

Pembuktian Validitas terkait Struktur Tes Potensi Akademik Pascasarjana (PAPS) Universitas Gadjah Mada

Examining Structural Validity of Tes Potensi Akademik Pascasarjana (PAPS) Universitas Gadjah Mada

Wahyu Widhiarso¹

Fakultas Psikologi, Universitas Gadjah Mada

Abstract. This study examined the construct validity of the Graduate Academic Potential Test (PAPS). Examination was performed on all existing PAPS series (6 forms) to identify the consistency of dimensionality structure of PAPS. Data of this study were analyzed using confirmatory factor analysis (CFA). The results of the analysis support assumption that the structure of the PAPS test is unidimensional. All of the model fit indices support the decision that unidimensional model fit the data. The study also examined factor loading that the non-verbal components, especially the quantitative components that had a higher factor weight than the other components.

Keywords: confirmatory factor analysis; PAPS; validity construct

Abstrak. Penelitian yang menguji validitas konstruk Tes PAPS belum pernah dilakukan. Penelitian ini bertujuan untuk menguji validitas konstruk Tes Potensi Akademik Pascasarjana (PAPS). Pengujian dilakukan pada semua seri PAPS yang ada (6 seri) untuk melihat konsistensi struktur dimensionalitas Tes PAPS. Analisis dilakukan dengan menggunakan analisis faktor konfirmatori (CFA). Hasil analisis membuktikan struktur Tes PAPS yang berbentuk unidimensi. Semua indeks ketepatan model memenuhi penyimpulan model pengukuran PAPS sebagai pengukuran yang bersifat unidimensi. Kajian terhadap bobot faktor menemukan bahwa komponen-komponen non-verbal terutama komponen kuantitatif memiliki bobot faktor yang lebih tinggi dibanding dengan komponen lainnya.

Keywords: analisis faktor konfirmatori; PAPS; validitas konstruk

Penggunaan tes yang mengukur potensi akademik yang berkaitan dengan kemampuan kognitif calon siswa dalam program admisi penerimaan siswa baru sudah dilakukan sejak lama. Di luar negeri beberapa tes telah dikembangkan untuk memenuhi keperluan tersebut seperti SAT

(*Scholastic Aptitude Test*), GMAT (*Graduate Management Admission Test*) atau LSAT (*Law School Admission Test*). Tes-tes yang mengukur potensi ini terbukti mampu memprediksi performansi siswa dengan baik ketika mereka diberikan kesempatan untuk menyelesaikan sekolahnya (Young, Klieger, Bochenek, Li, & Cline, 2014).

¹ Korespondensi mengenai artikel ini dapat melalui:
wahyu_psy@ugm.ac.id

Universitas Gadjah Mada (UGM) memiliki beberapa instrumen yang dipakai untuk menyeleksi calon mahasiswa, baik di tingkat sarjana maupun pascasarjana. Salah satu tes yang saat ini dipakai oleh UGM untuk menyeleksi mahasiswa di tingkat pascasarjana adalah tes Potensi Akademik Pascasarjana (PAPS). Tes PAPS merupakan salah satu instrumen untuk menyeleksi calon mahasiswa yang mengambil pendidikan di program magister atau yang sederajat serta program doktoral. Penyelenggaraan Tes PAPS dalam proses seleksi ini bertujuan untuk mengidentifikasi potensi akademik calon mahasiswa yang akan dapat mendukung kesuksesan mereka sekiranya menjalani program pascasarjana. Sama seperti halnya tes-tes yang lain, PAPS memerlukan pembuktian-pembuktian mengenai validitasnya dari hasil berbagai jenis pembuktian. Meskipun beberapa kajian (misalnya, Azwar & Ridho, 2013; Widhiarso & Haryanta 2015; Widhiarso & Haryanta 2016) telah dilakukan untuk membuktikan validitas Tes PAPS namun pengujian validitas dari pendekatan yang berbeda tetap diperlukan. Hal ini dikarenakan pengujian terhadap validitas sebuah tes perlu dilakukan secara berkesinambungan dari berbagai pendekatan.

Ada beberapa pertimbangan mengapa pembuktian validitas, khususnya validitas terkait dengan struktur Tes PAPS, perlu untuk dilakukan secara terus menerus meskipun tes telah selesai dikembangkan dan digunakan secara praktis. Pertama, konstruk ukur yang menjelaskan atribut-atribut yang diukur oleh sebuah tes sangat memungkinkan mengalami perubahan (Everson, 2003). Demikian juga Tes PAPS yang konstruk ukurnya mengalami perbaikan untuk memenuhi perkembangan zaman. Misalnya, jika pada awal-

awal versinya hanya menekankan pada penalaran kognitif maka pada versi-versi yang lebih akhir atribut baru disisipkan seperti unsur berpikir kritis. Penelitian sebelumnya menemukan bahwa tes inteligensi yang bersifat klasikal, misalnya *Standard Progressive Matrices* (SPM), kurang tepat jika dikenakan pada mahasiswa sehingga diperlukan tes lain yang dapat mengungkap kemampuan kognitif mahasiswa (Wulan, 1996). Tes yang dulunya dapat membedakan individu berdasarkan konstruk tertentu memungkinkan untuk kehilangan daya bedanya ketika sebagian besar orang di dalam populasi tidak memiliki perbedaan lagi dalam hal konstruk yang diukur. Oleh karena itu pengujian terhadap struktur ini merupakan bagian dari upaya untuk mengakomodasi perubahan tuntutan jika struktur tes kemudian dimodifikasi untuk menjawab tuntutan zaman.

Kedua, konstruk ukur PAPS berakar pada inteligensi yang memiliki struktur yang khas karena dikembangkan berdasarkan analisis faktor (Taub & McGrew, 2004; Ward, Ryan, & Axelrod, 2000). Tes-tes inteligensi memiliki komponen ukur dan struktur yang terorganisasi dengan sistematis. Sebagai contoh baterai tes kognitif Woodcock-Johnson IV memiliki struktur yang sangat sistematis. Struktur tersebut memuat bagian hierarki (*broad ability, narrow ability*) sekaligus bagian non hierarki (misalnya penalaran fluida dan memori kerja). Sebagai proksi dari tes inteligensi Tes PAPS memiliki struktur tersendiri, sama dengan tes-tes potensi akademik lainnya. Pertanyaan mengenai elemen non hierarki Tes PAPS yang memuat komponen verbal, kuantitatif dan figural sudah terjawab. Ketiga komponen yang mewakili media menalar tersebut memiliki dasar teori yang kuat sehingga tidak dipersoalkan penggunaannya.

Sebaliknya, struktur hierarki Tes PAPS belum banyak menemukan jawaban. Pertanyaan yang muncul mengenai Tes PAPS adalah jika diwujudkan dalam model analisis faktor konfirmatori, apakah berbentuk satu jenjang (*first order model*) ataukah dua jenjang (*second order model*).

Dari berbagai tipe validitas yang ada, validitas yang terkait dengan struktur tes merupakan memiliki keutamaan untuk dikaji pada Tes PAPS. Alasan pemilihan pengujian validitas tipe ini adalah karena belum ada penelitian dan publikasi yang menelaah struktur dari konstruk PAPS. Penelitian yang ada sebagian besar masih menelaah Tes PAPS dari level butir dan level komponen tes. Telaah terhadap konstruk atau bangunan Tes PAPS pernah dilakukan namun masih pada tataran teoretik. Telaah terhadap konstruk Tes PAPS dari sisi empirik, yaitu telaah konstruk yang didasarkan pada data di lapangan belum pernah dilakukan.

