Psychopolytan :Jurnal Psikologi
VOL. 5 No. 1, Agustus 2021

ISSN CETAK : 2614-5227
ISSN ONLINE : 2654-3672

Kesetaraan Skala Psikologi yang Disajikan Daring dan Luring: Kajian Literatur Deskriptif

Firmanto Adi Nurcahyo[1], Tience Debora Valentina[2]

[1,2]Program Studi Psikologi, Fakultas Kedokteran, Universitas Udayana
Jl. PB Sudirman Denpasar
[1]adinurcahyo@unud.ac.id, [2]tiencedebora@unud.ac.id

Abstrak

Pandemi Covid-19 menuntut berbagai aktivitas dijalankan secara berbeda dari biasanya. Skala psikologi yang sebelumnya disajikan secara *offline* (luring) diubah menjadi *online* (daring). Terdapat kelebihan dan kelemahan dalam penyajian skala psikologi secara daring. Salah satu kelebihannya adalah memungkinkan lebih banyak orang untuk mengisinya, sedangkan kekurangannya adalah sulitnya melakukan pengontrolan dalam kondisi terstandar. Pertanyaan yang mendasar adalah apakah kualitas skala psikologi yang disajikan secara daring setara dengan skala yang disajikan secara luring. Tulisan ini bertujuan untuk melakukan kajian literatur terhadap penelitian-penelitian yang membandingkan skala psikologi secara daring dan luring. Kajian literatur deskriptif dilakukan terhadap 10 penelitian dengan jumlah responden yang bervariasi. Hasil kajian literatur menunjukkan adanya kesetaraan skala daring dan luring, berdasarkan properti psikometrinya. Dari 10 penelitian, hanya satu penelitian yang menunjukkan adanya perbedaan struktur internal dari skala yang disajikan secara daring dan luring. Pemeriksaan terhadap properti psikometri suatu skala perlu dilakukan sebelum sebuah skala luring dapat diberikan secara daring.

Kata kunci: *skala psikologi; daring; luring*

Abstract

The Covid-19 pandemic requires various activities to be carried out differently than usual. The psychological scale that was previously presented offline was changed to online. There are advantages and disadvantages in presenting psychological scales online. One of the advantages was that it allows more people to fill it, while the disadvantage was that it was difficult to control under standardized conditions. The fundamental question is whether the quality of the psychological scale presented online was equivalent to the scale presented offline. This paper aims to conduct a literature review of studies comparing online and offline psychological scales. The descriptive literature review was conducted on 10 studies with varying numbers of respondents. The results of the literature review show that there was an equivalence of online and offline scales, based on their psychometric properties. Of the 10 studies, only one study showed differences in the internal structure of the online and offline scales. An examination of the psychometric properties of a scale needs to be conducted before an offline scale can be presented online.

Keywords: *psychological scale; online; offline*

PENDAHULUAN

Kondisi pandemi Covid-19 menuntut segala aktivitas dilakukan secara *online* atau dalam jaringan (daring), termasuk dalam penyajian skala psikologi. Penyajian skala secara daring tentunya melibatkan penggunaan komputer serta internet dalam bentuk halaman web. Aitem-aitem skala yang secara tradisional disajikan menggunakan kertas dan pensil secara *offline* atau luar jaringan (luring), diubah dalam bentuk format komputer. Responden dapat mengisi skala menggunakan perangkat lunak *browser* seperti *Mozilla's Firefox, Google's Chrome,* dan sebagainya. Setelah menjawab semua aitem skala, responden umumnya dapat mengklik tombol yang dapat menghasilkan skor dan hasilnya dapat ditampilkan melalui web (Barak, 2011).

Instrumen psikologi baik berupa tes kognitif maupun skala non kognitif dapat diwujudkan dalam bentuk daring. Tes kognitif dengan format pilihan ganda adalah jenis instrumen yang paling umum dipublikasikan melalui internet (Barak, 2011). Tes pilihan ganda dapat dinilai secara otomatis tanpa campur tangan manusia secara langsung. Bentuk instrumen lain yang

ISSN CETAK : 2614-5227
ISSN ONLINE : 2654-3672

umum diwujudkan secara daring adalah skala non kognitif yakni penilaian kepribadian dan sikap. Penggunaan format peringkat skala, misalnya *Likert*, pada skala non kognitif juga dapat diwujudkan melalui perangkat lunak komputer secara mudah dan cepat.

Terdapat berbagai kelebihan dalam penyajian skala secara daring. Skala yang disajikan secara daring memungkinkan lebih banyak orang untuk mengisinya (Buchanan, 2002). Ward et al. (2012) menunjukan bahwa kelebihan penyajian daring adalah lebih sedikit waktu yang dihabiskan untuk mengumpulkan data, pengumpulan data yang lebih bersih dan akurat, akses ke populasi yang besar dan beragam, dan peningkatan pengalaman dan anonimitas bagi responden. Selain itu, kelebihan penyajian skala dalam bentuk daring adalah fleksibilitas khususnya dalam hal waktu dan tempat. Kondisi tersebut memungkinkan responden mengisi skala ketika responden dalam kondisi nyaman (Naus et al., 2009). Kondisi responden yang nyaman dalam mengerjakan skala dapat mendukung validitas pengukuran. Kelebihan penyajian skala secara daring lainnya berkaitan dengan keakuratan penilaian skor mentah dan konversi standardisasi (Barak, 2011). Karena kedua hal tersebut dilakukan oleh perangkat lunak, maka hal tersebut dapat meningkatkan efisiensi dan menghindari potensi kesalahan manusia.

