*Jurnal Inovasi Pendidikan Fisika (JIPF)* *Vol. 05 No. 03, September 2016, 196-201*
*ISSN: 2302-4496*

# Pengembangan *Tree-Tier Diagnostic Test* pada Materi Gelombang Mekanik

**Nurul Fajriani Lutfiyah , Woro Setyarsih**
Pendidikan Fisika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Surabaya
Email: nurullutfiyah9982@gmail.com.

**Abstrak**

Salah satu hal yang harus dilaksanakan seorang guru adalah mengembangkan instrumen penilaian. Guru cenderung menggunakan tes formatif yang memiliki kelemahan tidak dapat mengetahui penyebab kegagalan siswa dalam memahami konsep fisika. Untuk mengetahui kelemahan tersebut perlu dilakukan sebuah tes yang dinamakan tes diagnostik Tinjauan ini perlu dilakukan penelitian agar diketahui kualitas instrumen soal yang dikembangkan. Metode penelitian yang digunakan adalah deskriptif kualitatif. Instrumen *three-tier diagnostic test* tersebut merupakan hasil adaptasi dari *The Wave Diagnostic Instrument* (WADI) dan *Mechanical Wave Conceptual Survey* (MWCS). Diperoleh data bahwa validitas internal dinyatakan valid dengan rerata 82,3%. Hasil validitas eksternal hanya 16 soal yang dinyatakan valid dari 25 soal yang diujikan. Indeks Alpha Cronbach untuk reliabilitas instrumen adalah $\alpha = 0{,}811$ yang dinyatakan reliabel. Taraf kesukaran soal tinggi dengan daya beda yang jelek ($D$ = 0,17).

**Kata Kunci:** *tree-tier diagnostic test*, pengembangan instrumen.

**Abstract**

One of the matter that teachers have should be done is instrument development. Teachers disposed use formatif test that posses weakness. It can't detect of reviewing causes the students failure in understanding physics concept. To detect that weakness needed a test called diagnostic test. It is important to know the quality of developing instrument. This research is using descriptive qualitative method. The instrument of three-tier diagnostic test is adopted by The Wave Diagnostic Instrument (WADI) and Mechanical Wave Conceptual Survey (MWCS). The result show that the internal validity is valid by the average as much as 82,3%. For external validity only 16 item expressed valid out of 25 item tested. The reliability index of Alpha cronbach is $\alpha = 0{,}811$ and it's expressed reliable. The difficulty levels are high and the discrimination index is poor ($D$ = 0,17).

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

## PENDAHULUAN

Mengacu pada Permendiknas no. 16 tahun 2007 tentang standar kualifikasi dan kompetensi guru, salah satu kompetensi pedagogi (ilmu pengajaran) yang harus dimiliki guru menjelaskan tentang penyelenggaraan penilaian dan evaluasi proses hasil belajar. Salah satu hal yang harus dilaksanakan seorang guru adalah mengembangkan instrumen penilaian dan evaluasi proses hasil belajar. Berdasarkan aspek pelaksanaan tes dari Permendikbud no. 66 (2013:5), guru cenderung menggunakan tes formatif yang bertujuan untuk mengetahui keberhasilan dalam proses pembelajaran. Kelemahan dari tes tersebut tidak dapat mengetahui sebab kegagalan siswa dalam belajar (Suwarto, 2013:5). Untuk mengetahui kelemahan-kelemahan siswa agar dapat dilakukan penanganan yang tepat, perlu dilakukan sebuah tes yang dinamakan tes diagnostik (Arikunto, 2013:50).

Data Puspendik pada hasil Ujian Nasional tahun 2015 mata pelajaran fisika jenjang SMA tingkat Kabupaten Gresik, dari lima kemampuan yang diujikan, penguasaan materi gelombang dan optik berada di urutan keempat yakni sebesar 74,36%. Penguasaan materi tertinggi terdapat pada materi pengukuran dan mekanika sebesar 81,52%. Sedangkan penguasaan materi terendah terdapat pada materi fisika modern sebesar 48,22%.