Validitas konstruk adalah syarat utama untuk mendukung interpretasi yang tepat terhadap skor tes. Di sisi lain, pertanyaan awal para pengguna tes seringkali lebih mengarah pada apa yang diukur oleh tes yang bersangkutan, apa dasar teorinya serta aspek-aspek apa saja di dimuat oleh tes (Rock, Werts, & Grandy, 1982). Semua pertanyaan ini dapat dijawab melalui validitas konstruk tes. Validitas konstruk adalah proses menyusun bukti mengenai hubungan antara tes dengan konsep (teori) yang dapat mendukung inferensi mengenai skor tes. Secara praktis prosedur analisisnya adalah peneliti mengembangkan sebuah model yang bersifat *a priori* atau *posteriori* yang menjelaskan struktur tes yang dikaitkan dengan data hasil pengukuran.

Identifikasi terhadap konstruk PAPS sangat diperlukan untuk menjawab

pertanyaan mengenai dengan interpretasi skor Tes PAPS maupun pertanyaan mengenai kesetaraan dengan tes lain yang mengukur abilitas yang sejenis (misalnya TPA Bappenas, TKDA Sertifikasi Dosen atau TIU Tes Seleksi CPNS). Meskipun secara teoritik konstruk yang diukur oleh Tes PAPS merupakan konstruk yang berstruktur unidimensi (tes yang mengukur atribut tunggal), interpretasi yang beredar di masyarakat menyebutkan bahwa dimensi-dimensi Tes PAPS (verbal, kuantitatif, figural) merupakan dimensi yang terpisah. Banyak orang awam yang menafsirkan skor tiap dimensi Tes PAPS merepresentasikan kemampuan yang bersifat independen satu dengan lainnya. Akibatnya mereka mengaitkan skor dimensi Tes PAPS dengan bidang tertentu, misalnya calon mahasiswa bidang pendidikan arsitektur diharapkan memiliki skor subtes figural yang tinggi sedangkan calon mahasiswa bidang pendidikan sastra diharapkan memiliki skor subtes verbal yang tinggi dan bidang matematika diharapkan memiliki skor kuantitatif yang tinggi. Masalah ini akan terjawab ketika konstruk teoretik Tes PAPS didukung keabsahannya oleh data di lapangan.

Permasalahan lain muncul dari pertanyaan mengenai paralelitas konstruk antara Tes PAPS dengan tes lain, misalnya Tes Potensi Akademik (TPA) yang dikembangkan oleh Bappenas dan Tes Kompetensi Dasar Akademik (TKDA) yang dikembangkan dengan HIMPSI. Beberapa instansi melihat ketiganya adalah tes yang sama sehingga mensyaratkan kepada calon pelamar untuk mendapatkan skor salah satu dari ketiga tes di atas. Salah satu universitas memiliki persyaratan bahwa calon mahasiswa telah memiliki skor tertentu untuk tes PAPS UGM atau TPA Bappenas. Adanya kebijakan mengenai Tes PAPS dan TPA

sebagai persyaratan yang substitutif menunjukkan bahwa kedua tes dilihat sebagai dua tes paralel yang mengukur konstruk yang sama. Asumsi mengenai kesetaraan kedua tes dapat dipenuhi melalui serangkaian uji psikometris, salah satunya adalah kesetaraan konstruk yang diukur. Meskipun konstruk atribut yang diukur oleh Tes PAPS merupakan modifikasi dari TPA namun dengan adanya perubahan komposisi atau komponen tes perbedaan konstruk ukur keduanya memungkinkan untuk terjadi.

Terakhir, pengujian konstruk merupakan sebuah keharusan karena konstruk potensi akademik mengalami perubahan. Sebagai contoh komposisi tes yang mengukur potensi akademik atau kecerdasan umum di luar negeri mengalami banyak perubahan. (misal: SAT, GRE, LSAT). Perubahan ini dilakukan untuk mengakomodasi definisi mengenai kesuksesan dalam bidang akademik sehingga tes yang memprediksinya mengalami modifikasi untuk menyesuaikan keperluan yang baru (Gre, Walters, & Kaufman, 2011). Kerangka pikir mengenai validitas tes potensi akademik saat ini mengarah pada pengujian validitas konstruk karena para peneliti berusaha untuk memecahkan masalah konstruk yang mendasari kemampuan yang diukur oleh tes tersebut (Jenkins, 2016). Ada dua kemampuan yang menjadi perdebatan di kalangan peneliti, yaitu inteligensi dan bakat (*aptitude*). Meski keduanya memiliki dasar yang cukup kuat karena pada dasarnya tes potensi akademik mengukur kemampuan yang bersifat potensial, bukan kemampuan aktual. Kedua kemampuan tersebut juga merupakan lebih terkait dengan performansi individu di masa yang akan datang daripada performansi di masa lalu. Penelitian mengenai validitas konstruk sangat diperlukan untuk mendukung

landasan yang dipakai oleh tes yang banyak dipakai untuk seleksi mahasiswa di perguruan tinggi (*college admission*) ini (Dienstag 2011). Pemahaman mengenai dimensi-dimensi di dalam tes dan kaitan antar dimensi tersebut memberikan manfaat yang penting untuk menentukan domain-domain pengukuran yang sesuai dengan kebutuhan.

Tes Potensi Akademik Pascasarjana (PAPS)

Tes PAPS menekankan pada pengukuran potensi dan bukan pada prestasi aktual. Hal ini dikarenakan Tes PAPS sama dengan tes-tes yang dipakai dalam admisi siswa yaitu bertujuan untuk memprediksi performansi individu di masa depan daripada mengukur performansi individu yang merefleksikan aktivitasnya di masa lalu (Kobrin, Camara, & Milewski, 2002). Konstruk yang diukur oleh tes ini terkait dengan potensi individu. Jika tes prestasi mengidentifikasi seberapa luas atau dalam pengetahuan individu mengenai pengetahuan secara spesifik, tes potensi menilai seberapa mudah individu dapat memperoleh pengetahuan yang tidak mereka memiliki. Tes yang mengukur potensi berusaha untuk mengidentifikasi seberapa baik individu dalam memahami dan menggunakan informasi dari berbagai ragam jenis sumber untuk mendukung tujuan mereka (Cook, 2003).

Tes PAPS dikembangkan dengan mengadaptasi konstruk-konstruk yang ukur oleh beberapa tes seperti *Scholastic Aptitude Test* (SAT). Tes ini kemudian yang menjadi dasar pengembangan tes-tes potensi akademik di Indonesia. Beberapa penulis menjelaskan bahwa tes-tes potensi akademik dikembangkan dengan berdasarkan teori bakat yang lebih mengarah pada bidang akademik sementara itu ahli lain mengatakan bahwa tes potensi akademik pengembangannya didasarkan

pada teori kecerdasan umum. Pada bagian berikut ini akan dijelaskan pengertian antara konstruk-konstruk yang terkait dengan Tes PAPS, yaitu kecerdasan, bakat dan prestasi. Dalam menguji validitas konstruk ada tiga hal yang perlu ditelaah, yaitu atribut yang diukur, dimensionalitas alat ukur dan model pengukuran. Atribut ukur menjelaskan atribut psikologis apa yang dijelaskan oleh konstruk yang dikaji. Berkaitan dengan konstruk yang diukur oleh TPA, setidaknya ada tiga atribut yang relevan, yaitu kecerdasan, bakat dan prestasi.

