



Volume 10  
Nomor 1, 2024  
Halaman: 1–13  
DOI:10.22146/gamajop.75080

Naskah masuk 7 Juni 2022  
Naskah revisi 20 Mei 2024  
Naskah diterima 20 Mei 2024  
Naskah terbit 31 Mei 2024

**Kata Kunci:**

analisis faktor konfirmatori; efek metode; Tes PAPS; validitas konstruk

**Keywords:**

*confirmatory factor analysis; construct validity; method effects; Tes PAPS*

\*Alamat korespondensi: Email:  
[wahyu\\_psy@ugm.ac.id](mailto:wahyu_psy@ugm.ac.id)

## Eksplorasi Kemunculan Efek Metode dalam Tes Potensi Kemampuan Kognitif *Exploring the Presence of Method Effect on Cognitive Aptitude Test*

Wahyu Widhiarso\* and Alifa Rahmi Khairunisa

Fakultas Psikologi, Universitas Gadjah Mada, Indonesia

**Abstrak**

Skor tes individu dapat dipengaruhi oleh tiga sumber variasi yaitu konstruk ukur, metode pengukuran, dan eror pengukuran. Idealnya, pengaruh terhadap skor dalam sebuah tes didominasi oleh konstruk ukur. Namun, dalam beberapa kasus terdapat pengaruh dari metode pengukuran yang disebabkan oleh keunikan yang terkait dengan metode pengukuran dalam tes tersebut. Pengaruh dari keunikan ini disebut sebagai efek metode yang dapat memengaruhi validitas struktural suatu alat ukur. Penelitian ini mengeksplorasi efek metode pada Tes PAPS UGM Seri E menggunakan analisis faktor konfirmatori dengan model bifaktor (*bifactor*). Hasil analisis ( $N = 2.170$ ) menunjukkan bahwa tidak terdapat efek metode dengan porsi yang cukup besar di dalam PAPS Seri E1 dan E2. Semua variasi skor didominasi oleh efek konstruk ukur daripada efek metode pengukuran yang ditunjukkan dengan nilai bobot faktor pada komponen lebih didominasi oleh konstruk ukur dibandingkan dengan metode pengukuran. Dapat disimpulkan bahwa skor yang didapatkan oleh peserta tes cenderung merepresentasikan kemampuan yang bersifat umum yaitu penalaran kognitif daripada kemampuan yang bersifat spesifik yang terkait dengan media pengukuran.

**Abstract**

Test scores can be influenced by three sources: measurement construct, measurement method, and measurement error. Ideally, a score should strongly reflect measurement construct. However, in some cases, there are contributions of measurement method because of the uniqueness that is not related to the measurement construct. This uniqueness is referred to as method effects, which affects the construct validity of the instrument. This study examined the construct validity by exploring the existence of method effects in Tes PAPS UGM Seri E using Confirmatory Factor Analysis of 2.170 participants with bifactor model. Analysis results suggest that structure of Tes PAPS Seri E are fit with General Mental Ability Theory (*g factor*). There are no method effects with large portion in Tes PAPS Seri E was found. Almost all variations in the component scores are dominated by scores derived from General Reasoning factor rather than Measurement Media construct factor. Factor loading on components that dominated by Measurement Media construct are compared to these components' factor loading to General Reasoning. As a result, there were no significant differences found, so it can be concluded that the contribution of these components to General Reasoning is not affected by contribution of components to Measurement Media construct.

Tes potensi kognitif merupakan tes performansi terstandar yang telah digunakan di seluruh dunia untuk persyaratan admisi baik di bidang industri maupun institusi pendidikan. Tes potensi kognitif dipilih karena mampu menunjukkan perbedaan individual terkait dengan penyelesaian masalah atau tugas-tugas secara umum serta menjadi prediktor kompetensi dan performansi akademis seseorang (Kobrin et al., 2008; Kuncel & Hezlett, 2007; Kuncel et al., 2004). Banyak jenis Tes Potensi Akademik (TPA) yang digunakan di seluruh dunia seperti *Armed Services Vocational Aptitude Battery test (ASVAB)*, *General Aptitude Test Battery (GATB)*,



© GamaJOP 2024. Ini adalah artikel Akses Terbuka, didistribusikan berdasarkan ketentuan lisensi Creative Commons Attribution (<https://creativecommons.org/licenses/by-sa/4.0/>), yang mengizinkan penggunaan kembali, distribusi, dan reproduksi tanpa batas dalam media apa pun, asalkan karya aslinya dikutip dengan benar.

*Scholastic Assessment Test (SAT)*, *Graduate Record Examination (GRE)*, *Graduate Management Admission Test (GMAT)*, *Law School Admission Test (LSAT)*, TPA Bappenas, dan sebagainya. Secara umum, subtes dalam tes-tes tersebut terdiri dari penalaran verbal, penalaran kuantitatif, penalaran figural/spasial, dan subtes khusus seperti penulisan analitis pada GMAT dan LSAT atau pengetahuan elektronika pada ASVAB. Subtes umum seperti verbal, kualitatif, dan figural memiliki harga bobot faktor yang cukup besar terhadap faktor umum sehingga banyak digunakan di berbagai tes potensi akademik (Jensen, 1982). Salah satu tes potensi akademik yang digunakan di Indonesia adalah Tes Potensi Akademik Pascasarjana (PAPS).

Penelitian terkait evaluasi Tes PAPS UGM sudah beberapa kali dilakukan. Widhiarso (2017) melakukan pengujian ketepatan model PAPS (Penalaran) berbasis Rasch. Hasilnya menunjukkan bahwa semua butir memiliki ketepatan model. Widhiarso and Haryanta (2015) meneliti tentang sejauh mana komponen subtes Verbal (sinonim dan antonim) mengukur domain kemampuan verbal yang sama. Ditemukan bahwa stimulus butir soal antonim lebih memerlukan proses kognitif yang kompleks daripada butir soal sinonim. Meski begitu, kedua stimulus yang berbeda ini tetap menunjukkan bahwa domain kemampuan verbal yang diukur tetap sama sehingga tidak bisa dibedakan menjadi dua metode yang memiliki keunikan tersendiri. Penelitian lain yang dilakukan terhadap PAPS adalah pengujian dimensionalitas. Pattipeilohy (2017) melakukan penelitian dengan menggunakan skor komposit tes terhadap PAPS seri A1-R dan A2-R dengan menggunakan analisis faktor eksploratori. Penelitian ini menemukan bahwa meskipun PAPS memiliki ketepatan model yang tinggi jika diuji dengan model unidimensi namun memiliki kecenderungan untuk memiliki struktur multidimensi (dua dimensi). Faktor pertama berkaitan dengan kemampuan non-verbal dan faktor kedua cenderung berkaitan dengan kemampuan verbal. Namun demikian, kedua dimensi ini memiliki korelasi yang sangat kuat sehingga peneliti menyimpulkan bahwa terdapat satu konstruk ukur yang lebih umum yaitu penalaran, meskipun ada kecenderungan munculnya dua faktor. Dari berbagai penelitian tersebut validitas terkait dengan struktur tes PAPS perlu diidentifikasi lebih lanjut, terlebih jika sebuah tes telah mengalami perubahan struktur dari sisi komponen-komponen konten yang dimuat dalam tes (Reeves & Marbach-Ad, 2016).

Selama beberapa puluh tahun terakhir, ada beberapa jenis validitas dalam teori psikometri yaitu validitas konten, validitas kriteria, dan validitas konstruk (Newton & Shaw, 2014). Akan tetapi, perkembangan terbaru menjelaskan bahwa validitas merupakan sebuah bukti mengenai kualitas alat ukur berdasarkan derajat ketepatan skor yang dihasilkan alat ukur dalam menjelaskan tujuan alat ukur tersebut serta proses pengembangannya (Plake & Wise, 2014). Berdasarkan konsep tersebut, validitas kemudian dievaluasi ulang dan dibagi menjadi lima jenis yaitu validitas berbasis bukti konten tes, proses respons, struktur internal tes, konsekuensi tes, dan keterkaitan

dengan variabel-variabel lain. Penelitian ini menekankan pada validitas konstruk atau dalam terminologi terbaru disebut sebagai validitas berbasis pada bukti struktur internal tes (Sireci & Benítez Baena, 2023).

Validitas konstruk atau yang sekarang disebut dengan validitas berbasis pada bukti struktur tes dapat membuktikan derajat ketepatan skor suatu tes dalam mengukur suatu konstruk teoretis yang diukur (Fogarty, 2008; Lewis & Sireci, 2022). Validitas ini dapat dievaluasi menggunakan berbagai teknik statistik seperti melihat korelasi antar tes atau komponen tes (metode korelasional), mengidentifikasi rerata bobot faktor komponen tes (*Average Variance Extracted/AVE*), atau menggunakan metode Model Persamaan Struktural atau *Structural Equation Model/SEM* (Tavakol & Wetzel, 2020). SEM merupakan prosedur analisis yang kompleks karena tidak hanya mengevaluasi model pengukuran akan tetapi juga mampu mengevaluasi model struktural yang menunjukkan keterkaitan antara alat ukur yang dikaji dengan alat ukur lainnya (Gefen et al., 2000). Terdapat dua prosedur umum untuk mengevaluasi validitas konstruk di dalam SEM yaitu Analisis Faktor Konfirmatori (*Confirmatory Factor Analysis/CFA*) dan Analisis Faktor Eksploratori (*Exploratory Factor Analysis/EFA*) (Haig, 2022).

