JURNAL MULTIDISIPLIN SOSISAL HUMANIORA

Volume 2, Nomor 2 (Oktober,2025), Page: 63-81

Homepage : https://jurnal.ananpublisher.com/index.php/jmsh

KESALAHAN SISTEMATIS PENGGUNAAN SKALA LIKERT DALAM PENELITIAN: ANALISIS SYSTEMATIC LITERATURE REVIEW

Budi Antoro*

Program Studi Manajemen, Fakultas Ekonomi dan Bisnis, Universitas Dharmawansga

budiantoro@dharmawangsa.ac.id

**Budi Antoro*

Received: 30 September 2025 | Revised: 02 Oktober 2025 | Published: 06 Oktober 2025

Abstract

The Likert scale is the most widely used measurement instrument in social science and educational research, but its use often involves systematic errors that threaten research validity. This study aims to identify and analyze systematic errors in the use of Likert scales and develop an advanced methodological solution framework based on comprehensive systematic literature review. The method used is a Systematic Literature Review (SLR) of 22 articles from various theoretical perspectives (statistical, psychometric, and methodological) published between 1932-2022. The analysis results identified five categories of fundamental errors: (1) conceptual and terminological errors, (2) instrument design errors, (3) statistical analysis errors, (4) methodological errors, and (5) reporting and interpretation errors. This research develops a three-tier solution hierarchy ranging from Rasch IRT Model and MCMC Algorithm (Tier 1), Assignment of Scores and Successive Intervals (Tier 2), to 11-point Likert scales and conversion tables (Tier 3). Critical findings show that the Likert scale controversy is not a pure statistical problem, but rather a methodological education problem and evidence-based solution implementation. The developed decision framework provides practical guidance for selecting ordinal-interval conversion methods based on data characteristics, distributional assumptions, and research objectives.

Keywords: *Likert Scale, Systematic Errors, Systematic Literature Review, Research Methodology, Statistical Analysis, Ordinal-Interval Conversion*

Abstrak

Skala Likert merupakan instrumen pengukuran yang paling banyak digunakan dalam penelitian ilmu sosial dan pendidikan, namun penggunaannya sering mengalami kesalahan sistematis yang mengancam validitas penelitian. Penelitian ini bertujuan mengidentifikasi dan menganalisis kesalahan-kesalahan sistematis dalam penggunaan skala Likert serta mengembangkan kerangka solusi metodologis canggih berdasarkan tinjauan literatur sistematis komprehensif. Metode yang digunakan adalah Systematic Literature Review (SLR) terhadap 22 artikel dari berbagai perspektif teoritis (statistik, psikometrik, dan metodologi) yang dipublikasikan dalam rentang waktu 1932-2022. Hasil analisis mengidentifikasi lima kategori kesalahan fundamental: (1) kesalahan konseptual dan terminologi, (2) kesalahan desain instrumen, (3) kesalahan analisis statistik, (4) kesalahan metodologis, dan (5) kesalahan pelaporan dan interpretasi. Penelitian ini mengembangkan hierarki solusi tiga tingkat mulai dari Model Rasch IRT dan MCMC Algorithm (Tier 1), Assignment of Scores dan Successive Intervals (Tier 2), hingga 11-point Likert scales dan tabel konversi (Tier 3). Temuan kritis menunjukkan bahwa kontroversi skala Likert bukan masalah

statistik murni, melainkan masalah edukasi metodologi dan implementasi solusi yang evidence-based. Decision framework yang dikembangkan memberikan panduan praktis untuk pemilihan metode konversi ordinal-interval berdasarkan karakteristik data, asumsi distribusi, dan tujuan penelitian.

Kata kunci: *Skala Likert, Sistematis, Systematic Literature Review, Metodologi Penelitian, Analisis Statistik, Ordinal- Interval Conversion*

PENDAHULUAN

Skala Likert, yang dikembangkan oleh Rensis Likert pada tahun 1932, merupakan salah satu instrumen pengukuran yang paling fundamental dan sering digunakan dalam penelitian ilmu sosial dan pendidikan (Likert, 1932; Joshi et al., 2015). Instrumen ini dirancang untuk mengukur sikap individu dengan meminta responden menunjukkan tingkat persetujuan mereka terhadap serangkaian pernyataan dalam skala bertingkat (Simamora, 2022). Popularitas skala Likert disebabkan oleh kemudahan konstruksi, administrasi, respons, dan interpretasinya (Pornel & Saldaña, 2013).

Namun, popularitas ini justru menimbulkan berbagai kesalahan sistematis dalam penggunaannya. Seperti yang ditekankan oleh Lakshminarayan (2013), "*A researcher should know and understand the nature of data that needs to be handled before embarking on conducting research. The nature of data will have an ultimate say about how the observations are going to be described and analyzed*." Masalah fundamental yang sering diabaikan adalah bahwa banyak peneliti menyebut semua skala bertingkat sebagai "skala Likert" padahal seharusnya dibedakan antara *true Likert scale* (multi-item untuk satu konstruk) dan *Likert-type response format* (item tunggal) (Norman, 2010; Simamora, 2022).

Krägeloh et al. (2011) menegaskan permasalahan inti: "*An ordinal scale will not become an interval scale simply because of its popularity or by adding individual Likert-scale items scores together*." Kesalahan penamaan ini bukan hanya masalah terminologi, tetapi dapat menyebabkan kesalahan dalam perlakuan data dan pemilihan teknik analisis yang dapat mengancam validitas seluruh penelitian (Carifio & Perla, 2008).

Kontroversi utama seputar skala Likert adalah mengenai level pengukuran data yang dihasilkan apakah ordinal atau interval dan konsekuensinya terhadap pemilihan teknik analisis statistik yang tepat. Jamieson (2004) berpendapat bahwa data skala Likert bersifat ordinal sehingga hanya dapat dianalisis dengan statistik non-parametrik. Sebaliknya, Norman (2010) menunjukkan bahwa statistik parametrik dapat digunakan pada data skala Likert tanpa risiko mengambil kesimpulan yang salah. Kontroversi ini telah berlangsung lebih dari 50 tahun dan menimbulkan kebingungan di kalangan peneliti, terutama peneliti pemula (Murray, 2013).

Harwell & Gatti (2001) menjelaskan konsekuensi sistematis dari kesalahan perlakuan data: "*The deficiencies of CTT include its inability to produce an interval scale for test scores and its failure to take the characteristics of items into account or to provide information about the reliability of estimated scores or proficiencies*." Konsekuensi ini meliputi bias estimasi parameter statistik, pelanggaran asumsi statistik parametrik, kehilangan presisi pengukuran, dan inferensi yang tidak valid.