Teori yang dipakai dalam pengembangan Tes PAPS adalah kombinasi antara teori mengenai inteligensi dan bakat. Masing-masing atribut ini akan ditelaah secara spesifik untuk mendapatkan gambaran yang lebih jelas mengenai konstruk ukur dari Tes PAPS. Pelibatan teori inteligensi dalam pengembangan Tes PAPS karena inteligensi merupakan potensi manusia yang bersifat umum. Dalam hal ini Tes PAPS memiliki kemiripan domain dengan domain-domain dalam tes inteligensi. Berbagai studi menunjukkan bahwa tes inteligensi adalah prediktor yang cukup efektif dan stabil dalam memprediksi performansi dalam berbagai bidang pekerjaan. Misalnya dalam *setting* akademik maupun lingkungan kerja. Tes inteligensi biasanya diadministrasikan dalam proses awal atau *screening* mengenai calon yang akan diterima. Kondisi ini menempatkan tes-tes kecerdasan dari luar negeri banyak yang diadaptasi ke budaya Indonesia. Di sisi lain, beberapa peneliti juga mulai mengembangkan tes sendiri yang secara khusus dapat diaplikasikan pada populasi Indonesia. Tes inteligensi memiliki kelebihan dibandingkan tes lain dalam hal profil kemampuan kognitif mengenai kekuatan dan kelemahan individu. Tes

inteligensi dapat memprediksi prestasi akademis siswa dan performansi kerja seorang karyawan. Tes inteligensi juga memberikan informasi yang bermanfaat bagi individu yang mengalami gangguan perkembangan. Hal ini dikarenakan *intelligence quotient* (IQ) dapat memberikan informasi batas-batas kemampuan seorang individu dan pengaruh dari gangguan yang dialami terhadap kemampuan belajar. Dalam lingkup sekolah, inteligensi dapat memberikan informasi apakah seorang siswa perlu mendapatkan program khusus atau tidak.

Dari sekian banyak teori mengenai inteligensi, teori inteligensi yang berkaitan dengan kecerdasan umum adalah teori yang paling dekat dengan konstruk ukur Tes PAPS. Frase kecerdasan umum (faktor g) selalu terkait dengan suatu variabel laten yang dihasilkan dari analisis inter-korelasi antara beberapa tes kecerdasan (Spearman, 1904). Faktor g telah digunakan untuk menjelaskan temuan bahwa skor pada berbagai tugas kognitif cenderung berkorelasi positif dengan satu sama lain. Perkembangan teori mengenai faktor g kemudian mengikuti pembagian dari John Horn dan Cattell mengenai teori kecerdasan fluida-terkristal. Kecerdasan fluida (gf) mengacu pada kemampuan untuk memecahkan masalah baru dan beradaptasi dengan situasi baru dan dianggap non verbal dan relatif adil budaya. Kecerdasan terkristal (gc) mengacu pada keterampilan dan pengetahuan yang diperoleh dan tergantung pada latar belakang pendidikan dan budaya. Tes yang mengukur faktor gf termasuk tidak terbatas pada tes berbentuk matriks atau figural sedangkan tes yang mengukur faktor gc juga tidak terbatas pada tes dengan menggunakan kosakata dan tes pengetahuan umum (Saltier, 1992). Sebuah tes yang mengukur

potensi akademik seperti SAT dapat dikatakan hampir pasti mencerminkan kombinasi kemampuan penalaran fluida dan terkristal (Engle, Tuholski, Laughlin, & Conway, 1999).

Teori kedua yang dipakai dalam pengembangan Tes PAPS adalah teori bakat. Bakat mengacu potensi individu untuk belajar tugas pada bidang tertentu berdasarkan pelatihan tertentu. Tes bakat yang dirancang untuk mengukur kemampuan subjek dalam belajar pada sebuah objek atau tugas yang spesifik atau memperoleh keterampilan yang spesifik. Meskipun bakat dan kemampuan secara umum memiliki perbedaan, dari sisi terminologinya kata bakat dan kemampuan sulit untuk dipisahkan. Pengukuran bakat biasanya dilakukan oleh sebuah tes yang diarahkan pada memprediksi seberapa mampu individu untuk menguasai pengetahuan atau keterampilan di bidang tertentu dapat diperoleh. Berangkat dari paparan di atas beberapa pakar menyatakan bahwa tes potensi akademik lebih menekankan pada abilitas yang terkait dengan bakat.

Kerangka pikir mengenai konstruk mana yang relevan dengan tes potensi akademik dapat dijawab oleh pengujian terhadap validitas konstruk tes. Struktur tes dan keterkaitan antara bagian-bagian di dalamnya akan menjawab apakah sebuah tes potensi akademik yang dikembangkan lebih merepresentasikan abilitas terkait dengan inteligensi atau bakat. Pada penelitian ini, peneliti menguji dan membandingkan dua model pengukuran (lihat Gambar 1). Pada gambar tersebut terlihat ada dua model yang akan diuji. Model 1 menunjukkan bahwa antara satu subtes dengan subtes lain dalam Tes PAPS mengukur atribut yang berbeda. Teori yang melatarbelakangi model ini adalah setiap subtes

merepresentasikan abilitas yang berbeda-beda sehingga ketiganya memiliki hubungan yang rendah. Salah satu teori yang dipakai adalah teori inteligensi berstrata. Pada Model 2 semua dimensi di dalam Tes PAPS mengukur atribut yang sama, yaitu kemampuan penalaran. Teori yang melatarbelakangi pengembangan model ini adalah bahwa subtes-subtes di dalam Tes PAPS merupakan suatu konteks stimulus yang dapat bersifat substitutif. Adanya sifat substitutif ini menunjukkan bahwa seorang individu yang memiliki penalaran yang tinggi akan bisa bekerja dalam berbagai media atau stimulus, baik stimulus berupa kata-kata (verbal), angka (numerik) maupun gambar (figural). Kedua model ini diuji untuk menentukan struktur pengukuran mana yang paling mewakili struktur Tes PAPS. Uji yang dipakai untuk membandingkan adalah uji kai-kuadrat.

Validitas konstruk

Kebanyakan variabel yang dikaji dalam penelitian psikologi merupakan sesuatu yang tidak ada atau nyata seperti halnya sebuah barang di pasar. Variabel merupakan konstruk teoretik yang fiktif dan memuat banyak perilaku (Domino & Domino, 2006). Dalam mempelajari konstruksi ini, peneliti biasanya menerjemahkannya ke dalam definisi yang bersifat operasi yang lebih spesifik. Bentuk yang paling operasional adalah sebuah tes psikologi. Misalnya, konstruk teoretis sebuah variabel kecerdasan dioperasionalkan menjadi sebuah tes yang khusus untuk mengukur inteligensi. Dalam proses memvalidasi tes aktivitas yang dilakukan tidak hanya menelaah tes akan tetapi juga menelaah konstruk yang diukur oleh tes tersebut.

Teknik yang banyak dipakai untuk mengidentifikasi validitas konstruk adalah

analisis faktor konfirmatori. Analisis faktor eksploratori dapat dipakai untuk menilai sejauh mana seperangkat butir di dalam sebuah domain konten tertentu mengukur atribut yang sama sekaligus struktur hubungan antar atribut ukur yang mendasarinya. Teknik seperti ini memiliki keterbatasan terutama karena ketidak-mampuannya untuk mengenali tepat tidaknya model dari struktur faktor yang dihasilkan. Di sisi lain, analisis faktor eksploratori melibatkan interpretasi pasca hasil analisis (*post-hoc*), sedangkan analisis faktor konfirmatori menetapkan hubungan apriori dan perbedaan antara dimensi di dalam tes. Indikator yang terbukti memuat secara jelas dalam analisis faktor eksploratori mungkin peneliti untuk menelaah apa yang menyebabkan model memiliki ketepatan yang rendah berdasarkan nilai bobot faktor baik pada level dimensi atau indikator yang biasanya disebabkan oleh rendahnya konsistensi eksternal dimensi-dimensi di dalam tes (Anderson & Gerbing, 1988).

