

Perbandingan Properti Psikometris Skala Kesehatan Mental (SKM) antara Metode Administrasi *Paper-and-Pencil* dan *Web-Based*

Mikael Reno Prasasto¹ & Wahyu Widhiarso²

^{1,2} Fakultas Psikologi Universitas Gadjah Mada

Abstract. Advancements on information technology had helped scientific development. For example, now we could present a test using computers rather than using printed papers. This research compares the differences between the psychometric properties of a paper-and-pencil test (PPT) and web-based test (WBT). This research used *Skala Kesehatan Mental* (SKM) which was given to 544 junior and senior high-school grade school. The scale has 32 items and administered with either paper or on a website. The result showed that each method has a different but not significant psychometric properties. Compared to PPT, WBT was able to deliver fewer cases of outliers and higher item-total correlation. Reliability coefficient on both method were satisfying and able to stay unidimensional on factor analysis. Item endorsability does not appear to have significant differences. More research on this topic is needed to further examine the effect between each method on test results.

Keywords: computer testing, web-based test, skala kesehatan mental

Abstrak. Era teknologi informasi membantu memudahkan pengembangan ilmu pengetahuan, salah satunya adalah dengan membuat sebuah alat tes yang dicetak di kertas menjadi alat tes digital. Penelitian ini membandingkan perbedaan antara administrasi tes dengan *paper-and-pencil* (PPT) dan *web-based* (WBT) ditinjau dari properti psikometris tes dan butirnya. Penelitian ini dilakukan pada subjek sejumlah 544 siswa tingkat sekolah menengah pertama dan sekolah menengah atas dengan menggunakan alat tes Skala Kesehatan Mental (SKM). Tes memiliki 32 butir dan disediakan dengan metode PPT dan WBT. Hasilnya menunjukkan bahwa kedua metode menghasilkan properti psikometris yang berbeda namun relatif sama. Dibandingkan dengan PPT, WBT mampu menghasilkan nilai *outlier* yang lebih sedikit, serta korelasi butir total lebih tinggi. Koefisien reliabilitas yang dihasilkan sama-sama memuaskan dan struktur faktor kedua metode tetap unidimensi. Derajat persetujuan kedua metode tidak memiliki perbedaan yang signifikan. Kajian lebih lanjut masih diperlukan untuk meneliti dampak perbedaan metode terhadap hasil tes.

Kata kunci: tes dengan komputer, tes online, skala kesehatan mental

Saat ini kita hidup di era teknologi informasi. Data statistik dari PBB, APJII,

dan Facebook yang menunjukkan bahwa jumlah pengguna internet aktif di Indonesia pada tahun 2015 mencapai 88,1 juta pengguna atau setara dengan 34 persen populasi Indonesia (We are social, 2015). Akses internet tertinggi ada pada

¹ Korespondensi mengenai isi artikel dapat dilakukan melalui: mikael.reno.p@mail.ugm.ac.id,
² Atau melalui wahyu_psy@ugm.ac.id

pengguna *smartphone* dengan sumbangan pemuatan laman web sebesar 70 persen dari semua akses laman web. Penggunaan komputer serta internet yang semakin mudah dijangkau masyarakat luas mendukung perkembangan pengukuran psikologi dengan menggunakan komputer (Kaplan & Saccuzzo, 2005).

Administrasi tes dengan teknologi komputer maupun gawai memiliki beberapa kelebihan, antara lain penghematan biaya penelitian karena tidak memerlukan pencetakan alat ukur di kertas (Vispoel, 2000), yang juga menjadikan metode ini lebih ramah lingkungan. Minnesota Pollution Control Agency (dalam Piaw Chua, 2012) menawarkan solusi untuk mengurangi penggunaan kertas yang salah satunya adalah dengan memindahkan media kertas ke media digital. Apabila berbagai institusi pendidikan mampu mengganti tes *paper-and-pencil* dengan menggunakan komputer, konsumsi kertas global akan menurun drastis (Piaw Chua, 2012). Pengurangan konsumsi kertas ini secara tidak langsung akan mengurangi gas rumah kaca dan konsumsi energi (DeRosa dalam Piaw Chua, 2012).

Kelebihan lain dari penggunaan komputer adalah administrasi alat tes yang lebih cepat dan efisien (Hart & Goldstein, 1986), ditambah dengan minimnya kesalahan skoring dan koding karena tidak ada proses enumerasi oleh manusia. Secara lebih jauh lagi, Kaplan dan Saccuzzo (2005) menyatakan bahwa komputer memungkinkan pengacakan nomor butir secara otomatis oleh komputer untuk setiap subjek yang berbeda, serta mampu menetapkan waktu menjawab atau menghitung waktu reaksi di setiap butir.

