

SKALA PENGUKURAN DAN JUMLAH RESPON SKALA LIKERT

(The Measurement Scale and The Number of Responses in Likert Scale)

Weksi Budiaji^{1*}

¹Dosen Fakultas Pertanian Universitas Sultan Ageng Tirtayasa
Jl. Raya Jakarta Km 4 Pakupatan Serang Banten

*Korespondensi: budiaji@untirta.ac.id

Diterima: 10 September 2013 / Disetujui: 15 Oktober 2013

ABSTRACT

The ease of likert scale on its construction as a measurement scale of individual traits must be cautioned to prevent some errors in the data analysis. Some researchers consider likert scale as an interval scale, while the others mention likert scale as an ordinal scale. After reviewing some papers from some different authors, we argue that a likert scale using total score of all items is an interval scale. On the other hand, items using likert format is an ordinal scale. The number of responses in the likert scale suggested is 7 based on responden's preference that of they like it the most. Moreover, the 7 response format has a good reliability, validity, discriminating power, and test-retest (stability) index.

Keywords: *interval, likert, measurement, ordinal, scale*

ABSTRAK

Kemudahan penyusunan skala likert sebagai alat ukur perilaku individu perlu diperhatikan dengan hati-hati untuk menghindari kesalahan dalam analisis data. Beberapa peneliti menggolongkan skala likert kedalam dua skala yang berbeda yaitu ada yang mengolongkan kedalam skala ordinal dan yang lain menggolongkan kedalam skala interval. Setelah mereview beberapa tulisan dari peneliti yang berbeda, penulis berpendapat bahwa skala likert yang menggunakan skor total dari semua butir pertanyaan adalah skala interval. Dilain pihak, butir pertanyaan yang menggunakan tipe data likert mempunyai skala pengukuran ordinal. Jumlah titik respon yang disarankan penulis adalah 7 berdasarkan atas preferensi responden yang paling responden suka terhadap jumlah respon ini. Disamping itu, jumlah titik respon 7 juga mempunyai reliabilitas, validitas, kekuatan diskriminasi, dan stabilitas yang cukup baik.

Kata kunci: likert, skala, pengukuran, ordinal, interval

PENDAHULUAN

Penelitian-penelitian tentang perilaku personal yang mengukur sifat-sifat individu selalu menggunakan alat ukur yang dirancang sendiri oleh peneliti, baik melalui *pre-test* maupun *eliciting* (bertanya pada ahli) karena tidak ada

alat ukur yang pasti. Alat ukur yang digunakan adalah butir-butir pertanyaan yang dianggap sebagai indikator dari perilaku tertentu misalnya pengetahuan atau sikap. Butir-butir pertanyaan tersebut akan direspon oleh individu-individu yang akan diukur perilakunya.

Penggunaan butir-butir pertanyaan mengikuti pedoman skala pengukuran yang digunakan.

Ada beberapa skala pengukuran yang dapat digunakan dalam merancang skala pengukuran pada penelitian perilaku misalnya skala thurstone, guttman, dan likert. Skala thurstone dapat digunakan untuk menduga preferensi individu dengan menggunakan nilai frekuensi responnya. Posisi dari butir-butir pertanyaan dapat diperoleh dengan mengambil rata-rata dari persentil sebaran normal baku berdasarkan proporsi preferensi responden terhadap sebuah butir pertanyaan (Lipovetsky 2007). Skala guttman menggunakan skala kumulatif dimana jika individu setuju pada butir pertanyaan tertentu, maka individu tersebut juga setuju pada semua butir pertanyaan lain yang lebih lemah (pertanyaan sebelumnya). Skala guttman jarang dipakai peneliti karena membutuhkan upaya yang lebih gigih untuk mendapatkan butir-butir pertanyaan yang valid (Uhlener 2002). Skala yang paling mudah digunakan adalah skala likert. Skala likert menggunakan beberapa butir pertanyaan untuk mengukur perilaku individu dengan merespon 5 titik pilihan pada setiap pertanyaan, sangat setuju, setuju, tidak memutuskan, tidak setuju, dan sangat tidak setuju (Likert 1932).

