



Properti Psikometri *Indonesia Desirable Responding Scale*

Rafli Sodiq Bagaskara^{*1}, Hafiyyan Izdiyar Iman^{2,3}, Nadim Alfian Assa^{2,3}, dan Whisnu Yudianta^{2,3}

¹Program Studi Psikologi, Fakultas Psikologi, Universitas Negeri Yogyakarta

Jl. Colombo No.1, Sleman, Daerah Istimewa Yogyakarta, Indonesia 55281

²Program Studi Psikologi, Fakultas Psikologi, Universitas Padjadjaran

Jl. Raya Bandung Sumedang KM.21, Sumedang, Jawa Barat, Indonesia 45363

³Faktor Test Publisher

Jl. Gitar No.26a, Bandung, Jawa Barat, Indonesia 40262

*E-mail: fiyyan.ok@gmail.com

Abstrak

Kepatutan sosial (*social desirability*) merupakan bias respons yang dapat mengancam validitas pengukuran *self-report*. Penelitian ini bertujuan untuk mengevaluasi properti psikometris dari *Indonesia Desirable Responding Scale* (IDRES), sebuah instrumen baru yang didasarkan pada model dua faktor Paulhus: *self-deceptive enhancement* (SDE) dan *impression management* (IM). Partisipan penelitian ini berjumlah 508 orang (M usia = 23 tahun; SD = 4.93) yang terlibat dalam asesmen daring berisiko rendah (*low stake*) maupun tinggi (*high stake*). Evaluasi psikometri dilakukan menggunakan pendekatan *classical test theory* (CTT), dan *item response theory* (IRT) melalui model Rasch untuk memperoleh bukti reliabilitas instrumen dan validitasnya. Hasil menunjukkan bahwa, berdasarkan pendekatan CTT, kedua dimensi memiliki reliabilitas yang memuaskan untuk SDE (α = .857; ω = .858) dan IM (α = .774; ω = .717). Namun demikian, analisis Rasch mengungkap bahwa dimensi IM memiliki reliabilitas *person* yang rendah (.552) dengan tingkat *endorsement* butir soal yang cenderung tinggi. Selanjutnya, pembuktian validitas internal dengan *confirmatory factor analysis* (CFA) menunjukkan bahwa model dua faktor memiliki kecocokan model (*fit*) yang jauh lebih baik (X^2 = 332, CFI = .973; TLI = .969; RMSEA = .044; SRMR = .069) dibandingkan model unidimensional (X^2 = 633; df = 170; p < .05; CFI = .922; TLI = .913; RMSEA = .073; SRMR = .102). Studi ini menyimpulkan bahwa meskipun IDRES adalah alat ukur yang reliabel dan terbukti secara struktural valid, dimensi IM memerlukan perbaikan lebih lanjut dengan penggunaan norma terpisah berdasarkan gender serta penyesuaian tingkat kesulitan butir soal untuk meminimalkan bias gender pengukuran kepatutan sosial.

Kata kunci: *differential item functioning*, IDRES, kepatutan sosial, model Rasch, psikometri.

Psychometric Property of the *Indonesia Desirable Responding Scale*

Abstract

Social desirability is a response bias that poses a threat to the validity of self-report measures. This study aims to evaluate the psychometric properties of the *Indonesia Desirable Responding Scale* (IDRES), a new instrument based on Paulhus' two-factor model: *self-deceptive enhancement* (SDE) and *impression management* (IM). A total of 508 participants (M age = 23 years; SD = 4.93) involved in both low-stakes and high-stakes online assessments recruited for this study. Psychometric evaluations were conducted using classical test theory (CTT) and item response theory (IRT) via the Rasch model to gather reliability and validity evidence. The results indicated that both dimensions demonstrated satisfactory internal consistency under the CTT approach, with SDE (α = .857; ω = .858) and IM (α = .774; ω = .717). However, Rasch analysis revealed that the IM dimension possessed low person reliability (.552), characterized by high item endorsement levels. Furthermore, validity evidence based on the internal structure by utilizing confirmatory factor analysis (CFA) demonstrated that the two-factor model achieved a significantly better fit (X^2 = 332; CFI = .973; TLI = .969; RMSEA = .044; SRMR = .069) compared to the unidimensional model (X^2 = 633; df = 170; p < .05; CFI = .922; TLI = .913; RMSEA = .073; SRMR = .102). This study concludes that while the IDRES is a reliable and structurally valid indigenous instrument, the IM dimension requires further refinement, specifically regarding the implementation of separate gender-based norms and the adjustment of item difficulty levels to minimize gender bias in measuring social desirability.

Keywords: *differential item functioning*, IDRES, psychometrics, Rasch model, social desirability.

Pendahuluan

Kepatutan sosial atau dikenal juga dengan *social desirability* telah banyak memengaruhi pengukuran yang bersifat *self-report*, baik dalam membuat klaim terlalu berlebihan (*overclaim*) maupun kurang dari situasi yang sesungguhnya (*underreport*) mengenai diri seseorang (Holtgraves, 2004). Berbagai penelitian menunjukkan bahwa individu dengan kepatutan sosial tinggi cenderung merespon positif secara berlebih pada pertanyaan yang dinilai positif oleh masyarakat, seperti dalam pelaporan perilaku hidup sehat (Carels, 2006; Klesges et al., 2004), pengukuran kepribadian dalam latar seleksi personel (Salgado, 2005), serta sikap yang positif terhadap pemilu (Silver et al., 1986). Sebaliknya, mereka cenderung merespon negatif untuk pertanyaan yang bernilai negatif di masyarakat, seperti pada kasus alkoholisme (Caputo, 2020), penyalahgunaan obat-obatan (Mensch & Kandel, 1988), maupun melaporkan kebangkrutan (Locander et al., 1976).

Konseptualisasi awal mengenai kepatutan sosial didominasi oleh pandangan unidimensional, seperti yang dihipotesiskan oleh Edwards (1957) serta Crowne & Marlowe (1960). Namun, paradigma ini mulai bergeser seiring dengan temuan empiris yang menyoroti kelemahan instrumen yang menggunakan model tunggal tersebut, salah satunya dalam hal sensitivitas butir soal dan relevansi skoring. Dengan *cutoff index* 35%, hanya 19 dari 33 butir soal yang dapat mendiskriminasi antara orang dengan skor rendah dan tinggi (Ballard et al., 1988). Selain itu, terdapat isu reliabilitas alat ukur, yaitu ketika berbagai sampel dari berbagai studi diujikan dengan *mixed-effects model* yang mempertimbangkan variabilitas antarstudi, koefisien reliabilitas untuk partisipan perempuan yang menginjak dewasa awal hanya mencapai .661, sementara untuk laki-laki hanya mencapai koefisien reliabilitas .526. Begitupun ketika diujikan secara *fixed-effects model*, rata-rata nilai koefisien reliabilitas hanya mencapai .680 (Beretvas et al., 2002). Di sisi lain, Paulhus (1984) memperkenalkan model multidimensional yang dianggap lebih komprehensif. Keunggulan model multidimensional ini selanjutnya diperkuat oleh studi-studi psikometri komparatif, seperti penelitian Nurumov et al. (2022) yang menemukan bahwa model multidimensional memiliki properti psikometri yang lebih superior ($X^2 = 249$; CFI = .943; TLI = .931; RMSEA = .035), sementara model unidimensional tidak menunjukkan nilai yang fit ($X^2 = 856$; CFI = .757; TLI = .709; RMSEA = .071). Dengan berbagai isu dimensionalitas yang muncul, Paulhus (1984) menemukan kecenderungan bahwa selama ini instrumen unidimensional kepatutan sosial hanya mampu melingkupi kecenderungan individu dalam memberikan respons positif agar dapat dinilai baik oleh orang lain atau *impression management* (IM), tetapi tidak menangkap dimensi lain yang menekankan kecenderungan menipu diri sendiri dengan pandangan positif yang melebih-lebihkan untuk menutupi kekurangan pribadi. Dimensi ini disebut sebagai *self-deceptive enhancement* (SDE) dalam *two-factor theory of social desirability* (Paulhus, 1984).