Penyajian skala secara daring juga memiliki berbagai kelemahan. Kelemahan penyajian tes secara daring terkait dengan berkurangnya kontrol dalam situasi skala (Barak, 2011). Secara umum tes dirancang untuk dilakukan secara terkontrol dengan kondisi terstandar. Penyajian tes secara daring tentunya menimbulkan kesulitan dalam melakukan pengontrolan. Kondisi ini dapat berdampak pada timbulnya masalah terkait identitas peserta. Peserta tes dapat berbuat curang atau membiarkan orang lain mengerjakan tes. Pemberian skala secara daring juga memungkinkan mendapatkan responden yang kurang bervariasi misalnya hanya memperoleh mayoritas perempuan, sehingga hasil penelitian sulit digeneralisasikan (Ward et al., 2012). Penyajian skala daring juga memungkinkan responden yang sama mengisi lebih dari satu kali (Naus et al., 2009). Pengembangan skala secara daring juga perlu memikirkan biaya pengembangan sistem daring dan biaya untuk menjaga sistem tersebut agar tetap terjaga dan mutakhir (Bartram, 2009).

Pengembangan skala daring nampaknya akan terus meningkat di tengah segala kelebihan dan kelemahannya. Tren dalam ilmu perilaku sosial bergeser ke arah penggunaan teknologi dan metode daring untuk melakukan penelitian survei (Ward et al., 2012). Hal itu tentunya memicu pengembangan skala yang semula berbentuk luring menjadi daring.

Pengembangan skala daring tentu perlu disambut dengan baik. Namun demikian, diperlukan perhatian khusus dalam pengembangan skala secara daring. Permasalahan yang seringkali timbul adalah skala daring yang dikembangkan seringkali diasumsikan memiliki kualitas yang sama dengan skala luring (Buchanan et al., 2005). Asumsi seperti ini dapat menyebabkan masalah yang dapat menjadi ancaman terhadap validitas internal (Lonsdale et al., 2006). Kondisi tersebut didukung oleh minimnya penelitian-penelitian yang menguji ekuivalensi skala daring dan luring (Naus et al., 2009). Oleh karena itu penelitian-penelitian komparasi antara skala daring dan luring perlu untuk dilakukan.

Penelitian-penelitian komparasi skala daring dan luring ditujukan untuk mengetahui apakah kualitas skala yang disajikan secara daring setara dengan skala yang disajikan secara luring. Salah satu aspek dalam menentukan kesetaraan antara penyajian secara daring dan luring adalah dalam kondisi rerata, dispersi, dan bentuk distribusi skor kedua hasil skala yang kira-kira sama (Lewis et al., 2009). Kesetaraan juga dilihat dari properti psikometri dari skala yang disajikan secara daring dan luring. Hal ini bisa diteliti dari hasil penyajian suatu skala psikologi yang disajikan dalam dua format yakni secara daring maupun secara luring. Oleh karena itu, tulisan

ISSN CETAK : 2614-5227
ISSN ONLINE : 2654-3672

ini bertujuan untuk melakukan kajian terhadap penelitian-penelitian yang membandingkan penyajian suatu skala psikologi secara daring dan luring.

METODE

Penelitian dilakukan dengan melakukan kajian literatur deskriptif. Pencarian literatur dilakukan melalui google scholar dengan kata kunci "*online test"*, "*online assessment"*, "*paper and pencil test", "traditional test",* dan "*psychology"*. Usaha pencarian tersebut menghasilkan 26 artikel yang membahas mengenai topik yang akan diteliti. Dari 26 artikel tersebut, sebanyak 10 artikel melakukan komparasi penyajian skala yang dilakukan secara daring dan luring. Oleh karena itu kajian literatur deskriptif dalam penelitian ini difokuskan terhadap 10 artikel penelitian tersebut yang dipublikasikan antara tahun 2000-2019. Judul penelitian, penulis, tahun publikasi, demografi responden, serta skala psikologi yang digunakan dalam artikel-artikel penelitian secara detail terdapat pada Tabel 1.

Tabel 1.
Daftar Penelitian Literatur Deskriptif

Judul	Penulis	Tahun Publikasi	Demografi Responden	Skala Psikologi
Personality Assessment via Internet: Comparing Online and Paper-and-Pencil Questionnaires	Hertel, G., Naumann, S., Konradt, U., & Batinic, B.	2000	136 secara daring (43% perempuan, 57% laki-laki, rerata usia 29.7) dan 112 secara luring (64% laki-laki, 36% perempuan, rerata usia 24.1)	Inventori kepribadian "*Big Five*", skala self-monitoring, skala kecemasan sosial
The Use of the Internet in Psychological Research: Comparison of Online and Offline Questionnaires	Riva, G., Teruzzi, T., & Anolli, L.	2003	203 secara daring (laki-laki 51,2%), perempuan 48,8%, usia rata-rata 23.8), 202 secara luring (laki-laki 63,4%, perempuan 36,6%, rerata usia 22.9)	Survei Penggunaan Komputer, Survei Sikap terhadap Internet
Nonequivalence of on-line and paper-and-pencil psychological tests: The case of the prospective memory questionnaire	Buchanan, T., Ali, T., Heffernan, T. M., Ling, J., Parrott, A. C., Rodgers, J., & Scholey, A. B.	2005	763(60,9% adalah perempuan, mayoritas berusia usia 21-25	*Prospective Memory Questionnaire* (PMQ)
Pixels vs. Paper:	Lonsdale, C.,	2006	117 daring, 97	*Athlete Burnout*

