

FITUR NON KONTEN DAN INTENSITAS KONTEN UKUR PADA BUTIR SKALA PSIKOLOGI

Wahyu Widhiarso^{1✉}, Safirah Hanifa²

Jurusan Psikologi, Fakultas Psikologi, Universitas Gadjah Mada, Indonesia

Info Artikel

Sejarah Artikel:
Disubmit 31 Agustus 2022
Direvisi 9 September 2022
Diterima 12 November 2022

Keywords:

Rasch, HEXACO, Non-content Feature, Psychological Scale

Abstrak

Selain mengukur suatu konten tertentu berdasarkan konstruk yang diukur, sebuah butir dalam skala psikologi memiliki berbagai macam atribut atau fitur yang tidak berkaitan dengan konten ukur (fitur non konten). Penelitian ini mengeksplorasi salah satu klasifikasi berdasarkan fitur non konten butir berupa ciri sifat dimanifestasikan (misalnya, aksi dan afeksi). Penelitian ini bertujuan untuk menemukan kaitan antara fitur non konten tersebut dengan intensitas atau kedalaman ukur suatu butir pada skala psikologi. Penelitian ini menggunakan data dari Proyek SAPA yang aktif melakukan penelitian terkait pengukuran kepribadian (N = 54.855). Hasil analisis menunjukkan bahwa tidak didapatkan keterkaitan antara fitur non konten (aksi, kognisi, afeksi, preferensi) dengan intensitas pengukuran yang menunjukkan probabilitas persetujuan responden terhadap pernyataan di dalam butir. Simpulan ini didukung dari hasil analisis yang dilakukan dengan menggunakan perbandingan parameter lokasi butir melalui analisis butir berbasis pemodelan Rasch. Hasil pengujian statistika menunjukkan tidak terdapat perbedaan dalam hal parameter lokasi butir dengan fitur non konten butir.

Abstract

Instead of assessing a specific content based on the construct being measured, item of a psychological scale has various attributes or features that are not related to the measured content (non-content related features). This study explores one of the classifications based on non-item content features in the form of how traits are manifested (e.g., action and affection). This study finds the relationship between these features and the intensity or depth of measurement of item on a psychological scale. This study uses data from the SAPA Project which is actively conducting research related to personality measurement across countries (N = 54,855). Analysis results show that there is no relationship between four item non-content features (action, cognition, affection, preference) and the item intensity that indicates the probability of respondents' agreement with the statements of item. This conclusion is supported by the finding the analysis carried out using a comparison of item location parameters obtained from modern item analysis based on Rasch modeling. The results of statistical tests show that there is no difference in terms of item location parameters with non-item content features.

© 2022 Universitas Negeri Semarang

✉ Alamat korespondensi:
Jurusan Psikologi, Fakultas Psikologi, Universitas Gadjah Mada,
Karang Malang, Caturtunggal, Depok, Sleman, Daerah Istimewa
Yogyakarta, Indonesia
wahyu_psy@ugm.ac.id

PENDAHULUAN

Selain mengukur atribut ukur sesuai dengan indikator yang diukur, suatu butir pada suatu skala psikologi memuat suatu konteks yang tidak berkaitan dengan tujuan utama penulisan butir (Bradburn, 2004). Misalnya, suatu butir pernyataan dapat memuat pelaporan perilaku (“*Saya menarik diri ...*”) ataukah penilaian mengenai perasaan individu (“*Saya merasa senang ketika...*”). Keberadaan konteks kemungkinan dapat menyebabkan adanya bias butir karena munculnya efek metode (Widhiarso, 2016), namun juga dapat memberikan manfaat terkait jangkauan ukur suatu butir (Ones & Viswesvaran, 1996). Pada argumen ini, pelibatan konteks dapat meningkatkan validitas konten karena butir-butir tersebut dapat langsung mengukur pada situasi yang diinginkan, misalnya lebih spesifik ataukah lebih umum. Misalnya, melibatkan kata “*di tempat kerja*” pada butir “*Saya bekerja keras*” dapat meningkatkan validitas konten butir karena langsung mengukur pada situasi yang spesifik (Bing, Whanger, & Vanhook, 2004; Schmit et al., 1995). Dua contoh di atas menunjukkan bahwa suatu butir memiliki suatu fitur yang tidak terkait dengan konten ukur. Pada tulisan ini, karakteristik tersebut dinamakan dengan fitur non konten butir.

Literatur banyak merekomendasikan bahwa suatu alat ukur harus memiliki butir dengan derajat ukur yang berbeda-beda (Drasgow et al., 2010). Intensitas butir ini dapat berkaitan dengan konteks atau fitur non konten dari suatu butir (Li et al., 2012). Pandangan lain juga menjelaskan bahwa pengembangan skala perlu mempertimbangkan adanya variasi tingkatan intensitas ini pada butir-butir di dalam skala karena banyaknya variasi akan meningkatkan presisi skor yang dihasilkan (Wilson, 2005). Salah satu contohnya adalah indikator pada skala depresi, indikator “*pemikiran bunuh diri*” memiliki intensitas atau tingkatan depresi lebih tinggi daripada indikator “*gangguan tidur*”

dalam memprediksi depresi. Metode pengembangan skala Thurstone (1938) mempertimbangkan hal tersebut dengan menilai kemampuan atau sifat individu menggunakan serangkaian butir dari berbagai tingkat kesulitan. Pada pengukuran kepribadian, istilah kesulitan butir (*items difficulty*) setara dengan derajat penilaian persetujuan terhadap konten (*level of endorsement*).

Semakin tinggi intensitas dari konten yang dimuat suatu butir, semakin tinggi tingkat kesulitan butir. Semakin tinggi tingkat kesulitan butir, semakin rendah derajat persetujuan responden terhadap konten yang dimuat. Intensitas konten dari butir-butir di dalam alat ukur, termasuk alat ukur kepribadian, harus bervariasi (Chernyshenko et al., 2007). Tujuan dari penggunaan butir dengan berbagai level intensitas adalah untuk meningkatkan presisi skor yang dihasilkan. Cara untuk memenuhi tujuan tersebut, yakni pengembang skala diharapkan menulis serangkaian butir yang memiliki derajat penilaian persetujuan terhadap konten dengan tingkatan rendah, sedang, dan tinggi (Drasgow et al., 2010). Adanya variasi intensitas pada butir skala dapat mempermudah dalam menentukan penilaian terhadap sifat individu. Misalnya pada skala depresi, variasi intensitas pada butir dapat memudahkan pengklasifikasian individu ke dalam depresi ringan, sedang, atau tinggi. Pada penelitian ini intensitas konten ukur butir diartikan sebagai informasi yang menunjukkan level suatu konten butir dalam memanifestasikan konstruk ukur. Sebagai contoh dalam konstruk perilaku agresif, “*memelototi*” menunjukkan intensitas agresif di level rendah sedangkan “*mencederai*” menunjukkan intensitas agresif level tinggi.