Siswa belajar secara formal melalui pendidikan sekolah atau informal melalui pengalaman sehari-hari sehingga siswa membentuk pengetahuannya sendiri (Kaltakci, dkk., 2015:989). Adakalanya pengetahuan yang dibentuk siswa berbeda dengan konsep yang dianut oleh para ahli, yang dikenal sebagai miskonsepsi (Suparno, 2013:4). Menurut Kirbulut dan Geban (2014:510), penelitian fisika mengenai pengembangan dan aplikasi three-tier diagnostic test juga masih sedikit.

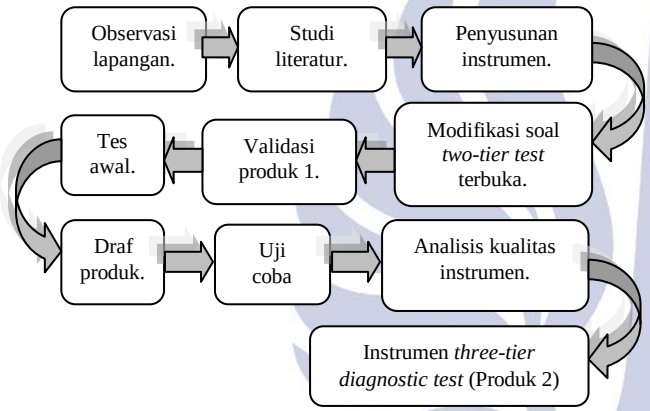
Pada instrumen *three-tier diagnostic test*, tingkatan pertama diberikan tes pilihan ganda untuk menjawab suatu fenomena. Tingkatan kedua, diberikan pilihan ganda yang terdiri dari alasan yang menguatkan jawaban fenomena tersebut. Tingkatan ketiga, diujikan untuk mengetahui tingkat kepercayaan diri siswa dalam menjawab pertanyaan pada dua tingkatan tes sebelumnya (Caleon dan Subramaniam, 2010:941; Kaltakci dan Eryilmaz, 2010:2; Kirbulut, 2014:510).

Berdasarkan uraian di atas, maka diperlukan penelitian pada pengembangan instrumen evaluasi siswa untuk mengetahui miskonsepsi siswa. oleh karena itu dilakukan penelitian dengan judul "Pengembangan

Instrumen *Tree-Tier Diagnostic Test* pada Materi Gelombang Mekanik". Penelitian dilakukan dengan adaptasi instrumen *The Wave Diagnostic Instrument* (WADI) dan *Mechanical Wave Conceptual Survey* (MWCS) agar diketahui kualitas instrumen yang akan digunakan dalam mengidentifikasi miskonsepsi siswa pada materi gelombang mekanik.

## METODE

Jenis penelitian yang digunakan adalah deskriptif kuantitatif. Penelitian deskriptif merupakan metode yang digunakan untuk menggambarkan keadaan variabel dan tidak mengadakan perubahan pada variabel-variabel bebas, tetapi menggambarkan kondisi apa adanya. Penggambaran kondisi bisa individual atau kelompok dan menggunakan angka-angka (Sukmadinata, 2015:54). Tahapan penelitian yang dilakukan adalah sebagai berikut.




Gambar 1. Tahapan kegiatan penelitian.

Penelitian ini dilakukan di SMA Negeri 1 Driyorejo. Tahap pelaksanaan penelitian dilakukan pada bulan Mei 2016 yakni pada semester genap tahun pelajaran 2015-2016. Uji coba dilakukan untuk menentukan kualitas instrumen dengan sampel 40 siswa. Lembar instrumen *three-tier diagnostic test* yang digunakan, telah diklasifikasikan menggunakan taksonomi Bloom lama. Lembar instrumen tes meliputi materi gelombang mekanik (gelombang pada tali dan gelombang bunyi).

Teknik analisis yang digunakan adalah;

### A. Uji Validitas Internal

Untuk menghitung validasi dari para ahli dibuat persentase dengan rumus:

$$V = \frac{\sum N}{T \times \sum P \times \sum R} \times 100\% \qquad (1)$$

(Sugiyono, 2015: 143-144)

Keterangan:

$V$ = persentase validitas instrumen.
$\sum N$ = jumlah penilaian validator.
$T$ = skor tertinggi dalam penilaian.
$\sum P$ = jumlah pertanyaan.
$\sum R$ = jumlah responden atau validator.

