

ANALISIS GENERALISABILITAS MULTI FASET PADA INSTRUMEN PENALARAN MATEMATIKA SMP

Shofan Fiangga¹, Yurizka Melia Sari²

¹ Program Studi Pendidikan Matematika, Universitas Negeri Surabaya

² Program Studi Penelitian dan Evaluasi Pendidikan, Pascasarjana Universitas Negeri Yogyakarta
shofanfiangga@unesa.ac.id

Abstrak

Penelitian ini bertujuan untuk mengetahui konsistensi internal atau reliabilitas instrumen penalaran matematika smp yang telah dikembangkan dengan menggunakan teori *2 facets generalisability* dengan model $p \times (i:r)$. Instrumen ini terdiri dari 5 indikator yang masing-masing indikator dinilai oleh rater. Metode analisis yang digunakan adalah analisis varians multi faset dengan *Generalisability study (G-study)* dan *Decision study (D-study)* dimana komponen variansnya adalah *person, rater, item*, interaksi *person* dan *rater*, dan kesalahan (*error*). *G-study* pada penelitian ini menggunakan rancangan bersarang (*nested design*) dan *D-study*-nya juga menggunakan rancangan bersarang (*nested design*). Hasil dari penelitian ini adalah didapatkan koefisien *G* untuk instrumen ini sebesar 0,67 dan hasil analisa *D-Study* menunjukkan bahwa instrumen yang dikembangkan memenuhi tingkat observasi yang dapat diterima untuk faset yang lebih luas, yaitu 0,95 tau 95% (>70%). Sehingga, Instrumen penalaran matematika yang menggunakan indikator 1, 2, 3, 4 dan 5 berdasarkan kerangka penalaran Killpatrick dengan menggunakan desain $p \times (i:r) - I$ Random, *R* Fixed ($P = 30, R = 5$ dan $I = 5$) telah memenuhi kriteria minimal yang dipersyaratkan yaitu 0,70 dan dapat diujicobakan pada skala lebih luas lagi.

Kata kunci: Generalisabilitas 2 facet, instrumen penalaran matematika, *G-Study*, *D-Study*

Abstract

The aims of this study is to determine the internal consistency or reliability coefficient of the developed mathematics reasoning instruments in Junior High School using generalizability theory multi facet, $P \times (I:R)$. The instrument consists of 5 indicators on which valued by each rater for each indicator. This research method is three-way analysis of variance of multi facet design on Generalisability study (*G-study*) and Decision study (*D-study*) by which the variance components are person, rater, item and the interaction between person and items, and the error. Both of *G-Study* and *D-Study* is using nested design. The result shows that the coefficient of *G-Study* is 0.66898. The coefficient for *D-Study* analysis suggest that the mathematics reasoning instrument has satisfied the level of observation on large facet with coefficient 0.95 or 95%. To sum up, raters have to use the indicators of reasoning instrument which consist of 1,2,3,4 and 5 components. The design of $P \times (I:R) - I$ Random, *R* Fixed ($P = 30, R = 5$ dan $I = 5$) has been satisfied the minimum criteria of conclusion which has been reach 0.70.

Keywords: multi-facet design, mathematics reasoning instrument, *G-Study*, *D-Study*

PENDAHULUAN

Konten matematika dan penalaran matematika merupakan dua hal yang tidak dapat dipisahkan. Konten matematika dipahami melalui penalaran dan penalaran dipahami dan dilatihkan melalui belajar matematika. Peningkatan kemampuan bernalar para siswa selama proses pembelajaran matematika di kelas menjadi sangat penting dan menentukan

keberhasilan mereka dan bangsa ini di masa depan. Saragih (2011, p. 109) menegaskan bahwa Kemampuan untuk menghadapi permasalahan, baik dalam permasalahan matematika maupun permasalahan dalam kehidupan nyata merupakan daya matematis (*mathematical power*), yang salah satunya adalah kemampuan penalaran.

Kemampuan penalaran sendiri sudah menjadi tujuan inti dari pembelajaran matematika di pendidikan dasar dan menengah. Di dalam Kurikulum Tingkat Satuan Pendidikan (Depdiknas, 2006) dinyatakan bahwa salah satu tujuan mata pelajaran matematika di sekolah untuk jenjang pendidikan dasar dan menengah adalah agar siswa mampu menggunakan penalaran pada pola dan sifat, melakukan manipulasi matematika dalam membuat generalisasi, menyusun bukti, atau menjelaskan gagasan dan pernyataan matematika, memecahkan masalah yang meliputi kemampuan memahami masalah (Permendiknas Nomor 22 Tahun 2006). Selain itu, pada Kurikulum 2013 juga menekankan pentingnya kemampuan penalaran dalam pembelajaran matematika seperti yang tercantum pada salah satu kemampuan inti yang harus dimiliki siswa SMP. Kemampuan yang harus dikuasai tersebut adalah mengolah, menalar, dan menyaji dalam ranah konkret dan ranah abstrak terkait dengan pengembangan dari yang dipelajarinya di sekolah secara mandiri, dan mampu menggunakan metode sesuai kaidah keilmuan (Kemendikbud, 2013). Sehingga, penalaran matematika merupakan dasar pengetahuan bagi siswa untuk mendapatkan atau menkonstruksi pengetahuan matematika.

