



## Karakteristik Instrumen Tes Keterampilan Berpikir Kritis Siswa SMA Pada Materi Fluida Statis Menggunakan Analisis Teori Respon Butir

Rini Andita, Iyon Suyana, Winny Liliawati

Artikel ini telah (akan) dipresentasikan pada kegiatan Seminar Nasional Fisika (Sinafi 9.0)

Universitas Pendidikan Indonesia, Bandung, Indonesia

23 September 2023

### Abstract

This study aims to obtain the characteristics of the critical thinking skills (KBK) test instrument which is constructed in 7 essay questions on static fluid material. The test instrument measures five KBK indicators and analyzed using item response theory (TRB). The research method is mixed methods with the exploratory sequential design. The expert validation were analyzed using the GRM by eirt, showing that the test was declared valid based on the information function which the range  $-2$  to  $+2$ . Field trials of 133 students were analyzed using the rasch model by winstep application. The Outfit MnSq, ZStd and PT-Measure Correlation values show that all questions are valid. The item reliability value of 0.94 means that the reliability of items is very good category, the person reliability value of 0.73 means that the reliability of persons is the good enough category, and the Cronbach Alpha value of 0.75 means that correlation is good category. For the level of difficulty, the most difficult item is item 7 with measure item value of 0.62 and the easiest item is item 1 with measure item value of  $-0.49$ . So, that the test can be declared valid and reliable for use with the item difficulty level being moderate.

**Keywords:** critical thinking skills · item response theory · static fluid

### INTRODUCTION

Abad ke-21 ditandai sebagai era globalisasi yang memberi dampak cukup luas dalam berbagai aspek kehidupan, termasuk tuntutan dalam penyelenggaraan pendidikan (Wijaya dkk., 2016). Dalam menghadapinya, siswa dituntut untuk memiliki keterampilan abad 21. Keterampilan tersebut biasa dikenal dengan 4C; *critical thinking, communication, collaboration, dan creativity* (National Educational Association, 2014). Tujuan diberlakukannya Kurikulum 2013 di Indonesia diantaranya adalah mengembangkan keterampilan berpikir tingkat tinggi (*higher order thinking skills*) atau HOTS, salah satu keterampilan berpikir tingkat tinggi adalah keterampilan berpikir kritis (*critical thinking skills*). Menurut Khaeruddin dkk. (2018) sebagian besar indikator pencapaian kompetensi dasar Kurikulum 2013 Mata Pelajaran Fisika SMA juga merupakan indikator keterampilan berpikir kritis. Maka dari itu keterampilan berpikir kritis sangat penting dalam pembelajaran terutama Fisika.

Tingkat keterampilan berpikir kritis siswa SMA berada pada kategori rendah (Nurhayati dkk., 2022). Hal ini disebabkan karena masih rendahnya pengetahuan dasar yang dimiliki oleh

✉ Rini Andita  
[riniandita.25@upi.edu](mailto:riniandita.25@upi.edu)

Winny Liliawati Iyon Suyana  
[winny@upi.edu](mailto:winny@upi.edu) [iyons@upi.edu](mailto:iyons@upi.edu)

Program Studi Pendidikan Fisika, FPMIPA UPI, Bandung, Indonesia.

**How to Cite:** Andita, R., & Suyana, Iyon., & Liliawati, W. (2023). Karakteristik Instrumen Tes Keterampilan Berpikir Kritis Siswa SMA Pada Materi Fluida Statis Menggunakan Analisis Teori Respon Butir. *Prosiding Seminar Nasional Fisika*, 2(1), 552-560. <http://proceedings.upi.edu/index.php/sinafi>

siswa dan keterampilan berpikir kritis kurang diukur atau diasah oleh guru mata pelajaran fisika (Hasanah, 2019). Salah satu cara untuk melihat atau mengetahui keterampilan berpikir kritis yang dimiliki oleh siswa dapat dilakukan dengan menggunakan sebuah tes dengan alat ukur yang relevan.

Berdasarkan studi pendahuluan yang dilakukan melalui wawancara dan studi literatur, ditemukan permasalahan yang umum terjadi pada jenjang SMA dimana minimnya instrumen penilaian yang sesuai untuk mengukur keterampilan berpikir kritis siswa. Penelitian sebelumnya telah mengembangkan suatu alat ukur keterampilan berpikir kritis siswa pada materi ajar suhu dan kalor (Ritdamaya, 2016). Selain itu, Mabrurroh (2017) juga telah mengkonstruksi alat ukur keterampilan berpikir kritis siswa materi gelombang bunyi. Alat ukur yang tersedia hanya dalam beberapa pokok bahasan, sedangkan pokok uji keterampilan berpikir kritis perlu dikembangkan dalam semua pokok bahasan fisika.