Secara lebih khusus, administrasi alat tes secara *online* atau menggunakan *website* memiliki kelebihan tambahan yakni tidak perlu instalasi terlebih dahulu

sehingga pengukuran dapat langsung dilakukan dengan menggunakan komputer atau gawai apapun yang memiliki koneksi internet. Dengan kelebihan ini, pengambilan data dalam skala besar dapat lebih mudah dilakukan.

Selain itu, semakin banyaknya perangkat elektronik dengan berbagai ukuran layar mendorong perkembangan *responsive web design*. Istilah *responsive web design* pertama kali disebutkan oleh Ethan Marcotte, yang merujuk pada metode desain situs internet yang membuat sebuah situs mampu menyusun ulang tata letak serta ukuran objek sesuai dengan ukuran layar perangkat (Natda, 2013). Dengan cara seperti ini, sebuah situs internet akan tetap nyaman untuk dibaca dan dilihat dari berbagai ukuran layar.

Penelitian untuk penerapan administrasi alat tes psikologi melalui situs internet sudah cukup banyak dilakukan. Sayangnya, penelitian mengenai metode administrasi ini hanya berhenti di penggunaan komputer serta dilakukan sebelum *smart-phone* menjadi barang yang lebih umum daripada komputer meja. Dengan adanya *smartphone*, akses ke dunia maya tidak lagi harus melalui komputer, ditambah dengan metode input berupa sentuhan pada layar.

Penelitian terdahulu yang dilakukan oleh Preckel dan Thiemann (2003) menyimpulkan bahwa hasil tes yang reliabel dan valid dapat diperoleh dari alat tes yang diadministrasikan melalui situs internet. Thomas, Ward, Hooper, dan Hannafin (1989) menemukan tingkat kecemasan yang terlihat lebih tinggi pada kelompok subjek yang mengerjakan alat tes melalui komputer namun tidak ada perbedaan hasil tes antara kedua kelompok. Sebaliknya, penelitian lain (Hart & Goldstein, 1986; Rosenfeld, Doherty, Vicino, Kantor, & Greaves, 1989) menemukan bahwa tidak ada perbedaan

yang signifikan dalam hal kecemasan antara kedua metode.

Temuan lain dari penelitian sebelumnya adalah alat tes yang diadministrasikan dengan komputer lebih mampu dalam menciptakan keterbukaan subjek terhadap butir (Hart & Goldstein, 1986). Penelitian komparasi metode *paper-based* dan *computer-based* dengan beberapa jenis tes menemukan bahwa respons yang diterima ekuivalen satu sama lain antar metode (Rosenfeld *et al.*, 1989).

Properti psikometris dapat dipengaruhi oleh beberapa kondisi ketika tes dilakukan. Kondisi yang dapat mempengaruhi properti psikometris adalah kemungkinan munculnya kecemasan karena berhadapan dengan perangkat teknologi (Kaplan & Saccuzzo, 2005), serta validitas tampak (Nunnally, 1970). Walaupun validitas tampak tidak memiliki bukti statistik, namun validitas tampak penting untuk dipenuhi terlebih dahulu sebelum membahas kualitas tes yang lainnya (Azwar, 2016b).

Properti psikometris adalah karakteristik tes yang dapat menjelaskan atribut sebuah instrumen. Atribut ini memberikan informasi mengenai kualitas sebuah tes baik di level butir maupun di level tes (Leong, 2008). Sebuah tes dikatakan memiliki kualitas yang bagus dan bisa dipakai secara praktis apabila memiliki reliabilitas dan validitas yang tinggi (Ginty, 2013). Reliabel berarti pengulangan pada subjek yang sama akan menghasilkan skor yang sama, sementara valid berarti tes benar mengukur variabel yang ingin diukur (Azwar, 2016b).

Properti psikometris di level tes menunjukkan kualitas tes secara keseluruhan baik di level subtes maupun tes. Suatu tes dikatakan memiliki kualitas yang baik apabila mampu mengukur variabel atau dimensi yang diukur (Clark & Watson, 1995). Berdasarkan teori skor murni klasik, ada beberapa parameter

properti psikometris di level tes, yakni reliabilitas, distribusi skor, dan struktur faktor.