Penalaran melibatkan alasan-alasan logis yang disajikan untuk menjelaskan atau menjustifikasi suatu solusi dari permasalahan ataupun untuk memperluas konjektur untuk sesuatu yang belum diketahui. Pada pembelajaran matematika di sekolah, penalaran diberikan melalui beberapa tahapan. mulai dari memahami pola-pola yang muncul, menentukan aturan-aturan maupun konjektur yang lebih umum, menguji konjektur dengan pola-pola baru, dan mengonstruksi bukti untuk mendukung konjektur yang dibuat (Lohman, 2007). Tahapan-tahapan tersebut dilalui di pembelajaran matematika di sekolah secara berkesinambungan sehingga penalaran dan pembuktian tidak dapat dipisahkan secara khusus (Stylianides, 2008).

Trends in International Mathematics and Science Study (TIMSS) merupakan salah satu studi internasional untuk mengevaluasi kualitas pendidikan pada peserta didik yang berusia 14 tahun. Berdasarkan benchmark Internasional 2011 capaian peserta didik Indonesia pada level rendah yang berarti menunjukkan rata-rata peserta didik Indonesia hanya mampu memahami dasar bilangan bulat dan desimal dan dapat melakukan perhitungan dasar, serta dapat mencocokkan tabel ke diagram batang dan piktograph dan membaca diagram garis sederhana (Rosnawati, 2013). Dari hasil TIMSS tersebut, perlu dikembangkan suatu instrumen yang bisa mengukur kemampuan penalaran dan pembuktian siswa tingkat dasar dan menengah.

Materi geometri dipilih pada instrumen penalaran matematika dikarenakan cakupan materi tersebut memfasilitasi siswa untuk melakukan proses penalaran dan mampu menunjukkan kinerjanya antara lain mengeksplorasi bangun-bangun geometri, menemukan sifat-sifat, menyusun konjektur kemudian mengujinya dengan strategi pembuktian. Selain itu, materi geometri juga memberikan kesempatan siswa melalui aktivitas penalaran yang mengacu pada NCTM (2000, p. 7) yaitu menganalisis masalah (*analyzing a problem*), menerapkan strategi (*implementing a strategy*) dan merefleksi solusi permasalahan (*reflecting on a solution to a problem*). Pembuktian dalam masalah geometri mempunyai kelengkapan bentuk representasi bukti dan berisi argumen berkesimpulan benar berdasarkan premis-premis yang terdapat dalam bidang geometri (Ndraha, 2013). Sehingga, masalah pembuktian yang dipilih dalam penelitian ini adalah masalah geometri. Penalaran siswa SMP dalam membuktikan permasalahan geometri perlu dideskripsikan agar kita dapat mengetahui sejauh mana siswa mengembangkan logikanya dalam meverifikasi dugaan-dugaan dalam pembuktian. Bagaimana siswa mencari beberapa strategi pembuktian geometri. Aktivitas penalaran dan jenis penalaran siswa dalam membuktikan permasalahan geometri juga menjadi fokus utama dalam pendeskripsian penalaran siswa. Karena dalam penalaran, siswa diharapkan dapat membuktikan masalah matematika yang memerlukan proses penalaran untuk mencari strategi-strategi yang tepat dan sesuai konsep-konsep matematika.

Siswa sekolah menengah pertama (SMP), ditinjau dari tahap perkembangan kognitif Piaget, berada pada tahap operasi formal (12 tahun ke atas) (Hudojo, 2001). Pada tahapan tersebut, siswa sudah mampu berpikir secara logis tanpa kehadiran benda-benda kongkret, dengan kata lain siswa sudah mampu melakukan abstraksi. Akan tetapi, fakta yang terdapat di lapangan menunjukkan bahwa siswa masih belum mampu melakukan abstraksi dengan baik dikarenakan pembelajaran konvensional yang terjadi selama ini. Padahal, pada tahap ini, siswa sudah memiliki potensi untuk berpikir kritis, kreatif, dan produktif (Prabowo & Ristiani, 2011). Sehingga, kemampuan berpikir siswa tersebut perlu dikembangkan melalui kegiatan investigasi, sehingga siswa memperoleh stimuli yang dapat mengaktifkan daya kreatif dan kritisnya untuk menyelesaikan masalah melalui instrumen penalaran matematika yang dikembangkan.

