

THE COMPARISON OF NUMBER RESPONSE CATEGORIES TOWARDS RELIABILITY OF MATHEMATICS DISPOSITION INSTRUMENT

Siti Muslihah Hadi

Jl. Semabung Baru Gg. Imam 3 Rt. 03 Rw. 03, Pangkal Pinang, Bangka
sitimuslihah.hadi@gmail.com

Abstract

This study aimed to look at the comparison of number response categories towards reliability of mathematics disposition instrument. The research method used in this study is an experimental method. While the design used is one way anava of 4 x 1 and research hypotheses were tested using one way analysis of variance (ANOVA). The participants of this study were 765 students from eleventh grade in 4 SMAN in Central Jakarta with sampling technique multi stage random sampling. The result concluded that: (1) there is an influence of number response categories towards reliability coefficient of mathematics disposition instruments, (2) reliability of seven categories response mathematics disposition instrument higher than reliability of five categories response mathematics disposition instrument, (3) reliability of seven categories response mathematics disposition instrument higher than reliability of four categories response mathematics disposition instrument , and (4) reliability of six categories response mathematics disposition instrument higher than reliability of four categories response mathematics disposition instrument. Implementation of the study about adding two or three categories on non test instrument can be implemented for mathematics subject for student of Senior High School because of reliability enhancement.

Keywords: *response categories, reliability, mathematics disposition instrument*

PERBANDINGAN BANYAK KATEGORI RESPON TERHADAP RELIABILITAS INSTRUMEN DISPOSISI MATEMATIKA

Siti Muslihah Hadi

Jl. Semabung baru Gg. Imam 3 Rt. 03 Rw. 03, Pangkal Pinang, Bangka
sitimuslihah.hadi@gmail.com

Abstrak

Penelitian ini bertujuan untuk melihat perbandingan banyak kategori respon terhadap reliabilitas instrumen disposisi matematika. Metode yang digunakan adalah metode eksperimen dengan desain *one way anava* 4×1 . Populasi adalah seluruh siswa SMAN di Jakarta Pusat tahun ajaran 2012-2013. Pengambilan sampel dilakukan dengan cara *multi stage random sampling* sehingga diperoleh 765 sampel siswa dari 4 SMA Negeri di Jakarta Pusat. Hipotesis penelitian diuji dengan menggunakan analisis varian (ANOVA) satu jalan. Hasil penelitian menyimpulkan bahwa: (1) terdapat pengaruh banyak kategori respon terhadap reliabilitas instrumen disposisi matematika (2) reliabilitas instrumen disposisi matematika dengan tujuh kategori respon lebih tinggi dari reliabilitas instrumen disposisi matematika dengan lima kategori respon, (3) reliabilitas instrumen disposisi matematika dengan tujuh kategori respon lebih tinggi dari reliabilitas instrumen disposisi matematika dengan empat kategori respon, (4) reliabilitas instrumen disposisi matematika dengan enam kategori respon lebih tinggi dari reliabilitas instrumen disposisi matematika dengan empat kategori respon. Hasil penelitian ini dapat menjadi masukan bagi guru matematika di Sekolah Menengah Atas dan bagi guru mata pelajaran lain pada umumnya untuk mempertimbangkan penambahan dua hingga tiga kategori respon pada instrumen non tes karena lebih efektif untuk meningkatkan koefisien reliabilitas.

Kata kunci: kategori respon, reliabilitas, disposisi matematika

PENDAHULUAN

Secara umum selain aspek kognitif, aspek afektif siswa atau sikap siswa merupakan salah satu hal yang harus diperhatikan dalam proses pembelajaran. Pendapat Krathwol seperti dikutip oleh Mardapi (1961: 3), Gibson dan Robbins seperti dikutip oleh Leonard dan Supardi (2010: 342), Popham (1995: 179), menggambarkan bahwa aspek sikap akan mempengaruhi seseorang dalam mengupayakan dirinya untuk melakukan sesuatu hal. Sebagai contoh jika siswa mempunyai rasa suka terhadap pelajaran matematika maka siswa akan mengupayakan dirinya untuk dapat memahami matematika dengan cara-cara yang siswa senangi pula.

Berbicara lebih khusus mengenai pembelajaran matematika, aspek afektif juga perlu diperhatikan oleh guru. Sebagaimana diketahui bahwa matematika yang objeknya berkaitan dengan angka-angka dan rumus-rumus secara tidak langsung membuat guru terfokus pada aspek kognitif sehingga seringkali aspek

afektif dalam pembelajaran matematika kurang diperhatikan oleh guru. Padahal sikap siswa terhadap pelajaran matematika merupakan salah satu faktor yang dapat menunjang keberhasilan siswa dalam belajar matematika. Hal ini seperti diperkuat oleh pendapat Begle seperti dikutip dalam Darhim (2004: 1) dan McLeod (1992: 575) bahwa sikap positif siswa terhadap matematika memberikan pengaruh positif dalam diri siswa untuk mencapai keberhasilan dalam belajar matematika. Dengan demikian jelaslah bahwa dalam pembelajaran matematika, siswa tidak hanya diupayakan untuk menguasai kompetensi dalam aspek kognitif, tetapi juga aspek afektif atau sikap.