Terdapat dua dimensi kepatutan sosial yang memiliki peran berbeda dalam asesmen industri. Rosse et al. (1998) menemukan bahwa *impression management* (IM) memoderasi validitas pengukuran ciri kepribadian dalam konteks seleksi kerja. Dalam hal ini, moderasi memiliki makna bahwa IM memengaruhi hubungan antara ciri kepribadian individu yang diukur dan seberapa valid hasil pengukurannya. IM dapat memperlemah atau memperkuat validitas pengukuran tersebut. Hal ini terjadi karena individu yang berusaha "membuat kesan baik" mungkin tidak menjawab dengan jujur, melainkan berusaha menyajikan diri mereka secara lebih positif. Temuan ini menunjukkan pentingnya pengontrolan *self-presentation* dalam proses pengambilan keputusan di bidang industri dan organisasi, seperti dalam seleksi karyawan, mutasi, hingga promosi jabatan. Berry et al. (2007) menemukan bahwa SDE mampu memoderasi hubungan tipe kepribadian ekstrasversi dan performa kerja. Pada individu dengan SDE tinggi, semakin tinggi ekstrasversi tidak diiringi kenaikan performa kerja. Sebaliknya pada individu dengan SDE rendah, makin tinggi ekstrasversi, maka makin tinggi pula performa kerja yang ditunjukkan.

Terdapat berbagai instrumen yang telah dikembangkan untuk mengukur kecenderungan individu dalam memberikan jawaban yang dianggap baik secara sosial terutama dalam konteks budaya negara-negara barat, meliputi *Edward Social Desirability* (Edwards, 1957), *Marlowe-Crowne Social Desirability Scale* (MCSDS) (Crowne & Marlowe, 1960), *Self-Deception Questionnaire* dan *Other-Deception Questionnaire*, *Brief Social Desirability Scale* (BSDS), *Overclaiming Questionnaire* (Bing et al., 2011), dan *The Balanced Inventory of Desirable Responding* (BIDR) (Hart et al., 2015). Namun, sejauh mana instrumen-instrumen tersebut relevan dan akurat bila diterapkan dalam konteks budaya yang berbeda, misalnya di Indonesia, menjadi pertanyaan penting. Seiring dengan meningkatnya perhatian terhadap keunikan budaya lokal dan potensi bias dalam pengukuran lintas budaya, sejumlah studi di Indonesia mulai mengkaji dinamika *social desirability* secara lebih kontekstual, baik dari segi pengaruhnya terhadap respons individu maupun upaya pengembangan instrumen yang lebih sesuai dengan karakteristik masyarakat Indonesia.

Literatur mengenai *social desirability* di Indonesia menunjukkan bahwa topik ini telah mendapat perhatian dari berbagai domain, mulai dari politik (Aspinall et al., 2019), etika dan gender (LeComte & Bernardi, 2008), pengaruh desain instrumen terhadap ketidakjujuran respon (Setiawati et al., 2024), analisa psikometri adaptasi instrumen MCSDS (Widhiarso, 2012), hingga pengembangan instrumen lokal (Oktapialdi, et al., 2018). Namun demikian, eksplorasi yang telah dilakukan masih memiliki sejumlah keterbatasan yang membuka ruang untuk penelitian lebih lanjut. Penelitian yang ada masih terfokus pada populasi tertentu yang spesifik seperti elit politik (Aspinall et al., 2019) atau mahasiswa (LeComte & Bernardi, 2008; Widhiarso, 2012), jumlah partisipan yang belum memadai untuk merepresentasikan norma Indonesia pada penelitian mengenai instrumen hasil pengembangan maupun adaptasi (Widhiarso, 2012; Oktapialdi et al., 2018), dan validasi instrumen yang belum teruji secara luas di berbagai konteks (Oktapialdi, et al., 2018). Selain itu, strategi mitigasi bias seperti variasi format butir soal, perbedaan media administrasi (daring vs luring), serta persepsi terhadap anonimitas, juga belum banyak dieksplorasi lebih lanjut (Setiawati et al., 2024).

Keterbatasan-keterbatasan yang telah disebutkan menegaskan pentingnya pengembangan instrumen *social desirability* yang valid, reliabel, lebih adil dalam konteks budaya Indonesia dan didasarkan pada kerangka teoritis yang lebih komprehensif dan representatif terhadap dinamika psikologis responden. Instrumen semacam ini krusial, tidak hanya untuk menjamin validitas data dalam asesmen psikologis tetapi juga untuk memastikan interpretasi terhadap data *self-report* benar-benar mencerminkan sikap dan perilaku responden. Hal ini menjadi lebih mendesak mengingat kecenderungan masyarakat Indonesia yang berakar pada nilai-nilai kolektivisme dan keharmonisan sosial sehingga dapat memperkuat bias kepatutan sosial dalam pengisian instrumen psikologis (Widhiarso, 2012). Penelitian ini bertujuan untuk mengevaluasi properti psikometri dari *Indonesia Desirable Responding Scale* (IDRES) sebagai instrumen pengukuran kepatutan sosial yang sesuai untuk digunakan di Indonesia, serta membandingkannya dengan instrumen lain yang terlebih dahulu dikembangkan dan secara umum digunakan di Indonesia.

Metode

Partisipan Penelitian

Penelitian ini melibatkan 508 partisipan berusia 17–44 tahun yang tinggal di seluruh wilayah Indonesia dengan 54% partisipan adalah perempuan. Partisipan penelitian ini merupakan peserta asesmen daring Factor Indonesia yang terbagi ke dalam dua konteks asesmen, yaitu pengambilan keputusan seperti seleksi (*high-stake assessment*; 42.5%) dan konteks evaluasi kepribadian (*low-stake assessment*; 57.5%). Partisipan berasal dari jenjang pendidikan SMP berjumlah 2 orang (0.4%), jenjang SMA berjumlah 285 orang (56.1%), jenjang diploma 13 orang (2.6%), jenjang sarjana 54 orang (10.6%), dan jenjang magister 6 orang (1.2%).

Instrumen

Dalam penelitian ini dua pengukuran *social desirability* yang digunakan adalah instrumen yang baru dikembangkan, yaitu *Indonesia Desirable Responding Scale* (IDRES), serta instrumen yang telah umum digunakan, yaitu *Balanced Inventory of Desirability Responding* (BIDR-16). *Indonesia Desirable Responding Scale* (IDRES) merupakan instrumen yang dikembangkan dalam bahasa Indonesia untuk mengukur konstruk kepatutan sosial berdasarkan teori *two-factors model* (Paulhus, 1984). IDRES terdiri dari 20 butir soal. Pada setiap butir soal, peserta disajikan sebuah pernyataan, kemudian peserta diminta menilai kesesuaian pernyataan tersebut dengan dirinya dalam format pilihan jawaban “Sesuai” atau “Tidak Sesuai”. IDRES terdiri dari 14 butir soal *favorable* dan 6 butir soal *unfavorable*. Butir soal *favorable* merupakan deskripsi aktivitas yang tidak realistis bahkan tidak mungkin bagi seorang individu untuk memiliki deskripsi tersebut, misalnya “Saya selalu berkata jujur kepada semua orang”. Pada butir soal *unfavorable*, peserta disajikan pernyataan yang bernuansa negatif atau menggambarkan perilaku yang tidak diinginkan, misalnya “Saya pernah memanfaatkan orang lain untuk kepentingan saya.”

