

**VERTICAL EQUATING SKOR TES DENGAN PENSKALAAN DIKOTOMUS  
PADA PELAJARAN MATEMATIKA  
DI SMP NEGERI 1 MARGA**

**Anak Agung Purwa Antara**  
IKIP Saraswati  
(purwa antara@gmail.com)

**ABSTRAK**

Penelitian ini bertujuan untuk mengetahui: (1) perkembangan kemampuan siswa pada pelajaran matematika dari kelas VII, VIII, dan IX SMP Negeri 1 Marga tahun 2017; (2) rerata kemampuan siswa kelas VII, VIII dan IX SMP Negeri Marga pada tes akhir semester 2017; dan (3) kualitas tes prestasi belajar yang dikembangkan dalam penelitian ini. Data penelitian adalah skor siswa Sekolah Menengah Pertama (SMP) Negeri 1 Marga Tabanan kelas VII, VIII dan IX pada tes sumatif tahun 2017. Model penskalaan yang digunakan adalah model dikotomus pada tes bentuk pilihan ganda yang dikembangkan menggunakan analisis model *rasch*. Ketiga paket tes dihubungkan dengan butir-butir bersama atau butir *anchor* sebanyak 20%. Hasil penelitian menunjukkan bahwa (1) kemampuan siswa ( $\theta$ ) pada pelajaran matematika mengalami peningkatan sesuai dengan meningkatnya peringkat kelas berturut-turut sebesar -0.600, 0.122, dan 0.439; (2) rerata prestasi belajar matematika siswa pada pelajaran matematika kelas VII, VIII, IX berturut-turut sebesar -0.600, -1.105, -0.591; (3) kualitas ketiga tes yang dikembangkan cukup baik dengan indikator nilai rerata logit item 0.00 yang menunjukkan instrumen memang mengukur, nilai Infit *MNSQ* dan Outfit *MNSQ* mendekati 1.00 (memenuhi kriteria), nilai Infit *ZSTD* dan Outfit *ZSTD* mendekati 0.00 (memenuhi kriteria), separation berturut-turut 3.86, 2.77, dan 3.03 yang berarti ketiga tes memiliki daya beda yang baik, dan nilai reliability berturut-turut 0.94, 0.88, dan 0.90 yang menunjukkan ketiga tes memiliki reliabilitas yang baik.

*Kata Kunci: Vertical Equating, Tes Model Rasch*

**ESTIMATING DEVELOPMENT STUDENT'S ABILITIES IN MATH  
TROUGH LEARNING ACHIEVEMENT TEST  
IN SMP NEGERI 3 TABANAN**

**ABSTRACT**

*This study aims to determine: (1) the development of students' abilities in math from grade VII, VIII, and IX Junior High School; (2) average of student ability class VII, VIII and IX Junior High School in final test semester; and (3) the quality of learning achievement test developed in this study. The research data is score junior secondary students in SMP Negeri 1 Marga Tabanan on summative tests in 2017. The research instrument is the mathematics achievement test multiple choice form that was developed using Rasch model. The third test packets associated with a grain anchor 20%. The results showed that (1) the ability of the student ( $\theta$ ) in math has increased along with the increasing class rank respectively at -0.600, 0.122, and 0.439; (2) the average mathematics achievement of students in math class VII, VIII, IX respectively -0.600, -1.105, -0.591; (3) the quality of third test developed quite well with indicators of the average value of the logit item 0.00 showing the instrument is measuring, the value Infit *MNSQ* and Outfit*

*MNSQ approaching 1.00 (conform with the criteria), value Infit ZSTD and Outfit ZSTD approaching 0.00 (conform with the criteria), separation respectively 3.86, 2.77, 3:03 that means three different tests have good power, reliability and value respectively 0.94, 0.88, and 0.90 showing the third test has good reliability.*