Salah satu aspek sikap yang perlu guru ketahui dalam ruang lingkup pelajaran matematika di sekolah yaitu disposisi matematika. Disposisi matematika merupakan salah satu tujuan pembelajaran matematika di sekolah. Dalam hal ini peserta didik diharapkan memiliki rasa ingin tahu, perhatian, dan minat dalam pelajaran matematika, serta sikap ulet dan percaya diri dalam pemecahan masalah. Kirkpatrick mengistilahkan sikap tersebut sebagai disposisi produktif yang merupakan salah satu komponen dalam kecakapan matematis (Widjajanti, 2011: 3).

Untuk mendapatkan informasi mengenai disposisi matematika peserta didik diperlukan instrumen yang dapat mengukur sikap. Instrumen yang digunakan yaitu berupa instrumen non tes yang menggunakan skala pengukuran dengan model yang diadaptasi dan dikembangkan dari skala Likert. Instrumen terbagi menjadi instrumen baku dan instrumen buatan sendiri. Instrumen baku atau standar telah dianggap baku, valid dan reliabel. Sementara untuk instrumen yang dibuat sendiri, perlu diujicobakan untuk mengetahui butir-butir instrumen yang valid dan mendapatkan instrumen yang reliabel. Untuk keperluan proses pembelajaran matematika di kelas, guru matematika hendaknya mampu membuat instrumen pengukuran sikap yang reliabel. Instrumen yang valid dan reliabel diperlukan untuk meyakinkan bahwa hasil penilaian atau evaluasi yang dilaksanakan informasinya benar-benar akurat.

Instrumen pengukuran sikap yang reliabel berarti bahwa instrumen tersebut memiliki reliabilitas tinggi, artinya bahwa instrumen tersebut memiliki derajat konsistensi atau stabilitas tinggi. Instrumen tersebut dapat mengukur apa yang hendak diukur secara akurat dan konsisten dengan hasil yang relatif sama walaupun digunakan berulang kali pada subjek yang sama. Berdasarkan pendapat Christensen (1988: 117), Popham (1995: 117), dan Nitko (1996: 63) bahwa reliabilitas mengacu pada ketetapan, kekonsistenan, atau keajegan alat ukur. Lebih lanjut menurut Popham, reliabilitas sama pengertiannya dengan konsistensi, dimana konsistensi dalam penilaian pendidikan terbagi menjadi tiga macam yaitu: (1) stabilitas; (2) format alternatif; dan (3) konsistensi internal. Dalam penelitian ini reliabilitas yang dimaksud adalah koefisien reliabilitas yang diestimasi dengan formula konsistensi internal *Alpha Cronbach* (α).

Untuk membuat instrumen pengukuran domain afektif atau sikap yang memiliki reliabilitas tinggi tidaklah mudah. Reliabilitas perangkat butir-butir

dalam pengembangan instrumen domain afektif dipengaruhi oleh beberapa faktor yaitu karakteristik sampel, homogenitas isi butir, jumlah butir, dan format respon. Format respon berkaitan dengan masalah jumlah kategori respon, label kategori respon, kutub positif dan negatif skala. Beberapa penelitian telah dilakukan mengenai format respon pada skala pengukuran sikap. Sebagai contoh yaitu penelitian yang telah dilakukan oleh Preston (2000: 1-15), Weng (2004: 956-972), Weijters dkk. (2010: 1-60), dan Kim (2010: 913-927). Penelitian-penelitian tersebut membahas pengaruh format respon yang meliputi label kategori dan banyak kategori respon terhadap hasil respon dan reliabilitas. Secara umum temuan penelitian tersebut bahwa label kategori yang jelas berupa pada semua kategori dapat mempengaruhi hasil respon. Demikian pula dengan peningkatan banyak kategori respon misalnya 3, 4, 5, hingga 11 dapat mempengaruhi peningkatan koefisien reliabilitas, namun hanya pada banyak kategori tertentu. Kategori yang terlalu banyak belum tentu akan menghasilkan koefisien reliabilitas yang tinggi pula.

