

**PENYETARAAN VERTIKAL TES MATEMATIKA SMP
DENGAN TEORI RESPONS BUTIR MODEL RASCH**

Devi Dwi Kurniawan, Djemari Mardapi
Prodi PEP PPs UNY, Universitas Negeri Yogyakarta
devidwikurniawan@gmail.com, djemarimardapi@yahoo.co.id

Abstrak

Penelitian ini bertujuan untuk mengetahui: (1) karakteristik perangkat tes prestasi belajar matematika, 2) penyetaraan vertikal tes prestasi belajar matematika, 3) tingkat kemampuan matematika siswa SMP kelas VII, VIII dan IX di Kota Pekalongan. Penelitian ini menggunakan pendekatan kuantitatif. Populasi penelitian adalah respon siswa SMP di Kota Pekalongan. Data yang berupa skor diolah dan disetarakan menggunakan desain *nonequivalent anchor test* dengan metode penyetaraan *mean & sigma* berdasarkan teori respons butir model Rasch. Hasil penelitian ini adalah: (1) diperoleh tiga perangkat tes matematika memiliki tingkat kesukaran dan reliabilitas dengan kategori baik; (2) hasil penyetaraan vertikal menunjukkan persamaan konversi untuk kelas VII ke kelas VIII adalah $Y^* = bx + 0,369$ serta persamaan konversi untuk kelas IX ke kelas VIII adalah $Y^* = bz + 0,337$; (3) kemampuan siswa kelas VIII lebih baik dibandingkan kemampuan siswa kelas VII, dan kemampuan siswa kelas IX lebih baik dibandingkan kemampuan siswa kelas VIII.

Kata kunci: *penyetaraan vertikal, tes matematika, teori responsbutir model Rasch*

***VERTICAL EQUATING OF MATHEMATICS TEST
WITH RASCH MODEL ITEM RESPONSE THEORY***

Devi Dwi Kurniawan, Djemari Mardapi
Prodi PEP PPs UNY, Universitas Negeri Yogyakarta
devidwikurniawan@gmail.com, djemarimardapi@yahoo.co.id

Abstract

This research aims to reveal; (1) the characteristics of the mathematics learning achievement tests, (2) the vertical equating of mathematics achievement tests, and (3) the level of mathematical ability of junior high school students of classes VII, VIII and IX in Pekalongan. The research used the quantitative approach. The population of this research was the responses of the students of junior high schools in Pekalongan. The data in the form of scores were processed and equalised using the nonequivalent anchor test with the mean and sigma equating method based on Rasch model item response theory. The results of this research are: (1) three math tests that have difficulty level good and reliability, (2) the vertical equating showing the conversion equation of class VII to VIII class is $Y^ = bx + 0.369$, and the conversion equation of class IX to VIII class is $Y^* = bz + 0.337$, and (3) the ability of class VIII students better is than that of class VII, and class IX students' ability is better than that of class VIII students'.*

Keywords: *vertical equating, mathematics test, Rasch model item response theory*

Pendahuluan

Pada abad 21 pendidikan menjadi faktor penting untuk dapat bersaing secara global. Menurut Mardapi (2012, p.12) upaya meningkatkan kualitas pendidikan dapat ditempuh melalui peningkatan kualitas pembelajaran dan kualitas sistem penilaiannya. Dengan demikian, dalam proses pendidikan di sekolah misalnya dalam pembelajaran matematika harus mengupayakan pelaksanaan proses pembelajaran dan penilaian dengan sebaik-baiknya. Proses pembelajaran matematika yang baik tentunya dapat dilakukan dengan memberikan keleluasaan bagi siswa untuk mengembangkan dan mengeksplorasi kemampuannya.

Schwartz (2005, p.1) mengemukakan keberhasilan dasar dari pendidikan matematika adalah untuk mendukung pengembangan kecakapan di matematika dari berbagai macam kondisi kehidupan. Kecakapan matematika siswa dalam kondisi kehidupan di Sekolah dapat dilihat pada saat siswa mengikuti Tes. Pelaksanaan tes pada dasarnya untuk menilai keberhasilan siswa selama proses pembelajaran. Tes sangat perlu dilakukan agar pendidik dalam hal ini guru dapat mengetahui prestasi belajar siswanya setelah diberikan materi pelajaran dalam proses pembelajaran. Oleh sebab itu, pembuatan tes yang baik perlu diupayakan dengan mempertimbangkan kemampuan siswa, agar tes yang dilakukan sebagai alat ukur untuk menguji prestasi belajar siswa dapat mencerminkan/menggambarkan kemampuan siswa sesungguhnya.

Tes pada mata pelajaran matematika berkaitan dengan materi matematika yang diajarkan dalam jenjang tertentu misalnya di Sekolah Menengah Pertama (SMP). Materi matematika bersifat berkesinambungan dan saling terkait, dimana materi yang diajarkan pada kelas sebelumnya sebagai fundasi awal untuk memperdalam serta melanjutkan dengan keterkaitan materi yang diajarkan pada kelas selanjutnya, sehingga perlu penguasaan materi yang baik oleh siswa, agar siswa dalam kelas selanjutnya dapat memahami materi dengan baik. Penguasaan materi matematika siswa dapat dilihat dari tes yang akan diujikan guru, tetapi kenyataan yang ada masih dijumpai dalam pengembangan pembuatan tes oleh guru untuk menguji penguasaan siswa kurang memperhatikan tingkat kemampuan siswa misalnya soal-soal tes matematika yang disusun untuk tes tanpa dilakukannya pengujian terlebih

dahulu untuk melihat apakah soal tersebut layak digunakan untuk tes atau tidak, sehingga kualitas tes yang dihasilkan kurang baik.

Pembuatan soal yang digunakan untuk ulangan akhir semester di Sekolah Menengah Pertama di Kota Pekalongan guru matematika yang bergabung dalam Musyawarah Guru Mata Pelajaran Matematika Kota Pekalongan harus menyusun tes dengan baik dengan memilih model tes yang akurat yang dapat mengukur kemampuan siswa dengan tepat, sehingga prestasi matematika siswa dalam penguasaan materi dapat diketahui guru dengan melihat skor yang diperoleh dalam pelaksanaan tes.

Menurut Kolen & Brennan (2004, p.5) skor dua tes yang berbeda dari dua kelompok atau lebih dapat diperbandingkan jika butir-butir soalnya setara dan memiliki skala yang sama. Proses ini dapat digambarkan jika para guru dalam melaksanakan tes di sekolah hasil skor dari tes dapat dibandingkan antar kelas maupun antar sekolah apabila tes yang diberikan kepada siswa setara (sama). Kesetaraan antar skor dapat dilakukan secara statistik. Suatu proses statistik yang digunakan untuk menghasilkan skala tunggal dari skor dua tes berbeda tersebut dinamakan penyetaraan atau *equating* (Kolen & Brennan, 1995, p.2). Sedangkan menurut Peteson, Kolen, Hoover (dalam Linn, 1989, p.242) mengemukakan penyetaraan adalah suatu proses empiris yang diperlukan untuk mentransformasi skor dari tes yang satu ke skor tes yang lain. Selanjutnya Hambleton, Swaminathan, Rogers (1991, p.123) menyatakan bahwa penyetaraan merupakan proses menstransformasi skor X ke matriks skor tes Y atau sebaliknya, sehingga dari hasil penyetaraan ini dapat dibandingkan. Berdasarkan pendapat yang dikemukakan para ahli tersebut dapat diambil kesimpulan bahwa penyetaraan mempunyai manfaat untuk menyamakan hasil tes dari berbagai variansi tes berbeda yang diujikan di Sekolah.

Penyetaraan tes dapat dilakukan dengan dua cara yakni horisontal maupun vertikal (Crocker & Algina, 2008, p.456). Penyetaraan yang dilakukan terhadap tes yang memiliki tingkat kesulitan berbeda pada tingkat kelas yang berbeda tapi mengukur perlakuan yang sama disebut penyetaraan vertikal. Penyetaraan vertikal dapat digunakan guru untuk mengungkap perkembangan kemampuan siswa, walaupun siswa tersebut berada pada tingkat kelas yang berbeda dan memiliki tingkat kemampuan

yang berbeda, asalkan tes yang digunakan mengukur perlakuan yang sama.

