameter or better known as the *Rasch Model*. Also, the resulting product was analyzed with a classical theoretical approach to see the difference in distinguishing power and effectiveness of the distractor. This research is a *Research and Development* by adapting the development model of Borg and Gall. The results showed that 34 items were declared valid according to experts with *Content Validity Ratio (CVR)* analysis. While the analysis based on the *Rasch Model* shows that 17 items are valid and the reliability of the questions is 0.90 in the good category, it can conclude that the FAss has a relatively moderate level of difficulty, a good difference in power, and a deception that works. The results of the implementation showed that most students mastered every level of TIPP except level 3a, 4d, and 4c relating to *matching*, *investigating*, and *experimenting*.

Keywords: *Fluids assessment, TIPP, rasch model*

1. Pendahuluan

Keberhasilan suatu proses pembelajaran dapat diketahui melalui suatu asesmen (penilaian). Asesmen juga dapat memberikan siswa informasi tentang kemajuan belajarnya sehingga siswa dapat memperbaiki perilaku belajarnya [1]. Dalam aplikasinya, asesmen memerlukan suatu instrumen. Instrumen penilaian harus berkualitas baik supaya mendapatkan informasi yang tepat. Berkaitan

dengan hal tersebut, pemerintah telah memberikan pedoman mengenai Standar Kualifikasi Akademik dan Kompetensi Guru mata pelajaran, yang menjelaskan bahwa salah satu kompetensi guru mata pelajaran adalah mengembangkan instrumen penilaian [2]. Mengingat pentingnya instrumen penilaian, maka seorang guru dituntut untuk dapat mengembangkan instrumen penilaian yang dapat mengukur kemampuan peserta didik, terutama dalam konteks pemecahan masalah. Diwujudkan dalam kurikulum 2013 sesuai dengan pernyataan Yani [3] yang menjelaskan bahwa gambaran Indonesia yang ingin dibentuk oleh kurikulum 2013 adalah manusia yang memiliki kemampuan berpikir kritis dan menalar terutama dalam konteks pemecahan masalah.

Kemampuan dalam memecahkan masalah salah satunya dipengaruhi oleh kemampuan proses kognitif. Untuk melatih kemampuan tersebut, dibutuhkan instrumen penilaian berupa soal yang dapat memfasilitasi dimensi pengetahuan dan proses kognitif dalam sebuah taksonomi [4]. Menggunakan taksonomi berpikir dapat memberikan gambaran sejauh mana level berpikir yang timbul dari soal-soal tersebut serta domain pengetahuan yang mungkin terlibat [5]. Pemilihan kata kerja operasional yang terdapat dalam taksonomi memegang peranan penting dalam indikator pencapaian agar konsep materi tersampaikan secara efektif. Pemilihan kata kerja operasional dapat disesuaikan dengan karakteristik mata pelajaran, kompetensi dasar, dan kedalaman materi. Berkaitan dengan hal ini, terdapat taksonomi yang berisi kata kerja operasional yang telah disesuaikan dengan karakteristik mata pelajaran fisika yakni *Taxonomy of Introductory Physics Problems* (TIPP). Berdasarkan hasil wawancara yang telah dilakukan kepada guru Fisika kelas XI di salah satu SMAN Kota Serang, diketahui bahwa taksonomi yang digunakan guru dalam menyusun ulangan harian dan ujian akhir semester adalah taksonomi Bloom Revisi. Beberapa guru kesulitan untuk memilih kata kerja operasional yang sesuai dengan materi yang ada pada mata pelajaran fisika. Selain itu, guru hanya menggunakan instrumen soal sebagai ulangan harian, dan ujian akhir semester untuk mengukur hasil belajar siswa.