ISSN CETAK : 2614-5227
ISSN ONLINE : 2654-3672

Comparing Online and Traditional Survey Methods in Sport Psychology	Hodge, K., & Rose, E. A.		luring. Usia rata-rata adalah 26,53 tahun, 51,40% adalah perempuan	*Questionnaire*
From paper to pixels: A comparison of paper and computer formats in psychological assessment	Naus, M. J., Philipp, L. M., & Samsi, M.	2009	Responden secara acak ditugaskan pada dua kondisi: daring diikuti luring (N = 40) dan luring diikuti daring (N = 36). Responden semua perempuan berusia antara 18-64 tahun (M = 24,01)	*Beck Depression Inventory*, Survei Kesehatan, Inventory Lima Faktor NEO
Psychometric comparison of paper-and-pencil and online personality assessments in a selection setting	Joubert, T., & Kriek, H. J.	2009	Mahasiswa secara daring 1091 (46,93% laki-laki, 53,07% perempuan, usia rata-rata 23,14), secara luring 1136 (43,57% laki-laki, 56,43% perempuan, rerata usia 23,70) Manager secara daring 1159 (73,51% laki-laki, 26,49% perempuan, usia rata-rata 42,89 , secara luring 950 (69,68% laki-laki, 30,32% perempuan, usia rata-rata 41,67)	*Occupational Personality Questionnaire* (OPQ32i)
Paper/Pencil Versus Online Data Collection: An Exploratory Study	Ward, P., Clark, T., Zabriskie, R., & Morris, T.	2012	184 mahasiswa (46 laki-laki dan 138 perempuan) secara *repeated measure* mengisi skala yang disajikan secara daring dan luring.	*Marlowe-Crowne scale, Morally Debatable Behaviors, The Way I Feel About Myself, Leisure Satisfaction Measure, Leisure Boredom,*

ISSN CETAK : 2614-5227
ISSN ONLINE : 2654-3672

				Satisfaction With Life Scale
Examination of the Equivalence of Self-Report Survey-Based Paper-and-Pencil and Internet Data Collection Methods	Weigold, A., Weigold, I. K., & Russell, E. J.	2013	181 siswa secara daring (66% wanita 34% pria, rerata usianya 21,20), 75 secara luring (wanita 73%) dan pria 27%, rerata usia 18,67)	*The International Personality Item Pool MarloweCrowne Social Desirability Scale The Computer Self-Efficacy Scale*
Comparing Paper-and-Pencil and Internet Survey Methods Conducted in a Combat-Deployed Environment	Eckford, R. D., & Barnett, D. L.	2016	680 secara daring (84.47% pria, 15,53% wanita), 133 secara luring (85.16% pria, 14,84% wanita)	*Combat exposure, Unit climate, Posttraumatic stress disorder (PTSD) symptoms, Depression symptoms, Perceived stigma and practical barriers to care.*
Comparing the Impact of Online and Paper-and-Pencil Administration of the Self-Determination Inventory: Student Report	Raley, S. K., Shogren, K. A., Rifenbark, G. G., Anderson, M. H., & Shaw, L. A.	2020	3591 pelajar secara daring (56.8% pria, 43.2% wanita), 1150 luring (51.4% pria, 48.6 wanita). Rerata usia 16,50.	*The Self-Determination Inventory: Student Report (SDI: SR)*

HASIL

Hasil kajian literatur terhadap 10 artikel penelitian terlihat pada Tabel 2. Metode penelitian yang dominan digunakan dalam kesepuluh penelitian adalah membandingkan dua kelompok (*between group*). Terdapat 7 penelitian (Eckford & Barnett, 2016; Hertel et al., 2000; Joubert & Kriek, 2009; Lonsdale et al., 2006; Raley et al., 2020; Riva et al., 2003; Weigold et al., 2013) yang membandingkan kelompok responden yang mengerjakan skala secara daring dan kelompok lain mengerjakan skala secara luring. Terdapat 2 penelitian (Naus et al., 2009; Ward et al., 2012) yang menggunakan metode *repeated measure*, salah satunya menggunakan desain eksperimen *within subject* dengan melakukan *counterbalance*. Penelitian Buchanan et al. (2005) menggunakan satu kelompok besar responden yang mengerjakan secara daring, yang hasilnya kemudian dikomparasikan dengan penelitian dengan metode luring yang telah ada sebelumnya.

Jumlah skala yang dipergunakan dalam kesepuluh penelitian cukup bervariasi. Empat dari artikel menggunakan satu skala psikologi. Penelitian Lonsdale et al. (2006) misalnya menggunakan *Athlete Burnout Questionnaire*, sedangkan Joubert dan Kriek (2009) menggunakan *Occupational Personality Questionnaire*. Enam artikel penelitian yang lain

ISSN CETAK : 2614-5227
ISSN ONLINE : 2654-3672

menggunakan beberapa macam skala psikologi. Sebagai contoh adalah penelitian Hertel et al. (2000) yang menggunakan Inventori kepribadian "*Big Five*", skala *self-monitoring*, skala kecemasan social. Contoh lain adalah penelitian Ward et al. (2012) yang menggunakan *Marlowe-Crowne scale, Morally Debatable Behaviors, The Way I Feel About Myself, Leisure Satisfaction Measure, Leisure Boredom,* dan *Satisfaction With Life Scale*.