Azwar dalam bukunya yang berjudul "Reliabilitas dan Validitas" (2016b) menjelaskan bahwa reliabilitas adalah estimasi sejauh mana hasil proses pengukuran dapat dipercaya. Reliabilitas merupakan salah satu parameter penentu baik buruk pengukuran yang dilakukan oleh sebuah alat tes. Koefisien reliabilitas dapat ditemukan dengan metode konsistensi internal menggunakan formula *Cronbach's alpha*.

Distribusi skor memuat informasi mengenai tendensi sentral dan variabilitas skor. Terdapat tiga jenis tendensi sentral, yakni rerata, median, dan modus. Rerata biasa dipilih karena memiliki stabilitas yang lebih tinggi serta mampu mengungkapkan lebih banyak informasi (Hadi, 2004). Skor yang terlalu ekstrem dan memiliki probabilitas kemunculan yang sangat kecil bila dibandingkan dengan distribusi nilai valid disebut sebagai *outlier* (Bethlehem dalam Widhiarso & Sumintono, 2016).

Deviasi standar menunjukkan bagaimana respons atau skor menyebar. Semakin tinggi harganya, menunjukkan respons subjek yang bervariasi, atau dengan kata lain *trait* yang dimiliki sangat bervariasi (Widhiarso, 2010). Deviasi standar yang baik adalah yang mampu memfasilitasi berbagai jenis subjek dari yang terendah sampai yang tertinggi, sehingga semakin tinggi deviasi standar, kualitas tes akan semakin baik (Berk & Theall, 2006).

Setiap tes merepresentasikan sebuah konstruk psikologis yang diukur (Clark & Watson, 1995). Setiap konstruk psikologi memiliki bagian atau elemen yang disebut dengan aspek atau dimensi. Hubungan antar elemen-elemen tersebut membentuk suatu struktur yang dinamakan dengan struktur faktor (Furr & Bacharach, 2013).

Properti psikometris di level butir merujuk pada kualitas sebuah butir. Tingkat kesulitan butir merujuk kepada seberapa mudah sebuah butir dijawab oleh subjek. Dalam skala psikologi, tingkat kesulitan butir disebut sebagai derajat persetujuan (*endorsability*) individu terhadap pernyataan di dalam butir (McCreary, Conrad, Conrad, Scott, Funk, & Dennis, 2013). Persetujuan atau *endorsability* dapat ditinjau dari statistik deskriptif butir berupa rerata dan deviasi standar butir.

Daya diskriminasi butir merupakan properti psikometris butir yang mampu menunjukkan seberapa kemampuan butir dalam membedakan subjek dengan kemampuan atau *trait* tinggi dan rendah (Widhiarso, 2010). Kalkulasi untuk mengestimasi daya diskriminasi dilakukan dengan membandingkan skor butir subjek yang masuk ke dalam kelompok tinggi dan kelompok rendah.

Administrasi *paper-and-pencil* merupakan cara penyajian tes yang menggunakan instrumen yang dicetak, ditulis, atau digambar di atas kertas. Untuk mengerjakan tes seperti ini, subjek diharuskan membaca dan menulis sendiri (Chadha, 2009). Administrator tes dapat menyediakan lembar yang berisi instrumen dan kolom jawaban, kemudian meminta subjek untuk membaca dan mengisi sendiri instrumen tes tersebut. Instruksi dapat diberikan secara verbal maupun tertulis di lembar instrumen.

Administrasi *web-based* merupakan cara penyajian secara *online*, yang berarti alat tes diberikan melalui jaringan internet, misalnya menggunakan situs internet (Conole & Warburton, 2005). Penggunaan situs internet memungkinkan akses dimanapun tanpa perlu pemasangan terlebih dahulu, serta bebas dalam memilih gawai yang akan digunakan.

Penggunaan metode yang berbeda disinyalir mampu memberikan hasil yang berbeda. Kemudahan membaca dan

tampilan yang lebih menarik diperkirakan mampu mengurangi tingkat kebosanan subjek dalam mengisi skala yang diberikan sehingga motivasi dalam mengerjakan tes lebih tinggi. Motivasi yang lebih tinggi akan membuat subjek memberikan respons yang lebih jujur, yang berarti tingkat respons asal yang didapat akan lebih rendah. Selain itu terdapat juga faktor anonimitas yang lebih terasa bila menggunakan situs internet.

Ranah penelitian psikometri di Indonesia saat ini belum cukup banyak mengkaji mengenai perbandingan penggunaan alat ukur *paper-based* dengan *computer-based* maupun *web-based*. Hipotesis dari penelitian ini adalah administrasi *web-based* menghasilkan harga properti psikometris yang berbeda dibanding dengan administrasi *paper-and-pencil*.