Dalam penelitian ini akan dikembangkan instrumen penilaian penalaran matematika untuk menilai dan mengukur kompetensi penalaran matematika siswa di SMP pada materi geometri. Pengembangan instrumen penilaian yang akan dikembangkan dimulai dengan tahap membangun *framework* dari penalaran matematika itu sendiri. *Framework* yang digunakan dalam penelitian ini mengacu kepada penelitian Killpatrick, Swafford dan Findell dimana

kompetensi matematika yang perlu dikembangkan dalam pengembangan kemampuan penalaran matematika di sekolah adalah *conceptual understanding* (pemahaman konsep), *procedural fluency* (kemampuan prosedural), *strategic competence* (kompetensi strategi), *adaptive reasoning* (penalaran adaptif) dan *productive disposition* (National Research Council, 2001).

Dalam pengembangan suatu instrumen pengukuran, suatu instrumen dikatakan reliabel jika dalam beberapa kali pelaksanaan pengukuran terhadap kelompok subjek yang sama diperoleh hasil yang relatif sama, selama aspek yang diukur belum berubah. Konsep validitas dan reliabilitas merupakan persyaratan yang sangat penting dan harus dipenuhi dalam pengembangan instrumen (Susilaningsih, 2011). Validitas dalam instrumen penelitian diartikan sebagai bukti teretik dan empirik dalam penafsiran skor tes sesuai dengan tujuan penggunaan tes. Mardapi (2004, p. 25) menjelaskan bahwa validitas didefinisikan sebagai ukuran seberapa cermat suatu instrumen penilaian melakukan fungsi ukurnya. Adapun sifat-sifat validitas suatu instrumen penilaian dipenuhi oleh ketercukupan, kebermaknaan dan kebergunaan dari simpulan spesifik dari skor tes yang diperoleh (Mardapi, 2012). Sehingga, dapat dikatakan bahwa konsep validitas berkaitan erat dengan relevansi dan akurasi. Relevansi merujuk kepada kemampuan instrumen dalam menerapkan fungsinya yaitu tujuan instrumen penilain tersebut dikembangkan. Sedangkan, akurasi berhubungan dengan ketepatan instrumen dalam mengidentifikasi aspek-aspek yang akan diukur secara tepat dalam menggambarkan kondisi yang sesungguhnya.

Reliabilitas atau keandalan merupakan koefisien yang menunjukkan tingkat keajegan atau konsistensi hasil pengukuran atau tes (Mardapi, 2012). Dalam hal ini, konsistensi suatu tes atau hasil pengukuran merujuk pada penggunaan alat ukur yang sama untuk subjek yang berbeda atau pada waktu berbeda tetapi kondisi pengukuran sama. Keandalan pada bidang pendidikan dan pengajaran sangat tidak mudah untuk diperoleh. Hal ini disebabkan oleh beberapa sumber kesalahan yaitu yang berasal dari alat ukur, yang diukur dan yang mengukur. Sehingga, instrumen yang memiliki reliabilitas yang tinggi apabila sumber kesalahan yang ada pada pengukuran dapat diminimalisir. Terdapat tiga jenis reliabilitas instrumen atau alat ukur berdasarkan cara dalam memperoleh data, yaitu konsistensi internal, stabilitas dan antar penilai (*inter rater*). Pada instrumen yang dikembangkan pada penelitian ini, jenis reliabilitas yang digunakan adalah reliabilitas antar penilai pada jenis tes performans pada penalaran matematika. Sedangkan, teori yang digunakan pada reliabilitas inter rater adalah teori generalisabilitas berdasarkan teori Brennan.

Teori generalisabilitas pada dasarnya terdiri dari *G-Theory* dan *D-Theory*. *G-theory* digunakan untuk mengestimasi besarnya koefisien reliabilitas antar penilai pada keadaan tertentu. Kemudian, hasil dari analisis *G-theory* selanjutnya digunakan pada analisis *D-theory*. Penggunaan metode generalisabilitas dalam penelitian di Indonesia jarang dilakukan, dalam hal ini dikarenakan keterbatasan pengetahuan terhadap penggunaan *software* GENOVA yang membantu analisis reliabilitas antar penilai. Sebenarnya, perhitungan manual dari teori generalisabilitas tidak hanya dapat dilakukan dengan GENOVA, karena program *Ms.Excel* dapat digunakan untuk membantu peneliti untuk melakukan metode generalisabilitas pada instrumen yang dikembangkan. Briesch, Chafouleas dan Johnson (2016), dalam penelitiannya menyatakan bahwa penggunaan teori generalisabilitas dalam data penilaian pada sistem sekolah telah banyak diaplikasikan pada beberapa tahun belakangan ini. Dalam penelitian ini, penulis menentukan koefisien keandalan instrumen penilaian dilakukan dengan menggunakan paket program komputer Genova berdasarkan teori generalizability yang dikembangkan oleh Crick dan Brennan pada tahun 1983 yang disebut dengan *A Generalized Analysis of Variance System*.