Balanced Inventory of Desirability Responding (BIDR-16) merupakan instrumen kepatutan sosial pertama yang dikembangkan oleh Paulhus (1984) yang membuktikan eksistensi *two-factor model of social desirability*. BIDR-16 terdiri dari 16 butir soal berupa pernyataan dengan format jawaban 6 skala *likert* dari “Sangat Tidak Setuju” hingga “Sangat Setuju”. Instrumen ini telah diadaptasi ke dalam Bahasa Indonesia oleh Setiawati et al. (2024). Dilaporkan bahwa reliabilitas adaptasi tes ini adalah .642 - .721. Dilaporkan juga bahwa adaptasi dari BIDR menunjukkan nilai yang tidak fit pada model dua faktor ($\chi^2/df = 8.136$; CFI = .825; TLI = .797; RMSEA = .119; SRMR = .059) dan nilai yang cukup fit pada model empat faktor ($\chi^2/df = 4.071$; CFI = .929; TLI = .912; RMSEA = .078; SRMR = .037) yang dijelaskan pada Widyastuti, et al (2024).

Teknik Analisis Data

Pada penelitian ini, pengujian properti psikometri, meliputi (1) analisis butir dan reliabilitas, mencakup di dalamnya *item endorsement*, daya pembeda dan konsistensi internal menggunakan pendekatan *classical test theory* (CTT) dan model Rasch; (2) keadilan (*fairness*) berdasarkan jenis kelamin, menggunakan *differential item functioning* (DIF) dengan metode area Raju dan (3) pengumpulan bukti validitas berdasarkan struktur internal dengan metode *principal component analysis* (PCA), *exploratory factor analysis* (EFA) dan *confirmatory factor analysis* (CFA) (4) pengumpulan bukti validitas berdasarkan hubungan dengan variabel lain menggunakan korelasi Pearson dan uji beda antara IDRES dan BIDR-16.

Evaluasi kualitas butir dan reliabilitas instrumen dilakukan dengan menggunakan dua pendekatan utama, yaitu CTT dan model Rasch. Sebelum estimasi reliabilitas, dilakukan analisis karakteristik butir melalui uji *item endorsement* untuk melihat proporsi keterpilihan jawaban dan uji daya pembeda. Daya pembeda diestimasi menggunakan korelasi butir-total (*item-total correlation*), di mana butir soal dianggap memiliki daya diskriminasi yang memuaskan jika memiliki nilai koefisien korelasi $\geq .30$ (Field, 2024). Dalam pendekatan CTT, reliabilitas konsistensi internal dihitung menggunakan *Cronbach's alpha* dan *McDonald's omega* untuk masing-masing dimensi IDRES dan keseluruhan butir soal. Nilai koefisien yang tinggi pada kedua indeks ini menunjukkan bahwa butir-butir soal dalam dimensi tersebut memiliki keselarasan yang baik dalam mengukur konstruk yang sama. Hal tersebut mengacu pada penemuan Kalkbrenner (2021), yang merupakan hasil sintesis rekomendasi dari berbagai literatur, bahwa nilai reliabilitas *Cronbach's alpha* yang dapat diterima sebagai panduan tentatif akan bukti konsistensi internal ada di rentang .70—.84 dan nilai yang kuat ada di angka $> .85$, sedangkan dalam penelitian yang sama, ditemukan juga rentang nilai yang dapat diterima untuk *McDonald's Omega* ada di rentang .65 - .80 dan nilai yang kuat ada di angka $> .80$.

Sementara itu, analisis model Rasch dilakukan untuk menilai karakteristik butir soal dan kesesuaian butir soal dengan model pengukuran Rasch. Indeks reliabilitas, seperti *person reliability* serta statistik *infit* dan *outfit mean square (MNSQ)* digunakan untuk mengevaluasi keandalan model dan kesesuaian respons peserta terhadap pola estimasi butir soal. Linacre (2002) menyebutkan nilai *infit* MNSQ dan *outfit* MNSQ pada rentang 0.5–1.5 menunjukkan butir soal yang produktif dalam pengukuran. Estimasi *difficulty index* juga dihitung untuk mengidentifikasi tingkat kesulitan relatif dari setiap butir soal dalam skala. Pendekatan model Rasch memberikan bukti tambahan mengenai kestabilan pengukuran serta kecocokan antara data dan model.

Keadilan dalam pengukuran berkaitan dengan keberadaan bias sistematis yang terdapat pada instrumen. Umumnya keadilan diinvestigasi menggunakan analisis DIF, yaitu perbedaan karakteristik respon butir soal antarkelompok. Pada penelitian ini, DIF dianalisis dalam pendekatan *item response theory* (IRT), yaitu teknik analisis perbedaan area Raju (Raju, 1988). DIF Raju mengidentifikasi bias dengan menghitung perbedaan area dari *item characteristic curve* (ICC) tiap kelompok berdasarkan model *item response theory* 1PL, 2PL, atau 3PL (Camilli & Shepard, 1994). Pada penelitian ini, DIF Raju mengkalkulasi bias berdasarkan model Rasch (setara dengan 1PL) pada kelompok. Ketika suatu butir soal tidak memiliki perbedaan area yang dihasilkan dari ICC tiap kelompok, berarti butir soal tersebut tidak memiliki DIF. *Effect size* pada teknik ini mirip dengan *effect size* dari Mantel-Haenszel, dan aturan delta ETS yang digunakan untuk mengklasifikasikan ukuran efek (Holland & Thayer, 1985). Aturan ini mengklasifikasikan *effect size* dari DIF ke dalam tiga kategori, yaitu “A” mewakili DIF yang dapat diabaikan ($\Delta_{\text{Raju}} < 1$), “B” mewakili DIF yang sedang ($1 \leq \Delta_{\text{Raju}} \leq 1,5$), dan “C” mewakili DIF yang besar ($\Delta_{\text{Raju}} \geq 1,5$).

Bukti validitas struktur internal dalam penelitian ini diawali dengan uji unidimensionalitas, dengan melakukan *principal component analysis* (PCA) pada masing-masing dimensi dengan metode penentuan faktor *parallel analysis*. Berdasarkan kriteria Reckase (1979), asumsi unidimensionalitas terpenuhi jika proporsi varians yang dijelaskan oleh komponen pertama (*variance explained by the first component*) minimal sebesar 20%. Hal ini mengindikasikan bahwa dimensi utama pengukuran dominan dibandingkan dimensi sekunder lainnya. Kemudian, dilakukan *exploratory factor analysis* (EFA) untuk seluruh butir guna mengeksplorasi pola pengelompokan butir secara alami. Kelayakan sampel untuk analisis faktor diuji menggunakan Kaiser-Meyer-Olkin (KMO) *measure of sampling adequacy* dan *Bartlett’s test of sphericity*. Nilai KMO $\geq .50$ (atau lebih ketat $\geq .60$) dan signifikansi pada uji Bartlett ($p < .05$) mengindikasikan bahwa data layak untuk difaktorkan (Kaiser, 1974; Tabachnick & Fidell, 2013). Terkait interpretasi hasil EFA, penelitian ini mengacu pada standar Tabachnick dan Fidell (2013) di mana *factor loading* sebesar .32 ditetapkan sebagai ambang batas minimum, sementara nilai di atas .55 dikategorikan baik, dan nilai di atas .71 dianggap istimewa.

