

Pengembangan dan Penyetaraan Instrumen Tes Kinematika Gerak Lurus dengan Metode Linear

Hafid Suyuti¹⁾, Sentot Kusairi²⁾, Sutopo²⁾

¹⁾SMKN 1 Blitar

²⁾Pendidikan Fisika–Universitas Negeri Malang

E-mail: hafidsuyuti2@gmail.com

Abstract: Implementation of assessment often takes several sets of questions. However, different tests will have a different difficulty indices even they were constructed on the same indicators. Equating or score converting in every test must be done to ensure the tests in same scale of measurement. Development of test instruments includes several steps: create multiple choice question base on students response of open ended question; consultation with experts to validation of contents, individual try out to assess aspects of language comprehension, limited field try out to assess the feasibility of items parameter and field try out to determine the conversion coefficients. In this study, equating method used is linear equating. According to the least error of equating is suggested using scoring conversion in test B (Y) to test A (X). Equation of conversion for first year highschool is: $X^* = 1,064 Y + 0,015$, for second year is $X^* = 0,834 Y + 0,455$, and for third year is $Y^* = 0,991 Y - 0,147$.

Key Words: test instrument, kinematics, straight motion, equating

Abstrak: Pelaksanaan penilaian membutuhkan beberapa set soal yang berbeda untuk mengukur kemampuan yang sama. Soal-soal tersebut memiliki tingkat kesukaran yang berbeda dan didasarkan pada kisi-kisi yang sama. Penyetaraan skor antar soal perlu dilakukan untuk memastikan set-set soal berada dalam skala pengukuran yang sama. Pengembangan dan penyetaraan instrumen tes meliputi tahapan pengembangan soal pilihan ganda yang didasarkan pada respon siswa pada soal uraian, validasi isi, uji coba perorangan aspek pemahaman bahasa, lapangan terbatas untuk menilai kelayakan soal dan uji coba lapangan untuk menentukan koefisien-koefisien penyetaraan dengan metode penyetaraan penyetaraan linear. Dengan memperhatikan nilai kesalahan baku penyetaraan yang terkecil disarankan proses konversi skor dilakukan dari skor soal kode B (Y) ke skor kode A (X) dengan persamaan konversi untuk kelas X adalah $X^* = 1,064 Y + 0,015$. Untuk kelas XI persamaan konversinya adalah $X^* = 0,834 Y + 0,455$. Untuk kelas XII persamaan konversinya adalah $Y^* = 0,991 Y - 0,147$.

Kata kunci: instrumen tes, kinematika, gerak lurus, penyetaraan

Bentuk tes pilihan ganda banyak digunakan dalam penilaian secara luas karena memiliki banyak keunggulan. Keunggulan yang paling mencolok dari tes ini adalah kemudahannya dalam proses koreksi (Reynolds dkk, 1999). Tes pilihan ganda juga menawarkan kemampuan cakupan (*diversity*) materi pelajaran yang luas (Osterlind, 1998; Kusairi, 2013). Lebih lanjut, pada tes pilihan ganda dapat dilakukan interpretasi yang cukup presisi

terkait validitas skor yang diperoleh peserta tes (Osterlind, 1998).

Selain keunggulan-keunggulan tersebut, tes pilihan ganda juga memiliki beberapa kelemahan. Beberapa ahli berpendapat bentuk tes uraian lebih baik dalam mengukur kognitif tingkat tinggi (Ebel & Frisbie, 1991), namun butir soal pilihan ganda dapat mengukur kemampuan berfikir tingkat tinggi jika dikonstruksi dengan baik (Osterlind, 1998). Selain itu,

pengembangan butir soal pilihan ganda yang baik membutuhkan keahlian khusus dan waktu penyusunan yang panjang (Hestenes & Wells, 1992; Hestenes dkk, 1992; Beichner, 1994; Rebello & Barrow, 2012). Bentuk tes pilihan ganda menyediakan pilihan jawaban sehingga dipandang mudah ditebak dan dicontek oleh siswa (Kusairi, 2013) serta membatasi peserta didik untuk mengekspresikan kreativitas atau mendemonstrasikan pemikiran imajinatif (Osterlind, 1998).

Permasalahan utama ketika suatu pengukuran dilakukan dengan menggunakan dua set soal yang berbeda adalah berbedanya tingkat kesukaran. Dua buah perangkat pengukuran yang berbeda akan memiliki tingkat kesukaran yang berbeda walaupun kedua perangkat didasarkan pada kisi-kisi yang sama (Hambleton dkk, 1991; Kartono, 2008). Konsep soal yang sama namun disajikan dalam konteks yang berbeda akan mempengaruhi kesulitan butir soal (Planinik, dkk., 2013). Lebih lanjut, skor yang dihasilkan pada pengukuran kemampuan awal dan akhir dengan menggunakan soal dengan tingkat kesukaran berbeda tidaklah menunjukkan perubahan kemampuan yang sebenarnya. Oleh karena itu perlu dilakukan proses penyetaraan sehingga skor-skor tes tersebut memiliki skala tunggal (*common scale*) (Kolen & Brennan, 1995).