Briesch, Swaniathan, Welsh dan Chafouleas (2014, p. 1) menjelaskan bahwa teori generalisabilitas dapat meningkatkan kualitas instrumen penelitian dengan kemampuan teori tersebut untuk menguji beberapa sumber varians untuk pembuatan keputusan dan pengujian konsistensi hasil dari koefisien generalisabilitas. Salah satu desain pada teori generalisabilitas adalah desain *multifacets universes of admissible observation*. Jenis interaksi yang ada pada desain tersebut terdiri atas keterkaitan P (*person*), I (*item*), dan R (*rater*) dengan desain interaksi disajikan pada Tabel 1. Desain Multifaset Teori Generalisabilitas berikut:

Tabel 1. Desain Multifaset Teori Generalisabilitas

<i>No</i>	<i>Design</i>	<i>Main Effects</i>	<i>Interaction Effects</i>
1	P x I x R	P, I, R	PI, PR, IR, dan PIR
2	P x (I : R)	P, R, (I : R)	PR, P(I:R)
3	P : I : R	P, I:P, I:R:P	-
4	P:(I:R)	P, I:P, R:P	IR:P

Pada teori ini terdapat dua komponen yang akan dicari yaitu *G-study* atau *generalization study* dan *D-study* atau *decision study*. Pada *G-study* dilakukan estimasi sejumlah varians komponen. Banyaknya komponen ditentukan oleh model yang digunakan. Sedangkan, *D-study* yang akan dilakukan menggunakan hasil dari *G-study* yang dilakukan sebelumnya. Brennan (2000) menjelaskan bahwa *D-study* menekankan estimasi, penggunaan, dan interpretasi dari varians komponen untuk membuat keputusan, dengan prosedur pengukuran yang baik Spesifikasi dari generalisasi universe *D-study* merupakan salah satu hal yang penting untuk dilakukan yang terdiri atas universe berlakunya generalisasi *D-study* dengan suatu prosedur

pengukuran tertentu. Salah satu penelitian yang mengkaji studi generalisabilitas 2 faset adalah Susongko (2010) dengan menggunakan desain $P \times I \times R$. Tulisan ini akan mengkaji metode analisis varian tiga jalur sebagai metode estimasi varian-varian komponen yang terlibat dalam estimasi koefisien generalizabilitas dan mengestimasi koefisien generalisabilitas dari hasil tes penalaran matematika di SMP secara matematis dan empirik dengan desain $P \times (I:R)$.

METODE

Jenis penelitian yang digunakan pada penelitian ini adalah jenis penelitian deskriptif dengan pendekatan kuantitatif yaitu *generalizability studi*. Instrumen yang dikembangkan akan dianalisis menggunakan studi generalisabilitas dua facet menggunakan *G-study* $P \times (I:R)$ dimana komponen variansnya adalah *person*, *rater*, *item*, interaksi *person* dan *rater*, dan kesalahan. *G study*-nya dan *D-study*-nya menggunakan rancangan bersarang (*nested design*).

Desain *G-Study* yang digunakan adalah $P \times (I:R)$, yaitu item yang digunakan bersarang pada penilai (*rater*) dalam melakukan penilaian kompetensi penalaran matematika siswa. Pendapat rater terhadap item yang akan dinilai merujuk pada cara penilai untuk mengukur kompetensi penalaran matematika siswa. Efek utama dalam desain penelitian berdasarkan analisis varians efek acak adalah p , r , $r:i$ dan efek interaksinya adalah varians pi dan varians pr bersarang pada i ($pr:i$). Adapun rumus varians komponen pada dengan interaksi $p, r:i, e$ adalah sebagai berikut.

$$\sigma_{p,r,i,e}^2 = \sigma_p^2 + \sigma_{r,i,e}^2$$

Keterangan:

p = person

r = guru/rater

i = item

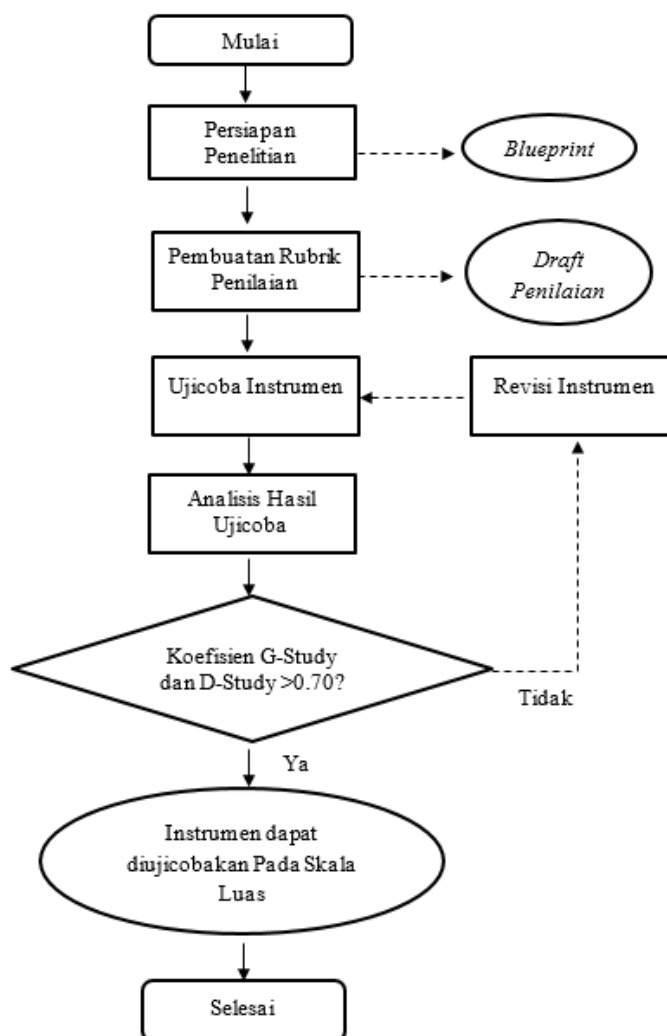
$r : i$ = rater bersarang pada item

e = kesalahan

Penentuan koefisien keandalan instrumen penilaian dilakukan dengan menggunakan paket program komputer Genova berdasarkan teori generalisabilitas yang dikembangkan oleh Crick dan Brennan pada tahun 1983 yang disebut dengan *A Generalized Analysis of Variance System*. *G-study* dilakukan dengan mengestimasi sejumlah varians komponen yang ditentukan oleh model yang digunakan. Hasil dari *G-study* digunakan pada *D-study*. Menurut Brennan (dalam Retnowati, 2009), *D-study* menekankan estimasi, penggunaan, dan interpretasi dari varians komponen untuk membuat keputusan, dengan prosedur pengukuran yang baik.

Adapun langkah-langkah atau tahapan prosedur penelitian analisis reliabilitas generalisabilitas instrumen penalaran matematika sebagai berikut: (1) pembuatan rancangan atau *blueprint* instrumen penalaran matematika untuk siswa SMP melalui pengkajian indikator,

silabus, rubrik penilaian, skoring siswa; (2) penyusunan item-item instrumen penalaran matematika berdasarkan kajian melalui rancangan yang telah dibuat; (3) penyusunan rubrik panduan penilaian bagi rater atau penilai penalaran matematika siswa; (4) validasi instrumen penalaran matematika oleh *expert judgement*; (5) ujicoba instrumen penelitian pada sampel siswa SMP di Surabaya; (6) perhitungan estimasi koefisien generalisabilitas pada data ujicoba berbantuan program komputer GENOVA; (7) penganalisisan hasil estimasi reliabilitas dari instrumen penalaran matematika; (8) pengambilan kesimpulan dari hasil analisis dan menentukan apakah instrumen dapat diujicobakan pada skala luas. Hasil kesimpulan didasarkan kepada indeks kehandalan yang dikenal dengan indeks reliabilitas. Besarnya indeks minimum 0.70, bila indeks lebih kecil dari 0.70 maka kesalahan pengukuran akan melebihi batas (Basrowi & Siskandar, 2012). Adapun rancangan terperinci dalam prosedur penyusunan instrumen beserta analisis generalisabilitas dijelaskan pada Gambar 1. Prosedur Penelitian



Gambar 1. Prosedur Penelitian

Instrumen penilaian akan diujicobakan kepada tiga puluh siswa (*P*) yang dinilai oleh lima *expert* sebagai *rater* (*R*) atau penilai terhadap kompetensi penalaran matematika siswa. Tiap *rater* menilai kompetensi yang berbeda-beda dan tiap kompetensi terdiri dari empat item/aspek (*I*) yang dinilai. Adapun indikator pada setiap kompetensi adalah *Conceptual Understanding* (Pemahaman Konsep), *Procedural Fluency* (Kemahiran Prosedural), *Strategic Competence* (Kompetensi Strategis), *Adaptive Reasoning* (Penalaran Adaptif) dan *Productive Disposition* (Sikap Produktif)

HASIL DAN PEMBAHASAN

Hasil Penelitian

Data yang diperoleh dari 30 subjek (*Person*) serta hasil penilaian dari 5 Guru (*Rater*) terhadap masing-masing dari 5 butir soal (*Item*) dianalisis menggunakan *software GENOVA*. Adapun hasil analisis *G-study* dari data ujicoba instrumen penilaian kompetensi penalaran matematika siswa dengan menggunakan program Genova adalah sebagai berikut.