Setiap skor yang dihasilkan dari pengukuran mengandung informasi mengenai atribut yang diukur (varians terkait konstruk yang diukur) dan varians terkait dengan eror pengukuran. Eror pengukuran dapat bersifat sistematis atau non sistematis. Eror yang bersifat sistematis memiliki pola yang jelas sehingga dapat diidentifikasi penyebab kemunculannya pada pengukuran, sedangkan eror pengukuran non sistematis bersifat acak dan faktornya sulit diidentifikasi. Eror yang bersifat sistematis memiliki keragaman yang tidak dapat dijangkau oleh model. Terdapat dua penyebab munculnya varians eror sistematis yang dapat menjelaskan keragaman skor yang didapatkan oleh individu, yaitu varians terkait dengan konstruk ukur lain (*irrelevant construct*) dan varians terkait dengan metode (Viswanathan, 2005). Beberapa peneliti menamakan varians ini dengan nama konstruk metode (Vigneau & Cormier, 2008).

Konstruk metode adalah suatu konstruk yang muncul dari karakteristik butir-butir soal yang satu dengan lainnya memiliki kesamaan berdasarkan metode yang dipakai untuk mengukur (Ding et al., 2022; Podsakoff et al., 2012). Peranan konstruk metode ini disebut efek metode (*method effect*). Dalam hal ini, efek metode dapat bersifat positif karena sifatnya komplementer, namun dapat bersifat negatif karena meningkatkan ketidaktepatan pengukuran (Castille & Williams, 2022; Widhiarso & Hanifa, 2023). Furr and Bacharach (2013) berpendapat jika suatu pengukuran dilakukan dengan dua atau lebih metode dan memiliki korelasi dengan selain berdasarkan konstruk yang diukur dapat diasumsikan pengukuran terjangkit efek metode. Lebih sederhana, Podsakoff et al. (2003) menyebut munculnya varians sistematis yang bukan berasal dari konstruk ukur melainkan berasal dari metode pengukuran ini sebagai efek metode atau varians metode. Efek metode banyak dipengaruhi oleh banyak faktor, fak-

tor yang paling besar adalah format butir yang dipakai dalam suatu alat ukur (Widhiarso & Hanifa, 2023).

Metode atau media pengukuran didefinisikan Campbell and Fiske (1959) sebagai beberapa hal yang berada dalam suatu pengukuran sebagai bagian dari prosedur, berupa konten butir, format respons, instruksi umum dan fitur-fitur umum lain yang memengaruhi keseluruhan pengerjaan tes, karakteristik penguji, kondisi lingkungan tes, dan alasan subjek mengikuti tes. Sejalan dengan pendapat ini, Lance et al. (2009) menjelaskan metode sebagai pendekatan alternatif untuk menerjemahkan hasil pengukuran konstruk laten. Metode merupakan cara untuk mengukur, seperti kesamaan konten butir, struktur, atau format alat ukur yang memengaruhi kecenderungan subjek dalam memilih respons tertentu, kesesuaian arah pernyataan dengan konstruk (Marsh & Hocevar, 1988) ataupun sumber pengukuran (Urbina, 2004).

Penelitian ini mengeksplorasi sejauh mana komponen-komponen dalam Tes Potensi Akademik Pascasarjana (PAPS) Universitas Gadjah Mada memiliki keunikan yang tidak terkait dengan konstruk ukur. Keunikan tersebut dijelaskan sebagai efek metode karena bukan berasal dari konstruk Penalaran umum melainkan berasal dari prosedur pengukuran atau media pengukuran yaitu subtes-subtes dalam Tes PAPS seri E. Efek metode ditunjukkan oleh dominannya bobot faktor masing-masing komponen ke arah konstruk media pengukuran. Semakin besar efek metode maka semakin rendah ketepatan ukur dari Tes PAPS UGM. Jika teridentifikasi efek metode dalam tiap komponen, selanjutnya dilakukan analisis apakah efek metode tersebut bersifat komplementer dan menguatkan keberadaan konstruk penalaran umum, atau efek metode tersebut justru menguatkan keberadaan konstruk media pengukuran yaitu penalaran verbal, kuantitatif, maupun figural sehingga dianggap sebagai bias metode (Podsakoff et al., 2012). Konstruk laten menangkap varians sistematis yang bersumber dari perbedaan individual yang merepresentasikan atribut yang diukur. Jika varians metode sistematis tidak dikontrol, varians ini akan disatukan dengan varians sifat sistematis dalam konstruk. Hal ini tidak bisa dibiarkan karena dapat menyebabkan persepsi yang salah tentang kecukupan reliabilitas skala dan validitas konvergen.

Penelitian ini menggunakan analisis faktor konfirmatori (CFA) dengan menggunakan model bifaktor (*bifactor*). Model bifaktor merupakan CFA yang memiliki kelebihan mampu mendekomposisikan varians skor menjadi beberapa elemen tergantung dari sumbernya. Dalam hal ini sumber yang dimaksud adalah sumber yang berasal dari atribut yang diukur, metode yang dipakai untuk mengukur, serta eror pengukuran. Dalam beberapa model, metode tersebut dapat digantikan dengan waktu pengukuran (Eid, 2000).

Beberapa penelitian telah menggunakan model bifaktor untuk mengevaluasi validitas konstruk alat ukur yang mereka teliti. (Dombrowski et al., 2015) menggunakan model bifaktor untuk mengevaluasi validitas konstruk pada WISC (*Wechsler Intelligence Scale for Children*) Edisi Kelima. Hasil evaluasi ini mendukung keberadaan faktor

g, namun tidak ada bukti untuk membedakan faktor verbal, Gf (*Fluid Reasoning*), ataupun visual-spasial. Faktor g memberikan sumbangan pada sebagian besar varians total dan varians subtes pada WISC Edisi Kelima. (Urbán et al., 2014) menggunakan model bifaktor untuk mengevaluasi model pengukuran dan struktural dari *Derogatis Symptom Checklist* (SCL-90-R) dan *Brief Symptom Inventory* (BSI). Evaluasi ini menunjukkan bahwa model bifaktor mendukung konsep keparahan gejala global (*global symptom severity*) dan faktor gejala spesifik. Selain itu, penelitian yang dilakukan oleh (Jovanović, 2015) untuk mengevaluasi validitas struktural dari *Mental Health Continuum-Short Form* (MHC-SF) menunjukkan bahwa model bifaktor menghasilkan kecocokan terbaik untuk data di dua sampel. Hasil penelitian menunjukkan bahwa faktor umum kesejahteraan (*well-being*) memberikan sumbangan yang lebih besar daripada tiga faktor khusus pada kesejahteraan. Penelitian-penelitian ini menunjukkan bahwa model bifaktor dapat digunakan untuk mengevaluasi alat ukur psikologi dan mengonseptualisasikan konstruk psikologi, baik alat ukur yang bersifat klinis, skala sikap, maupun alat ukur kognitif (Luo & Al-Harbi, 2016).

Berdasarkan penjelasan tersebut maka model bifaktor memiliki kelebihan dibanding dengan metode CFA biasa yang hanya mampu mendekomposisikan varians skor berdasarkan varians konstruk yang diukur serta eror pengukuran (Wang et al., 2017). Model bifaktor dapat sangat berguna dalam membandingkan peranan sekumpulan faktor spesifik (misalnya metode) dengan peranan atribut ukur (faktor umum) dalam menghasilkan skor yang didapatkan individu. Sebuah alat ukur dapat dikatakan memiliki bukti kuat dalam hal validitas konstruk ketika porsi varians faktor umum yang diukur lebih besar dibanding dengan porsi varians yang bersumber dari metode (Vallevand et al., 2023). Selain harga varians, perbandingan juga dapat dilakukan dengan menggunakan bobot faktor (*factor loading*). Hal ini menunjukkan keunikan penelitian ini dibandingkan penelitian sebelumnya. Penelitian ini tidak hanya melakukan evaluasi alat ukur berdasarkan struktur konstruk latennya saja namun juga melihat seberapa besar peranan konstruk lain (konstruk metode) dalam memengaruhi ketepatan struktur pengukuran yang dimodelkan. Sebuah tes selalu berkaitan dengan metode pengukuran yang dilakukan. Apabila penggunaan metode tersebut justru memengaruhi jawaban yang diberikan oleh peserta tes, maka skor yang dihasilkan dari tes tersebut mengalami penurunan dari aspek validitasnya. Jika hal ini terjadi, varians skor lebih banyak berkaitan dengan aspek-aspek yang tidak berkaitan dengan kemampuan. Hasil dari penelitian ini diharapkan dapat menjadi evaluasi bagi pengembang alat ukur. Struktur pengukuran yang dikaji dalam penelitian ini adalah struktur pengukuran PAPS. Selama ini penelitian mengenai PAPS hanya memusatkan pada evaluasi struktur pengukuran yang menekankan pada efek dari konstruk (*construct effect*) dan belum ada yang melibatkan efek dari metode yang dipakai (*method effect*). Dengan menemukan peranan dari keduanya dalam satu model pengukuran, peneliti akan

dapat membandingkan efek mana yang dominan dalam mendukung perolehan skor oleh peserta tes.

## Metode

### Partisipan

Partisipan penelitian ini merupakan calon mahasiswa pascasarjana UGM yang mengikuti tes PAPS pada tahun 2017. Jumlah partisipan pada data yang dianalisis sebanyak 2.170, dengan rincian seri E1 sebanyak 1.083 dan seri E2 sebanyak 1.087. Data skor Tes PAPS ini merupakan data sekunder yang didapatkan dari Unit Pengembangan Alat Psikodiagnostika (UPAP) Fakultas Psikologi UGM sebagai lembaga resmi penyelenggara Tes PAPS.