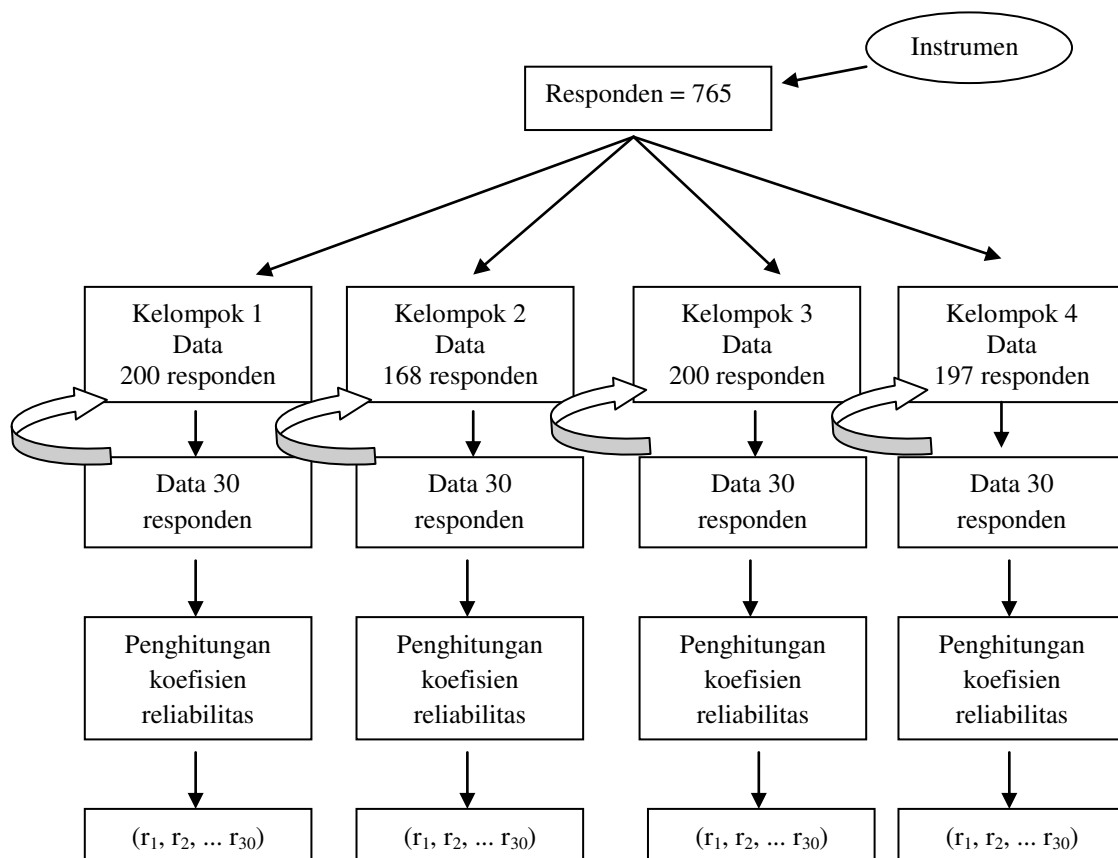
Selain itu dalam penelitian tersebut ada juga yang menyinggung tentang kategori titik tengah turut berpengaruh atau tidak. Skala dengan banyak kategori respon ganjil misalnya 5 atau 7 menyediakan titik tengah, sebaliknya skala dengan banyak kategori respon genap misalnya 4 atau 6 tanpa titik tengah. Ada yang menganggap bahwa titik tengah hanya akan memberikan kesempatan pada responden untuk tidak cenderung pada salah satu kutub skala. Sementara itu ada pula praktisi yang tetap menyediakan titik tengah dengan alasan menghindari kesalahan sistematis yang membuat tidak tergambaranya perasaan responden secara akurat karena responden seperti dipaksa memilih salah satu kutub skala. Dengan demikian pilihan jumlah kategori respon genap atau ganjil dapat mempengaruhi data respon yang diperoleh.

Berdasarkan penelitian-penelitian yang telah dilakukan, terlihat bahwa banyak kategori respon mungkin cukup berpengaruh terhadap reliabilitas instrumen pengukuran sikap. Namun semakin banyak kategori tidak pula menjamin semakin tinggi koefisien reliabilitas. Pilihan banyak kategori genap atau ganjil juga dapat berpengaruh pada hasil respon. Oleh karena itu, masih perlu adanya penelitian-penelitian lanjutan tentang banyak kategori respon yang kaitannya dengan koefisien reliabilitas instrumen. Selain itu dari empat tokoh yang disebutkan di atas, hanya satu di antaranya yang melakukan penelitian di bidang pendidikan. Dalam rangka memperkaya penelitian tentang teori pengukuran sikap di bidang pendidikan khususnya bidang pendidikan matematika maka penelitian ini mengkaji tentang perbandingan banyak kategori respon terhadap koefisien reliabilitas instrumen disposisi matematika. Adapun tujuan penelitian ini adalah membandingkan koefisien reliabilitas antara instrumen disposisi matematika dengan tujuh kategori respon, enam kategori respon, lima kategori respon, dan empat kategori respon. Sehingga nantinya dari hasil perbandingan tersebut, guru matematika dapat mengembangkan instrumen pengukuran sikap dengan banyak kategori respon yang beragam

hingga tujuh kategori, tidak hanya terpaku pada skala dengan lima kategori respon.