Dalam proses pembuatan tes tersebut, perlu adanya analisis butir soal dengan tujuan untuk mengetahui mutu soal sehingga dapat memberikan informasi mengenai karakteristik dari butir tes tersebut ataupun analisis empirisnya. Untuk menganalisis butir soal pada instrumen tes terdapat dua teori pengukuran yaitu teori tes klasik atau CTT (*Classical Test Theory*) dan teori tes modern atau sering disebut TRB (Teori Respon Butir) atau IRT (*Item Response Theory*). Berikut adalah kelebihan *Item Response Theory* (IRT) dalam Hambleton & Jones, (1993) diantaranya adalah statistik butir soal yang tidak bergantung pada kelompok peserta tes, skor mendeskripsikan kemampuan peserta tes dan tidak bergantung pada tingkat kesulitan tes, model tes yang dapat digunakan untuk menghubungkan butir soal dengan tingkat kemampuan peserta tes, model tes yang tidak memerlukan tes secara paralel dalam menentukan reliabilitas.

IRT terdiri atas dua jenis tes yaitu dikotomi (dua kategori) seperti benar atau salah dan politomi (lebih dari dua kategori) seperti penilaian skala *Likert*. Dalam menganalisis tes dikotomi terdapat tiga jenis model yang dapat digunakan, diantaranya model 1 parameter logistik (1PL), model 2 parameter logistik (2PL) dan model 3 parameter logistik (3PL) (Hambleton, 1991). Sedangkan terdapat beberapa jenis model analisis untuk tes politomi diantaranya *Graded Response Model* (GRM) atau model respon bergradasi, *Modified Graded Response Model* (M-GRM) atau model modifikasi respon bergradasi, *Partial Credit Model* (PCM) atau model kredit parsial, *Generalized Partial Credit Model* (G-PCM) atau model generalisasi kredit parsial dan *Rating Scale Model* (RSM) atau model skala penilaian. Rasch Model (setara dengan model 1 PL) juga dapat digunakan untuk analisis data politomi.

IRT ini dikembangkan oleh para ahli untuk mengatasi kekurangan-kekurangan yang ada pada CTT. Namun, belum banyak peneliti yang menggunakan IRT atau teori respon butir ini, sehingga perlu adanya penelitian dalam bidang Pendidikan yang menggunakan teori respon butir agar estimasi karakteristik suatu objek akan lebih teliti.

Materi fisika sebagian besar bersifat abstrak dan membutuhkan tingkat pemahaman dan analisis siswa yang cukup tinggi, salah satunya materi fluida statis sehingga dapat membantu siswa untuk memahami dan menganalisis suatu masalah yang terdapat pada materi tersebut. Pada penelitian kali ini materi pembahasan yang akan dikembangkan adalah fluida statis. Pemilihan materi ajar yang berbeda untuk penelitian kali ini dilandasi pada kebutuhan untuk memperkaya instrumen tes yang dapat digunakan untuk menilai keterampilan berpikir kritis siswa.

Keterampilan berpikir kritis pada penelitian ini didasarkan pada keterampilan berpikir kritis yang dikemukakan oleh Ennis (2011). Terdapat lima aspek keterampilan berpikir kritis diantaranya yaitu klarifikasi dasar (*elementary clarification*), dasar dalam mengambil keputusan atau dukungan (*the basis for the decision basic support*), inferensi (*inference*), klarifikasi lanjut (*advanced clarification*), serta strategi dan taktik (*strategies and tactics*).

Keterampilan berpikir kritis pada siswa SMA terutama materi fisika perlu dilatihkan dan dievaluasi sebagai tuntutan keterampilan abad 21. Namun, instrumen tes keterampilan berpikir kritis belum banyak tersedia terutama yang melatih materi Fisika dan telah dikarakterisasi. Oleh karena itu, peneliti memutuskan untuk melakukan penelitian dengan memperoleh karakteristik instrumen tes keterampilan berpikir kritis siswa SMA pada materi fluida statis berdasarkan analisis Teori Respon Butir.