### Instrumen Penelitian

Tes PAPS dikembangkan oleh Unit Pengembangan Alat Psikodiagnostika (UPAP) Fakultas Psikologi Universitas Gadjah Mada dan dikhususkan untuk seleksi penerimaan mahasiswa pascasarjana. Tes PAPS dirancang untuk mengungkap potensi keberhasilan seseorang jika mendapat kesempatan belajar lebih lanjut di jenjang pascasarjana (Widhiarso & Haryanta, 2015). Spesifikasi Tes PAPS serupa dengan GRE Test maupun SAT yang mengukur Penalaran umum. Tes ini awalnya memiliki 3 subtes yaitu subtes Verbal (V), subtes Kuantitatif (K), dan subtes Penalaran (P). Namun pada tahun 2016, tim pengembang Tes PAPS mengembangkan versi terbaru Tes PAPS dengan mengubah spesifikasi tes pada subtes ketiga. Subtes ketiga yang bernama subtes penalaran diubah menjadi subtes figural karena konstruk penalaran merupakan unsur dasar yang diukur dalam Tes PAPS UGM.

Setiap seri Tes PAPS UGM dikembangkan sebagai tes yang setara satu sama lain. Pada seri sebelumnya, Tes PAPS UGM memiliki reliabilitas yang memuaskan yaitu sebesar 0,92 (Belinda, 2015). Skor komponen Tes PAPS diperoleh dari menjumlahkan seluruh butir yang dijawab dengan benar pada setiap komponen. Jawaban benar akan mendapatkan skor 1 dan jawaban salah akan mendapatkan skor 0. Tidak ada pengurangan skor jika ada jawaban yang salah. Pada setiap soal terdapat satu jawaban benar di antara lima pilihan jawaban. Kemudian skor ini dikonversi menjadi skor standar dengan rata-rata sebesar 500 dan simpangan baku sebesar 100.

Versi terbaru PAPS dikenal dengan PAPS seri E dan mulai diadministrasikan pada tahun 2017. Pengujian terhadap validitas konstruk Tes PAPS Seri E ini belum banyak dilakukan. Padahal untuk disebut sebagai alat tes yang baik, pengujian validitas harus dilakukan secara berkala dan melihat dari berbagai aspek psikometris. Pengujian validitas merupakan proses yang terus menerus karena berbagai bukti empiris yang baru sangat mungkin mengalami perubahan. Proses pengujian ini dibutuhkan untuk mengonfirmasi apakah alat tes mengukur konstruk yang sesuai dengan tujuan awal penyusunannya. Penelitian ini menggunakan Seri E Tes PAPS UGM yang memuat 13 komponen tes yang dibagi ke dalam subtes verbal, subtes kuantitatif, dan subtes figural. Penelitian ini dilakukan pada seri ini karena seri ini merupakan versi

pengembangan terbaru yang merupakan perbaikan dari Tes PAPS di seri sebelumnya. Terdapat 50 butir soal dalam setiap subtes sehingga secara keseluruhan terdapat 150 butir soal dengan spesifikasi yang dijelaskan dalam Tabel 1.

**Tabel 1**

*Spesifikasi Tes PAPS UGM Seri E*

| Subtes      | Komponen                 | Jumlah Soal |
|-------------|--------------------------|-------------|
| Verbal      | Padanan Kata             | 15          |
|             | Lawan Kata               | 15          |
|             | Analogi Kata             | 15          |
|             | Analitis                 | 5           |
| Kuantitatif | Deretan Angka            | 10          |
|             | Aritmetika               | 10          |
|             | Konsep Aljabar           | 10          |
|             | Perbandingan Kuantitatif | 10          |
|             | Penalaran Geometrika     | 10          |
| Figural     | Diagram                  | 15          |
|             | Seriasi                  | 15          |
|             | Klasifikasi              | 10          |
|             | Analogi Gambar           | 10          |

### Prosedur Analisis Data

Analisis dalam penelitian ini menggunakan perangkat lunak MPLUS 7.0 untuk menganalisis model pengukuran dan model struktural dalam Tes PAPS UGM Seri E. Analisis yang digunakan adalah Analisis Faktor Konfirmatori (*Confirmatory Factor Analysis/CFA*) dengan model bifaktor. CFA digunakan untuk mengetahui kesesuaian struktur Tes PAPS seri E dengan data. Struktur Tes PAPS Seri E1 dan E2 secara teoretis diasumsikan memiliki satu dimensi pengukuran yaitu Penalaran umum atau Faktor *g*. Ketika komponen di dalam tes mengukur suatu konstruk di luar penalaran umum, maka dapat dikatakan terjadi ketidaktepatan model antara model teoretis dengan data lapangan.

Selain itu, CFA digunakan untuk melihat kontribusi setiap komponen tes (misalnya padanan kata, analogi kata, dst.) dalam menjelaskan konstruk yang hendak diukur, yaitu penalaran umum sekaligus melihat besarnya kontribusi skor setiap komponen yang mengarah ke konstruk media pengukuran. Model bifaktor akan membantu peneliti mendapatkan informasi secara simultan mengenai dua informasi: (a) besarnya nilai bobot faktor (*factor loading*) dari setiap komponen tes dalam menjelaskan konstruk penalaran umum dan (b) besarnya nilai bobot faktor dari komponen dalam menjelaskan media pengukuran.

Dalam penelitian ini, pengukuran yang terganggu oleh efek metode ditunjukkan oleh temuan ketika bobot faktor yang menjelaskan konstruk ukur nilainya lebih rendah dibanding dengan bobot faktor yang menjelaskan media pengukuran. Kondisi ini menunjukkan bahwa perolehan skor individu bukan lebih dominan disebabkan oleh

kemampuan utama yang diukur akan tetapi lebih kepada kemampuan lain yang berkaitan dengan metode pengukuran. Pada penelitian ini unit analisis yang menjadi pokok kajian adalah komponen tes sehingga ketika ditemukan sebuah komponen tes memiliki bobot faktor konstruk ukur yang lebih rendah daripada metode ukur, maka komponen tersebut disimpulkan pengukurannya terganggu oleh efek metode. Semakin besar selisih dari kedua bobot faktor tersebut (bobot faktor dari metode ukur lebih besar) menunjukkan semakin besar efek metode yang dihasilkan. Pada kondisi ini dapat disimpulkan bahwa pengukuran terhadap konstruk ukur melalui komponen tes tersebut sangat dipengaruhi oleh keragaman individual yang kurang terkait dengan konstruk ukur.

### Hasil

Tabel 2 menunjukkan hasil estimasi ketepatan model terhadap skor Tes PAPS Seri E1 dan E2 dengan menggunakan estimasi *Robust Maximum Likelihood* (MLR). MLR adalah parameter estimasi yang meminimalisasi dampak dari ketidaknormalan data (Enders, 2001). Penggunaan MLR ini meningkatkan akurasi Maximum Likelihood biasa karena MLR menyesuaikan ulang eror standar (SE). Kline (2011) merekomendasikan beberapa indeks ketepatan model yang mewakili masing-masing kelompok indeks ketepatan, yaitu Chi Kuadrat, *Comparative Fit Index* (CFI), *Goodness of Fit Index* (GFI), *Root Mean Squared Error of Approximation* (RMSEA), dan *Standardized Root Mean-squared Residual* (SRMR). Namun, Hu and Bentler (1999) tidak merekomendasikan penggunaan GFI pada ukuran sampel yang besar dan merekomendasikan penggunaan TLI (Tucker-Lewis Index) sebagai pelengkap. Penggunaan SRMR yang dilengkapi dengan CFI, TLI, dan RMSEA direkomendasikan dalam penelitian yang menggunakan estimasi (ML) agar mengakomodasi kemungkinan kesalahan spesifikasi model dan bobot faktor. Menurut Hu and Bentler (1999) data dapat dikatakan tepat atau sesuai dengan model jika nilai CFI dan TLI >0,90, serta memiliki nilai RMSEA dan SRMR <0,08. (Lihat Tabel 2)

Hasil estimasi ketepatan model terhadap skor Tes PAPS seri E1 ( $N = 1.083$ ) menunjukkan nilai Chi Square sebesar 173,466 dengan derajat kebebasan sebesar 52 dan penyesuaian MLR sebesar 0,989. Kemudian CFI dan TLI yang diperoleh sebesar 0,964 dan 0,945, RMSEA sebesar 0,046 dan SRMR sebesar 0,026. Hal yang serupa juga ditunjukkan oleh hasil estimasi ketepatan model terhadap skor Tes PAPS seri E2 ( $N = 1.087$ ). Nilai Chi Square yang diperoleh sebesar 155,698 dengan derajat kebebasan sebesar 52 dan penyesuaian MLR sebesar 0,966. Kemudian CFI dan TLI yang diperoleh sebesar 0,968 dan 0,952, RMSEA sebesar 0,043 dan SRMR sebesar 0,026. Berdasarkan indeks ketepatan model tersebut, dapat disimpulkan bahwa struktur Tes PAPS seri E1 dan seri E2 sudah sesuai dengan teori yang digunakan dalam pengembangan tes.

Hasil analisis faktor konfirmatori menunjukkan harga bobot faktor masing-masing komponen terhadap kon-

struk Penalaran umum dan konstruk media pengukuran pada Tes PAPS Seri E1 seperti pada Gambar 1. Komponen Aritmetika menunjukkan kontribusi terbesar terhadap Penalaran umum dengan harga bobot faktor sebesar 0,715. Hal ini menunjukkan bahwa komponen Aritmetika merupakan indikator yang kuat dalam mengukur Penalaran umum. Sedangkan komponen padanan kata menunjukkan harga bobot faktor paling kecil sebesar 0,315 terhadap Penalaran umum sehingga dapat dikatakan komponen padanan kata kurang baik dalam mengukur Penalaran umum.