Metode

Peneliti menggunakan data yang telah didapatkan pada penelitian sebelumnya, yakni data dari Penelitian Unggulan Perguruan Tinggi (PUPT) yang dihibahkan Direktorat Pendidikan Tinggi kepada Fakultas Psikologi Universitas Gadjah Mada dengan ketua peneliti Prof. Dr. Amitya Kumara, M.S., Psi. Judul penelitian tersebut adalah "Epidemiologi Kesehatan Mental Anak dan Remaja di Sekolah". Peneliti terlibat dalam penelitian tersebut sebagai asisten peneliti bidang teknologi informasi dan olah data.

Proses PUPT ini telah berjalan selama dua tahun, dengan populasi penelitian siswa sekolah (SD, SMP, SMA) di Daerah Istimewa Yogyakarta yang dipilih berdasarkan sekolah. Penelitian ini hanya menggunakan sampel dari jenjang Sekolah Menengah Pertama (SMP) dan Sekolah Menengah Atas (SMA). Sampel tahun pertama diambil dari Kota Yogyakarta menggunakan metode *purposive random sampling* dari tingkat

sekolah sampai ke tingkat kelas, dan mendapatkan 221 subjek. Sampel tahun kedua diambil dari Kota Yogyakarta dan Kabupaten Bantul dengan metode *multi-stage random sampling* dari tingkat kecamatan sampai ke tingkat kelas, dan mendapatkan 323 subjek. Jumlah total subjek dalam penelitian ini adalah 544 subjek. Jumlah siswa SMP dan SMA setara untuk setiap kelompok.

Tidak ada kekhasan dalam hal perkembangan subjek pada kedua lokasi. Remaja pada kedua lokasi berada pada tahapan perkembangan remaja yang normal. Hal ini dilihat dari nilai UN serta kemampuan penggunaan komputer yang tidak berbeda jauh antara kedua lokasi.

Instrumen penelitian yang diguna-

pengambilan data tahun pertama, serta 0,772 pada pengambilan data tahun kedua.

Analisis data dilakukan dengan menggunakan teori tes klasik. Perangkat lunak yang digunakan adalah *Predictive Analytics Software* (PASW) 18 yang dijalankan di sistem operasi Windows 10. Pada level butir dilakukan uji t sampel berpasangan (*paired-sample t-test*) untuk melihat apakah ada perbedaan dari hasil data yang didapat antara kedua metode pengambilan data yang dilakukan untuk setiap butir tes.

Hasil

Tabel statistik deskriptif data penelitian menunjukkan bahwa terdapat beberapa perbedaan dari skor empirik kedua

Tabel 1.
Statistik Deskriptif Data Penelitian

Metode	Jumlah subjek	Jumlah butir	Skor Empirik			
			Min	Maks	Rerata	SD
Web-based (WBT)	221	32	85	159	124,73	12,238
Paper-and-pencil (PPT)	323	32	61	150	121,50	11,572

kan merupakan alat ukur yang dikembangkan oleh tim PUPT dengan ketua peneliti Prof. Dr. Amitya Kumara, M.S., Psi., yang diberi nama Skala Kesehatan Mental (SKM). SKM menggunakan skala model likert untuk pencatatan respon, dengan lima opsi jawaban yang bergerak dari Sangat Tidak Sesuai sampai dengan Sangat Sesuai. Skoring untuk respon bergerak dari 1 sampai dengan 5. Reliabilitas instrumen dilihat mengguna-

metode (lihat tabel 1). Perbandingan statistik deskriptif antar kedua metode administrasi pengukuran ini menunjukkan adanya sedikit perbedaan. Pada skor empirik terdapat perbedaan menonjol pada skor minimal. Selain itu, tidak ada perbedaan yang terlalu jauh pada skor yang didapatkan dengan kedua metode.