Keywords: *Vertical Eqquating, Rasch Model Test*

## PENDAHULUAN

Persaingan global di berbagai sektor kehidupan membawa konsekuensi akan tuntutan penyediaan sumber daya manusia yang berkualitas sehingga mampu bersaing dalam kehidupannya. Hal ini berarti bahwa sekolah sebagai suatu lembaga pendidikan harus selalu meningkatkan kualitasnya agar lulusan yang dihasilkan memiliki kompetensi yang memadai. Menghadapi tantangan ini pemerintah Indonesia telah melakukan berbagai upaya di antaranya peningkatan kompetensi guru melalui sertifikasi dan penyempurnaan kurikulum pada seluruh jenjang pendidikan. Perubahan kurikulum hendaknya jangan dipandang sebagai suatu momok, karena kurikulum adalah sesuatu yang dinamis, yang selalu akan berubah mengikuti tuntutan kualitas sumber daya yang mesti dihasilkan agar dapat mengikuti perkembangan kehidupan di masa yang akan datang.

Ada empat komponen yang selalu menjadi perhatian dalam setiap perubahan kurikulum yaitu (1) penentuan syarat kelulusan peserta didik yang lebih dikenal dengan standar kompetensi lulusan yaitu kompetensi yang harus dimiliki peserta didik setelah mengikuti suatu proses pembelajaran; (2) materi pelajaran yang mesti diberikan agar kompetensi yang mesti

dimiliki tersebut dapat tercapai; (3) metode atau strategi pembelajaran yang mesti digunakan dalam proses pembelajaran agar materi pelajaran yang diberikan dalam rangka mencapai kompetensi yang diinginkan dapat terserap dengan baik oleh peserta didik; dan (4) evaluasi yang mesti dilakukan agar dapat diketahui apakah tujuan pendidikan yang diharapkan tersebut benar-benar telah tercapai.

Evaluasi dalam lembaga pendidikan meliputi proses penilaian (asesmen) dan proses pengukuran. Menurut Mardapi (2008: 5), suatu lembaga pendidikan dapat memulai peningkatan kualitas lulusan yang dihasilkan melalui peningkatan kualitas program pembelajaran dan peningkatan kualitas penilaian yang dilakukan guru di dalam kelas. Hal tersebut memberi makna bahwa guru sebagai komponen yang bersentuhan langsung dengan peserta didik (siswa) hendaknya mampu mengembangkan potensi-potensi yang dimiliki oleh siswanya, baik melalui pelaksanaan program pembelajaran yang dilaksanakan di dalam kelas, maupun melalui model penilaian yang digunakan untuk mengukur keberhasilan program pembelajaran yang dilaksanakan. Model penilaian yang digunakan guru, hendaknya mampu memberikan informasi yang

lengkap dan akurat tentang perkembangan kemampuan siswa selama mengikuti program pembelajaran, mampu memberi informasi tentang hasil belajar yang telah dicapai siswa setelah mengikuti proses pembelajaran, sehingga dapat dilakukan evaluasi untuk perbaikan menuju pencapaian tujuan pendidikan yang diharapkan (Antara, 2014: 2).

Salah satu materi pelajaran pada kurikulum Sekolah Menengah Pertama (SMP) adalah matematika. Matematika adalah kelompok ilmu dasar yang mesti dikuasai siswa. Materi pada pelajaran matematika sangat kompleks, abstrak dan bersifat hierarkis. Hal ini berarti bahwa penguasaan siswa terhadap materi matematika pada kelas tertentu, berkaitan dengan penguasaan materi matematika pada kelas sebelumnya. Dengan demikian untuk memperoleh informasi perkembangan kemampuan siswa pada pelajaran matematika di SMP perlu diketahui bagaimana perkembangan kemampuan matematika siswa pada seluruh peringkat kelas.

Guru sering dihadapkan pada suatu masalah dalam melakukan pengukuran terhadap kemampuan siswanya pada suatu mata pelajaran. Masalah tersebut terjadi apabila terdapat perbedaan kemampuan siswa yang ekstrim dalam satu kelas. Siswa yang pintar kemungkinan menjawab seluruh butir soal yang diberikan dan bahkan mampu menjawab dengan benar butir soal pada tingkat kelas di atasnya, sebaliknya siswa yang lemah kemungkinan tidak mampu menjawab seluruh butir soal yang diberikan sehingga guru tidak

memperoleh informasi apa-apa tentang kemampuannya.