Dalam pengembangan tes, analisis faktor konfirmatori sangat diperlukan untuk baik untuk mengidentifikasi struktur tes maupun untuk melengkapi atau memperkuat temuan dari analisis faktor eksploratori yang dilakukan sebelumnya. Analisis faktor konfirmatori merupakan jenis analisis yang berbasis pemodelan, terutama pemodelan persamaan struktural (SEM) yang dapat dipakai untuk menilai ketepatan sebuah dengan data atau dengan model. Dalam analisis ini model yang dikembangkan dibandingkan dengan model nihil (*null model*) yaitu ketika semua butir memuat dimensi-dimensi yang terpisah, model dengan dimensi tunggal umum dan model yang memuat sejumlah indikator yang merupakan manifestasi dari dimensi yang sama (Joreskog & Sorbom, 1990). Model

yang memuat dimensi majemuk ini membatasi setiap indikator agar merupakan bagian dari (*load*) dimensi yang relevan dan bukan merupakan bagian dari dimensi yang tidak relevan. Namun demikian dalam berbagai kasus, ada indikator yang memanifestasikan dimensi di luar dimensi yang dimanifestasikan olehnya. Kasus ini dinamakan dengan fenomena lintas bobot (*cross loading*).

Analisis faktor konfirmatori dilakukan dengan menggunakan matriks butir varians-kovarians daripada matriks korelasi (Harvey, Billings, & Nilan, 1985). Ketepatan model yang diajukan dalam analisis faktor konfirmatori dapat ditunjukkan melalui harga statistik. Statistik kai-kuadrat memungkinkan penilaian ketepatan pada model yang spesifik, serta perbandingan antara dua model. Semakin kecil kai-kuadrat, semakin baik tinggi ketepatan model yang diuji. Kai-kuadrat yang harganya sebesar tiga kali lebih besar derajat kebebasan model yang diuji dapat menunjukkan penerimaan model tetapi lebih disarankan jika harga kai-kuadrat mendekati harga derajat kebebasan pada model yang diuji (Thacker, Fields, & Tetric, 1989). Harga kai-kuadrat yang tidak signifikan diharapkan oleh peneliti karena hasil ini menunjukkan bahwa perbedaan antara matriks varians-kovarians model yang diajukan oleh peneliti dan matriks varians-kovarians model berdasarkan data yang didapatkan dari lapangan cukup kecil. Selain itu, harga kai-kuadrat yang signifikan diinginkan jika model yang diuji dibandingkan dengan model yang menjadi kompetitor-nya. Namun demikian perlu di catat bahwa uji kai-kuadrat cukup sensitif terhadap ukuran sampel. Oleh karena itu para pakar mengembangkan indeks ketepatan model yang dapat melengkapi keterbatasan indeks kai-kuadrat tersebut.

Indeks-indeks tersebut dibuktikan tidak terpengaruh oleh ukuran sampel. Dengan adanya indeks pelengkap tersebut maka harga kai-kuadrat yang signifikan mungkin tidak menjadi masalah jika indeks ketepatan tambahan sudah cukup memadai.

Penelitian ini bertujuan untuk menguji validitas konstruk Tes PAPS. Validitas konstruk merupakan informasi mengenai struktur internal di dalam tes. Validitas menggambarkan bagaimana antar dimensi atau komponen tes membentuk sebuah kombinasi yang membentuk satu kesatuan konstruk yang utuh. Penelitian ini mengaplikasikan analisis faktor konfirmatori (*confirmatory factor analysis/CFA*) untuk menguji konstruk Tes PAPS. Penelitian sangat penting dilakukan karena hasilnya dapat menjawab beberapa permasalahan teoritis maupun praktis mengenai hasil Tes PAPS. Sebagai contoh, jika konstruk Tes PAPS merupakan konstruk yang bersifat multidimensional, maka skor tiap dimensi dapat berdiri sendiri yang tidak terkait dengan dimensi lainnya. Sebaliknya, jika ditemukan bahwa konstruk Tes PAPS bersifat unidimensi, maka skor tiap dimensi tidak dapat dilepaskan dengan dimensi lainnya. Dengan kata lain, tafsiran terhadap skor Tes PAPS harus meliputi semua dimensi. Keutamaan lainnya dari penelitian ini terkait dengan relevansi Tes PAPS untuk menjelaskan perbedaan individual terkait dengan potensi akademik. Meskipun dari tataran butir-butirnya Tes PAPS telah dikaji validitasnya akan tetapi dari tataran komposit tes, Tes PAPS juga perlu dikaji. Pengujian terhadap validitas di level tes ini memberikan banyak manfaat terkait dengan fungsi tes.

Metode

Data penelitian

Data yang dianalisis melalui penelitian ini didapatkan dari data sekunder yang diperoleh dari database UPAP, hasil administrasi Tes PAPS mulai dari tahun 2012 hingga 2016. Data yang dipakai dalam penelitian ini adalah hasil administrasi Tes PAPS masih menggunakan format lama yaitu tes dengan waktu 100 menit untuk 120 soal.

Dalam melakukan analisis faktor konfirmasi (CFA) peneliti menggunakan teknik persamaan struktural pemodelan (SEM) yang diaplikasikan pada semua seri Tes PAPS. Analisis ini tidak hanya memungkinkan penyelidikan secara simultan terhadap hubungan antara variabel dependen dan independen yang kompleks, akan tetapi juga identifikasi adanya dimensi atau komponen yang minor (memiliki kontribusi rendah terhadap konstruk yang diukur) dan dimensi yang relatif sempit dalam kaitannya dengan konstruk yang diukur.

Tabel 1.
Subtes dan Komponen Tes PAPS

Subtes	Komponen
Verbal	Padanan Kata
	Lawan Kata
	Analogi Kata
	Pemahaman Wacana
Kuantitatif	Deretan Angka
	Aritmetika
	Konsep Aljabar
Penalaran	Perbandingan Kuantitatif
	Logis
	Diagram
	Analitis
	Serial
	Klasifikasi

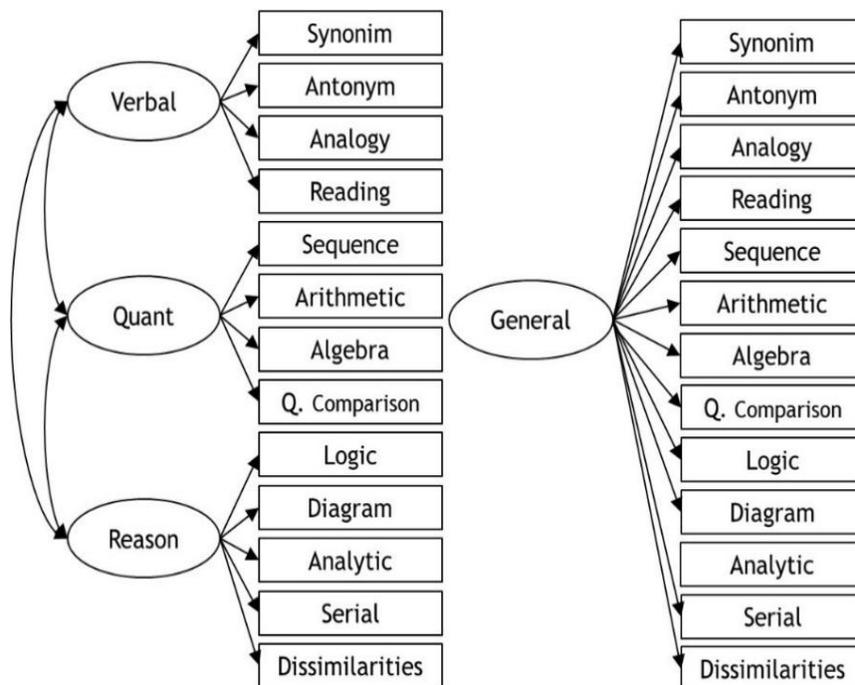
Instrumen

Penelitian ini menggunakan Tes Potensi Akademik Pascasarjana (PAPS). Tes ini memuat 120 soal yang terdiri dari 13 komponen tes yang merepresentasikan jenis tes. Berdasarkan konteks atau penggunaan media yang dipakai dalam mengukur, Tes PAPS memiliki tiga subtes (Verbal, Kuantitatif dan Penalaran). Tes ini dikembangkan oleh Tim Pengembangan PAPS yang diketuai oleh Saifuddin Azwar. Properti psikometris Tes PAPS menurut berbagai laporan penelitian telah sesuai dengan batasan yang ditetapkan oleh literatur psikometris. Misalnya, reliabilitas pengukuran yang dihasilkan dari administrasi Tes PAPS untuk semua seri telah berada di atas 0,80. Semua butir memiliki daya diskriminasi yang mencukupi (di atas 0,20) dan tingkat kesulitan butir-butir Tes PAPS terdistribusi secara merata dari mudah, sedang hingga tinggi (Widhiarso, Azwar, Suharti, & Haryanta, 2015). Tingkat kesulitan butir tidak terkait dengan

komponen tes, karena di dalam setiap komponen tes terdapat butir yang mudah, sedang hingga sulit.