EFFECT	VARIANCE COMPONENTS IN TERMS OF G STUDY UNIVERSE (OF ADMISSIBLE OBSERVATIONS) SIZES					VARIANCE COMPONENTS IN TERMS OF D STUDY UNIVERSE (OF GENERALIZATION) SIZES				
	VARIANCE COMPONENTS FOR SINGLE OBSERVATIONS	FINITE UNIVERSE COR-RECTIONS	D STUDY SAMPLING FRE-QUENCIES	VARIANCE COMPONENTS FOR MEAN SCORES		VARIANCE COMPONENTS FOR SINGLE OBSERVATIONS	FINITE UNIVERSE COR-RECTIONS	D STUDY SAMPLING FRE-QUENCIES	VARIANCE COMPONENTS FOR MEAN SCORES	
				ESTIMATES	STANDARD ERRORS				ESTIMATES	STANDARD ERRORS
P	7.20738	1.0000	1	7.20738	2.34391	7.20738	1.0000	1	7.20738	2.34391
R	0.00000E+00	1.0000	5	0.00000E+00	0.01269	0.00000E+00	1.0000	5	0.00000E+00	0.01269
I:R	0.00000E+00	1.0000	10	0.00000E+00	0.00943	0.00000E+00	1.0000	10	0.00000E+00	0.00943
PR	7.15921	1.0000	5	1.43184	0.25841	7.15921	1.0000	5	1.43184	0.25841
PI:R	10.67215	1.0000	10	1.06721	0.07220	10.67215	1.0000	10	1.06721	0.07220

QFM = QUADRATIC FORM

Gambar 2. Hasil Analisis G-Study pada Genova

Kondisi diatas berkaitan dengan *universe of admissible observation* (objek pengukuran penilaian) untuk kompetensi penalaran matematika siswa. Efek yang dominan dalam mempengaruhi variansi kesalahan pengukuran adalah komponen *person*. Sedangkan, sumber variansi *person cross item nested rater* ($P \times I : R$) merupakan komponen varian yang paling dominan. Penulis menduga bahwa siswa yang berperan sebagai *person* dan *expert* yang menjadi *rater* atau penilai baru mengenal model dan konstruk alat penilaian yang dikembangkan.

Interpretasi dari hasil *G-Study* menunjukkan tentang tingkat kebermaknaan instrumen kompetensi penalaran siswa. Koefisien G dari komponen-komponen penilaian tersebut telah memenuhi untuk kepentingan faset pengukuran yang berkaitan dengan objek pengukuran (*universe of admissible observations*) pada pengukuran kompetensi penalaran matematika siswa. Selain itu, hal ini menunjukkan bahwa secara keseluruhan pengembangan instrumen

penilaian kompetensi penalaran matematika siswa dapat diterima untuk digunakan melakukan penilaian pada faset yang lebih luas yang ditunjukkan oleh **indeks koefisien G sebesar 0,67**. (lihat Gambar 3. Indeks Koefisien *G-Study* pada Genova)

	VARIANCE	STANDARD DEVIATION	STANDARD ERROR OF VARIANCE	
UNIVERSE SCORE	7.20738	2.68466	2.34391	
EXPECTED OBSERVED SCORE	10.77365	3.28232	2.33242	
LOWER CASE DELTA	3.56627	1.88846	0.27785	GENERALIZABILITY COEFFICIENT = 0.66898 (2.02098)
UPPER CASE DELTA	3.56627	1.88846	0.26906	PHI = 0.66898 (2.02098)
MEAN	0.35912	0.59927		

NOTE: SIGNAL/NOISE RATIOS ARE IN PARENTHESES

GENOVA VERSION 3.1 PAGE 11

Gambar 3. Indeks Koefisien *G-Study* pada Genova

Langkah selanjutnya adalah menggunakan hasil *G-study* untuk menganalisis *D-study*. Tujuan analisis *D-study* adalah untuk menjawab pertanyaan rancangan *D-study* yang mana harus dipilih dan seberapa banyak butir komponen penilaian harus dicakup sebagai sarana mengukur dan menilai penalaran matematika siswa SMP sehingga dapat menunjukkan kebermaknaan untuk faset yang lebih luas. Hasil analisis *D-study* mengungkapkan bahwa pada penilaian proses untuk mencapai kesepahaman dan kesepakatan yang memenuhi tingkat observasi yang dapat diterima untuk faset yang lebih luas, yaitu 0,70.

Data dibawah ini (Gambar 4) merupakan hasil analisis *D-Study* untuk menentukan rancangan *D-study* yang mana harus dipilih dan seberapa banyak butir komponen penilaian harus dicakup sebagai sarana mengukur dan menilai kompetensi penalaran siswa sehingga dapat menunjukkan kebermaknaan untuk faset yang lebih luas.