Uji normalitas. Perbandingan antar jumlah kasus *outlier* dan ekstrem antar kedua metode menunjukkan lebih

Tabel 2.
Skor Outlier dan Ekstrem

Metode	Jumlah subjek	Jumlah Outlier		Jumlah Nilai Ekstrem	
		N	Persen	N	Persen
<i>Web-based</i>	221	3	1,36%	0	0,00%
<i>Paper-and-pencil</i>	323	6	1,86%	2	0,62%

kan reliabilitas *Cronbach's alpha* yang menghasilkan nilai sebesar 0,874 pada

rendahnya kasus *outlier* dan ekstrem pada metode WBT. Pada metode WBT hanya

ditemukan 3 kasus *outlier* atau 1,36% dari distribusi, sementara pada metode PPT ditemukan 6 kasus *outlier* atau 1,86% dari distribusi, dan 2 kasus ekstrem atau 0,62% dari distribusi. Maka persentase jumlah kasus skor *outlier* dan ekstrem pada metode WBT adalah 1,36% dari distribusi, sedangkan pada metode PPT sebesar 2,48% dari distribusi. Jumlah ini berbeda cukup jauh dengan perbedaan hampir dua kali lipat (lihat Tabel 2).

Setelah seluruh kasus *outlier* dibuang, ditemukan nilai minimal yang meningkat serta nilai maksimal yang menurun. Terjadi juga sedikit peningkatan pada rerata, yakni sebesar 0,19 poin untuk metode WBT dan 0,77 poin untuk metode PPT. Deviasi standar kedua metode mengalami penurunan, sebesar 0,695 poin untuk WBT dan 1,756 poin untuk PPT (lihat tabel 3). Hasil uji normalitas menunjukkan bahwa kedua data normal,

Butir dengan nilai muatan faktor yang bernilai di bawah 0,3 dapat diabaikan karena tidak dianggap signifikan (Kline, 2014). Hasilnya, terdapat 2 butir SKM yang tidak masuk ke dalam dimensi pertama, yakni butir SKM12 dan SKM18. Sementara itu, metode PPT mengeliminasi lima butir yakni butir SKM7, SKM9, SKM12, SKM13, dan SKM18, serta terdapat satu butir yang memiliki nilai terbalik atau negatif yaitu butir SKM23.

Perbandingan deskriptif skor butir. Butir SKM pada metode WBT seluruhnya memiliki skor rerata butir yang berada di atas rerata hipotetik, yang bergerak dari 3,087 s.d. 4,54. Metode PPT menghasilkan skor rerata butir yang sedikit berbeda, dengan skor yang bergerak dari 2,175 s.d. 4,510. Nilai rendah ini didapatkan oleh butir SKM23. Properti deviasi standar butir kedua metode memiliki perbedaan yang tidak terlalu jauh. Metode WBT

Tabel 3.

Statistik Deskriptif Data Penelitian Tanpa Kasus *Outlier*

Metode	Jumlah subjek	Jumlah butir	Skor Empirik			
			Min	Maks	Rerata	SD
Web-based (WBT)	218	32	91	154	124,92	11,543
Paper-and-pencil (PPT)	315	32	96	148	122,27	9,816

dengan harga $p > 0,05$ (lihat tabel 4).

Reliabilitas pengukuran. Cronbach's *alpha* alat ukur yang ditemukan dengan metode WBT adalah $r=0,874$, sementara untuk metode PPT adalah $r=0,772$. Koefisien reliabilitas yang didapatkan pada metode WBT lebih memuaskan daripada yang didapatkan dari metode PPT. Terdapat perbedaan sebesar 0,1 poin pada koefisien reliabilitas kedua metode. Walaupun begitu, dengan pendekatan yang diberikan oleh Nunnally (1970), PPT dan WBT memiliki nilai reliabilitas yang sama-sama memuaskan.

Struktur faktor. Butir dalam SKM diekstraksi ke dalam empat faktor secara terpisah untuk masing-masing metode.

Tabel 4.

Hasil Uji Normalitas Tanpa Kasus *Outlier*

Metode	Z KS	Sig.
<i>Web-based</i>	0,851	0,508
<i>Paper-and-pencil</i>	1,489	0,153

Keterangan: KS = Kolmogorov Smirnov

memiliki deviasi standar skor butir yang bergerak dari 0,560 s.d. 1,061 sementara metode PPT memiliki deviasi standar skor butir yang bergerak dari 0,614 s.d. 1,820. Korelasi butir total WBT bergerak dari 0,209 s.d. 0,612 sementara pada PPT bergerak dari -0,554 dan 0,462. Apabila menggunakan *cut-off point* 0,3 untuk korelasi butir total, maka terdapat 3 butir

pada WBT dan 7 butir pada PPT yang harus dieliminasi.