Menurut Naga (1992: 347), cara untuk mengatasi masalah tersebut adalah dengan memberikan siswa yang kemampuannya lambat tes yang berada pada tingkat di bawahnya, sedangkan siswa yang kemampuannya cepat, diberikan tes pada tingkat di atasnya. Kemudian melalui penyamaan skala, skor yang mereka peroleh dikonversi kembali pada kedudukan semula. Kolen & Brennan (2004: 5) menyatakan bahwa, skor dua tes berbeda dari dua kelompok atau lebih dapat diperbandingkan jika butir-butir soalnya setara dan memiliki skala yang sama. Kesetaraan antar skor dapat dilakukan secara statistik.

Menurut Angoff (1971: 563), dua kelompok skor, satu kelompok pada instrumen X dan yang lain pada instrumen Y dengan instrumen X dan instrumen Y mengukur hal yang sama dan memiliki derajat reliabilitas sama, dapat dipertimbangkan ekuivalen jika rank-rank persentilnya yang bersesuaian adalah sama. Hal ini memberikan dasar untuk melakukan pertukaran skor butir soal pada dua instrumen atau lebih yang memiliki kelompok skor ekuivalen. Hubungan antar skor dari dua instrumen tes merupakan bentuk transformasi dari skor salah satu instrumen tes kepada skor instrumen tes lainnya. Hubungan (*linking*) antar instrumen tes dapat dikelompokkan menjadi tiga yaitu penyetaraan (*equiting*), *concordance*, dan prediksi (*prediction*) (Kollen & Brennan, 2004). Ketiga hubungan ini dibedakan oleh konstruk tes

dan distribusinya.

Hollan & Dorans (2006: 187) menyatakan bahwa, “*we use the term linking to refer the general idea of a transformation between the scores from one test and those of another. We will divide linking methods into three basic categories called predicting, scale aliging, and equating*”. *Predicting* bertujuan memprediksi hasil terbaik skor peserta pada tes yang satu (Y) berdasarkan instrumen tes lain (X) dan informasi lainnya berkenaan peserta tersebut. *Scaling* bertujuan untuk menstranformasikan skor-skor instrumen tes X dan tes Y ke suatu skala bersama (*common scale*) sehingga keduanya dapat saling diperbandingkan. *Equating* bertujuan menghasilkan skor dari instrumen tes yang satu (X) terhadap tes lainnya (Y) sehingga skor-skor keduanya dapat saling bertukar.

Perbandingan skor antara tes yang berbeda tetapi mengukur kemampuan yang sama, merupakan hal yang penting dalam pengukuran. Perlu dilakukan penyetaraan (*equating*) untuk membandingkan skor yang didapat dari dua tes yang berbeda. Menurut Kolen dan Brennan, (1995: 2, 2004: 2), “*equating is a statistical process that is used to adjust scores on test forms so that scores on the forms can be use interchangeably*”. Sedangkan menurut Peterson, Kolen, & Hoover (Linn, 1989: 242), “*equating methods are empirical procedures for establishing a relation between raw scores on two test forms that can that be used to express the scores on one form in terms of the scores on the other form*”. Pernyataan tersebut dapat dimaknai

bahwa, *equating* atau penyetaraan adalah prosedur empiris yang melibatkan proses statistik untuk menyesuaikan skor-skor dari dua atau lebih instrumen tes sehingga dapat dipergunakan saling bertukar. Tujuan penyetaraan skor tes adalah untuk menyamakan skor tes, sehingga skor tersebut dapat diperbandingkan. Suatu skor dapat diperbandingkan dengan skor yang lain, jika keduanya mengukur karakteristik kemampuan yang sama dan dinyatakan dalam metrik yang sama. Dengan demikian walaupun kedua skor itu mengukur karakteristik kemampuan yang sama tetapi kalau metrik atau skalanya berbeda maka kedua skor tersebut tidak dapat dibandingkan.