Prosedur analisis data

Data yang didapatkan dalam penelitian ini dianalisis dengan menggunakan teknik analisis faktor konfirmatori (CFA). Alasan penggunaannya adalah karena Tes PAPS menggunakan teori inteligensi umum dalam pengembangannya. Dalam hal ini teori inteligensi umum yang menekankan pada satu atribut ukur, yaitu penalaran kognitif. Dengan demikian model yang diharapkan muncul adalah model unidimensi sehingga model ini merupakan prioritas pertama untuk diuji ketepatannya terhadap data yang didapatkan dari hasil pengukuran. Model kedua yang diuji adalah model multidimensi dengan menempatkan subtes sebagai dimensi yang berbeda karena diasumsikan mengukur atribut yang kurang memiliki keterkaitan antara satu dengan lainnya.



Gambar 1. Struktur hipotetik Tes PAPS

Indeks ketepatan data yang dijadikan sebagai pendukung penyimpulan ketepatan model dengan data beserta harga minimalnya adalah sebagai berikut: CFI (*Comparative Fit Index*), dan TLI (*Tucker Lewis Index*) yang diharapkan harganya di atas 0,95 untuk menunjukkan model yang tepat dengan data. Indeks RMSEA (*Root Mean Square Error Approximation*) yang harganya harus di bawah 0,08 (Hu & Bentler, 1999). Untuk mempermudah proses estimasi parameter butir dan indeks ketepatan model, unit analisis pada model yang diuji adalah skor komponen tes yang didapatkan dari rerata skor butir pada satu komponen yang sama. Proses ini dinamakan dengan pemaketan butir (Widhiarso & Suhapti, 2018) atau dinamakan juga dengan *item parcelling* (Matsunaga, 2008), yaitu proses untuk mendapatkan satu skor komposit yang didapatkan dari seperangkat butir.

Perbandingan antar model dilakukan secara deskriptif dan statistik. Uji secara deskriptif dilakukan dengan menggunakan perbandingan harga kai-kuadrat, koefisien *Akaike's Information Criterion* (AIC); *Bayesian Information Criterion* (BIC); *Sample-Size-Adjusted BIC* (aBIC) sedangkan uji secara statistik dilakukan dengan menggunakan uji Wald (Armstrong & Shi, 2009) dan Vuong (Vuong, 1989). Program lunak yang dipakai untuk mengerjakan semua prosedur analisis dilakukan dengan menggunakan MPLUS 7.4.

Hasil

Uji ketepatan model

Hasil analisis terhadap enam tes mendapatkan ketepatan model yang cukup tinggi pada model unidimensi. Hasil selengkapnya dapat dilihat pada Tabel 1. Semua harga indeks ketepatan model

memenuhi kriteria mengenai model yang tepat dengan data. Harga CFI dan TLI berada di atas 0,90 sedangkan harga indeks RMSEA di bawah 0,08. Hasil ini menunjukkan bahwa model unidimensi maupun multidimensi sesuai dengan data. Harga statistik kai-kuadrat yang didapatkan pada semua model sangat besar. Besarnya statistik ini cenderung menyimpulkan adanya kesenjangan antara model dengan data (model tidak sesuai dengan data) karena harga statistik yang melebihi titik kritis penolakan hipotesis nihil cenderung menolak model yang tepat dengan data (Hu & Bentler, 1999). Namun oleh karena ukuran sampel yang dipakai dalam data ini sangat besar (melebihi 1000) maka harga statistik kai-kuadrat ini tidak dipakai untuk menentukan tepat tidak model yang diajukan karena statistik kai-kuadrat sangat sensitif terhadap ukuran sampel yang besar. Harga kai-kuadrat dipakai beserta harga indeks *log-likelihood* dipakai untuk membandingkan ketepatan model antara model unidimensi dengan multidimensi.

Hasil perbandingan ketepatan model dengan menggunakan uji Wald menunjukkan dari enam seri Tes PAPS yang dikaji, semua menunjukkan bahwa model multidimensi memiliki ketepatan kecuali Tes PAPS Seri B1 yang tidak menemukan perbedaan yang signifikan antara model unidimensi dan multidimensi (Wald = 0,34; $p > 0,05$). Hasil yang sama juga didapatkan dari uji melalui teknik Vuong. Secara deskriptif ketidaksetaraan ketepatan antara model unidimensi dan model multidimensi terlihat dari selisih harga kai-kuadrat antar kedua model pada seri Tes PAPS selain Seri PAPS B1. Hasil uji ketepatan model dan perbandingan tingkat ketepatan model dapat disimpulkan bahwa model pengukuran Tes PAPS adalah model pengukuran multidimensi

Tabel 2.
Indeks Ketepatan dan Uji Perbandingan Ketepatan Model Tes PAPS

Seri PAPS	Model	Kai-Kuadrat	LL	CFI	TLI	AIC	BIC	aBIC	SRMR	RMSEA LB	RMSEA UB	Wald
A1 (R)	Unidimensi	913,95	10757,20	0,91	0,89	-21432,40	-21190,69	-21320,96	0,043	0,07	0,08	14,47*
	Multidimensi	648,91	10889,72	0,94	0,92	-21691,45	-21432,05	-21571,85	0,036	0,06	0,07	
A2 (R)	Unidimensi	836,86	10795,75	0,92	0,90	-21509,49	-21267,78	-21398,05	0,041	0,07	0,07	44,99*
	Multidimensi	595,51	10916,42	0,94	0,93	-21744,85	-21485,45	-21625,25	0,034	0,05	0,06	
B1	Unidimensi	375,46	9222,80	0,93	0,92	-18363,60	-18130,71	-18260,98	0,033	0,05	0,05	0,34
	Multidimensi	329,76	9245,65	0,94	0,93	-18403,30	-18153,37	-18293,16	0,031	0,05	0,05	
B2	Unidimensi	357,52	8384,12	0,94	0,93	-16686,23	-16451,60	-16581,87	0,031	0,05	0,05	50,77*
	Multidimensi	222,52	8451,62	0,97	0,96	-16815,23	-16563,43	-16703,23	0,024	0,03	0,04	
C1	Unidimensi	478,16	4280,16	0,92	0,90	-8478,32	-8268,55	-8398,78	0,044	0,07	0,08	55,45*
	Multidimensi	338,37	4350,05	0,95	0,93	-8612,11	-8386,98	-8526,75	0,037	0,06	0,07	
C2	Unidimensi	394,64	6576,81	0,96	0,95	-13071,63	-12842,44	-12972,70	0,029	0,05	0,06	23,88*
	Multidimensi	298,80	6624,74	0,97	0,96	-13161,47	-12915,52	-13055,31	0,025	0,05	0,05	

Catatan: LL = Indeks Log-likelihood; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion; aBIC = Sample-Size-Adjusted BIC; RMSEA = Point estimate of root mean squared error of approximation; RMSEA LB = Lower bound of the 90% interval konfidensi pada RMSEA; RMSEA UB = Upper bound of the 90% interval konfidensi pada RMSEA; Wald = Statistik Uji Wald; * = $p < 0,05$.

akan tetapi antar dimensi memiliki keterkaitan yang tinggi atau yang dinamakan dengan tes yang bersifat multidimensi akan tetapi memiliki dimensi yang memiliki keterkaitan (*multidimensional tests with correlated*). Konsep ini selengkapnya dapat dilihat pada Furr dan Bacharach (2008).