Penyetaraan vertikal pada tingkat kelas yang berbeda pada pelajaran matematika akan membantu guru memperoleh informasi tentang kemampuan siswa pada mata pelajaran matematika, sehingga perkembangan kemampuan siswa dapat diamati. Hal tersebut mengisyaratkan betapa pentingnya melakukan penyetaraan vertikal bagi sekolah sebab melalui penyetaraan vertikal dapat diperoleh informasi tentang kemampuan siswa dan berbagai kecenderungannya. Informasi tersebut akan menjadi acuan dalam usaha perbaikan menuju peningkatan kualitas hasil belajar khususnya di Sekolah Menengah Pertama di Kota Pekalongan.

Penyetaraan vertikal tes matematika dengan teori respons butir model Rasch dilakukan di Kota Pekalongan untuk melihat seberapa besar kontribusi kebermanfaatan penyetaraan tes pada ulangan akhir semester matematika SMP di Kota Pekalongan. Rasch (1980, p.16) adalah orang pertama yang mengembangkan model logistik satu parameter. Pada model Rasch, orang diberi karakteristik tingkat kemampuan laten dan butir diberi karakteristik tingkat kesukaran. Probabilitas menjawab benar suatu butir adalah fungsi dari perbandingan antara tingkat kemampuan dan kesukaran butir.

Pada teori respons butir (IRT), kemampuan laten merupakan konstruk statistik yang diperoleh dari data empiris. Pendekatan IRT menggunakan tiga informasi butir tes: taraf kesukaran, nilai daya beda butir, dan faktor terkaan. Model Rasch hanya menggunakan taraf kesukaran butir. Pada model tersebut probabilitas jawaban benar naik ketika kemampuan siswa naik. Probabilitas jawaban salah naik ketika kemampuan siswa turun. Model Rasch mempunyai model ciri khusus yang tidak dimiliki model lain. Pertama, probabilitas jawaban benar tergantung pada perbedaan antara parameter kemampuan siswa dan parameter kesukaran butir. Kedua, parameter kemampuan siswa dan parameter kesukaran butir dapat diestimasi secara terpisah. Parameter kemampuan siswa yang diestimasi tidak bergantung pada distribusi butir tes. Parameter kesukaran butir yang diestimasi tidak bergantung pada distribusi kemampuan siswa. Ketiga, skor mentah cukup untuk mengestimasi parameter kemampuan. Keempat, luas di bawah fungsi informasi selalu konstan. Untuk jumlah kategori respons tertentu, semua butir tes memberikan sum-

bangun sama terhadap tinggi fungsi informasi sepanjang kontinum kemampuan.

Berbagai penelitian yang telah dilakukan tentang penyetaraan diantaranya hasil penelitian Antara (2014, p.2) yang berjudul penyetaraan vertikal dengan pendekatan klasik dan IRT pada siswa sekolah dasar, menunjukkan bahwa penggunaan beberapa metode penyetaraan memberikan informasi bahwa nilai kemampuan meningkat seiring meningkatnya peringkat kelas. Sedangkan hasil penelitian Sugeng (2010, p.289) yang berjudul penyetaraan vertikal model kredit parsial soal matematika SMP, menunjukkan keakuratan penyetaraan semakin meningkat seiring meningkatnya panjang tes. Dari hasil penelitian-penelitian tersebut menunjukkan bahwa penyetaraan tes baik digunakan dalam mengukur prestasi belajar siswa. Berdasarkan uraian-uraian teori dan permasalahan-permasalahan serta diperkuat hasil penelitian yang telah ada dapat disimpulkan bahwa dengan penyetaraan tes matematika di SMP akan dapat mengukur kemampuan siswa dengan lebih baik sehingga diharapkan dapat meningkatkan prestasi belajar matematika selama proses pembelajaran di Sekolah.

Penelitian ini bertujuan untuk mengetahui; (1) karakteristik perangkat tes prestasi belajar matematika siswa SMP kelas VII, VIII dan IX di Kota Pekalongan yang telah disusun berdasarkan teori respons butir model Rasch, (2) penyetaraan vertikal tes prestasi belajar matematika siswa SMP kelas VII ke kelas VIII dan kelas VIII ke kelas IX di Kota Pekalongan yang telah disusun berdasarkan teori respons butir model Rasch, (3) tingkat kemampuan matematika siswa SMP kelas VII, VIII dan IX di Kota Pekalongan.

Metode Penelitian

Penelitian ini merupakan jenis penelitian eksploratif dengan pendekatan kuantitatif. Mengeksplorasi kemampuan siswa dalam tes prestasi belajar matematika tingkat Sekolah Menengah Pertama (SMP).

Penelitian ini dilakukan di Kota Pekalongan, khususnya di Sekolah Menengah Pertama (SMP) di bawah naungan Dinas Pendidikan Pemuda dan Olahraga Kota Pekalongan Tahun Pelajaran 2014/2015. Penelitian ini dilakukan pada bulan Februari 2015 – Maret 2015.

Populasi dalam penelitian ini adalah respons siswa SMP di Kota Pekalongan terha-

dap tes prestasi belajar matematika. Pengambilan sampel menggunakan teknik *stratified proporsional random sampling*, untuk menentukan sekolah dan kelas yang dipilih sebagai sampel penelitian. Alasan peneliti menggunakan teknik ini dikarenakan populasi dalam penelitian ini berstrata, yaitu kelas VII, VIII, dan IX, kemudian secara proporsional diambil siswa pada setiap strata. Selanjutnya adalah penentuan sekolah secara random atau acak. Langkahnya adalah mengelompokkan sekolah kedalam tiga kategori, yaitu sekolah dengan kategori atas; Kategori tengah; kategori bawah. Hal ini diperoleh berdasarkan data dari Dinas Pendidikan Pemuda dan Olahraga Kota Pekalongan serta hasil diskusi dengan beberapa guru matematika yang bernaung dalam Asosiasi Musyawarah Guru Mata Pelajaran (MGMP) matematika di Kota Pekalongan. Setelah strata sekolah teridentifikasi dilakukan pemilihan sekolah secara acak pada setiap kategori sekolah. Dalam penelitian ini siswa kelas VII, VIII, dan IX pada 5 sekolah terpilih akan dijadikan sampel penelitian yakni SMP Negeri 2 Pekalongan mewakili strata atas, SMP Negeri 8 Pekalongan dan SMP Negeri 15 Pekalongan mewakili strata tengah serta SMP Negeri 5 Pekalongan dan SMP Negeri 11 Pekalongan mewakili strata bawah. Besar ukuran sampel penelitian disajikan pada tabel 2 di bawah ini.

Tabel 1. Sampel Penelitian

No	Sekolah	Kelas		
		VII	VIII	IX
1	SMP N 2 Pekalongan	108	108	95
2	SMP N 5 Pekalongan	99	80	79
3	SMP N 8 Pekalongan	109	109	100
4	SMP N 11 Pekalongan	126	121	105
5	SMP N 15 Pekalongan	72	85	62
Jumlah Siswa		514	503	441
Jumlah Sampel		1458		

Penelitian ini menggunakan data empiris berupa pola respon siswa yang bersifat dikotomus, sebab data dikotomus bentuk pilihan ganda lebih objektif dan handal untuk melihat respon siswa, tanpa dipengaruhi subjektivitas penilai. Data empiris yang digunakan yaitu, hasil tes mata pelajaran matematika SMP di Kota Pekalongan Tahun Ajaran 2014/2015. Selain itu, pemilihan tes bentuk pilihan ganda didasarkan pada kemampuan bentuk tes pilihan

ganda mampu mengungkap materi yang luas dan digunakan pada tes berskala besar.

Data diperoleh dari respons jawaban tes matematika siswa kelas VII, VIII, dan IX di SMP Kota Pekalongan Tahun Pelajaran 2014/2015. Data respon jawaban siswa diperoleh dari masing-masing sekolah yang digunakan sebagai sampel penelitian, yakni SMP N 2 Pekalongan, SMP N 5 Pekalongan, SMP N 8 Pekalongan, SMP N 11 Pekalongan, dan SMP N 15 Pekalongan.

Instrumen tes matematika model dikotomus untuk kelas VII, VIII, dan IX SMP di Kota Pekalongan disusun berdasarkan pokok bahasan sesuai dengan kurikulum yang digunakan. Ketiga perangkat tes memuat butir-butir soal yang sama sebagai butir anchor. Banyaknya butir anchor ditentukan minimal 20% untuk pilihan ganda. Dasar yang digunakan penetapan banyaknya butir anchor mengacu pada teori bahwa banyak butir anchor minimal 20% (Hambleton, Swaminathan, & Rogers, 1991, p.135).