Butir-butir dalam skala psikologi diharapkan mampu menjangkau berbagai macam aspek individu tidak hanya berdasarkan konten yang diukur akan tetapi juga berdasarkan aspek yang tidak terkait langsung dengan konten (Furr & Bacharach, 2013). Pada

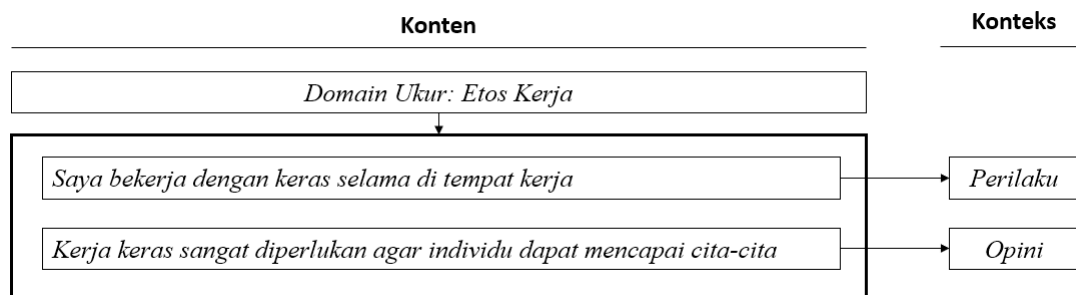
penelitian ini, aspek yang terkait dengan atribut ukur dinamakan dengan dimensi ukur, sedangkan aspek yang tidak terkait langsung dengan atribut ukur dinamakan dengan fitur non-konten. Contoh dari fitur non-konten adalah konteks, situasi, waktu, atau hal-hal lain yang bersifat umum yang berlaku pada semua pada atribut psikologi. Fitur non-konten merupakan variasi sudut pandang terhadap sifat yang dimanifestasikan pada indikator yang diukur (DeVellis, 2011). Loevinger (1957) menyatakan bahwa kumpulan butir skala yang dipilih harus mewakili dimensi ukur dan fitur non-konten disesuaikan dengan semua konten yang membentuk sifat yang akan diukur. Selain itu, butir skala juga harus disesuaikan pada semua teori alternatif yang diketahui dari sifat tersebut.

Contoh lain dari penggunaan fitur non-konten adalah penggunaan pernyataan positif dan negatif pada skala psikologi untuk mengurangi bias respons. Responden memiliki kecenderungan untuk memilih jawaban setuju pada butir skala yang memiliki pernyataan positif. Hal ini menyebabkan butir skala yang mengandung pernyataan positif memiliki skor yang cenderung lebih tinggi daripada butir skala dengan pernyataan negatif (Salazar, 2015). Kecenderungan tersebut menyebabkan eror pengukuran yang dapat mengganggu interpretasi terhadap hasil pengukuran. Eror tersebut berupa gambaran perilaku yang terlalu baik. Campuran pernyataan positif dan negatif memaksa responden untuk lebih fokus memilih jawaban, serta mengurangi adanya bias respons yang ekstrem dan kecenderungan menyetujui semua pernyataan (*acquiescent bias and yes-saying*) (Qasem et al., 2014).

Penggunaan fitur non-konten juga didemonstrasikan oleh Peabody (1970) yang membedakan butir dalam psikologi dapat dibagi menjadi butir yang memiliki aspek evaluatif dan deskriptif. Aspek evaluatif dapat menggambarkan pernyataan secara umum, sedangkan aspek deskriptif dapat

menggambarkan pernyataan secara umum dan lebih spesifik. Contoh pernyataan kalimat evaluatif, yakni "*Stoberi ini memiliki rasa manis,*" dan contoh kalimat pernyataan deskriptif adalah "*Stoberi ini besar, berwarna merah, dan mengandung banyak air*" (Saucier et al., 2001). Peabody (1970) menyimpulkan aspek evaluatif dan deskriptif sama pentingnya. Namun aspek deskriptif lebih memiliki kemampuan dalam mengungkap kepribadian karena responden lebih mudah untuk menyesuaikan pernyataan dengan keadaan diri.

Beberapa ahli ada yang juga mengembangkan aspek non konten dalam suatu butir, salah satunya adalah Bäckström et al. (2009). Dia mengajukan fitur *evaluation neutralization*, yakni penggunaan kalimat netral untuk butir skala. Netralisasi evaluasi diajukan karena beberapa butir skala memuat kalimat yang memotivasi responden untuk memberi jawaban berdasarkan norma sosial. Butir skala yang mengandung frase positif dibingkai menjadi sedikit positif dan butir yang mengandung frase negatif dibingkai menjadi sedikit negatif sehingga memiliki frase netral. Misalnya dari kalimat "*Mudah sedih*" diubah menjadi "*Mudah bereaksi pada hal yang terjadi*", dan "*Kurang peduli dengan orang lain*" diubah menjadi "*Percaya bahwa orang lain dapat menjaga dirinya sendiri*". Konsep ini dianggap dapat mengurangi efek kecenderungan sosial dalam memberi penilaian pada butir skala. Gambar 1 menjelaskan contoh bagaimana sebuah butir yang mengukur domain ukur yang sama dapat memuat konteks yang berbeda. Dapat dikatakan bahwa suatu butir memiliki dua atribut (fitur) yaitu fitur konten dan fitur non konten (konteks). Butir "*Saya bekerja dengan keras selama di tempat kerja*" cenderung menggunakan konteks aksi atau perilaku sedangkan butir "*Kerja keras sangat diperlukan agar individu dapat mencapai cita-cita*" memuat konteks opini.



Gambar 1. Contoh Atribut (Fitur) Butir yang terkait dengan Konten dan Konteks

Konsep kerangka kerja yang dikemukakan oleh beberapa ahli berfungsi untuk membantu dalam memetakan butir skala dari indikator perilaku agar sesuai dengan tujuan ukur. Karakteristik butir memiliki peran penting dalam menentukan validitas alat ukur, misalnya sensitivitasnya dalam mengungkap perilaku lebih spesifik, dan mengurangi kecenderungan sosial. Melalui penelitian ini penulis menawarkan pemetaan fitur non-konten butir skala psikologi berdasarkan aspek perilaku, preferensi, afeksi, keyakinan diri, dan kognitif. Dasar dari klasifikasi ini adalah literatur-literatur tentang kepribadian yang menunjukkan bahwa kepribadian itu sesuatu yang kompleks.

Aspek pertama dari fitur non-konten butir adalah perilaku. Penggunaan aspek ini didasarkan pada literatur yang menunjukkan bahwa perilaku merupakan manifestasi dari kepribadian. Perilaku merupakan ucapan verbal atau gerakan yang dapat dideteksi oleh pengamat (*observer*) melalui proses sensori normal (Furr, 2009). Rosenbleth dan kawan-kawan (1943) mengklasifikasikan perilaku menjadi dua jenis, yaitu perilaku aktif dan perilaku pasif. Perilaku aktif adalah gerakan berdasarkan energi internal, sedangkan perilaku pasif adalah gerakan yang disebabkan oleh energi eksternal. Paparan di atas menunjukkan bahwa fitur non-konten dalam bentuk perilaku memiliki dasar yang kuat.

Aspek kedua dari fitur non-konten yakni aspek kognitif. Aspek kognitif banyak disinggung dalam literatur yang membahas masalah kepribadian (Gruszka et al., 2010).

Kognitif berkaitan dengan bagaimana orang memahami, mempelajari, mengingat, dan berpikir mengenai suatu informasi (Sternberg, 2012). Aspek ketiga dari fitur non-konten adalah afeksi. Funder (2001) menyatakan bahwa afeksi merupakan salah satu bentuk manifestasi kepribadian. Aspek afeksi merupakan proses kognitif berupa penilaian terhadap stimulus yang dinilai berdasarkan memori dan kriteria subjektif yang dimiliki seseorang. Proses kognitif tersebut menimbulkan reaksi emosional yang menyenangkan atau tidak menyenangkan (Jung, 1975). Proses kognitif tersebut juga melibatkan kesadaran seseorang untuk menilai perasaan yang sedang dialami (Damasio, 1999).