*Equating* tes dapat diklasifikasikan menjadi dua, yaitu *equating* horizontal dan vertikal (Hambleton & Swaminathan, 1985: 197; Crocker & Algina, 1986: 456; Felan, 2002: 6). Penyetaraan horizontal adalah penyetaraan tes yang dirancang dengan harapan paralel dalam hal isi, dan tingkat kesukaran item, serta diberikan pada kelompok-kelompok peserta tes yang memiliki distribusi kemampuan yang diharapkan sama (Crocker & Algina, 1986: 456; Hambleton & Swaminathan, 1985: 197), seperti ujian Akhir nasional (UAN), dan seleksi penerimaan mahasiswa baru (SPMB). Sedangkan *equating* vertikal adalah suatu usaha untuk menyetarakan skor tes pada dua naskah atau lebih yang dirancang secara sengaja berbeda dalam tingkat kesukaran butir tetapi masih mengukur karakteristik yang sama, diujikan pada peserta tes yang mempunyai distribusi kemampuan berbeda (Crocker & Algina,

1983: 456; Hambleton & Swaminathan, 1985: 198; Felan, 2002: 6).

Definisi tentang *equating* vertikal dikemukakan banyak ahli di antaranya, “*vertical equating involves two or more tests that are supposed to measure the same trait and are designed to be different in difficulty*”. (Crocker & Algina, 1986: 473). Sedangkan Kolen & Brennan (1995: 3) menyatakan bahwa, *equating* vertikal adalah penyetaraan skor tes dengan *content* tidak berbeda tetapi kelompok peserta tes berasal dari tingkat kelas yang berbeda dan skor tes yang demikian dapat digunakan saling bertukar. *Content* tidak berbeda dimaksudkan adalah isi tes dan karakteristik kemampuan yang diukur sama. Isi tes dapat dilihat dari kisi-kisi tes yaitu matrik yang menunjukkan bahan tes serta tingkat berpikir yang terlibat dalam mengerjakan tes. Isi tes mengacu kepada tema, kata-kata, format butir, tugas atau pertanyaan dan pedoman penskorannya (Mardapi, 2008: 9).

*Equating* sangatlah penting bagi guru maupun lembaga penyelenggara pendidikan atau sekolah karena melalui *equating* dapat diketahui perbandingan kemampuan siswa dalam suatu mata pelajaran dibandingkan dengan sekolah lainnya, juga perkembangan kemampuan siswanya dalam suatu mata pelajaran pada seluruh peringkat kelas.

Informasi tentang perkembangan kemampuan siswa pada suatu mata pelajaran pada seluruh peringkat kelas akan sangat bermanfaat baik bagi guru maupun lembaga penyelenggara pendidikan dalam mengambil keputusan baik dalam

rangka perbaikan program pembelajaran maupun proses pembelajaran yang akan dilakukan berikutnya sehingga kualitas hasil belajar dapat ditingkatkan. Perkembangan kemampuan siswa dalam suatu mata pelajaran dapat diestimasi melalui penyetaraan vertikal skor tes yang diberikan pada seluruh peringkat kelas.

Penelitian ini bertujuan untuk mengetahui: (1) perkembangan kemampuan siswa pada pelajaran matematika dari kelas VII, VIII, dan IX SMP; (2) prestasi belajar matematika siswa pada pelajaran matematika kelas VII, VIII, IX SMP; dan (3) kualitas tes dalam penelitian ini yang dikembangkan dengan analisis model rasch.. Subjek penelitian adalah siswa kelas VII, VIII, dan IX SMP Negeri 3 Tabanan, dengan menggunakan tes dengan penskalaan dikotomis pada model tes bentuk pilihan ganda.

Secara teoretis, penelitian ini diharapkan dapat memberikan kontribusi bagi guru matematika dalam memperoleh informasi tentang perkembangan kemampuan siswa pada suatu mata pelajaran, dan memberikan informasi tentang analisis perkembangan kemampuan siswa menggunakan metode penyetaraan secara vertikal. Secara praktis hasil penelitian ini dapat dimanfaatkan oleh guru dalam menentukan strategi, metode pembelajaran untuk meningkatkan kualitas hasil belajar siswa khususnya pada pelajaran matematika.

## **METODE PENELITIAN**

Penelitian ini adalah penelitian empirik tentang perkembangan ke-

mampuan siswa yang dilakukan melalui *equating* (penyetaraan) vertikal. Secara umum pendekatan yang digunakan adalah pendekatan kuantitatif, dan dalam pelaksanaannya terdiri atas beberapa bagian penelitian yang saling terkait yaitu penyusunan instrumen penelitian, analisis karakteristik instrumen yang disusun, dan analisis tentang perkembangan kemampuan siswa.