Komparasi dilakukan terhadap hasil skala psikologi yang disajikan secara daring dan luring. Kesetaraan diperoleh dengan membandingkan skor rerata skala yang disajikan secara daring dan luring. Kesetaraan rerata dapat dicontohkan hasil penelitian Weigold et al. (2013) yang menemukan rerata skor penyajian luring=36.45, 25.77, 37.327, 36.33, 34.61, sedangkan daring=35.96, 23.81, 36.19, 38.60, 36.40, untuk dimensi *Extraversion, Neuroticism, Openness to Experience, Agreeablenes,* dan *Conscientiousness.* Kesetaraan rerata juga dikuatkan melalui uji t seperti pada penelitian Naus et al. (2009) yang menunjukkan tidak ada perbedaan signifikan untuk BDI (t=-1.004, p=0.318) maupun survei kesehatan (t=-1.333, p=0.187) antara skala daring dan luring. Hasil uji perbedaan pada penelitian Ward et al. (2012) juga menunjukkan tidak adanya perbedaan skor yang signifikan antara penyajian daring dan luring pada *the Marlowe-Crowne scale* (t=-1.03, p>0.05), *Leisure Boredom scale* (t=-1.62, p>0.05), serta *the Satisfaction With Life Scale* (t=1.37, p>0.05).

Kesetaraan reliabilitas antara skala daring dan luring diperoleh dalam beberapa artikel penelitian. Hal ini dicontohkan pada penelitian Weigold et al. (2013) yang menemukan kesetaraan reliabilitas format luring (koefisien Alpha =0.89, 0.86, 0.81, 0.76, 0.80) dan daring (koefisien Alpha =0.84, 0.85, 0.76, 0.76) 0.81) pada dimensi *Extraversion, Neuroticism, Openness to Experience, Agreeablenes,* dan *Conscientiousness.* Hasil penelitian Joubert & Kriek (2009) menemukan reliabilitas pada kondisi luring antara 0.60-0.84 (rerata 0.73), sedangkan pada kondisi daring 0.60-0.91 (rerata 0.74). Kesetaraan koefisien reliabilitas Alpha juga ditemukan dalam penelitian Eckford dan Barnett (2016) yakni luring =0.95, 0.88, 0.94, dan daring =0.96, 0.91, 0.94 pada skala PCL, *depression*, dan stigma.

Kesetaraan struktur internal dilakukan dengan komparasi dilakukan dengan *Exploratory Factor Analysis.* Hasil penelitian Riva et al. (2003) dengan Survei Penggunaan Komputer menghasilkan 4 faktor yang menyumbang 31,69% dari total varians dalam sampel daring dan 31,92% dalam sampel luring. Sementara itu, hasil analisis faktor dalam penelitian Hertel et al. (2000) menghasilkan 5 faktor yang pada penyajian luring dapat menjelaskan 42% dari total varians, sedangkan pada kondisi daring dapat menjelaskan 62% dari total varians.

Kesetaraan struktur internal dilakukan dengan pengukuran invariansi yang dilakukan melalui *Multigrup Confirmatory Factor Analysis*. Hasil penelitian Lonsdale et al. (2006) menunjukkan tidak adanya perbedaan yang signifikan (p>0.05) pada $\chi 2$ serta CFI antara model daring dan luring, yang menunjukkan bahwa muatan faktor setara pada kedua kelompok. Hasil model fit dalam penelitian Raley et al. (2020) menunjukkan perbedaan CFI -0.001 yang berarti adanya invariansi antara kondisi daring dan luring. Joubert & Krie (2009) melakukan pemodelan persamaan struktural dalam penelitiannya dan menghasilkan dukungan bahwa matriks kovariansi daring dan luring adalah identik (CFI yang diperoleh untuk Studi 1 adalah 0,985 dengan RMSEA 0,015, sedangkan CFI yang diperoleh untuk Studi 2 adalah 0,993 dengan RMSEA 0,012).

Perbandingan penyajian skala secara daring dan luring juga menghasilkan adanya ketidaksamaan hasil. Perbedaan skala daring dan luring nampak pada penelitian Buchanan et al, (2005) yakni hasil analisis faktor menunjukkan bahwa 4 faktor ditemukan dalam format luring, sedangkan 2 faktor ditemukan dalam format daring. Selain itu, terdapat penelitian yang

ISSN CETAK : 2614-5227
ISSN ONLINE : 2654-3672

menunjukkan tidak semua rerata skor pada skala yang digunakan dalam penelitian menunjukkan kesetaraan. Hal ini dicontohkan dalam penelitian Naus et al. (2009) yang menunjukkan ada perbedaan signifikan antara format daring dan luring untuk aspek *Openness to Experience* (t=-8.667, p=0.000), *Agreeableness* (t=2.064, p=0.042), dan *Conscientiousness* (t=2.677, p=0.009). Dalam segi waktu pengerjaam, 2 penelitian menunjukkan skala daring dikerjakan lebih cepat dibandingkan secara luring. Hasil penelitian Weigold et al. (2013) menunjukkan kondisi luring membutuhkan waktu lebih lama (rerata 13.76) dibandingkan kondisi daring (rerata 9.98). Kondisi senada juga ditunjukkan dalam penelitian Eckford dan Barnett (2016) yang menunjukkan bahwa responden dalam kondisi daring menyelesaikan skala dalam waktu lebih singkat (rerata 17,20) dibandingkan dengan kondisi luring (rerata 19,84). Perbedaan skala daring dan luring juga ditemukan dalam penelitian Ward et al. (2012) yang menunjukkan bahwa semua rerata pada skor total dan subskala lebih tinggi pada luring dibandingkan pada daring. Kondisi ini mengindikasikan bahwa responden cenderung memberikan respon dengan bias sosial pada metode luring.