Setelah eksplorasi EFA, pengujian dilanjutkan dengan analisis *confirmatory factor analysis* (CFA) untuk menguji kesesuaian model teoritis terhadap data empiris. Analisis CFA menggunakan metode estimasi *weighted least squares means and variance adjusted* (WLSMV) yang dirancang untuk data ordinal dengan kategori respon dalam jumlah kecil (Muthén, 1997), seperti skoring ordinal dikotomis pada IDRES. Metode ini didasarkan pada korelasi *polychoric* antarvariabel teramati yang menghasilkan estimasi *loading* yang lebih akurat dibandingkan metode *maximum likelihood* (ML) yang cenderung menghasilkan estimasi *loading* yang rendah pada data ordinal (Beauducel & Herzberg, 2006; Li, 2016). Terdapat dua model yang diujikan, model unidimensi yang mengasumsikan bahwa seluruh butir soal menggambarkan satu faktor umum, yaitu *social desirability*, dan model dua faktor *self-deceptive enhancement* dan *impression management* dari Paulhus (1984). Model *second-order factor* tidak diujikan dalam penelitian ini mengingat Kline (2023) menyebutkan bahwa model dua faktor dan model *second-order* merupakan model yang setara. Evaluasi model dilakukan dengan memperhatikan nilai fit seperti χ^2/df , *comparative fit index* (CFI), *Tucker-Lewis index* (TLI), *root mean square error of approximation* (RMSEA), dan *standardized root mean square residual* (SRMR). Berdasarkan Hu dan

Bentler (1999), nilai *cut-off* untuk CFI dan TLI yang dianggap baik adalah $\geq .95$, sedangkan untuk RMSEA nilai *cut-off* yang baik adalah $\leq .06$ dan untuk SRMR adalah $\leq .08$. Kline (2023) juga menyarankan nilai yang sama, di mana CFI dan TLI harus lebih besar atau sama dengan .90 untuk menunjukkan fit model yang baik, sementara RMSEA sebaiknya lebih kecil dari .08.

Bukti validitas konstruk berdasarkan hubungan dengan variabel lain dianalisis untuk mengevaluasi keterkaitan IDRES dengan instrumen lain yang sama. Dalam penelitian ini, dilakukan analisis korelasi Pearson antara dua dimensi IDRES, yaitu *self-deceptive enhancement* dan *impression management*, dengan dimensi sepadan pada BIDR-16. Analisis dilakukan melalui penyusunan tabel interkorelasi antardimensi. Korelasi yang tinggi antara dimensi yang sesuai dari IDRES dan BIDR-16 (misalnya, antara SDE-IDRES dan SDE-BIDR) diinterpretasikan sebagai bukti validitas konvergen. Sementara itu, korelasi yang lebih rendah antara dimensi yang berbeda (nilai korelasi .567 antara SDE-IDRES dan IM-BIDR, dibandingkan dengan nilai korelasi .749 antara SDE-IDRES dan SDE-BIDR) mendukung validitas diskriminan, yaitu sejauh mana konstruk yang berbeda dapat dibedakan secara empiris.

Hasil

Internal Consistency

Instrumen IDRES terdiri dari 20 butir soal yang terbagi menjadi dimensi *self-deceptive enhancement* (SDE) dan *impression management* (IM) masing-masing terdiri dari 10 butir soal. Hasil dari tingkat *endorsement*, daya pembeda dan internal konsistensi alpha (α) dan omega (ω) ditampilkan pada tabel 1. Berdasarkan kriteria dari Ebel & Frisbie (1991), dimensi SDE menunjukkan rentang tingkat *endorsement* yang lebar dengan daya beda yang baik termasuk kategori sangat baik ($> .4$). Seluruh butir soal pada dimensi IM memiliki tingkat *endorsement* di atas 50%, hal inimenunjukkan bahwa butir soal cenderung mudah disetujui. Di antara 10 butir soal, terdapat tiga butir soal IM yang berada dalam kategori yang cukup baik ($> .3$) menurut Ebel & Frisbie (1991). Butir soal “*Saya selalu terbuka dengan segala bentuk kritik dari orang lain*” memiliki tingkat *endorsement* tertinggi dengan daya pembeda yang termasuk dalam kategori cukup baik (*endorsement* = .917; daya beda = .323). Hasil ini menunjukkan butir soal pada kedua dimensi ini mampu membedakan peserta secara memadai. Konsistensi internal dari kedua dimensi ini tergolong baik dengan nilai *Cronbach's alpha* .857 untuk SDE dan .774 untuk IM, nilai *McDonald's omega* .858 untuk SDE dan .780 untuk IM. Nilai *standard error of measurement* (SEM) tercatat sebesar .119 untuk SDE dan .117 untuk IM.

Ketika ditinjau sebagai satu kesatuan skala (unidimensional) dengan total 20 butir soal ($M = 12.41$; $SD = 4.91$), performa instrumen terlihat kokoh dengan koefisien reliabilitas yang lebih tinggi, yakni *Cronbach's alpha* sebesar .874 dan *McDonald's omega* sebesar .875. Estimasi kesalahan pengukuran (SEM) pada skala total ini pun tercatat sangat rendah, yaitu sebesar .087. Lebih lanjut, butir-butir soal menunjukkan rentang tingkat *endorsement* yang luas (.281–.917), serta indeks daya pembeda yang memuaskan (.298 – .609), mengindikasikan bahwa seluruh butir soal

Tabel 1. Reliabilitas Konsistensi Internal IDRES dengan Pendekatan CTT

	<i>N</i>	<i>M</i>	<i>SD</i>	Tingkat <i>endorsement</i>	Daya pembeda	Konsistensi internal		
						α	ω	SEM
Unidimensional	20	12.41	4.91	.281–.917	.298–.609	.874	.875	.087
<i>Self-deceptive enhancement</i> (SDE)	10	5.146	3.138	.281–.717	.417–.632	.857	.858	.119
<i>Impression management</i> (IM)	10	7.266	2.463	.531–.917	.323–.607	.774	.780	.117

N = Jumlah butir soal; *M* = mean; *SD* = standard deviation; α = *Cronbach's alpha*; ω = *McDonald's omega*; SEM = *standard error of measurement* (dihitung dari nilai α)

Tabel 2. Reliabilitas Konsistensi Internal IDRES dengan Pendekatan Model Rasch

Dimensi	<i>N</i>	<i>Range difficulty index</i>	<i>Mean difficulty index</i>	<i>SD difficulty index</i>	<i>Infit</i>	<i>Outfit</i>	<i>Person reliability</i>	<i>Item reliability</i>
Unidimensional	20	-3.254– +1.332	-0.831	1.188	0.852– 1.238	0.723– 1.537	.819	.99
<i>Self-deceptive enhancement</i>	10	-1.577– 1.560	-0.132	1.124	0.866– 1.209	0,729– 1,873	.749	.99
<i>Impression management</i>	10	-3.272– 0.175	-1.518	0.956	0.833– 1.093	0,690– 1,184	.552	.98

berkontribusi positif terhadap pengukuran konstruk secara utuh. Secara keseluruhan, instrumen menunjukkan karakteristik psikometrik yang memadai berdasarkan reliabilitas internal maupun daya pembeda, tetapi catatan khusus perlu diberikan pada dimensi IM terkait sebaran tingkat *endorsement* butir soal yang cenderung mengelompok pada area persetujuan tinggi.