Tes PAPS Seri E1 memiliki tujuh komponen yang terindikasi memiliki efek metode karena besarnya harga bobot faktor ke arah konstruk media pengukuran lebih dari 0,30. Komponen-komponen tersebut adalah Padanan Kata, Lawan Kata, Konsep Aljabar, Perbandingan Kuantitatif, Penalaran Geometrika, Klasifikasi, dan Analogi Gambar. Namun, dari ketujuh komponen ini hanya terdapat tiga komponen yang memiliki harga bobot faktor ke arah konstruk media pengukuran lebih besar daripada ke arah konstruk Penalaran umum. Tiga komponen tersebut adalah Padanan Kata, Penalaran Geometrika, dan Klasifikasi.

Gambar 2 menunjukkan harga bobot faktor masing-masing komponen terhadap konstruk Penalaran umum dan konstruk media pengukuran pada Tes PAPS Seri E2. Komponen Aritmetika menunjukkan kontribusi terbesar terhadap Penalaran umum dengan harga bobot faktor sebesar 0,718 sehingga dapat dikatakan sebagai indikator yang kuat dalam mengukur Penalaran umum. Hal ini menunjukkan hal yang sama dengan Tes PAPS seri E1. Komponen padanan kata menunjukkan harga bobot faktor paling kecil sebesar 0,315 terhadap Penalaran umum sehingga dapat dikatakan komponen padanan kata kurang berkontribusi secara maksimal dalam mengukur Penalaran umum. (Lihat Gambar 1) dan (Gambar 2)

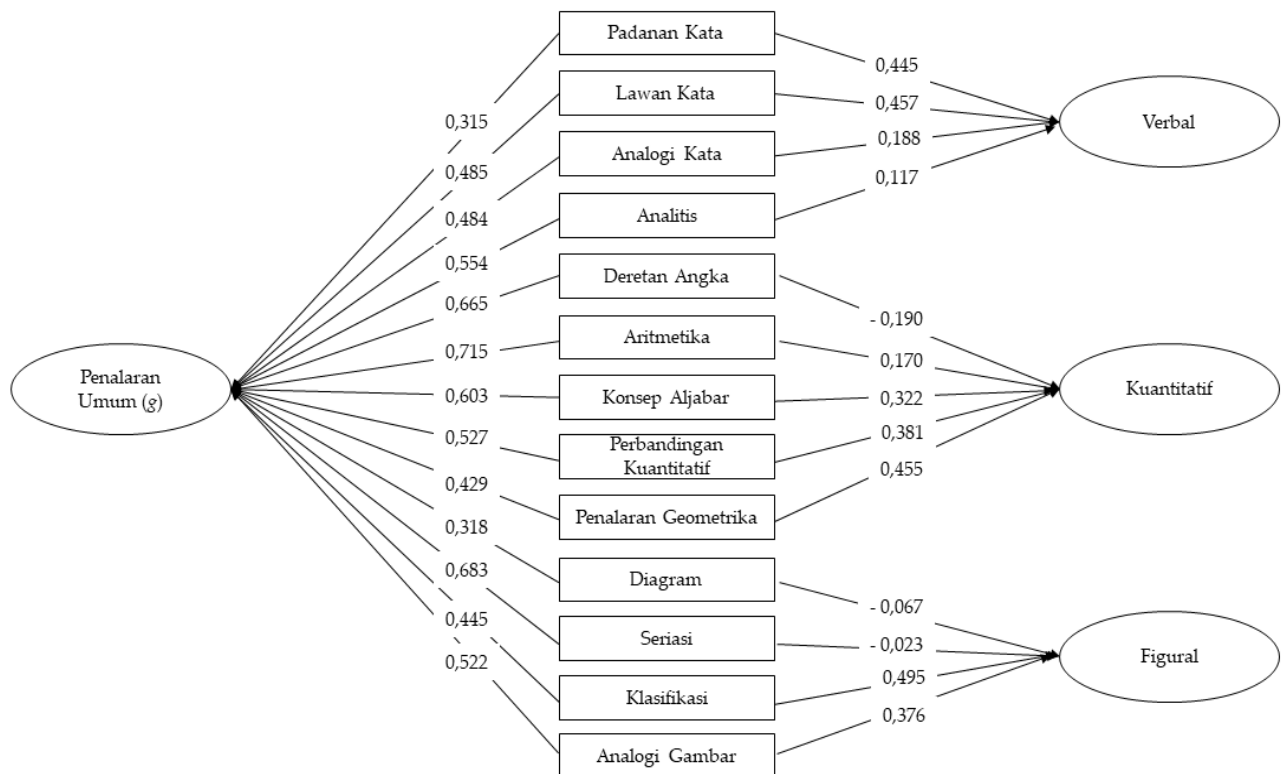
Serupa dengan Tes PAPS Seri E1, pada Tes PAPS Seri E2 terdapat dua komponen yang terindikasi memiliki efek metode Penalaran Geometrika dan Klasifikasi Gambar. Kemudian, terdapat tiga komponen yang memiliki harga bobot faktor ke arah konstruk media pengukuran lebih besar daripada ke arah konstruk Penalaran umum yaitu Padanan Kata, Penalaran Geometrika, dan Klasifikasi. Berbeda dengan seri E1, pada seri E2 terdapat satu komponen yang harga bobot faktornya tidak signifikan secara statistik baik ke arah konstruk Penalaran umum maupun konstruk media pengukuran yaitu komponen Diagram (bobot faktor = 0,294;  $p < 0,05$ ). Berdasarkan informasi mengenai perbandingan harga bobot faktor yang mengarah pada konstruk Penalaran umum dan konstruk media pengukuran, dapat disimpulkan bahwa efek metode muncul pada sebagian kecil komponen Tes PAPS UGM Seri E1 dan E2.

Sebagai analisis tambahan, dilakukan analisis statistik deskriptif untuk melihat tingkat kesulitan setiap komponen. Tingkat kesulitan sebuah butir soal berkaitan

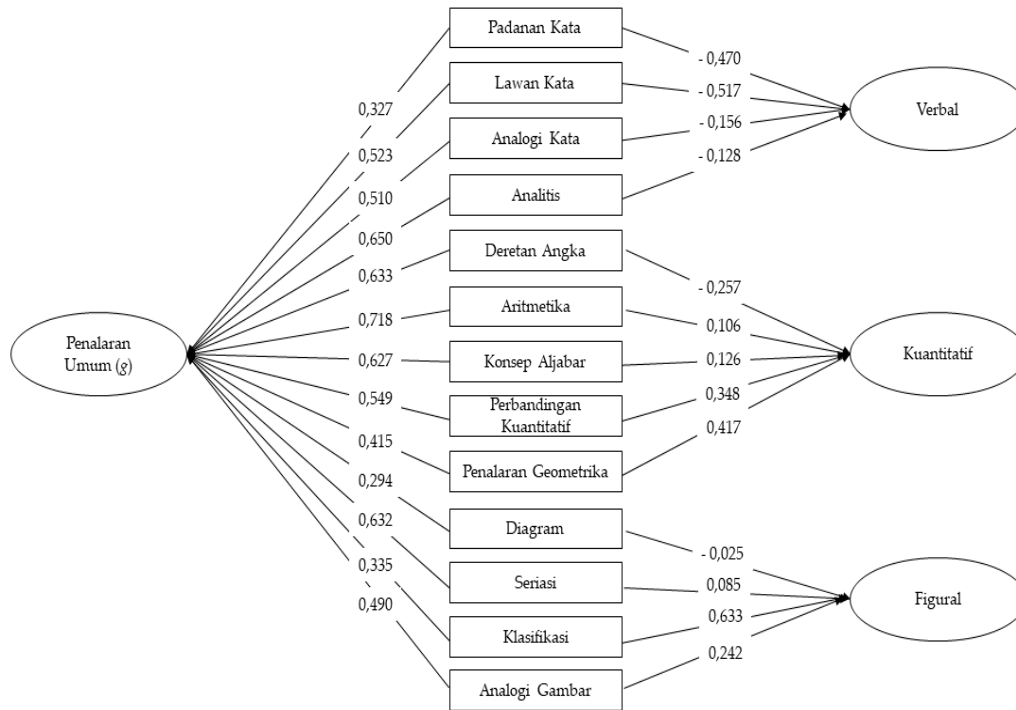
**Tabel 2**  
*Indeks Ketepatan Model pada Tes PAPS Seri E1*

| Indeks Ketepatan Model            | Nilai Kritis | Hasil Model Seri E1 | Hasil Model Seri E2 |
|-----------------------------------|--------------|---------------------|---------------------|
| Chi Square                        | -            | 173,466             | 155,698             |
| Degrees of Freedom                | -            | 52                  | 52                  |
| p-Value                           | -            | 0,000               | 0,000               |
| Scaling Correction Factor for MLR | -            | 0,989               | 0,966               |
| CFI                               | >0,90        | 0,964               | 0,968               |
| TLI                               | >0,90        | 0,945               | 0,952               |
| RMSEA                             | <0,08        | 0,046               | 0,043               |
| SRMR                              | <0,08        | 0,026               | 0,026               |