Penelitian tentang penyetaraan telah banyak dilakukan. Penelitian-penelitian ini mencakup bagaimana melakukan penyetaraan tes pilihan ganda (Rustam, 2000; Herukusumo, 2011), penyetaraan tes uraian (Kartono, 2006) dan penyetaraan tes campuran (Kartono, 2008) pada tes prestasi belajar. Sugeng (2010) melakukan penyetaraan vertikal model kredit parsial pada soal matematika SMP. Beberapa peneliti yang lain melakukan penelitian tentang pengaruh metode (Liu, 2008; Moses, 2007; Tsai, 2001), pengaruh ukuran sampel terhadap variansi skor hasil penyetaraan (Puhan, 2008) dan pengaruh perulangan terhadap penyetaraan skor (Kim, 2012).

Penelitian tentang penggunaan penyetaraan dalam pelaksanaan penilaian dalam penelitian Pendidikan Fisika masih sulit ditemukan. Sebagian besar, soal-soal tersebut disusun dalam satu paket soal. Beberapa contoh instrumen tes yang telah dikembangkan adalah FCI (Hestenes dkk, 1992), MBT (Hestenes & Wells, 1992), TUG-F (Beichner, 1994), MWCS (Tongchai dkk, 2009), dan FMCE (Rebello & Barrow, 2012). Topik kinematika gerak lurus sangat penting dalam mempelajari fisika lebih

lanjut. Topik ini sangat mendasar pada mata pelajaran fisika yang digunakan untuk landasan konsep fisika yang lain seperti gaya (Hestenes & Wells, 1992; Rosenblatt & Heckler, 2011). Kinematika melibatkan analisis grafik yang seringkali digunakan sebagai “bahasa pembelajaran” dalam fisika (Beichner, 1994). Oleh karena itu kinematika gerak lurus selalu dipelajari pada tahun pertama pada semua jenjang yang mempelajari fisika. Oleh karena itu, instrumen tes yang terdiri dari dua set soal yang dilanjutkan dengan proses penyetaraan skor topik kinematika gerak lurus dipilih untuk dikembangkan.

Terdapat dua pendekatan penyetaraan skor, penyetaraan dengan pendekatan klasik dan dengan pendekatan modern. Penyetaraan dengan pendekatan klasik dipandang paling cocok dilakukan di tingkat sekolah, karena pendekatan klasik tidak mensyaratkan jumlah *testee* yang terlalu banyak (Linn, 1989). Proses penyetaraan dengan pendekatan klasik dapat dilaksanakan dengan mempergunakan 2 metode, yaitu penyetaraan secara linear dan penyetaraan secara ekuipersentil (Hambleton dkk, 1991). Metode penyetaraan linear menghubungkan skor konversi dengan skor asalnya melalui suatu fungsi linear, sedangkan metode penyetaraan ekuipersentil mengkonversi skor berdasarkan jenjang persentil yang sama (Livingstone, 2004).

Terdapat tiga jenis rancangan penyetaraan tes, yaitu rancangan kelompok tunggal (*single group design*), rancangan kelompok acak (*random group design*), dan rancangan dengan butir tes jangkar (*anchor test group design*) (Von Davier, dkk., 2003). Pada *single group design*, satu kelompok sampel yang sama diuji dua kali dengan paket tes yang berbeda. Misal paket tes tersebut paket X dan paket Y. Paket X diadministrasikan pertama pada semua peserta tes, kemudian disusul dengan paket Y. Idealnya untuk menyetarakan skor dari beberapa perangkat tes, perangkat tes-perangkat tes tersebut diberikan pada responden yang sama. Kenyataan di lapangan, rancangan ini sulit dilakukan karena adanya faktor kelelahan, dan adanya faktor latihan untuk tes berikutnya.

Rancangan yang kedua adalah *random group design*. Paket-paket tes pada desain ini diadministrasikan secara random/acak kepada semua peserta tes. Misal ada dua paket yang diujikan yaitu paket X dan paket Y yang dibagikan secara bergantian. Peserta tes pertama akan mendapat paket X, peserta kedua mendapat paket Y, peserta ketiga mendapat paket X, dan seterusnya. Dengan

cara ini menjamin perangkat tes akan terdistribusi pada dua grup peserta tes dengan kemampuan yang relatif sama. Kelebihan dari desain ini adalah setiap peserta tes hanya menempuh satu paket tes sehingga faktor kelelahan (*fatigue effect*) peserta tes dapat dihilangkan. Kelebihan yang lain adalah memungkinkan beberapa paket diujikan dalam waktu yang bersamaan. Kekurangan *random group design* adalah terdapat kemungkinan bias yang disebabkan tidak mudah untuk membuat distribusi kemampuan dua kelompok peserta tes yang benar-benar ekuivalen.