Preferensi merupakan aspek keempat dari fitur non-konten yang ditawarkan pada penelitian ini. Aspek ini digunakan berdasarkan literatur yang menunjukkan bahwa preferensi merupakan manifestasi kepribadian (Werner & Pervin, 1986). Preferensi menunjukkan kecenderungan untuk mempertimbangkan apakah sesuatu diinginkan atau tidak diinginkan. Hal ini biasanya terjadi ketika proses pengambilan keputusan (Zajonc, 1980). Preferensi seseorang dipengaruhi oleh memori, lingkungan, atau kepercayaan (Dijksterhuis et al., 2006). Aspek kelima dari fitur non-konten butir adalah keyakinan diri. Aspek tersebut merupakan salah satu manifestasi kepribadian (Schyns & von Collani, 2002). Keyakinan diri merupakan kepercayaan mengenai kemampuan diri sendiri untuk mencapai tujuan. Menurut teori

keyakinan diri, seseorang hanya berusaha pada hal-hal yang dipercayai dapat dicapai dan tidak berusaha pada hal-hal yang dipercayai akan gagal untuk dicapai. Literatur-literatur di atas menunjukkan bahwa manifestasi kepribadian dapat dilihat dari aspek kognitif, afeksi, perilaku (Revelle, 2008), preferensi (Werner &

Pervin, 1986), dan keyakinan diri (Schyns & von Collani, 2002). Literatur tersebut menunjukkan bahwa kelima aspek yang ditawarkan layak dijadikan sebagai aspek fitur non-konten berdasarkan dasar literatur yang dipaparkan di atas.

Tabel 1.
Klasifikasi Fitur non-konten Pernyataan di dalam Butir

Kategori	Keterangan	Contoh Pernyataan
Perilaku	Ucapan verbal atau gerakan yang dapat dideteksi oleh pengamat (<i>observer</i>) melalui proses sensori normal.	Menertawakan lelucon. (<i>Laugh at a joke.</i>) Melakukan banyak pekerjaan. (<i>I do just enough work to get by.</i>)
Kognitif	Berkaitan dengan bagaimana orang memahami, mempelajari, mengingat, dan berpikir mengenai suatu informasi.	Sulit berkonsentrasi pada pekerjaan ketika saya mendengar berita baik. (<i>Have difficulty concentrating on the job at hand when I hear good news.</i>) Saya sulit memahami gagasan yang abstrak. (<i>I have difficulty understanding abstract ideas.</i>)
Afeksi	Merupakan proses kognitif berupa penilaian terhadap stimulus yang dinilai berdasarkan memori dan kriteria subjektif yang dimiliki seseorang.	Saya mudah merasa marah (<i>I get angry easily.</i>) Saya suka membaca buku dan artikel yang menantang. (<i>I enjoy reading challenging books and articles.</i>)
Preferensi	Menunjukkan kecenderungan untuk mempertimbangkan apakah sesuatu diinginkan atau tidak diinginkan, biasanya terjadi ketika proses pengambilan keputusan.	Saya cenderung memilih kandidat politik liberal (<i>I tend to vote for liberal political candidates.</i>) Saya lebih suka untuk tidak terlalu dekat dengan orang lain. (<i>I prefer not to get too close to people.</i>)
Keyakinan Diri	Merupakan kepercayaan mengenai kemampuan diri sendiri untuk mencapai tujuan.	Menganggap diri saya sama seperti orang lain. (<i>See myself as an average person.</i>) Saya mampu mengendalikan keinginan saya. (<i>I am able to control my cravings.</i>)

Kajian mengenai eksplorasi fitur non konten dari sebuah alat ukur perlu untuk dilakukan karena merupakan bagian dari studi mengenai validitas dalam pengembangan alat ukur. Menurut APA, AERA, & NCME (2014), validitas menunjukkan derajat dari suatu bukti atau teori yang dipakai sebagai landasan dalam pengembangan alat ukur dapat mendukung interpretasi terhadap skor yang dihasilkan dari pengukuran yang dilakukan Proses melakukan validasi terhadap alat ukur dan fiturnya (misalnya konstruk ukur, skor, interpretasi) merupakan proses yang terus dilakukan meskipun alat ukur tersebut sudah selesai

dalam proses pengembangannya. Validitas saat ini ditekankan pada aspek pembuktian.

Ada beberapa dimensi yang dapat dipakai untuk mendukung pembuktian, yaitu pembuktian berdasarkan konten alat ukur, pembuktian berdasarkan proses respons, pembuktian berdasarkan struktur internal, pembuktian berdasarkan kaitan konstruk ukur dengan konstruk ukur lain, dan pembuktian berdasarkan konsekuensi dari pengukuran yang dilakukan. (1) Pembuktian berdasarkan konten alat ukur menjelaskan sejauh mana domain-domain yang diukur oleh alat ukur dapat diakomodasi oleh alat ukur. Jika sebuah alat ukur merepresentasikan semua domain

ukur dengan tepat maka alat ukur tersebut telah memiliki bukti dari aspek konten ukur. Ada ukuran yang dapat dipakai untuk mengevaluasi konten alat ukur, yaitu relevansi, ketepatan dan kepentingan dari domain ukur yang dilibatkan. (2) Pembuktian validitas berdasarkan proses respons menunjukkan kesamaan proses psikologis yang dilakukan oleh individu dalam menjawab atau menanggapi sebuah butir sudah sesuai dengan proses psikologis yang terkait dengan alat ukur. Misalnya dalam suatu tes yang mengukur kemampuan penalaran, jawaban yang diberikan oleh subjek harus berkaitan dengan proses penalaran. (3) Pembuktian berbasis pada struktur internal menekankan pada pembuktian berbasis pada ketepatan penjabaran konstruk ukur menjadi dimensi-dimensi yang sifatnya lebih operasional serta mampu menjelaskan kaitan antar dimensi-dimensi tersebut dengan tepat. Misalnya kaitan antar dimensi tersebut membentuk struktur unidimensi atau multidimensi. (4) Pembuktian berdasarkan kaitan dengan konstruk ukur lain menekankan pada ketepatan pengembang alat ukur dalam memposisikan konstruk yang diukur oleh alat ukur yang dikembangkan dengan konstruk-konstruk yang sudah ada. Misalnya, secara teoretis, konstruk ukur yang merepresentasikan sebuah kemampuan harus memiliki kaitan yang tinggi dengan konstruk ukur yang juga merepresentasikan sebuah kemampuan, namun memiliki kaitan yang rendah dengan konstruk ukur yang mengukur pada aspek ciri sifat kepribadian. (5) Pembuktian berdasarkan konsekuensi menekankan pada ketepatan dari hasil pengukuran yang dilakukan dengan dampak dan akibatnya. Skor yang valid akan memberikan konsekuensi yang sejalan, misalnya skor tes potensi kemampuan yang valid memberikan konsekuensi terhadap prestasi atau kompetensi individu, dan bukan sebaliknya. Pembuktian berdasarkan konsekuensi ini sangat berkaitan erat dengan daya prediksi dari sebuah alat ukur terhadap kriteria yang ditetapkan. Kajian terhadap fitur

non konten sebuah butir mendukung upaya pembuktian dari aspek struktur alat ukur. Penggunaan fitur-fitur non konten butir yang bervariasi akan meningkatkan proporsi varians yang relevan dengan konstruk (*construct relevant variance*) dalam sebuah skor.

Penelitian ini bertujuan untuk membandingkan intensitas ukur suatu butir ditinjau dari fitur non konten suatu butir. Intensitas ukur suatu butir dapat dilihat melalui tingkat persetujuan (*endorsability*) responden terhadap butir. Dalam hal ini fitur non konten yang dipakai dalam penelitian ini didasarkan pada bagaimana ciri sifat kepribadian tersebut diekspresikan. Ada lima jenis ekspresi yang ditelaah pada penelitian ini yaitu perilaku (aksi), preferensi, afeksi, efikasi, dan kognitif. Butir yang memiliki tingkat persetujuan yang tinggi oleh responden menunjukkan bahwa butir tersebut memiliki intensitas yang rendah. Sebaliknya, butir-butir yang memiliki tingkat persetujuan yang rendah adalah butir-butir yang memiliki intensitas yang rendah karena hanya orang-orang yang memiliki ciri sifat di tingkat tinggi saja yang menyetujui butir tersebut.