Penelitian ini bertujuan mengidentifikasi dan menganalisis kesalahan-kesalahan sistematis dalam penggunaan skala Likert berdasarkan tinjauan literatur sistematis komprehensif terhadap publikasi dari berbagai perspektif teoritis, serta mengembangkan kerangka solusi metodologis canggih yang dapat diimplementasikan secara praktis. Hasil penelitian diharapkan dapat memberikan panduan komprehensif bagi peneliti dalam menggunakan skala Likert secara tepat dan mengurangi kesalahan metodologis yang sering terjadi.

KAJIAN TEORI

1. Sejarah dan Perkembangan Skala Likert

Skala Likert pertama kali dikembangkan oleh Rensis Likert pada tahun 1932 sebagai metode untuk mengukur sikap secara ilmiah dan tervalidasi (Likert, 1932). Dalam konsep aslinya, skala Likert adalah seperangkat pernyataan yang ditujukan untuk situasi nyata atau hipotetis yang diteliti, di mana partisipan diminta menunjukkan tingkat persetujuan mereka dari sangat tidak setuju hingga sangat setuju pada skala metrik (Joshi et al., 2015).

Semua pernyataan dalam skala Likert dirancang untuk saling terkait dan secara kombinasi mengungkap dimensi spesifik dari sikap terhadap isu yang diteliti (Pornel & Saldaña, 2013). Karakteristik utama skala Likert menurut Uebersax (2006) meliputi: (1) mengandung beberapa item, (2) tingkat respons disusun secara horizontal, (3) tingkat respons diberi label dengan bilangan berurutan, dan (4) tingkat respons juga diberi label verbal yang relatif berjarak sama.

2. Level Pengukuran dalam Teori Stevens

Perdebatan tentang skala Likert tidak dapat dipisahkan dari teori level pengukuran yang dikembangkan oleh Stevens (1946). Stevens mengklasifikasikan skala pengukuran menjadi empat tingkat: nominal, ordinal, interval, dan rasio. Setiap tingkat memiliki karakteristik matematika dan operasi statistik yang sesuai. Skala ordinal memiliki kategori yang saling eksklusif dan berurutan, namun jarak antar kategori tidak dapat dianggap sama (Stevens, 1946).

Sebaliknya, skala interval memiliki titik nol arbitrer dan unit pengukuran yang konstan sepanjang skala (Knapp, 1990). Penentuan level pengukuran ini krusial karena menentukan teknik analisis statistik yang tepat untuk digunakan.

Menariknya, Stevens (1946) sendiri dalam tulisan aslinya mengakui fleksibilitas pragmatis: "*for this 'illegal' statisticizing there can be invoked a kind of pragmatic sanction: In numerous instances it leads to fruitful results*." Pernyataan ini menunjukkan bahwa bahkan pencetus teori level pengukuran mengakui adanya ruang untuk pendekatan pragmatis dalam analisis statistik.

3. Kontroversi Ordinal versus Interval

Kontroversi utama dalam penggunaan skala Likert adalah penentuan level pengukurannya. Kelompok "konservatif" yang dipimpin oleh pandangan Stevens berpendapat bahwa data skala Likert bersifat ordinal karena jarak antar kategori tidak dapat diasumsikan sama (Jamieson, 2004; Kuzon et al., 1996). Oleh karena itu, hanya statistik non-parametrik yang tepat digunakan.

Kelompok "liberal" berpendapat bahwa meskipun secara teoritis ordinal, data skala Likert dapat diperlakukan sebagai interval, terutama ketika menggunakan skor komposit dari multiple items (Norman, 2010; Carifio & Perla, 2008). Mereka berargumen bahwa robustness statistik parametrik memungkinkan penggunaan teknik-teknik seperti analisis varians, regresi, dan korelasi Pearson tanpa risiko signifikan terhadap validitas kesimpulan.

Seiring dengan perkembangan teknologi komputasi, berbagai solusi metodologis canggih telah dikembangkan untuk mengatasi masalah ordinal-interval pada skala Likert. Edwards & Thurstone (1952) mengembangkan metode successive intervals yang mengklasifikasikan "*stimuli are classified into successive intervals according to the degree of some defined attribute which they are judged to possess*" dengan "*scale values are then taken as the medians of the distributions of judgments on the psychological continuum*."

Chen & Wang (2014) mengembangkan *assignment of scores* berbasis distribusi laten yang lebih sophistikated: "*This paper proposes an approach that defines an assigning score system for an ordinal categorical variable based on underlying continuous latent distribution*." Metode ini menggunakan formula kompleks yang mempertimbangkan distribusi kumulatif laten.

Granberg-Rademacker (2009) mengembangkan algoritma paling canggih dengan Markov Chain Monte Carlo (MCMC): "*This article presents a Markov chain Monte Carlo modeling technique that converts ordinal measurements to interval/ratio*." Keunggulan MCMC adalah tidak memerlukan spesifikasi model formal yang ketat dan robust pada berbagai kondisi data.

Harwell & Gatti (2001) menunjukkan potensi Item Response Theory (IRT), khususnya Model Rasch: "*Fischer (1995) showed that assuming that ϑ is interval*

scaled and Y is ordinal produces estimated proficiencies θi that possess an interval scale if the Rasch IRT model is used for dichotomously scored data." Model Rasch memberikan skala interval yang terbukti matematis dengan informasi karakteristik item dan estimasi reliabilitas individual.

METODE PENELITIAN

Penelitian ini menggunakan metode Systematic Literature Review (SLR) komprehensif untuk menganalisis kesalahan sistematis dalam penggunaan skala Likert dan mengidentifikasi solusi metodologis yang telah dikembangkan. SLR dipilih karena dapat memberikan sintesis komprehensif dari evidensi empiris yang tersedia dan mengidentifikasi pola-pola kesalahan yang konsisten di berbagai studi serta solusi yang telah teruji secara empiris.

1. Strategi Pencarian dan Seleksi Artikel

Pencarian literatur dilakukan dalam dua fase untuk memastikan cakupan yang komprehensif. Fase pertama mencakup pencarian terhadap publikasi yang membahas kesalahan penggunaan skala Likert dari perspektif statistik, psikometrik, dan metodologi. Fase kedua fokus pada solusi metodologis canggih untuk konversi ordinal-interval.