Adapun kriteria interpretasi skor hasil persentase adalah:

Tabel 1 Kriteria Interpretasi Skor.

| Persentase | Kriteria |
|---|---|
| 0% - 25% | Tidak Valid |
| 26% - 50% | Cukup Valid |
| 51% - 75% | Valid |
| 76% - 100% | Sangat Valid |

(Sugiyono, 2015:144)

### B. Uji Validitas Eksternal

#### 1. Uji Validitas Instrumen

Uji validitas instrumen *three-tier diagnostic test* menggunakan *t-test* dengan *pooled varian* dengan derajat kebebasan $(dk) = n_1 + n_2 - 2$.

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \qquad (2)$$

(Sugiyono, 2015:273)

Keterangan:

$\overline{X}_1$ = rerata skor kelompok atas.
$\overline{X}_2$ = rerata skor kelompok bawah.
$n_1$ = jumlah anggota kelompok atas.
$n_2$ = jumlah anggota kelompok bawah.
$s_1^2$ = varians kelompok atas.
$s_2^2$ = varians kelompok bawah.

Menindak lanjuti bukti empiris dari validitas isi dan konstruk berdasarkan hasil uji coba 1, maka dapat ditentukan sebagai berikut.

#### a. Validitas Empiris Konten

Uji validitas empiris konten *three-tier diagnostic test* menggunakan identifikasi persentase dari *false negatives* dan *false positives* (Syahrul, 2015:51). Menurut Hestenes dan Halloun (1995:3) validitas empiris konten suatu instrumen akan terpenuhi jika false negatives kurang dari 10%.

#### b. Validitas Empiris Konstruk

Validitas empiris konstruk dihitung menggunakan persamaan korelasi *pearson product moment (r).*

$$r_{xy} = \frac{\sum xy}{\sqrt{(\sum x^2)(\sum y^2)}} \qquad (3)$$

(Arikunto, 2013:85)

Keterangan:

$x$ = selisih dari jumlah keseluruhan jawaban benar pada *two-tier* tiap item soal dengan rerata jawaban benar pada semua item soal.
$y$ = selisih dari jumlah keseluruhan jawaban yakin pada tingkat ketiga tiap item soal

dengan rerata jawaban yakin pada semua item soal.

Hasil korelasi dari *pearson product moment* dapat dikategorikan seperti pada Tabel 2.

Tabel 2. Kriteria Korelasi *Pearson Product Moment.*

| Interval Koefisien | Tingkat Hubungan |
|---|---|
| 0,000 – 0,199 | Sangat rendah |
| 0,200 – 0,399 | Rendah |
| 0,400 – 0,599 | Sedang |
| 0,600 – 0,799 | Kuat |
| 0,800 – 1,000 | Sangat kuat |

(Sugiyono, 2015:257)

**C. Uji Reliabilitas Instrumen**

Menurut Suwarto (2013:102), reliabilitas tes bentuk *multiple choice* menggunakan perhitungan rumus Alpha Cronbach.

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum s^2{}_i}{s^2{}_x}\right) \quad (4)$$

(Suwarto, 2013:102)

Keterangan:
$\alpha$ = indeks Alpha Cronbach
$k$ = jumlah butir tes.
$\sum s^2{}_i$ = jumlah semua varian butir pembentuk tes.
$s^2{}_x$ = varian skor total tes.

Hasil indeks Alpha Cronbach dapat ditentukan kriterianya seperti tabel 3.

Tabel 3. Kriteria Indeks Alpha Cronbach

| Rentang Nilai α | Kriteria |
|---|---|
| 0,00 – 0,50 | Tidak dapat diterima |
| 0,51 – 0,60 | Jelek |
| 0,61 – 0,70 | Diragukan |
| 0,71 – 1,00 | Dapat diterima |

Diadaptasi dari Türker (2005:43)

**D. Uji Tingkat Kesukaran Butir Soal**

Persamaan yang digunakan untuk menentukan tingkat kesukaran dengan proporsi menjawab benar adalah:

$$p = \frac{\sum x}{S_m N} \quad (5)$$

(Suwarto, 2013:105)

Keterangan:
$p$ = tingkat kesukaran.
$\sum x$ = banyaknya siswa yang menjawab benar.
$S_m$ = nilai maksimum.
$N$ = jumlah peserta tes.