Uji t sampel berpasangan. Perbedaan pada rerata dan SD tidak signifikan, sementara perbedaan pada r-it signifikan (lihat tabel 5). Temuan ini berarti tingkat kesulitan atau derajat persetujuan butir tidak berbeda antara kedua metode, sementara daya diskriminasi butir berbeda secara nyata. Analisis tambahan dilakukan untuk melihat seberapa besar dampak penggunaan situs internet terhadap properti psikometris alat tes bila dibandingkan dengan penggunaan PPT.

dan WBT tidak mendapat dukungan dari data sehingga hipotesis ditolak.

Terdapat perbedaan pada beberapa properti psikometris antara pengambilan data menggunakan PPT dengan WBT. Pada level tes, ditemukan *outlier* yang lebih banyak pada metode PPT. Menurut Widhiarso dan Sumintono (2016) adanya nilai *outlier* maupun ekstrem disebabkan oleh subjek yang menjawab secara asal-asalan. Temuan ini menunjukkan bahwa ada kemungkinan subjek menjawab secara lebih asal-asalan apabila mengerjakan tes yang diadministrasikan dengan metode

Tabel 5.
Hasil Uji t Sampel Berpasangan

	Perbedaan Berpasangan		t	df	Sig.
	Rerata	SD			
Rerata	-0,075094	0,214055	-1,985	31	0,056
SD	0,040563	0,216464	1,060	31	0,297
r-it	-0,104750	0,157326	-3,766	31	0,001

Besaran dampak atau *effect size* merupakan istilah yang digunakan untuk menyebut sekelompok metode penghitungan besaran dampak perlakuan (Becker, 2000). Salah satu besaran dampak yang dapat digunakan adalah Cohen's *d* (Cohen, 1988). Besaran dampak yang didapatkan untuk rerata skor butir adalah ($d = -0,158$) dan ($r^2 = 0,006$). Besaran dampak untuk deviasi standar butir adalah ($d = 0,221$) dan ($r^2 = 0,012$). Sementara besaran dampak untuk korelasi butir total adalah ($d = -0,723$) dan ($r^2 = 0,116$). Sesuai dengan standar dari Cohen, metode WBT memiliki besaran dampak yang kecil terhadap rerata skor butir dan deviasi standar butir dengan perbedaan yang tidak signifikan. Sementara itu, metode WBT memiliki dampak yang besar dan signifikan terhadap korelasi butir total.

Diskusi

Secara umum hipotesis penelitian ini yang mengatakan bahwa terdapat perbedaan properti psikometris antara metode PPT

PPT ketimbang metode WBT. Distribusi dapat menjadi normal dengan mengeliminasi kasus *outlier*. Uji reliabilitas juga dilakukan pada level tes. Ditemukan nilai reliabilitas yang lebih tinggi pada metode WBT. Walaupun lebih tinggi, tidak terdapat perbedaan yang jauh antara reliabilitas skala pada metode WBT dengan metode PPT. Keduanya memiliki koefisien reliabilitas yang memuaskan dengan harga di atas 0,70. Oleh karena itu, kedua metode dianggap sama baiknya dalam menghasilkan koefisien reliabilitas yang memuaskan.

Secara jelas terlihat perbedaan properti psikometris butir antara kedua metode. Hal ini dikarenakan properti psikometris pada teori tes klasik bergantung pada sampel yang digunakan. Namun analisis dengan menggunakan uji *t* menemukan bahwa tidak semua perbedaan signifikan. Pada properti rerata dan deviasi standar butir, tidak terdapat perbedaan yang signifikan.

Koefisien korelasi butir total pada metode WBT secara garis besar lebih tinggi dibanding metode PPT. Hal ini berarti penggunaan metode WBT menghasilkan nilai daya beda butir yang lebih baik daripada metode PPT. Penyebab jelas dari perbedaan ini tentu pada sampel karena daya diskriminasi butir bergantung pada sampel.

Terdapat butir yang menemukan muatan faktor negatif, yakni butir SKM23 dengan bunyi, "saya merasa kegiatan-kegiatan di sekolah membosankan." Muatan faktor butir tersebut pada metode WBT adalah 0,365 sedangkan pada metode PPT -0,661. Koefisien korelasi butir total dari butir tersebut pada metode PPT juga menemukan nilai negatif, yakni -0,554. SKM23 merupakan butir *unfavorable*. Butir *unfavorable* berarti skor butir harus dibalik ketika penghitungan skor agar sesuai dengan konstruk yang diukur. Maka, respons yang diharapkan untuk butir ini adalah jawaban dengan nilai rendah atau ketidaksetujuan terhadap butir. Namun, nilai negatif pada muatan faktor dan korelasi butir total mengindikasikan hal yang sebaliknya. Respons yang muncul pada metode PPT cenderung ke menyetujui butir, tidak seperti pada metode WBT. Dengan kata lain, subjek pada metode PPT merasa kegiatan sekolah mereka membosankan sementara pada metode WBT tidak.