Taxonomy of Introductory Physics Problems (TIPP) didesain khusus untuk mengklasifikasikan masalah-masalah fisika. TIPP di dalamnya terdapat 4 level berpikir sistem kognitif yang dapat melingkupi kemampuan berpikir kritis dalam memecahkan masalah yaitu: *Retrieval* (menimbulkan kembali), *Comprehension* (pemahaman), *Analysis* (analisis), *Knowledge Utilization* (penggunaan pengetahuan) [6]. Hal ini menunjukkan bahwa TIPP dapat melatih kemampuan pemecahan masalah yang tersusun secara hirarki dari hal yang sederhana (mengingat) hingga ke hal yang rumit. Penelitian yang dilakukan oleh Sutiadi & Kurniawati [5] mengenai analisis butir soal Ujian Nasional bidang Fisika berdasarkan TIPP, menunjukkan bahwa hanya 5 tipe soal berdasarkan TIPP dari 25 tipe soal yang terdapat pada buku-buku teks fisika. Oleh karena itu, dibutuhkan *Fluids Assessment* berupa instrumen soal berbasis TIPP.

Instrumen yang baik harus memenuhi beberapa parameter. Menurut Sumintono & Widhiarso [7] menjelaskan “Kesempatan untuk dapat menyelesaikan satu soal dengan benar bergantung pada rasio antara kemampuan orang dan tingkat kesulitan soal”. Artinya, jumlah jawaban benar siswa tidak bisa dijadikan tolak ukur pemahaman siswa, karena tidak sama antara siswa yang dapat mengerjakan soal sulit dengan siswa yang hanya dapat mengerjakan soal mudah. Karena peluang untuk dapat menjawab soal dengan benar dipengaruhi oleh kemampuan dan tingkat kesulitan soal, maka dibutuhkan pendekatan yang melibatkan satu Parameter Logistik (PL) berupa tingkat kesulitan butir. *Rasch Model* hanya melibatkan 1 PL yaitu menekankan pada tingkat kesulitan [7]. Oleh karena itu, dibutuhkan *Rasch Model* untuk menganalisis FAss agar menghasilkan informasi yang lebih tepat. Berdasarkan pertimbangan sebelumnya, maka dikembangkan *Fluids Assessment* (FAss) berdasarkan *Taxonomy of Introductory Physics Problems* (TIPP) dengan menggunakan analisis *Rasch Model*. Tujuan pengembangan ini adalah menghasilkan produk *Fluids Assessment* (FAss) berdasarkan *Taxonomy of Introductory Physics Problems* (TIPP) yang layak untuk digunakan.

2. Metode

Penelitian ini merupakan *Research and Development* (RnD) dengan mengadaptasi model pengembangan Borg & Gall [8]. Langkah pengembangan diadaptasi menjadi 8 langkah dari 10 langkah disesuaikan dengan kebutuhan penelitian. Adapun langkah pengembangannya sebagai berikut: Pertama, pengumpulan data dan studi literatur. Pada tahap ini, dilakukan wawancara kepada guru fisika kelas XI di salah satu SMAN Kota Serang. Kemudian, mengkaji jurnal, buku, dan berbagai

referensi yang berkaitan dengan *Fluids Assessment* dan *Taxonomy of Introductory Physics Problems* (TIPP). Kedua, deskripsi dan desain produk. Pada tahap ini, dihasilkan berupa kisi-kisi, instrumen *Fass* berdasarkan TIPP. Ketiga, validasi produk oleh ahli. Pada tahap ini dilakukan validasi butir soal oleh dua dosen ahli dan satu guru fisika yang bertujuan untuk menilai dan memberikan masukan serta untuk mengevaluasi produk sebelum uji coba terbatas. Hasil dari analisis validitas isi dilihat dari skor yang diperoleh dari penilaian validitas oleh ahli yang dihitung menggunakan persamaan (1) [9].

