

Identifikasi Kualitas Instrumen Soal *Three-Tier Diagnostic Test* Pada Materi Suhu, Kalor dan Perpindahannya

Hellmy Nur Pratama Annuari Putri, Suliyannah, Lydia Rohmawati

Jurusan Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya

Email: hellmyputri94@gmail.com

Abstrak

Tujuan dari penelitian ini adalah mendeskripsikan kualitas instrumen soal yang telah dikembangkan, yakni tipe *three-tier diagnostic test* terdiri atas tiga tingkat untuk menilai pemahaman konsep siswa. Metode pengembangan ini menggunakan metode pengembangan yang dikemukakan oleh Peter Fenrich. Hasil analisis kualitas instrumen dinyatakan baik dengan kategori valid menurut validator sebesar 88,19%, indeks Alpha Cronbach untuk reliabilitas adalah 0,98 dinyatakan reliabel, daya beda soal 55% tergolong jelek, serta taraf kesukaran 80% tergolong sukar.

Kata Kunci: *three-tier diagnostic test*, kualitas instrumen.

Abstract

The main purposes are to describe instrument quality and analyze a certain problem that cause a misconception especially in heat and temperature concept, *Three-tier diagnostic test* consisting of three-tier items for assessing student's understanding of states of matter concepts. This instrument was developed by Peter Fenrich. This instrument was validated by the validator with amount 88,19%, the value of Alpha Cronbach index for reliability are 0,98 that present a reliable value, different strength item has 55% that can be classified as a bad differentiation, also the value of a difficulty level is 80% that classified difficult.

Keywords: *three-tier diagnostic test*, instrumen quality.

PENDAHULUAN

Belajar merupakan perubahan perilaku yang relatif tetap karena adanya latihan dan pengalaman serta merupakan proses kognitif (Forgas dan Bower, 1987:53). Dalam proses belajar tidak lepas adanya kesulitan belajar untuk mengetahui kesulitan belajar dan hasil belajar, dilakukan evaluasi atau penilaian yang merupakan proses lanjutan. Baik buruknya hasil belajar dan kesulitan belajar dapat dilihat dari hasil evaluasi. Salah satu contoh dari hasil evaluasi adalah berdasarkan wawancara guru SMAN 1 Soko Mojokerto pada bulan april 2016 yang menyatakan ulangan harian fisika yang mencapai KKM 75%. Hal ini mencerminkan rata-rata siswa di SMAN 1 Soko Mojokerto telah mencapai tujuan pembelajaran yang dibuat oleh guru akan tetapi, di dalam proses mencapai kompetensi dasar siswa masih mengalami kesulitan belajar. Ulangan harian yang diberikan kebanyakan hanya pengaplikasian (ranah C_3). Hasil tersebut belum mencerminkan tentang pemahaman konsep siswa secara menyeluruh dan belum dapat mendeteksi letak kesulitan belajar siswa. Berdasarkan contoh tersebut dapat disimpulkan bahwa evaluasi disusun mengacu pada tujuan belajar. Tujuan belajar merupakan pokok dari kegiatan belajar mengajar akan tetapi untuk mengetahui kesulitan belajar diperlukan alat evaluasi khusus yakni tes diagnostik.

Tes diagnostik adalah tes yang digunakan untuk mengetahui kelemahan-kelemahan siswa sehingga dapat dilakukan penanganan yang tepat (Suharsimi, 2005:34). Strategi untuk mengetahui kesulitan belajar terlebih dahulu harus mengetahui penyebab kesulitan tersebut. Salah satu contohnya adalah miskonsepsi dan tidak tahu