Penyusunan kisi-kisi tes bertujuan untuk mendiskripsikan ruang lingkup dan tekanan tes beserta bagian-bagiannya, sehingga menjadi petunjuk efektif dalam penulisan soal. Penulisan soal merupakan realisasi dari penjabaran indikator dan jenis tingkah laku yang diukur menjadi pertanyaan-pertanyaan yang karakteristiknya sesuai dengan perincian dalam kisi-kisi. Soal dibuat sesuai kisi-kisi tes matematika SMP yang telah disusun untuk masing-masing kelas yakni kelas VII, kelas VIII, dan kelas IX. Dalam setiap perangkat tes matematika terdapat 10 butir bersama (*anchor*).

Analisis kualitatif tes yang meliputi validitas isi dan keterbacaan soal melibatkan ahli dalam bidang pendidikan matematika, ahli dalam bidang pengukuran pendidikan dan ahli bidang evaluasi pendidikan. Selain itu analisis soal juga melibatkan guru dan beberapa siswa SMP untuk mendapatkan informasi keterbacaan soal. Perangkat tes yang telah dikonsultasikan kepada para ahli direvisi dengan mempertimbangkan penilaian dan masukan-masukan dari para ahli yang telah menelaah perangkat tes. Setelah melakukan revisi dihasilkan perangkat tes matematika yang siap diuji cobakan.

Ujicoba soal dilakukan untuk mendapatkan informasi empirik tentang karakteristik butir dan kemampuan butir. Dalam ujicoba soal dilaksanakan disalah satu SMP di Kota Pekalongan yang bukan menjadi sampel penelitian yakni SMP IT Assalam Boarding School Peka-

longan. Pelaksanaan uji coba melibatkan siswa kelas VII, kelas VIII, dan kelas IX. Hasil analisis uji coba dianalisis menggunakan program Winsteps untuk melihat tingkat kesukaran dan reliabilitas tes yang telah disusun. Butir-butir soal yang tidak baik digugurkan atau direvisi. Hasil uji coba instrumen disajikan pada Tabel 2 berikut ini.

Tabel 2. Hasil Uji Coba Instrumen

Karakteristik	Kelas		
	VII	VIII	IX
Jumlah Item	30	30	30
Jumlah Siswa	104	102	95
Soal Baik	23	24	23
Soal di Revisi	7	6	7
Soal Tidak Baik	-	-	-
Reliabilitas	0,87	0,80	0,89

Berdasarkan hasil analisis terhadap uji coba instrumen di atas, tidak terdapat soal yang tidak layak digunakan, hanya terdapat beberapa soal yang perlu direvisi, hal ini dikarenakan hasil output analisis menunjukkan bahwa semua item berada pada interval $[-2,+2]$. Hal ini sesuai dengan yang dikemukakan Hambleton & Swaminathan (1985, p.36) item yang baik jika nilai tingkat kesukaran dalam interval $[-2,+2]$ yang terbagi dalam tiga kategori yakni Soal dikategorikan sukar apabila b mendekati $+2$, kategori mudah apabila b sekitar -2 dan soal dikategorikan sedang apabila $-1,00 < b < +1,00$. Soal-soal yang perlu dilakukan revisi yakni soal-soal yang berada dalam kategori sukar dan mudah. Untuk soal kelas VII yang perlu direvisi yakni soal nomor 3, 4, 12, 16, 26, 27 dan 28. Untuk soal kelas VIII yang perlu direvisi yakni soal nomor 1, 7, 8, 13, 14 dan 26. Untuk soal kelas IX yang perlu direvisi yakni soal nomor 7, 9, 14, 18, 20, 21 dan 25. Selanjutnya soal-soal yang direvisi pada masing-masing kelas, disusun kembali dan digunakan untuk pengambilan data penelitian. Dari hasil uji coba juga diketahui keefisienan waktu yang sesuai digunakan untuk menyelesaikan seperangkat tes.

Desain penyetaraan vertikal tes matematika SMP menggunakan *common-item non-equivalent groups design* atau desain tes anchor. Penentuan butir anchor dilakukan dengan mempertimbangkan butir yang memenuhi syarat dan distribusi materi soal menurut peringkat kelas. Dalam penelitian ini, peneliti

membuat seperangkat tes, kemudian diterapkan pengukuran kepada peserta. Pengukuran ini akan menampilkan hasil dalam bentuk skor. Selanjutnya, bagaimana membandingkan dua skor dari dua tes yang berbeda. Jadi jika dua peserta masing-masing mengikuti tes berbeda, maka bagaimana cara membandingkan skor siswa. Hal ini dapat dilakukan dengan cara penyetaraan.

Penelitian ini akan menyetarakan tiga perangkat tes matematika yang berbeda tingkatannya. Perangkat tes meliputi materi kelas VII, VIII, dan IX. Siswa kelas VII mengerjakan tes X dengan butir anchor V1, kelas VIII mengerjakan tes Y dengan anchor V1 dan V2, sedangkan kelas IX mengerjakan tes Z dengan butir anchor V2. Masing-masing paket tes terdiri dari 30 butir dikotomus, sedangkan butir anchor V1 dan V2 masing-masing berjumlah 10 item.

Analisis data butir tes dilakukan dengan pendekatan teori respons butir dengan menggunakan bantuan program Winsteps. Penyetaraan tes juga dilakukan dengan pendekatan teori respons butir dengan bantuan program Winsteps. Winsteps merupakan perangkat lunak (*software*) yang diciptakan khusus untuk analisis butir soal. Program ini disusun berdasarkan pendekatan model Rasch. Hasil analisis butir dengan menggunakan program Winsteps memberikan informasi tentang skor, kemampuan, indeks kesalahan pengukuran, infit, dan outfit, serta tingkat kesukaran baik untuk butir (*item*) maupun peserta tes (*person*).

Hasil Penelitian dan Pembahasan

Hasil analisis penelitian meliputi: (1) Uji asumsi teori respons butir (2) Karakteristik perangkat tes matematika. (3) penyetaraan skor dan skala; (4) penyetaraan kemampuan siswa. Analisis yang digunakan dalam penelitian ini menggunakan pendekatan teori respons butir model Rasch menggunakan bantuan program Winsteps.

Uji Asumsi Teori Respons Butir

Uji asumsi unidimensi dilakukan untuk mengetahui apakah tes yang digunakan mengukur satu macam kemampuan, dalam penelitian ini berarti mengukur kemampuan siswa SMP pada mata pelajaran matematika. Uji asumsi unidimensi dilakukan melalui analisis faktor menggunakan program SPSS statistics 21. Sebelum melakukan analisis faktor dilakukan pengujian kelayakan analisis menggunakan uji

KMO-MSA dan uji Bartlet's pada tiap tes. Uji KMO-MSA digunakan untuk melihat kecukupan sampel, sedangkan uji Bartlet's untuk normalitas data yang digunakan. Hasil analisis disajikan pada Tabel 3.

Tabel 3. Uji Analisis Faktor Tes Kelas VII, VIII, dan IX

Tes	KMO	Bartlett's Test
Kelas VII	0,917	0,000
Kelas VIII	0,890	0,000
Kelas IX	0,917	0,000

Tabel 4 menunjukkan hasil analisis empiris pada tes kelas VII dengan nilai KMO-MSA sebesar 0,917 dan nilai signifikansi uji Bartlet sebesar 0,000. Nilai KMO-MSA yang dihasilkan dari ketiga perangkat tes kelas VII, kelas VIII dan kelas IX sudah memenuhi syarat untuk melakukan analisis faktor. Hal ini telah sesuai dengan syarat untuk analisis faktor adalah KMO-MSA > 0,5 dan uji Bartlet < 0,5.

Menurut Hambleton & Swaminathan (1985, p.16) uji unidimensi terpenuhi jika tes mengukur satu dimensi yang dominan yang mengukur kemampuan yang sama. Untuk mendapatkan butir yang mengukur dimensi yang sama, dilakukan proses ekstraksi sehingga dihasilkan beberapa faktor. Setiap faktor yang terbentuk memiliki nilai eigen dan faktor dengan nilai eigen di atas 1,00 dipertahankan. Nilai eigen tes kelas VII, VIII dan IX dapat dilihat pada Tabel 4.