$$CVR = \frac{\left(n_e - \frac{N}{2}\right)}{\frac{N}{2}} \quad (1)$$

Dimana, *CVR* adalah *Content Validity Ratio*; n_e adalah jumlah ahli yang memberi respon esensial; dan N adalah jumlah total ahli. Acuan hasil perhitungan yaitu $CVR > 0$, maka butir soal dikatakan valid [10]. Keempat, revisi tahap I. Pada tahap ini, dilakukan perbaikan mengenai isi soal, dan kesesuaian indikator soal berdasarkan validitas ahli sebelum uji coba di lapangan. Kelima, uji coba terbatas. Pada tahap ini responden yang terlibat sebanyak 34 mahasiswa jurusan Pendidikan Fisika Untirta. Hasil dari analisis uji coba terbatas atau validitas empiris dilihat dengan menggunakan program *Quest*. Nilai validitas dapat ditentukan dengan ketentuan *fit item*. Dalam program *Quest*, suatu butir soal dikatakan *fit* terhadap model dengan batas kisaran INFIT MNSQ sebesar 0,77-1,33 dan nilai outfit t sebesar ≥ 2 [11]. Hasil dari analisis reliabilitas dilihat dengan menggunakan program *Quest*. Nilai reliabilitas pada program *Quest* ditinjau dari nilai estimasi *item* dan nilai estimasi *case* (responden). Ketentuan nilai reliabilitas antara responden dengan butir soal disajikan pada tabel 1.

Tabel 1. Klasifikasi Nilai Reliabilitas [10].

Nilai	Keterangan
<0,67	Lemah
0,67-0,80	Cukup
0,80-0,90	Bagus
0,91-0,94	Bagus sekali
>0,94	Istimewa

Hasil dari analisis tingkat kesukaran butir soal dilihat dengan menggunakan program *Quest*. Tingkat kesukaran dapat dilihat dari ketentuan nilai *threshold* yang disajikan dalam tabel 2.

Tabel 2. Ketentuan Tingkat Kesukaran Butir Soal [12].

Nilai <i>Threshold</i>	Keterangan
$b > 2$	Sangat sukar
$1 < b \leq 2$	Sukar
$-1 < b \leq 1$	Sedang
$-1 > b \geq 2$	Mudah
$b < -2$	Sangat mudah

Hasil dari analisis pengujian daya beda butir soal dilihat dengan menggunakan program *Iteman*. Hasil pengujian daya beda dapat dilihat dari ketentuan nilai *korelasi point biserial* yang disajikan dalam tabel 3.

Tabel 3. Ketentuan Korelasi Point Biserial [13].

Korelasi Point Biserial	Keterangan
0,4 - 1,00	Baik
0,30 - 0,39	Sedang (tidak perlu revisi)
0,20 - 0,29	Cukup baik (perlu revisi)
Negatif - 0,19	Buruk (ditolak)

Hasil dari analisis efektivitas distraktor dilihat dengan menggunakan program *IteMan*. Sebuah pengecoh dikatakan berfungsi dengan baik jika paling sedikit dipilih oleh 5% pengikut tes [14]. Ketentuan nilai efektivitas distraktor disajikan dalam tabel 4.

Tabel 4. *Ketentuan Nilai Efektivitas Distraktor [16].*

No	Distraktor yang Tidak Berfungsi	Keterangan
1	0	Sangat Baik
2	1	Baik
3	2	Cukup
4	3	Kurang Baik
5	4	Tidak Baik

Keenam, revisi tahap II. Pada tahap ini, dilakukan perbaikan soal berdasarkan hasil analisis uji coba terbatas. Ketujuh, uji coba lapangan. Uji coba dilakukan terhadap 33 peserta didik kelas XI. Kedelapan, diseminasi. Instrumen tes yang dikembangkan yaitu berupa *Fluids Assessment* berbasis *Taxonomy of Introductory Physics Problems (TIPP)* dalam bentuk pilihan ganda dan disajikan secara daring menggunakan *google form*.