Perbandingan bobot faktor

Bobot faktor menunjukkan hubungan antara indikator dengan atribut ukur, dalam hal ini skor laten yang dibentuk oleh semua komponen tes. Identifikasi terhadap bobot faktor pada semua seri Tes PAPS yang dikaji pada penelitian ini menunjukkan bahwa semua komponen tes mampu merefleksikan atribut yang diukur dengan baik. Simpulan ini didapatkan dari harga bobot faktor yang semuanya berada di atas 0,30 (lihat Tabel 3).

Berdasarkan analisis faktor konfirmatori model multidimensi didapatkan bahwa subtes penalaran kuantitatif memiliki bobot faktor yang paling tinggi dibanding dengan subtes penalaran

figural dan verbal. Hasil uji Wald menunjukkan perbedaan yang signifikan pada semua seri Tes PAPS. Misalnya pada seri PAPS A1 didapatkan harga statistik Wald adalah 59,314 ($p < 0,05$) untuk uji perbedaan antara bobot faktor verbal dan kuantitatif, 4,486 ($p < 0,05$) untuk verbal dan figural serta 59,314 ($p < 0,05$) untuk kuantitatif dan figural. Secara umum jika diurutkan dari subtes yang memiliki bobot faktor tertinggi adalah subtes kuantitatif, figural dan verbal.

Pada bobot faktor komponen ditemukan juga temuan yang mendukung temuan di atas. Bobot faktor komponen-komponen penalaran kuantitatif memiliki bobot faktor yang lebih tinggi (misalnya aritmetika dan konsep aljabar) dibanding dengan bobot faktor komponen-komponen di subtes yang lain. Secara spesifik pada internal komponen-komponen di dalam subtes juga ditemukan temuan yang konsisten. Pada subtes verbal, komponen yang paling dominan dalam merepresentasikan kemampuan yang diukur Tes PAPS adalah Analogi Kata.

Tabel 3.

Bobot Faktor dari Komponen pada Masing-Masing Seri Tes PAPS

Subtes	Komponen Tes	Form Tes PAPS					
		A1(R)	A2(R)	B1	B2	C1	C2
Penalaran Verbal	Padanan Kata	0.379	0.308	0.316	0.371	0.464	0.429
	Lawan Kata	0.497	0.494	0.342	0.342	0.547	0.556
	Analogi Kata	0.551	0.559	0.638	0.558	0.602	0.621
	Pemahaman Wacana	0.452	0.501	0.559	0.445	0.482	0.596
Penalaran Kuantitatif	Deretan Angka	0.654	0.723	0.553	0.511	0.622	0.701
	Aritmetika	0.717	0.748	0.558	0.726	0.753	0.736
	Konsep Aljabar	0.714	0.668	0.459	0.523	0.719	0.624
	Perb. Kuantitatif	0.555	0.676	0.471	0.587	0.621	0.641
Penalaran Figural	Logis	0.508	0.511	0.442	0.470	0.598	0.504
	Diagram	0.501	0.368	0.440	0.392	0.552	0.317
	Analitis	0.594	0.521	0.487	0.516	0.688	0.618
	Serial	0.635	0.525	0.575	0.509	0.619	0.720
	Klasifikasi	0.379	0.308	0.316	0.371	0.464	0.429

Komponen ini memiliki bobot faktor yang tinggi melebihi komponen-komponen lainnya di subtes yang sama. Sementara itu pada subtes kuantitatif, komponen aritmetika dan aljabar merupakan komponen dengan bobot faktor yang cenderung tinggi. Di sisi lain, pada subtes penalaran, penalaran serial cenderung konsisten memberikan kontribusi paling tinggi dalam pengukuran kemampuan penalaran akademik. Secara keseluruhan, dari semua komponen Tes PAPS yang ada, komponen-komponen dalam subtes kuantitatif merupakan komponen yang paling banyak memberikan kontribusi dalam pengukuran penalaran akademik.

Diskusi

Penelitian ini bertujuan untuk membuktikan validitas konstruk Tes PAPS. Prosedur yang dipakai adalah dengan menguji apakah struktur Tes PAPS bersifat unidimensi. Struktur unidimensi menunjukkan bahwa antara komponen satu dengan komponen lain dalam sebuah tes memiliki hubungan yang erat. Secara praktis unidimensionalitas ini dapat ditunjukkan melalui ketepatan model unidimensi pada data yang dilibatkan. Hasil analisis membuktikan bahwa model unidimensionalitas Tes PAPS didukung oleh data. Simpulan ini didasarkan pada pemenuhan harga indeks ketepatan terhadap kriteria yang ditetapkan. Misalnya harga CFI dan TLI di atas 0,90 dan harga RMSEA dan SRMR di bawah 0,08. Namun demikian pada beberapa seri, ketepatan model yang tinggi ini dilakukan melalui pelibatan korelasi antar error.

Temuan ini mendukung beberapa pernyataan yang dilontarkan oleh para ahli pengukuran tes potensi. Misalnya, pernyataan bahwa setiap tes yang mengukur kemampuan kognitif hampir

pasti dapat dikatakan akan mengukur satu atribut kemampuan saja (Gustafsson, 1988). Penyebabnya adalah karena setiap kemampuan yang diukur oleh subtes memiliki keterkaitan dengan kemampuan yang diukur oleh subtes lainnya. Setiap subtes di dalam tes potensi tidak bersifat independen. Misalnya mengukur kemampuan verbal akan terkait dengan kemampuan bahasa (penguasaan kosakata), mengukur kemampuan penalaran kuantitatif juga akan mengukur kemampuan matematika. Di sisi lain, performansi kognitif individu dipengaruhi oleh variasi dari sejumlah kemampuan yang memiliki tingkat yang berbeda-beda dan memiliki struktur hierarkis (abilitas rendah hingga tinggi). Adanya fakta ini memunculkan fenomena bahwa tes potensi hanya mengukur satu atribut kemampuan. Namun demikian persoalan mengenai bagaimana mengoperasionalkan hasil pengukuran dalam bentuk skor tes. Skor tes dapat diwujudkan dalam satu skor komposit dapat juga kemampuan-kemampuan lain yang juga diukur perlu dilaporkan sebagai skor subtes. Misalnya sub skor kemampuan bahasa dan kemampuan matematika. Perhatian terhadap hal ini akan berkaitan dengan validitas konstruk dan validitas prognostik sebuah tes. Dari titik inilah penelitian untuk menggali informasi mengenai struktur faktor dalam konstruk ukur merupakan sebuah upaya yang sangat penting (Åberg-Bengtsson, 2005).