Tabel 4. Nilai Eigen Tes Matematika Kelas VII, VIII, dan IX

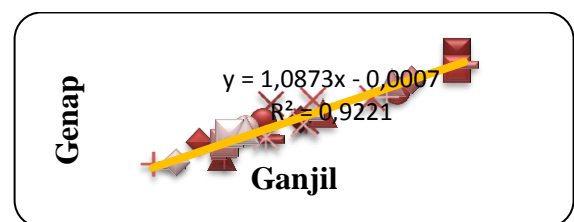
Jumlah Faktor	Nilai Eigen		
	Kelas VII	Kelas VIII	Kelas IX
1	6,460	5,599	6,836
2	1,388	1,643	1,657
3	1,296	1,314	1,271
4	1,205	1,215	1,153
5	1,107	1,193	1,142
6	1,077	1,167	1,106
7	1,052	1,109	1,052
8	1,023	1,041	1,026
9	0,982	1,004	0,963

Hasil analisis faktor pada Tabel 5 di atas menunjukkan pada kelas VII terdapat 8 nilai eigen yang lebih dari 1, pada kelas VIII terdapat

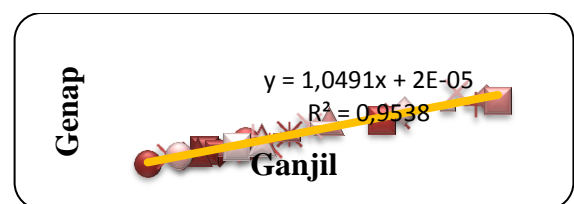
9 nilai eigen yang lebih dari 1, dan pada kelas IX terdapat 8 nilai eigen yang lebih dari 1, sehingga dikatakan bahwa dari ke 30 butir pada setiap kelas tersebut dapat membentuk 8 faktor pada kelas VII, 9 faktor pada kelas VIII dan 8 faktor pada kelas IX. Hasil analisis juga menunjukkan bahwa faktor 1 merupakan faktor yang dominan karena memiliki nilai eigen sebesar 6,460 pada kelas VII, nilai eigen sebesar 5,599 dan nilai eigen sebesar 6,836 pada kelas IX, nilai eigen faktor 1 pada setiap kelas hampir 3 kali dari nilai eigen faktor yang kedua, sehingga dapat dikatakan tes matematika kelas VII, VIII, dan IX bersifat unidimensi.

Uji asumsi independensi lokal otomatis terbukti, setelah dibuktikan dengan unidimensionalitas data respons peserta terhadap suatu tes (Heri Retnowati, 2014, p.7). Penjelasan tersebut dapat dimaknai bahwa dengan telah terbukti asumsi unidimensi sehingga asumsi independensi lokal juga telah terpenuhi.

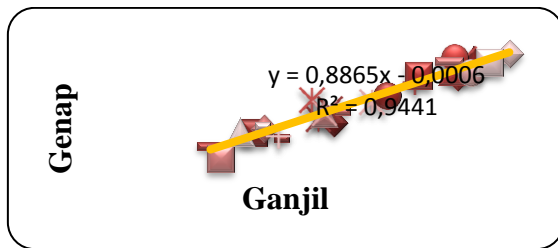
Uji invariansi parameter bertujuan untuk mengetahui apakah karakteristik item tidak berubah meskipun dijawab oleh kelompok siswa yang berbeda. Begitu pula untuk kelompok siswa yang sama, estimasi kemampuan tidak akan berubah meskipun item soal berubah-ubah. Jadi, ada dua uji invariansi parameter yakni invariansi parameter butir dan invariansi parameter kemampuan siswa. Pengujian dilakukan menggunakan diagram pencar scree plot untuk tiap parameter butir. Hasil invariansi parameter butir dan parameter kemampuan kelas VII, kelas VIII dan kelas IX dapat dilihat pada gambar 1, 2, 3, 4, 5 dan 6.



Gambar 1. Scree Plot Invariansi Parameter Tingkat Kesukaran Kelas VII

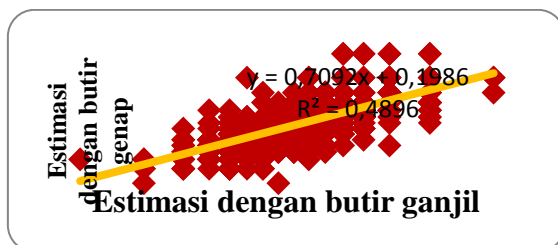


Gambar 2. Scree Plot Invariansi Parameter Tingkat Kesukaran Kelas VIII

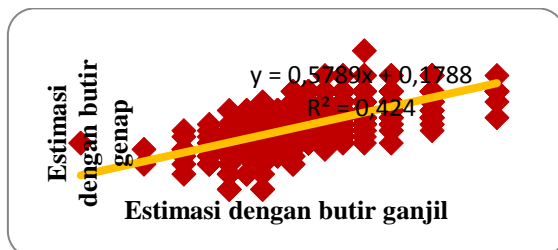


Gambar 3. Scree Plot Invariansi Parameter Tingkat Kesukaran Kelas IX

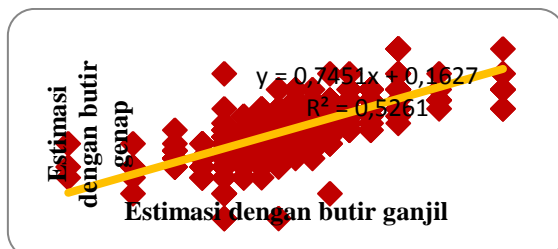
Pada gambar 1, 2, dan 3 di atas menunjukkan secara keseluruhan plot menyebar pada garis diagonal. Plot yang menyebar pada garis diagonal berarti parameter butir tingkat kesukaran bersifat invariansi.



Gambar 4. Scree Plot Invariansi Parameter Kemampuan Kelas VII



Gambar 5. Scree Plot Invariansi Parameter Kemampuan Kelas VIII



Gambar 6. Scree Plot Invariansi Parameter Kemampuan Kelas I

Pada gambar 4, 5, dan 6 di atas menunjukkan secara keseluruhan plot mendekati garis diagonal. Plot yang menyebar pada garis diagonal berarti parameter kemampuan bersifat invariansi.

Karakteristik Perangkat Tes Matematika

Penelitian ini menghasilkan tiga perangkat tes matematika yang telah diuji empiris secara kuantitatif dan kualitatif. Ketiga perangkat adalah tes matematika kelas VII, kelas VIII dan kelas IX yang masing-masing berjumlah 30 butir soal.

Karakteristik perangkat tes matematika dianalisis dengan pendekatan teori respons butir model Rasch menggunakan bantuan program Winsteps. Hasil keluaran (*output*) analisis butir soal dengan program Winsteps mencakup informasi tentang skor (*score*), kemampuan, dan indeks kesalahan pengukuran (*standard error*). Tampilan lain yang disajikan oleh Winsteps adalah beberapa grafik dan tabel, termasuk tabel konversi skor ke kemampuan. Hal ini dipilih sesuai dengan kebutuhan dalam mencapai tujuan penelitian.

Tes Matematika Kelas VII

Analisis tes matematika kelas VII meliputi kecocokan model, parameter butir dan ciri peserta, fungsi informasi butir dan SEM. Hasil analisis menggunakan program Winsteps diperoleh beberapa informasi. Butir yang dianalisis sebanyak 30 dan peserta 514 siswa. Butir dengan tingkat kesukaran paling tinggi adalah butir 2 dengan tingkat kesukaran 1,51 dan butir paling mudah adalah butir nomor 16 dengan tingkat kesukaran -1,13.

Responden yang memiliki kemampuan tertinggi pada siswa adalah peserta nomor 78 dengan kemampuan 3,63, dan peserta dengan kemampuan terendah peserta nomor 406 dengan kemampuan -2,84. Pada output hasil analisis pemilihan peserta yang diterima sesuai dengan data adalah peserta dengan *outfit mean square* ($0,5 < \text{MNSQ} < 1,5$) dan nilai *point measure correlation* (*Pt Mean Corr*) tidak negatif. Hasil analisis menunjukkan bahwa terdapat 29 siswa yang tidak sesuai dengan kedua kategori.