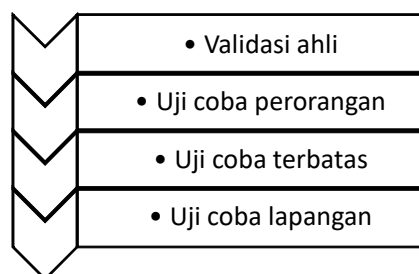
Rancangan penyetaraan *tes anchor test group design* dirancang untuk mengatasi kelemahan *random group design*. Pada rancangan ini masing-masing kelompok mengerjakan perangkat tes yang berbeda dan setiap perangkat tes memiliki beberapa butir soal yang sama (*common item/anchor item*/butir tes jangkar). Pada desain ini terdapat dua variasi, yaitu jika *common item/anchor item* diperhitungkan dalam pemberian skor disebut *internal common item* dan jika *common item/anchor item* tidak diperhitungkan dalam pemberian skor disebut *external common item* (Yu & Osborn, 2005). Butir tes jangkar (*common item*) terdiri dari sejumlah butir, yang merupakan miniatur dari kedua tes yang disetarakan dalam arti memiliki keserupaan sedekat mungkin, baik konten maupun kedalaman materi dengan kedua tes yang disetarakan (Petersen, 1989). Banyaknya butir soal jangkar disarankan sekitar antara 20 % sampai 25 % dari jumlah butir soal dalam satu perangkat tes (Hambleton dkk, 1991).

METODE

Secara garis besar, penelitian dan pengembangan ini mengadaptasi model 4D (Thiagarajan dkk, 1974). Tahapan pendefinisian dimulai dengan analisis ujung depan yang merupakan analisis kebutuhan pengembangan instrumen tes yang dilanjutkan dengan penyetaraan sebagaimana sudah dibahas pada bagian sebelumnya. Selanjutnya didefinisikan lingkup survey, yang mana dipilih materi kinematika gerak lurus yang dipelajari pada tahun pertama sekolah menengah atas. Lebih lanjut, ditentukan indikator kompetensi pada delapan sub-pokok bahasan kinematika gerak lurus antara lain: jarak dan perpindahan; laju dan kecepatan; percepatan; analisis diagram gerak; analisis grafik x-t; analisis grafik v-t; GLB; dan GLBB. Proses perancangan dimulai dengan membuat soal uraian (*open ended*) sejumlah 20 soal yang disebarkan ke siswa kelas XII SMAN. Jawaban siswa dianalisis dan

dijadikan salah satu acuan untuk menyusun soal pilihan ganda.

Validasi ahli dilakukan untuk menguji validitas isi dan kesetaraan tingkat kesulitan antara dua paket tes. Uji coba perorangan dilakukan oleh 6 siswa untuk menguji aspek pemahaman bahasa dan keterbacaan grafik. Pada proses validasi ahli dan uji coba perorangan terdapat beberapa revisi kecil yang dilakukan terkait tata bahasa dan tampilan gambar/grafik. Uji coba terbatas dilakukan oleh pada akhir Desember 2014 dengan responden siswa 101 kelas XI SMAN untuk menguji validitas empirik soal sebelum dimanfaatkan. Berdasarkan hasil uji coba lapangan terbatas dilakukan seleksi sehingga soal akhir terdiri dari 43 soal dimana setiap set terdiri 26 soal dengan 9 buah soal bersama (*anchor*). Alur penelitian dan pengembangan instrumen disajikan pada Gambar 1.



Gambar 1. Diagram Alir Pengembangan

Uji coba lapangan dilakukan di SMAN 4 Surabaya dengan melibatkan 209 siswa kelas X, XI dan XII. Analisis statistik baku dilakukan terhadap respon siswa dalam rangka mendapatkan nilai reliabilitas item tes. Proses statistik terkait penyetaraan juga dilakukan. Penyetaraan mengacu pada Crocker & Algina (1986). Penyetaraan dilakukan terhadap skor X dan skor Y yang melibatkan skor Z. Skor X adalah nilai siswa yang mengerjakan soal kode A, skor Y adalah nilai siswa yang mengerjakan soal kode B dan skor Z adalah nilai pada soal *anchor*. Skor X, Y dan Z memiliki rentang nilai seragam antara 0-4, disesuaikan dengan sistem penilaian pada kurikulum 2013.