Faktor lain yang menyebabkan belum maksimalnya struktur unidimensionalitas Tes PAPS dapat disebabkan oleh munculnya struktur hierarkhis di dalam konstruk ukur Tes PAPS. Gustafsson, Westerlund, dan Wedman (1992) meneliti sepuluh pengukuran dengan menggunakan *Swedish Scholastic Aptitude Test* (SweSAT). Mereka menemukan untuk konstruk

dengan struktur dua faktor yang bersifat heirarkis. Selain faktor umum yang memuat seluruh komponen tes, mereka juga menemukan beberapa sub faktor. Setelah dikaji kontennya sub faktor yang mereka temukan adalah (a) pengetahuan, yang terkait dengan penguasaan kosa kata, membaca, dan wawasan mengenai informasi umum (GI) dan (b) teknik studi (STECH) yang terkait kemampuan untuk memahami sebuah wacana. Mereka juga menemukan bahwa faktor umum memiliki korelasi yang tinggi dengan dua komponen tes, yaitu tes diagram, tabel, dan peta (DTM) dan tes kecukupan data yang (DS) yang diasumsikan untuk mengukur kemampuan analitis.

Melalui bobot faktor tes, penelitian ini mendapatkan bahwa komponen-komponen kuantitatif cenderung memiliki bobot faktor yang tinggi dibanding dengan komponen verbal maupun figural. Komponen verbal secara konsisten meskipun memiliki derajat yang dapat diterima secara metodologis (>0.30) cenderung paling rendah di semua seri tes yang dikaji. Di dalam internal subtes verbal, bobot faktor soal tipe padanan kata dan lawan kata harganya konsisten di bawah tipe soal lainnya, misalnya analogi verbal dan pemahaman wacana (*critical reading*). Hasil ini didukung oleh pernyataan Ward (1982) yang mengatakan bahwa komponen padanan dan lawan kata merupakan tes yang cenderung mengukur kemampuan verbal dibanding dengan subtes yang menekankan pada penalaran. Berbeda dengan pendapat tersebut, Sincoff dan Sternberg (1987) mengatakan bahwa tes sinonim dan antonim memiliki korelasi yang tinggi dengan kemampuan kognitif secara umum. Namun demikian, kedua tes tersebut lebih banyak mengukur hasil belajar daripada penalaran. Artinya,

kedua komponen tes tersebut sangat berkaitan dengan faktor belajar dibanding dengan faktor belajar yang merupakan fokus dari Tes PAPS.

Di dalam subtes verbal, jenis tes yang paling memberikan kontribusi terhadap variasi skor adalah tes analogi. Rendahnya keterkaitan antara tes analogi dengan masalah verbal memungkinkan menjadi penyebabnya. Namun demikian, hasil penelitian ini berkebalikan dengan pernyataan Atkinson yang mengatakan bahwa tes analogi lebih banyak mengukur aspek pengetahuan daripada penalaran (Zwick, 2004). Melibatkan tes dengan menggunakan unsur verbal merupakan tantangan tersendiri dalam tes potensi karena potensi masuknya unsur pengetahuan di dalam skor tes ini sangat besar. Perbedaan tingkat pendidikan, budaya dan latar belakang antara individu memungkinkan untuk memengaruhi keberhasilan mereka dalam mengatasi soal-soal yang terkait dengan kosa kata yang merupakan representasi dari kemampuan verbal.

Di dalam subtes kuantitatif, subtes yang paling memberikan kontribusi dalam variasi skor adalah tes aritmetika. Bobot faktor yang dihasilkan dari tes jenis ini paling tinggi dibanding dengan jenis-jenis tes lainnya. Tes aritmetika memiliki beberapa fitur yang khas dibanding dengan tes lainnya karena di dalamnya terkandung aspek berpikir secara sistematis dan strategis, kecepatan dalam memproses dan ketelitian. Kekhasan ditambah dengan minimnya faktor verbal yang turut memengaruhi sehingga skor yang dihasilkan lebih banyak dipengaruhi oleh faktor spontanitas. Temuan ini memiliki kesamaan (misal: Visser, Ashton, & Vernon, 2006) dan juga berbeda (misal: Wallbrown, Carmin, & Barnett, 1988) dengan hasil penelitian sebelumnya.

Kesamaan dan perbedaan ini disebabkan oleh seberapa jauh bentuk tes aritmetika yang dipakai menekankan pada penalaran aritmetika (*arithmetic demand*). Semakin besar unsur penalarannya semakin besar bobot faktor yang dimuat oleh tes aritmetika karena pengaruh kemampuan matematika terhadap perolehan skor semakin minim.

Dari sisi pengembangan tes potensi kognitif, secara umum hasil penelitian ini menunjukkan faktor umum (*g factor*) yang ditargetkan untuk diukur tetap mendominasi varians skor yang didapatkan oleh peserta tes. Dengan kata lain, pengaruh variasi dan keunikan dari tiap jenis-jenis tes yang dilibatkan relatif minim. Minimnya keunikan dari komponen subtes serta subtes juga terlihat dari munculnya kasus Heywood dalam analisis. Kasus Heywood muncul karena ada korelasi yang sangat besar antar variabel dalam model yang menyebabkan munculnya hasil statistika yang janggal seperti harga varians yang negatif dan koefisien estimasi terstandar yang melebihi 1. Hasil seperti ini merupakan hal yang umum dalam pemodelan dengan menggunakan tes inteligensi (Flanagan & Dixon, 2005; Taub & McGrew, 2004).

Hasil penelitian ini menegaskan hubungan substansial antara skor Tes PAPS dan faktor *g* dalam teori inteligensi. Sebelumnya, Brodnick dan Ree (1995) melakukan penelitian untuk mengidentifikasi hubungan antara faktor *g*, variabel sosial ekonomi, nilai tes prestasi dan variabel terkait dengan potensi akademik dengan menggunakan pendekatan pemodelan struktural untuk menguji hubungan. Mereka menemukan harga bobot faktor terkait kecerdasan umum pada dua subtes yaitu matematika dan verbal pada tes SAT sangat tinggi. Hal ini menunjukkan kedua kemampuan tersebut merupakan repre-

sentasi dari kecerdasan. Namun demikian beberapa penulis mengajukan kritik atas temuan tersebut (Frey & Detterman, 2004). Salah satunya adalah karena peneliti tersebut menggunakan tes potensi akademik yang dikembangkan dengan justifikasi mereka terhadap faktor *g*, bukan faktor *g* yang diukur dari tes-tes inteligensi yang terstandarisasi. Paparan mengenai apakah faktor *g* ini adalah sama dengan faktor *g* yang diperoleh dari tes kecerdasan terstandar tidak dijelaskan secara rinci oleh penulis.

Kesimpulan

Temuan dari penelitian ini menunjukkan bahwa Tes PAPS merupakan tes yang cenderung bersifat unidimensi karena kecenderungannya dalam mengukur satu atribut kemampuan yang bersifat tunggal lebih besar daripada mengukur atribut kemampuan yang bersifat majemuk. Temuan ini menunjukkan bahwa unidimensionalitas merupakan suatu derajat dan bukan merupakan fenomena yang bipolar antara dua kutub: unidimensi ataukah multidimensi.

Saran

Penelitian lanjutan untuk mengeksplorasi proporsi faktor umum dapat melibatkan model-model pengukuran lanjutan misalnya model bifaktor (Reise, 2012) atau model-model yang menggunakan pendekatan *multi-trait multi-method* dari Eid (Eid, Lischetzke, Nussbeck, & Trierweiler, 2003). Model-model pengukuran tersebut dapat mengidentifikasi seberapa besar porsi penalaran umum dan penalaran yang lebih spesifik di dalam skor yang muncul dari Tes PAPS (lihat Widhiarso, 2016).