Tabel 2.
Hasil Kajian Literatur Deskriptif

Judul	Penulis	Hasil Kajian
Personality Assessment via Internet: Comparing Daring and Paper-and-Pencil Questionnaires	Hertel, G., Naumann, S., Konradt, U., & Batinic, B.	Reliabilitas Alpha penyajian secara daring (>0.7) setara dengan penyajian luring (>0.7). Struktur faktorial (melalui EFA) penyajian secara luring ekuivalen dengan penyajian secara daring. Hasil analisis faktor data daring dan luring menghasilkan 5 faktor, pada penyajian luring menjelaskan 42%, pada daring menjelaskan 62% varians.
The Use of the Internet in Psychological Research: Comparison of Daring and Luring Questionnaires	Riva, G., Teruzzi, T., & Anolli, L.	Tidak ada perbedaan signifikan yang ditemukan dalam struktur faktorial (EFA). Misalnya pada Survei Penggunaan Komputer pada penyajian daring dan luring menghasilkan 4 faktor yang menyumbang 31,69% dari total varians dalam sampel daring dan 31,92% dalam sampel luring. Reliabilitas Cronbach's Alpha antara daring dan luring tidak berbeda jauh misalnya reliabilitas Survei Penggunaan Komputer, sampel daring 0,75; sampel luring 0,83. Beberapa subskala daring memuat item selain yang termasuk dalam luring.
Nonequivalence of on-line and paper-and-pencil psychological skala ts: The case of the prospective memory questionnaire	Buchanan, T., Ali, T., Heffernan, T. M., Ling, J., Parrott, A. C., Rodgers, J., & Scholey, A. B.	Hasil analisis faktor menunjukkan bahwa 4 faktor ditemukan dalam format luring, 2 faktor yang ditemukan dalam format daring
Pixels vs. Paper: Comparing Daring	Lonsdale, C., Hodge, K., &	Analisis faktor konfirmatori multigroup menunjukkan bahwa tidak ada perbedaan

ISSN CETAK : 2614-5227
ISSN ONLINE : 2654-3672

and Traditional Survey Methods in Sport Psychology	Rose, E. A.	signifikan dalam struktur faktor serta struktur rerata laten dari ABQ. Tidak ada perbedaan yang signifikan ($p > 0.05$) pada $\chi 2$ atau CFI antara kedua model, yang menunjukkan bahwa muatan faktor setara pada kedua kelompok.
From paper to pixels: A comparison of paper and computer formats in psychological assessment	Naus, M. J., Philipp, L. M., & Samsi, M.	Hasil analisi uji-t sampel berpasangan menunjukkan tidak ada perbedaan signifikan untuk BDI (t=-1.004, p=0.318) maupun survei kesehatan (t=-1.333, p=0.187). Tidak ada perbedaan untuk aspek *Neuroticism* (t=-1.035, p=0.304) dan *Extraversion* dari NEO .(t=1.356, p=0.179) Ada perbedaan signifikan ditemukan antara format kertasdan format komputer untuk aspek *Openness to Experience* (t=-8.667, p=0.000), *Agreeableness* (t=2.064, p=0.042), dan *Conscientiousness* (t=2.677, p=0.009) dari NEO. Kesetaraan reliabilitas NEO serupa untuk kedua format (daring = .87, .82, .67, .70, .86, dan luring = .87, .85, .70, .76, .84, untuk N, E, O, A, dan C)
Psychometric comparison of paper-and-pencil and daring personality assessments in a selection setting	Joubert, T., & Kriek, H. J.	Hasil koefisien Alpha sangat mirip untuk daring dan luring. Reliabilitas pada kondisi luring antara 0.60-0.84 (rerata 0.73) dan pada kondisi daring 0.60-0.91 (rerata 0.74) Pemodelan persamaan struktural menunjukkan dukungan bahwa matriks kovariansi daring dan luring adalah identik. CFI yang diperoleh untuk Studi 1 adalah 0,985 dan RMSEA 0,015. CFI diperoleh untuk Studi 2 (manajer) 0,993 dan RMSEA 0,012.
Paper/Pencil Versus Daring Data Collection: An Exploratory Study	Ward, P., Clark, T., Zabriskie, R., & Morris, T.	Tidak ada perbedaan skor yang signifikan antara skala daring dan luring pada the *Marlowe-Crowne scale* (t=-1.03, p>0.05), *Leisure Boredom scale* (t=-1.62, p>0.05), serta *the Satisfaction With Life Scale* (t=1.37, p>0.05). Terdapat perbedaan pada *The Morally Debatable Behaviors scale* (t=2.12, p<0.05), *The Way I Feel About Myself scale* (t=3.70, p<0.01), dan *Leisure Satisfaction Measure* (t=3.09, p<0.01). Semua rerata pada skor total dan subskala lebih tinggi pada luring, yang mengindikasikan bahwa responden cenderung memberikan respon dengan bias sosial pada metode luring.