Pada bagian pengembangan instrumen, disusun instrumen (tes) prestasi belajar matematika model pilihan ganda untuk kelas VII, VIII dan IX SMP (tes sumatif) yang disusun berdasarkan pokok bahasan bilangan, aljabar, geometri, pengukuran, statistika dan peluang sesuai dengan Kurikulum 2013. Ketiga paket tes tersebut memuat butir-butir soal yang sama sebagai butir *anchor*. Banyaknya butir *anchor* ditentukan sebanyak 20%. Dasar yang digunakan penetapan banyaknya butir *anchor* mengacu pada teori bahwa banyak butir *anchor* minimal 20% (Hambleton, Swaminathan, & Rogers, 1991:135).

Uji validitas isi dilakukan oleh *expert* yang melibatkan dua ahli dalam bidang pendidikan matematika dan pengukuran. Teknik analisis yang digunakan adalah teknik yang dikembangkan oleh Gregory (2000: 123). Selain itu analisis soal juga melibatkan 10 orang siswa pada masing-masing tingkat kelas untuk mendapatkan informasi keterbacaan soal. Perbaikan dan penyempurnaan draf soal yang dikembangkan dilakukan sesuai dengan petunjuk *expert*. Instrumen (tes) yang telah diperbaiki selanjutnya diujicobakan

di SMP Negeri 1 Marga Tabanan untuk memperoleh validitas konstruk.

Analisis data uji coba menggunakan pemodelan *Rasch* dengan bantuan program *winsteps* (Linarce, 1994: 328). Indikator yang digunakan adalah *infit-outfit meansquare*, *outfit* dan *infitzstandard*, *item reliability*, *item separation*, dan *unidimensionality*. Kriteria *Outfit Mean Square* adalah  $0.5 < MNSQ < 1.5$ ; *Outfit Z-Standard* yang diterima adalah  $-2.0 < ZSTD < 2.0$ ; *reliability*  $\geq 0,7$ ; *separation*  $\geq 0.2$  dan kriteria dimensialitas yang diterima adalah *raw variance explained by measure*  $\geq 20\%$  dan *unexplained variance*  $< 10\%$ . Pengumpulan data hasil penelitian dilakukan dengan teknik random dengan sampel sebanyak 88 orang kelas VII, 59 orang kelas VIII, dan 69 orang kelas IX.

Bagian berikutnya adalah melakukan proses penyamaan skala pengukuran melalui *equating* terhadap data hasil ujian akhir semester (tes sumatif) dengan metode *mean&sigma*. Keakuratan metode diuji dengan *Root Mean Square Difference (RMSD)*. Informasi tentang perkembangan kemampuan siswa dalam pelajaran matematika dapat diperoleh dari hasil penyetaraan skor dan skala kemampuan siswa pada ketiga kelas yang dilibatkan.

## HASIL PENELITIAN

### a. Pengembangan Instrumen Penelitian

Uji validitas isi dilakukan untuk mengetahui kesesuaian secara logika antara butir-butir soal yang disusun dengan kisi-kisi yang dikembangkan. Dengan teknik Gregory diperoleh nilai validitas isi

(*content validity; CV*) untuk tes kelas VII, VIII, dan IX berturut-turut sebesar 0.775; 0.750; dan 0.725. Kriteria yang digunakan  $CV > 0.700$ . Dengan demikian ketiga tes yang dikembangkan memiliki validitas isi (*content validity*) yang baik sehingga dapat dilakukan proses selanjutnya.

Uji validitas konstruk terhadap data hasil ujicoba menggunakan program *winsteps* disajikan pada Tabel 1.