SUMMARY OF D STUDY RESULTS FOR SET OF CONTROL CARDS NO. 003												
		SAMPLE SIZES			VARIANCES							
D STUDY DESIGN NO	INDEX= UNIV.- INF.	SP INF.	R 5	I INF.	UNIVERSE SCORE	EXPECTED OBSERVED SCORE	LOWER CASE DELTA	UPPER CASE DELTA	MEAN	GEN. COEF.	PHI	
003-001	30	5	1		8.63922	10.77365	2.13443	2.13443	0.35912	0.80188	0.80188	
003-002	30	5	2		8.63922	9.70643	1.06721	1.06721	0.32355	0.89005	0.89005	
003-003	30	5	3		8.63922	9.35070	0.71148	0.71148	0.31169	0.92391	0.92391	
003-004	30	5	4		8.63922	9.17283	0.53361	0.53361	0.30576	0.94183	0.94183	
003-005	30	5	5		8.63922	9.06611	0.42689	0.42689	0.30220	0.95291	0.95291	

GENOVA VERSION 3.1 PAGE 45

Gambar 4. Hasil Analisis *D-Study* pada Genova

Dengan mencermati setiap tahap rancangan *D-study* pada komposisi besar sampel tertentu maka akan dapat diperoleh informasi koefisien G dan juga diperoleh informasi berapa kenaikan indeks kebermaknaan pada koefisien G setelah satu butir komponen penilaian dilibatkan untuk mengukur atau menilai. Untuk menjawab pertanyaan ini dan tujuan tersirat

didalamnya analisis pada setiap hasil *D-study* dapat digunakan. Uraian berikut memaparkan hasil-hasil analisis *D-study* ini.

Pembahasan

Pada komponen penilaian kompetensi guru matematika $P \times (I:R)$ Design – I *Random*, R *Fixed* (disain *D-study* nomor 003-001 dengan $P = 30$, $R = 5$ dan $I = 1$) maka tingkat atau koefisien kesepahaman dan kesepakatan (reliabilitas dalam koefisien G) sebesar 0,80188, Artinya penilai memiliki tingkat kesepahaman dan kesepakatan terhadap penggunaan konstruk instrumen penilaian yang dipakai sebesar 80%. Jika penilai menggunakan dua indikator (desain *D-study* nomor 003-002, dengan $P = 30$, $R = 5$ dan $I = 2$) yakni indikator/kompetensi 1 dan 2, maka tingkat atau koefisien kesepahaman dan kesepakatan sebesar 0,89005; demikian seterusnya untuk rancangan 003-003 diperoleh koefisien sebesar 0,92391. Berdasarkan kenyataan ini maka dapat dikatakan bahwa untuk mencapai kesepahaman dan kesepakatan yang memenuhi tingkat observasi yang dapat diterima untuk faset yang lebih luas, yaitu $>0,70$, penilai harus menggunakan indikator 1, 2, 3, 4 dan 5 karena dalam konteks ini jika 5 indikator digunakan maka akan dicapai koefisien kesepahaman dan kesepakatan sebesar 0,95291 atau 95% ($>70\%$).

Hasil yang telah didapatkan mendukung penelitian studi generalisabilitas yang dilakukan oleh Susongko (2010) dalam menganalisis generalisabilitas pada studi empirik tes prestasi belajar kimia bentuk uraian atau testlet yang menunjukkan bahwa analisis varian tiga jalur model $P \times I \times R$ sangat baik bila digunakan untuk mengestimasi besarnya komponen-komponen varian yang terlibat dalam studi generalizabilitas sehingga dapat digunakan pada faset yang lebih luas meskipun dengan desain yang berbeda. Selain itu, penelitian pada makalah ini juga mendukung analisis generalisabilitas yang dilakukan oleh Retnowati (2009) dalam pengembangan instrumen penilaian karya seni lukis anak di sekolah dasar menggunakan desain $P \times (I:R)$ dan didapatkan hasil koefisien Genova secara keseluruhan adalah 0,71 dimana nilai rata-rata ini telah memenuhi kriteria minimum yang disyaratkan yaitu 0,70 untuk dilakukan ujicoba pada skala luas.

SIMPULAN DAN SARAN

Berdasarkan hasil penelitian dan pembahasan dapat disusun kesimpulan dilihat dari pengembangan instrument dan faktor reliabilitas instrumen. Instrumen yang dikembangkan mengacu pada framework pengembangan kemampuan penalaran matematika di sekolah adalah *conceptual understanding* (pemahaman konsep), *procedural fluency* (kemampuan prosedural),

strategic competence (kompetensi strategi), *adaptive reasoning* (penalaran adaptif) dan *productive disposition*. Pada instrumen pada penelitian ini, teori generalisabilitas berdasarkan teori Brennan digunakan untuk mengukur reliabilitas antar penilai (*inter-rater reliability*) untuk memastikan reliabilitas antar penilai pada jenis tes performans pada penalaran matematika.