3. Hasil dan Pembahasan

Pengembangan *Fluid Assessment* berbasis *Taxonomy of Introductory Physics Problems (TIPP)* diperoleh beberapa hasil diantaranya:

3.1. Validitas Isi

Berdasarkan hasil analisis angket validasi, dosen ahli dan guru memberikan kesimpulan bahwa beberapa butir soal yang telah dikembangkan valid untuk digunakan. Hal ini dibuktikan dengan hasil perhitungan penilaian menggunakan persamaan CVR yaitu lebih dari nol sehingga 34 dari 57 butir soal valid untuk digunakan pada uji coba terbatas.

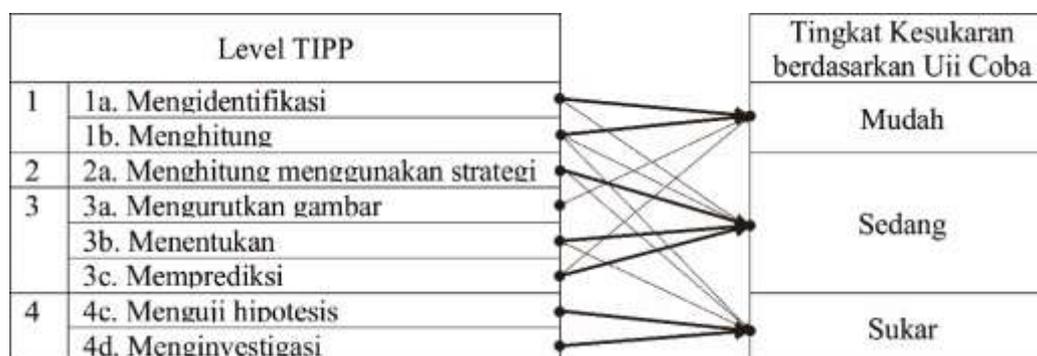
3.2. Validitas Empiris

Berdasarkan dari hasil analisis menggunakan program *Quest*, terdapat 2 butir soal yang tidak valid yaitu butir soal nomor 5 dan 6, karena tidak memenuhi kriteria *item fit*. Soal-soal tersebut tidak memenuhi kriteria *item fit* karena banyak responden yang menjawab benar, ataupun soal tersebut terlalu mudah sehingga semua responden dapat mengerjakan. Selain itu ada beberapa responden yang kurang termotivasi dan kurang serius saat mengerjakan tes, dengan berpikir penilaian tidak mempengaruhi prestasi di kelas. Menyebabkan banyak responden yang menebak. Serta suasana dan kondisi kelas pada saat tes yang tidak kondusif menyebabkan banyak responden yang mencontek dalam memberikan jawaban. Hal ini sesuai dengan teori yang menjelaskan bahwa adanya soal tidak valid dan memiliki nilai reliabilitas rendah dapat disebabkan oleh kemampuan menebak peserta didik yang tinggi dalam menjawab tes yang diberikan [17].

Reliabilitas soal bagus dengan perolehan nilai sebesar 0,90. Sedangkan reliabilitas responden masuk dalam kategori lemah dengan nilai 0,53. Rendahnya nilai reliabilitas dikarenakan ukuran sampel yang dijadikan responden kurang begitu banyak. Apabila sampel terlalu sedikit, variasi hirarki pada penggaris logit juga hanya sedikit [18]. Sehingga, butir-butir soal belum rata menjangkau individu dengan kemampuan ditingkat tinggi sampai ke yang rendah, ataupun belum rata menjangkau tingkatan kesukaran dari yang mudah ke yang sukar. Selain itu, rendahnya nilai reliabilitas juga disebabkan oleh kemampuan menebak jawaban yang tinggi. Semakin tinggi kemampuan menebak responden, tentunya akan memperlemah daya pembeda soal dan menyebabkan semakin homogen skor yang diperoleh [17].