Hasil analisis menunjukkan informasi yang menunjukkan tingkat kesukaran, kesalahan pengukuran, *infit*, *outfit* dan *Pt measure*. Pada output hasil analisis perangkat tes kelas VII pemilihan item yang diterima sesuai dengan data adalah item dengan *outfit mean square* ($0,5 < \text{MNSQ} < 1,5$) dan nilai *point measure correlation* (*Pt Mean Corr*) tidak negatif. Hasil analisis menunjukkan bahwa terdapat 1 butir yang tidak sesuai dengan kategori. Butir tersebut adalah Butir nomor 9 yang memiliki

tingkat kesukaran -0,34, kesalahan pengukuran 0,10, outfit MNSQ 1,60, outfit ZSTD 7,2 dan Pt Measure 0,12.

Hasil analisis menunjukkan kriteria tingkat kesukaran butir yang terbagi dalam tiga kategori. Soal dikategorikan sukar apabila b mendekati +2, dan kategori mudah apabila b sekitar -2. Soal dikategorikan sedang apabila $-1,00 < b < +1,00$. Hasil karakteristik butir soal perangkat tes matematika kelas VII disajikan pada Tabel 5.

Tabel 5. Karakteristik Butir Tes Kelas VII

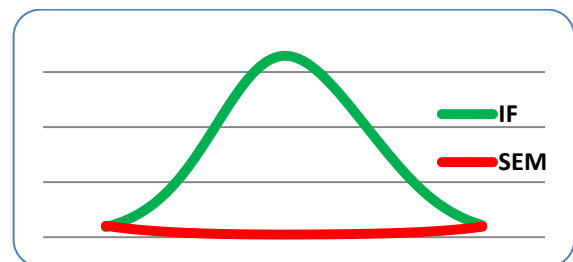
Butir	Tingkat Kesukaran	Kriteria
1	0,03	Sedang
2	1,51	Sukar
3	-0,75	Sedang
4	0,12	Sedang
5	0,03	Sedang
6	-0,14	Sedang
7	-0,74	Sedang
8	-0,35	Sedang
9	-0,34	Sedang
10	-0,66	Sedang
11	1,30	Sukar
12	0,20	Sedang
13	0,34	Sedang
14	0,60	Sedang
15	0,79	Sedang
16	-1,13	Mudah
17	-0,77	Sedang
18	1,39	Sukar
19	0,98	Sedang
20	-0,57	Sedang
21	0,86	Sedang
22	-0,02	Sedang
23	-0,33	Sedang
24	-0,31	Sedang
25	0,67	Sedang
26	0,11	Sedang
27	-1,07	Mudah
28	-1,01	Mudah
29	-0,43	Sedang
30	-0,33	Sedang

Hasil analisis pada Tabel 5 di atas menunjukan butir soal yang dikategorikan mudah sebanyak 3 butir, kategori butir sedang sebanyak 24 butir dan kategori butir sukar sebanyak 3 butir.

Fungsi Informasi Butir Kelas VII

Fungsi informasi butir berguna untuk menentukan mutu perangkat tes yang terdiri atas sejumlah butir. Pada model Rasch, fungsi informasi butir tertinggi akan dicapai apabila

yang merespon butir memiliki kemampuan setara dengan tingkat kesukarannya. Dengan kata lain fungsi informasi butir maksimum ketika parameter ciri peserta $\theta = b_i$. Hasil perhitungan menunjukkan nilai fungsi informasi maksimum sebesar 16,4487 logit pada sekitar -0,2. SEM dalam tes ini sebesar 0,2466 SEM berbanding terbalik dengan fungsi informasi tes. Hal ini berarti siswa kelas VII akan memberikan informasi yang baik, dengan kesalahan pengukuran terkecil, apabila dikerjakan oleh siswa yang memiliki $\theta = -0,2$. Grafik hubungan fungsi informasi butir dan SEM disajikan pada Gambar 7.



Gambar 7. Hubungan Fungsi Informasi Butir dan SEM Tes Kelas VII

Pada Gambar 7 di atas menyajikan nilai fungsi informasi dari suatu tes matematika kelas VII dengan 30 butir. Perangkat tes ini memiliki nilai fungsi informasi yang lebih tinggi dibandingkan kesalahan pengukuran pada rentang -3 sampai +3. Dengan demikian perangkat tes ini sesuai untuk peserta tes pada rentang kemampuan tersebut.

Tes Matematika Kelas VIII

Hasil analisis perangkat tes matematika kelas VIII menggunakan program Winsteps diperoleh beberapa informasi. Butir yang dianalisis sebanyak 30 dan peserta 503 siswa. Butir dengan tingkat kesukaran paling tinggi adalah butir 19 dengan tingkat kesukaran 1,56 dan butir paling mudah adalah butir nomor 6 dengan tingkat kesukaran -1,21.

Responden yang memiliki kemampuan tertinggi pada siswa adalah peserta nomor 476 dengan kemampuan 3,71, dan peserta dengan kemampuan terendah peserta nomor 69 dengan kemampuan -2,09. Pada output hasil analisis pemilihan peserta yang diterima sesuai dengan data adalah peserta dengan *outfit mean square* (MNSQ) $0,5 < MNSQ < 1,5$ dan nilai *point measure correlation* (Pt Mean Corr) tidak negatif. Hasil analisis menunjukkan bahwa terdapat 28 siswa yang tidak sesuai dengan kedua kategori.

Hasil analisis menunjukkan informasi yang menunjukkan tingkat kesukaran, kesalahan pengukuran, *infit*, *outfit* dan *Pt measure*. Pada output hasil analisis pemilihan item yang diterima sesuai dengan data adalah item dengan *outfit mean square* (MNSQ) $0,5 < MNSQ < 1,5$ dan nilai *point measure correlation* (Pt Mean Corr) tidak negatif. Hasil analisis menunjukkan tidak terdapat item yang tidak sesuai dengan kedua kategori.

Hasil analisis menunjukkan kriteria tingkat kesukaran butir yang terbagi dalam tiga kategori. Soal dikategorikan sukar apabila b mendekati $+2$, dan kategori mudah apabila b sekitar -2 . Soal dikategorikan sedang apabila $-1,00 < b < +1,00$. Hasil karakteristik butir soal kelas VIII disajikan pada Tabel 6.

Tabel 6. Karakteristik Butir Tes Kelas VIII

Butir	Tingkat Kesukaran	Kriteria
1	-0,69	Sedang
2	-0,71	Sedang
3	-0,78	Sedang
4	0,81	Sedang
5	-0,01	Sedang
6	-1,21	Mudah
7	1,38	Sukar
8	-0,67	Sedang
9	-0,51	Sedang
10	-0,91	Sedang
11	0,66	Sedang
12	-0,23	Sedang
13	-0,98	Sedang
14	-0,33	Sedang
15	-0,34	Sedang
16	-0,56	Sedang
17	-0,68	Sedang
18	-0,09	Sedang
19	1,56	Sukar
20	1,55	Sukar
21	0,41	Sedang
22	1,56	Sukar
23	0,26	Sedang
24	-0,91	Sedang
25	-0,20	Sedang
26	1,17	Sukar
27	0,31	Sedang
28	0,95	Sedang
29	-0,50	Sedang
30	-0,30	Sedang

Hasil analisis pada Tabel 6 di atas menunjukkan butir soal yang dikategorikan mudah sebanyak 1 butir yakni butir 6, kategori butir sedang sebanyak 24 butir.

Hasil tes dapat digunakan untuk menentukan penyetaraan tes anchor. Pada tes ini yang merupakan butir anchor adalah butir nomor 11 sampai 20. Oleh sebab itu yang perlu diperhatikan untuk analisis lebih lanjut adalah butir 11 sampai dengan 20.

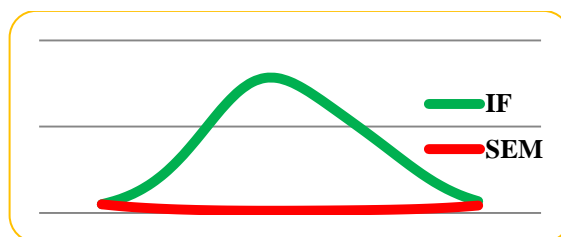
Hasil analisis diperoleh beberapa informasi untuk penyetaraan tes. Rangkuman penyetaraan soal kelas VII dengan soal kelas VIII yang merupakan butir anchor dapat dilihat pada Tabel 7. Selanjutnya dapat ditentukan persamaan kesetaraan antara tes anchor kelas VII ke tes anchor kelas VIII.