METODE PENELITIAN

Pendekatan dalam penelitian ini menggunakan pendekatan kuantitatif dengan menggunakan desain *one way anava* 4×1 . Populasi dalam penelitian ini adalah siswa kelas XI SMA Negeri di Jakarta Pusat. Pengambilan sampel siswa yang mengisi instrumen dilakukan dengan cara *multi stage random sampling*. Dari skor kelompok siswa yang telah mengisi instrumen, diambil 30 data siswa secara acak untuk dihitung koefisien reliabilitasnya misalnya r_1 . Setelah itu data 30 responden dikembalikan lagi ke kelompok masing-masing, kemudian diambil lagi secara acak 30 responden dan dihitung kembali koefisien reliabilitasnya, misalkan nilai nya r_2 . Demikian seterusnya replikasi dilakukan sebanyak 30 kali sehingga diperoleh data koefisien reliabilitas sebanyak 30 untuk setiap kelompok. Jumlah total koefisien reliabilitas yang akan dianalisis adalah sebanyak 120 reliabilitas (dengan 30 reliabilitas pada tiap selnya). Dengan demikian yang menjadi unit analisis pada penelitian ini adalah koefisien reliabilitas instrumen.



Gambar 1. Diagram Alur Prosedur Penelitian

Instrumen dalam penelitian ini menggunakan instrumen disposisi matematika. Pada instrumen tersebut dilakukan validasi konstruk oleh 20 orang panelis dengan menggunakan rumus *Lawshe* dengan 60 butir pernyataan yang divalidasi ada 5 butir yang tidak cocok sehingga direvisi. Sedangkan validasi empiris setiap set instrumen yang berbeda kategori respon diujicobakan pada 300 responden menggunakan rumus korelasi Product Moment perhitungan reliabilitasnya menggunakan rumus alpha Cronbach didapatkan koefisien 0,77; 0,75; 0,73; 0,73 yang berarti dinyatakan reliabel.

HASIL PENELITIAN

Hasil penelitian ini menunjukkan bahwa nilai rata-rata koefisien reliabilitas instrumen disposisi matematika dengan tujuh kategori respon adalah 0,887 dengan standar deviasi 0,037. Nilai rata-rata koefisien reliabilitas instrumen disposisi matematika dengan enam kategori respon adalah 0,876 dengan standar deviasi 0,036. Nilai rata-rata koefisien reliabilitas instrumen disposisi matematika dengan lima kategori respon adalah 0,854 dengan standar deviasi 0,045. Nilai rata-rata koefisien reliabilitas instrumen disposisi matematika dengan empat kategori respon adalah 0,844 dengan standar deviasi 0,055. Adapun hasil statistik deskriptif empat kelompok data koefisien reliabilitas dapat dilihat pada tabel 1.

Tabel 1. Rekapitulasi Statistik Deskriptif Koefisien Reliabilitas

Statistik	Kategori Respon			
	7 (X ₁)	6 (X ₂)	5 (X ₃)	4 (X ₄)
N	30	30	30	30
Rata-rata	0,887	0,876	0,854	0,844
SD	0,037	0,036	0,045	0,055

Dari pengujian persyaratan analisis dengan uji normalitas menggunakan uji *Lilliefors* dan uji homogenitas menggunakan uji *Bartlett* didapatkan bahwa data penelitian berasal dari populasi yang berdistribusi normal dan mempunyai variansi yang sama (homogen). Teknik analisis data untuk menguji hipotesis menggunakan Analisis Varian (ANOVA) satu jalan. Sedangkan untuk mengetahui signifikansi perbedaan dari masing-masing kelompok dilakukan uji lanjut menggunakan Uji Tukey.

Berdasarkan pengujian hipotesis diperoleh bahwa terdapat perbedaan rata-rata koefisien reliabilitas antara instrumen dengan tujuh kategori respon, enam kategori respon, lima kategori respon, dan empat kategori respon.