**Gambar 1**  
*Hasil CFA dengan model bifaktor terhadap Tes PAPS Seri E1*



**Gambar 2**  
 Hasil CFA dengan model bifaktor terhadap Tes PAPS Seri E2



dengan varians skor yang dihasilkan dari butir tersebut. Butir yang memiliki tingkat kesulitan yang optimal akan menghasilkan varians skor yang lebih besar dibanding dengan butir yang memiliki tingkat kesulitan sangat rendah atau sangat tinggi. Selanjutnya, varians skor ini nantinya akan memengaruhi seberapa besar porsi varians utama dari sebuah skor, yaitu varians yang bersumber dari kemampuan peserta tes. Hasil analisis menunjukkan bahwa tingkat kesulitan tidak memiliki keterkaitan dengan kemunculan efek metode. Efek metode dalam penelitian ini muncul baik pada komponen tes yang memiliki tingkat kesulitan rendah maupun tinggi. Pada Tabel 3, statistik deskriptif terhadap 1.083 sampel pada Tes PAPS Seri E1 menunjukkan komponen dengan tingkat kesulitan terendah adalah Analitis (rerata skor = 4,00; persen = 0,80) dan Deretan Angka (rerata skor = 7,52; persen = 0,75). Skor rata-rata individu dalam komponen Analitis adalah 4,00 dengan persentase individu yang menjawab benar sebesar 80%. Sementara itu, skor rata-rata individu dalam komponen Deretan angka adalah 7,52 dengan persentase individu yang menjawab benar sebesar 75%. Komponen dengan tingkat kesulitan paling tinggi adalah Penalaran Geometrika (rerata skor = 3,18; persen = 0,32) dan Diagram (rerata skor = 5,42; persen = 0,36). Penjelasan mengenai harga statistik ini seterusnya sama dengan penjelasan sebelumnya. (Lihat tabel 3) Serupa dengan Tes PAPS Seri E1, pada Tes PAPS Seri E2 ini komponen dengan tingkat kesulitan terendah adalah Deretan Angka ( $M = 6,57; p = 0,66$ ) dan Lawan Kata ( $M = 9,39; p =$

**Tabel 3**  
 Statistik Deskriptif Tes PAPS UGM Seri E1 dan E2

| Komponen                 | Seri E1 |            | Seri E2 |            |
|--------------------------|---------|------------|---------|------------|
|                          | Rerata  | Persentase | Rerata  | Persentase |
| Padanan Kata             | 7,33    | 0,49       | 7,41    | 0,49       |
| Lawan Kata               | 9,00    | 0,60       | 9,39    | 0,63       |
| Analogi Kata             | 7,37    | 0,49       | 7,76    | 0,52       |
| Analitis                 | 4,00    | 0,80       | 2,61    | 0,52       |
| Deretan Angka            | 7,52    | 0,75       | 6,57    | 0,66       |
| Aritmetika               | 5,34    | 0,53       | 5,04    | 0,50       |
| Konsep Aljabar           | 3,95    | 0,40       | 4,51    | 0,45       |
| Perbandingan Kuantitatif | 4,92    | 0,49       | 4,33    | 0,43       |
| Penalaran Geometrika     | 3,18    | 0,32       | 2,87    | 0,29       |
| Diagram                  | 5,42    | 0,36       | 5,33    | 0,36       |
| Seriasi                  | 8,21    | 0,55       | 8,74    | 0,58       |
| Klasifikasi              | 6,02    | 0,60       | 5,52    | 0,55       |
| Analogi Gambar           | 6,30    | 0,63       | 3,87    | 0,39       |

0,63). Serupa dengan Seri E1, komponen dengan tingkat kesulitan tertinggi pada Seri E2 adalah Penalaran Geometrika ( $M = 2,87$ ;  $p = 0,36$ ) dan Diagram ( $M = 5,33$ ;  $= 0,36$ ).

### Diskusi

Berdasarkan hasil analisis dengan *Confirmatory Factor Analysis* (CFA) yang menggunakan model bifaktor didapatkan temuan bahwa secara umum tidak terdapat efek metode dengan porsi yang cukup besar di dalam konstruk media pengukuran PAPS Seri E1 dan E2. Hampir semua variasi di dalam skor komponen lebih didominasi oleh skor yang bersumber dari Penalaran umum daripada konstruk media pengukuran yang diwakili oleh jenis-jenis subtes pada Tes PAPS.

Secara umum, pada kedua seri PAPS di tiga komponen tes yaitu Padanan Kata, Penalaran Geometrika, dan Klasifikasi Gambar ditemukan harga bobot faktor yang signifikan ke arah konstruk media pengukuran dan lebih besar daripada ke arah konstruk Penalaran umum. Namun, setelah harga bobot faktor tersebut dibandingkan dengan harga bobot faktor yang mengarah ke konstruk Penalaran umum, tidak ditemukan adanya perbedaan yang signifikan ( $p < 0,050$ ). Artinya, sumbangan komponen-komponen tersebut ke konstruk Penalaran umum tidak dipengaruhi oleh sumbangan komponen ke konstruk media pengukuran.

Selanjutnya, perlu dilakukan eksplorasi lebih lanjut penyebab tiga komponen pada tes PAPS Seri E dapat terjangkau efek metode dengan melihat contoh butir soal yang disajikan. Pertama, komponen Padanan Kata memiliki kontribusi besar terhadap konstruk Penalaran Verbal karena soal dalam padanan kata memiliki bentuk yang sederhana sehingga hanya membutuhkan pengetahuan terhadap suatu kata. Komponen Padanan Kata pada Tes PAPS Seri E secara umum menggunakan kata dalam konteks percakapan sehari-hari yang serupa dengan komponen Lawan Kata. Namun, komponen Lawan Kata memiliki performa yang lebih baik dalam mengukur Penalaran umum karena peserta harus melakukan proses penalaran yang lebih panjang dalam memilih jawaban yang terhubung secara konteks dengan soal yang disajikan (Widhiarso & Haryanta, 2015). Sedangkan butir dalam komponen Padanan Kata hanya perlu memilih jawaban yang menurut peserta paling mendekati konteks kata pada soal berdasarkan pengetahuan peserta. Contoh butir soal Padanan Kata dapat dilihat pada Gambar 3

Besarnya kontribusi komponen Penalaran Geometrika terhadap konstruk Penalaran Kuantitatif dapat dijelaskan oleh beberapa hal berikut. Pertama, soal dalam komponen Penalaran Geometrika mengombinasikan penalaran ruang dan penalaran matematis sehingga memerlukan persepsi statis dan dinamis dalam memahami informasi spasial (Hart et al., 2018). Selain itu, dilihat dari contoh butir pada (Lihat Gambar 4), soal Penalaran Geometrika

**Gambar 3**

Contoh Butir Soal Padanan Kata

|   |
|---|
| <p>Instruksi:<br/>Soal berikut terdiri dari satu kata yang dicetak dengan huruf kapital dan diikuti oleh lima pilihan jawaban. Pilihlah satu di antara pilihan jawaban yang mempunyai ARTI SAMA atau PALING DEKAT dengan arti kata yang dicetak dengan huruf kapital.</p> |
| <p>BUSUNG</p> <p>(A) lapar<br/>(B) besar<br/>(C) kosong<br/>(D) timbunan<br/>(E) gembung</p>  |

dalam Tes PAPS Seri E ini juga memerlukan pengetahuan rumus matematika dari bangun datar maupun bangun ruang yang spesifik sehingga tugas kognitif yang diperlukan berbeda dengan penalaran karena peserta harus menghafal rumus-rumus yang mungkin diperlukan dalam menyelesaikan soal. Sehingga, komponen Penalaran Geometrika tidak hanya mengukur Penalaran umum, namun juga memerlukan pengetahuan dan kemampuan mengingat rumus-rumus matematika lanjut. Menurut Primi (2001) tingkat kesulitan pada soal-soal Geometrika tergantung pada empat hal yaitu jumlah gambar yang digunakan, jumlah aturan atau rumus yang harus digunakan menyelesaikan tugas dari stimulus gambar yang diberikan, kompleksitas aturan, dan kompleksitas persepsi dari stimulus yang diberikan. (Lihat Gambar 4)

**Gambar 4**

Contoh Butir Soal Penalaran Geometrika

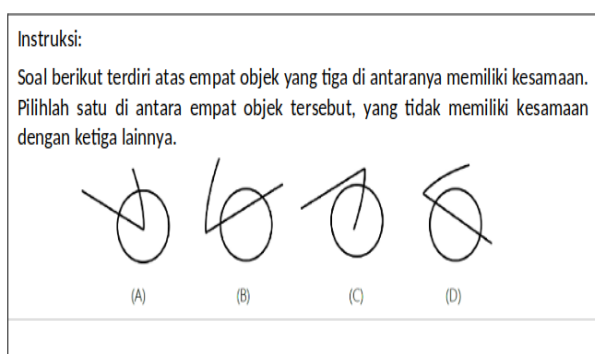
|   |
|---|
| <p>Instruksi:<br/>Pilihlah satu jawaban yang tepat dari lima pilihan yang disediakan.</p>   |
| <p>Sebuah bujur sangkar memiliki panjang sisi sebesar <math>\sqrt{7}</math> cm. Luas lingkaran maksimal yang dapat masuk ke dalam bujur sangkar tersebut adalah...</p> <p>(A) 7<br/>(B) 22<br/>(C) 107<br/>(D) 154<br/>(E) 1078</p> |

Komponen Klasifikasi memiliki kontribusi yang besar terhadap konstruk penalaran figural karena sifat soalnya yang membutuhkan kemampuan konsep spasial yang kompleks dalam mencari persamaan dan perbedaan dari pilihan jawaban yang tersedia. Hasil penelitian ini sejalan dengan penelitian Johnson dan kolega (2004) bahwa bentuk soal yang melibatkan rotasi figural secara mental maupun soal spasial secara umum memang memiliki



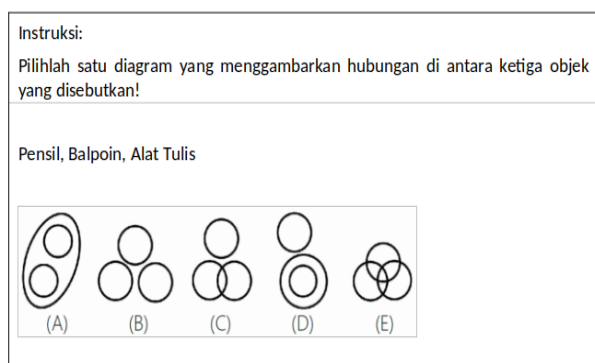
harga bobot faktor ke faktor  $g$  lebih kecil daripada komponen lain. Selain itu, dilihat dari contoh butir pada Gambar 5, soal Klasifikasi pada Tes PAPS UGM seri E dapat menimbulkan interpretasi yang bersifat subjektif dalam mengelompokkan objek-objek pada pilihan jawaban dan lebih membutuhkan ketelitian daripada penalaran dalam melihat bentuk-bentuk yang disajikan pada pilihan jawaban.