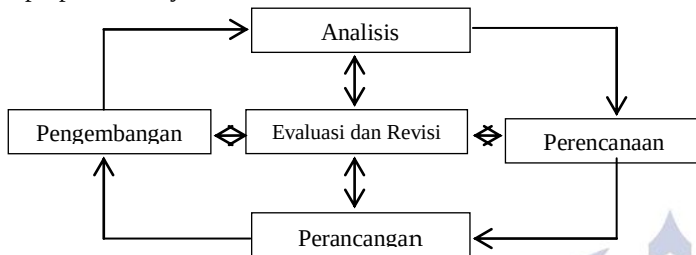
konsep. Hingga saat ini sangat sulit membedakan antara miskonsepsi dan tidak tahu konsep, apabila keduanya tidak dibedakan maka akan sulit untuk menanggunalangnya Tes diagnostik ini digunakan untuk membedakan dari paham konsep, paham konsep sebagian, menebak, miskonsepsi, tidak paham konsep dan tidak tahu konsep dan penyebab miskonsepsi yang berasal dari siswa. Tes diagnostik yang akan digunakan adalah tipe *three-tier* karena tipe ini dapat mengidentifikasi konsepsi siswa secara lengkap dan efisien waktu.

Berdasarkan penelitian Alwan (2011) menyatakan miskonsepsi topik suhu dan kalor responden sukar mengidentifikasi perbedaan suhu dan kalor. Didukung pula hasil penelitian dari Eryilmaz (2010) yang bahwa miskonsepsi untuk materi suhu dan kalor di Turkey pada jenjang Stara 1 yaitu 5% sub bab suhu dan 48% sub bab kalor *master student* 4% sub bab suhu dan 49% sub bab kalor dan untuk program doktor 7% sub bab suhu dan 40% sub bab kalor. Dari hasil penelitian-penelitian tersebut menyatakan bahwa materi suhu, kalor dan perpindahan kalor hampir semua siswa memiliki pemahaman konsep sendiri bahkan masih banyak yang mengalami miskonsepsi. Berdasarkan paparan tersebut maka dilakukan pengembangan instrumen soal pada materi suhu, kalor, dan perpindahannya.

METODE

Jenis penelitian ini merupakan penelitian pengembangan dengan metode pengembangan yang dikemukakan oleh Peter Fenrich. Dalam penelitian ini

mengembangkan instrumen tes diagnostik tipe *three-tier* pada materi suhu, kalor, dan perpindahannya. Penelitian ini diawali dengan analisis dan diakhiri dengan analisis kembali dalam satu siklus. Tujuan penelitian ini adalah menganalisis kualitas instrumen yang telah dikembangkan sehingga dapat menganalisis miskonsepsi siswa pada materi suhu, kalor, dan perpindahannya.



Gambar: Model siklus instruksional pengembangan oleh Fernrich

Sumber: Fernrich, 2004:43

Penelitian ini dilakukan di SMA Negeri 1 Sooko Mojokerto pada semester genap tahun ajaran 2015/2016 dan pengambilan data dilaksanakan pada bulan April 2014. Sampel penelitiannya adalah 2 kelas sebanyak 50 siswa untuk uji coba 1 yakni mengukur kualitas instrumen dan satu kelas sebanyak 30 siswa untuk uji coba 2 yakni untuk mendiagnosis miskonsepsi siswa.

Pengambilan data dilakukan dengan menggunakan lembar tes diagnostik. Hasil tersebut digunakan untuk mendeskripsikan kualitas instrumen yang telah dikembangkan. Kualitas instrumen dilihat dari validitas, reliabilitas, daya beda, dan taraf kesukaran. Validitas tes dibagi menjadi dua yakni validitas internal dan validitas eksternal. Validitas internal dianalisis menggunakan persentase dengan rumus sebagai berikut:

$$P = \frac{S_R}{N \times P_A \times R} \times 100\% \quad (1)$$

Keterangan:

P : Persentase validitas

S_R : Jumlah skor responden

N : Skor tertinggi dalam angket

P_A : Jumlah pertanyaan dalam angket

R : Jumlah responden

interpretasi hasil persentase seperti pada Tabel 1

Tabel 1 Interpretasi hasil validitas internal

Persentase	Kriteria Interpretasi Skor
0%-39%	Tidak valid
40%- 69%	Cukup valid
70%-89%	Valid
90%-100%	Sangat valid

Sumber: Diadaptasi dari Riduwan (2012:15)

Validitas tiap butir soal dianalisis menggunakan SPSS (*subject package for social science*). Valid secara empiris konstan jika memenuhi *false negatives* dan kurang *false positives* dari 10%. Validitas empiris konstruk dianalisis dengan menggunakan *pearson product moment* (r).