Kemudahan penggunaan skala likert menyebabkan skala ini lebih banyak digunakan oleh peneliti. Kelly and Tincani (2013), misalnya, menggunakan skala likert untuk mengukur perilaku kerjasama individu yaitu dengan mengukur variabel ideologi, perspektif, pelatihan pribadi, dan pelatihan orang lain. Di bidang pertanian, skala likert juga sering digunakan untuk mengukur preferensi individu seperti pada preferensi konsumen terhadap penerimaan produk makanan yang telah dimodifikasi (Herath *et al.* 2013) dan preferensi petani terhadap karakteristik tanaman gandum yang ingin diusahakan (Nelson 2013).

Kemudahan yang ada pada penyusunan skala likert harus diperhatikan dengan hati-hati agar analisis lanjutan terhadap butir-butir respon tepat. Permasalahan yang sering terjadi adalah kebingungan dalam penggolongan skala likert ke dalam dua skala pengukuran yang berbeda yaitu diantara ordinal dan interval. Perdebatan antara penggolongan skala likert kedalam dua golongan skala yang berbeda telah berlangsung cukup lama karena penggolongan skala pengukuran yang berbeda akan berdampak pada analisis yang dapat digunakan pada skala likert. Tujuan penelitian ini adalah untuk mereview penggolongan skala likert yang telah dikemukakan beberapa peneliti. Kemudian, jumlah optimal titik respon pada skala likert juga akan dibahas, berjumlah genap atau ganjil.

METODE PENELITIAN

Penelitian ini adalah penelitian literatur. Literatur yang digunakan sebagai sumber adalah literatur yang akan dihubungkan dengan masalah yang ingin diteliti yaitu mengenai penggolongan skala likert dan jumlah respon skala likert.

HASIL DAN PEMBAHASAN

Ordinal vs Interval

Pertama, skala likert digolongkan kedalam skala ordinal. Jamieson (2004) menyatakan bahwa kategori respon pada skala likert mempunyai tingkatan tetapi jarak diantara kategori tidak dapat dianggap sama, sehingga skala likert adalah kelas skala ordinal. Jika analisis statistik yang digunakan salah, peluang kesimpulan salah juga meningkat. Jika skala likert dianggap sebagai skala ordinal, perhitungan rata-rata dan standar deviasi tidak tepat. Pada kelas ini, statistik yang dapat digunakan adalah median atau modus untuk menghitung ukuran pemusatannya, sedangkan variasi data dapat dilihat menggunakan frekuensi dari jawaban responden,

sehingga statistika parametrik tidak dapat diterapkan pada data ini. Jika skala ini dianalisis dengan statistika parametrik, maka langkah ini adalah salah, kesalahan pertama dari tujuh kesalahan yang umum dilakukan (Kuzon *et al.* 1996). Barua (2013) juga telah mengusulkan cara-cara dalam penentuan titik potong pada skala likert sebagai skala ordinal yang tidak mempunyai standar baku pada skala ordinal ini.

Dilain pihak, beberapa peneliti menganggap bahwa skala likert adalah skala pengukuran interval. Carrafio and Rocco (2007) menyatakan bahwa skala likert dapat menghasilkan skala pengukuran interval. Hal yang sama diungkapkan oleh Boone and Boone (2012) yang menjelaskan bahwa skala likert dapat dianalisis menggunakan statistika parametrik misalkan ANOVA (*analysis of variance*) maupun uji t. Carrafio and Rocco (2007) dan Boone and Boone (2012) mempunyai pendapat yang sama karena ada perbedaan yang cukup mendasar antara skala likert dan data likert (tipe/ format data likert).

Skala likert adalah skala pengukuran yang dikembangkan oleh Likert (1932). Skala likert mempunyai empat atau lebih butir-butir pertanyaan yang dikombinasikan sehingga membentuk sebuah skor/nilai yang merepresentasikan sifat individu, misalkan pengetahuan, sikap, dan perilaku. Dalam proses analisis data, komposit skor, biasanya jumlah atau rata-rata, dari semua butir pertanyaan dapat digunakan. Penggunaan jumlah dari semua butir pertanyaan valid karena setiap butir pertanyaan adalah indikator dari variabel yang direpresentasikannya. Boone and Boone (2012) memberikan contoh skala likert untuk mengukur sifat individu dalam hal kebiasaan memakan makanan

sehat seperti pada Tabel 1. Analisis dari jawaban pertanyaan pada Tabel 1 dapat menggunakan skor total dari jumlah jawaban tiap pertanyaan, sehingga skor yang diperoleh berkisar antara 5 sampai dengan 25. Karena skor 5 sampai dengan 25 tersebut berskala interval, analisis data parametrik dapat digunakan.