**Gambar 5**  
Contoh Butir Soal Klasifikasi



Komponen Diagram memiliki kontribusi yang kecil baik ke konstruk Penalaran umum maupun Penalaran Figural. Dilihat dari contoh soal pada Gambar 6, interpretasi terhadap butir-butir soal dalam komponen diagram dapat bersifat subjektif berdasarkan pengetahuan dan pengalaman subjek terhadap objek yang disebutkan. Di sisi lain, pengalaman dan wawasan partisipan terhadap objek-objek tersebut sangat memengaruhi jawaban. Butir soal pada komponen Diagram juga tidak memiliki konteks spasial. Kondisi ini membuat komponen Diagram pada Tes PAPS Seri E2 menjadi kurang optimal dalam menjelaskan konstruk Penalaran umum dan Penalaran Figural. Lihat Gambar 6

**Gambar 6**  
Contoh butir soal Diagram



Temuan lainnya dalam penelitian ini adalah komponen Aritmetika pada Tes PAPS Seri E1 dan E2 konsisten sebagai komponen dengan kontribusi terbesar ke konstruk Penalaran umum dengan harga bobot faktor sebesar 0,715 dan 0,718. Hal ini konsisten dengan hasil penelitian Tirre dan Pena (1993) bahwa komponen Penalaran Aritmetika (*Arithmetic Reasoning*) dan Pengetahuan Matematika (*Math Knowledge*) pada tes ASVAB (*Armed Services Vocational Aptitude Battery*) menjadi jenis komponen yang berkontribusi besar terhadap konstruk Penalaran umum.

Hasil penelitian ini mendukung bahwa terdapat efek metode pada sebagian komponen pada Tes PAPS UGM seri E. Beberapa komponen terbukti memiliki skor yang tidak hanya menjelaskan konstruk Penalaran umum, tapi juga menjelaskan konstruk media pengukuran yaitu Subtes pada Tes PAPS UGM Seri E. Namun setelah dieksplorasi lebih lanjut, keunikan-keunikan skor tersebut bersifat komplementer karena tidak adanya perbedaan signifikan antara harga bobot faktor ke dua konstruk tersebut. Ketepatan model pada Tes PAPS seri E1 dan E2 pun sangat memuaskan dan tidak begitu terpengaruh oleh munculnya efek metode, sehingga efek metode ini tidak dapat dikategorikan sebagai bias metode.

Analisis tambahan dilakukan dengan melihat tingkat kesulitan komponen. Terdapat beberapa kesamaan dan perbedaan di antara Seri E1 dan E2. Pada Seri E1 dan E2, komponen Penalaran Geometrika dan Diagram merupakan komponen dengan kesulitan tertinggi. Perbedaan dapat ditemukan pada komponen Penalaran Geometrika, Analitis, dan Deretan Angka. Pada Seri E2, hasil analisis menunjukkan komponen Penalaran Geometrika memiliki tingkat kesulitan yang tinggi (persen  $< 0,30$ ), sedangkan pada seri E1 komponen ini memiliki tingkat kesulitan sedang. Kemudian, pada Seri E1 ditemukan bahwa komponen Analitis dan Deretan Angka merupakan komponen dengan tingkat kesulitan rendah. Sedangkan pada Seri E2 tidak terdapat komponen dengan tingkat kesulitan rendah (persen  $> 0,70$ ). Selain itu, pada seri E1 tidak terdapat komponen dengan kesulitan tinggi.

Tingkat kesulitan dapat dipengaruhi oleh bentuk soal yang disajikan. Penyelesaian soal yang kompleks akan memerlukan kemampuan yang berbeda dengan penyelesaian soal yang hanya perlu memilih di antara alternatif jawaban yang disediakan (Ward, 1982). Komponen Analitis dan Deretan Angka menjadi komponen dengan tingkat kesulitan terendah karena soal dalam kedua komponen ini menyediakan informasi-informasi awal yang dapat digunakan oleh partisipan untuk menjawab pertanyaan yang diberikan sehingga peserta tinggal memilih jawaban yang memenuhi informasi-informasi tersebut. Sedangkan komponen Penalaran Geometrika dan Diagram menjadi komponen dengan tingkat kesulitan tertinggi karena soal dalam kedua komponen ini tidak menyediakan informasi awal baik berupa fakta atau aturan pola, sehingga partisipan harus memecahkan soal dengan analisis yang lebih mendalam. Selain itu, pemecahan soal pada komponen Penalaran Geometrika melibatkan rumus dan prinsip bangun datar serta bangun ruang, sedangkan komponen

Diagram melibatkan perspektif individu dalam menghubungkan objek-objek yang disebutkan. Menurut Jensen (1982), kompleksitas atau kesulitan butir soal merupakan hal yang secara langsung dapat memengaruhi pengukuran terhadap  $g$ . Semakin sulit suatu butir soal, maka semakin baik dalam mengukur  $g$ . Namun, temuan penelitian ini membantah hal tersebut. Komponen Penalaran Geometrika dan Diagram sebagai dua komponen tersulit di kedua Seri Tes PAPS menunjukkan bobot faktor yang rendah ke arah Penalaran umum. Berdasarkan temuan yang ada, pengukuran yang tepat pada faktor  $g$  ditunjukkan oleh bentuk soal yang memerlukan penalaran atau proses kognitif yang lebih tinggi. Ini didukung temuan bahwa penggunaan kemampuan manipulasi mental dalam penyelesaian soal menunjukkan ketepatan pengukuran pada faktor  $g$  (Gottfredson, 2002).

Berdasarkan pengujian ketepatan model, penelitian ini juga mendukung keberadaan faktor  $g$  yang telah bertahun-tahun diperdebatkan. Tes PAPS Seri E menunjukkan unidimensionalitas sehingga seluruh komponen didasari satu faktor yang sama. Menurut Spearman (1904) terdapat satu fungsi pokok yang dapat mengonseptualisasikan seluruh cabang kemampuan mental manusia yang disebut dengan *General Mental Ability* atau faktor  $g$  (Penalaran umum). Mendukung pernyataan ini Jensen (1998) menjelaskan bahwa faktor  $g$  merupakan satu faktor yang mendukung individu dalam menyelesaikan berbagai macam tugas yang berhubungan dengan kemampuan mental umum. Kemampuan mental umum dapat berupa kemampuan belajar, pemahaman terhadap instruksi atau bagaimana individu menyelesaikan masalah.

Faktor  $g$  adalah kemampuan yang bersifat umum, maka dari itu tingkat atau jenis pendidikan tidak berpengaruh terhadap kemampuan umum (Jensen, 1982). Seseorang yang memiliki kemampuan  $g$  yang tinggi akan memiliki performa dan keahlian penyelesaian tugas yang lebih baik dibandingkan individu lain yang memiliki spesifikasi pendidikan yang sama dengannya. Maka, Tes PAPS UGM Seri E sebagai tes yang mengukur kemampuan umum sudah sesuai dengan tujuan pembuatannya yaitu agar calon mahasiswa dengan potensi berkembang terbaik dalam bidang akademik dapat memperoleh kesempatan belajar lebih lanjut di jenjang pascasarjana terlepas dari latar belakang pendidikannya serta dapat digunakan untuk syarat admisi lintas disiplin umum.

Jika dilihat dari perspektif model CHC (Cattell-Horn-Carroll) yang merupakan pengembangan terkini dari teori GMA dari Spearman, secara umum Tes PAPS dapat dikategorikan sebagai tes yang mengukur  $Gf$  (*Fluid Reasoning*). Beberapa penelitian telah membuktikan secara statistik,  $Gf$  dan  $g$  memiliki performa yang identik sehingga sulit untuk dipisahkan (Gustafsson, 1984) (Gustafsson, 1984). Model CHC merupakan model inteligensi hierarkis yang menggambarkan kemampuan kognitif dalam tiga strata (Taub & McGrew, 2004). CHC memiliki lebih dari 80 kemampuan terbatas (*narrow*) pada strata 1 dan 16 kemampuan luas (*broad*) pada strata 2 yang dikembangkan dari teori faktor  $g$  Spearman, teori  $Gf$ - $Gc$  Cattell-Horn, dan teori strata Carroll (Flanagan & Dixon, 2014).