Model Rasch

Tabel 2 menunjukkan hasil analisis Rasch terhadap kedua dimensi IDRES. Pada tabel ini, terdapat *difficulty item* yang dalam bentuk tes ini dapat diinterpretasikan sebagai ambang batas *endorsment* atau tingkat persetujuan seseorang untuk menjawab hal tersebut, di mana nilai yang makin positif mengindikasikan bahwa butir tersebut makin sulit untuk disetujui. Hasil ini menunjukkan sebaran indeks kesulitan butir atau ambang batas persetujuan soal-soal SDE memiliki rata-rata -0.132 dan standar deviasi 1.124, sedangkan IM memiliki rata-rata -1.518 dan standar deviasi 0.956. Nilai tersebut menunjukkan bahwa butir-butir soal dalam dimensi SDE relatif lebih seimbang dalam mengukur tingkat kecenderungan kepatutan sosial, sementara butir soal pada dimensi IM cenderung lebih mudah disetujui oleh partisipan.

Lebih lanjut, nilai *infit* dan *outfit* untuk mayoritas butir soal pada kedua dimensi berada dalam batas toleransi yang umum diterima menurut Bond & Fox (2013), yaitu 0.7–1.3. Hanya terdapat masing-masing satu butir soal SDE

Tabel 3. Factor Loading Hasil Eksplorasi Faktor IDRES

Item	Faktor	
	1	2
SDE1	.532	
SDE2	.728	
SDE3	.571	
SDE4	.705	
SDE5	.449	
SDE6	.552	
SDE7	.656	
SDE8	.654	
SDE9	.483	
SDE10	.562	

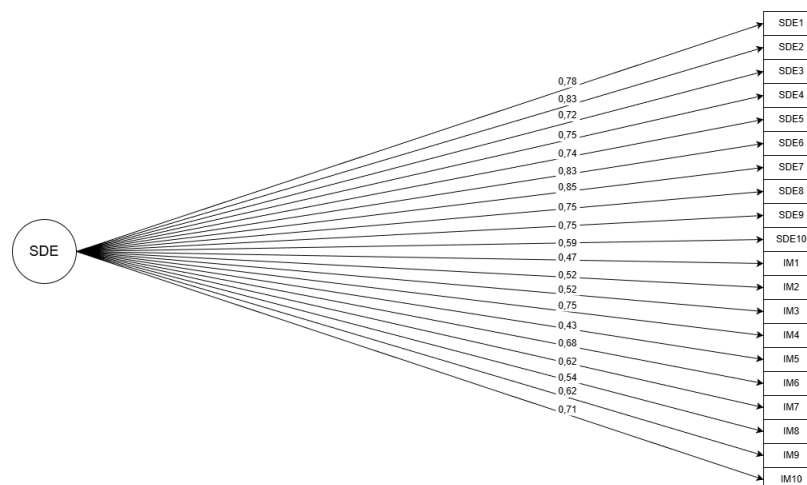
Item	Faktor	
	1	2
IM1		.387
IM2		.403
IM3		.405
IM4		.708
IM5		.497
IM6		.596
IM7		.459
IM8		.565
IM9		.517
IM10		.391

(*outfit* MNSQ = 1.873) dan IM (*outfit* MNSQ = 0.690) yang berada di luar batas tersebut yang menunjukkan bahwa butir-butir soal tersebut memiliki kesesuaian model yang cukup baik terhadap asumsi Rasch. Nilai *person reliability* pada dimensi SDE tercatat sebesar .749, hal ini mencerminkan kemampuan skala untuk membedakan individu dengan tingkat kecenderungan SDE yang berbeda secara memadai. Sementara itu, dimensi IM memiliki *person reliability* sebesar .552, yang menunjukkan keandalan sedang. Namun, hasil yang kontras ditemukan pada stabilitas butir soal, di mana nilai *item reliability* tercatat sangat tinggi, yakni .99 untuk dimensi SDE dan .98 untuk dimensi IM. Angka yang mendekati sempurna ini mengindikasikan bahwa hierarki tingkat kesulitan (*endorsability*) butir soal pada kedua dimensi telah terkalibrasi dengan sangat presisi dan stabil sehingga urutan butir soal tersebut dapat diandalkan walaupun diujikan pada sampel lain. Temuan ini memberikan indikasi bahwa meskipun keseluruhan model cukup sesuai, masih dibutuhkan perhatian dan pengembangan lebih lanjut pada dimensi IM.

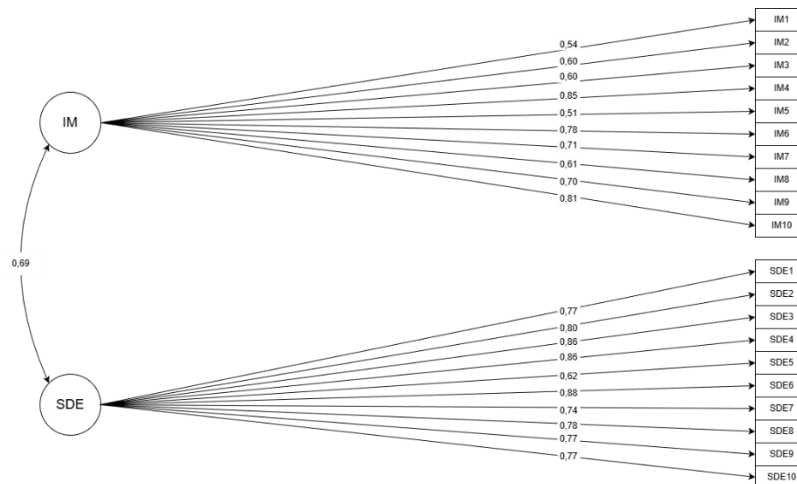
Struktur Internal IDRES

Sebelum dilakukan pengujian kesesuaian model (*model fit*), struktur dimensi diuji terlebih dahulu menggunakan *principal component analysis* (PCA) dan *exploratory factor analysis* (EFA). Analisis PCA pada masing-masing dimensi secara terpisah menunjukkan bahwa dimensi *impression management* (IM) didominasi oleh satu komponen utama yang menjelaskan varians sebesar 33.8%. Sementara itu, pada dimensi *self-deceptive enhancement* (SDE), komponen utama mampu menjelaskan varians sebesar 31.8%, dengan komponen sekunder sebesar 24.3%. Mengacu pada kriteria Reckase (1979), kedua dimensi tersebut dianggap telah memenuhi asumsi unidimensionalitas karena proporsi varians yang dijelaskan oleh komponen utama melebihi ambang batas minimal 20%. Hal ini mengindikasikan bahwa meskipun dimensi SDE memiliki keragaman yang lebih kompleks, kedua dimensi tersebut secara dominan mengukur satu atribut tunggal yang mendasarinya.

Secara keseluruhan, evaluasi terhadap asumsi menunjukkan bahwa data layak dianalisis, dibuktikan oleh nilai KMO sebesar .903 dan hasil uji *Bartlett's test of sphericity* yang signifikan ($p < .001$). Sementara itu, dari *exploratory factor analysis* (EFA) menggunakan ekstraksi *minimum residual* dengan rotasi Oblimin, dapat disimpulkan bahwa IDRES memiliki struktur dua dimensi yang cukup kokoh. Hal ini terlihat dari terbentuknya dua faktor dengan *eigenvalue* di atas 1, dan butir-butir soal yang mendukung faktornya masing-masing dengan konsistensi internal yang sebagian besar tergolong cukup baik (rentang dimensi SDE .449–.728; rentang dimensi IM .387–.708), seperti tertera pada Tabel 3.