Penyetaraan dari skor perangkat tes kode A (X) ke skor perangkat tes kode B (Y) pada model penyetaraan linear pada desain tes jangkar dilakukan dengan rumus:

$$Y^* = \frac{\sigma^2_Y + b^2_{YZ}(\sigma^2_Z - \sigma^2_{Z2})}{\sigma^2_X + b^2_{XZ1}(\sigma^2_Z - \sigma^2_{Z1})} (X - \mu_X + b_{XZ1}(\mu_Z - \mu_{Z1})) + \mu_Y + b_{YZ2}(\mu_Z - \mu_{Z2}) \dots \dots \dots (1)$$

Dengan cara yang sama, penyetaraan dari skor perangkat tes kode B (Y) ke skor perangkat tes kode A (X) dilakukan dengan rumus

$$X^* = \frac{\sigma_x^2 + b_{xz1}^2(\sigma_z^2 - \sigma_{z1}^2)}{\sigma_y^2 + b_{yz2}^2(\sigma_z^2 - \sigma_{z2}^2)} (Y - \mu_y + b_{yz2}(\mu_z - \mu_{z2})) + \mu_x + b_{xz1}(\mu_z - \mu_{z1}) \dots \dots \dots (2)$$

Proses pemilihan persamaan penyetaraan dilakukan dengan memperhatikan kesalahan baku penyetaraan. Kesalahan baku penyetaraan dari X ke Y dilakukan dengan rumus:

$$SEE[Y(X_0)] = \frac{\sigma_y^2}{2} \left(\frac{1}{N_x} + \frac{1}{N_y} \right) \left[2 + \frac{(X_0 - \mu_x)^2}{\sigma_x^2} \right] \dots \dots \dots (3)$$

Kesalahan baku penyetaraan dari Y ke X dilakukan dengan rumus:

$$SEE[X(Y_0)] = \frac{\sigma_x^2}{2} \left(\frac{1}{N_x} + \frac{1}{N_y} \right) \left[2 + \frac{(Y_0 - \mu_y)^2}{\sigma_y^2} \right] \dots \dots \dots (4)$$

HASIL

Proses penelitian dan pengembangan diawali dengan menyusun dan mengedarkan sejumlah soal uraian untuk menjarang konsep Fisika yang dimiliki siswa. Berikut contoh soal uraian yang diedarkan dengan indikator kompetensi “Membedakan jarak tempuh dan besar perpindahan”.

Eko berlari mengelilingi lapangan berbentuk lingkaran. Jika Eko berlari sebanyak 2,5 kali putaran, dan jari-jari lapangan 7 m. Tentukanlah jarak dan perpindahan yang ditempuh Eko?

Respon siswa dianalisis dengan memilah dan mengelompokkan ide yang diadaptasi dari *axial coding* (Strauss & Corbin, 1998) sebagaimana disajikan pada Tabel 1. Proses analisis ini meliputi mengidentifikasi ide pokoknya, mengode respon jawaban, dan mengelompokkan respon jawaban yang sejenis dalam kategori yang sama.

Tabel 1. Respon Jawaban Siswa

Kategori Respon Jawaban	Jumlah	Persentase Jawaban
Jarak sebagai panjang lintasan, dan perpindahan adalah jalur terpendek dari posisi awal dan akhir	8 ^a	24 ^a
Siswa berpikir bahwa jarak adalah luas dari lingkaran, dan tidak memiliki ide bagaimana menghitung perpindahan	1	3
Siswa memahami bahwa jarak adalah panjang lintasan, namun masih salah dalam memahami perpindahan. Mereka menganggap perpindahan sebagai keliling.	3	9
Siswa memahami bahwa jarak adalah panjang lintasan, namun masih salah dalam memahami perpindahan. Mereka menganggap perpindahan sebagai jarak dalam setengah putaran	11	32
Siswa salah dalam memilih rumus yang digunakan. Mereka berpikir bahwa keliling sebagai ?? = 2????	11	32

^aJawaban benar

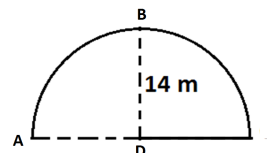
Langkah selanjutnya dibuat 4 buah soal pilihan ganda yang terdiri dari dua indikator soal. Indikator soal pertama adalah “disajikan data berupa gerak benda dalam lintasan dua dimensi, siswa dapat menentukan besar perpindahan” dengan soal pada Gambar 2 dan Gambar 3. Pengecoh pada soal pilihan ganda mengacu pada Tabel 1.

Perhatikan gambar di bawah ini!

Bila benda bergerak menyusuri lintasan A->B->C->D, maka besar perpindahan benda adalah ...