## METODE

### Metode dan Desain Penelitian

Penelitian ini menggunakan metode campuran atau *mix methods*, yaitu pendekatan penelitian yang menggabungkan atau menghubungkan metode penelitian kuantitatif dan kualitatif. Desain penelitian *The Exploratory Sequential Design* dengan tiga tahap utama, tahap pertama yaitu mengumpulkan dan menganalisis data kualitatif, tahap kedua yaitu tahap mengumpulkan dan menganalisis data kuantitatif, dan yang terakhir adalah interpretasi data. Data kualitatif meliputi studi pendahuluan, analisis materi dan indikator KBK, perancangan produk instrumen tes keterampilan berpikir kritis, serta justifikasi pakar. Materi fluida statis sebagai konten yang diperoleh dari hasil studi literatur sumber-sumber berupa artikel, jurnal, buku dan lain-lain. tahap mengumpulkan dan mengolah data kuantitatif dengan cara melakukan pengujian instrumen tes keterampilan berpikir kritis di lapangan yakni kepada siswa. Pengujian ini dilakukan dua tahap yaitu uji skala kecil dan uji lapangan untuk memperoleh data kuantitatif berupa jawaban dari setiap butir soal yang disusun berdasarkan indikator keterampilan berpikir kritis. Uji skala kecil dilakukan untuk uji awal produk dan menilai keterbacaan soal sehingga akan terbentuk produk final untuk digunakan pada uji lapangan. Data yang didapat dari uji lapangan kemudian digunakan untuk melakukan analisis serta interpretasi mengenai karakteristik instrumen tes keterampilan berpikir kritis pada materi fluida statis jenjang SMA.

### Sasaran Penelitian

Penelitian ini dilakukan terhadap 133 orang siswa SMA kelas XI yang telah mempelajari materi fluida statis. Uji lapangan instrumen tes dilakukan pada bulan Mei 2023 di tiga sekolah berbeda di kota Bandung.

### Teknik Pengumpulan Data

Data kualitatif diperoleh dari tahap studi pendahuluan berupa wawancara dan studi literatur, analisis tujuan dan spesifikasi tes, perancangan instrumen tes, serta tahap justifikasi pakar berupa saran perbaikan terhadap instrumen tes. Data kualitatif diperoleh dari justifikasi pakar, uji skala kecil, dan uji lapangan. Justifikasi pakar menghasilkan data berupa *skala likert* yang dianalisis menggunakan *graded response models* (GRM) berbantuan *software eirt*. Sedangkan data yang diperoleh dari hasil uji lapangan dianalisis menggunakan *Rasch Model* berbantuan aplikasi *winstep*.

Berisi jenis penelitian, waktu dan tempat penelitian, target/sasaran, subjek penelitian, prosedur, instrumen dan teknik analisis data serta hal-hal lain yang berkaitan dengan cara penelitiannya. target/sasaran, subjek penelitian, prosedur, data dan instrumen, dan teknik pengumpulan data, serta teknik analisis data serta hal-hal lain yang berkaitan dengan cara penelitiannya dapat ditulis dalam sub-subbab, dengan sub-subheading. Sub-subjudul tidak perlu diberi notasi, namun ditulis dengan huruf kecil berawalkan huruf kapital, TNR-11 unbold, rata kiri.

Khususnya untuk penelitian kualitatif, waktu dan tempat penelitian perlu dituliskan secara jelas (untuk penelitian kuantitatif, juga perlu). Target/subjek penelitian (untuk penelitian kualitatif) atau populasi-sampel (untuk penelitian kuantitatif) perlu diurai dengan jelas dalam bagian ini. Perlu juga dituliskan teknik memperoleh subjek (penelitian kualitatif) dan atau teknik samplingnya (penelitian kuantitatif).