Tabel 7. Penyetaraan Tes Anchor Kelas VII ke Kelas VIII

Nomor Butir	Tes Anchor-1 Kelas VII	Tes Anchor-1 Kelas VIII
1	1,30	0,66
2	0,20	-0,23
3	0,34	-0,98
4	0,60	-0,33
5	0,79	-0,34
6	-1,13	-0,56
7	-0,77	-0,68
8	1,39	-0,09
9	0,98	1,56
10	-0,57	1,55

Fungsi Informasi Butir Kelas VIII

Hasil analisis dengan model Rasch perangkat tes matematika kelas VIII memiliki nilai fungsi informasi maksimum 15,6386 logit pada sekitar $-0,4$ dan SEM sebesar 0,2529. Hal ini berarti siswa kelas VIII akan memberikan informasi yang baik, dengan kesalahan pengukuran terkecil, apabila dikerjakan oleh siswa yang memiliki kemampuan sekitar $-0,4$. Grafik hubungan fungsi informasi butir dan SEM disajikan pada Gambar 8.



Gambar 8. Hubungan Fungsi Informasi Butir dan SEM Tes Kelas VIII

Pada gambar 8 di atas menyajikan nilai fungsi informasi dari suatu tes matematika kelas VIII dengan 30 butir. Perangkat tes ini

memiliki nilai fungsi informasi yang lebih tinggi dibandingkan kesalahan pengukuran pada rentang -3 sampai +3. Dengan demikian perangkat tes matematika kelas VII ini sesuai untuk peserta tes pada rentang kemampuan tersebut.

Tes Matematika Kelas IX

Hasil analisis perangkat tes matematika kelas IX menggunakan program Winsteps diperoleh beberapa informasi. Butir yang dianalisis sebanyak 30 dan peserta 441 siswa. Butir dengan tingkat kesukaran paling tinggi adalah butir nomor 28 dengan tingkat kesukaran yakni 1,58 dan butir paling mudah adalah butir nomor 20 dengan tingkat kesukaran yakni -1,89.

Responden yang memiliki kemampuan tertinggi pada siswa adalah peserta nomor 220 dengan kemampuan 5,07 dan peserta dengan kemampuan terendah peserta nomor 410 dengan kemampuan -3,09. Kemampuan terendah peserta tes disebabkan mungkin tidak dapat melakukan upaya penuh pada semua butir karena kondisi pengujian atau faktor pribadi (Jin & Wang, 2014, p.196). Pada output hasil analisis pemilihan peserta yang diterima sesuai dengan data adalah peserta dengan *outfit mean square* (MNSQ) $0,5 < MNSQ < 1,5$ dan nilai *point measure correlation* (Pt Mean Corr) tidak negatif. Hasil analisis menunjukkan bahwa terdapat 36 siswa yang tidak sesuai dengan kedua kategori.

Hasil analisis menunjukkan informasi tingkat kesukaran, kesalahan pengukuran, *infit*, *outfit* dan *Pt measure*. Pada output hasil analisis perangkat tes matematika kelas IX pemilihan item yang diterima sesuai dengan data adalah item dengan *outfit* MNSQ $0,5 < MNSQ < 1,5$ dan *Pt Mean Corr* tidak negatif. Hasil analisis menunjukkan bahwa terdapat 5 item yang tidak sesuai dengan kedua kategori. Item tersebut adalah soal nomor 8, 17, 21, 22, dan 26. Kelima soal yang tidak sesuai kategori tersebut memiliki *outfit* MNSQ $> 1,50$.

Hasil analisis menunjukkan kriteria tingkat kesukaran butir yang terbagi dalam tiga kategori. Soal dikategorikan sukar apabila b mendekati +2, dan kategori mudah apabila b sekitar -2. Soal dikategorikan sedang apabila $-1,00 < b < +1,00$. Hasil karakteristik butir soal perangkat tes matematika kelas IX disajikan pada Tabel 9.

Tabel 8. Karakteristik Butir Tes Kelas IX

Butir	Tingkat Kesukaran	Kriteria
1	-0,61	Sedang
2	-1,25	Mudah
3	1,25	Sukar
4	-0,41	Sedang
5	-0,65	Sedang
6	0,15	Sedang
7	0,53	Sedang
8	-1,22	Mudah
9	-1,72	Mudah
10	0,88	Sedang
11	0,66	Sedang
12	0,95	Sedang
13	1,22	Sukar
14	-0,36	Sedang
15	1,18	Sukar
16	1,13	Sukar
17	-0,41	Sedang
18	1,09	Sukar
19	-1,13	Mudah
20	-1,89	Mudah
21	-0,53	Sedang
22	-0,50	Sedang
23	-0,04	Sedang
24	1,03	Sukar
25	-1,16	Mudah
26	-1,13	Mudah
27	1,38	Sukar
28	1,58	Sukar
29	1,31	Sukar
30	-1,32	Mudah

Hasil analisis pada Tabel 8 di atas menunjukkan butir soal yang dikategorikan mudah sebanyak 8 butir, kategori butir sedang sebanyak 13 butir, dan kategori butir sukar sebanyak 9 butir

Hasil tes dapat digunakan untuk menentukan penyetaraan tes anchor. Kolom yang digunakan untuk keperluan penyetaraan adalah kolom nomor butir dan kolom tingkat kesukaran. Pada tes ini butir yang merupakan *anchor item* adalah butir nomor 21 sampai 30. Oleh sebab itu yang perlu diperhatikan untuk analisis lebih lanjut adalah *measure* butir 21 sampai dengan 30.

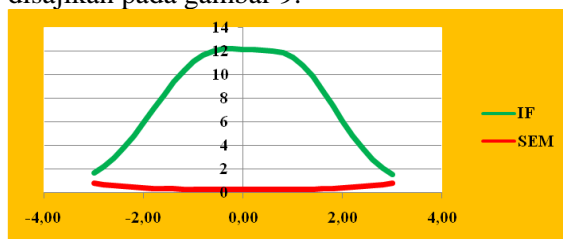
Hasil analisis diperoleh beberapa informasi untuk penyetaraan tes. Rangkuman penyetaraan soal kelas VIII dengan soal kelas IX, yang merupakan anchor, dapat dilihat pada Tabel 9. Selanjutnya dapat ditentukan persamaan kesetaraan antara tes anchor kelas VII ke tes anchor kelas VIII.

Tabel 9. Penyetaraan Tes Anchor Kelas VIII ke Kelas IX

Nomor Butir	Tes Anchor-2 Kelas IX	Tes Anchor-2 Kelas VIII
1	-0,53	0,41
2	-0,50	1,56
3	-0,04	0,26
4	1,03	-0,91
5	-1,16	-0,20
6	-1,13	1,17
7	1,38	0,31
8	1,58	0,95
9	1,31	-0,50
10	-1,32	-0,30

Fungsi Informasi Butir Kelas IX

Berdasarkan hasil analisis perangkat tes matematika kelas IX memiliki nilai fungsi informasi maksimum 12,2344 logit pada sekitar -0,2 dan SEM sebesar 0,2859. Hal ini berarti siswa kelas IX akan memberikan informasi yang baik, dengan kesalahan pengukuran terkecil, apabila dikerjakan oleh siswa yang memiliki kemampuan sekitar -0,2. Grafik hubungan fungsi informasi butir dan SEM disajikan pada gambar 9.



Gambar 9. Hubungan Fungsi Informasi Butir dan SEM Tes Kelas IX

Pada gambar 9 di atas menyajikan nilai fungsi informasi dari suatu tes matematika kelas IX dengan 30 butir. Perangkat tes ini memiliki nilai fungsi informasi yang lebih tinggi dibandingkan kesalahan pengukuran pada rentang -3 sampai +3. Dengan demikian perangkat tes ini sesuai untuk peserta tes pada rentang kemampuan tersebut.

Penyetaraan Skor dan Skala

Penyetaraan merupakan prosedur empiris yang diperlukan untuk menstransformasi skor suatu tes ke skor tes yang lain. Melalui proses penyetaraan yang dilakukan secara benar dapat diukur perkembangan dan pemetaan berbagai kecenderungan serta mendapat dan menggabungkan informasi dari perangkat tes berbeda. Penyetaraan yang dapat mengukur perkembangan adalah penyetaraan vertikal.

Hasil penyetaraan memungkinkan melakukan perbandingan atau konversi antara dua kemampuan yang mengukur perangkat tes yang berbeda. Cara memperoleh skor yang sama, kemampuan skor tes X dan Y dikonversikan pada skala utama (*common score scale*).