Tabel 2. Hasil Perhitungan ANAVA Satu Jalan

Sumber Varians	JK	Db	RJK	F _{hitung}	F _{tabel} ($\alpha=0,05$)	Keterangan
Antar	0,035	3	0,012	5,940	2,687	Signifikan
Dalam	0,225	116	0,002			
Total	0,260	119				

Berdasarkan hasil analisis varian (ANAVA) pada tabel 2 diperoleh $F_{hitung} = 5,940 > F_{tabel}$ pada $\alpha_{0,05} = 2,687$ yang berarti terdapat pengaruh banyak kategori respon terhadap koefisien reliabilitas instrumen disposisi matematika. Oleh karena itu dapat dilakukan pengujian hipotesis lanjutan menggunakan uji Tukey. Adapun ringkasan hasil uji Tukey dapat dilihat pada tabel berikut:

Tabel 3. Hasil Pengujian Uji Tukey

Kelompok	Q _{hitung}	Q _{tabel}	Kesimpulan
		$\alpha=0,05$	
X ₁ dan X ₂	1,311	3,850	Terima H ₀
X ₁ dan X ₃	4,027	3,850	Tolak H ₀
X ₁ dan X ₄	5,316	3,850	Tolak H ₀
X ₂ dan X ₃	2,716	3,850	Terima H ₀
X ₂ dan X ₄	4,005	3,850	Tolak H ₀
X ₃ dan X ₄	1,289	3,850	Terima H ₀

Berdasarkan tabel 4 dapat dilihat bahwa untuk kelompok X₁ dan X₃ nilai $Q_{hitung} > Q_{tabel}$. Ini berarti rata-rata koefisien reliabilitas instrumen disposisi matematika dengan tujuh kategori respon lebih tinggi dari koefisien reliabilitas instrumen disposisi matematika dengan lima kategori respon. Untuk kelompok X₁ dan X₄ nilai $Q_{hitung} > Q_{tabel}$. Ini berarti rata-rata koefisien reliabilitas instrumen disposisi matematika dengan tujuh kategori respon lebih tinggi dari koefisien reliabilitas instrumen disposisi matematika dengan empat kategori respon. Untuk kelompok X₂ dan X₄ nilai $Q_{hitung} > Q_{tabel}$. Ini berarti rata-rata koefisien reliabilitas instrumen disposisi matematika dengan enam kategori respon lebih tinggi dari koefisien reliabilitas instrumen disposisi matematika dengan empat kategori respon. Sementara itu untuk kelompok X₁ dan X₂ nilai $Q_{hitung} < Q_{tabel}$, ini berarti tidak terdapat perbedaan rata-rata koefisien reliabilitas antara instrumen disposisi matematika dengan tujuh kategori respon dan instrumen disposisi matematika dengan enam kategori respon. Untuk kelompok X₂ dan X₃ nilai $Q_{hitung} < Q_{tabel}$, ini berarti tidak terdapat perbedaan rata-rata koefisien reliabilitas antara instrumen disposisi matematika dengan enam kategori respon dan

instrumen disposisi matematika dengan lima kategori respon. Untuk kelompok X_3 dan X_4 nilai $Q_{hitung} < Q_{tabel}$, ini berarti tidak terdapat perbedaan rata-rata koefisien reliabilitas antara instrumen disposisi matematika dengan lima kategori respon dan instrumen disposisi matematika dengan empat kategori respon.

PEMBAHASAN

Hasil penelitian menunjukkan bahwa secara keseluruhan terdapat rata-rata koefisien reliabilitas antara instrumen disposisi matematika dengan tujuh kategori respon, enam kategori respon, lima kategori respon, dan empat kategori respon. Tujuh kategori respon yang disediakan pada instrumen yaitu Sangat Setuju (SS), Setuju (S), Agak Setuju (AS), Netral (N), Kurang Setuju (KS), Tidak Setuju (TS), dan Sangat Tidak Setuju (STS). Enam kategori respon yang disediakan pada instrumen yaitu Sangat Setuju (SS), Setuju (S), Agak Setuju (AS), Kurang Setuju (KS), Tidak Setuju (TS), dan Sangat Tidak Setuju (STS). Lima kategori respon yang disediakan pada instrumen yaitu Sangat Setuju (SS), Setuju (S), Netral (N), Tidak Setuju (TS), dan Sangat Tidak Setuju (STS). Empat kategori respon yang disediakan pada instrumen yaitu Sangat Setuju (SS), Setuju (S), Tidak Setuju (TS), dan Sangat Tidak Setuju (STS). Makin beragam kategori respon yang disediakan maka makin leluasa pula siswa menggambarkan respon sesuai dengan kecenderungan sikap siswa. Dapat pula pilihan jawaban atau kategori respon yang berbeda menyebabkan siswa memberikan respon yang berbeda pula, sehingga respon yang dihasilkan akan makin beragam yang pada akhirnya dapat mempengaruhi koefisien reliabilitas. Hal ini sesuai dengan pendapat Aiken (2002: 43) bahwa skor responden terhadap suatu instrumen dengan jumlah kategori tertentu berhubungan dengan variansi. Variansi berhubungan dengan reliabilitas instrumen. Skor pada instrumen yang memiliki jumlah kategori respon yang lebih besar cenderung memiliki variansi yang lebih besar pula dibandingkan dengan skor pada instrumen yang memiliki jumlah kategori respon yang lebih kecil. Variansi yang lebih besar akan menghasilkan reliabilitas yang lebih tinggi. Dengan demikian, antara instrumen disposisi matematika dengan tujuh, enam, lima, dan empat kategori respon menghasilkan keragaman respon yang berbeda pula. Adanya perbedaan tersebut kemudian menyebabkan perbedaan rata-rata koefisien reliabilitas instrumen disposisi matematika dengan kategori respon yang berbeda.