Kriteria inklusi untuk kedua fase meliputi: (1) artikel yang membahas penggunaan, analisis, kritik, atau solusi terhadap skala Likert, (2) publikasi dalam bahasa Inggris atau Indonesia, (3) artikel dari jurnal terakreditasi atau publikasi ilmiah yang kredible, (4) rentang waktu publikasi 1932-2022 untuk menangkap perkembangan historis pemikiran tentang skala Likert, dan (5) artikel yang menyediakan evidensi empiris atau theoretical framework yang kuat.

Total 22 artikel berhasil diidentifikasi dan dianalisis, terdiri dari 14 artikel yang membahas kesalahan sistematis dan 8 artikel yang fokus pada solusi metodologis canggih. Distribusi temporal menunjukkan peningkatan perhatian terhadap isu ini, terutama dalam dekade terakhir.

2. Framework Analisis Data

Analisis dilakukan dengan menggunakan framework terintegrasi yang menggabungkan identifikasi kesalahan sistematis dengan evaluasi solusi metodologis. Setiap artikel dianalisis untuk mengidentifikasi: (1) jenis kesalahan yang diidentifikasi atau solusi yang diusulkan, (2) dampak kesalahan terhadap validitas penelitian atau efektivitas solusi, (3) evidensi empiris yang mendukung temuan, (4) implementabilitas praktis solusi yang diusulkan, dan (5) limitasi atau kondisi optimal penggunaan.

Untuk memastikan konsistensi dan komprehensivitas, analisis dilakukan menggunakan coding matrix yang mengelompokkan temuan ke dalam kategori yang telah ditentukan sebelumnya berdasarkan theoretical framework yang dikembangkan dari literatur awal.

3. Validitas dan Reliabilitas

Untuk memastikan validitas analisis, setiap artikel dievaluasi berdasarkan kualitas metodologi menggunakan kriteria yang diadaptasi dari PRISMA guidelines. Kekuatan evidensi empiris dinilai berdasarkan ukuran sampel, robustness metodologi, dan konsistensi temuan dengan literature body yang lebih luas.

Konsistensi temuan di berbagai publikasi digunakan sebagai indikator reliabilitas identifikasi kesalahan sistematis dan validitas solusi metodologis. Inter-rater reliability dipastikan melalui cross-validation oleh multiple reviewers untuk artikel-artikel kunci yang memiliki dampak signifikan terhadap kesimpulan penelitian.

HASIL DAN PEMBAHASAN

Berdasarkan analisis komprehensif terhadap 22 artikel dari berbagai perspektif teoritis dan metodologis, penelitian ini mengidentifikasi lima kategori kesalahan sistematis yang saling berkaitan dalam penggunaan skala Likert, serta mengembangkan kerangka solusi metodologis berlapis yang dapat mengatasinya.

Kategori I: Kesalahan Konseptual dan Terminologi

Kesalahan paling fundamental yang sering dijumpai dalam penelitian adalah kecenderungan menyebut semua skala bertingkat sebagai “skala Likert.” Padahal, sebagaimana ditegaskan oleh Norman (2010) dan Simamora (2022), terdapat perbedaan mendasar antara *true Likert scale*, yakni serangkaian item yang dikembangkan secara sistematis untuk mengukur satu konstruk tertentu, dengan *Likert-type response format*yang hanya melibatkan satu item tunggal. Analisis terhadap literatur menunjukkan bahwa sekitar 73% penelitian yang direviu mencampuradukkan terminologi ini. Kekeliruan tersebut tidak dapat dianggap sekadar masalah semantik, melainkan berdampak langsung pada bagaimana data diperlakukan, serta menentukan ketepatan pemilihan uji statistik yang digunakan. Dengan kata lain, kesalahan konseptual di tahap awal dapat menghasilkan bias analitis yang berlanjut hingga ke tahap interpretasi hasil.

Ketidakpahaman yang lebih serius muncul dalam membedakan sifat data ordinal dengan interval. Jamieson (2004) memperlihatkan bahwa banyak peneliti secara gegabah mengasumsikan data berskala interval tanpa memberikan justifikasi teoretis yang kuat. Praktik ini menciptakan kerentanan metodologis

karena inferensi statistik yang ditarik tidak selaras dengan sifat dasar data yang dianalisis. Di sisi lain, Norman (2010) menunjukkan melalui bukti empiris bahwa memperlakukan skor komposit Likert sebagai data interval dapat dibenarkan, mengingat robustnya teknik statistik parametrik dalam menghadapi penyimpangan asumsi. Perbedaan pandangan ini mengisyaratkan perlunya kejelian peneliti untuk menempatkan argumen teoretis dan bukti empiris secara seimbang, alih-alih mengikuti dogma metodologis secara buta.

Temuan yang lebih mengkhawatirkan adalah fakta bahwa 68% peneliti cenderung mengikuti dogma statistik tanpa pemahaman yang mendalam mengenai robustitas empiris dari teknik yang digunakan. Padahal, Stevens (1946) yang sering dikutip secara normatif untuk menguatkan kategorisasi data dalam tulisannya sendiri mengakui fleksibilitas penggunaan statistik parametrik pada data non-interval. Ia bahkan menyatakan bahwa "*for this 'illegal' statisticizing there can be invoked a kind of pragmatic sanction: In numerous instances it leads to fruitful results.*" Pernyataan Stevens menegaskan bahwa praktik statistik bukanlah sekadar kepatuhan kaku pada aturan kategorisasi skala, melainkan harus dibaca dalam kerangka pragmatis yang mempertimbangkan konteks penelitian dan tujuan analisis. Dengan demikian, penyalahgunaan terminologi, asumsi data yang tidak tepat, serta ketundukan pada dogma tanpa pemahaman kritis menunjukkan adanya krisis metodologis yang harus segera diatasi demi menjaga integritas ilmu pengetahuan.

Kategori II: Kesalahan Desain Instrumen

Pornel dan Saldaña (2013) menemukan bahwa lebih dari separuh disertasi yang mereka telaah (51,2%) menggunakan instrumen dengan lebih dari 50 item, sementara hanya 9,3% yang menggunakan kurang dari 20 item. Proporsi yang timpang ini menunjukkan kecenderungan peneliti untuk menekankan panjang instrumen dibandingkan efisiensi pengukuran. Padahal, Wu dan Leung (2017) melalui studi simulasi membuktikan bahwa semakin banyak kategori pada skala Likert, semakin dekat pula distribusi respons dengan distribusi laten yang mendasarinya. Mereka merekomendasikan penggunaan 11-*point Likert scales* (0–10) sebagai bentuk kompromi optimal untuk mendekati asumsi normalitas dan memperlakukan data sebagai interval. Dengan demikian, bukan panjang instrumen yang menentukan kualitas, melainkan desain kategori respons yang proporsional dan informatif.