Figur 1. Model unidimensional pada IDRES



Figur 2. Model CFA Paulhus pada IDRES

Analisis CFA dilakukan untuk mengevaluasi struktur internal dari IDRES dengan membandingkan dua model teoritis, yaitu model unidimensional dan model dua faktor sebagaimana diusulkan oleh Paulhus (1984). Model unidimensional mengasumsikan bahwa semua butir soal mengukur satu konstruk umum dari kepatutan sosial (lihat Figur 1), sedangkan model dua faktor memisahkan butir soal ke dalam dua dimensi utama (lihat Figur 2), yaitu SDE dan IM.

Hasil analisis CFA menggunakan metode estimasi WLSMV pada Tabel 4 menunjukkan bahwa model unidimensional menunjukkan nilai *fit* yang cukup baik ($\chi^2 = 633$; $df = 170$; $p < .05$; CFI = .922; TLI = .913; RMSEA = .073; SRMR = .102). Model dua faktor menunjukkan nilai *fit* yang jauh lebih baik ($\chi^2 = 332$; $df = 169$; $p < .05$; CFI = .973; TLI = .969; RMSEA = .044; SRMR = .069). Temuan ini mendukung struktur dua faktor dari IDRES yang konsisten dengan kerangka teori Paulhus, serta menunjukkan bahwa pemisahan antara SDE dan IM memberikan model pengukuran yang lebih akurat dan sesuai dengan data empiris.

Hubungan IDRES dengan BIDR-16

Analisis korelasi Pearson menunjukkan bahwa dimensi SDE pada IDRES memiliki korelasi paling tinggi dengan dimensi yang sama dari BIDR-16 ($r = 0,749$; $p < .001$), dibandingkan dengan korelasinya terhadap dimensi IM dari IDRES ($r = .532$, $p < .001$) maupun dari BIDR-16 ($r = .567$, $p < .001$). Demikian pula, dimensi IM pada IDRES menunjukkan korelasi tertinggi dengan dimensi IM pada BIDR-16 ($r = .643$, $p < .001$), lebih tinggi daripada korelasinya dengan dimensi SDE dari BIDR ($r = .606$, $p < .001$) dan IDRES ($r = .532$, $p < .001$).

Tabel 4. Interkorelasi IDRES dan BIDR-16

Dimensi	<i>M</i>	<i>SD</i>	1	2	3	4
IDRES						
1. <i>Self-deceptive enhancement</i>	5.146	3.135	(.857)			
2. <i>Impression management</i>	7.266	2.460	.532***	(.774)		
BIDR-16						
3. <i>Self-deceptive enhancement</i>	44.980	9.490	.749***	.532***	(.742)	
4. <i>Impression management</i>	40.569	9.512	.567***	.643***	.606***	(.743)

Keterangan: Nilai yang dicetak tebal dalam tanda kurung menunjukkan *Cronbach's alpha*

* $p < .05$; ** $p < .01$; *** $p < .001$

Tabel 5. *Differential item functioning* dimensi SDE

Item	<i>Raju statistic</i>	<i>p</i>	Δ_{Raju}	Kategori
sde_01	-0.997	.319	0.598	A
sde_02	0.156	.876	-0.089	A
sde_03	0.498	.618	-0.302	A
sde_04	-0.097	.923	0.056	A
sde_05	0.321	.748	-0.200	A
sde_06	-0.796	.426	0.498	A
sde_07	-2.186	.029	1.283	B
sde_08	1.680	.093	-0.971	A
sde_09	0.983	.326	-0.621	A
sde_10	0.445	.656	-0.253	A

Tabel 6. *Differential item functioning* dimensi IM

Item	<i>Raju statistic</i>	<i>p</i>	Δ_{Raju}	Kategori
im_01	-2.892	.004	1.509	C
im_02	-0.436	.663	0.332	A
im_03	-1.856	.063	0.964	A
im_04	0.617	.537	-0.368	A
im_05	-3.014	.003	1.615	C
im_06	-0.389	.698	0.271	A
im_07	3.072	.002	-1.838	C
im_08	1.967	.049	-1.919	C
im_09	0.249	.803	-0.147	A
im_10	0.765	.444	-0.419	A

Keadilan Antargender IDRES

Berdasarkan Tabel 5 dan 6, tidak ditemukan efek DIF (kategori A) pada 90% butir soal SDE dan 60% butir soal IM. Pada dimensi SDE, satu butir soal yang tersisa (10%) memiliki efek DIF moderat (kategori B). Sementara itu, empat butir soal yang tersisa pada dimensi IM (40%) memiliki efek DIF besar (kategori C) yang mengindikasikan adanya bias pada keempat butir soal tersebut. Hasil ini menunjukkan bahwa sebagian besar butir soal tidak menunjukkan adanya bias yang signifikan antara kelompok gender. Namun empat butir soal pada dimensi IM dalam kategori bias di level tinggi perlu menjadi perhatian atas prinsip keadilan.

Pembahasan

Penelitian ini bertujuan untuk mengembangkan dan mengevaluasi properti psikometris dari IDRES yang mencakup reliabilitas konsistensi internal baik menggunakan CTT dan model Rasch, pengumpulan bukti validitas, baik berdasarkan struktur internal, hubungan dengan variabel lain, maupun (*fairness*) keadilan pengukuran berdasarkan kelompok gender. Hasil penelitian ini menunjukkan IDRES telah memiliki konsistensi internal yang memenuhi kriteria dengan kemampuan yang cukup baik dalam membedakan individu pada tiap dimensi. Terdapat catatan pada dimensi IM, yaitu butir soal pada dimensi ini cenderung mudah disetujui. Temuan ini mereplikasi penelitian sebelumnya yang juga menemukan *item endorsement* dimensi IM cenderung lebih tinggi dari dimensi SDE (Tully & Bailey, 2017; Setiawati, et al., 2024).

Dimensi IM pada IDRES ditemukan memiliki konsistensi internal yang lebih rendah, bahkan dalam pendekatan model Rasch dimensi IM masih menunjukkan *person reliability* yang kurang baik. Perbedaan reliabilitas antara CTT dan model Rasch dapat dijelaskan melalui perbedaan metode dari kedua pendekatan. Rasch menggunakan rata-rata *variance error* dari skor yang didapat individu, sedangkan CTT langsung menggunakan varians dari skor peserta (Schumacker & Smith Jr, 2007). Melalui sudut pandang Rasch, butir soal pada dimensi IM cenderung memiliki indeks kesulitan atau ambang batas persetujuan yang rendah sehingga memungkinkan peserta mendapatkan skor yang lebih tinggi. Hal ini menyebabkan kurangnya informasi yang dapat diberikan butir soal pada skor peserta yang tinggi, dan berdampak pada rata-rata *variance error* yang tinggi. Dengan demikian, dapat disimpulkan bahwa butir soal pada dimensi IM dalam IDRES belum mampu membedakan secara optimal individu dengan tingkat *impression management* yang tinggi. Hal ini terjadi karena tingkat kesulitan yang rendah menyebabkan keterbatasan perolehan informasi di rentang skor tersebut. Rendahnya fungsi informasi ini secara teoritis berimplikasi pada kenaikan nilai *standard error of estimation*, yang menandakan penurunan presisi pengukuran bagi responden dengan kemampuan tinggi (Hambleton

et al., 1991). Temuan ini mengindikasikan perlunya peninjauan ulang terhadap kualitas butir pada dimensi IM, seperti menambahkan butir dengan tingkat kesulitan lebih tinggi atau yang merepresentasikan strategi IM yang lebih subtil sehingga instrumen dapat mencakup variasi kecenderungan IM secara lebih akurat dan konsisten.