Berdasarkan data respon siswa terhadap instrumen dengan tujuh kategori yang menyediakan pilihan netral dan instrumen dengan enam kategori respon yang tanpa pilihan netral, tidak mempengaruhi siswa dalam memberikan respon sesuai dengan kecenderungan sikap siswa. Dengan adanya pilihan netral tidak berarti siswa lebih cenderung untuk memilih kategori netral sebagai gambaran kecenderungan sikap siswa terhadap butir-butir pada instrumen yang diberikan. Namun demikian, kategori netral masih tetap dipilih oleh sejumlah siswa untuk butir-butir tertentu. Meskipun menyebabkan variasi skor siswa

tetapi perbedaannya tidak signifikan. Hal ini diperkuat oleh kesimpulan Widhiarso (2010: 2) bahwa skor skala yang menyediakan kategori titik tengah dengan yang tidak memiliki kategori titik tengah, tidak memiliki perbedaan yang berarti. Reliabilitas pengukuran dan validasi butir tidak mengalami perbedaan, tetapi yang berbeda adalah varian skor. Beberapa penelitian yaitu oleh Mattel & Jacoby seperti dikutip oleh Chang (1994: 205-206) dan Dawes (2001: 1-17) memberikan hasil yang serupa bahwa titik tengah tidak sungguh-sungguh dapat mempengaruhi validitas dan reliabilitas. Dengan demikian tidak terdapat perbedaan rata-rata koefisien reliabilitas antara instrumen dengan tujuh kategori respon dan instrumen dengan enam kategori respon.

Instrumen dengan tujuh kategori respon yang memiliki dua pilihan jawaban lebih banyak dari instrumen dengan lima kategori respon menyebabkan respon siswa terhadap instrumen dengan tujuh kategori lebih bervariasi. Sehingga variansi skor responden juga akan meningkat. Meningkatnya variansi berhubungan dengan meningkatnya koefisien reliabilitas. Hal ini serupa dengan pendapat yang dikemukakan oleh Guilford seperti dikutip oleh Kirkendall (1980: 382) bahwa reliabilitas meningkat sebagaimana banyak respon pilihan meningkat, dengan reliabilitas paling besar ditemukan pada tujuh kategori respon. Pendapat serupa juga diungkapkan oleh Dawes (2001: 1-17) dalam artikelnya bahwa skala dengan tujuh kategori menghasilkan data dengan variansi yang lebih tinggi dari skala dengan lima kategori. Dengan demikian rata-rata koefisien reliabilitas instrumen disposisi matematika dengan tujuh kategori respon lebih tinggi dari rata-rata koefisien reliabilitas instrumen disposisi matematika dengan lima kategori respon.

Instrumen dengan tujuh kategori respon yang memiliki tiga pilihan jawaban lebih banyak dari instrumen dengan empat kategori respon menyebabkan respon siswa terhadap instrumen dengan tujuh kategori lebih bervariasi. Sehingga variansi skor responden juga akan meningkat. Meningkatnya variansi berhubungan dengan meningkatnya koefisien reliabilitas. Hal ini serupa dengan pendapat yang dikemukakan oleh Guilford seperti dikutip oleh Kirkendall dkk., (1980: 382) bahwa reliabilitas meningkat sebagaimana banyak respon pilihan meningkat, dengan reliabilitas paling besar ditemukan pada tujuh kategori respon. Selain itu pendapat serupa juga dikemukakan oleh Mueller (1986: 12) bahwa beberapa pembuat skala menggunakan tujuh kategori, dan beberapa lainnya lebih memilih empat atau enam kategori. Pengurangan jumlah kategori respon akan mengurangi penyebaran skor (mengurangi variansi) dan cenderung mengurangi reliabilitas. Meningkatkan jumlah kategori menambah variansi. Dengan kata lain jumlah kategori respon yang sedikit dalam hal ini empat kategori akan menyebabkan variansi rendah dan cenderung menyebabkan reliabilitas lebih rendah. Sebaliknya, peningkatan jumlah kategori respon akan meningkatkan variansi sehingga reliabilitas pun meningkat. Dalam artikelnya Preston (2000: 1-15) menyebutkan bahwa skor paling reliabel dihasilkan dari skala 7 kategori dan 10 kategori. Dengan demikian, rata-rata

koefisien reliabilitas instrumen disposisi matematika dengan tujuh kategori respon lebih tinggi dari rata-rata koefisien reliabilitas instrumen disposisi matematika dengan lima kategori respon.