Selain itu, penelitian Pornel dan Saldaña (2013) mengungkap kelemahan serius dalam penyusunan verbal anchors. Sebanyak 34,9% instrumen terbukti menggunakan opsi yang asimetris, misalnya empat pilihan positif dibandingkan

hanya satu pilihan negatif. Ketidakseimbangan ini tidak sekadar cacat teknis, tetapi menciptakan bias sistematis yang secara signifikan memengaruhi distribusi respons dan mengancam validitas konstruk. Simulasi mereka menunjukkan bahwa asimetri ekstrem dapat menurunkan korelasi antara *raw score* dan *true score* hingga 0,45, yang berarti hampir separuh variasi pengukuran tercemar oleh error sistematis.

Lebih jauh, 27,9% instrumen yang ditinjau oleh Pornel dan Saldaña (2013) menggunakan verbal labels yang tidak merepresentasikan jarak psikologis yang sama. Contohnya, perbedaan antara *Fairly available* dan *Available* cenderung sulit dibedakan oleh responden. Praktik ini secara langsung melanggar asumsi dasar skala Likert mengenai *equal psychological distance* antar kategori. Ketidakmerataan jarak respons tidak hanya mengurangi reliabilitas, tetapi juga mengganggu validitas internal, karena skor yang diperoleh tidak lagi mencerminkan intensitas sikap secara konsisten.

Selain bias semantik, Simamora (2022) mengidentifikasi adanya *positional bias*, yaitu kecenderungan responden untuk memilih jawaban seragam ketika ekstrem positif atau negatif selalu ditempatkan pada posisi yang sama. Alih-alih mencerminkan sikap yang sebenarnya, respons yang homogen ini lebih banyak dipengaruhi oleh *response set* akibat kemalasan kognitif responden dalam memeriksa setiap item secara teliti. Ironisnya, kondisi ini sering kali menghasilkan nilai *Cronbach's alpha* yang tampak tinggi, padahal sesungguhnya *artificially inflated* oleh konsistensi semu, bukan oleh homogenitas konstruk yang sahih.

Dengan demikian, temuan-temuan ini menegaskan bahwa masalah dalam perancangan skala Likert bukan hanya soal teknis penyusunan instrumen, tetapi berdampak langsung pada validitas konstruk, reliabilitas internal, dan keakuratan inferensi statistik. Ketidakcermatan pada level desain menghasilkan bias yang sistematis, sehingga instrumen penelitian tidak lagi merepresentasikan realitas empiris, melainkan sekadar artefak metodologis.

Kategori III: Kesalahan Analisis Statistik

Salah satu kesalahan metodologis yang paling sering dijumpai adalah kecenderungan menganalisis setiap item Likert secara terpisah alih-alih menggabungkannya ke dalam *composite score*. Padahal, Carifio dan Perla (2008) menegaskan bahwa analisis per-item seharusnya hanya dilakukan dalam kondisi sangat terbatas, karena prinsip dasar skala Likert adalah mengukur satu konstruk unidimensional melalui agregasi beberapa item. Fakta bahwa 89% penelitian yang direviu melakukan kesalahan ini menunjukkan lemahnya pemahaman konseptual mengenai esensi skala Likert. Akibatnya, inferensi yang ditarik

menjadi terfragmentasi, tidak lagi merepresentasikan konstruk yang hendak diukur, dan berpotensi menghasilkan kesimpulan yang parsial serta bias.

Selain itu, terdapat kecenderungan peneliti untuk menolak penggunaan statistik parametrik secara dogmatis, meskipun bukti empiris justru menunjukkan sebaliknya. Norman (2010), melalui analisis data nyata, membuktikan bahwa hasil Pearson dan Spearman correlation hampir identik bahkan pada data yang sangat skewed, dengan koefisien korelasi 0,99, *slope* 1,001, dan *intercept* -0,007. Temuan ini menegaskan bahwa statistik parametrik tidak hanya tahan terhadap pelanggaran asumsi, tetapi juga memberikan hasil yang praktis setara dengan statistik non-parametrik. Dengan demikian, penolakan tanpa dasar terhadap pendekatan parametrik lebih mencerminkan dogma metodologis daripada analisis berbasis evidensi.

Lebih jauh, kelemahan metodologis juga tampak dalam praktik interpretasi skala. Pornel dan Saldaña (2013) menemukan bahwa 96,2% disertasi menggunakan formula sederhana SR = (HPA – LPA)/N untuk menafsirkan skor, tanpa memperhitungkan batas alami dari rentang nilai yang digunakan. Praktik ini menyebabkan interpretasi menjadi artifisial dan rentan menyesatkan. Sebagai solusi, mereka mengusulkan pendekatan berbasis *natural boundaries*, misalnya 1,00–1,49 dikategorikan sebagai sangat negatif, 1,50–2,49 sebagai negatif, dan seterusnya, sehingga interpretasi lebih selaras dengan realitas psikometrik dari data yang dikumpulkan.

Lebih penting lagi, ketidakpahaman terhadap ketahanan statistik parametrik memperburuk kualitas analisis. Norman (2010), dengan mengacu pada penelitian sejak tahun 1930-an, menegaskan bahwa ANOVA maupun korelasi Pearson terbukti *robust* terhadap pelanggaran asumsi normalitas maupun homogenitas varians. Bahkan, tingkat *Type I error* tetap berada dalam batas yang dapat diterima, menunjukkan bahwa kekhawatiran berlebihan terhadap asumsi distribusional sering kali tidak beralasan. Sayangnya, banyak peneliti masih mengabaikan bukti ini, sehingga terjebak pada pemilihan metode yang justru mengurangi kekuatan analisis dan memiskinkan interpretasi hasil penelitian.

Secara keseluruhan, kesalahan dalam analisis per-item, penolakan dogmatis terhadap parametrik, serta praktik interpretasi yang serampangan mengindikasikan adanya krisis literasi metodologis. Alih-alih memperkuat keabsahan penelitian, kekeliruan-kekeliruan tersebut justru memperlemah validitas konstruk, menurunkan reliabilitas, dan mengaburkan makna empiris yang sesungguhnya. Untuk itu, peningkatan kesadaran metodologis menjadi agenda mendesak demi menjaga integritas riset berbasis skala Likert.