Teori tentang *aquiescence* dan *social desirability* (Widhiarso, 2010) dapat menjelaskan fenomena pada butir SKM23. Penjelasan pertama, butir tersebut berada pada nomor urut 23 dari 32 butir pada skala *paper-and-pencil*, sehingga butir tersebut berada pada urutan belakang. Siswa yang telah malas atau lelah mengisi skala pada akhirnya akan mengisi secara asal-asalan sehingga butir yang terletak paling belakang memiliki potensi paling rawan untuk munculnya *aquiescence* atau respons asal tanpa melihat pertanyaan.

Penjelasan kedua, metode WBT mampu mengacak butir secara otomatis untuk setiap pengambilan data sehingga urutan butir pada setiap subjek unik. Oleh karena itu, butir SKM23 akan memiliki peluang *aquiescence* yang lebih kecil karena tidak selalu terletak di urutan belakang. Namun, butir tersebut dapat dikatakan memiliki *social desirability* yang cukup tinggi. Siswa akan merasa takut bila menjawab hal yang tidak diinginkan oleh sekolah. Oleh karena itu, walaupun peluang *aquiescence* bisa ditekan, masih ada faktor *social desirability* sehingga subjek pada metode WBT akan cenderung menyatakan kegiatan di sekolah tidak membosankan.

Penelitian ini juga menemukan adanya 21 jawaban kosong (omit) yang diberikan oleh 15 subjek pada metode PPT. Pada metode WBT tidak terdapat jawaban kosong karena sistem mengharuskan subjek untuk mengisi seluruh butir.

Pengamatan di lapangan ketika pengambilan data berlangsung memberikan gambaran tentang perbedaan kedua metode. Pengambilan data menggunakan metode WBT hanya membutuhkan waktu sekitar 5 s.d. 10 menit untuk 32 butir. Sementara pengambilan data menggunakan metode PPT membutuhkan waktu yang relatif lebih lama, yakni 10 s.d. 15 menit. Pengambilan data menggunakan metode WBT sendiri terdapat perbedaan apabila perangkat yang digunakan berbeda. Pengambilan data menggunakan perangkat *smartphone* pribadi siswa membutuhkan waktu yang relatif lebih singkat daripada menggunakan komputer. Peneliti menganggap hal ini disebabkan karena metode input *human interface device (HID)* yang berbeda.

Kesimpulan

Berdasarkan hasil penelitian, maka dapat disimpulkan bahwa tidak terdapat perbedaan properti psikometris antara

metode PPT dan WBT. Kedua metode pada dasarnya menghasilkan properti psikometris yang berbeda namun relatif sama.

Melihat adanya beberapa perbedaan properti psikometris tes dan butir, sebuah penelitian harus memperhatikan metode yang digunakan saat mengambil data. Oleh karena itu, disarankan untuk tidak mencampur metode pengambilan data karena hasil yang didapatkan mungkin berbeda. Selain itu, uji coba alat tes sebaiknya dilakukan dengan metode yang sama dengan metode yang akan dilakukan untuk mengambil data.

Perlu digarisbawahi bahwa pengambilan data menggunakan komputer ataupun gawai lainnya harus mempertimbangkan ketersediaan sarana serta kemampuan subjek dalam mengoperasikan sarana tersebut. Tidak semua subjek dapat diberikan perlakuan yang sama. Bukan tidak mungkin hasil yang didapatkan kurang baik karena ketidakmampuan subjek mengoperasikan alat dalam mengerjakan tes.

Walaupun demikian, metode WBT secara garis besar mampu memberikan properti psikometris yang kurang lebih sama dengan metode PPT. Oleh karena itu, peneliti ingin mendorong penggunaan teknologi untuk mengurangi penggunaan kertas.

Saran

Penelitian berikutnya dapat menggunakan metode eksperimen untuk melihat secara lebih jelas dampak penggunaan administrasi WBT dibandingkan dengan metode PPT. Selain itu, parameter yang diungkap dapat ditambah lagi dengan menambah properti psikometris lain atau menggunakan teori respons butir.

Adapun hal yang perlu dikaji adalah apabila tes WBT dilakukan tanpa pengawasan. Apabila lebih banyak data telah

terkumpul mengenai perbandingan penggunaan metode WBT dengan PPT, dapat dilakukan penelitian pengambilan metode WBT yang dilakukan dengan pengawasan dan tanpa pengawasan.