Kedekatan antara bentuk soal pada komponen-komponen tes PAPS dengan kemampuan-kemampuan pada struktur CHC mendukung pengembangan tes PAPS ke arah model CHC. Komponen yang terindikasi efek metode memiliki kedekatan bentuk dengan kemampuan selain  $Gf$  pada CHC. Seperti komponen Padanan Kata memiliki kedekatan dengan kemampuan  $Gc$  (*Comprehension-Knowledge*).  $Gc$  pada CHC merupakan kemampuan yang muncul karena proses belajar, budaya, dan pengalaman hidup (Horn & Blankson, 2005). Komponen penalaran Geometrika memiliki kedekatan dengan kemampuan  $Gq$  (*Quantitative Knowledge*).  $Gq$  menunjukkan kemampuan menggunakan informasi kuantitatif dan memanipulasi simbol numerik untuk menyelesaikan masalah (Flanagan & Dixon, 2014). Komponen Klasifikasi memiliki kedekatan dengan kemampuan  $Gv$  (*Visual Processing*).  $Gv$  merupakan kemampuan menggunakan informasi gambar dan spasial untuk menyelesaikan masalah (Schneider & McGrew, 2012). Pengembangan ke arah model CHC ini masih perlu evaluasi lebih lanjut pada komponen-komponen lain pada Tes PAPS UGM dengan memperhatikan kemampuan terbatas dan kemampuan luas pada CHC.

Hasil analisis ini juga menunjukkan model bifaktor dapat digunakan untuk menjelaskan keragaman data pada Tes PAPS Seri E. Kesimpulan ini didapatkan dari Indeks Ketepatan Model yang seluruhnya memenuhi kriteria. Hasil ini konsisten baik pada Seri E1 maupun Seri E2. Model bifaktor memiliki keunggulan dibandingkan model CFA bertingkat dalam menjelaskan Tes PAPS UGM Seri E. Pertama, model bifaktor mengakomodasi sumber-sumber skor dari dua tipe faktor yang berbeda, dalam hal ini adalah konstruk Penalaran umum dan konstruk media pengukuran. Kedua, model bifaktor dapat membandingkan kontribusi masing-masing faktor terhadap skor yang didapatkan oleh individu. Ketiga, model ini cocok untuk menjelaskan Tes PAPS UGM karena tes PAPS UGM melibatkan satu konstruk ukur dan beberapa konstruk metode. Keempat, model bifaktor dapat menunjukkan varians dari konstruk umum dan varians dari konstruk-konstruk khusus secara simultan sehingga memberikan dasar yang kuat untuk memahami konstruk psikologi yang diukur dan pengukuran yang tepat (Reise, 2012).

Model bifaktor juga dapat mengakomodasi pengembangan lanjutan pada tes PAPS UGM Seri E. Komponen dengan harga bobot faktor yang tinggi ke arah konstruk media pengukuran dapat dikembangkan menjadi model CHC seperti yang telah dijelaskan pada penjelasan sebelumnya atau menjadi alat ukur kemampuan khusus. Misalnya, komponen dengan harga bobot faktor yang besar ke arah Kuantitatif dapat dikembangkan lebih lanjut menjadi alat ukur untuk evaluasi kemampuan matematis. Begitu pula dengan konstruk lain.

Keterbatasan pada penelitian ini adalah CFA model Bifaktor tidak mengakomodasi korelasi antar media pengukuran. Namun, terdapat model lain yang dapat mengetahui faktor korelasi tersebut. Misalnya, model CFA-CTC(M-1)<sub>k</sub> atau *Confirmatory Factor Analysis - Correlated Trait Correlated* (Method Minus 1)<sub>k</sub> di mana  $k$

adalah metode yang dipilih sebagai standar pembandingan (Eid, 2000), Kelebihan model CFA-CTC(M-1)k selain dapat menunjukkan korelasi antar metode, model ini mampu mengetimasi komponen varians karena efek konstruk ukur, efek konstruk metode, dan efek residual.

### Kesimpulan

Berdasarkan hasil Analisis Faktor Konfirmatori, terdapat beberapa komponen yang terindikasi memiliki efek metode. Analisis lebih lanjut menunjukkan bahwa komponen tersebut memiliki harga bobot faktor yang signifikan ke arah konstruk media pengukuran dan lebih besar daripada harga bobot faktor ke arah konstruk Penalaran umum. Setelah dieksplorasi lebih lanjut, sumbangan ketiga komponen tersebut ke arah konstruk Penalaran umum tidak dipengaruhi sumbangan ke arah konstruk media pengukuran. Selain itu, indeks ketepatan model pada struktur Tes PAPS Seri E menunjukkan bahwa Seri E1 dan E2 sudah sesuai dengan teori Kemampuan Mental Umum (*g factor*). Dilihat dari dua bukti tersebut dapat disimpulkan bahwa efek metode yang muncul pada Tes PAPS Seri E tidak dapat dikategorikan sebagai bias metode.

Temuan lain dari hasil penelitian ini di antaranya adalah dilihat dari tingkat kesulitan komponen soal, terdapat persamaan dan perbedaan antara Tes PAPS UGM Seri E1 dan E2. Selain itu, penelitian ini menunjukkan bahwa semakin kompleks atau sulit suatu tes tidak berarti tes tersebut semakin tepat dalam mengukur faktor *g*. Hal yang memengaruhi ketepatan pengukuran terhadap faktor *g* adalah tingkat proses kognitif atau penalaran yang diperlukan dalam memecahkan permasalahan dalam butir soal.

### Saran

Peneliti merekomendasikan adanya perbaikan butir soal pada komponen yang memiliki efek metode agar Tes PAPS UGM Seri E dapat memiliki performa yang lebih baik dalam mengukur Penalaran umum. Terdapat beberapa alternatif perbaikan yang dapat dilakukan pada komponen Padanan Kata, Penalaran Geometrika, dan Klasifikasi. Pada komponen Padanan Kata dapat diberikan butir soal berupa menemukan padanan kata yang sesuai untuk konteks yang disajikan. Konteks disajikan berupa sebuah kalimat dengan satu kata yang disorot, kemudian peserta diminta memberikan sebuah kata yang sesuai dengan konten kata yang disorot tersebut. Pada komponen Penalaran Geometrika, terkait butir soal dengan rumus matematika lanjut dapat disajikan rumus dalam butir soal sehingga peserta hanya perlu memecahkan soal tanpa perlu menghafal rumus tertentu. Pada komponen Klasifikasi dapat diberikan suatu pola visual bangun datar yang melibatkan warna seperti pada soal komponen Seriasi agar peserta tidak perlu memecahkan soal dengan ketelitian sudut atau jarak pada stimulus yang disajikan. Alternatif alternatif perbaikan tersebut diharapkan dapat meningkatkan ketepatan pengukuran pada Tes PAPS Seri E.

Peneliti juga menemukan bahwa kemampuan model bifaktor yang sangat baik dalam menjelaskan Tes PAPS

UGM Seri E. Penggunaan model bifaktor juga mendukung pengembangan Tes PAPS UGM Seri E untuk penyusunan alat ukur dengan model CHC maupun alat ukur kemampuan yang lebih spesifik. Kemudian, komponen Aritmetika konsisten sebagai komponen yang sangat baik dalam menjelaskan Penalaran umum dan komponen Diagram tidak memiliki sumbangan yang signifikan dalam menjelaskan Penalaran umum dan media pengukuran pada Tes PAPS UGM Seri E2. Temuan-temuan tambahan ini dapat dieksplorasi lebih lanjut dalam penelitian selanjutnya.

### Pernyataan

#### Ucapan Terima Kasih

Peneliti mengucapkan terima kasih kepada Fakultas Psikologi UGM yang memberikan hibah untuk mendukung penelitian ini dan kepada Unit Pengembangan Alat Psikodiagnostika (UPAP) yang telah menyediakan data

### Pendanaan

Penyusunan tulisan ini merupakan bagian dari keluaran kegiatan dari hibah penelitian yang disediakan oleh Fakultas Psikologi UGM


### Kontribusi Penulis

Penulis pertama menyiapkan dan mereviu manuskrip publikasi dan analisis data, penulis kedua menyiapkan manuskrip, mengeksplorasi landasan teoretis dan penyiapan data

### Konflik Kepentingan

Peneliti menyatakan bahwa tidak ada konflik kepentingan dalam melakukan penelitian maupun menyiapkan manuskrip publikasi

### Orcid ID

Wahyu Widhiarso  <https://orcid.org/0000-0002-0187-6536>  
Alifa Rahmi Khairunisa

### Pustaka

- Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*(2), 81–105. <https://doi.org/10.1037/h0046016>
- Castille, C., & Williams, L. J. (2022). To Partial or Not? Re-Examining the Unmeasured Latent Method Construct (ULMC). *Academy of Management Proceedings*, *2022*(1). <https://doi.org/10.5465/ambpp.2022.10998abstract>
- Ding, C. G., Chen, C.-F., & Jane, T.-D. (2022). Improving the performance of the unmeasured latent method construct technique in common method variance detection and correction. *Journal of Organizational Behavior*, *44*(3), 519–542. <https://doi.org/10.1002/job.2673>
- Dombrowski, S. C., Canivez, G. L., Watkins, M. W., & Beaujean, A. A. (2015). Exploratory bifactor analysis of the Wechsler Intelligence Scale for Children—Fifth Edition with the 16 primary and secondary subtests. *Intelligence*, *53*. <https://doi.org/10.1016/j.intell.2015.10.009>
- Eid, M. (2000). A multitrait-multimethod model with minimal assumptions. *Psychometrika*, *65*(2), 241–261. <https://doi.org/http://dx.doi.org/10.1007/bf02294377>