Uji reliabilitas tes dengan menggunakan alpha cronbach karena bentuk tes yang diujikan berupa *multiple choice*. Uji taraf kesukaran soal dengan menggunakan persamaan (2)

$$p = \frac{\sum X}{S_m N} \quad (2)$$

Keterangan:

P : indeks kesukaran atau proporsi menjawab benar

$\sum X$: jumlah mahasiswa yang menjawab benar

S_m : skor maksimum

N : jumlah siswa

Indeks kesukaran soal dapat ditentukan dengan kriteria berdasarkan Tabel 2

Tabel 2 Kriteria indeks kesukaran soal

Rentang nilai p	Interpretasi
$0,00 \leq p \leq 0,30$	Butir soal sukar
$0,31 \leq p \leq 0,70$	Butir soal sedang
$0,71 \leq p \leq 1,00$	Butir soal mudah

Sumber : Suharsimi (2013:225)

Uji daya beda dapat dianalisis dengan menggunakan persamaan (3)

$$D = \frac{B_A}{J_A} - \frac{B_B}{J_B} \quad (3)$$

Keterangan :

D : indeks diskriminasi

J_A : jumlah mahasiswa tes kelompok atas

J_B : jumlah mahasiswa tes kelompok bawah

B_A : jumlah kelompok atas yang menjawab benar pada soal tertentu

B_B : jumlah kelompok bawah yang menjawab benar pada soal tertentu

Hasil perhitungan indeks diskriminasi di atas, dapat diklasifikasikan daya beda untuk tiap soal seperti pada Tabel 3

Tabel 3 : Klasifikasi daya pembeda

Rentang nilai D	Interpretasi
0,00 – 0,20	Jelek
0,21 – 0,40	Cukup
0,41 – 0,70	Baik
0,71 – 1,00	Baik sekali

Sumber : Suharsimi (2013:232)

HASIL DAN PEMBAHASAN

Hasil penilaian validasi oleh validator dalam bentuk persentase dapat ditabulasikan pada Tabel 4

Tabel 4: Persentase penilaian oleh validator

Elemen yang divalidasi	Persentase	Kriteria
Materi	87,5%	Valid
Konstruk	89,5%	Valid
Bahasa	87,5%	Valid
Rerata	88,19%	Valid

Berdasarkan Tabel 4 menggambarkan bahwa validitas internal berdasarkan penilaian dan telaah validator adalah valid berdasarkan kesesuaian materi dengan kompetensi dasar, konstruksi dan bahasa

Hasil uji coba soal di SMAN 1 Sooko Mojokerto dengan 50 responden dapat digunakan untuk mengukur validitas eksternal. Dari 20 soal yang dikembangkan dinyatakan valid dengan menggunakan SPSS sebanyak 19 butir soal. Hasil perhitungan validitas empiris konten diperoleh *false negatives* 5.9% dan *false positives* 6.1% sehingga dikatakan valid secara empiris konten. Hasil tersebut dapat ditabulasikan pada Tabel 5

Tabel 5: Jumlah False Positives (FP) dan false negatives (FN) tiap soal

No. Soal	FN	FP	No. Soal	FN	FP
1	1	1	11	2	6
2	0	0	12	0	1
3	7	1	13	4	3
4	1	3	14	3	2
5	3	9	15	5	5
6	6	3	16	2	4
7	6	1	17	0	5
8	0	4	18	0	5
9	1	1	19	12	1
10	4	3	20	2	3
Jumlah	59	61			
Persentase	5,9%	6,1%			

Butir soal nomor 2 tidak terdapat *false negatives* dan *false positives* hal ini mencerminkan bahwa siswa lebih sedikit mengalami miskonsepsi sehingga siswa telah paham konsep, disisi lain tidak seutuhnya siswa paham konsep saja akan tetapi masih terdapat tidak paham konsep dan kurang paham konsep. Sedangkan untuk butir soal nomor 19 banyak siswa yang miskonsepsi dikarenakan *false negatives* siswa sudah dapat menjelaskan alasan yang mendukung jawaban akan tetapi salah dalam memilih jawaban sehingga siswa dapat dikatakan ceroboh. Sedangkan untuk butir soal nomor 5 paling banyak mengalami *false positives* hal ini dapat dikatakan siswa sebenarnya sudah paham konsep tersebut akan tetapi tidak dapat menjelaskan alasan yang tepat.