Tabel 1. Ringkasan Hasil Running *Winsteps*

Komponen		Tes Kls VII	Tes Kls VIII	Tes Kls IX
Nilai rerata logit item		0.00	0.00	0.00
Mean-square ( <i>MNSQ</i> ):	<i>Infit</i>	1.00	1.00	1.01
	<i>Outfit</i>	1.06	1.13	1.03
<i>Zstandard</i> ( <i>ZSTD</i> ):	<i>Infit</i>	-0.2	-0.4	-0.2
	<i>Outfit</i>	0.00	-0.10	-0.2
<i>Item separation</i>		3.86	2.77	3.03
Dimen- sionality	<i>Raw variance explained by measure <math>\geq</math> 20%</i>	23.3	22.4	23.1
	<i>Unexplned variance in 1st contrast</i>	8.4	19.2	12.2
	<i>Unexplned variance in 2nd contrast</i>	7.1	9.1	8.8
	<i>Unexplned variance in 3rd contrast</i>	6.0	6.7	7.7
	<i>Unexplned variance in 4th contrast</i>	4.5	5.8	5.8
	<i>Unexplned variance in 5th contrast</i>	4.1	4.8	4.4
	Alpha Cronbach		0.94	0.88

Tabel 1 menunjukkan bahwa kualitas keempat tes yang dikembangkan cukup baik. Hal tersebut dapat dilihat dari indikator nilai rerata logit item disekitar 0.00 yang menunjukkan instrumen memang mengukur. Nilai *Infit MNSQ* dan *Outfit*

*MNSQ* mendekati 1.00, serta nilai *Infit ZSTD* dan *Outfit ZSTD* mendekati 0.00 yang menunjukkan bahwa data yang dihasilkan *fit* dengan model yang dikembangkan atau dengan kata lain, data yang dihasilkan sesuai dengan tujuan pengukurannya.

Nilai separation berturut-turut 3.86, 2.77, dan 3.03 yang berarti ketiga tes memiliki daya beda yang baik karena ketiga tes yang dikembangkan dapat membedakan kemampuan siswa lebih dari dua kelompok. Demikian pula nilai Cronbach Alpha berturut-turut 0.94, 0.88, dan 0.90 yang menunjukkan ketiga tes cukup handal atau memiliki reliabilitas yang baik.

## b. Analisis Data Penelitian

Data hasil penelitian disajikan secara deskriptif pada Tabel 2 sebagai berikut:

Tabel 2. Deskripsi Data

	Kelas VII	Kelas VIII	Kelas IX
Mean	18,773	12,169	15,319
Skor Maksimum	29	17	28
Skor Minimum	5	3	7
Median	19	13	16

Skor minimum tes kelas VII, VIII dan IX sebesar 0 dan skor maksimum sebesar 40 sehingga rerata skor ideal adalah 20. Data hasil pengukuran yang disajikan pada Tabel 1 menunjukkan bahwa rerata skor (*mean*) kelas VII, VIII, dan IX berturut-turut sebesar 18.773, 12.169, dan 15.319 yang berarti rerata tersebut berada dibawah rerata ideal. Jika dilihat dari nilai median yang diperoleh tampak bahwa nilai rerata kelas VII lebih kecil dari median, sedangkan nilai rerata kelas VIII dan kelas

IX lebih besar dari median. Hal ini berarti bahwa sebagian besar siswa memiliki kemampuan di bawah rerata. Tes kelas VII memiliki rentang skor antara 5 sampai dengan 29, tes kelas VIII memiliki rentang antara 3 sampai dengan 17, dan tes kelas IX memiliki rentang skor antara 7 sampai dengan 28 yang berarti bahwa varians skor cukup baik karena rentang skor cukup besar.

Analisis perkembangan kemampuan siswa pada pelajaran matematika dilakukan dengan melakukan *equating* (penyetaraan) skala kemampuan siswa kelas VII, VIII dan IX dengan menggunakan metode *mean & sigma* dengan teknik *over lapping*. Pada metode *mean & sigma* konstanta penyetaraan dihitung berdasarkan nilai *mean* dan simpangan baku dari tingkat kesulitan butir. Rangkuman hasil penghitungan disajikan pada Tabel 3.