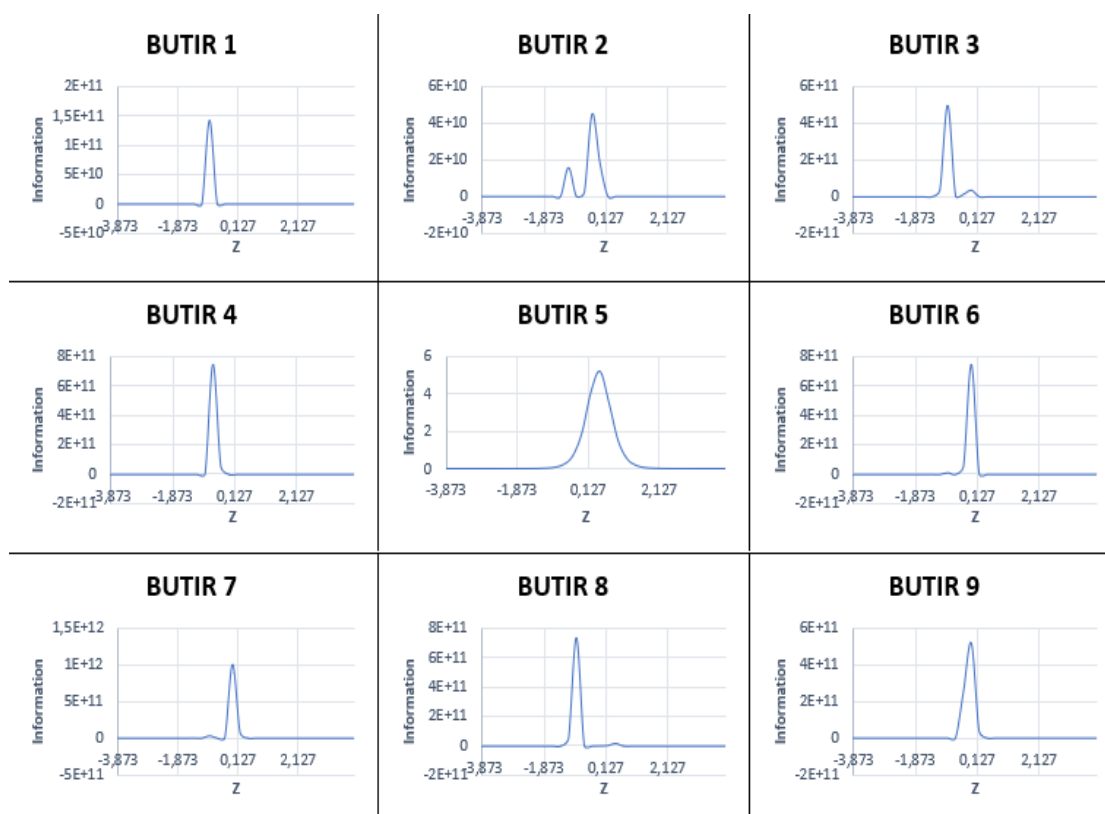
Prosedur perlu dijabarkan menurut tipe penelitiannya. Bagaimana penelitian dilakukan dan data akan diperoleh, perlu diuraikan dalam bagian ini. Untuk penelitian eksperimental, jenis rancangan (experimental design) yang digunakan sebaiknya dituliskan di bagian ini. Macam data, bagaimana data dikumpulkan, dengan instrumen yang mana data dikumpulkan, dan bagaimana teknis pengumpulannya, perlu diuraikan secara jelas dalam bagian ini.

Bagaimana memaknakan data yang diperoleh, kaitannya dengan permasalahan dan tujuan penelitian, perlu dijabarkan dengan jelas. (Catatan: Sub-subbab bisa berbeda, menurut jenis atau pendekatan penelitian yang digunakan. Jika ada prosedur atau langkah yang sifatnya sekuensial, dapat diberi notasi (angka atau huruf) sesuai posisinya).

## HASIL DAN PEMBAHASAN

Instrumen tes keterampilan berpikir kritis dikonstruksi menghasilkan 9 soal *essay* yang kemudian mendapatkan validasi dari lima orang ahli. Aspek penilaian terhadap setiap butir soal pada tahap justifikasi pakar diantaranya yaitu mengenai kesesuaian indikator KBK yang digunakan, kesesuaian informasi dan konsep yang diberikan pada soal, tata bahasa yang digunakan, serta kesesuaian rubrik penilaian dalam mengukur jawaban siswa. Data hasil justifikasi pakar adalah berupa *skala likert* (1-5) dan saran perbaikan soal.

Data kuantitatif hasil justifikasi pakar diolah menggunakan GRM berbantuan *eirt* dengan hasil berupa tabel fungsi informasi seperti pada Gambar 1. Menurut Hambleton dan Swaminathan (1985), suatu tes dikatakan valid jika puncak kurva fungsi informasi berada pada rentang -2 sampai +2. Berdasarkan Gambar 1, puncak terkeskrem *Z-score* butir 1 bernilai -0,825, butir 2 bernilai -0,317, butir 3 bernilai -0,825, butir 4 bernilai -0,571, butir 5 bernilai +0,444, butir 6 bernilai -0,063, butir 7 bernilai -0,063, butir 8 bernilai -0,825, dan butir 9 bernilai -0,063. Sehingga diketahui bahwa puncak informasi terekestrem untuk semua butir soal berada pada rentang -0,825 sampai +0,444. Dapat disimpulkan bahwa instrumen tes keterampilan berpikir kritis dikatakan valid untuk setiap butir soal. Pada justifikasi pakar ini empat ahli menyatakan tes yang telah disusun layak digunakan untuk uji coba lapangan setelah revisi dan satu ahli menyatakan tes layak untuk uji coba lapangan tanpa revisi



Gambar 1. Fungsi Informasi Butir Soal Hasil Justifikasi Pakar

Selanjutnya dilakukan uji skala kecil terhadap 25 siswa yang dipilih dengan cara *purposive sampling* untuk memilih partisipan berdasarkan kriteria tertentu yaitu partisipan telah mempelajari materi fluida statis dan bersedia menjadi partisipan penelitian.