ISSN CETAK : 2614-5227
ISSN ONLINE : 2654-3672

Examination of the Equivalence of Self-Report Survey-Based Paper-and-Pencil and Internet Data Collection Methods	Weigold, A., Weigold, I. K., & Russell, E. J.	Kelompok daring dan luring ekuivalen dalam rerata. Kesetaraan rerata untuk kedua format (luring = 36.45, 25.77, 37.327, 36.33, 34.61, dan daring = 35.96, 23.81, 36.19, 38.60, 36.40, untuk E, N, O, A, dan C) Sebagian besar variabel menunjukkan koefisien Cronbach Alpha yang setara. Kesetaraan reliabilitas untuk kedua format (luring = 0.89, 0.86, 0.81, 0.76, 0.80, dan daring =0.84, 0.85, 0.76, 0.76, 0.81), untuk E, N, O, A, dan C) Ada perbedaan waktu penyelesaian total, kondisi luring membutuhkan waktu lebih lama (rerata 13.76) dibandingkan kondisi daring (rerata 9.98)
Comparing Paper-and-Pencil and Internet Survey Methods Conducted in a Combat-Deployed Environment	Eckford, R. D., & Barnett, D. L.	Ada kesetaraan dalam reliabilitas Cronbach Alpha pada metode daring dan luring. Kesetaraan reliabilitas Alpha untuk kedua format (luring =0.95, 0.88, 0.94, dan daring =0.96, 0.91, 0.94, untuk skala PCL, depression, stigma) Ada ekuivalensi rerata antara penyajian daring dan luring. Kesetaraan rerata untuk kedua format (luring =28.34, 1.62, 2.19, dan daring =30.08, 1.68, 2.13, untuk skala PCL, depression, stigma) Responden dalam kondisi daring menyelesaikan skala dalam waktu lebih singkat (rerata 17,20) dibandingkan dengan kondisi luring (rerata 19,84).
Comparing the Impact of Daring and Paper-and-Pencil Administration of the Self-Determination Inventory: Student Report	Raley, S. K., Shogren, K. A., Rifenbark, G. G., Anderson, M. H., & Shaw, L. A.	Hasil pengujian invariansi pengukuran mendukung penggunaan set aitem yang sama pada kelompok daring dan luring. Untuk kelompok daring CFI= 0,94 dan RMSEA=0.06, sedangkan untuk kelompok luring, CFI = 0,90, RMSEA=0.06. Hasil model fit menunjukkan perbedaan CFI -0.001 yang berarti adanya invariansi antara daring dan luring.

DISKUSI

Terdapat berbagai metode yang dipakai para peneliti untuk mengetahui kesetaraan skala yang dilakukan secara daring dan luring. Hasil penelitian menunjukkan bahwa para peneliti lebih banyak menggunakan desain *between-group designs* yakni dengan membandingkan dua kelompok (Creswell, 2012). Ini dilakukan dengan cara memberikan suatu skala psikologi pada satu kelompok secara daring, serta memberikan skala psikologi yang sama pada satu kelompok yang lain secara luring. Skor skala pada kedua kelompok kemudian dibandingkan dan dianalisis untuk menentukan sejauhmana kesetaraan yang antara skala yang disajikan secara daring dan

ISSN CETAK : 2614-5227
ISSN ONLINE : 2654-3672

luring.

Within-group design juga dapat dipakai untuk mengetahui kesetaraan skala yang dilakukan secara daring dan luring. Desain *within-group* memungkinkan peneliti untuk menggunakan satu kelompok responden saja (Creswell, 2012). Dari kesepuluh artikel, terdapat dua penelitian yang menggunakan *within-group design* yakni pada penelitian Naus et al. (2009) dan Ward et al. (2012). Secara khusus kedua penelitian tersebut menggunakan *repeated measure* yakni semua responden mengikuti semua perlakuan yang diberikan (Creswell, 2012). Dengan demikian semua responden diminta untuk mengerjakan skala secara daring, selanjutnya responden yang sama diminta untuk mengerjakan skala secara luring. Hasil pengerjaan skala dengan cara penyajian yang berbeda itu kemudian dikomparasikan.

Pedoman diperlukan dalam melakukan komparasi antara penyajian skala secara daring dan luring. *The International Test Commission* (2006) menyatakan bahwa skala daring yang dikembangkan dari skala luring yang sudah ada sebelumnya perlu memenuhi beberapa kriteria. Secara khusus kedua skala harus (1) memiliki reliabilitas sebanding, (2) berkorelasi satu sama lain, (3) berkorelasi sebanding dengan kriteria eksternal, dan (4) menghasilkan skor rerata dan deviasi standar yang sebanding atau telah dikalibrasi dengan tepat untuk menghasilkan skor sebanding.

Secara umum hasil kajian literatur menunjukkan adanya ekuivalensi skala yang disajikan secara daring dan luring. Ini dapat dilihat berdasarkan rerata skor dari skala yang disajikan secara daring dan luring. Kesetaraan rerata dan deviasi standar pada dua skala menunjukkan kedua skala tersebut memiliki sifat paralel (Azwar, 2012). Skala yang diberikan secara daring paralel dengan tes yang dilakukan secara luring jika hasil penyajian kedua tes tersebut dapat menghasilkan skor yang setara.