Landasan berpikir yang dipakai mengasumsikan bahwa jenis-jenis ekspresi ciri sifat kepribadian (misalnya aksi, afeksi) antara satu dengan lainnya memiliki perbedaan adalah sebagai berikut. *Pertama*, jenis-jenis tersebut berkaitan dengan kemudahan entitas tersebut untuk diekspresikan. Ekspresi secara kognitif lebih mudah dibanding dengan ekspresi melalui aksi atau perilaku empiris. Misalnya, mendukung kejujuran lebih mudah diekspresikan dibanding dengan melakukan kejujuran. *Kedua*, klasifikasi berdasarkan ekspresi individu berkaitan dengan intensitas atau kedalaman jenis tersebut terkait dengan ciri sifat kepribadian yang diukur. Sebagai contoh, kedalaman derajat keramahan (Keramahan) individu lebih mudah ditunjukkan melalui perasaan dibanding dengan preferensi yang ditunjukkan dengan minat sosial.

METODE

Penelitian ini dilakukan dengan menggunakan data sekunder berdasarkan penelitian yang dilakukan oleh Condon dan kolega (2017). Terdapat dua jenis kelompok data pada penelitian ini, yaitu *Replicational Sample* dan *Confirmatory Sample*. Data *Replicational Sample* dikumpulkan pada tanggal 26 Juli 2014 hingga 7 Februari 2017, sedangkan data *Confirmatory Sample* dikumpulkan pada tanggal 22 Desember 2015 hingga tanggal 7 Februari 2017. Pengumpulan data tersebut menggunakan Synthetic Aperture Personality Assessment (“SAPA”), yakni alat pengumpulan data *online* secara kolaboratif untuk menilai konstruksi psikologis di berbagai domain kepribadian. Data tersebut digunakan untuk mengevaluasi struktur konstruk kepribadian. Dalam konteks teori kepribadian modern, kepribadian digambarkan dengan istilah Big Five (*Conscientiousness*, Keramahan, *Neuroticism*, *Openness*, and *Extraversion*). Sekitar kurang lebih 400 dimensi kepribadian dimasukkan ke dalam penelitian yang dilakukan, sebagian dimensi tersebut dipilih secara eksplisit karena sebagian besar dari butir tersebut termasuk ke dalam skala kepribadian yang sering digunakan. Jumlah butir skala kepribadian yang digunakan dalam penelitian tersebut adalah 696 butir. Salah satu skala yang digunakan dalam penelitian tersebut adalah skala IPIP-HEXACO dengan jumlah butir sebanyak 240 butir.

Partisipan dari yang *Replication Sample* memiliki jumlah 54.855 partisipan dengan jumlah perempuan sebanyak 62% dan jumlah laki-laki 38%. Pada *Confirmatory Sample* terdapat partisipan sebanyak 48.350 dengan jumlah perempuan 63% dan jumlah laki-laki 37%. Partisipan pada penelitian Condon dkk., (2017) memiliki kisaran usia 15 tahun hingga 65 tahun ke atas. Partisipan yang digunakan penelitian ini sebanyak 23.681 orang. Jumlah tersebut merupakan jumlah partisipan yang mengisi skala IPIP-HEXACO

dari penelitian yang dilakukan oleh Condon dkk., (2017).

Pada penelitian ini, kepribadian diukur dengan menggunakan skala kepribadian IPIP-HEXACO (Lee & Ashton, 2004). IPIP-HEXACO merupakan skala pelaporan diri yang terdiri dari 240 butir skala untuk mengukur domain *honesty-humility*, *emotion*, *extraversion*, Keramahan, *conscientiousness*, *openness to experience*. Masing-masing domain kepribadian diwakili oleh 40 butir skala. Contoh butir dalam skala ini adalah “*Keep others at a distance*”, “*Have excellent ideas*”, “*Let myself be influenced by others.*” Format respons untuk skala ini menggunakan lima poin skala. Skor 1 untuk menyatakan “sangat tidak setuju”, skor 2 untuk menyatakan “tidak setuju”, skor 3 untuk menyatakan “netral”, skor 4 untuk menyatakan “setuju”, dan skor 5 untuk menyatakan “sangat setuju”. Reliabilitas skala ini diestimasi dengan menggunakan teknik konsistensi internal dengan hasil mulai dari 0,73 sampai 0,88 (Ashton et al., 2014).

Tahap pertama dari penelitian ini adalah penilaian butir-butir oleh *rater* dengan tujuan untuk memperoleh informasi mengenai klasifikasi butir skala IPIP-HEXACO ke dalam lima fitur non-konten, yaitu aspek perilaku, kognitif, afeksi, preferensi, dan keyakinan diri. Butir skala IPIP-HEXACO yang berjumlah 240 butir dikelompokkan ke dalam enam dimensi kepribadian IPIP-HEXACO, yaitu kejujuran (*honesty-humility*), emosi (*emotion*), ekstrasversi (*extraversion*), keramahan (Keramahan), ketelitian (*conscientiousness*), keterbukaan (*openness to experience*). Masing-masing dimensi kepribadian HEXACO memiliki jumlah butir yang sama, yakni terdiri dari 40 butir skala. Setiap butir yang terdapat pada masing-masing dimensi akan diklasifikasikan oleh tiga *rater* yang berbeda. Jumlah *rater* dalam penelitian ini adalah 18 orang (6 peneliti di Fakultas Psikologi UGM, 12 mahasiswa psikologi pascasarjana). Setiap *rater* bertugas untuk mengklasifikasikan 40

butir skala dengan dimensi kepribadian pada IPIP-HEXACO yang ditentukan oleh peneliti. Mereka diberi panduan berupa pengertian aspek-aspek yang terdapat pada fitur non-konten dan contoh butir skala pada masing-masing aspek untuk memudahkan dalam pengklasifikasian butir skala. *Rater* hanya boleh memilih satu dari lima aspek yang telah ditentukan. Misalnya butir skala yang berbunyi “*I enjoy reading challenging books and articles*” harus diklasifikasikan ke dalam satu aspek, misalnya afeksi karena terdapat kata “*enjoy*” yang mengindikasikan perasaan menikmati sesuatu meskipun butir tersebut juga memanasifestasikan preferensi .

Data penelitian ini dianalisis melalui dua tahap, yaitu tahap kalibrasi butir, identifikasi statistik deskriptif tiap kombinasi kategori (dimensi x fitur) dan pengujian statistik rerata tingkat kesulitan di tiap kombinasi antar kategori. Kalibrasi butir dilakukan dengan menggunakan analisis butir berbasis teori respons butir (*item response theory/IRT*). Model IRT yang dipakai untuk mengkalibrasi adalah model kredit parsial (*partial credit model/PCM*). Pemilihan ini didasarkan pada hasil penelitian yang menunjukkan bahwa model ini sangat tepat diterapkan pada pengukuran kepribadian dengan menggunakan skala psikologi yang datanya bersifat politomi (Widhiarso, 2010). Hasil kalibrasi akan menghasilkan informasi mengenai lokasi butir (*item location*) yang didapatkan dari rerata nilai ambang (*threshold value*) yang merupakan nilai tingkat kesulitan antar opsi pilihan butir. Oleh karena merupakan rerata dari setiap nilai ambang, nilai lokasi butir menunjukkan tingkat kesulitan untuk di level butir. Semakin tinggi nilainya, semakin tinggi tingkat kesulitannya dan semakin rendah peluang responden untuk menyetujui pernyataan dalam butir tersebut.