Disisi lain, tipe/format data likert adalah satu pertanyaan yang berdiri sendiri yang menggunakan format jawaban seperti pada skala likert. Jika seorang peneliti menggunakan butir pertanyaan yang banyak tetapi berdiri sendiri (bukan merupakan suatu indikator dari sifat tertentu misalkan sikap, kebiasaan atau pengetahuan), maka peneliti tersebut tidak dapat menggabungkan nilai dari jawaban tiap pertanyaan. Tabel 2 memperlihatkan contoh butir-butir pertanyaan yang berdiri sendiri, sehingga penjumlahan dari jawaban tidak dapat dilakukan. Pada format data likert, penggunaan analisis data parametrik tidak disarankan karena format data likert mempunyai skala kelas ordinal. Analisis data yang dapat digunakan misalnya median, modus, dan frekuensi.

Perbedaan yang mendasar dari kedua golongan skala (interval dan ordinal) pada skala likert adalah penggunaan skor total butir pertanyaan dan skor masing-masing butir pertanyaan. Skor total dari semua butir pertanyaan disebut *test score* yang merupakan hal utama yang menjadi perhatian dalam *classical test theory* (CTT) (McDonald 1999; Baker 2001). Penulis berpendapat, merujuk pada CTT yang berkembang saat Likert (1932) muncul, bahwa skala likert (bukan tipe data likert) adalah skala interval, sehingga *factor model* (model yang menggunakan analisis faktor) dapat diterapkan.

Tabel 1 Contoh skala likert sebagai skala interval

No	Pertanyaan	Sangat Tidak Setuju	Tidak Setuju	Netral	Setuju	Sangat Setuju
1	Saya makan makanan sehat secara teratur					
2	Saat membeli makanan di supermarket saya tidak memperdulikan makan 'sampah'					
3	Saat menyiapkan makan, saya memperhatikan kandungan lemak dalam makanan					
4	Saat menyiapkan makan, saya memperhatikan kandungan lemak dalam makanan					
5	Diet makanan sehat penting bagi keluarga saya					

Sumber: Boone and Boone (2012)

Tabel 2 Contoh format data likert sebagai skala ordinal

No	Pertanyaan	Sangat Tidak Setuju	Tidak Setuju	Netral	Setuju	Sangat Setuju
1	Program pelatihan sangat baik untuk menambah pengalaman saya					
2	Orang tua saya mendukung program pelatihan yang saya ikuti					
3	Keterlibatan saya dalam pelatihan akan memberikan perubahan pada saya					
4	Pembimbing pelatihan selalu ada untuk membimbing saya					
5	Teman-teman pelatihan akan mempengaruhi pekerjaan saya dimasa mendatang					

Penulis tidak sependapat dengan Carrafio and Rocco (2007) yang menyarankan agar menghindari analisis butir pertanyaan. Pada *modern test theory* yaitu *item response theory* (IRT), butir pertanyaan dapat dibandingkan satu sama lain sehingga butir pertanyaan dapat diurutkan berdasarkan tingkat kesulitannya (Embretson and Reise, 2000) dengan menggunakan model IRT. Model-model IRT tersebut dapat diperoleh dengan menggunakan model peluang dengan konsep yang berdasarkan jarak (Andrich 1988; Polak 2011; Budiaji 2012). Skala likert sama dengan *rating scale model* pada IRT karena mempunyai format jumlah respon yang

sama pada semua butir pertanyaan (Ostini and Nering, 2006).

Jumlah Titik Respon

Skala likert dikembangkan pertama kali menggunakan 5 titik respon yaitu sangat setuju, setuju, tidak memutuskan, tidak setuju, dan sangat tidak setuju (Likert 1932). Kemudian, beberapa peneliti mencoba mengembangkan jumlah titik yang lain dan membandingkannya. Dawes (2002) membandingkan jumlah 5 titik dengan 11 titik yang menghasilkan rata-rata yang sama setelah kedua skala tersebut diskala ulang. Jumlah titik 5 dan 11 adalah sebanding, artinya respon yang

menjawab pertanyaan dengan jumlah titik respon 5 dapat dikonversi kedalam jumlah titik respon 11 tanpa ada masalah. Tetapi, jumlah titik 11 mempunyai variasi yang lebih tinggi dibandingkan 5 titik, sehingga untuk analisis regresi lebih disukai karena keragamannya yang tinggi.