Tabel 3. Nilai Konstanta Penyetaraan Metode *Mean & Sigma*

Penyetaraan	<i>A</i>	<i>β</i>
Kelas VII ke Kelas VIII	1,937	0,533
Kelas VII ke Kelas IX	1,109	0,561
Kelas VIII ke Kelas IX	0,795	0,162

Rerata kemampuan siswa ( $\theta$ ) hasil *equating* (penyetaraan) dan simpangan baku dari ketiga tes yang disetarakan dihitung menggunakan bantuan program *excel*. Hasil penyetaraan kemudian disubstitusikan ke rumus  $\theta_y^* = a\theta_x + \beta$ , sehingga diperoleh persamaan konversi penyetaraan kemampuan seperti rangkuman Tabel 4. Berdasarkan hasil perhitungan tersebut kemudian dapat dibandingkan

rerata kemampuan siswa pada suatu kelas dengan kelas yang lainnya.

Tabel 4. Persamaan Konversi Penyetaraan Kemampuan Metode *Mean & Sigma*

<i>Equating</i> (Penyetaraan)	Persamaan Konversi
Dari Kelas VII ke Kelas VIII	$\theta^* = 1,937 \theta + 0,533$
Dari Kelas VII ke Kelas IX	$\theta^* = 1,109 \theta + 0,561$
Dari Kelas VIII ke Kelas IX	$\theta^* = 0,795 \theta + 0,162$

Hasil perhitungan rerata kemampuan siswa dan simpangan bakunya pada semua *equating* disajikan pada Tabel 5. Perbandingan kemampuan siswa pada ketiga kelas sebelum dan sesudah *equating* dapat dilihat pada Gambar 2. Pada Gambar 2 tampak bahwa kemampuan siswa mengalami peningkatan sesuai dengan meningkatnya kelas.

Tabel 5. Rerata  $\theta$  dan Simpangan Baku Hasil Penyetaraan Metode *Mean & Sigma*

<i>Equating</i> (Penyetaraan)	Rerata Kemampuan ( $\theta$ )	Simpangan Baku
Theta kelas VII	-0.600	0.467
Konversi $\theta$ (VII ke VIII)	0.122	0.693
Konversi $\theta$ (VII ke IX)	0.439	0.740
Theta kelas VIII	-1.105	0.502
Konversi $\theta$ (VIII ke IX)	0.814	0.983
Theta kelas IX	-0.591	0.709



Gambar 2  
Grafik Rerata Kemampuan dan Simpangan Baku Hasil Penyetaraan Metode *Mean & Sigma*



Akurasi penyetaraan pada metode *mean & sigma* ditentukan dengan cara membandingkan *Root Mean Square Difference (RMSD)* dari kemampuan siswa, sebelum dan sesudah penyetaraan.

Penghitungan *RMSD* menggunakan rumus:

$$RMSD(q) = \sqrt{\frac{\sum_{i=1}^N (\hat{q}_i - q_i)^2}{N}}$$

dimana,

N: banyaknya examinee

$\hat{q}_i$ : kemampuan peserta ke i setelah disetarakan

$q_i$ : kemampuan peserta ke i sebelum disetarakan

Rangkuman hasil perhitungan *RMSD* disajikan pada Tabel 6. Pada Tabel 6 dapat dilihat bahwa nilai rerata *RMSD* penyetaraan dengan metode *mean & sigma* sebesar 0.367 yang menunjukkan hasil yang cukup akurat..

Tabel 6. Rangkuman Hasil Perhitungan *RMSD* Metode *Mean & Sigma*

Penyetaraan	RMSD
Kelas VII ke Kelas VIII	0.285
Kelas VII ke Kelas IX	0.436
Kelas VIII ke Kelas IX	0.379
Rerata RMSD	0.367

## SIMPULAN

Hasil *equating* (penyetaraan) skor siswa pada mata pelajaran matematika dari kelas VII ke kelas VIII dan kelas VIII ke kelas IX di SMP Negeri 1 Marga Tabanan memberikan informasi bahwa

kemampuan siswa mengalami peningkatan seiring dengan meningkatnya peringkat kelas. Peningkatan yang cukup tajam terjadi dari kelas VIII ke kelas IX. Hal ini mungkin disebabkan karena siswa kelas IX diberikan pelajaran tambahan menjelang ujian kelulusan.