Hasil uji skala kecil dari instrumen tes sebanyak 9 soal *essay* kemudian diolah menggunakan *Rasch Model* berbantuan aplikasi *winstep*. Dari output tabel selanjutnya dianalisis nilai *outfit MnSq*, *ZStd* dan *PT-Mea Corr*. Rentang yang diterima untuk *Outfit Mean Square (MnSq)* adalah  $0,5 < MnSq < 1,5$ , untuk  $-2 < ZStd < 2$ , dan untuk *Point Measure Correlation (PT-Mea Corr)* adalah  $0,4 < PT-Mea Corr < 0,85$ . Menurut Sumintono & Widhiarso (2014) Item instrumen dikatakan valid berdasarkan *Rasch Model* apabila memenuhi minimal dua dari tiga kriteria.

Tabel 1. Validitas Uji Skala Kecil

| Butir ke- | <i>MnSq</i> |               | <i>ZStd</i> |               | <i>PT-Mea Corr</i> |               | Validasi    |
|-----------|-------------|---------------|-------------|---------------|--------------------|---------------|-------------|
|           | Nilai       | Ket.          | Nilai       | Ket.          | Nilai              | Ket.          |             |
| 1         | 1,18        | <i>Fit</i>    | 0,6         | <i>Fit</i>    | 0,25               | <i>Misfit</i> | Valid       |
| 2         | 1,18        | <i>Fit</i>    | 0,6         | <i>Fit</i>    | 0,41               | <i>Fit</i>    | Valid       |
| 3         | 0,62        | <i>Fit</i>    | -1,1        | <i>Fit</i>    | 0,41               | <i>Fit</i>    | Valid       |
| 4         | 0,91        | <i>Fit</i>    | -0,3        | <i>Fit</i>    | 0,39               | <i>Misfit</i> | Valid       |
| 5         | 1,52        | <i>Misfit</i> | 1,6         | <i>Fit</i>    | 0,31               | <i>Misfit</i> | Tidak Valid |
| 6         | 0,53        | <i>Fit</i>    | -2,0        | <i>Misfit</i> | 0,60               | <i>Fit</i>    | Valid       |
| 7         | 0,85        | <i>Fit</i>    | -0,5        | <i>Fit</i>    | 0,65               | <i>Fit</i>    | Valid       |
| 8         | 1,69        | <i>Misfit</i> | 0,9         | <i>Fit</i>    | -0,27              | <i>Misfit</i> | Tidak Valid |
| 9         | 0,99        | <i>Fit</i>    | 0,41        | <i>Fit</i>    | 0,42               | <i>Fit</i>    | Valid       |

Ket: *Fit*: nilai diterima, *Misfit*: nilai tidak diterima,

Dari hasil analisis data yang telah dilakukan, terdapat 2 dari 9 soal instrumen tes KBK yang tidak valid yaitu nomor 5 dan 8 karena terdapat dua dari tiga kategori yang tidak *fit*. Pada uji skala kecil siswa diberikan angket keterbacaan soal yang akan menjadi pertimbangan dalam memutuskan soal akan direduksi atau tidak. Berdasarkan hasil analisis validitas uji skala kecil

soal yang tidak valid yaitu nomor 5 dan 8 akan direduksi dari instrumen tes KBK Fluida Statis dengan dukungan data angket keterbacaan soal, sehingga soal yang diujikan pada tahap uji lapangan adalah sebanyak 7 soal *essay*.

Uji lapangan dilakukan terhadap 133 siswa Kelas XI SMA pada tiga sekolah berbeda yang berada di kota Bandung. Hasil uji lapangan dari instrumen tes kemudian diolah menggunakan *Rasch Model* berbantuan aplikasi *winstep*. Karakteristik instrument tes yang akan dianalisis adalah validitas, reliabilitas, dan tingkat kesukaran. Nilai validitas akan dianalisis berdasarkan output tabel yaitu nilai *outfit MnSq*, *ZStd* dan *PT-Mea Corr*.

Tabel 2. Validitas Uji Lapangan

| Butir ke- | <i>MnSq</i> |            | <i>ZStd</i> |               | <i>PT-Mea Corr</i> |            | Validasi |
|-----------|-------------|------------|-------------|---------------|--------------------|------------|----------|
|           | Nilai       | Ket.       | Nilai       | Ket.          | Nilai              | Ket.       |          |
| 1         | 0,75        | <i>Fit</i> | -2,2        | <i>Misfit</i> | 0,68               | <i>Fit</i> | Valid    |
| 2         | 1,06        | <i>Fit</i> | 0,6         | <i>Fit</i>    | 0,69               | <i>Fit</i> | Valid    |
| 3         | 0,96        | <i>Fit</i> | -0,3        | <i>Fit</i>    | 0,48               | <i>Fit</i> | Valid    |
| 4         | 0,79        | <i>Fit</i> | -1,9        | <i>Fit</i>    | 0,63               | <i>Fit</i> | Valid    |
| 5         | 1,08        | <i>Fit</i> | 0,7         | <i>Fit</i>    | 0,57               | <i>Fit</i> | Valid    |
| 6         | 1,12        | <i>Fit</i> | 1,0         | <i>Ft</i>     | 0,68               | <i>Fit</i> | Valid    |
| 7         | 1,30        | <i>Fit</i> | 2,4         | <i>Misfit</i> | 0,70               | <i>Fit</i> | Valid    |

Berdasarkan Tabel 2, validitas item untuk instrumen tes keterampilan berpikir kritis hasil analisis validasi empiris ini menunjukkan bahwa semua soal dinyatakan valid.