Dawes (2008) kemudian membandingkan jumlah titik 5, 7 dan 10. Jumlah titik respon pada skala likert akan berpengaruh pada rata-rata, ukuran penyebaran, dan bentuk sebaran. Rata-rata yang sama diperoleh saat skala likert menggunakan 5 titik dan 7 titik ketika keduanya diskala ulang. Jumlah titik 5 dan 7 juga sebanding. Tetapi jumlah titik 5 dan 7 tidak dapat dibandingkan dengan jumlah titik 10 karena jumlah titik 5 dan 7 menghasilkan rata-rata yang lebih tinggi dibandingkan 10.

Pembahasan yang lebih lengkap mengenai jumlah titik pada skala likert dijelaskan oleh (Preston and Colman 2000). Jumlah titik 2, 3, dan 4 menghasilkan indeks reliabilitas, validitas, dan kekuatan diskriminasi yang jelek. Hasil ini berbeda nyata dibandingkan jumlah titik 5, 6, dan 7 yang mempunyai indeks reliabilitas, validitas, dan kekuatan diskriminasi lebih baik. Ukuran stabilitas (test-retest validitas) menunjukkan jumlah titik 2, 3, dan 4 mempunyai stabilitas yang paling buruk dan semakin bagus saat jumlah titik meningkat menjadi 7, 8, 9 atau 10. Kemudian, ukuran stabilitas cenderung turun pada jawaban responden dengan jumlah titik lebih dari 10. Menurut responden, jumlah titik yang disukai adalah 10, 7, dan 9 sedangkan jumlah titik 5, 7, dan 10 lebih mudah dijawab. Jumlah titik 2, 3, dan 4 cukup cepat dijawab oleh responden, tetapi menurut responden kurang merepresentasikan jawaban responden. Preston and Colman (2000) kemudian menyarankan tingkatan jumlah titik yang disarankan adalah 7 kemudian 9 dan 10 berdasarkan kriteria yang telah disebutkan.

Selain menggunakan validitas dan reliabilitas, penggunaan kriteria lain

yaitu sensitivitas dan linearitas diberikan oleh Hofmans *et al.* (2007). Empat jumlah titik diuji dengan kedua kriteria ini yaitu jumlah titik 5, 7, 10, dan 15. Hasil yang diperoleh menunjukkan bahwa sensitivitas dan linearitas jumlah titik 5, 7, 10, dan 15 sama. Meskipun mempunyai sensitivitas dan linearitas yang sama, Hofmans *et al.* (2007) menyarankan jumlah titik 7 karena jumlah titik ini lebih disukai oleh responden.

Bias sosial menjadi pertimbangan Garland (1991) saat menentukan jumlah titik respon. Jumlah titik respon genap (jumlah titik 4) lebih disarankan daripada jumlah titik respon ganjil (jumlah titik 5) karena bias sosial dapat dikurangi. Bias sosial yang dimaksud adalah keinginan untuk menyenangkan interviewer atau perasaan agar dianggap penolong oleh interviewer karena mau menjadi responden. Responden akan cenderung menjawab dengan memilih netral (titik ditengah). Penelitian Garland (1991) perlu dikembangkan lagi karena menurut Preston and Colman (2000) jumlah titik kurang dari 5 mempunyai kriteria yang jelek dalam hal reliabilitas, validitas, kekuatan diskriminasi dan stabilitas.

Hasil-hasil penelitian mengenai jumlah titik respon dari skala likert menunjukkan bahwa jumlah titik 5, 7, 11 adalah sebanding. Jumlah titik 5, 7, 10, dan 15 juga sama dalam hal linearitas dan sensitivitas. Penulis menyarankan penggunaan jumlah titik 7 karena indeks validitas, reliabilitas, kekuatan diskriminasi, dan stabilitasnya baik dan disukai oleh responden.