(catatan: bila dibutuhkan, gunakan nilai $\pi = \frac{22}{7}$)

- a. 58 m
b. 44 m
c. 28 m
d. 20 m
e. 14 m



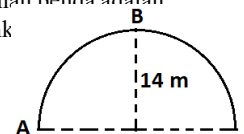
Gambar 2. Soal pilihan ganda kode A dengan indikator soal “disajikan data berupa gerak benda dalam lintasan dua dimensi, siswa dapat menentukan besar perpindahan”

Perhatikan gambar di bawah ini!

Bila benda bergerak menyusuri lintasan AàBàC, maka besar perpindahan benda adalah

(catatan: bila dibutuhkan

- a. 44 m
b. 28 m
c. 20 m
d. 18 m
e. 14 m

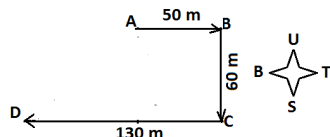


Gambar 3. Soal pilihan ganda kode B dengan indikator soal “disajikan data berupa gerak benda dalam lintasan dua dimensi, siswa dapat

Indikator soal kedua adalah “Disajikan data berupa gerak benda dalam lintasan dua dimensi, siswa dapat menentukan jarak tempuh” dengan konstruksi soal seperti Gambar 4 dan Gambar 5.

Sebagaimana ilustrasikan pada gambar di bawah, seorang anak dari titik A berjalan ke arah timur sejauh 50 m menuju titik B dan dilanjutkan ke titik C yang berada di selatan B sejauh 60 m. Jika anak tersebut kemudian berjalan ke arah barat sejauh 130 m menuju titik D, maka jarak tempuh perjalanan ini adalah

- 20 m
- 80 m
- 100 m
- 140 m
- 240 m

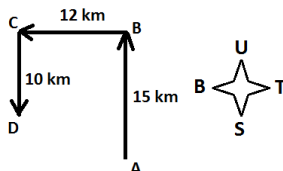


Gambar 4. Soal pilihan ganda kode A dengan indikator soal “disajikan data berupa gerak benda dalam lintasan dua dimensi, siswa dapat menentukan besar perpindahan”

Perhatikan gambar di bawah ini!

Rute perjalanan sebuah perlombaan sepeda mengikuti alur peta di bawah. Bermula dari titik start A ke arah utara 15 km menuju titik B, dilanjutkan 12 km ke arah barat menuju titik C, dan berakhir di titik D yang berada 10 km di selatan C. Jarak tempuh perlombaan sepeda tersebut adalah ...

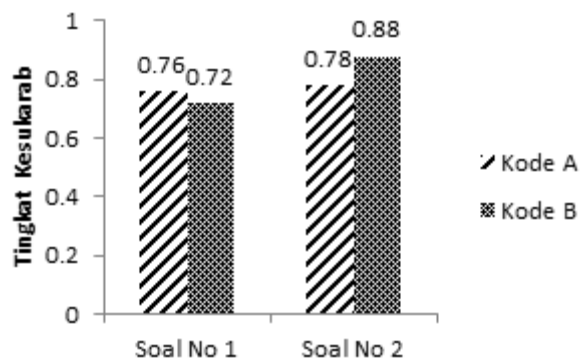
- 5 km
- 7 km
- 13 km
- 17 km
- 37 km



Gambar 5. Soal pilihan ganda kode B dengan indikator soal “disajikan data berupa gerak benda dalam lintasan dua dimensi, siswa dapat menentukan besar perpindahan”

Analisis statistik baku yang digunakan mengacu pada Ebel & Frisbie (1991) dan Kubiszyn & Borich (2003). Analisis statistik dilakukan pada data uji lapangan terbatas dengan responden sebanyak 101 siswa kelas XI SMAN 1 Blitar. Kriteria kelayakan adalah nilai tingkat kesukaran soal terletak antara 0,2 - 0,8. Nilai kurang dari 0,2 mengindikasikan bahwa item soal terlalu sulit, sedangkan nilai tingkat kesukaran lebih dari 0,8 menandakan bahwa item soal terlalu mudah. Daya beda bernilai lebih besar dari 0,2. Koefisien poin biserial bernilai lebih besar dari 0,2. Item tes yang nilainya kurang dari 0,2 dimaknai memiliki daya beda yang jelek.

Analisis tingkat kesukaran pada soal nomor 1 dan 2 soal kode A dan B sebagaimana Gambar 2, 3, 4 dan 5 ditunjukkan Gambar 6.



Gambar 6. Tingkat Kesukaran Item Tes No 1 dan 2 Kode A dan B

Soal nomor 2 kode B memiliki tingkat kesukaran 0,88 yang berarti bahwa soal terlalu mudah sehingga tidak digunakan dalam tahap selanjutnya. Sebagai gantinya soal nomor 2 kode A digunakan secara bersama (*sebagai soal anchor*). Hasil analisis secara keseluruhan menghasilkan 43 soal dari 60 soal yang diujikan dinilai layak. Soal-soal ini kemudian disusun ulang menjadi 2 set soal dengan jumlah soal 26 buah dengan 9 buah soal *anchor*.