Tingkat kesukaran butir soal dilihat pada nilai *threshold*. Kemudian setelah dikategorikan berdasarkan tingkat kesukaran *Rasch Model*, butir-butir soal tersebut dikorelasikan dengan *Taxonomy*

of *Introductory Physics Problems*. Hasil korelasi antara tingkat kesukaran hasil uji coba dengan TIPP dapat dilihat pada gambar 1.



Gambar 1. Hubungan antara level TIPP dengan tingkat kesukaran hasil uji coba

Berdasarkan Gambar 1, dari 34 butir soal, beberapa soal menunjukkan kesesuaian tiap level TIPP dengan tingkat kesukaran secara empiris, kecuali pada level 3a. Level 3a seharusnya masuk ke dalam kategori tingkat kesukaran sedang, namun kenyataannya masuk ke dalam kategori tingkat kesukaran mudah. Artinya, banyak dari responden yang telah menguasai kemampuan mengurutkan gambar terutama pada materi Bernoulli, sedangkan kemampuan menguji hipotesis, serta menginvestigasi merupakan kemampuan tertinggi yang belum banyak dikuasai beberapa responden. Hal ini sesuai dengan pendapat Teodorescu et al. [19] mengenai level kognitif TIPP yaitu kemampuan menguji hipotesis, serta menginvestigasi merupakan kemampuan tertinggi yang masuk ke dalam level keempat TIPP.

Parameter lain yang dianalisis yakni daya beda namun menggunakan pendekatan klasik. Hasil analisis dapat dilihat pada tabel 5.

Tabel 5. Distribusi Hasil Pengujian Daya Beda

No	Korelasi Point Biserial	Butir Soal	Keterangan
1	0,4-1,00	4, 7, 12, 21	Baik
2	0,30-0,39	2, 3, 5, 19, 25, 26, 28, 32, 33, 34	Sedang (tidak perlu revisi)
3	0,20-0,29	16, 18, 27, 31	Cukup baik (perlu revisi)
4	Negatif- 0,19	1, 6, 8, 9, 10, 11, 13, 14, 15, 17, 20, 22, 24, 29, 30	Buruk (ditolak)

Berdasarkan tabel 5, terdapat 18 soal dengan 4 soal yang masih perlu direvisi yang sudah dapat membedakan antara responden yang telah menguasai materi fluida dinamis dengan responden yang belum menguasai materi fluida dinamis. Serta terdapat 15 soal yang belum bisa membedakan antara responden yang telah menguasai materi fluida dinamis dengan responden yang belum menguasai materi fluida dinamis. Hal tersebut dikarenakan beberapa faktor, yang diantaranya yaitu kunci jawaban butir tidak tepat, jawaban pengecoh tidak berfungsi, dan beberapa butir soal terlalu sulit sehingga banyak responden yang menebak. Semakin tinggi kemampuan menebak responden, tentunya akan memperlemah daya pembeda soal dan menyebabkan semakin homogen skor yang diperoleh [17].

Parameter selanjutnya yang dianalisis yakni efektivitas distraktor dapat dilihat pada tabel 6.

Tabel 6. Distribusi Hasil Efektivitas Distraktor

No	Distraktor yang Tidak Berfungsi	Butir Soal	Keterangan
1	0	9, 12, 13, 17, 20, 29,	Sangat Baik
2	1	2, 11, 14, 15, 16, 18, 22, 26, 30, 31, 32, 34	Baik
3	2	4, 23, 25, 27, 28, 33	Cukup
4	3	1, 3, 10, 19, 21, 24	Kurang Baik
5	4	5, 6, 7, 8	Tidak Baik

Berdasarkan tabel 6, terdapat 24 soal dengan pengecoh yang berfungsi, dapat mengecoh responden yang memang tidak mengetahui kunci jawaban yang tersedia. Juga terdapat 6 soal dengan pengecoh yang masih perlu direvisi karena memiliki pengecoh yang kurang baik, serta 4 soal yang perlu dibuang karna memiliki pengecoh yang tidak baik. Hal tersebut karena dipengaruhi oleh beberapa faktor. Diantaranya, ada responden yang pintar dan lebih memilih pengecoh sebagai jawaban yang benar daripada kunci jawaban.