Tahap kedua dilakukan dengan mengidentifikasi statistik deskriptif untuk masing-masing butir-butir yang berada pada kategori yang sama. Kategori pertama adalah

dimensi ukur yang terdiri dari enam dimensi sesuai dengan dimensi ukur HEXACO, Kategori kedua adalah fitur butir yang didapatkan dari penilaian para rater terhadap fitur butir. Ada lima kategori fitur butir, yaitu kognitif, afektif, efikasi perilaku dan preferensi. Tahap ketiga dilakukan dengan menguji rerata tiap kombinasi kategori dengan menggunakan uji statistik. Uji statistik yang dipakai adalah uji Welch yang dipilih karena jumlah butir antar satu kategori dengan kategori lainnya cenderung tidak seimbang. Literatur menunjukkan bahwa uji Welch tahan terhadap heterogenitas varians (Howell, 2010).

HASIL DAN PEMBAHASAN

Pada bagian ini akan dipaparkan hasil kalibrasi alat ukur dengan menggunakan pemodelan Rasch. Indeks yang dipakai dari hasil kalibrasi tersebut adalah tingkat kesulitan butir dalam satuan logit. Perbandingan rerata tingkat kesulitan antar fitur ini dilakukan untuk mengidentifikasi konsistensi tingkat kesulitan suatu fitur pada dimensi yang berbeda.

Hasil analisis butir pada dimensi kejujuran menunjukkan bahwa semua tingkat kesulitan butir cenderung sedang yang ditunjukkan dengan nilainya yang mendekati 0. Butir-butir yang mengandung fitur aksi dan kognisi cenderung mudah disetujui dibanding dengan efikasi dan preferensi dengan jarak yang cukup jauh. Tidak didapatkan perbedaan parameter lokasi butir antar fitur pada dimensi ini ($\chi^2 = 3.77$; $p > .05$). Butir-butir dengan fitur efikasi yang misalnya ditunjukkan dengan penggunaan kata yang menunjukkan kelebihan responden (misalnya, “*saya mampu ...*”, “*saya memiliki kapasitas ...*”) menunjukkan intensitas kejujuran yang tinggi. Kondisi ini berbeda dengan fitur perilaku yang sebagian besar justru menunjukkan kejujuran di level rendah karena mendapatkan persetujuan dari banyak orang.

Pada aspek emosi, butir-butir di dalamnya cenderung memiliki tingkat persetujuan yang sedang. Fitur yang memiliki

tingkat persetujuan yang rendah adalah preferensi ($M = .19$) dan fitur dengan tingkat persetujuan tinggi adalah efikasi ($b = -.25$). Tidak didapatkan perbedaan parameter lokasi butir antar fitur pada dimensi ini ($\chi^2 = 8.38$; $p > .05$). Tren nilai lokasi butir antar fitur pada dimensi ini sedikit berbeda dengan dimensi kejujuran. Butir-butir dengan fitur efikasi pada dimensi ini memiliki persetujuan yang rendah, berbeda dengan ketika pada dimensi kejujuran. Butir dengan fitur efikasi seperti “*membutuhkan bantuan*”, “*mudah panik*” atau “*merasakan emosi orang lain*” cenderung menjelaskan emosi di level rendah. Di sisi lain, butir dengan fitur afeksi seperti “*jarang merasa sedih*” atau “*gelisah dengan ketidaknyamanan*” memiliki intensitas yang tinggi karena cenderung hanya disetujui oleh responden dengan level emosi yang tinggi.

Butir-butir pada dimensi ekstraversi terletak pada lokasi butir sekitar 0 yang

menunjukkan kecenderungannya merepresentasikan intensitas yang sedang ($\text{min} = .40$; $\text{maks} = .20$). Temuan yang agak berbeda dengan dimensi lain adalah lokasi butir dengan fitur afeksi yang rendah, artinya aspek afeksi (misalnya, tertawa, tersenyum atau merasa nyaman) merepresentasikan level rendah dalam ekstraversi. Sebaliknya, ekstraversi level tinggi banyak ditunjukkan oleh butir dengan fitur aksi, seperti banyak bicara, berenergi dan mendekati orang lain. Tidak didapatkan perbedaan parameter lokasi butir antar fitur pada dimensi ini ($\chi^2 = 5.74$; $p > .05$). Butir “*menumpahkan perasaan*” ($b = .69$) menjadi butir yang menunjukkan derajat tertinggi keekstraversian seseorang karena hanya responden dengan tingkat ekstraversi yang tinggi yang cenderung menyetujui butir ini.

Tabel 1.
Statistik Deskriptif Parameter Lokasi Butir (dalam Satuan Logit)

Dimensi	Fitur	Rerata	Jumlah	Dimensi	Fitur	Rerata	Jumlah
Kejujuran	aksi	-0.107	18	Ketelitian	Aksi	0.049	21
	efikasi	0.223	6		Efikasi	0.000	4
	kognisi	-0.082	9		Kognisi	0.053	3
	preferensi	0.191	7		Preferensi	-0.098	12
	Afeksi	0.162	13		Afeksi	0.245	2
Emosi	Aksi	-0.015	8	Keterbukaan	Aksi	0.208	5
	Efikasi	-0.249	12		Efikasi	-0.139	8
	Kognisi	0.087	3		Kognisi	-0.171	7
	Preferensi	0.188	4		Preferensi	0.043	18
	Afeksi	-0.395	4		Afeksi	0.547	3
Ekstraversi	Aksi	0.200	10	Keramahan	Aksi	-0.210	12
	Efikasi	-0.027	11		Efikasi	0.031	16
	Kognisi	-0.098	5		Kognisi	0.170	5
	Preferensi	0.038	10	Preferensi	-0.113	4	

Keterangan: Rerata tingkat kesulitan menunjukkan tingkat persetujuan responden terhadap suatu butir. Tingkat kesulitan pada tabel ini menggunakan skala logit yang bergerak antara mendekati -4 hingga 4, meskipun kadang melebihi harga ini karena sifatnya yang asimtotik.

Pada dimensi keramahan, semua fitur memiliki tingkat rata-rata persetujuan yang mendekati nol ($\text{min} = -.21$; $\text{maks} = .55$). Dua butir yang memiliki tingkat persetujuan yang

tinggi adalah “*menerima orang apa adanya*” ($b = -1,16$; fitur preferensi) dan “*memaafkan orang lain*” ($b = -.75$; fitur aksi). Sebaliknya, contoh butir yang menunjukkan orang dengan

level keramahan yang tinggi adalah “*mencintai musuh*” ($b = .81$; fitur preferensi) dan “*bereaksi terhadap kritik*” ($b = .68$; fitur afeksi). Tidak didapatkan perbedaan parameter lokasi butir antar fitur pada dimensi ini ($\chi^2 = 8.38$; $p > .05$). Berbeda dengan ekstraversi, fitur aksi pada dimensi keramahan justru merepresentasikan level keramahan yang rendah dibanding dengan fitur afeksi. Hal ini dikarenakan perilaku yang dilibatkan pada skala ini cenderung perilaku umum yang banyak dilakukan oleh orang awam yang berbudaya. Sebaliknya, status emosi seperti “*tetap santai menghadapi kritik*” atau “*jarang marah*” menunjukkan level keramahan yang tinggi.