PEMBAHASAN

Secara keseluruhan, analisis statistik terhadap hasil pengujian lapangan instrumen tes kinematika gerak lurus menunjukkan bahwa tes reliabel. Reliabilitas dihitung dengan rumus KR-20 menghasilkan nilai 0,767 untuk soal kode A dan 0,744 untuk soal kode B. Nilai rata-rata daya beda adalah 0,31 untuk soal kode A dan 0,29 untuk soal kode B yang menunjukkan bahwa soal memiliki daya beda yang baik.

Tingkat kesukaran soal berbeda bergantung pada jenjang kelas siswa. Tingkat kesukaran rata-rata soal untuk siswa kelas X 0,32, kelas XI 0,34, kelas XII adalah 0,36. Angka ini menunjukkan bahwa siswa kelas XII secara umum lebih mudah menyelesaikan soal dibandingkan dengan siswa kelas XI ataupun kelas X. Hal tersebut dapat dimaklumi karena siswa kelas XII lebih matang dalam konsep fisika secara keseluruhan. Siswa dengan jenjang lebih tinggi telah memahami “bahasa fisika” yang dibutuhkan untuk menjawab soal dengan benar.

(Tongchai, 2009). Sejalan dengan itu adanya hirarki respon siswa yang mana untuk memahami suatu konsep tertentu dibutuhkan suatu pemahaman konsep lain sebagai suatu prasyarat (Rosenblatt & Heckler, 2011).

Meskipun berasal dari indikator soal yang sama, dua buah soal memiliki tingkat kesukaran yang berbeda. Hal ini sejalan dengan pernyataan Hambleton, dkk. (1991) yang menyatakan bahwa dua buah perangkat pengukuran yang berbeda akan memiliki tingkat kesukaran yang berbeda walaupun kedua perangkat didasarkan pada kisi-kisi yang sama. Lebih lanjut, konsep soal yang sama namun disajikan dalam konteks yang berbeda juga akan mempengaruhi kesulitan butir soal (Planinik, dkk., 2013).

Penggunaan rumus yang lebih kompleks merupakan masalah siswa yang lain. Diperoleh nilai rata-rata tingkat kesulitan rata-rata untuk materi GLBB 0,22 yang lebih rendah dibandingkan dengan GLB 0,47. Permasalahan akan kompleksitas ini dialami oleh semua jenjang siswa, mulai siswa kelas X, XI dan XII. Nilai tingkat kesukaran untuk kelas X adalah 0,22, kelas XI adalah 0,22 dan kelas XII adalah 0,23.

Analisis grafik merupakan bagian lain dari kesulitan yang dialami siswa. Data menunjukkan bahwa rata rata tingkat kesukaran analisis grafik $x-t$ adalah 0,31, dan rata-rata tingkat kesukaran analisis grafik $v-t$ adalah 0,32. Nilai ini lebih kecil dari rata-rata tingkat kesukaran pada materi GLB 0,47 maupun analisis diagram gerak 0,39 apalagi jika dibandingkan dengan jarak dan perpindahan yang nilai rata-rata tingkat kesukarannya mencapai 0,57. Sejalan dengan temuan ini Hale (2000) menyatakan bahwa miskonsepsi grafik pada kinematika merupakan kejadian yang umum pada siswa. Kesulitan dalam menginterpretasikan grafik kinematika juga masih dialami mahasiswa yang telah mendapatkan mata kuliah dasar-dasar kalkulus (McDermott, dkk., 1987). Lebih lanjut Beichner (1994) menyatakan bahwa disamping menginterpretasikan grafik siswa juga mengalami kesulitan dalam menyajikan data dalam grafik.

Analisis distraktor digunakan untuk melihat respon siswa terhadap soal yang diberikan. Soal yang dibahas dengan rinci dipilih soal kode B nomor 1 dan 2 dengan indikator kompetensi membedakan jarak tempuh dan besar perpindahan yang konstruksi soalnya sesuai dengan Gambar 3. Sebanyak 39% siswa menjawab B yang merupakan kunci jawaban,

58% siswa yang lain menjawab A, sedangkan opsi yang lain tidak signifikan. Pilihan jawaban A adalah panjang lintasan A à B à C, sedangkan pilihan jawaban B adalah diameter lingkaran.

Indikator soal nomor 2 adalah “Disajikan data berupa gerak benda dalam lintasan dua dimensi, siswa dapat menentukan jarak tempuh” dengan konstruksi untuk soal sesuai Gambar 4. Sebanyak 71% siswa menjawab benar E, 16% menjawab C, sedangkan opsi lainnya tidak terlalu signifikan. Pilihan jawaban E adalah panjang lintasan A à B à C à D, sedangkan pilihan jawaban C adalah panjang vektor perpindahan AD.