Kepustakaan

- Azwar, S. (2016). *Reliabilitas dan validitas* (Edisi IV). Yogyakarta: Pustaka Pelajar.
- Becker, L. A. (2000). *Effect size (ES)*. <http://web.uccs.edu/lbecker/Psy590/es.htm>
- Berk, R. A., & Theall, M. (2006). *Thirteen strategies to measure college teaching: A consumer's guide to rating scale construction, assessment, and decision making for faculty, administrators, and clinicians*: Stylus Publishing, LLC.
- Chadha, N. K. (2009). *Applied psychometry*. New Delhi: Sage Publications India.
- Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7(3), 309-311.
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Cohen, R. J. & Swerdlik, M. E. (2009). *Psychological testing and assessment: An introduction to tests and measurement* (Edisi 7): McGraw-Hill.
- Conole, G. & Warburton, B. (2005). A review of computer-assisted assessment. *ALT-J*, 13(1), 17-31. doi: 10.1080/0968776042000339772
- Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: An introduction*. Sage.
- Ginty, A. T. (2013). Psychometric Properties. In M. D. Gellman & J. R. Turner (Eds.), *Encyclopedia of Behavioral Medicine* (pp. 1563-1564). New York, NY: Springer New York.

- Hadi, S. (2004). *Statistik* (Edisi II. Vol. 1). Yogyakarta: ANDI.
- Hart, R. R., & Goldstein, M. A. (1986). Computer-assisted psychological assessment. *Computers in Human Services*, 1(3), 69-75. doi: 10.1300/J407v01n03_05
- Kaplan, R. M., & Saccuzzo, D. P. (2005). *Psychological testing: Principles, applications, and issues* (Edisi 6). Belmont: Thomson Wadsworth.
- Leong, F. T. (2008). *Encyclopedia of counseling* (Vol. 2): Sage.
- McCreary, L. L., Conrad, K. M., Conrad, K. J., Scott, C. K., Funk, R. R., & Dennis, M. L. (2013). Using the Rasch measurement model in psychometric analysis of the family effectiveness measure. *Nursing research*, 62(3), 149-159. doi: 10.1097/NNR.0b013e31828eafe6
- Natda, K. V. (2013). Responsive web design. *Eduvantage*, 1(1).
- Nunnally, J. C. (1970). *Introduction to psychological measurement*. Tokyo: Kogakusha Co. Ltd.
- Piaw Chua, Y. (2012). Effects of computer-based testing on test performance and testing motivation. *Computers in Human Behavior*, 28(5), 1580-1586. doi: 10.1016/j.chb.2012.03.020
- Preckel, F., & Thiemann, H. (2003). Online-versus paper-pencil version of a high potential intelligence test. *Swiss Journal of Psychology*, 62(2), 131-138.
- Rosenfeld, P., Doherty, L. M., Vicino, S. M., Kantor, J., & Greaves, J. (1989). Attitude assessment in organizations: Testing three microcomputer-based survey systems. *The Journal of General Psychology*, 116(2), 145-154. doi: 10.1080/00221309.1989.9711119
- Thomas J. Ward, J., Hooper, S. R., & Hannafin, K. M. (1989). The effect of computerized tests on the performance and attitudes of college students. *Journal of Educational Computing Research*, 5(3), 327-333. doi: 10.2190/4U1D-VQRM-J70D-JEQF
- Vispoel, W. P. (2000). Computerized versus paper-and-pencil assessment of self-concept: Score comparability and respondent preferences. *Measurement and Evaluation in Counseling and Development*, 33(3), 130.
- Watson, D. (1992). Correcting for acquiescent response bias in the absence of a balanced scale. *Sociological Methods & Research*, 21(1), 52-88. doi: 10.1177/0049124192021001003
- Widhiarso, W. (2010). *Analisis butir dalam pengembangan pengukuran psikologi*. Diambil dari http://widhiarso.staff.ugm.ac.id/files/widhiarso_2001_-_analisis_butir_dalam_pengembangan_pengukuran_psikologi.pdf
- Widhiarso, W., & Sumintono, B. (2016). Examining response aberrance as a cause of outliers in statistical analysis. *Personality and Individual Differences*, 98, 11-15. doi: 10.1016/j.paid.2016.03.099
- We Are Social. (2015). *Digital, Social & mobile in Southeast Asia*. Diunduh dari <https://wearesocial.com/special-reports/digital-southeast-asia-2015>