Indeks kesukaran sering diklasifikasikan seperti Tabel 4.

Tabel 4. Klasifikasi Kesukaran Butir Soal.

| Rentang Nilai $p$ | Interpretasi |
|---|---|
| 0,00 – 0,30 | Soal sukar |
| 0,31 – 0,70 | Soal sedang |
| 0,71 – 1,00 | Soal mudah |

( Arikunto, 2013:225)

**E. Uji Daya Beda Butir Soal**

Indeks daya pembeda yang digunakan adalah:

$$D = \frac{\sum A}{n_A} - \frac{\sum B}{n_B} \quad (6)$$

(Suwarto, 2013:109)

Keterangan:
$D$ = Indeks daya pembeda.
$\sum A$ = Banyaknya siswa yang menjawab benar pada kelompok atas.
$\sum B$ = Banyaknya siswa yang menjawab benar pada kelompok bawah.
$n_A$ = banyaknya siswa kelompok atas.
$n_B$ = banyaknya siswa kelompok bawah.

Klasifikasi daya pembeda dapat dilihat pada Tabel 5.

Tabel 5. 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 |

(Arikunto, 2013:232)

## HASIL DAN PEMBAHASAN

Berdasarkan penelitian yang dilakukan, diperoleh hasil sebagai berikut:

**1. Validitas Tes Diagnostik**

**a. Validitas Internal**

Pada tahap ini, instrumen soal divalidasi oleh dua dosen fisika Unesa. Adapun hasil penilaian validasi dari validator yakni:

Tabel 6. Rekapitulasi Penilaian Validitas Internal Tes Diagnostik.

| No. | Elemen yang Divalidasi | Persentase (%) | Kriteria |
|---|---|---|---|
| **Ranah Materi** | | | |
| 1 | Butir-butir soal sesuai dengan materi gelombang mekanik (gelombang tali dan gelombang bunyi). | 87,5 | Sangat Valid |
| 2 | Butir-butir soal yang digunakan sesuai dengan indikator soal. | 75 | Valid |
| 3 | Urutan tiap butir soal sesuai dengan urutan materi. | 87,5 | Sangat Valid |
| 4 | Batasan pertanyaan, jawaban, dan penjelasan alasan yang diharapkan jelas. | 87,5 | Sangat Valid |

UNESA Universitas Negeri Surabaya

| No. | Elemen yang Divalidasi | Persentase (%) | Kriteria |
|---|---|---|---|
| | **Rerata** | **84,4** | **Sangat Valid** |
| **Ranah Konstruk** | | | |
| 1 | Petunjuk tes diagnostik gelombang mekanik dinyatakan jelas. | 87,5 | Sangat Valid |
| 2 | Kesesuaian kriteria butir soal dengan Taksonomi Bloom dan kompetensi dasar. | 75 | Valid |
| 3 | Setiap butir tes diagnostik dapat mengidentifikasi konsepsi siswa. | 87,5 | Sangat Valid |
| 4 | Pilihan alasan yang disajikan dapat mengungkapkan penyebab miskonsepsi yang bersumber dari siswa. | 100 | Sangat Valid |
| 5 | Pengecoh pada pilihan alasan bersifat rasional dan homogen dengan jawaban tingkat pertama. | 100 | Sangat Valid |
| 6 | Tabel, grafik, gambar, dan sejenisnya bersesuaian dengan masalah yang disajikan. | 100 | Sangat Valid |
| | **Rerata** | **91,7** | **Sangat Valid** |
| **Ranah Bahasa** | | | |
| 1 | Kalimat soal menggunakan bahasa yang baik dan benar. | 62,5 | Valid |
| 2 | Ragam kalimat atau pernyataan tidak menimbulkan penafsiran ganda. | 62,5 | Valid |
| 3 | Pernyataan setiap butir tes dinyatakan jelas dan komunikatif. | 87,5 | Sangat Valid |
| | **Rerata** | **70,8** | **Valid** |

Berdasarkan Tabel 6 dapat diperoleh rerata validitas internal sebesar 82,3% sehingga penilaian validitas internal oleh validator adalah valid.