- Enders, C. K. (2001). Enders, C. K. The performance of the full information maximum likelihood estimator in multiple regression models with missing data. *Educational and Psychological Measurement*, 61(5). <https://doi.org/10.1177/00131644011615001>
- Flanagan, D. P., & Dixon, S. G. (2014). The Cattell-Horn-Carroll theory of cognitive abilities. <https://doi.org/10.1002/9781118660584.ese0431>
- Fogarty, G. J. (2008). Principles and applications of educational and psychological testing. In *Adult Educational Psychology* (pp. 351–383). BRILL. [https://doi.org/10.1163/9789087905552\\_015](https://doi.org/10.1163/9789087905552_015)
- Furr, R. M., & Bacharach, V. R. (2013). *Psychometrics: An Introduction 2nd Ed.* Sage.
- Gefen, D., Straub, D., & Boudreau, M.-C. (2000). Structural equation modeling and regression: Guidelines for research practice. *Communications of the Association for Information Systems*, 4. <https://doi.org/10.17705/1cais.00407>
- Gottfredson, L. (2002). Where and why g matters: Not a mystery. *Human Performance*, 15(1), 25–46. [https://doi.org/10.1207/s15327043hup1501&02\\_03](https://doi.org/10.1207/s15327043hup1501&02_03)
- Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8(3), 179–203. [https://doi.org/10.1016/0160-2896\(84\)90008-4](https://doi.org/10.1016/0160-2896(84)90008-4)
- Haig, B. (2022). Repositioning construct validity theory: From nomological networks to pragmatic theories, and their evaluation by explanatory means. <https://doi.org/10.31234/osf.io/k54b6>
- Hart, Y., Dillon, M. R., Marantan, A., Cardenas, A. L., Spelke, E., & Mahadevan, L. (2018). The statistical shape of geometric reasoning. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-018-30314-y>
- Horn, J. L., & Blankson, N. (2005). Foundations for better understanding of cognitive abilities. In D. P. F. bibinitperiod P. L. H. (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues*. Guilford Press.
- Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6(1). <https://doi.org/10.1080/10705519909540118>
- Jensen, A. R. (1982). Reaction time and psychometric g. In *A Model for Intelligence* (pp. 93–132). Springer Berlin Heidelberg. [https://doi.org/10.1007/978-3-642-68664-1\\_4](https://doi.org/10.1007/978-3-642-68664-1_4)
- Jovanović, V. (2015). Structural validity of the Mental Health Continuum-Short Form: The bifactor model of emotional, social and psychological well-being. *Personality and Individual Differences*, 75. <https://doi.org/10.1016/j.paid.2014.11.026>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford.
- Kobrin, J., Patterson, B., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). *Validity of the SAT for Predicting First-Year College Grade Point Average. College Board Research Report 2008(5)*. The College Board.
- Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science*, 315. <https://doi.org/https://psycnet.apa.org/doi/10.1126/science.1136618>
- Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, careerpotential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, 86. <https://doi.org/http://10.1037/0022-3514.86.1.148>
- Lance, C. E., Baranik, L. E., Lau, A. R., & Scharlau, E. A. (2009). If it ain't trait it must be method: (Mis)application of the multitrait-multimethod design in organizational research. In *Statistical and methodological myths and urban legends: Doctrine, verity and fable in the organizational and social sciences*. Taylor & Francis Group.
- Lewis, J., & Sireci, S. G. (2022). Digital Module 30: Validity and Educational Testing: Purposes and Uses of Educational Tests. *Educational Measurement: Issues and Practice*, 41(4). <https://doi.org/10.1111/emip.12533>
- Luo, Y., & Al-Harbi, K. (2016). The Utility of the Bifactor Method for Unidimensionality Assessment When Other Methods Disagree: An Empirical Illustration. *SAGE Open*, 6(4), 215824401667451. <https://doi.org/10.1177/2158244016674513>
- Marsh, H. W., & Hocevar, D. (1988). A new, more powerful approach to multitrait multimethod analyses: Application of second-order confirmatory factor analysis. *Journal of Applied Psychology*, 73(1), 107–117. <https://doi.org/10.1037//0021-9010.73.1.107>
- Newton, P. E., & Shaw, S. D. (2014). *Validity in educational and psychological assessment*. SAGE Publications Ltd. <https://doi.org/10.4135/9781446288856>
- Pattipeilohy, F. W. C. (2017). *Pengujian Validitas Konstruk Tes Potensi Akademik Pascasarjana (PAPS) melalui Analisis Faktor Eksploratori* [Doctoral dissertation, Skripsi, Universitas Gadjah Mada].
- Plake, B. S., & Wise, L. L. (2014). What is the role and importance of the revised AERA, APA, NCME standards for educational and psychological testing? *Educational Measurement: Issues and Practice*, 33(4), 4–12. <https://doi.org/10.1111/emip.12045>
- Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology*, 63(1), 539–569. <https://doi.org/10.1146/annurev-psych-120710-100452>
- Podsakoff, P. M., MacKenzie, S. B., Podsakoff, N. P., & Lee, J. Y. (2003). The mismeasure of management and its implications for leadership research. *The Leadership Quarterly*, 14(6), 615–656. <https://doi.org/10.1016/j.leaqua.2003.08.002>
- Primi, R. (2001). Complexity of geometric inductive reasoning tasks: Contribution to the understanding of fluid intelligence. *Intelligence*, 30(1). [https://doi.org/10.1016/S0160-2896\(01\)00067-8](https://doi.org/10.1016/S0160-2896(01)00067-8)
- Reeves, T. D., & Marbach-Ad, G. (2016). Contemporary Test Validity in Theory and Practice: A Primer for Discipline-Based Education Researchers (R. Nehm, Ed.). *CBE—Life Sciences Education*, 15(1). <https://doi.org/10.1187/cbe.15-08-0183>
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667–696. <https://doi.org/10.1080/00273171.2012.715555>
- Schneider, W. J., & McGrew, K. S. (2012). The Cattell–Horn–Carroll model of intelligence. In D. P. F. bibinitperiod P. L. H. (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues*. Guilford Press.
- Sireci, S., & Benítez Baena, I. (2023). Evidence for test validation: A guide for practitioners. *Psicothema*, 35(3). <https://doi.org/http://10.7334/psicothema2022.477>
- Spearman, C. (1904). “General Intelligence,” objectively determined and measured. *The American Journal of Psychology*, 15(2), 201. <https://doi.org/10.2307/1412107>
- Taub, G. E., & McGrew, K. S. (2004). A confirmatory factor analysis of Cattell-Horn-Carroll theory and cross-age invariance of the Woodcock-Johnson Tests of Cognitive Abilities III. *School Psychology Quarterly*, 19(1), 72–87. <https://doi.org/10.1521/scpq.19.1.72.29409>
- Tavakol, M., & Wetzel, A. (2020). Factor Analysis: a means for theory and instrument development in support of construct validity. *International Journal of Medical Education*, 11, 245–247. <https://doi.org/10.5116/ijme.5f96.0f4a>
- Urbán, R., Kun, B., Farkas, J., Paksi, B., Kökönyei, G., Unoka, Z., Felvinczi, K., Oláh, A., & Demetrovics, Z. (2014). Bifactor structural model of symptom checklists: SCL-90-R and Brief Symptom Inventory (BSI) in a non-clinical community sample. *Psychiatry Research*, 216. <https://doi.org/10.1016/j.psychres.2014.01.027>
- Urbina, S. (2004). *Essentials of psychological testing*. John Wiley & Sons, Inc.
- Vallevand, A., Manthey, D. E., Askew, K., Hartman, N. D., Burns, C., Strowd, L. C., & Violato, C. (2023). Assessing clinical competence: a multitrait-multimethod matrix construct validity study. *Advances in Health Sciences Education*,

- 29(2), 567–585. <https://doi.org/10.1007/s10459-023-10269-0>
- Vigneau, F., & Cormier, S. (2008). The factor structure of the state-trait anxiety inventory: An alternative view. *Journal of Personality Assessment*, 90(3), 280–285. <https://doi.org/http://dx.doi.org/10.1080/00223890701885027>
- Viswanathan, M. (2005). *Measurement error and research design*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412984935>
- Wang, Y., Kim, E. S., Dedrick, R. F., Ferron, J. M., & Tan, T. (2017). A Multilevel Bifactor Approach to Construct Validation of Mixed-Format Scales. *Educational and Psychological Measurement*, 78(2), 253–271. <https://doi.org/10.1177/0013164417690858>
- Ward, W. C. (1982). A comparison of free-response and multiple-choice forms of verbal aptitude tests. *Applied Psychological Measurement*, 6(1), 1–11. <https://doi.org/10.1177/014662168200600101>
- Widhiarso, W. (2017). Mengakomodasi efek metode dalam pengujian validitas konstruk melalui analisis faktor konfirmatori. *Psikologia : Jurnal Psikologi*, 1(1), 37. <https://doi.org/10.21070/psikologia.v1i1.478>
- Widhiarso, W., & Hanifa, S. (2023). Fitur Non Konten dan Intensitas Konten Ukur pada Butir Skala Psikologi. *Intuisi : Jurnal Psikologi Ilmiah*, 14(2), 10–24. <https://doi.org/10.15294/intuisi.v14i2.37616>
- Widhiarso, W., & Haryanta. (2015). Examining method effect of synonym and antonym test in verbal abilities measure. *Europe's Journal of Psychology*, 11(3), 419–431. <https://doi.org/10.5964/ejop.v11i3.865>