Pada dimensi ketelitian, sebagian besar butir memiliki tingkat kesulitan yang sedang. Semua fitur pada dimensi ini memiliki nilai lokasi butir yang hampir setara yang menunjukkan daya persetujuan yang setara antara satu fitur dengan fitur lainnya. Fitur dengan tingkat daya persetujuan tertinggi adalah preferensi ($M = -.10$), sedangkan fitur kognisi adalah yang terendah ($M = .05$). Tidak didapatkan perbedaan parameter lokasi butir antar fitur pada dimensi ini ($\chi^2 = 1.23$; $p > .05$). Temuan ini menunjukkan bahwa representasi ketelitian individu tidak berkaitan dengan fitur manifestasinya. Level ketelitian individu yang rendah dapat ditunjukkan perilaku, misalnya “*menyelesaikan tugas*” ($b = .70$, fitur aksi, level ketelitian rendah) dan “*menjaga ruangan tetap rapi*” ($b = .73$, fitur aksi, level ketelitian tinggi). Hal yang sama juga terjadi pada fitur preferensi. “*menginginkan kualitas*” ($b = -.57$, fitur preferensi, level ketelitian rendah) dan “*menyukai perencanaan*” ($b = .65$; fitur preferensi, level ketelitian rendah).

Tren parameter lokasi butir pada dimensi keterbukaan tidak jauh beda dengan temuan pada dimensi lain dengan lokasi parameter butir cenderung sedang ($\min = -.14$; $\max = .25$). Temuan yang cukup menarik terletak pada kesetaraan nilai lokasi afeksi dan aksi yang cenderung setara. Temuan ini menunjukkan bahwa ekspresi individu dalam

hal ciri sifat keterbukaan baik melalui aksi (misalnya “*menyelami kondisi saat ini*”) dan (misalnya “*tidak mempermasalahkan masalah politik*”) merepresentasikan level keterbukaan yang setara. Tidak didapatkan perbedaan parameter lokasi butir antar fitur pada dimensi ini ($\chi^2 = 5.69$; $p > .05$).

Tujuan dari penelitian ini adalah membandingkan tingkat kesulitan butir skala psikologi berdasarkan fitur non-konten butir. Penelitian ini mengajukan beberapa fitur yang merepresentasikan properti non-konten, antara lain aspek perilaku, preferensi, afeksi, keyakinan diri, dan kognitif. Secara umum, hasil penelitian menunjukkan bahwa kelima fitur non-konten tersebut memiliki proporsi yang berbeda pada dimensi Keramahan, *Conscientiousness*, dan *Emotion*. Ketiga dimensi di atas menunjukkan hasil yang tidak konsisten. Hasil tersebut dipengaruhi oleh jumlah butir skala yang cenderung berbeda pada setiap aspeknya. Terdapat beberapa aspek yang memiliki jumlah dominan pada setiap dimensi. Selain itu, pada dimensi Keramahan juga tidak ditemukan aspek kognitif dalam butir skalanya.

Berdasarkan analisis hasil pada dimensi Keramahan ditemukan bahwa aspek afeksi termasuk ke dalam kategori paling sulit pada sub skala ini. Sedangkan aspek keyakinan diri dan aspek perilaku termasuk ke dalam kategori mudah, serta preferensi termasuk ke dalam kategori paling mudah dalam dimensi tersebut. Pada dimensi *Conscientiousness* ditemukan bahwa aspek afeksi termasuk ke dalam kategori sulit dan aspek kognitif termasuk ke dalam kategori paling sulit pada dimensi ini. Sedangkan aspek perilaku dan aspek preferensi termasuk ke dalam kategori mudah, serta aspek keyakinan diri termasuk ke dalam kategori paling mudah dalam sub skala ini. Pada dimensi *Emotion* ditemukan bahwa aspek keyakinan diri termasuk ke dalam kategori paling sulit dalam kelompok dimensi ini dan aspek perilaku termasuk ke dalam kategori sulit. Selain itu aspek kognitif dan

aspek preferensi termasuk ke dalam kategori mudah, serta aspek afeksi termasuk ke dalam kategori paling mudah dalam sub skala dimensi *Emotion*.

Hasil yang tidak konsisten pada setiap dimensi menunjukkan bahwa fitur non konten pada butir skala psikologi tidak mempengaruhi secara langsung tergantung dari konstruk yang diukur. Secara keseluruhan ditemukan kesamaan hasil pada dimensi Keramahan dan dimensi *Conscientiousness*, yaitu aspek afeksi pada kedua dimensi tersebut termasuk ke dalam kategori sulit. Berbeda dengan kedua aspek tersebut, pada dimensi *Emotion* ditemukan bahwa aspek afeksi termasuk ke dalam kategori paling mudah. Selain itu juga ditemukan kesamaan bahwa aspek preferensi pada ketiga dimensi di atas termasuk ke dalam kategori mudah.

Aspek afeksi merupakan proses kognitif berupa penilaian terhadap stimulus yang dinilai berdasarkan memori dan kriteria subjektif yang dimiliki seseorang. Proses kognitif tersebut menimbulkan reaksi emosional yang menyenangkan atau tidak menyenangkan (Jung, 1975). Proses kognitif tersebut juga melibatkan kesadaran seseorang untuk menilai perasaan yang sedang dialami (Damasio, 1999). Butir skala aspek afeksi pada dimensi Keramahan dan *Conscientiousness* memiliki kesamaan, yakni keduanya termasuk ke dalam kategori butir yang sulit. Contoh butir dalam aspek afeksi pada dimensi Keramahan, antara lain “Merasa sedih ketika orang lain mengubah cara saya mengatur berbagai hal” (“*Get upset if others change way I arranged things*”), “Saya terganggu dengan kesalahan orang lain” (“*Am annoyed by others' mistakes*”), dan “Saya jarang merasa tersinggung” (“*Rarely get irritated*”). Kemudian contoh butir dalam aspek afeksi pada dimensi *Conscientiousness* adalah “Suka merapikan” (“*Like to tidy up*”), “Tidak menyukai pekerjaan yang tidak sempurna” (“*Dislike imperfect works*”), “Mudah kehilangan minat pada tugas yang saya

kerjakan” (“*Quickly lose interest in the tasks I start*”).

Rendahnya derajat penilaian persetujuan terhadap konten pada kedua aspek di atas dikarenakan butir tersebut kurang memaparkan objektivitas dalam mengungkapkan pernyataan mengenai perasaan. Misalnya pada butir dimensi Keramahan, yakni menyatakan “Saya jarang merasa tersinggung”, butir tersebut tidak menyertakan objek penyebab perasaan tersinggung. Contoh lainnya pada dimensi *Conscientiousness* terdapat pernyataan “Suka merapikan” tidak menjelaskan objek yang akan dirapikan. Kurangnya objektivitas dalam mengungkap perasaan yang muncul membuat seseorang kesulitan untuk mendefinisikannya (Smith, 1914).

Selain itu, untuk menentukan perasaan menyenangkan atau tidak menyenangkan pada orang lain juga tergantung pada situasi yang mendasarinya (Madell & Ridley, 1997). Misalnya pada pernyataan “Tidak menyukai pekerjaan yang tidak sempurna”, kata ‘tidak sempurna’ pada butir ini tidak menjelaskan konteks yang spesifik. Pernyataan pada butir yang kurang spesifik menyebabkan responden kesulitan dalam menentukan jawaban yang sesuai dengan dirinya (Kline, 2015).

Temuan khusus pada dimensi Keramahan adalah butir dalam aspek afeksi cenderung mengungkapkan pernyataan tentang perasaan seseorang terhadap perilaku orang lain. Contohnya pada pernyataan “Merasa sedih ketika orang lain mengubah cara saya mengatur berbagai hal”. Pernyataan tersebut juga menjadi penyebab aspek afeksi pada dimensi Keramahan termasuk ke dalam kategori sulit karena seseorang cenderung tidak mudah dalam mengungkapkan perasaannya mengenai perilaku orang lain (Burns, 1989).