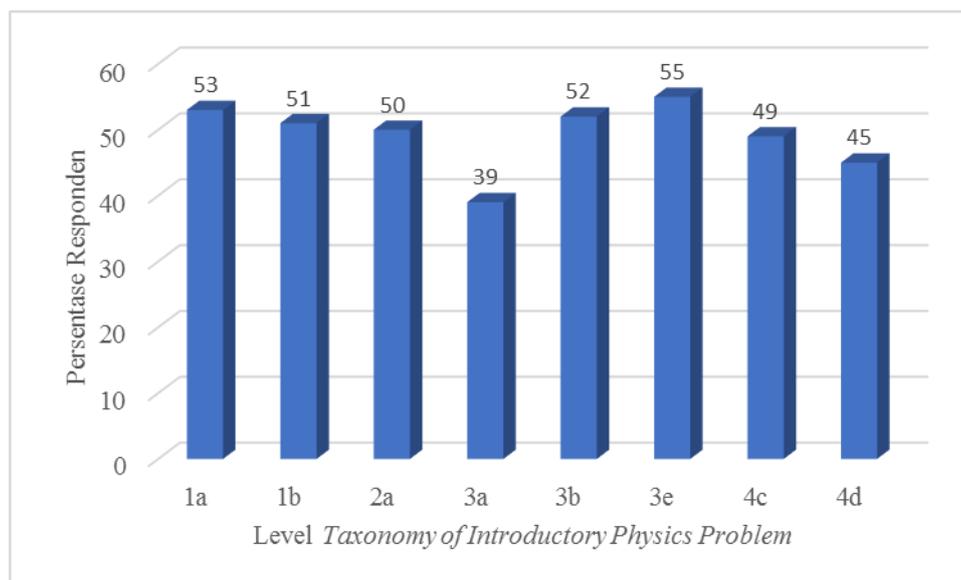
8	0-8	0.971	-0.413	-0.164	a	0.029	0.413	0.164	?
					b	0.971	-0.413	-0.164	*
	CHECK THE KEY				c	0.000	-9.000	-9.000	
	b was specified, a works better				d	0.000	-9.000	-9.000	
					e	0.000	-9.000	-9.000	
					Other	0.000	-9.000	-9.000	

Gambar 3. Hasil analisis butir nomor 8

Misalnya pada gambar 3, soal nomor 8, alternatif A lebih banyak dipilih dibandingkan dengan kunci jawaban B. Tidak berfungsinya pengecoh juga disebabkan oleh kunci jawaban yang terlihat lebih mencolok sehingga seluruh responden memilih kunci jawaban tersebut sebagai jawaban yang benar. Sekalipun soal terlalu sukar atau terlalu mudah apabila setiap pengecoh (distribusi jawaban) pada soal tersebut menunjukkan kondisi distribusi jawaban merata atau distribusi jawabannya logis dan daya pembedanya negatif (kecuali kunci) maka soal-soal tersebut masih memenuhi syarat untuk diterima [14][15]. Oleh karena itu, pada penelitian ini, seluruh butir soal yang telah valid, memiliki daya beda negatif, memiliki pengecoh yang berfungsi, dapat dipakai untuk melatih kemampuan pemecahan masalah siswa pada materi fluida dinamis.

3.3. Implementasi Fass berbasis TIPP

Hasil dari impleementasi Fass berbasis TIPP pada 33 responden yakni siswa SMA dapat ditunjukkan pada gambar 4.