Properti psikometri suatu skala psikologi menjadi hal yang penting diperhatikan ketika suatu skala akan digunakan. Skala yang disajikan secara daring secara psikometri dapat diterima jika ditunjukkan secara empiris, bukan diasumsikan (Barak, 2011). Properti psikometri tersebut yang akan menentukan apakah suatu skala memiliki kualitas yang baik atau tidak. Properti psikometri yang umum digunakan dalam penyelidikan skala adalah validitas dan reliabilitas.

Validitas merupakan properti dari interpretasi dan penggunaan skor skala yang didukung oleh bukti-bukti yang tepat (Kane, 2013). Salah satu sumber bukti validitas dapat dilihat berdasarkan struktur internal *(American Educational Research Association* et al., 2014). Analisis struktur internal tes dapat menunjukkan sejauh mana hubungan antara aitem tes dan komponen tes mengkonfirmasi konstrak yang mendasarinya. Faktor analisis umumnya dipergunakan untuk mengevaluasi struktur internal suatu skala psikologi.

Hasil kajian literatur menunjukkan adanya bukti kesetaraan struktur internal skala yang disajikan secara daring dan luring. Kesetaraan tersebut diperoleh berdasarkan analisis faktor seperti dalam penelitian (Hertel et al., 2000) dan (Riva et al., 2003). Secara khusus, Ward et al. (2012) dan Raley et al. (2020) melakukan analisis konfirmatori multigrup untuk mengetahui kesetaraan skor yang dihasilkan dari penyajian secara daring dan luring. Analisis konfirmatori faktor multigrup digunakan untuk mengetahui sejauh mana kesamaan antara suatu kelompok dengan kelompok lain (Byrne, 2016). Hal ini diikuti dengan uji invariansi untuk menentukan apakah pemuatan faktor dari setiap item skala ekuivalen dalam dua kelompok yang berbeda.

Properti psikometri juga ditunjukkan dengan reliabilitas. Dalam hal ini, komparasi perlu dilakukan terhadap koefisien reliabilitas skala yang disajikan secara daring dan luring. Secara umum hasil kajian literatur menunjukkan adanya kesetaraan reliabilitas antara skala yang

ISSN CETAK : 2614-5227
ISSN ONLINE : 2654-3672

disajikan secara daring dan luring. Komparasi dapat dilihat berdasarkan reliabilitas misalnya dengan melihat perbedaan Cronbach's Alpha <+/- .10 (Weigold et al., 2013). Adanya kesetaraan reliabilitas antara penyajian skala daring dan luring tersebut mengkonfirmasi persyaratan yang diberikan *The International Test Commission* (2006).

Perbedaan struktur internal dimungkinkan terjadi dalam usaha komparasi skala daring dan luring. Hasil penelitian Riva et al. (2003) menunjukkan penyajian skala secara daring dan luring bisa setara secara struktur faktor tetapi tidak selalu identik. Dalam penelitian tersebut, beberapa subskala daring memuat item selain yang termasuk dalam luring. Perbedaan lebih nyata terlihat pada penelitian Buchanan et al. (2005) yang menunjukkan ada perbedaan jumlah faktor yang terbentuk dari hasil penyajian daring dan luring. Meski banyak variabel yang bisa mempengaruhi hasil tersebut seperti misalnya jumlah sampel serta metode yang dilakukan, hasil penelitian tersebut menggarisbawahi pentingnya validasi skala daring. Tanpa pemeriksaan properti psikometrinya, peneliti tidak dapat memastikan bahwa skala yang dilakukan secara daring benar-benar mengukur konstrak yang hendak diukur.

Dalam segi waktu pengerjaam, dua penelitian menunjukkan bahwa skala daring dikerjakan lebih cepat dibandingkan secara luring. Responden skala luring dimungkinkan lebih lama mengerjakan karena adanya distraksi dari lingkungan, sementara responden skala daring dimungkinkan lebih terpacu karena adanya batasan waktu yang jelas dalam sistem sehingga membuat responden termotivasi untuk segera menyelesaikannya (Weigold et al., 2013). Perbedaan kecepatan pengerjaan daring ini menjadi dukungan untuk melakukan tes secara daring pada masa mendatang (Eckford & Barnett, 2016).

Perbedaan penyajian skala secara daring dan luring juga terkait dengan dalam bias sosial. Responden merasakan tingkat anonimitas yang berbeda ketika mengerjakan skala secara daring dan luring. Dalam kondisi luring, responden merasa lebih memiliki resiko dalam anonimitas sehingga mereka cenderung mengikuti keinginan sosial dalam menjawab skala (Ward et al., 2012). Sebaliknya, anonimitas dalam kondisi daring membuat responden lebih jujur dalam memberikan jawaban dalam skala.

Suatu skala psikologi yang umumnya disajikan secara luring dapat disajikan secara daring. Pemeriksaan terhadap properti psikometri suatu skala perlu dilakukan sebelum sebuah skala luring dapat diberikan secara daring. Hal ini bisa dilakukan dengan melakukan komparasi antara penyajian skala secara daring dan luring. Jika hasil komparasi menunjukkan adanya kesetaraan serta properti psikometri yang kuat, maka skala dapat disajikan secara daring. Dengan demikian, penyajian skala psikologi dapat dilakukan secara daring dengan tetap memperhatikan properti psikometri dari skala tersebut.