Karakteristik instrumen penilaian instrumen penalaran matematika yang mencakup *generalizability theory (G-theory)* dan *Decision theory (D-theory)* ini dapat diketahui. Koefisien G untuk instrumen ini sebesar 0,66898 dan koefisien D-Study sebesar 0,95. Lebih jauh lagi analisis *D-Study* menghasilkan kesimpulan bahwa untuk mencapai kesepahaman dan kesepakatan yang memenuhi tingkat observasi yang dapat diterima untuk faset yang lebih luas, yaitu $>0,70$, penilai harus menggunakan indikator 1, 2, 3, 4 dan 5. Desain $P \times (I:R) - I$ Random, R Fixed ($P = 30$, $R = 5$ dan $I = 5$) telah memenuhi kriteria minimal yang dipersyaratkan yaitu 0,95.

DAFTAR PUSTAKA

- Basrowi, & Siskandar. (2012). *Evaluasi Belajar Berbasis Kinerja*. Bandung: Karya Putra Darwati.
- Brennan, R. L. (2000, December 1). Performance Assessments from the perspective of Generalizability Theory. *Applied Psychological Measurement*, 24(4), 339-353. doi:<https://doi.org/10.1177/01466210022031796>
- Briesch, A., Chafouleas, S. M., & Johnson, A. (2016). Use of Generalizability Theory Within K–12 School-Based Assessment: A Critical Review and Analysis of the Empirical Literature. *Journal Applied Measurement in Education*, 29(2), 83-107. doi:<http://dx.doi.org/10.1080/08957347.2016.1138955>
- Briesch, A., Swaminathan, H., Welsh, M., & Chafouleas, S. M. (2014). Generalizability theory: A practical guide to study design, implementation, and interpretation. *Journal of School Psychology*, 52(1), 13-35. doi:10.1016/j.jsp.2013.11.008.
- Hudojo, H. (2001). *Pengembangan Kurikulum dan Pembelajaran Matematika*. Malang: Universitas Negeri Malang.
- Kemendikbud. (2013). Lampiran Permendikbud Nomor 65 Tahun 2013 Tentang Kerangka Dasar dan Struktur Kurikulum SMP/MTs. Jakarta: Kementerian Pendidikan dan Kebudayaan RI.
- Lohman, D. (2007). *Reasoning and Intelligence*. New York: Cambridge University Press.
- Mardapi, D. (2004). *Penyusunan Tes Hasil Belajar*. Yogyakarta: Program Pascasarjana Universitas Negeri Yogyakarta.
- Mardapi, D. (2012). *Pengukuran, Penilaian dan Evaluasi Pendidikan*. Yogyakarta: Nuha Litera.
- National Research Council. (2001). *Adding it up: Helping children learn* (Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education ed.). (J. Kilpatrick, J. Swafford, & B. Findell, Eds.) Mathematics Learning Study Committee, Center for Education, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy

- Press. Retrieved from https://www.pearsoned.com/wp-content/uploads/m14-res_practice.pdf
- NCTM. (2000). *Principles and Standards for School Mathematics*. USA: NCTM.
- Ndraha, F. (2013). "Nilai Strategis memandang Bukti Geometri sebagai Prosep dalam Pembelajaran. *Seminar Nasional Matematika dan Pendidikan Matematika UNY*. Yogyakarta: UNY.
- Permendiknas Nomor 22 Tahun 2006. (2006, Mei 23). *Standar Isi untuk Satuan Pendidikan Dasar dan Menengah*. Jakarta.
- Prabowo, A., & Ristiani, E. (2011). Rancang Bangun Instrumen Tes Kemampuan Keruangan Pengembangan Tes Kemampuan Keruangan Hubert Maier dan Identifikasi Penskoran Berdasar Teori Van Hiele. *Jurnal Kreano ISSN: 2086-2334*, 2(2).
- Retnowati, T. H. (2009). Pengembangan Instrumen Penilaian Proses dan Produk Karya Seni Lukis. *Jurnal Penelitian dan Evaluasi Pendidikan*, 12(2).
- Rosnawati, R. (2013). Kemampuan Penalaran Matematika Siswa SMP Indonesia Pada TIMSS 2011. *Prosiding Seminar Nasional Penelitian, Pendidikan dan Penerapan MIPA 18 Mei 2013*. Yogyakarta: Universitas Negeri Yogyakarta.
- Saragih, S. (2011, November). Pengaruh Pendekatan Pembelajaran dan Locus of Control terhadap Kemampuan Penalaran Matematika Siswa. *Jurnal Kependidikan*, 41(2), 108-119.
- Stylianides, G. (2008). An Analytic Framework of Reasoning-and-Proving. *For the Learning of Mathematics*, 28(1), 9-16.
- Susilaningsih, E. (2011). Instrumen Penilaian Praktikum Kimia dan Estimasi Reliabilitasnya Dengan Koefisien Generalisabilitas. *Prosiding Seminar Nasional Kimia dan Pendidikan Kimia VI* (pp. 87-97). Solo: Program Studi Kimia, FKIP UNS.
- Susongko, P. (2010). *Studi Generalizabilitas Tes Tipe Dua Facet dengan Menggunakan Analisis Varian Tiga Jalur*. Oseatek.