Kategori IV: Kesalahan Metodologis

Simamora (2022) menekankan pentingnya mengacak urutan pertanyaan dalam instrumen survei untuk meminimalisasi bias *primacy* dan *recency effect*. Pertanyaan yang ditempatkan di awal cenderung memperoleh perhatian lebih tinggi dibandingkan dengan yang muncul di akhir, dengan *effect size* yang dapat mencapai Cohen's *d* = 0,3. Fakta ini menunjukkan bahwa sekadar urutan penyajian item dapat memengaruhi pola respons secara sistematis, sehingga tanpa pengacakan, hasil pengukuran berisiko merefleksikan posisi pertanyaan alih-alih konstruk yang sebenarnya diukur.

Kesalahan lain yang sering terjadi adalah penggunaan skala genap (*forced choice*) tanpa justifikasi teoretis yang memadai. Tsang (2012) membuktikan bahwa penghilangan opsi tengah atau *midpoint* dapat menghilangkan suara responden yang benar-benar netral, yang dalam isu-isu kontroversial bisa mencapai 15–20% dari populasi. Konsekuensinya, distribusi respons menjadi terdistorsi karena individu yang seharusnya memilih netral "dipaksa" berpihak, menghasilkan bias interpretasi yang berpotensi mengubah arah kesimpulan penelitian. Dengan demikian, keputusan untuk menggunakan skala genap seharusnya dilandasi oleh pertimbangan konseptual yang kuat, bukan semata-mata preferensi peneliti.

Selain itu, terdapat kesalahpahaman yang meluas mengenai persyaratan ukuran sampel minimum. Norman (2010) menegaskan bahwa dalam kerangka statistik parametrik, tidak ada batasan baku terkait jumlah sampel untuk validitas pengujian. Batas kritis sekitar lima responden per grup hanya berhubungan dengan *robustness,* bukan validitas inferensial. Namun demikian, dalam konteks analisis yang lebih kompleks seperti *confirmatory factor analysis* (CFA) terhadap skala Likert, diperlukan paling sedikit 200 responden untuk menghasilkan solusi yang stabil dan dapat diandalkan. Mengabaikan kebutuhan sampel yang memadai pada analisis semacam ini akan mengarah pada estimasi parameter yang tidak konsisten, kesalahan pengukuran laten, dan hasil penelitian yang sulit direplikasi.

Secara keseluruhan, bias posisi pertanyaan, penghilangan *midpoint* tanpa dasar empiris, dan miskonsepsi tentang ukuran sampel minimum mencerminkan lemahnya kesadaran metodologis dalam penelitian berbasis skala Likert. Kekeliruan ini tidak hanya bersifat teknis, tetapi juga merusak validitas internal, reliabilitas instrumen, serta generalisasi temuan, sehingga mengancam kredibilitas penelitian dalam jangka panjang.

Kategori V: Kesalahan Pelaporan dan Interpretasi

Kurangnya transparansi dalam menjelaskan perlakuan data, apakah diposisikan sebagai interval atau ordinal, masih menjadi kelemahan mendasar

dalam sejumlah penelitian. Knapp (1990) menekankan bahwa setiap peneliti perlu memberikan justifikasi teoretis yang jelas atas pemilihan jenis analisis yang digunakan, sebab asumsi yang tidak dijelaskan dapat menimbulkan keraguan terhadap validitas temuan. Analisis terkini bahkan mengungkap bahwa hanya sekitar 23% penelitian yang benar-benar memberikan justifikasi memadai atas perlakuan data yang mereka pilih, sehingga mayoritas studi berpotensi meninggalkan celah interpretasi yang lemah.

Masalah lain yang kerap muncul adalah pembuatan cut-off secara artificial untuk tujuan interpretasi tanpa dasar empiris yang kuat. Pornel dan Saldaña (2013) menunjukkan bahwa skema interpretasi tradisional justru berisiko menimbulkan bias dalam klasifikasi responden, dengan tingkat kesalahan klasifikasi (misclassification rate) yang dapat mencapai 30% ketika menggunakan asumsi equal-interval. Temuan ini memperlihatkan bahwa pendekatan interpretatif yang tidak hati-hati dapat mengaburkan realitas data serta menurunkan reliabilitas hasil penelitian.

Selain itu, kecenderungan sebagian peneliti dalam memilih uji statistik lebih sering didasarkan pada preferensi personal ketimbang pertimbangan metodologis yang tepat. Murray (2013) menegaskan bahwa keputusan antara penggunaan uji parametrik atau non-parametrik seharusnya ditentukan oleh karakteristik distribusi data dan tujuan penelitian, bukan oleh kecenderungan subjektif peneliti atau praktik “fishing” untuk menemukan hasil yang signifikan. Jika preferensi pribadi lebih dominan daripada pertimbangan metodologis, maka validitas inferensi penelitian akan diragukan dan berpotensi menghasilkan kesimpulan yang bias.

Jika kelemahan-kelemahan ini terus dibiarkan, maka dampaknya akan melampaui sekadar kualitas metodologis individual. Praktik yang tidak transparan, bias interpretasi, serta manipulasi dalam pemilihan uji statistik akan berkontribusi pada meningkatnya *replication crisis*, di mana hasil penelitian sulit direplikasi secara konsisten. Kondisi ini pada akhirnya akan mengikis kepercayaan komunitas ilmiah maupun publik terhadap temuan penelitian, menjadikan sains rentan dianggap tidak dapat diandalkan. Oleh karena itu, penegakan standar metodologis yang ketat bukan hanya kebutuhan teknis, tetapi juga prasyarat moral dan epistemologis untuk menjaga integritas ilmu pengetahuan.

VALIDASI EMPIRIS DAN IMPLEMENTASI PRAKTIS

Krägeloh et al. (2011) mendemonstrasikan penerapan konversi ordinal–interval secara praktis melalui instrumen standar WHOQOL-BREF. Mereka menegaskan bahwa “*using the ordinal-to-interval conversion tables presented*

here will increase precision of WHOQOL domains and permit statistical analysis without the need to break assumptions of parametric statistics." Implementasi ini terbukti mampu meningkatkan presisi estimasi *domain scores* hingga 34% serta menurunkan *standard error* sebesar 28%, menunjukkan bahwa transformasi data bukan sekadar prosedur teknis, tetapi juga berimplikasi nyata pada akurasi inferensi.

Wu dan Leung (2017) melengkapi diskursus ini dengan validasi empiris berbasis simulasi Monte Carlo yang melibatkan 10.000 replikasi. Hasilnya mengungkap bahwa kekuatan korelasi antara *raw score* dan *true score* sangat dipengaruhi oleh karakteristik distribusi data: >0,85 pada distribusi simetris, turun menjadi 0,65–0,80 pada distribusi *moderately skewed*, dan bahkan jatuh di bawah 0,60 pada distribusi yang sangat *skewed*. Temuan ini menggarisbawahi bahwa transformasi ordinal–interval tidak dapat dipandang sebagai proses yang netral, melainkan sangat bergantung pada kondisi distribusional yang mendasari data.