**b. Validitas Eksternal**

Uji validitas keseluruhan butir instrumen *three-tier diagnostic test* dilakukan menggunakan *t-test*. Didapatkan $t_{hitung} = 6{,}67$ sedangkan nilai $t_{tabel}$ dengan $dk$ = 20 dan taraf signifikan 5% adalah 2,086. Karena $t_{hitung} > t_{tabel}$ maka dapat disimpulkan bahwa instrumen *three-tier diagnostic test* sebanyak 25 soal yang telah dikembangkan adalah **valid**.

Uji validitas tiap butir soal dilakukan menggunakan pengujian SPSS *(Subject Package for Social Science)* yang mengorelasikan antara total skor tiap butir soal dari jawaban 40 siswa dengan total skor yang diperoleh tiap siswa. Hasil pengujian tersebut ditabulasikan pada Tabel 7.

Tabel 7. Hasil Pengujian Menggunakan SPSS.

| No. Soal | *Pearson Correlation* | Kriteria |
|---|---|---|
| 1 | 0,968 | Valid |
| 2 | 0,812 | Valid |
| 3 | 0,986 | Valid |
| 4 | 0,879 | Valid |
| 5 | 0,882 | Valid |
| 6 | 0,976 | Valid |
| 7 | [a] | Tidak Valid |
| 8 | 0,893 | Valid |
| 9 | [a] | Tidak Valid |
| 10 | [a] | Tidak Valid |
| 11 | [a] | Tidak Valid |
| 12 | 0,866 | Valid |
| 13 | [a] | Tidak Valid |
| 14 | 0,686 | Valid |
| 15 | 0,693 | Valid |
| 16 | 0,686 | Valid |
| 17 | 0,900 | Valid |
| 18 | [a] | Tidak Valid |
| 19 | 0,933 | Valid |
| 20 | [a] | Tidak Valid |
| 21 | [a] | Tidak Valid |
| 22 | [a] | Tidak Valid |
| 23 | 0,808 | Valid |
| 24 | 0,908 | Valid |
| 25 | 0,991 | Valid |

**Keterangan:** [a] merupakan tanda bahwa tidak dapat dilakukan perhitungan karena tidak ada satu siswa pun dari sampel yang dapat menjawab butir soal tersebut.

Berdasarkan Tabel 7, dari 25 soal yang diujikan hanya 16 soal yang dinyatakan valid yakni delapan soal untuk materi gelombang tali dan delapan soal untuk materi gelombang bunyi.

Uji validitas eksternal selanjutnya dikuatkan oleh uji validitas empiris konten dan konstruk sebagai berikut:

**1) Validitas Empiris Konten**

Pengujian empiris konten dilakukan dengan menggunakan *false positives* (FP) dan *false negatives* (FN). Validitas empiris konten akan terpenuhi atau dinyatakan valid

jika *false negatives* kurang dari 10%. Perhitungan validitas empiris konten dapat ditabulasikan seperti Tabel 8.

Tabel 8. Perhitungan *False Positives* (FP) dan *False Negatives* (FN) Tiap Butir Soal.

| No. Soal | FP | FN |
|---|---|---|
| 1 | 3 | 0 |
| 2 | 2 | 1 |
| 3 | 1 | 10 |
| 4 | 9 | 0 |
| 5 | 2 | 1 |
| 6 | 0 | 1 |
| 7 | 1 | 1 |
| 8 | 0 | 4 |
| 9 | 0 | 6 |
| 10 | 0 | 0 |
| 11 | 2 | 1 |
| 12 | 1 | 2 |
| 13 | 2 | 7 |
| 14 | 3 | 0 |
| 15 | 0 | 5 |
| 16 | 0 | 2 |
| 17 | 0 | 3 |
| 18 | 2 | 1 |
| 19 | 3 | 1 |
| 20 | 1 | 2 |
| 21 | 2 | 7 |
| 22 | 0 | 4 |
| 23 | 0 | 1 |
| 24 | 11 | 1 |
| 25 | 1 | 1 |
| **Jumlah** | **46** | **62** |
| **Persentase** | **4,6%** | **6,2%** |

Berdasarkan Tabel 8 diperoleh data *false positives* dan *false negatives* kurang dari 10%. Hal ini menyatakan bahwa instrumen *three-tier diagnostic test* materi gelombang mekanik adalah valid.