Untuk reliabilitas dilakukan analisis untuk *item reliability*, *person reability*, dan reliabilitas korelasi reliabilitas instrumen (*Cronbach Alpha*).

Berdasarkan Sumintono & Widhiarso (2014) nilai *Reliability (Item/Person)* >0,94 dengan interpretasi istimewa, 0,91-0,94 bagus sekali, 0,81-0,90 bagus, 0,67-0,80 cukup, dan <0,67 dengan interpretasi lemah. Sedangkan untuk *Reliability Cronbach Alpha* menurut Sumintono & Widhiarso (2014) nilai *Alpha Cronbach* >0,8 dengan kategori bagus sekali, 0,7-0,8 bagus, 0,6-0,7 cukup, 0,5-0,6 jelek, dan <0,5 dengan interpretasi buruk. *Item reliability* didapatkan dari *output table: Measured Item* seperti pada Gambar 2.

SUMMARY OF 7 MEASURED Item

|                         | TOTAL SCORE | COUNT   | MEASURE | MODEL ERROR | INFIT MNSQ | ZSTD             | OUTFIT MNSQ | ZSTD |
|-------------------------|-------------|---------|---------|-------------|------------|------------------|-------------|------|
| MEAN                    | 381.7       | 133.0   | .00     | .09         | 1.01       | .0               | 1.01        | .0   |
| S.D.                    | 49.6        | .0      | .36     | .00         | .19        | 1.7              | .18         | 1.5  |
| MAX.                    | 449.0       | 133.0   | .62     | .09         | 1.31       | 2.4              | 1.30        | 2.4  |
| MIN.                    | 296.0       | 133.0   | -.49    | .08         | .73        | -2.6             | .75         | -2.2 |
| REAL RMSE               | .09         | TRUE SD | .35     | SEPARATION  | 3.89       | Item RELIABILITY | .94         |      |
| MODEL RMSE              | .09         | TRUE SD | .35     | SEPARATION  | 4.06       | Item RELIABILITY | .94         |      |
| S.E. OF Item MEAN = .15 |             |         |         |             |            |                  |             |      |

Gambar 2. Item reliability dari output table: Measured Item

Nilai *item reliability* adalah 0,94 sehingga dapat disimpulkan reliabilitas item termasuk kategori bagus sekali. *Person reliability* bisa didapatkan dari *output table: Measured Person* seperti pada Gambar 3.



SUMMARY OF 133 MEASURED Person

|                           | TOTAL SCORE | COUNT   | MEASURE | MODEL ERROR | INFIT |                    | OUTFIT |      |
|---------------------------|-------------|---------|---------|-------------|-------|--------------------|--------|------|
|                           |             |         |         |             | MNSQ  | ZSTD               | MNSQ   | ZSTD |
| MEAN                      | 20.1        | 7.0     | .51     | .38         | 1.00  | .0                 | 1.01   | .0   |
| S.D.                      | 5.9         | .0      | .80     | .05         | .49   | 1.0                | .49    | 1.0  |
| MAX.                      | 32.0        | 7.0     | 2.46    | .58         | 2.38  | 2.3                | 2.43   | 2.3  |
| MIN.                      | 8.0         | 7.0     | -1.32   | .34         | .18   | -2.5               | .19    | -2.4 |
| REAL RMSE                 | .42         | TRUE SD | .69     | SEPARATION  | 1.65  | Person RELIABILITY |        | .73  |
| MODEL RMSE                | .38         | TRUE SD | .71     | SEPARATION  | 1.85  | Person RELIABILITY |        | .77  |
| S.E. OF Person MEAN = .07 |             |         |         |             |       |                    |        |      |

Gambar 3. Person Reliability dari output table: Measured Person

Nilai *person reliability* adalah 0,73 sehingga dapat disimpulkan reliabilitas peserta tes termasuk kategori cukup. Selanjutnya yaitu nilai *Cronbach Alpha* yang bisa didapatkan dari *output table: Misfit order* seperti pada Gambar 4.