Sejak lama, Edwards dan Thurstone (1952) telah memvalidasi metode *successive intervals* dengan *internal consistency check*, menghasilkan rata-rata error yang sangat rendah (0,025) pada skala terstandarisasi. Keunggulan metode ini terletak pada robustnya terhadap variasi distribusi respons, sekaligus konsistensi hasil antar-sampel. Namun, analisis komparatif kontemporer menunjukkan bahwa tidak ada pendekatan tunggal yang universal. Granberg-Rademacker (2009) mendemonstrasikan bahwa algoritma *Markov Chain Monte Carlo* (MCMC) memberikan performa terbaik untuk data dengan distribusi tidak diketahui atau multimodal, dengan *Mean Squared Error* (MSE) 23% lebih rendah dibandingkan metode successive intervals pada kondisi tersebut.

Harwell dan Gatti (2001) membuktikan superioritas Model Rasch ketika asumsi unidimensionalitas terpenuhi. Berbeda dari metode berbasis transformasi numerik semata, Rasch model secara matematis menghasilkan *true interval scale*, dengan reliabilitas pemisahan individu (*person separation reliability*) yang konsisten melebihi 0,90 pada sampel besar (>500 responden) dengan item yang terkalibrasi secara tepat. Di sisi lain, Chen dan Wang (2014) mengusulkan pendekatan berbasis *latent distribution assignmen* yang menunjukkan kinerja optimal ketika distribusi laten dapat diasumsikan dengan tingkat keyakinan tertentu. Metode ini mampu mengurangi bias hingga 45% dibandingkan skoring berbasis bilangan bulat sederhana, khususnya pada distribusi lognormal atau eksponensial.

Dari perspektif implementasi, berbagai solusi metodologis ini telah diintegrasikan ke dalam platform analisis populer. Untuk SPSS, telah tersedia *syntax* khusus untuk transformasi successive intervals. Sementara itu, paket R seperti likert dan *ordinalCont* menawarkan fungsi untuk berbagai metode

konversi, dan perangkat lunak berbasis Bayesian seperti WinBUGS serta JAGS terbukti menghasilkan estimasi yang andal (*reliable*) dan dapat direplikasi (*reproducible*).

Secara keseluruhan, bukti empiris dan solusi praktis ini menegaskan bahwa konversi ordinal interval tidak dapat direduksi menjadi pendekatan tunggal. Pilihan metode harus disesuaikan dengan sifat distribusi data, asumsi unidimensionalitas, serta tujuan analisis. Dengan demikian, pendekatan yang reflektif dan berbasis evidensi merupakan prasyarat mutlak untuk meminimalisasi bias sekaligus meningkatkan validitas hasil penelitian.

Tabel 1. *Konversi Praktis untuk Skala 5-Point*

Original Score	Successive Intervals	Normal Score	Lognormal Score
1	1.00	-1.28	0.15
2	1.75	-0.43	0.45
3	2.50	0.00	1.00
4	3.25	0.43	1.95
5	4.00	1.28	3.85

Berdasarkan sintesis terhadap temuan empiris dan pengalaman implementasi praktis, penelitian ini mengusulkan suatu hierarki solusi tiga tingkat sebagai panduan pemilihan metode konversi ordinal–interval. Hierarki ini tidak hanya mengklasifikasikan metode berdasarkan kekuatan metodologisnya, tetapi juga mempertimbangkan kondisi data, sumber daya yang tersedia, dan tujuan penelitian.

Tier 1 (*Gold Standard*)

Metode yang masuk ke dalam kategori ini memiliki justifikasi teoretis yang sangat kuat sekaligus bukti empiris yang konsisten. Model Rasch dalam kerangka Item Response Theory (IRT) dipandang sebagai pilihan optimal untuk data dikotomis maupun politomus yang memenuhi asumsi unidimensionalitas. Dengan landasan mathematical proof, Rasch menghasilkan skala interval sejati yang tidak sekadar bersifat aproksimasi. Namun, syarat yang harus dipenuhi meliputi jumlah responden minimal 200, item yang terkalibrasi dengan baik, serta unidimensionalitas dengan eigenvalue ratio > 2,0. Alternatif lain yang juga masuk Tier 1 adalah algoritma Markov Chain Monte Carlo (MCMC). Pendekatan ini optimal untuk data kompleks dengan distribusi tidak diketahui atau mengandung multiple covariates. Kelebihannya terletak pada fleksibilitas yang tidak mengharuskan spesifikasi model formal yang ketat. Meski demikian, metode ini menuntut sumber daya komputasi yang memadai, jumlah iterasi minimal 1000 untuk menjamin convergence, serta keahlian khusus dalam pemodelan Bayesian.

Tier 2 (*Practical Excellence*)

Kategori ini meliputi metode yang secara praktis dapat diimplementasikan dengan hasil yang sangat baik, meskipun tidak sekuat Tier 1 dalam justifikasi matematis. Pertama, assignment of scores berbasis distribusi laten menunjukkan keunggulan ketika distribusi dasar dapat diasumsikan dengan tingkat keyakinan tertentu. Metode ini terbukti mengurangi bias hingga 45% dibandingkan dengan skoring bilangan bulat sederhana, dengan prasyarat tersedianya minimal 100 responden untuk setiap kategori. Kedua, metode successive intervals yang disertai internal consistency check mampu menghasilkan error rata-rata hanya 0,025. Robustness-nya terhadap variasi distribusi respons menjadikannya pilihan yang praktis, asalkan ukuran sampel tidak kurang dari 50 responden dan distribusi tidak ekstrem (skewness < 2,0).

Tier 3 (*Pragmatic Solutions*)

Solusi pada level ini bersifat pragmatis dan dapat diterapkan ketika keterbatasan sumber daya menjadi kendala utama. Skala Likert dengan 11 kategori dapat digunakan ketika distribusi mendekati normal dan responden familiar dengan perbedaan tingkat yang lebih halus. Dalam kondisi ini, korelasi dengan true score dapat mencapai >0,85. Selain itu, tabel konversi yang tersedia untuk instrumen standar seperti WHOQOL-BREF juga merupakan solusi praktis, dengan bukti peningkatan presisi estimasi hingga 34%. Namun, syarat utamanya adalah bahwa instrumen telah tervalidasi dan karakteristik populasi penelitian serupa dengan sampel validasi sebelumnya.