KESIMPULAN

Hasil kajian literatur dalam tulisan ini menunjukkan bahwa terdapat kesetaraan skala psikologi yang disajikan secara daring dan luring. Kesetaraan tersebut diwujudkan dalam skor rerata, koefisien reliabilitas, serta bukti validitas struktur internal. Limitasi dari kajian literatur ini adalah pada jumlah studi yang terbatas serta metode pembandingan yang dilakukan hanya secara deksriptif. Penelitian lebih lanjut diharapkan dapat dilakukan dengan menambahkan jumlah artikel serta penggunaan metode analisis yang lebih akurat misalnya meta analisis atau *systematic literature review*.

ISSN CETAK : 2614-5227
ISSN ONLINE : 2654-3672

DAFTAR PUSTAKA

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Azwar, S. (2012). *Reliabilitas dan Validitas*. Pustaka Pelajar.

Barak, A. (2011). Internet-based Psychological Testing and Assessment. In *Online Counseling* (pp. 225–255). Elsevier. https://doi.org/10.1016/B978-0-12-378596-1.00012-5

Bartram, D. (2009). *The Advantages and Disadvantages of On-line Testing* (S. Cartwright & C. L. Cooper, Eds.; Vol. 1). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199234738.003.0011

Buchanan, T. (2002). Online assessment: Desirable or dangerous? *Professional Psychology: Research and Practice*, *33*(2), 148–154. https://doi.org/10.1037/0735-7028.33.2.148

Buchanan, T., Ali, T., Heffernan, T. M., Ling, J., Parrott, A. C., Rodgers, J., & Scholey, A. B. (2005). Nonequivalence of on-line and paper-and-pencil psychological tests: The case of the prospective memory questionnaire. *Behavior Research Methods*, *37*(1), 148–154. https://doi.org/10.3758/BF03206409

Byrne, B. M. (2016). *Structural Equation Modeling with Amos, Basic Concepts, Applications, and Programming*. Routledge.

Creswell, J. W. (2012). *Educational research. Planning, conducting, and evaluating quantitative and qualitative research* (Fourth Edition). Pearson Education, Inc.

Eckford, R. D., & Barnett, D. L. (2016). Comparing Paper-and-Pencil and Internet Survey Methods Conducted in a Combat-Deployed Environment. *Military Psychology*, *28*(4), 209–225. https://doi.org/10.1037/mil0000118

Hertel, G., Naumann, S., Konradt, U., & Batinic, B. (2000). Personality Assessment via Internet: Comparing Online and Paper-and-Pencil Questionnaires. In *Online Social Sciences* (pp. 137–154).

Joubert, T., & Kriek, H. J. (2009). Psychometric comparison of paper-and-pencil and online personality assessments in a selection setting. *SA Journal of Industrial Psychology*, *35*(1), 11 pages. https://doi.org/10.4102/sajip.v35i1.727

Kane, M. T. (2013). Validating the Interpretations and Uses of Test Scores: Validating the Interpretations and Uses of Test Scores. *Journal of Educational Measurement*, *50*(1), 1–73. https://doi.org/10.1111/jedm.12000

Lewis, I., Watson, B., & White, K. M. (2009). Internet versus paper-and-pencil survey methods in psychological experiments: Equivalence testing of participant responses to health-related messages. *Australian Journal of Psychology*, *61*(2), 107–116. https://doi.org/10.1080/00049530802105865

Lonsdale, C., Hodge, K., & Rose, E. A. (2006). Pixels vs. Paper: Comparing Online and Traditional Survey Methods in Sport Psychology. *Journal of Sport and Exercise Psychology*, *28*(1), 100–108. https://doi.org/10.1123/jsep.28.1.100

Naus, M. J., Philipp, L. M., & Samsi, M. (2009). From paper to pixels: A comparison of paper and computer formats in psychological assessment. *Computers in Human Behavior*, *25*(1), 1–7. https://doi.org/10.1016/j.chb.2008.05.012

Raley, S. K., Shogren, K. A., Rifenbark, G. G., Anderson, M. H., & Shaw, L. A. (2020). Comparing the Impact of Online and Paper-and-Pencil Administration of the Self-Determination Inventory: Student Report. *Journal of Special Education Technology*, *35*(3), 133–144. https://doi.org/10.1177/0162643419854491

Riva, G., Teruzzi, T., & Anolli, L. (2003). The Use of the Internet in Psychological Research:

ISSN CETAK : 2614-5227
ISSN ONLINE : 2654-3672

Comparison of Online and Offline Questionnaires. *CyberPsychology & Behavior*, *6*(1), 73–80. https://doi.org/10.1089/109493103321167983

The International Test Commission. (2006). International Guidelines on Computer-Based and Internet-Delivered Testing. *International Journal of Testing*, *6*(2), 143–171. https://doi.org/10.1207/s15327574ijt0602_4

Ward, P., Clark, T., Zabriskie, R., & Morris, T. (2012). Paper/Pencil Versus Online Data Collection: An Exploratory Study. *Journal of Leisure Research*, *44*(4), 507–530. https://doi.org/10.1080/00222216.2012.11950276

Weigold, A., Weigold, I. K., & Russell, E. J. (2013). Examination of the equivalence of self-report survey-based paper-and-pencil and internet data collection methods. *Psychological Methods*, *18*(1), 53–70. https://doi.org/10.1037/a0031607