Berdasarkan analisis distraktor soal nomor 1 dan 2 kode B dapat diketahui bahwa siswa mengalami kesulitan membedakan jarak tempuh dan perpindahan. Siswa menganggap perpindahan sama dengan jarak tempuh. Lebih lanjut, analisis terhadap soal 3 dan 4 menunjukkan bahwa karena siswa tidak dapat membedakan antara jarak tempuh dan perpindahan, maka siswa juga tidak mampu membedakan konsep kelajuan dan kecepatan. Analisis terhadap soal-soal yang lain akan dibahas pada artikel yang lain.

Perhitungan skor akhir didasarkan pada soal-soal yang memiliki tingkat kesulitan 0,2 – 0,8, daya beda ³ 0,2, dan koefisien point biserial ³ 0,2. Skor jawaban siswa pada butir-butir soal yang layak digunakan diperoleh persamaan konversi skor. Persamaan konversi dipilih dengan memperhatikan nilai kesalahan baku penyetaraan (SEE) terkecil. Dalam penyetaraan perangkat tes telah disarankan persamaan konversi skor untuk kelas X dari perangkat tes kode B (Y) ke perangkat tes kode A (X), yaitu $X^* = 1,064 Y + 0,015$. Untuk kelas XI disarankan persamaan konversi skor dari perangkat tes kode B (Y) ke perangkat tes kode A (X), yaitu $X^* = 0,834 Y + 0,455$. Untuk kelas XII disarankan persamaan konversi skor dari perangkat tes kode A (X) ke perangkat tes kode B (Y), yaitu $Y^* = 0,991 X - 0,147$. Dari persamaan tersebut dapat dilihat bahwa set soal kode A lebih mudah dan juga memberikan kesalahan baku penyetaraan yang lebih kecil dibandingkan set soal kode B, sehingga disarankan dilakukan konversi ke set soal kode A.

SIMPULAN

Penelitian ini mengembangkan dua buah instrumen tes kinematika gerak lurus dan penyetaraannya didasarkan pada kurikulum sekolah

menengah atas. Hasil evaluasi instrumen tes menunjukkan bahwa tes reliabel dan valid. Instrumen tes kinematika gerak lurus ini memiliki tujuh sub-pokok bahasan. Salah satu dari dua set instrumen tes kinematika gerak lurus dapat digunakan digunakan oleh guru sebelum instruksi atau ketika proses pembelajaran untuk mengidentifikasi konsepsi alternatif dan adaptasi proses pembelajaran pengajaran sedangkan set yang lain dapat digunakan di akhir proses pembelajaran untuk mengevaluasi peningkatan kemampuan yang dicapai oleh siswa.