Di pihak lain, butir skala dengan aspek afeksi pada dimensi *Emotion* memiliki pernyataan dengan konteks yang lebih spesifik daripada butir skala pada kedua dimensi

lainnya. Butir skala pada dimensi *Emotion* tersebut melibatkan konten emosi. Emosi menjelaskan perasaan pada kategori yang lebih spesifik (Brown, 2014). Contoh pernyataan pada dimensi ini, antara lain “Merasa takut ketika berjalan di kota yang memiliki tingkat kejahatan yang tinggi” (“*Would fear walking in a high-crime part of a city*”), “Mudah merasa sedih ketika mendengarkan peristiwa yang menyedihkan” (“*Immediately feel sad when hearing of an unhappy event*”), dan “Jarang merasa sedih ketika membaca cerita yang menyedihkan” (“*Seldom feel weepy while reading sad story*”). Ketiga pernyataan tersebut menjelaskan keadaan yang spesifik. Pernyataan yang spesifik tersebut cenderung lebih memudahkan responden dalam memberikan penilaian persetujuan terhadap butir (Kline, 2015).

Aspek preferensi menunjukkan kecenderungan untuk mempertimbangkan apakah sesuatu diinginkan atau tidak diinginkan (Zajonc, 1980). Preferensi seseorang dipengaruhi oleh memori, lingkungan, dan nilai yang dianut (Dijksterhuis et al., 2006). Contoh butir skala pada aspek preferensi, yaitu “Mementingkan kualitas” (“*Demand quality*”) dan “Membiarkan sesuatu terjadi apa adanya” (“*Prefer to just let things happen*”). Butir skala tersebut memiliki derajat penilaian persetujuan terhadap konten yang tinggi yang artinya responden mudah dalam memberi penilaian pada butir. Tingginya derajat penilaian tersebut terjadi karena butir skala pada aspek preferensi berkaitan dengan pengambilan keputusan. Pengambilan keputusan dalam aspek preferensi melibatkan nilai dan pengalaman yang telah dimiliki seseorang. Adanya nilai dan pengalaman tersebut memudahkan seseorang untuk menentukan sesuatu yang sesuai dan menguntungkan bagi diri sendiri (Hoeffler & Ariely, 1999; Kahneman & Tversky, 1981). Selain itu, seseorang cenderung memilih sesuatu yang dianggap menyenangkan dan menguntungkan (Smallman & Roese, 2008).

Keterbatasan penelitian ini antara lain jumlah sebaran butir skala tidak seimbang pada setiap aspeknya, misalnya jumlah butir skala aspek perilaku pada dimensi *Conscientiousness* memiliki jumlah butir lebih banyak dibanding dengan aspek lainnya. Keterbatasan lainnya, yakni pada dimensi Keramahan tidak terdapat aspek kognitif sehingga tidak memiliki proporsi aspek yang sama dengan dimensi lainnya.

Para peneliti sudah mengembangkan konsep sebagai acuan kerangka kerja dalam memetakan fitur non-konten, misalnya dengan menamakannya dengan atribut butir (Mael, 1991). Kerangka ini dapat dipakai untuk penulisan butir. Pengembangan fitur non-konten dapat meningkatkan manfaat dalam penulisan butir. Standardisasi kerangka kerja dengan cara kontekstualisasi butir dapat memfasilitasi presentasi diri responden yang lebih akurat. Tes kepribadian yang tidak memiliki konteks pada butirnya dapat memberi interpretasi yang berbeda pada responden dan kurang dapat memprediksi perilaku dalam konteks tertentu. Misalnya tes kepribadian untuk memprediksi perilaku kerja karyawan, akan lebih baik apabila melibatkan konteks lingkungan kerja pada butir skala. Butir skala yang dikontekstualisasikan secara spesifik dapat meningkatkan prediksi perilaku (Bing, Whanger, & Vanhook, 2004). Penggunaan fitur non-konten dalam bentuk kontekstualisasi yang spesifik dapat meningkatkan validitas prediktif tes kepribadian (Schmit et al., 1995). Secara keseluruhan uraian tersebut menunjukkan bahwa fitur non-konten berupa kontekstualisasi butir skala dapat memberikan manfaat berupa spesifikasi perilaku dan meningkatkan validitas prediktif.

Secara umum penelitian ini menemukan bahwa peranan variasi fitur-fitur non konten butir terhadap perubahan intensitas konten butir tidak cukup besar. Temuan ini ditunjukkan dengan temuan bahwa antara satu tipe fitur dan tipe fitur lainnya memiliki daya persetujuan yang setara. Kesetaraan properti

psikometris fitur-fitur butir terkait dengan level persetujuan terhadap pernyataan suatu butir dapat menjadi dasar bagi penulis soal untuk melibatkan berbagai macam fitur non konten terkait dengan ragam manifestasi dari ciri sifat yang diukur. Tujuannya adalah untuk meningkatkan jangkauan ukur yang tidak hanya merepresentasikan satu jenis manifestasi saja akan tetapi juga berbagai macam manifestasi. Sebagai contoh penulis butir tidak hanya membuat pernyataan yang memantapkan perilaku saja akan tetapi juga afeksi atau preferensi.

Penggunaan butir dengan melibatkan berbagai macam fitur sangat disarankan oleh para pakar dalam bidang pengukuran. Loevinger (1957) menyarankan agar ketika menulis suatu butir, pengembang skala melibatkan berbagai macam perspektif untuk dilibatkan dalam butir yang ditulis agar skala yang dikembangkan menghasilkan informasi yang komprehensif. Bing dan kawan-kawan (2004) merekomendasikan agar penulis butir melibatkan fitur non konten butir berdasarkan tingkat kespesifikan butir (umum vs. spesifik). He (2019) merekomendasikan penggunaan berbagai macam sisi (*facet*) untuk pengukuran ciri sifat kepribadian. Namun demikian, pemaknaan terhadap sisi suatu konstruk ini dapat diartikan dua hal. Sisi konstruk (*facet*) dapat diartikan sebagai bagian dari konten, namun memungkinkan untuk diartikan sebagai fitur non konten.

Penelitian ini memiliki beberapa keterbatasan. Pertama, keterbatasan terkait dengan jumlah butir yang tidak setara antar dimensi. Hasil kategorisasi menunjukkan bahwa jumlah butir dengan fitur yang sama berbeda-beda. Sebagai contoh pada dimensi kejujuran, jumlah butir yang mengandung fitur aksi lebih banyak dibanding dengan fitur afeksi. Penelitian lebih lanjut dapat menggunakan kriteria tertentu agar didapatkan sebaran butir dengan fitur non konten yang setara. Misalnya mengadaptasi butir yang ada menjadi alat ukur baru dengan jumlah fitur

yang setara. Kedua, data SAPA project memiliki tingkat kelengkapan data yang rendah karena pengadministrasiannya dilakukan dengan teknik *missing by design*. Hal ini dikarenakan responden merespons alat ukur lain selain HEXACO. Oleh karena itu teknik analisis yang dilakukan hanya dapat dilakukan dengan menggunakan analisis dengan pendekatan teori tes modern atau melakukan imputasi data terlebih dahulu.

SIMPULAN

Penulisan butir pada sebuah skala psikologi seringkali dilakukan tanpa berpegang pada sebuah acuan sehingga penulis cenderung menulis butir berdasarkan pengalamannya. Penelitian ini membuktikan bahwa fitur non konten sebuah butir merupakan fitur yang tidak mengganggu struktur sebuah tes. Antara satu fitur dengan fitur lainnya memiliki karakteristik properti psikometris yang setara. Oleh karena fitur non konten butir dapat dipakai sebagai acuan dalam penulisan butir pada skala psikologi agar komposisi butir di dalam skala bersifat sistematis dan memiliki variasi yang cukup besar dalam mengukur aspek-aspek konstruk ukur di luar dari konten yang telah ditetapkan.