Teori respon butir memiliki sifat utama yakni invarian pada parameternya. Parameter ciri peserta adalah invarian terhadap butir, demikian pula ciri butir adalah invarian terhadap peserta. Penyetaraan skala dilakukan apabila parameter butir maupun parameter peserta belum diketahui. Estimasi terhadap parameter ini dapat dilakukan secara linier, dengan model logistik. Pada penelitian ini penentuan persamaan penyetaraan menggunakan rancangan tes anchor secara linear model logistik satu parameter. Pada model ini daya beda butir dianggap konstan, sehingga pada persamaan konversi hanya ada satu konstanta penskalaan. Penyetaraan berdasarkan teori respons butir menggunakan bantuan program Winsteps yang mengacu pada Model Rasch.

Hasil perhitungan rerata terhadap tingkat kesukaran tes anchor (V1), diperoleh besaran rerata tingkat kesukaran tes yang dikerjakan oleh siswa kelas VII sebesar 0,313; dan rerata tes yang dikerjakan siswa kelas VIII sebesar 0,056. Dengan demikian konstanta translasi butir sebesar 0,369. Persamaan konversi tingkat kesukaran butir kelas VII ke kelas VIII adalah:

$$Y^* = b_x + 0,369$$

Misalkan butir nomor 1 tes kelas VII memiliki $b_x = 0,03$ maka $Y^* = 0,39$. Hal ini sedikit berbeda dengan hasil konversi secara langsung menggunakan program yang sebesar -0,69. Perbedaan sebesar -1,09 masih di bawah 0,5 sehingga masih dapat diterima. Hasil penyetaraan tingkat kesukaran butir tersebut menunjukkan bahwa tingkat kesukaran butir kelas VII mengalami penurunan apabila dikonversi terhadap kelas VIII. Hal ini apabila tes kelas VII dikerjakan oleh siswa kelas VIII maka tingkat kesukarannya menjadi berkurang. Cara yang sama dapat dilakukan untuk menentukan penyetaraan skala tingkat kesukaran butir kelas VII ke kelas VIII yang lain.

Besar rerata tingkat kesukaran tes anchor (V2) yang dikerjakan oleh siswa kelas IX = 0,275; dan rerata tingkat kesukaran tes yang dikerjakan oleh siswa kelas VIII = 0,062. Dengan demikian konstanta translasi butir

sebesar 0,337. Persamaan konversi tingkat kesukaran butir kelas IX ke kelas VIII adalah:

$$Y^* = b_z + 0,337$$

Misalkan butir nomor 1 tes kelas IX memiliki $b_z = -0,61$ maka $Y^* = -0,27$. Hal ini sedikit berbeda dengan hasil konversi secara langsung menggunakan program yang sebesar -0,69. Perbedaan sebesar -0,42 masih di bawah 0,5 sehingga masih dapat diterima. Hasil penyetaraan tingkat kesukaran butir tersebut menunjukkan bahwa tingkat kesukaran butir kelas IX mengalami peningkatan apabila dikonversi terhadap kelas VIII. Hal ini apabila tes kelas IX dikerjakan oleh siswa kelas VIII maka tingkat kesukarannya menjadi meningkat atau lebih sulit. Cara yang sama dapat dilakukan untuk menentukan penyetaraan skala tingkat kesukaran butir kelas IX ke kelas VIII yang lain.

Hasil penyetaraan dapat juga dilihat pada hasil analisis program Winsteps menunjukkan hasil konversi antara tingkat kesukaran butir tes kelas VII dengan kelas VIII. Hasil akhir menunjukkan bahwa tingkat kesukaran butir tes kelas VII mengalami penurunan apabila dikonversikan terhadap kelas VIII. Hal ini berarti, apabila tes kelas VII ini dikerjakan oleh siswa kelas VIII, maka tingkat kesukarannya menjadi berkurang. Hasil akhir juga menunjukkan bahwa tingkat kesukaran butir tes kelas IX mengalami peningkatan apabila dikonversikan terhadap kelas VIII. Hal ini berarti, tes kelas IX ini dikerjakan oleh siswa kelas VIII, maka tingkat kesukarannya menjadi lebih tinggi.

Tingkat kesukaran butir yang telah disetarakan skalanya dapat diperbandingkan satu dengan yang lain. Dengan dilakukan penyetaraan skala ini, dapat diketahui butir yang memiliki tingkat kesukaran paling tinggi dan paling rendah. Hasil analisis menunjukkan bahwa perangkat tes kelas VII, kelas VIII, dan kelas IX mempunyai tingkat kesukaran yang hampir sama. Hal ini ditunjukkan oleh skor rerata tingkat kesukaran perangkat tes yang telah dikonversikan dalam skala WITs yaitu sebesar 498,90, kelas VIII sebesar 503,25 dan kelas VII sebesar 494,00.

Penentuan keakuratan penyetaraan menggunakan *root mean square difference* (RMSD) dari nilai estimasi parameter tingkat kesukaran (b) sebelum dan sesudah disetarakan pada butir 2 diperoleh hasil analisis untuk perangkat tes kelas VII sebesar 0,098, perangkat kelas VII sebesar 0,024 dan perangkat kelas IX sebesar 0,026. Berdasarkan hasil analisis RMSD pada butir 2 pada ketiga perangkat tes

tersebut menunjukkan nilai RMSD sangat kecil sehingga dapat disimpulkan bahwa metode penyetaraan yang digunakan cukup baik. Penyetaraan dengan pendekatan teori respons butir memberikan hasil penyetaraan yang lebih konsisten. Hal ini sesuai yang dikemukakan Powers & Kolen (2014, p.1) mengemukakan keakuratan hasil penyetaraan sangat penting ketika membandingkan skor ujian dalam bentuk pengujian. Hal tersebut dikarenakan dalam teori respons butir terjadi invariansi diantara butir dan peserta, atau dapat dikatakan bahwa ciri butir adalah invarian sekalipun pesertanya berbeda-beda serta ciri peserta adalah invarian sekalipun butirnya berbeda-beda.

Penyetaraan Kemampuan Siswa

Hasil analisis program Winsteps memperlihatkan kesetaraan antara kemampuan siswa kelas VII dengan kemampuan siswa kelas VIII. Hasil penyetaraan menunjukkan bahwa kemampuan siswa kelas VII menjadi lebih rendah setelah disesuaikan dengan skala kemampuan siswa kelas VIII. Hasil analisis kesetaraan antara kemampuan siswa kelas IX ke kemampuan siswa kelas VII menunjukkan bahwa kemampuan siswa kelas IX menjadi lebih tinggi setelah disesuaikan dengan skala kemampuan siswa kelas VIII.

Persamaan konversi untuk kemampuan mengikuti persamaan konversi ciri butir. Konstanta penskalaannya sebesar 0,369 untuk konversi kemampuan kelas VII ke kelas VIII, dan -0,03 untuk konversi kemampuan kelas IX ke kelas VIII. Persamaan konversi kemampuan kelas VII ke kelas VIII adalah:

$$\theta_y = \theta_x + 0,369$$

misalkan peserta kelas VII memiliki $\theta_x = -3,59$ maka $\theta_y = -3,22$. Hal ini sedikit berbeda dengan hasil konversi secara langsung menggunakan program yang sebesar -3,63. Perbedaan sebesar 0,41 masih di bawah 0,5 sehingga masih dapat diterima. Hasil penyetaraan tingkat kemampuan tersebut menunjukkan bahwa tingkat kemampuan siswa kelas VII mengalami penurunan apabila dikonversikan terhadap kemampuan siswa kelas VIII. Hal ini apabila siswa kelas VII mengerjakan tes kelas VIII maka kemampuannya akan mengalami penurunan. Cara yang sama dapat dilakukan untuk menentukan penyetaraan kemampuan siswa kelas VII ke kelas VIII yang lain.

Persamaan konversi kemampuan kelas IX ke kemampuan kelas VIII adalah:

$$\theta_y^* = \theta_z + 337$$

misalkan peserta kelas IX memiliki $\theta_x = -3,85$ maka $\theta_y = -3,51$. Hal ini sedikit berbeda dengan hasil konversi secara langsung menggunakan program yang sebesar $-3,63$. Perbedaan sebesar $0,12$ masih di bawah $0,5$ sehingga masih dapat diterima. Hasil penyetaraan tingkat kemampuan tersebut menunjukkan bahwa tingkat kemampuan siswa kelas IX mengalami peningkatan apabila dikonversikan terhadap kemampuan siswa kelas VIII. Hal ini apabila siswa kelas IX mengerjakan tes kelas VIII maka tingkat kemampuannya akan mengalami meningkat. Cara yang sama dapat dilakukan untuk menentukan penyetaraan kemampuan siswa kelas IX ke kelas VIII yang lain.