KESIMPULAN

Skala likert (bukan tipe data likert) yang mengukur sifat-sifat (*traits*) individu misalnya pengetahuan atau sikap dengan menggunakan skor total dari butir pertanyaan adalah skala pengukuran interval. Disisi lain, butir pertanyaan yang menggunakan tipe data likert mempunyai skala pengukuran ordinal. Jumlah titik respon pada

skala likert yang disarankan adalah 7 titik respon karena lebih disukai responden dan mempunyai kriteria validitas, reliabilitas, kekuatan diskriminasi, dan stabilitas yang cukup baik.

DAFTAR PUSTAKA

- Andrich A. 1988. A Rating Formulation for Ordered Response Categories. *Psychometrika* 43:561 – 573
- Baker FB. 2001. *The Basic of Item Response Theory*. ERIC: USA
- Barua A. 2013. Methods for Decision-Making in Survey Questionnaires Based on Likert Scale. *Journal of Asian Scientific Research* 3 (1) pp: 35 – 38
- Boone Jr, Harry N, and Deborah A. Boone. 2012. Analyzing Likert Data. *Journal of Extension* 50 (2)
- Budiaji W. 2012. Item Response Models from A Distance Perspective [Thesis]. Leiden University: Leiden
- Carrafio J, and J Rocco. 2007. Ten common misunderstandings, misconceptions, persistent myths and urban legends about likert scales and likert response formats and their antidotes. *Journal of Social Sciences* 3 (3) pp: 106-116
- Dawes J. 2002. Five Point vs Eleven Point Scales: Does It Make Difference To Data Characteristics? *Australasian Journal of Market Research* 10 (1)
- Dawes J. 2008. Do data characteristics change according to the number of scale points used? An experiment using 5 point, 7 point and 10 point scales. *International Journal of Market Research* 50 (1)
- Embretson, Susan E, and Steven P Reise. 2000. *Item Response Theory for Psychologists*. Lawrence Erlbaum: New Jersey
- Garland R. 1991. The Mid-Point on a Rating Scale: Is it Desirable? *Marketing Bulletin 2 Research Note* 3 pp: 66-70
- Herath HMTK, JMM Udugama and UK. Jayasinghe-Mudalige. 2013. Women Consumer Preferences for Socially Responsible Food Production Atributes: Evidence from Urban Supermarket Setting in the Anuradhapura District.. *The Journal of Agricultural Sciences* 8 (2) pp: 57 – 69
- Hofmans, Joeri, Peter Theuns, and O Mairesse. 2007. Impact of the Number of Response Categories on Linearity and Sensitivity of Self-Anchoring Scales: A Functional Measurement Approach. *Methodology* 3 (4) pp: 160–169
- Jamieson S. 2004. Likert scales: How to (ab)use them. *Medical education*, 38(12): 1217-1218
- Kelly, Amy, and M Tincani. 2013. Collaborative Training and Practice among Applied Behavior Analysts who Support Individuals with Autism Spectrum Disorder. *Education and Training in Autism and Developmental Disabilities* 48(1) pp: 120–131
- Kuzon WM Jr, Urbanchek MG, and McCabe S. 1996. The seven deadly sins of statistical analysis. *Ann Plastic Surg* 37 pp: 265–72
- Likert RA. 1932. Technique for the measurement of attitudes. *Archives of Psychology*, 140 pp: 1-55
- Liposvetsky S. 2007. Thurstone Scaling in Order Statistics. *Mathematical and Computer Modelling* 45 pp: 917 – 926
- McDonald, Roderick P. 1999. *Test Theory: A Unified Treatment*. Lawrence Erlbaum: New Jersey
- Nelson, K Margaret. 2013. Analysis of Farmer Preferences for Wheat Variety Traits in Ethiopia: A Gender-Responsive Study. [Thesis]. Cornell University: New York
- Ostini, Remo, and Michael L. Nering. 2006. *Polytomous Item Response Theory Models*. Sage: California

- Polak M. 2011. Item Analysis of Single-Peaked Response Data: The Psychometric Evaluation of Bipolar Measurement Scales [PhD thesis]. Leiden University: Leiden
- Preston, Carolyn C, and Andrew M Colman. 2000. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica* 104: 1-15
- Uhlnaer, Lorraine M. 2002. The Use of the Guttman Scale in Development of a Family Bussiness Index. *Research Report H2002003*. SCALES: Zoertemeer