Gambar 4. Diagram hubungan antara level TIPP terhadap persentase jawaban benar responden

Berdasarkan gambar 4, dapat dijelaskan bahwa siswa mendapat kendala pada butir soal dengan level 3a (39%), 4d (45%) dan 4c (49%). Level 3a yaitu *matching* (mencocokkan) memiliki indikator yaitu membandingkan, mengatur, membedakan, mencocokkan, mengurutkan, menyusun, dan menyusun ulang besaran kuantitas fisik, membandingkan fenomena fisika, persamaan, dan pernyataan terkait fisika.



Gambar 5. Butir soal nomor 19

Misalnya pada gambar 5, soal nomor 19 dengan sub-bab debit, siswa dituntut dapat mengurutkan informasi berupa gambar berdasarkan perbedaan serta hubungan antar komponen. Namun, hasil implementasi menunjukkan bahwa banyak dari siswa yang kesulitan menjawab soal tersebut. Hal tersebut dapat disebabkan oleh beberapa faktor. Antara lain menurut Teodorescu et al. [19] bahwa soal dengan indikator mengurutkan informasi masih jarang ditemukan dalam buku pelajaran fisika. Akibatnya, siswa tidak pernah melihat dan terlatih dalam mengerjakan jenis soal seperti ini.

Level 4d yaitu *investigating* (menyelidiki) memiliki indikator yaitu menyelidiki atau menginvestigasi peristiwa masa lalu, sekarang, atau masa depan atau melakukan penyelidikan mengenai informasi, fenomena, pertanyaan spesifik. Pada soal level 4d, siswa dituntut untuk melakukan peninjauan apa yang mungkin terjadi disertai alasan atau argumen yang tepat. Namun, hasil implementasi menunjukkan bahwa banyak dari siswa yang kesulitan menjawab soal tersebut. Hal tersebut dapat disebabkan oleh beberapa faktor. Antara lain menurut Teodorescu et al. [19] bahwa

soal dengan indikator menginvestigasi tidak ada dalam buku pelajaran fisika. Akibatnya, siswa tidak pernah melihat dan terlatih dalam mengerjakan jenis soal menginvestigasi.

Level 4c yaitu *experimenting* (melakukan eksperimen) memiliki indikator yaitu menghasilkan dan menguji hipotesis untuk tujuan memahami fenomena, menggunakan aturan pembuktian yang mematuhi pengujian hipotesis statistik, membuat prediksi berdasarkan prinsip-prinsip yang diketahui atau dihipotesiskan, merancang cara untuk menguji prediksi tersebut. Pada soal level 4c, siswa dituntut untuk menguji hipotesis metode yang paling tepat. Namun, hasil implementasi menunjukkan bahwa banyak dari siswa yang kesulitan menjawab soal tersebut. Hal tersebut dapat disebabkan oleh beberapa faktor. Antara lain menurut Teodorescu et al. [19], bahwa baik buku pelajaran fisika maupun literatur penelitian tidak mengandung banyak jenis soal seperti ini. Akibatnya, siswa tentu juga tidak pernah melihat dan terlatih dalam mengerjakan jenis soal seperti ini.

Berdasarkan hasil analisis yang telah dibahas, dapat disimpulkan bahwa 17 butir soal *Fluids Assessment* berdasarkan *Taxonomy of Introductory Physics Problems* (TIPP) yang dikembangkan, layak digunakan untuk melatih kemampuan pemecahan masalah siswa. Butir-butir soal tersebut yaitu butir soal nomor 2, 3, 4, 12, 16, 18, 19, 21, 23, 25, 26, 27, 28, 31, 32, 33, dan 34. Pada penerapannya, kemampuan mencocokkan pada level 3a, kemampuan melakukan eksperimen pada level 4c, serta kemampuan menginvestigasi pada level 4d, merupakan kemampuan tertinggi yang belum banyak dikuasai oleh siswa, yang disebabkan oleh jarang ditemukannya soal dengan indikator tersebut pada buku pelajaran siswa.