Berdasarkan data yang diperoleh, respon siswa yang mengisi instrumen dengan enam kategori respon dan instrumen dengan lima kategori respon tidak terlalu berbeda sehingga rata-rata koefisien reliabilitasnya pun tidak signifikan berbeda. Aiken seperti dikutip oleh Whidiarso (2010: 2) menyebutkan bahwa nilai konsistensi internal (α) tidak berubah secara sistematis dengan meningkatnya jumlah kategori respon. Dengan kata lain, tidak setiap penambahan pilihan kategori pada instrumen akan menyebabkan peningkatan nilai koefisien reliabilitas. Seperti hasil studi Matell & Jacoby (Chang, diakses Nopember 2011) bahwa skala dengan lima kategori memiliki koefisien reliabilitas yang lebih tinggi dari skala dengan enam kategori. Dengan demikian tidak terdapat perbedaan rata-rata koefisien koefisien reliabilitas yang signifikan antara instrumen disposisi matematika dengan enam kategori respon dan instrumen dengan lima kategori respon.

Instrumen dengan enam kategori respon yang memiliki dua kategori lebih banyak dari instrumen dengan empat kategori respon menyebabkan respon siswa terhadap instrumen dengan tujuh kategori lebih bervariasi. Sehingga variansi skor responden juga akan meningkat. Meningkatnya variansi berhubungan dengan meningkatnya koefisien reliabilitas. Sehingga dapat dikatakan bahwa meningkatnya banyak kategori respon akan memberikan nilai koefisien reliabilitas. Hal ini diperkuat oleh hasil studi Chang seperti dikutip oleh Chomeya (2010: 402) bahwa validitas dan reliabilitas skala Likert dengan enam titik lebih tinggi dari skala Likert dengan empat titik (makin banyak skala memberikan nilai yang makin tinggi). Dengan demikian, rata-rata koefisien reliabilitas instrumen disposisi matematika dengan enam kategori respon lebih tinggi dari rata-rata koefisien reliabilitas instrumen dengan empat kategori respon.

Seperti halnya pada kasus instrumen dengan tujuh kategori dan enam kategori respon, bahwa perbedaan antara instrumen disposisi matematika dengan lima kategori respon dan empat kategori respon yaitu pada pilihan kategori N atau disebut pula dengan kategori titik tengah. Dapat dikatakan bahwa instrumen dengan lima kategori respon menyediakan titik tengah sementara instrumen dengan empat kategori respon tanpa titik tengah. Berdasarkan data yang diperoleh bahwa perbedaan skor siswa yang mengisi kedua instrumen tersebut tidaklah signifikan sehingga tidak menyebabkan perbedaan koefisien reliabilitas yang juga signifikan. Hal ini diperkuat oleh kesimpulan Widhiarso (2010: 2) bahwa skor skala yang menyediakan kategori titik tengah dengan yang tidak memiliki kategori titik tengah, tidak memiliki perbedaan yang berarti. Reliabilitas pengukuran dan validasi butir tidak mengalami perbedaan, tetapi yang berbeda adalah varian skor. Beberapa penelitian oleh Mattel & Jacoby (1971) dan Dawes (2001) memberikan hasil yang

serupa bahwa titik tengah tidak sungguh-sungguh dapat mempengaruhi validitas dan reliabilitas. Dengan demikian koefisien tidak terdapat perbedaan rata-rata reliabilitas antara instrumen dengan lima kategori respon kurang dan instrumen empat kategori respon.

SIMPULAN

Pada umumnya lima kategori merupakan jumlah kategori respon yang biasanya digunakan dalam penskalaan pendidikan atau psikologi. Namun temuan pada penelitian ini menunjukkan penambahan dua kategori atau tiga kategori respon pada instrumen disposisi matematika dapat menghasilkan koefisien reliabilitas yang lebih tinggi dari koefisien reliabilitas instrumen yang sebelumnya. Sehingga instrumen non tes dengan enam atau tujuh kategori respon dapat menjadi alternatif baru yang dapat diterapkan guru di sekolah. Pemilihan jumlah kategori respon yang digunakan tidak hanya berdasarkan pertimbangan kebutuhan dan karakteristik responden tetapi juga mengenai reliabilitas instrumen. Penyediaan pilihan kategori yang beragam memberikan kesempatan kepada siswa untuk dapat memberikan preferensi yang lebih sesuai dengan gejala sikap siswa sesungguhnya. Sehingga respon yang dihasilkan juga akan semakin beragam yang kemudian berpengaruh pada nilai koefisien reliabilitas instrumen non tes.