Hasil analisis CFA pada IDRES menunjukkan bahwa model dua faktor Paulhus (1984) memiliki indeks fit yang lebih tinggi dibanding model unidimensi. Temuan ini konsisten dengan pandangan Paulhus yang mengkritik model unidimensional dari kepatutan sosial dan mengusulkan pemisahan antara dimensi SDE dan IM dalam model kepatutan sosial yang lebih modern. Keuntungan dari pemisahan dua faktor ini adalah memperkaya pengetahuan akan proses kepatutan sosial yang terjadi di lapangan, yaitu ketika individu mungkin memberikan gambaran diri yang positif secara berlebihan (SDE) dan atau memanipulasi respons untuk menciptakan kesan positif (IM), alih-alih hanya menangkap salah satu dimensi jika menggunakan model dari teori terdahulu (Crowne & Marlowe, 1960; Edwards, 1957). Ketika menggunakan model dua faktor dalam evaluasi kepatutan sosial individu, pengguna instrumen dapat mengetahui sumber dari bias respons dan kadarnya secara lebih akurat. Hal ini membuat interpretasi menjadi lebih akurat tentang validitas respons dan menerapkan koreksi yang lebih tepat sasaran berdasarkan jenis bias yang terdeteksi. Dengan instrumen yang memisahkan dimensi SDE dan IM, seperti yang dilakukan pada IDRES, secara umum dapat menjadi instrumen kontrol dalam suatu penelitian atau pengujian yang berbentuk *self-report* dalam berbagai konteks aplikasi. Sebagai contoh, dalam konteks seleksi kerja, IM dapat membantu mengidentifikasi kandidat yang melebih-lebihkan kualitas diri mereka untuk meninggalkan impresi yang memudahkan mereka mendapat pekerjaan (Bourdage et al., 2018). Sementara itu, penggunaan SDE di lingkungan klinis menemukan bahwa ketergantungan obat berkaitan dengan skor SDE yang tinggi (Weissman & Gorlin, 2023). Penelitian sebelumnya menemukan hasil yang berbeda. IM dapat memprediksi ketidakpuasan *actual/ideal self* (Uziel, 2014) dan kelainan kepribadian anti-sosial (Tully & Bailey, 2017), sedangkan SDE dapat memprediksi selisih *actual/ideal-self* yang lebih sempit (Uziel, 2014) dan kelainan kepribadian narsistik (Tully & Bailey, 2017). Temuan ini juga mereplikasi penelitian sebelumnya yang menunjukkan keunggulan model dua faktor dalam menjelaskan kepatutan sosial. Contohnya yaitu Paulhus (1984) yang menjadi dasar pengembangan instrumen BIDR, Gilles (2013) yang merekomendasikan penggunaan skor komposit dimensi SDE dan skor komposit dimensi IM, serta Lanyon & Carle (2007) yang menemukan SDE dan IM sebagai pengukuran yang memuaskan untuk masing-masing konstruk maupun untuk menggambarkan kepatutan sosial secara umum berdasarkan data *concurrent validity*. Selain itu, hubungan antardimensi IDRES dengan BIDR-16 pada penelitian ini juga menambah keyakinan bahwa keduanya mengukur konstruk yang sama dengan diferensiasi antardimensi yang cukup jelas.

Analisis DIF mengungkap temuan menarik dalam penelitian ini, yaitu sebagian butir soal IM menunjukkan indikasi bias. Ditemukan bahwa empat dari sepuluh butir soal IM termasuk ke dalam bias gender kategori C (bias besar) dan satu butir soal SDE bias gender kategori B (bias moderat). Hal ini sejalan dengan penelitian oleh Dalton dan Ortegren (2011), yang menemukan bahwa perempuan lebih rentan terpengaruh kepatutan sosial dalam merespon pertanyaan di dimensi IM. Banyaknya bias pada dimensi IM dapat dijelaskan melalui perspektif sosiokultural. Penelitian oleh Guadagno dan Cialdini (2007) menemukan bahwa terdapat perbedaan strategi IM yang dipengaruhi oleh ekspektasi peran gender.

Pengembangan dan evaluasi properti psikometri IDRES ini tidak terlepas dari limitasi-limitasi, seperti keterbatasan demografis sampel yang diperoleh. Lebih dari setengah responden memiliki latar belakang pendidikan SMA, dan hampir sepertiga responden tidak terdokumentasi latar belakang pendidikannya. Selain itu, terdapat keterbatasan geografis, yaitu representasi regional Indonesia belum merata. Ukuran sampel yang terbatas juga menjadi batasan jika selanjutnya hendak dilakukan penormaan untuk skala nasional. Masalah bias gender yang ditemukan menjadi salah satu keterbatasan ketika hendak melakukan interpretasi. Norma yang digunakan antarkelompok gender

menjadi tidak mungkin diseragamkan. Untuk mendapatkan validitas instrumen dengan penggunaan *mixed-gender* akan memerlukan koreksi khusus seperti penyesuaian interpretasi atau penggunaan norma terpisah (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014).

Simpulan

Hasil penelitian ini menunjukkan bahwa *Indonesia Desirable Responding Scale* (IDRES) memiliki properti psikometrik yang memadai berdasarkan pendekatan *classical test theory* dan model Rasch. Skala ini terbukti valid secara struktural dalam mengukur dua dimensi terpisah, yaitu *self-deceptive enhancement* (SDE) dan *impression management* (IM) dengan konsistensi internal yang stabil. Meskipun demikian, catatan kritis ditemukan pada dimensi IM. Hasil analisis menunjukkan tingkat *endorsement* butir soal yang cenderung tinggi pada dimensi ini, hal ini mengimplikasikan bahwa butir-butir soal tersebut relatif mudah disetujui oleh responden. Fenomena ini berpotensi memunculkan penurunan sensitivitas skala dalam membedakan individu yang memiliki tingkat *impression management* (IM) sangat tinggi. Selain itu, terdeteksinya bias respons antargender (DIF) pada dimensi yang sama memperkuat perlunya kehati-hatian dalam interpretasi skor alat ukur ini. Sebagai tindak lanjut dan implikasi praktis, penggunaan IDRES saat ini disarankan menggunakan norma terpisah berdasarkan jenis kelamin (*split norms*) untuk meminimalkan dampak bias gender. Untuk pengembangan jangka panjang, direkomendasikan agar dilakukan revisi atau penambahan butir soal pada dimensi IM dengan fokus pada peningkatan taraf serta formulasi kalimat dan konteks yang lebih netral secara gender. Dengan perbaikan tersebut, IDRES diharapkan dapat berfungsi optimal sebagai instrumen asesmen kepatutan sosial yang valid, adil, dan sensitif dalam konteks budaya Indonesia.