Kemampuan peserta tes yang telah disetarakan dapat diperbandingkan satu dengan yang lain. Dengan dilakukan penyetaraan skala ini, dapat mengetahui peserta yang memiliki tingkat kemampuan paling tinggi dan siswa yang memiliki kemampuan paling rendah. Hasil menunjukkan bahwa kemampuan siswa kelas IX lebih tinggi dibandingkan kemampuan kelas VIII maupun kelas VII. Hal ini ditunjukkan oleh skor rerata tingkat kemampuan peserta tes yang telah dikonversikan dalam skala WITs, yaitu $507,85$ untuk kelas IX, $507,73$ untuk kelas VIII, dan $507,53$ untuk kelas VII.

Hasil akhir dari penelitian ini menunjukkan bahwa perangkat tes matematika kelas IX merupakan tes dengan tingkat kesulitan sedikit lebih tinggi dibanding dua perangkat tes yang berada pada level di bawahnya, yaitu tes matematika kelas VIII maupun tes matematika kelas VII. Hal ini ditunjukkan oleh skor rerata tingkat kemampuan peserta tes yang telah dikonversikan dalam skala WITs yaitu $507,85$ untuk kelas IX, $507,73$ untuk kelas VIII, dan $507,53$ untuk kelas VII.

Selain itu berdasarkan hasil konversi melalui pendekatan teori respon butir, diperoleh hasil bahwa kemampuan siswa kelas VIII lebih tinggi dibanding kemampuan siswa kelas VII dan kemampuan siswa kelas IX lebih tinggi dibandingkan kemampuan siswa kelas VIII, atau dapat dikatakan bahwa siswa kelas IX mempunyai kemampuan yang paling tinggi. Hal ini dikarenakan rerata kemampuan siswa kelas IX lebih tinggi dari pada dua kelas di bawahnya, sehingga meskipun mengerjakan tes dengan tingkat kesukaran yang tinggi tetapi karena sesuai dengan kemampuannya maka hasilnya juga akan baik.

Hal lain yang mendukung hasil penelitian adalah dalam pelaksanaan penelitian dapat dilihat bahwa kelas IX lebih serius dalam mengerjakan tes. Selain itu sebelum penelitian dilakukan, beberapa sekolah telah melaksanakan tes pendalaman materi untuk melihat kesiapan siswa kelas IX dalam menghadapi ujian nasional, sehingga siswa kelas IX lebih siap dalam mengerjakan soal.

Simpulan dan Saran

Simpulan

Hasil analisis kuantitatif dengan pendekatan teori respon butir model Rasch pada perangkat tes matematika kelas VII yang berjumlah 30 butir terdapat kategori mudah 3 butir, kategori sedang terdapat 24 dan kategori sukar terdapat 3 butir. Pada perangkat tes matematika kelas VIII yang berjumlah 30 butir terdapat kategori mudah 1 butir, kategori sedang 24 butir dan kategori sukar 5 butir. Pada perangkat tes matematika kelas IX yang berjumlah 30 butir terdapat kategori mudah 8 butir, kategori sedang 13 butir dan kategori sukar 9 butir.

Hasil penyetaraan vertikal dengan teori respon butir diperoleh persamaan konversi untuk kelas VII ke kelas VIII adalah $Y^* = bx + 0,369$ dan persamaan konversi untuk kelas IX ke kelas VIII adalah $Y^* = bz + 0,337$.

Hasil penyetaraan tingkat kemampuan matematika siswa dapat disimpulkan bahwa kemampuan siswa kelas VIII lebih tinggi dibandingkan kemampuan siswa kelas VII, dan kemampuan siswa kelas IX lebih tinggi dibandingkan kemampuan siswa kelas VIII. Sehingga dapat disimpulkan bahwa kelas IX mempunyai kemampuan paling tinggi dibandingkan dua level kelas di bawahnya.

Saran

Hasil penelitian merekomendasikan beberapa saran yaitu; (1) Bagi pihak sekolah khususnya guru matematika tingkat SMP dapat menggunakan instrumen tes matematika ini untuk mengetahui tingkat perkembangan kemampuan siswa maupun untuk menyiapkan siswa menghadapi ujian semester, ujian kenaikan kelas maupun ujian nasional, (2) Bagi Dinas Pendidikan Kota Pekalongan untuk membuat program terencana yang dapat meningkatkan kemampuan siswa khususnya kemampuan siswa kelas VII yang mempunyai tingkat kemampuan paling rendah, dan (3) Bagi pene-

litian selanjutnya perlu sekiranya dirancang instrumen tes dengan butir bersama (*anchor item*) antara kelas VII, kelas VIII dan kelas IX, maupun item bersama antara kelas VII dan kelas IX, sehingga dapat memperoleh persamaan konversi kelas VII ke kelas IX dan sebaliknya. Sehingga dapat mengungkap langsung perkembangan kemampuan siswa dari kedua kelas tersebut.

Daftar Pustaka

- Anak Agung Purwa Antara. (2014). *Penyetaraan vertikal dengan pendekatan klasik dan item response theory pada siswa sekolah dasar*. Artikel Disertasi Doktor. Yogyakarta: Program Pascasarjana Universitas Negeri Yogyakarta.
- Badan Standar Nasional Pendidikan. (2010). *Paradigma pendidikan nasional abad xxi*. Jakarta: BSNP.
- Becker, Kirk A & Bergstrom, Betty A. (2013). *Test administration models*. Journal of Practical Assessment, Research & Evaluation, Vol 18, Number 14, pp.1-7.
- Crocker, Linda., & Algina, James. (2008). *Introduction to classical and modern test theory*. New York: Holt, Rinehart And Winston, Inc.
- Dinas Pendidikan Pemuda dan Olahraga Kota Pekalongan. (2014). *Daftar SMP*. Diambil pada tanggal 4 Agustus 2014, dari <http://dindikporakotapekalongan.or.id/>
- Djemari Mardapi. (2012). Pengukuran penilaian dan evaluasi pendidikan. Yogyakarta: Nuha Medika.
- Hambleton, Ronald. K., & Swaminathan, H. (1985). *Item response theory*. Boston: Kluwer, Inc.
- Hambleton, Ronald. K., Swaminathan, H., & Rogers, H. Jane. (1991). *Fundamentals of item response theory*. California: Sage Publications, Inc.
- Heri Retnowati. (2014). *Teori respon butir dan penerapannya*. Yogyakarta: Nuha Medika
- Jin, Kuan-Yu., & Wang, Wen-Chung. (2014). *Item response theory models for performance decline during testing*. Journal of Educational Measurement, Vol 51, Number 2, pp.178-200.
- Kementerian Pendidikan dan Kebudayaan. (2014). *Materi pelatihan implementasi kurikulum 2013 mata pelajaran matematika SMP/MTs*. Jakarta: BPSDMPK-PMP Kemendikbud.
- Kolen, Michael J., & Brennan, Robert L. (1995). *Test equating: methods and practices*. New York: Springer-Verlag, Inc.
- Kolen, Michael J., & Brennan, Robert L. (2004). *Test equating: scaling and linking methods and practices*. (2nd Ed). New York: Springer-Verlag, Inc.
- Peterson, N.S., Kolen, M.J., & Hoover, H.D. (1989). *Scaling, norming, and equating*. Dalam R.L. Linn (Eds), *Educational Measurement*. New York: Macmillan Publishing Company.
- Powers, Sonya., & Kolen, Michael J. (2014). *Evaluating equating accuracy and assumptions for groups that differ in performance*. Journal of Educational Measurement, Vol 51, Number 2, pp.39-56.
- Rasch, Georg. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: The University of Chicago Press.
- Schwartz, Sydney L. (2005). *Teaching young children mathematics*. London: Praeger Publishers.
- Sugeng. (2010). Penyetaraan vertikal model kredit parsial soal matematika SMP. *Jurnal Penelitian dan Evaluasi Pendidikan, Vol 14, Nomor 2*, 289-308.