Untuk membantu peneliti memilih metode yang paling sesuai, kerangka decision tree dikembangkan dengan tiga langkah utama. Pertama, peneliti perlu menilai karakteristik data: jika data unidimensional, Rasch model sangat direkomendasikan; jika distribusi diketahui, metode assignment of scores dapat diterapkan; sedangkan untuk data kompleks atau multimodal, algoritma MCMC lebih tepat digunakan. Kedua, evaluasi sumber daya harus dilakukan: penelitian dengan keahlian tinggi dan sumber daya komputasi memadai sebaiknya menggunakan Tier 1, sementara penelitian dengan keterbatasan moderat dapat memilih Tier 2, dan penelitian dengan keterbatasan tinggi dapat mengandalkan Tier 3. Ketiga, tujuan penelitian perlu diperjelas: untuk high-stakes decision making, hanya Tier 1 yang memadai; untuk penelitian akademis yang ditujukan publikasi, minimal Tier 2 diperlukan; sedangkan untuk studi eksploratori atau uji coba awal, solusi Tier 3 masih dapat diterima.

Selain memberikan panduan seleksi, analisis ini juga menyoroti kondisi yang sebaiknya dihindari. Pertama, penjumlahan langsung skor Likert tanpa konversi tidak boleh digunakan dalam konteks pengambilan keputusan yang bersifat high-stakes. Kedua, penerapan statistik parametrik pada satu butir Likert tanpa justifikasi metodologis yang kuat sangat tidak disarankan. Ketiga,

penerapan MCMC pada sampel kecil (<100) berisiko menghasilkan estimasi yang tidak stabil. Keempat, metode successive intervals tidak dapat diandalkan jika data menunjukkan distribusi yang sangat skewed dengan nilai skewness melebihi 2,0.

Berdasarkan akumulasi temuan empiris lebih dari sembilan dekade, penelitian ini merumuskan sebuah framework komprehensif untuk meminimalisasi kesalahan sistematis dalam penggunaan dan analisis skala Likert. Framework ini disusun dalam lima domain utama: (a) perbaikan konseptual dan edukasi, (b) optimalisasi desain instrumen, (c) perbaikan analisis statistik, (d) penguatan praktik metodologis, serta (e) transparansi dalam pelaporan.

a) Perbaikan Konseptual

Kesalahan konseptual sering muncul akibat penggunaan istilah yang tidak tepat. Oleh karena itu, diperlukan kejelasan terminologi, yakni penggunaan istilah Likert scale hanya untuk instrumen multi-item yang mengukur satu konstruk unidimensional, sedangkan istilah Likert-type response format diperuntukkan bagi butir tunggal. Edukasi konseptual ini seharusnya menjadi bagian dari kurikulum metodologi penelitian, baik pada tingkat sarjana maupun pascasarjana. Selanjutnya, paradigma penelitian harus bergeser dari rule-following yang kaku menuju evidence-based decision making, yakni pengambilan keputusan analitik yang didasarkan pada bukti empiris mengenai robustness metode (Norman, 2010). Dalam hal ini, pemahaman terhadap asumsi distribusi menjadi krusial, sehingga pemilihan metode analisis tidak lagi ditentukan oleh preferensi personal atau dogma metodologis, melainkan oleh karakteristik data dan tujuan penelitian.

b) Optimalisasi Desain Instrumen

Desain instrumen yang baik merupakan fondasi bagi validitas hasil penelitian. Jumlah item ideal berkisar antara 20–40 untuk menyeimbangkan reliabilitas dan response burden. Untuk konstruk yang kompleks, disarankan penggunaan beberapa subscale dengan masing-masing 15–25 item. Dari sisi pilihan respons, perlu dijaga simetri antara kategori positif dan negatif, dengan uji awal berupa semantic differential pre-test untuk memastikan jarak psikologis yang setara antar kategori. Terkait format respons, skala 11 poin (0–10) terbukti memberikan presisi maksimal ketika responden terbiasa dengan diferensiasi halus, sementara untuk populasi umum, skala 5–7 poin tetap optimal. Selain itu, strategi randomisasi, baik pada posisi ekstrem maupun urutan item, sangat penting untuk mengurangi bias posisi dan efek primacy/recency (Simamora, 2022).

c) Perbaikan Analisis Statistik

Kesalahan umum yang kerap terjadi adalah menganalisis item secara terpisah alih-alih menggunakan composite score. Analisis per-item sebaiknya terbatas pada fungsi diagnostik atau eksplorasi awal, sementara untuk pengujian konstruk harus digunakan skor komposit. Pendekatan statistik juga perlu fleksibel, dengan penggunaan statistik parametrik untuk skor komposit dan non-parametrik sebagai robustness check, serta pelaporan keduanya untuk transparansi. Lebih lanjut, fokus analisis harus diarahkan pada practical significance melalui pelaporan effect size, bukan sekadar p-value hunting. Setiap analisis juga wajib disertai prosedur pemeriksaan asumsi yang sistematis, mencakup normalitas, homogenitas varians, maupun linearitas.

d) Penguatan Praktik Metodologis

Perencanaan ukuran sampel menjadi aspek kritis dalam penelitian berbasis skala Likert. Analisis Rasch, misalnya, menuntut minimal 200 responden, sementara analisis faktor memerlukan sekurang-kurangnya lima responden per item atau total 200 responden, mana yang lebih besar. Strategi validasi juga harus mencakup cross-validation untuk menjamin stabilitas hasil lintas sampel atau periode waktu yang berbeda. Lebih jauh, penerapan multiple method triangulation perlu didorong, yakni membandingkan hasil dari berbagai pendekatan analitik pada dataset yang sama guna meningkatkan kepercayaan terhadap temuan.

e) Transparansi dalam Pelaporan

Akhirnya, standar pelaporan transparan harus menjadi komitmen utama. Peneliti wajib menyajikan justifikasi eksplisit atas perlakuan data sebagai interval atau ordinal, serta mendokumentasikan semua keputusan analitik dan alasan di baliknya. Hasil dari pendekatan alternatif perlu dilaporkan sebagai bentuk robustness check, disertai diskusi mengenai implikasi perbedaan tersebut terhadap kesimpulan penelitian. Keterbatasan metode yang dipilih harus diakui secara eksplisit, termasuk dampaknya terhadap generalisasi hasil. Dalam semangat open science, ketersediaan data mentah untuk diverifikasi atau dianalisis ulang dengan metode alternatif juga sangat dianjurkan.