Person RAW SCORE-TO-MEASURE CORRELATION = 1.00  
 CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .75

Gambar 4. Cronbach Alpha dari output table: Summery Item

Nilai *person reliability* adalah 0,75 sehingga dapat disimpulkan reliabilitas peserta tes termasuk kategori bagus.

Pada pemodelan Rasch, tingkat kesukaran butir soal dikategorikan berdasarkan *item measure* dan nilai simpangan baku (SD). Menurut Sumintono & Widhiarso (2014) tingkat kesekaran tersebut dibagi menjadi beberapa kategori diantaranya yaitu *measure logit < -SD logit* dengan katgori sangat mudah,  $-SD \logit \leq \text{measure logit} < 0$  dengan kategori mudah, *measure logit = 0* dengan kategori sedang,  $0 < \text{measure logit} \leq SD \logit$  dengan kategori sukar, dan *measure logit > SD* dengan kategori sangat sukar. *Measure logit* dan *SD logit* bisa didapat dari *output table: misfit order*. Nilai SD yang didapat adalah 0,36 sehingga tingkat kesukaran soal adalah sebagai berikut.

Tabel 3. Tingkat Kesukaran Soal

| Butir ke- | Item Measure | Keterangan |
|-----------|--------------|------------|
| 1         | -0,49        | Mudah      |
| 2         | 0,12         | Mudah      |
| 3         | 0,31         | Mudah      |
| 4         | 0,00         | Sedang     |
| 5         | -0,23        | Sukar      |
| 6         | -0,33        | Sukar      |
| 7         | 0,62         | Sukar      |

Apabila diurutkan berdasarkan nilai *item measure*, tingkat kesukaran butir soal apabila diurutkan dari soal paling sukar sampai paling mudah adalah 7, 3, 2, 4, 5, 6, 1. Sehingga dapat disimpulkan bahwa butir soal yang paling sukar adalah soal nomor 7 dan butir soal paling mudah adalah soal nomor 1.

## SIMPULAN

Simpulan pada penelitian ini secara umum yaitu mengenai karakteristik instrumen tes keterampilan berpikir kritis materi Fluida Statis menggunakan analisis teori respon butir. Hasil validasi ahli terhadap instrumen tes dianalisis menggunakan TRB *graded response models* (GRM) berbantuan *irt*, menunjukkan bahwa tes KBK dinyatakan valid berdasarkan puncak fungsi informasi yang berada pada rentang -2 sampai +2. Uji coba lapangan terhadap 133 siswa dianalisis menggunakan TRB *Rasch Model* berbantuan aplikasi *winstep* yang kemudian dianalisis berdasarkan validitas, reliabilitas, dan tingkat kesukaran. Nilai *Outfit MnSq*, *ZStd* dan *PT-Measure Correlation*, menunjukkan bahwa semua soal *fit* dan dapat dinyatakan valid. Nilai *item reliability* sebesar 0,94 berarti reliabilitas butir soal termasuk kategori bagus sekali, nilai *person reliability* sebesar 0,73 berarti reliabilitas peserta tes termasuk kategori cukup bagus, serta untuk nilai *Cronbach Alpha* sebesar 0,75 berarti reliabilitas secara keseluruhan termasuk kategori bagus dan instrumen tes KBK reliabel untuk digunakan. Untuk tingkat kesukaran, butir soal tersulit adalah butir 7 dengan nilai *item measure* 0,62 serta butir paling mudah yaitu butir 1 dengan nilai *item measure* -0,49.