Rerata kemampuan ( $\theta$ ) siswa pada mata pelajaran matematika berturut-turut sebesar -0.600; -1,105 dan -0.591 sekala logit. Hal ini menunjukkan bahwa sebagian besar siswa memiliki kemampuan matematika yang tergolong rendah. Rerata kemampuan siswa kelas VIII lebih rendah dari rerata kemampuan siswa kelas VII dan kelas IX.

Penelitian ini menghasilkan tiga perangkat tes prestasi belajar untuk ujian akhir semester (sumatif) kelas VII, VIII, dan IX SMP dengan karakteristik tingkat kesulitan berturut-turut  $-2.19 \leq b \leq 3.03$  (cenderung sulit);  $-1.97 \leq b \leq 3.04$  (cenderung sulit); dan  $-2.52 \leq b \leq 2.44$  (sedang), memiliki daya beda berturut-turut sebesar 4; 3; dan 3, serta memiliki nilai reliabilitas (konsistensi internal) yang bagus dengan nilai berturut-turut sebesar 0.940; 0.880; dan 0.900.

## UCAPAN TERIMA KASIH

Terima kasih penulis ucapkan kepada Bapak Kepala Dinas Pendidikan Pemuda dan Olah raga Kabupaten Tabanan atas ijin yang diberikan dalam melakukan penelitian ini, bapak kepala SMP Negeri 1 Marga dan teman-teman guru yang terlibat dalam penelitian ini yang tak dapat penulis sebutkan satu persatu kami ucapkan juga terimakasih yang tak terhingga atas kerja samanya, bantuan

dan dukungannya sehingga penelitian ini dapat berlangsung sesuai dengan rencana. Terima kasih pula kami ucapkan kepada Prof. Dr. Dra Ni Nyoman Karmini, M.Hum atas dukungannya sehingga hasil penelitian ini dapat dipublikasikan. Semoga hasil penelitian ini bermanfaat utamanya bagi dunia pendidikan.

## DAFTAR PUSTAKA

- Antara.A.A.P. (2015). *Penyetaraan Vertikal Dengan Pendekatan Klasik dan Item Response Theory pada Siswa Sekolah Dasar*. Jurnal Himpunan Evaluasi Pendidikan Indonesia (HEPI). 19 (1): 13-24. Online: <http://journal.uny.ac.id/index.php/jpep>. (Diakses pada tanggal 25 Desember 2016).
- Angoff, W.H. (1971). Scales, norms and equivalent scores. Dalam R. L. Thorndike (Ed.), *Educational measurement* (2<sup>nd</sup> ed., pp. 508-600). Washington, DC: American Council on Education.
- Baker, F.B. (2001). *The basics of item response theory*. College Park, MD: ERIC Clearinghouse on Assessment and Evaluation.
- BNSP. (2006). *Standar isi untuk satuan pendidikan dasar dan menengah*. Jakarta. Online: <https://masdwijanto.files.wordpress.com/2011/03/buku-standar-isi-smp.pdf>. (Diakses pada tanggal 26 Desember 2016).
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehard and Winston Inc.
- Depdiknas, (2003). *Undang-Undang RI Nomor 20, Tahun 2003, tentang Sistem Pendidikan Nasional*.
- Felan, G.D. (2002, February). *Test equating: Mean, linear, equipercentile, and item response theory*. Paper presented at the annual meeting of the Southwest Educational Research Association, Austin, TX.
- Gregory, R.J. (2007). *Psychological testing: history, principles, and applications* (4<sup>th</sup> ed). Boston: Pearson Education, Inc.
- Hambleton, R. K., Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer Inc.
- Hambleton, R. K., Swaminathan, H., & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc
- Kolen, M.J., & Brennan, R.L (1995). *Test equating: Methods and practices*. New York: Springer-verlag.
- Linarce, J.M. (1994). Sample size and item calibration stability. *Rasch measurement transactions*. 88, 355-383
- Naga. D.S. (1992). *Pengantar teori skor pada pengukuran pendidikan*. Jakarta: Gunadarma.
- Peterson, N.S., Kolen, M.J., & Hoover, H.D. (1989). Scaling, norming, and equating. Dalam R.L. Linn (Eds), *Educational Measurement* (pp. 221-262). New York: Macmillan Publishing Company.