DAFTAR PUSTAKA

- APA, AERA, & NCME. (2014). *The Standards for Educational and Psychological Testing*. American Educational Research Association
- Ashton, M. C., Lee, K., & de Vries, R. E. (2014). The HEXACO Honesty-Humility, Agreeableness, and Emotionality Factors: A Review of Research and Theory. *Personality and Social Psychology Review, 18*(2), 139-152. Doi: 10.1177/1088868314523838
- Bäckström, M., Björklund, F., & Larsson, M. R. (2009). Five-factor inventories have a major general factor related to social desirability which can be reduced by

- framing items neutrally. *Journal of Research in Personality*, 43(3), 335-344. Doi: 10.1016/j.jrp.2008.12.013
- Bing, M. N., Whanger, J. C., Davison, H. K., & VanHook, J. B. (2004). Incremental Validity of the Frame-of-Reference Effect in Personality Scale Scores: A Replication and Extension. *Journal of Applied Psychology*, 89(1), 150-157. Doi: 10.1037/0021-9010.89.1.150
- Bing, M. N., Whanger, J. C., & Vanhook, J. B. (2004). Incremental Validity of the Frame-of-Reference Effect in Personality Scale Scores : A Replication and Extension. 89(1), 150-157. Doi: 10.1037/0021-9010.89.1.150
- Bradburn, N. M. (2004). *Asking questions : the definitive guide to questionnaire design*. John Wiley & Sons, Inc.
- Burns, D. D. (1989). *The Feeling Good Handbook*. William Morrow and Company.
- Chernyshenko, O. S., Stark, S., Drasgow, F., & Roberts, B. W. (2007). Constructing personality scales under the assumptions of an ideal point response process: toward increasing the flexibility of personality measures. *Psychological Assessment*, 19(1), 88-106. Doi: 10.1037/1040-3590.19.1.88
- Condon, D. M., Roney, E., & Revelle, W. (2017). A SAPA Project Update: On the Structure of phrased Self-Report Personality Items. *Journal of Open Psychology Data*, 5(1). Doi: 10.5334/jopd.32
- Damasio, A. (1999). *The Feeling of What Happens*. Hartcourt, Inc.
- DeVellis, R. F. (2011). *Scale development: Theory and applications*. SAGE Publications, Inc.
- Dijksterhuis, A., Bos Maarten, W., Nordgren Loran, F., & van Baaren Rick, B. (2006). On Making the Right Choice: The Deliberation-Without-Attention Effect. *Science*, 311(5763), 1005-1007. Doi: 10.1126/science.1121629
- Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 Years After Likert: Thurstone Was Right! *Industrial and Organizational Psychology-Perspectives on Science and Practice*, 3(4), 465-476. Doi: 10.1111/j.1754-9434.2010.01273.x
- Drasgow, F., Chernyshenko, O. S., & Stark, S. (2010). 75 years after Likert: Thurstone was right! *Industrial and Organizational Psychology: Perspectives on Science and Practice*, 3(4), 465-476. Doi: 10.1111/j.1754-9434.2010.01273.x
- Funder, D. C. (2001). Personality. *Annual Review of Psychology*, 52(1), 197-221. Doi: 10.1146/annurev.psych.52.1.197
- Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: An Introduction*. SAGE Publications, Inc.
- Gruszka, A., Matthews, G., & Szymura, B. (2010). *Handbook of Individual Differences in Cognition Attention, Memory, and Executive Control*. Springer Science.
- He, T.-h. (2019). Personality Facets, Writing Strategy Use, and Writing Performance of College Students Learning English as a Foreign Language. *SAGE Open*, 9(3), 2158244019861483. Doi: 10.1177/2158244019861483
- Howell, D. C. (2010). *Statistical Methods for Psychology*. Cengage Wadsworth.
- Kline, P. (2015). *A Handbook of Test Construction*. Routledge.
- Lee, K., & Ashton, M. C. (2004). Psychometric properties of the HEXACO personality inventory. *Multivariate Behavioral Research*, 39(2), 329-358. Doi: 10.1207/s15327906mbr3902_8
- Li, F., Cohen, A., & Shen, L. (2012). Investigating the Effect of Item Position in Computer-Based Tests.

- Journal of Educational Measurement*, 49(4), 362-379. Doi: 10.1111/j.1745-3984.2012.00181.x
- Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, 3, 635-694. Doi: 10.2466/pr0.3.7.635-694
- Mael, F. A. (1991). A conceptual rationale for the domain and attributes of biodata items. *Personnel Psychology*, 44(4), 763-792.
- Ones, D. S., & Viswesvaran, C. (1996). Bandwidth–fidelity dilemma in personality measurement for personnel selection. *17(6)*, 609-626. Doi: 10.1002/(sici)1099-1379(199611)17:6<609::Aid-job1828>3.0.Co;2-k
- Peabody, D. (1970). Evaluative and descriptive aspects in personality perception: A reappraisal. *Journal of Personality and Social Psychology*, 16(4), 639-646. Doi: 10.1037/h0030259
- Qasem, N., Ali, M., Gul, A., & Bilal, S. (2014). Effect of Items Direction (Positive or Negative) on the Factorial Construction and Criterion Related Validity in Likert Scale. *Khazar Journal of Humanities and Social Sciences*, 17(3), 77-84.
- Revelle, W. (2008). The Contribution Of Reinforcement Sensitivity Theory To Personality Theory. In P. J. Chorr (Ed.), *Reinforcement Sensitivity Theory of Personality*. Cambridge University Press.
- Rosenblueth, A., Wiener, N., & Bigelow, J. (1943). Behavior, Purpose and Teleology. *10(1)*, 18-24. Doi: 10.1086/286788
- Salazar, M. S. (2015). El dilema de combinar ítems positivos y negativos en escalas. *Psicothema*, 27(2), 192–199. Doi: 10.7334/psicothema2014.266
- Saucier, G., Ostendorf, F., & Peabody, D. (2001). The Circumplex of Personality Adjectives. *69(4)*, 537-582. Doi: 10.1111/1467-6494.694155
- Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, 80, 607-620. Doi: 10.1037/0021-9010.80.5.607
- Schyns, B., & von Collani, G. (2002). A new occupational self-efficacy scale and its relation to personality constructs and organizational variables. *European Journal of Work and Organizational Psychology*, 11(2), 219-241. Doi: 10.1080/13594320244000148
- Sternberg, R. J. (2012). *Cognitive Psychology*. Wadsworth.
- Thurstone, L. L. (1938). *Primary mental abilities*. University of Chicago Press.
- Werner, P. D., & Pervin, L. A. (1986). The Content of Personality-Inventory Items. *Journal of Personality and Social Psychology*, 51(3), 622-628. Doi: 10.1037/0022-3514.51.3.622
- Widhiarso, W. (2010). *Model Politomi dalam Teori Respons Butir*. Universitas Gadjah Mada.
- Widhiarso, W. (2016). Mengakomodasi Efek Metode dalam Pengujian Validitas Konstruk Melalui Analisis Faktor Konfirmatori. *Jurnal Psikologi Psikologia*, 1(1), 37--51. Doi: 10.21070/psikologia.v1i1.478
- Wilson, M. (2005). *Constructing Measures. An Item Response Modeling Approach*. Lawrence Erlbaum Associates Publishers.
- Zajonc, R. B. (1980). Feeling and thinking: Preferences need no inferences. *American Psychologist*, 35(2), 151-175. Doi: 10.1037/0003-066X.35.2.151