DAFTAR RUJUKAN

- Beichner, R. J.. 1994. Testing Student Interpretation of Kinematics Graphs, *American Journal of Physics*. North Carolina.
- Crocker, L. and Algina, J.. 1986. *Introduction to classical and modern test theory*. New York: Holt, Rineart and Winston, Inc.
- Ebel, R. L., and Frisbie, D. A.. 1991. *Essentials of Educational Measurement Fifth Edition*. Yew Jersey: Prentice Hall Inc.
- Hale, P. 2000. *Kinematics and Graphs: Students' difficulties and CBLs. The Mathematics teacher*. Pomona: California.
- Hambleton, R.K., Swaminathan, H., Rogers, H.J..1991. *Fundamentals of Items Response Theory*. California: Sage Publication. Inc.
- Herukusumo, A. K.. 2011. Penyetaraan (Equating) Ujian Akhir Sekolah Berstandar Nasional (UASBN) dengan Teori Tes Klasik. *Jurnal Pendidikan dan Kebudayaan*, Vol. 17, Nomor 4, Juli 2011.
- Hestenes, D. & Wells, M.. 1992. A Mechanics Baseline Test. *The Physics Teacher*, Vol. 30, March 1992, p. 159-166.
- Hestenes, D., Wells, M. & Swackhamer, G.. 1992. Force Concept Inventory. *The Physics Teacher*, Vol. 30.
- Kartono. 2008. Penyetaraan tes Model campuran butir dikotomus dan politomus pada tes prestasi belajar. *Jurnal Penelitian dan Evaluasi Pendidikan*, No 2, Tahun XII.
- Kim, S. 2012. Investigating Repeater Effects on Chained Equipercentile Equating With Common Anchor Items. *Journal of Educational Measurement*. (Online), (<http://web.ebscohost.com/ehost>), diakses tanggal 6 Nopember 2013.
- Kolen, M. J., & Brennan, R. L.. 1995. *Test equating: Methods and practices*. New York: Springer-Verlag.
- Kubiszyn, T., & Borich, G. D. 2003. *Educational testing and measurement: Classroom application and practice* (7th ed.). John Wiley & Sons.
- Kusairi, S.. 2013. *Asesmen Pembelajaran Sains*. Malang: Modul tidak dipublikasikan dalam rangka Teachers Quality Imprevement Programm (TEQIP) kerjasama Universitas Negeri Malang dan Pertamina
- Linn, R L.. 1989. *Educational measurement*. New York: Macmillan Publishing Company
- Liu, J. 2008. A Comparison of the Kernel Equating Method with Traditional Equating Methods Using SAT Data., *Journal of Educational Measurement*. (Online) dalam Ebscohost, (<http://web.ebscohost.com/ehost>), diakses tanggal 6 Nopember 2013.
- Livingstone. L. 2004. *Equating Test Scores (Without IRT)*. Princeton: ETS
- McDermott, L.C., Rosenquist, M. L. and Van Zee, E .H. 1987. Student difficulties in connecting graphs and physics: Examples from kinematics. *American Journal of Physics* 55, 503-513.
- Moses, T. 2007. Using Kernel Equating to Assess Item Order Effectson Test Scores. *Journal of Educational Measurement*. (Online), (<http://web.ebscohost.com/ehost>), diakses tanggal 6 Nopember 2013.
- Osterlind, S.J.. 1998. *Constructing Test Items: Multiple-Choic, Constructed-Rensponse Performance, and other Formats*. New York: Kluwer Academic Publishers.
- Planinik, M., Ivanjek, L., & Susac, A.. 2013. Comparison of university students' understanding of graphs in different contexts. *Physical Review Special Topics –Physics Education Research*. (Online). (<http://prst-per.aps.org/pdf/PRSTPER/v9/i2/e020103>), diakses 25 september 2013.
- Puhan, G. 2008. Small-Sample Equating Using a Single-Group Nearly Equivalent Test (SiGNET) Design. *Journal of Educational Measurement*. (Online), dalam Ebscohost, (<http://web.ebscohost.com/ehost>), diakses tanggal 6 Nopember 2013.
- Rebello, C. M. & Barrow, L. H.. 2012. *Relationships Between Undergraduates Argumentation Skills, Conceptual Quality of Problem Solutions, and Problem Solving Strategies In Introductory Physics*. Disertasi. Columbia. Faculty of the Graduate School at the University of Missouri.
- Reynolds, C. R., Livingston, R. B., & Willson, V. 1999. *Measurement and assessment in education (2nd ed.)*. USA: Pearson.
- Rosenblatt, R. & Heckler, Andrew F.. 2011. Systematic study of student understanding of the relationships between the directions of force, velocity, and acceleration in one dimension.

- Physics Review Special Teacher*
- Rustam. 2000. *Penyetaraan perangkat tes matematika program D2 PGSD UT*. Lemlit Universitas Terbuka. Laporan Penelitian tidak dipublikasikan.
- Strauss, A., & Corbin, J.. 1998. *Basic of qualitative research: Techniques and procedures for developing grounded theory (2nd ed.)*. Thousand oaks, CA: Sage Publications.
- Sugeng. 2010. Penyetaraan vertikal model kredit parsial soal matematika smp. *Jurnal Penelitian dan Evaluasi Pendidikan* (Online), (<http://journal.uny.ac.id/index.php/jpep/article/view/1083/866>) diakses tanggal 24 Nopember 2013.
- Thiagarajan, S., Dorothy, S., Melvyn, I., & Semmel. 1974. *Instructional development for Traning Teachers of Exceptional Children*. Blomington Indiana: Indiana University.
- Tongchai, A., Sharma, M.D., Johnston, I. D., Arayathanitkul, K. & Soankwan, C.. 2009. Developing, Evaluating and Demonstrating the Use of a Conceptual Survey in Mechanical Waves. *International Journal of Science Education*, 31: 18, 2437-2457.
- Tsai, T. H. 2001. A Comparison of Bootstrap Standard Errors of IRT Equating Methods for the Common-Item Nonequivalent Groups Design. *Applied Measurement In Education*, California: Lawrence Erlbaum Associates, Inc. (Online), dalam Ebscohost, (<http://web.ebscohost.com/ehost>), diakses tanggal 6 Nopember 2013
- Von Davier, A., Holland, P.W., & Dorothy, T.. 2003. *The Kernel Method of test equating (Statistitics for Social Science and Behavioral Science)*. New York:Springer. (Online), (<http://en.bookfi.org>) diakses tanggal 25 Noember 2013.
- Yu, H. C., & Osborn, S. E.. 2005. Test equating by common items and common subjects: Concepts and applications. *Practical Assessment, Research & Evaluation*, 10, 4, 1-9.