## DAFTAR PUSTAKA

- Arikunto & Suharsimi. (2009). *Prosedur Penelitian Suatu Pendekatan Praktik*. Jakarta: PT Rineka Cipta.
- Ennis, R. H. (1993). Critical Thinking Assessment. *Theory Into Practice*, 32(3), 179–186. [Http://Www.Jstor.Org/Stable/1476699](http://www.jstor.org/stable/1476699)
- Erdiani, G. (2019). *Karakterisasi Instrumen Tes Keterampilan Berpikir Kritis Dengan Analisis Model Rasch Pada Materi Alat Optik*. (Skripsi). Universitas Pendidikan Indonesia.
- Facione, P. A. (2001). *Test Manual: Thr Test of Everyday Reasoning-;/A Measure of Thinking Skills*. Millbrae, CA: Insight Assessment/Thr California Academic Press.
- Fisher, A. (2001). *Critical Thinking an Introduction*. United Kingdom: Cambridge University Press.
- Khaeruddin, Bunga D. A. & Jasruddin. (2018). Analisis Keterampilan Berpikir Kritis Pada Kompetensi Dasar Kurikulum 2013 Mata Pelajaran Fisika SMA. *Prosiding Seminar Nasional Lembaga Penelitian Universitas Negeri Makassar*. (hlm. 178-181).
- Hasanah, I. (2019). *Analisis Kemampuan Berpikir Kritis Siswa SMAN Tangerang Selatan Pada Materi Hukum Newton Tentang Gerak*. (Skripsi). Universitas Islam Negeri Hidayatullah, Jakarta.
- In Bloom, B. S. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York: Longmans, Green.
- Istiyono, E. (2019). Developing IRT-Based Physics Critical Thinking Skill Test: A CAT to Answer 21<sup>st</sup> Century Challenge. *International Journal of Instruction*. 12(2). 267-280.
- Kanginan, M. 2014. *Fisika 2 Untuk SMA/MA Kelas XI*. Jakarta: Erlangga
- Kementerian Pendidikan dan Kebudayaan. (2016). *Silabus Mata Pelajaran Fisika Sekolah Menengah Atas/ Madrasah Aliyah (SMA/MA) Mata Pelajaran Fisika*. Jakarta: Kemendikbud.
- Kementerian Pendidikan dan Kebudayaan. (2018). *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 37 Tahun 2018 tentang Kompetensi Inti dan Kompetensi Dasar Pelajaran pada Kurikulum 2013 pada Pendidikan Dasar dan Pendidikan Menengah*. Jakarta: Kemendikbud.
- Mukhtar, M. & Haniin, K. (2019). *Modul Penyusunan Soal Keterampilan Berpikir Tingkat Tinggi*. Jakarta: Kemendikbud.
- Mabrurroh, F. (2017). *Konstruksi Instrumen Tes Keterampilan Berpikir Kritis Terkait Materi Gelombang Bunyi*. (Tesis). Sekolah Pascasarjana, Universitas Pendidikan Indonesia, Bandung.



- National Educational Association. (2014). *Priparing 21<sup>st</sup> Century Stdents for a Global Society*. Washington D.C.: NEA
- Nurhayati, Asniar & Khaeruddin. (2022). Analisis Keterampilan Berpikir Kritis dalam Pembelajaran Fisika Siswa di SMAN 11 Makassar. *Jurnal Sains dan Pendidikan Fisika*. 18(2). 140-151. doi: <https://doi.org/10.35580/jspf.v18i2.31622>
- Paul & Elder. (2006). *The Miniature Guide to Critical Thinking Concept and Tools: The Foundation of Critical Thinking*. Tersedia: [www.criticalthinking.org](http://www.criticalthinking.org)
- Ramalis, T.R. (2015). Karakteristik Pengembangan Tes Keterampilan Berpikir Kritis Bumi dan Antariksa untuk Calon Guru. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*, 1(2). 51-58.
- Rasch, D., Kubinger, K. D., & Yanagida, T. (2011). *Using R and SPSS*. Chichester: John Wiley & Sons, Ltd.
- Ritdamaya, D. (2016). Konstruksi Instrumen Tes Keterampilan Berpikir Kritisss Terkait Materi Suhu dan Kalor. *Jurnal Penelitian & Pengembangan Pendidikan Fisika*. 2(1). 87-96. doi: <https://doi.org/10.21009/1.02212>
- Sudjana. (2016). *Metoda Statistika*. Bandung: Tarsito.
- Sugiyono. (2017). *Statistika untuk Penelitian*. Bandung: CV Alfabeta.
- Sumintono, B. & Widhiarso, W. (2014). *Aplikasi Pemodelan Rasch pada Assesment Pendidikan*. Cimahi: Trim Komunikata.
- Tiruneh, Mieke & Weldessie. (2015). *Measuring Critical Thinking in Physics: Magnetism*. *International Journal of Science and Mathematics Education on Springer*. doi: <https://doi.org/10.1007/s10763-016-9723-0>
- Watson, G., & Glaser, E. (2009). Watson Glaser Critical Thinking Appraisal. Pearson, 24–26. [http://www.pearsonvue.com/phnro/wg\\_practice.pdf](http://www.pearsonvue.com/phnro/wg_practice.pdf)
- Wijaya, E. Y., Sudjimat, D. A. & Nyoto, A. (2016). Tranformasi Pendidikan Abad 21 Sebagai Tuntutan Pengembangan Sumber Daya Manusia di Era Global. *Prosiding Seminar Nasional Pendidikan Matematika*. (hlm. 263-278).
- Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. New Jersey: Lawrence Erlbaum Associates.
- Zubaidah, S. (2010). Berpikir Kritis: Kemampuan Berpikir Tingkat Tinggi yang Dapat Dikembangkan melalui Pembelajaran Sains. *Makalah Seminar Nasional Sains dengan Tema Optimalisasi Sains untuk Memberdayakan Manusia*. Pascasarjana Unesa. (hlm 1-14)