**2) Validitas Empiris Konstruk**

Pengujian validitas empiris konstruk dilakukan dengan menggunakan rumus korelasi *pearson product moment (r)* dimana mengorelasikan antara skor siswa yang menjawab benar pada *two-tier* dengan skor *confident tier*. Diperoleh hasil perhitungan $r_{xy} = 0{,}502$. Selanjutnya hasil tersebut dikategorikan seperti pada Tabel 2, sehingga validitas empiris konsruk tergolong sedang.

**2. Reliabilitas Tes Diagnostik**

Uji reliabilitas tes bentuk *multiple choice* menggunakan perhitungan rumus Alpha Cronbach (persamaan 4) untuk jumlah butir soal yang valid. Diperoleh indeks Alpha Cronbach $\alpha = 0{,}811$. Selanjutnya hasil tersebut dikategorikan seperti pada Tabel 3, sehingga dapat disimpulkan bahwa instrumen *three-tier diagnostic test* sebanyak 16 soal yang telah valid adalah dapat diterima (**reliabel**).

**3. Taraf Kesukaran Butir Soal**

Pengujian taraf kesukaran butir 16 soal yang valid dihitung menggunakan persamaan 5. Hasil pengujian tersebut dapat ditabulasikan seperti Tabel 9..

Tabel 9. Taraf Kesukaran Butir Soal yang Valid.

| No. Soal | Rentang Nilai *P* | Kriteria | Ranah Kognitif |
|---|---|---|---|
| 1 | 0,28 | Sukar | C4 |
| 2 | 0,10 | Sukar | C4 |
| 3 | 0,50 | Sedang | C4 |
| 4 | 0,10 | Sukar | C4 |
| 5 | 0,08 | Sukar | C4 |
| 6 | 0,40 | Sedang | C2 |
| 8 | 0,10 | Sukar | C4 |
| 12 | 0,08 | Sukar | C4 |
| 14 | 0,03 | Sukar | C5 |
| 15 | 0,03 | Sukar | C5 |
| 16 | 0,03 | Sukar | C2 |
| 17 | 0,10 | Sukar | C6 |
| 19 | 0,15 | Sukar | C2 |
| 23 | 0,05 | Sukar | C3 |
| 24 | 0,10 | Sukar | C3 |
| 25 | 0,55 | Sedang | C1 |

Berdasarkan data dari tabel tersebut, 81,25% soal dikategorikan sukar. Hal ini dapat mengindikasikan bahwa siswa masih belum sepenuhnya menguasai konsep gelombang mekanik dengan baik.

**4. Daya Beda Butir Soal**

Pengujian daya beda butir 16 soal yang valid dihitung menggunakan persamaan 6. Hasil pengujian daya beda dapat ditabulasikan seperti tabel 10.

Tabel 10. Daya Beda Butir Soal yang Valid.

| No. Soal | Indeks Diskriminasi | Kriteria |
|---|---|---|
| 1 | 0,36 | Cukup |
| 2 | 0,09 | Jelek |
| 3 | 0,10 | Jelek |
| 4 | 0,18 | Jelek |
| 5 | 0,27 | Cukup |
| 6 | 0,20 | Jelek |
| 8 | 0,00 | Jelek |
| 12 | 0,18 | Jelek |
| 14 | 0,09 | Jelek |
| 15 | 0,10 | Jelek |
| 16 | 0,10 | Jelek |
| 17 | 0,25 | Cukup |
| 19 | 0,45 | Baik |
| 23 | 0,00 | Jelek |
| 24 | 0,18 | Jelek |
| 25 | 0,11 | Jelek |
| **Rerata** | **0,17** | **Jelek** |

Berdasarkan Tabel 10, indeks diskriminasi yang diperoleh untuk tes diagnostik sebesar 0,17 sehingga dikategorikan jelek. Hal ini dapat diartikan bahwa instrumen soal tidak dapat membedakan antara siswa kelas atas dan kelas bawah. Menurut Suwarto (2013:124) instrumen tes diagnostik yang memiliki daya beda yang rendah tetap dapat digunakan.