4. Simpulan

Berdasarkan hasil analisis butir soal *Fluids Assessment* berdasarkan *Taxonomy of Introductory Physics Problems* (TIPP) yang dikembangkan, diperoleh 17 butir soal yang layak digunakan untuk melatih kemampuan pemecahan masalah siswa. Dengan kategori, reliabilitas mahasiswa diperoleh hasil 0,53 berada pada kategori lemah dengan reliabilitas soal diperoleh hasil 0,90 bagus, serta diperoleh tingkat kesukaran relatif sedang, menunjukkan daya beda yang baik, dan pengecoh yang berfungsi. Hasil implementasi menunjukkan bahwa sebagian besar siswa menguasai setiap level TIPP kecuali level 3a, 4d dan 4c yang berkaitan dengan *matching*, *investigating*, dan *experimenting*.

Daftar Pustaka

- [1] Kusairi S 2012 *Jurnal Penelitian dan Evaluasi Pendidikan* 3 p 68–87
- [2] Permendiknas 2007 Standar Kualifikasi Akademik dan Kompetensi Guru 1–32
- [3] Yani A 2014 *Mindset Kurikulum 2013* (Bandung: Alfabeta, CV)
- [4] Kuswana and Sunaryo W 2013 *Taksonomi Berpikir* (Bandung: PT Remaja Rosdakarya Offset)
- [5] Sutiadi A and Kurniawati R 2015 Analisis Butir Soal Ujian Nasional SMA Bidang Fisika Tahun 2014 Menggunakan Taxonomy of Introductory Physics Problem *Prosiding Pertemuan Ilmiah XXIX HFI Jateng & DIY* (April) 306–309
- [6] Fabby C and Koenig K M 2015 *Journal of STEM Education* 16 4 p 20–27
- [7] Sumintono B and Widhiarso W 2015 *Aplikasi Pemodelan Rasch: pada Assesment Pendidikan* (Cimahi: Penerbit Trim Komunikata)
- [8] Borg W R and Gall M. D 1989 *Educational Research: An Introduction, Fifth Edition* (New York: Longman)
- [9] Lawshe C H 1975 *A Personnel Psychology* 28 4 p 563–575
- [10] Widhiarso W 2010 *Prosedur Pengujian Validitas Isi melalui Indeks Rasio Validitas Isi (CVR)* (Yogyakarta: Fakultas Psikologi UGM)
- [11] Adams R J and Khoo S T 1996 *Quest: The interactive test analysis system version 2.1* (Melbourne: ACER)
- [12] Setyawarno D 2016 *Upaya Peningkatan Kualitas Butir Soal dengan Analisis Aplikasi Quest* (Yogyakarta: Universitas Negeri Yogyakarta)

- [13] Ebel R L and Frisbie D A 1986 *Using Test and Item Analysis to Evaluate and Improve Test Quality. Essentials of Educational Measurement* (New Jersey: Prentice Hall Inc)
- [14] Hidayati K 2012 *Manual Item And Test Analisis (Iteman): Pedoman Penggunaan Iteman* (Yogyakarta: Universitas Negeri Yogyakarta)
- [15] Ibnu M, Indriyani B, Inayatullah A and Guntara Y 2019 *Prosiding Seminar Nasional Pendidikan FKIP* **2** 1 p 205-210
- [16] Fitriatun A and Sukanti 2016 *Jurnal Kajian Pendidikan Akuntansi Indonesia* **3** p 1–11
- [17] Aida N and Hamdani A S 2017 *Suska Journal of Mathematics Education* **3** 2 p 130–139
- [18] Erdiani G and Liliawati W 2019 *Karakterisasi Instrumen Tes Keterampilan Berpikir Kritis dengan Analisis Model Rasch pada Materi Alat Optik* (Bandung: *Prosiding Seminar Nasional Fisika*) **1** 1 p 99–108
- [19] Teodorescu R E, Bennhold C, Feldman G and Medsker L 2013 *Physical Review Special Topics - Physics Education Research* **9** 1 p 1–20