Pengujian validitas empiris konstruk yaitu dengan cara mengkorelasikan skor siswa yang menjawab benar pada *two-tier* dengan *confident tier*. Diperoleh diperoleh koefisien korelasi bernilai positif yang besarnya $r_{xy} = 0,43$ sehingga skor *confident tier* lebih tinggi (Kirbulut, 2014:517) dan validitas empiris konstruk tergolong sedang.

Diperoleh indeks Alpha Cronbach sebesar 0,98 dengan jumlah soal 20 soal sehingga 20 soal instrumen tes diagnostik dapat diterima atau reliabel.

Uji daya beda dengan menggunakan 27% kelas atas dan 27% kelas bawah diperoleh 55% dengan kriteria jelek yang berarti tidak dapat membedakan siswa pandai dan kurang pandai di kelas X MIA 3 SMAN 1 Sooko Mojokerto, meskipun diperoleh indeks diskriminan rendah tetap dapat digunakan (Suwanto, 2013:124).

Uji tingkat kesukaran soal diperoleh indeks 80% dikategorikan dalam kategori sulit, meskipun sulit instrumen suhu, kalor, dan perpindahannya tetap dapat digunakan (Abayneh, 2012:27).

Berdasarkan hasil uji taraf kesukaran dan daya beda soal fungsi tes diagnostik yang dikembangkan ini berada pada tes diagnostik 4, yang berarti tes diagnostik ini dapat digunakan untuk mengetahui tingkat penguasaan siswa terhadap bahan yang diajarkan sehingga tes diagnostik yang telah dikembangkan ini memiliki kesamaan dengan tes formatif (Suharsimi, 2013:49). Dapat dikatakan bahwa pengujian daya beda dan taraf kesukaran soal berdasarkan kemampuan responden, sehingga apabila di ujikan di tempat berbeda maka akan memperoleh hasil yang berbeda. Meskipun daya beda dan taraf kesukaran jelek dan sukar tetap digunakan karena fungsi tes diagnostik adalah untuk mengetahui kesulitan belajar siswa bukan untuk membedakan kemampuan antar siswa (Suwanto, 2013:124).

Rekapitulasi hasil pengujian kualitas instrumen berdasarkan validitas, taraf kesukaran soal, daya beda dengan reliabilitas 0,98. Dapat ditabulasikan seperti Tabel 6.

Tabel 6: Rekapitulasi kualitas butir soal

No. Soal	Validitas	Taraf kesukaran	Daya beda
1	Valid	Sedang	Cukup
2	Valid	Sukar	Jelek
3	Valid	Sukar	Cukup
4	Valid	Sedang	Cukup
5	Valid	Sukar	Jelek
6	Valid	Sukar	Jelek
7	Valid	Sedang	Cukup
8	Valid	Sukar	Jelek
9	Valid	Sukar	Cukup
10	Valid	Sukar	Cukup
11	Valid	Sukar	Jelek
12	Valid	Sukar	Jelek
13	Tidak Valid	Sukar	Jelek
14	Valid	Sedang	Jelek
15	Valid	Sukar	Jelek
16	Valid	Sukar	Cukup
17	Valid	Sukar	Jelek
18	Valid	Sukar	Jelek
19	Valid	Sukar	Cukup
20	Valid	Sukar	Cukup

SIMPULAN

Kualitas instrumen tes diagnostik tipe *three-tier* yang telah dikembangkan dinyatakan baik dengan rincian sebagai berikut:

- Validitas internal yang divalidasi oleh dua dosen ahli sebesar 88,19% tergolong kategori valid.
- Validitas eksternal yakni uji validitas empiris konstruk menggunakan *pearson product moment* diperoleh $r_{xy} = 0,43$ tergolong sedang dan hasil uji validitas empiris konstruk diperoleh *false positives* 6,1% dan *false negatives* 5,9% maka instrumen tersebut dikatakan valid.
- Uji reliabilitas dengan menggunakan Alpha Cronbach sebesar 0,98 yang artinya dapat diterima (reliabel).
- Uji taraf kesukaran 80% tergolong sukar dan uji daya beda 55% tergolong jelek.