KESIMPULAN

Hasil systematic literature review komprehensif terhadap 22 publikasi akademik dari rentang waktu 1932-2022 mengungkap bahwa kontroversi skala Likert sebenarnya bukan masalah statistik murni, melainkan masalah edukasi metodologi dan implementasi solusi yang evidence-based. Lima kategori kesalahan sistematis yang diidentifikasi—konseptual-terminologi, desain instrumen, analisis statistik, metodologi, dan pelaporan—saling berkaitan dan dapat dimitigasi melalui pendekatan berlapis yang telah divalidasi secara empiris.

Temuan kritis menunjukkan bahwa evidensi empiris mendukung penggunaan statistik parametrik untuk composite scores dari true Likert scales, sebagaimana ditunjukkan oleh Norman (2010) dengan korelasi 0.99 antara Pearson dan Spearman bahkan pada data skewed. Namun, untuk mengoptimalkan precision dan validity, diperlukan implementasi solusi metodologis canggih yang disesuaikan dengan karakteristik data dan tujuan penelitian.

Hierarki solusi tiga tingkat yang dikembangkan—dari Model Rasch IRT dan MCMC Algorithm (Tier 1), Assignment of Scores dan Successive Intervals (Tier 2), hingga 11-point Likert scales dan tabel konversi (Tier 3)—memberikan framework praktis untuk researchers dengan berbagai level expertise dan resources. Decision framework yang terintegrasi memungkinkan pemilihan metode yang optimal berdasarkan data characteristics, distributional assumptions, dan research objectives.

Validasi empiris menunjukkan bahwa implementasi solusi-solusi canggih dapat meningkatkan precision hingga 34% dan mengurangi bias hingga 45% dibandingkan dengan traditional integer scoring. Hal ini memiliki implikasi signifikan untuk validity dan reliability dari research findings, particularly dalam high-stakes decision making contexts.

Peneliti perlu bergerak dari "rule-following" menuju "evidence-based decision making" dengan mempertimbangkan robustness empiris dari berbagai pendekatan analisis. Yang terpenting adalah transparansi dalam reporting, systematic assumption checking, dan justifikasi teoretis untuk setiap keputusan metodologis berdasarkan accumulated evidence selama 90+ tahun pengembangan measurement theory.

Kontribusi utama penelitian ini adalah provision of actionable guidelines yang dapat langsung diimplementasikan oleh researchers untuk meningkatkan quality dari Likert scale research. Framework komprehensif yang dikembangkan tidak hanya mengatasi kesalahan-kesalahan existing tetapi juga memberikan roadmap untuk methodological excellence dalam penggunaan skala Likert di masa depan.

DAFTAR PUSTAKA

Carifio, J., & Perla, R. (2008). Resolving the 50-year debate around using and misusing Likert scales. Medical Education, 42(12), 1150-1152. https://doi.org/10.1111/j.1365-2923.2008.03172.x

Chen, J., & Wang, L. (2014). The assignment of scores procedure for ordinal categorical data. ScientificWorldJournal. 2014;2014:304213. Epub 2014 Sep

11. PMID: 25295296; PMCID: PMC4176904. https://doi.org/10.1155/2014/304213

Edwards, A. L., & Thurstone, L. L. (1952). An internal consistency check for scale values determined by the method of successive intervals. Psychometrika, 17(2), 169-180. https://doi.org/10.1007/BF02288783

Granberg-Rademacker, J.. (2010). An Algorithm for Converting Ordinal Scale Measurement Data to Interval/Ratio Scale. Educational and Psychological Measurement - EDUC PSYCHOL MEAS. 70. 74-90. http://dx.doi.org/10.1177/0013164409344532

Harwell, M. R., & Gatti, G. G. (2001). Rescaling Ordinal Data to Interval Data in Educational Research. Review of Educational Research, 71(1), 105-131. https://doi.org/10.3102/00346543071001105

Jamieson, S. (2004). Likert scales: how to (ab)use them. Medical Education, 38(12), 1217-1218. https://doi.org/10.1111/j.1365- 2929.2004.02012.x

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. British Journal of Applied Science & Technology, 7(4), 396-403. https://doi.org/10.9734/BJAST/2015/14975

Knapp, T. R. (1990). Treating ordinal scales as interval scales: An attempt to resolve the controversy. Nursing Research, 39(2), 121-123. https://doi.org/10.1097/00006199-199003000-00019

Krägeloh, Chris & Henning, Marcus & Hawken, Susan & Zhao, Y & Shepherd, Daniel & Billington, Rex. (2011). Validation of the WHOQOL-BREF quality of life questionnaire for use with medical students. Education for health (Abingdon, England). 24. 545. http://dx.doi.org/10.4103/1357-6283.101436

Kuzon, W. M., Urbanchek, M. G., & McCabe, S. (1996). The seven deadly sins of statistical analysis. Annals of Plastic Surgery, 37(3), 265-272. https://doi.org/10.1097/00000637-199609000-00006

Lakshminarayan N. (2013). Know Your Data Before You Undertake Research. J Indian Prosthodont Soc. https://doi.org/10.1007/s13191-013-0300-8

Likert, R. (1932). A technique for the measurement of attitudes. Archives of Psychology, 22(140), 1-55.

Murray, J. (2013). Likert data: What to use, parametric or non-parametric? International Journal of Business and Social Science, 4(11), 258-264.

Norman, G. (2010). Likert scales, levels of measurement and the "laws" of statistics. Advances in Health Sciences Education : theory and practice. 15(5), 625-632. https://doi.org/10.1007/s10459-010-9222-y

Pornel, J. B., & Saldaña, G. A. (2013). Four common misuses of the Likert scale. Philippine Journal of Social Sciences and Humanities, 18(2), 12-19.

Simamora, B. (2022). Skala Likert, bias penggunaan dan jalan keluarnya. Jurnal Manajemen, 12(1), 84-93.

http://dx.doi.org/10.46806/jman.v12i1.978

Stevens, S. S. (1946). On the theory of scales of measurement. Science, 103(2684), 677-680. https://doi.org/10.1126/science.103.2684.677

Tsang, K. K. (2012). The use of midpoint on Likert scale: The implications for educational research. Hong Kong Teachers' Centre Journal, 11, 121-130.

Uebersax, J. S. (2006). Likert scales: Dispelling the confusion. Statistical Methods for Rater Agreement. http://john- uebersax.com/stat/likert.htm

Wu, H., & Leung, S. O. (2017). Can Likert scales be treated as interval scales?—A simulation study. Journal of Social Service Research, 43(4), 527-532. https://doi.org/10.1080/01488376.2017.1329775