## PENUTUP

### Simpulan

Instrumen yang disusun telah divalidasi dengan hasil sebagai berikut: Hasil validitas internal adalah valid dengan rerata penelitian sebesar 82,3%. Hasil validitas eksternal untuk tiap butir soal menggunakan pengujian SPSS hanya 16 soal yang dinyatakan valid dari 25 soal yang diujikan yakni delapan soal untuk materi gelombang tali dan delapan soal untuk materi gelombang bunyi. Indeks Alpha Cronbach untuk reliabilitas instrumen adalah $\alpha = 0{,}811$ yang dinyatakan reliabel. Taraf kesukaran soal dikategorikan sukar dan daya beda yang jelek ($D$ = 0,17).

### Saran

Sampel yang digunakan untuk pengujian sebaiknya dalam jumlah yang relatif besar. Hal tersebut dikarenakan semakin banyak sampel yang diambil, maka data yang akan diperoleh semakin akurat

## DAFTAR PUSTAKA

Arikunto, Suharsimi. 2013. *Dasar-Dasar Evaluasi Pendidikan*. Edisi kedua. Jakarta: Bumi Aksara.

Caleon, I., dan Subramaniam, R. 2010. "Development and Application of a Three-Tier Diagnostic Test to Assess Secondary Students' Understanding of Waves". *International Journal of Science Education*. Vol. 32 (7): pp. 939-961.

Hestenes, David and Ibrahim Halloun. 1995. Interpreting the Force Concept Inventory: A response to Huffman and Heller. *The Physics Teacher*. 33, pp. 502-506.

Kaltakci, D., Eryilmaz, A., dan McDermott, L.C. 2015. "A Review and Comparison of Diagnostic Instruments to Identify Students' Misconceptions in Science". *Eurasia Journal of Mathematics, Science & Technology Education*. Vol. 11 (5): pp. 989-1008.

Kemendikbud. 2015. *Laporan Hasil Ujian Nasional*, (Online), (http://un.kemdikbud.go.id/, diakses 7 Maret 2016).

Kirbulut, Z.D., dan Geban, O. 2014. "Using Three-Tier Diagnostic Test to Assess Students' Misconception of States of Matter". *Eurasia Journal of Mathematics, Science & Technology Education*. Vol. 10(5): pp. 509-521.

Mendiknas. 2007. *Standar Kualifikasi Akademik dan Kompetensi Guru*, (Online), (http://hukum.unsrat.ac.id/men/mendiknas_16_2007.pdf, diakses 15 Februari 2016)

Pertiwi, Chyta Anindya. 2015. *Profil Konsepsi Siswa Terhadap Hukum Newton dengan Menggunakan Instrumen Force Concept Inventory (FCI)*. Skripsi tidak diterbitkan. Surabaya: PPs Universitas Negeri Surabaya.

Sugiyono. 2015. *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.

Suhendi, Herni Y. 2014. *Penerapan Model ECIRR Berbantuan Media Simulasi Virtual untuk Meningkatkan Pemahaman Konsep dan Mengidentifikasi Miskonsepsi Siswa*. Tesis tidak diterbitkan. Bandung: PPs Uninersitas Pendidikan Indonesia.

Sukmadinata, Nana S. 2015. *Metode Penelitian Pendidikan*. Bandung: Remaja Roesdakarya Offset.

Suparno, Paul. 2013. *Miskonsepsi dan Perubahan Konsep dalam Pendidikan Fisika*. Jakarta: Gramedia.

Suwarto. 2013. *Pengembangan Tes Diagnostik dalam Pembelajaran*. Yogyakarta: Pustaka Pelajar.

Syahrul, D.A. 2015. *Diagnosis Konsepsi Siswa dengan Three-Tier Diagnostic Test Berbasis Revised Bloom's Taxonomy pada Materi Dinamika Rotasi*. Skripsi tidak diterbitkan. Surabaya: PPs Universitas Negeri Surabaya.

Türker, Fatma. 2005. *Developing a Three-Tier Test to Assess High School Students' Misconceptions Concerning Force and Motion*. Tesis. Middle East Technical University.

UNESA Universitas Negeri Surabaya