DAFTAR PUSTAKA

- Abayneh. 2012. Diagnosing The Diagnostics: Misconceptions of Twelfth Grade Students on Selected Chemistry Concepts in Two Preparatory Schools in Eastern Ethiopia. *ECJE* .2 (2)
- Azwar, Saifudin.1998. *Sikap Manusia Teori Dan Pengukurannya*. Liberty: Yogyakarta
- Alwan, Almadi Ali. 2010. *Misconception Of Heat and Temperature Among Physic Students*, (Online https://www.academia.edu/3585334/MISCONCEPTION_IN_PHYSICS_AMONG_HIGH_SCHOOL_STUDENTS, diunduh pada 27 oktober 2015)
- Depdiknas. 2008. *Kamus Besar Bahasa Indonesia*. Gramedia : Jakarta
- Dindar, Alya Cetin dan Geban, Omer. 2011. Development of A Three-Tier Test to Assess High School Students' Understanding of Acids and Bases. *Procedia Social And Behavior Sciences*. 15(2011),600-604
- Eryilmaz, Ali. *Development and Application of Three-Tier Heat and Temperature Test: Sample of Bachelor And Graduate Student*, (online, <http://www.ejmste.com/Makale.aspx?kimlik=1984>, diunduh pada 7 November 2105)
- Fenrich, Peter. 2004. Instructional Design Tips For Virtual Teaching Practical Skills. *Proceeding Of The 2004 Informing Science and IT Education Joint Conference*
- Georgio, Helen. 2010. *A Report on Preliminary Diagnostic for Identifying Thermal Physics Conception of Tertiary Student*, (online, <http://openjournals.library.usyd.edu.au/index.php/CAL/article/viewFile/4700/5482>, diunduh pada 31 oktober 2015)
- Gurel, Derya Kaltakci; Ali Eryilmaz; Lilian Christie McDermott. 2015. A Review And Comparison of Diagnostic Instrument to Identify Students' Misconception In Science. *Eurasia Journal of Mathematics, Science & Technology Education*. 11 (5), pp 989-1008
- Hastenes, David dan Ibrahim Halloun. 1995. Interpreting The Force Concept Inventory: A Response To Huffman And Heller. *The Physics Teacher*. 33, pp.502-506
- Kirbulut Demet, Zubeyde dan Geban, Omer. 2014. Using Three-Tier Diagnostic Test To Assess Student Misconceptions Of States Mater. *Eurasia Journal of Mathematics & Technology Education*. 10(5), 509-521
- Riduwan. 2012. *Skala Pengukuran Variabel-Variabel Penelitian*. Bandung: Alfabeta
- Sudjana. 2005. *Metode Statistika Edisi 6*. Bandung: Tarsito
- Suparno, Paul. 2013. *Miskonsepsi Perubahan Konsep Dalam Pendidikan Fisika*. Jakarta : Grasindo
- Suwarto, 2013. *Pengembangan Tes Diagnostic Dalam Pembelajaran*. Yogyakarta: Pustaka Belajar.
- Suharsimi Arikunto. 2013. *Dasar – Dasar Evaluasi Pendidikan Edisi 2*. Jakarta: Bumi Askara
- Sugiyono. 2011. *Statistika Untuk Penelitian*. Bandung: Alfabeta
- Sugiyono. 2015. *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta
- Türker, Fatma. 2005. *Developing a Three-Tier Test to Assess High School Students' Misconceptions Concerning Force and Motion*. Tesis. Middle East Technical University