Daftar Pustaka

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Aspinall, Edward & Fossati, Diego & Muhtadi, Burhanuddin & Warburton, Eve. (2019). Elites, masses, and democratic decline in Indonesia. *Democratization*, 27, 1-22. 10.1080/13510347.2019.1680971.
- Ballard, R., Crino, M. D., & Rubinfeld, S. (1988). Social desirability response bias and the Marlowe-Crowne social desirability scale. *Psychological Reports*, 63(1), 227-237.
- Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, 13(2), 186-203.
- Beretvas, S. N., Meyers, J. L., & Leite, W. L. (2002). A reliability generalization study of the Marlowe-Crowne Social Desirability Scale. *Educational and Psychological Measurement*, 62(4), 570-589.
- Berry, C. M., Page, R. C., & Sackett, P. R. (2007). Effects of self-deceptive enhancement on personality-job performance relationships. *International Journal of Selection and Assessment*, 15(1), 94-109.
- Bing, M. N., Kluemper, D., Davison, H. K., Taylor, S., & Novicevic, M. (2011). Overclaiming as a measure of faking. *Organizational Behavior and Human Decision Processes*, 116(1), 148-162.
- Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.
- Bourdage, J. S., Roulin, N., & Tarraf, R. (2018). "I (might be) just that good": Honest and deceptive impression management in employment interviews. *Personnel Psychology*, 71(4), 597-632.

<https://doi.org/10.1111/peps.12285>

- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items* (Vol. 4). Sage.
- Caputo, A. (2020). The role of self-deceptive enhancement and impression management in adolescents' self-reported attitudes toward drinking, willingness to drink and risky alcohol consumption. *Journal of Substance Use*, 26(3), 268–274.
- Carels, R. A., Cacciapaglia, H. M., Rydin, S., Douglass, O. M., & Harper, J. (2006). Can social desirability interfere with success in a behavioral weight loss program? *Psychology and Health*, 21(1), 65–78.
- Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, 24(4), 349.
- Dalton, D., & Ortegren, M. (2011). Gender differences in ethics research: The importance of controlling for the social desirability response bias. *Journal of Business Ethics*, 103(1), 73–93. <https://doi.org/10.1007/s10551-011-0843-8>
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement*, Prentice-Hall. Inc. [https://doi.org/10.1016/0022-4405\(73\)90055-9](https://doi.org/10.1016/0022-4405(73)90055-9)
- Edwards, A. L. (1957). *The social desirability variable in personality assessment and research*. Dryden Press.
- Field, A. (2024). *Discovering Statistics Using IBM SPSS Statistics* (6th ed.). London: SAGE Publications.
- Guadagno, R. E., & Cialdini, R. B. (2007). Gender differences in impression management in organizations: A qualitative review. *Sex Roles*, 56, 483–494.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage Publications.
- Hart, C. M., Ritchie, T. D., Hepper, E. G., & Gebauer, J. E. (2015). The balanced inventory of desirable responding short form (BIDR-16). *Sage Open*, 5(4), 2158244015621113.
- Holland, P. W., & Thayer, D. T. (1985). An alternate definition of the ETS delta scale of item difficulty. *ETS Research Report Series*, 1985(2), i–10.
- Holtgraves, T. (2004). Social desirability and self-reports: Testing models of socially desirable responding. *Personality and Social Psychology Bulletin*, 30(2), 161–172.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36. <https://doi.org/10.1007/BF02291575>
- Kalkbrenner, Michael. (2021). Alpha, omega, and h internal consistency reliability estimates: Reviewing these options and when to use them. *Counseling Outcome Research and Evaluation*. 14. 1-12. [10.1080/21501378.2021.1940118](https://doi.org/10.1080/21501378.2021.1940118).
- Klesges, L. M., Baranowski, T., Beech, B., Cullen, K., Murray, D. M., Rochon, J., & Pratt, C. (2004). Social desirability bias in self-reported dietary, physical activity and weight concerns measures in 8- to 10-year-old African-American girls: results from the Girls Health Enrichment Multisite Studies (GEMS). *Preventive medicine*, 38 Suppl, S78–S87.
- Kline, R. B. (2023). *Principles and practice of structural equation modeling* (5th ed.). Guilford Press.
- Lanyon, R. I., & Carle, A. C. (2007). Internal and external validity of scores on the Balanced Inventory of Desirable Responding and the Paulhus Deception Scales. *Educational and Psychological Measurement*, 67(5), 859–876.
- LeComte, C. L., & Bernardi, R. A. (2008). Impressions of Questionable Marketing Practices in Indonesia: The Influence of Gender and Social Desirability Response Bias. *Electric Journal of Business Ethics and Organization Studies*.
- Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48, 936–949.
- Linacre, J. (2002) What Do Infit and Outfit, Mean-Square and Standardized Mean? *Rasch Measurement Transactions*, 16, 878.
- Locander, W., Sudman, S., & Bradburn, N. (1976). An investigation of interview method, threat and response

- distortion. *Journal of the American Statistical Association*, 71(354), 269–275.
- Mensch, B. S., & Kandel, D. B. (1988). Underreporting of substance use in a national longitudinal youth cohort: Individual and interviewer effects. *Public Opinion Quarterly*, 52(1), 100–124.
- Muthén, B. O. (1997). Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. *Psychometrika*.
- Nurumov, K., Hernández-Torrano, D., Ait Si Mhamed, A., & Ospanova, U. (2022). Measuring social desirability in collectivist countries: A psychometric study in a representative sample from kazakhstan. *Frontiers in psychology*, 13, 822931. <https://doi.org/10.3389/fpsyg.2022.822931>
- Oktapialdi, R., Tarigan, M., & Musthofa, M. A. (2018). Pengembangan skala social desirability. *Jurnal Psikologi Insight*, 2(1), 33–42.
- Paulhus, D. L. (1984). Two-component models of socially desirable responding. *Journal of Personality and Social Psychology*, 46(3), 598.
- Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika*, 53(4), 495–502.
- Reckase, Mark. (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational and Behavioral Statistics*. 4. 207-230. 10.3102/10769986004003207.
- Rosse, J. G., Stecher, M. D., Miller, J. L., & Levin, R. A. (1998). The impact of response distortion on preemployment personality testing and hiring decisions. *Journal of Applied Psychology*, 83(4), 634.
- Salgado, Jesus. (2005). Personality and social desirability in organizational settings: Practical implications for work and organizational psychology. *Papeles Del Psicólogo*. 26.
- Schumacker, R. E., & Smith Jr, E. V. (2007). A Rasch perspective. *Educational and Psychological Measurement*, 67(3), 394–409.
- Setiawati, F. A., Widyastuti, T., Fathiyah, K. N., & Nabila, T. S. (2024). Minimizing Social Desirability in Questionnaires of Non-Cognitive Measurements. *European Journal of Psychology and Educational Research*, 7(1), 33–43.
- Silver, B. D., Anderson, B. A., & Abramson, P. R. (1986). Who overreports voting? *American Political Science Review*, 80(2), 613–624.
- Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics (6th ed.). Boston, MA: Pearson.
- Tully, R. J., & Bailey, T. (2017). Validation of the Paulhus Deception Scales (PDS) in the UK and examination of the links between PDS and personality. *Journal of Criminological Research, Policy and Practice*, 3(1), 38–50.
- Uziel, L. (2014). Impression management (“lie”) scales are associated with interpersonally oriented self-control, not other-deception. *Journal of Personality*, 82(3), 200–212.
- Widhiarso W. (2012). Evaluasi Properti Psikometris Skala Kepatutan Sosial Adaptasi dari Marlowe-Crowne Social Desirability Scale. *Jurnal Insan*. 13 (3). 138-148.
- Widyastuti, T., Setiawati, F. A., Darajat, A. A., & Utomo, R. C. (2024). Selecting number of factors of the indonesian version of the Balanced Inventory of Desirable Responding Short Form (BIDR-16): is it two or four-factor?. *OSF Preprints*.
- Weissman, S., & Gorlin, E. I. (2023). Exploring the (mal)adaptive consequences of self-deceptive enhancement: A narrative review. *Journal of Theoretical Social Psychology*, 2023, Article 2382005. <https://doi.org/10.1155/2023/2382005>