DAFTAR PUSTAKA

- Aiken, Lewis R. (2002). *Attitudes and Related Psychosocial Constructs: Theories, Assessment, and Research*. California: Sage Publications.
- Chang, Lei. (1994). "Psychometric Evaluation of 4-Point and 6-Point Likert Type Scales in Relation to Reliability and Validity." Di dalam *Jurnal Applied Psychological Measurement online*: <http://conservancy.umn.edu/bitstream/117059/1/v18n3p205.pdf>.
- Chomeya, Rungson. (2010). "Quality of Psychology Test Between Likert Scale 5 and 6 Points." Di dalam *Jurnal of Social Science online*: <http://thescipub.com/pdf/10.3844/jssp.2010.399.403>.
- Christensen, Larry S. (1988). *Experimental Methodology*. Boston: Allyn and Bacon, Inc.
- Darhim. (2013). "Pengaruh Pembelajaran Matematika Kontekstual Terhadap Sikap Siswa Sekolah Dasar." Di dalam *Jurnal online*: http://file.upi.edu/Direktori/FPMIPA/JUR._PEND._MATEMATIKA/195503031980021-DARHIM/Makalah_Artikel/JurnalSikapSiswa.pdf.

- Dawes, John. (2001). "Five Point vs Eleven Point Scales." Di dalam *Jurnal of Market Research*, Vol. 10(1).
- Kim, Saehee. (2010). "The Influence of Likert Scale Format on Response Result, Validity, and Reliability of Scale." Di dalam *Jurnal of Korean Society of Clothing and Textiles*, Vol. 34(6).
- Kirkendall, Don R., Joseph J. Gruber, dan Robert E. Johnson. (1980). *Measurement and Evaluation for Physical Educators*. Iowa: Brown Company Publishers.
- Leonard dan Supardi U.S. (2010). "Pengaruh Konsep Diri, Sikap Siswa Pada Matematika, Dan kecemasan Siswa Terhadap Hasil Belajar Matematika," *Jurnal Cakrawala Pendidikan*, Tahun XXIX(3).
- McLeod, Douglas B. (1992) *Handbook of Research on Mathematics Teaching and Learning* New York: MacMillan Publishing Company.
- Mardapi, Djemari. (2013). "Penilaian Karakter." Di dalam *Bahan tulisan Penilaian Pendidikan Karakter Universitas Negeri Yogyakarta online*: http://staff.uny.ac.id/sites/default/files/tmp/Penilaian_karakter.
- Mueller, Daniel J. (1986). *Measuring Social Attitude*. New York: Teachers College Press.
- Nitko, Antony J. (1996). *Educational Assessment of Students*. New Jersey: Prentice-Hall Inc.
- Popham, W. James. (1995). *Classroom Assessment, What Teacher Needs to Know*. Boston: Allyn and Bacon.
- Preston, C. C., & Colman, A. M. (2000). "Optimal Number of Response Categories in Rating Scales: Reliability, Validity, Discriminating Power, and Respondent Preferences." Di dalam *Jurnal Acta Psychologica*, 104.
- Weijters, Bers, Elke Cabooter, dan Niels Schillewaert. (2010). "The Effect of Rating Scales Rormat on Response Styles." Di dalam *Jurnal Forthcoming IJRM*, Vol. 27(3).
- Weng, Li Jen. (2004). "Impact of the Number of Response Categories and Anchor Labels on Coefficient Alpha and Test Retest Reliability." Di dalam *Jurnal Educational and Psychological Measurement*, Vol. 64(6).

Widhiarso, Wahyu. (2010). "Pengembangan Skala Psikologi." Di dalam *Makalah online*: http://widhiarso.staff.ugm.ac.id/files/widhiarso_2010_respon_alternatif_tengah_pada_skala_likert.pdf.

Widjajanti, Djamilah Bondan. (2011). "Mengembangkan Kecakapan Matematis Mahasiswa Calon Guru Matematika Melalui Strategi Perkuliahan Kolaboratif Berbasis Masalah." Di dalam *Makalah pada Prosiding Seminar Nasional Penelitian, Pendidikan, dan Penerapan MIPA online*: <http://staff.uny.ac.id/sites/default/files/131569335/Makalah%20Djamilah%20Semnas%2014%20MEI%202011.pdf>.