Kepustakaan

- Åberg-Bengtsson, L. (2005). Separating quantitative and analytic dimensions in the Swedish Scholastic Aptitude Test (SweSAT). *Scandinavian Journal of Educational Research*, 49(4), 359–383. doi: [10.1080/00313830500202892](https://doi.org/10.1080/00313830500202892)
- Anderson, J. C., & Gerbing, D. W. (1988). Structural Equation Modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, 103(3), 411–423.
- Armstrong, R. D., & Shi, M. (2009). A parametric cumulative sum statistic for person fit. *Applied Psychological Measurement*, 33(5), 14–19.
- Azwar, S., & Ridho, A. (2013). Abilitas komposit dalam tes potensi. *Jurnal Psikologi*, 40(2), 127–142.
- Brodnick, R. J., & Ree, M. J. (1995). A structural model of academic performance, socioeconomic status, and Spearman's g . *Educational and Psychological Measurement*, 55(4), 583–594. doi: [10.1177/0013164495055004006](https://doi.org/10.1177/0013164495055004006)
- Cook, M. D. (2003). *Personnel selection. Adding value through people*. Chichester: John Wiley & Sons.
- Dienstag, J. L. (2011). The medical college admission test – Toward a new balance. *New England Journal of Medicine*, 365(21), 1955–1957. doi: [10.1056/NEJMp1110171](https://doi.org/10.1056/NEJMp1110171)
- Domino, G., & Domino, M. L. (2006). *Psychological testing an introduction*. Cambridge: Cambridge University Press.
- Eid, M., Lischetzke, T., Nussbeck, F. W., & Trierweiler, L. I. (2003). Separating trait effects from trait-specific method effects in multitrait – multimethod models: A Multiple-Indicator CT-C(M-1) model. *Psychological Methods*, 8(1), 38–60. doi: [10.1037/1082-989X.8.1.38](https://doi.org/10.1037/1082-989X.8.1.38)
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, 128(3), 309–331. doi: [10.1037/0096-3445.128.3.309](https://doi.org/10.1037/0096-3445.128.3.309)
- Everson, H. T. (2003). Innovation and change in the SAT: A design framework for future college admission tests. In R. Zwick (Ed.), *The SAT rethinking the future of standardized testing in university admissions*. New York, NY: Routledge.
- Flanagan, D. P. & Dixon, S. G. (2005). The Cattell-Horn-Carroll theory of cognitive abilities. *Contemporary Intellectual Assessment: Theories, Tests, and Issues*, 136–181. doi: [10.1002/9781118660584.ese0431](https://doi.org/10.1002/9781118660584.ese0431)
- Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g ? *Psychological Science*, 15(6), 373–378. doi: [10.1111/j.0956-7976.2004.00687.x](https://doi.org/10.1111/j.0956-7976.2004.00687.x)
- Furr, R. M., & Bacharach, V. R. (2008). *Psychometrics: An introduction*. Los Angeles: Sage Publications.
- Gre, E. T. S., Walters, A. M., & Kaufman, J. C. (2011). The role of noncognitive constructs and other background variables in graduate education. *ETS Research Reports Series*, (RR-11-12 April), 1–131.
- Gustafsson, J. E. (1988). Hierarchical models of individual differences and cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*. Hillsdale, NJ.: Lawrence Erlbaum Associates, Inc.
- Gustafsson, J. E., Westerlund, A., & Wedman, I. (1992). The dimensionality of the Swedish Scholastic Aptitude

- Test. *Scandinavian Journal of Educational Research*, 36(1), 21–39. doi: [10.1080/0031383920360102](https://doi.org/10.1080/0031383920360102)
- Harvey, R. J., Billings, R. S., & Nilan, K. J. (1985). Confirmatory factor analysis of the job diagnostic survey: Good news and bad news. *Journal of Applied Psychology*, 70(3), 461–468. doi: [10.1037/0021-9010.70.3.461](https://doi.org/10.1037/0021-9010.70.3.461)
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. doi: [10.1080/10705519909540118](https://doi.org/10.1080/10705519909540118)
- Jenkins, D. J. (2016). *The predictive validity of the General Scholastic Aptitude Test (GSAT) for first-year students in information technology*. University of Zululand, Kwadlangezwa.
- Joreskog, K. G., & Sorbom, D. (1990). Model search with Tetrad-II and Lisrel. *Sociological Methods & Research*, 19(1), 93–106.
- Kobrin, J. L., Camara, W. J., & Milewski, G. B. (2002). *Jennifer L. Kobrin, Wayne J. Camara, and Glenn B. Milewski. College board research report*. New York, NY.
- Matsunaga, M. (2008). Item parceling in Structural Equation Modeling: A primer. *Communication Methods and Measures*, 2(4), 260–293. doi: [10.1080/19312450802458935](https://doi.org/10.1080/19312450802458935)
- Reise, S. P. (2012). The rediscovery of bifactor measurement models, 667–696. doi: [10.1080/00273171.2012.715555](https://doi.org/10.1080/00273171.2012.715555)
- Rock, D. A., Werts, C., & Grandy, J. (1982). Construct validity of the gre aptitude test across populations: An empirical confirmatory study. *ETS Research Report*, 78–1P, 81–57.
- Sincoff, J. B., & Sternberg, R. J. (1987). Two faces of verbal ability. *Intelligence*, 11(4), 263–276. doi: [10.1016/0160-2896\(87\)90010-9](https://doi.org/10.1016/0160-2896(87)90010-9)
- Spearman, C. (1904). General intelligence, objectively determined and measured. *American Journal of Psychology*, 15, 201–293.
- 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. *School Psychology Quarterly*, 19(1), 72–87.
- Thacker, J. W., Fields, M. W., & Tetrick, L. E. (1989). The factor structure of Union Commitment: An application of confirmatory factor analysis. *Journal of Applied Psychology*, 74(2), 228–232. doi: [10.1037/0021-9010.74.2.228](https://doi.org/10.1037/0021-9010.74.2.228)
- Visser, B. A., Ashton, M. C., & Vernon, P. A. (2006). Beyond g: Putting multiple intelligences theory to the test, 34, 487–502. doi: [10.1016/j.intell.2006.02.004](https://doi.org/10.1016/j.intell.2006.02.004)
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica*, 57, 307–333.
- Wallbrown, F. H., Carmin, C. N., & Barnett, R. W. (1988). Investigating the construct validity of the multidimensional aptitude battery. *Psychological Reports*, 62(3), 871–878. doi: [10.2466/pr0.1988.62.3.871](https://doi.org/10.2466/pr0.1988.62.3.871)
- Ward, L. C., Ryan, J. J., & Axelrod, B. N. (2000). Confirmatory Factor Analyses of the WAIS-III standardization data. *Psychological Assessment*, 12(3), 341–345. doi: [10.1037//1040-3590.12.3.341](https://doi.org/10.1037//1040-3590.12.3.341)
- Ward, W. C. (1982). A comparison of free-response and multiple-choice forms of verbal aptitude tests. *Applied Psychological Measurement*, 6(1), 1–11. doi: [10.1177/014662168200600101](https://doi.org/10.1177/014662168200600101)

- 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](https://doi.org/10.21070/psikologia.v1i1.478)
- Widhiarso, W., Azwar, S., Suhapti, R., & Haryanta. (2015). Analisis dan penyempurnaan Tes PAPS Seri A1. *Seri Technical Report UPAP*, 2(2), 1–7.
- 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. doi: [10.5964/ejop.v11i3.865](https://doi.org/10.5964/ejop.v11i3.865)
- Widhiarso, W., & Haryanta. (2016). Comparing the performance of synonym and antonym tests in measuring verbal abilities. *TPM - Testing, Psychometrics, Methodology in Applied Psychology*, 23(3), 335–345. doi: [10.4473/TPM23.3.5](https://doi.org/10.4473/TPM23.3.5)
- Widhiarso, W. & Suhapti, R. (2018). Penggunaan testlet dalam pengembangan tes psikologi. *Insan Jurnal Psikologi dan Kesehatan Mental*, 3(1).
- Wulan, R. (1996). SPM untuk mengukur inteligensi. *Jurnal Psikologi*, 2, 67–73.
- Young, J. W., Klieger, D., Bochenek, J., Li, C., & Cline, F. (2014). The validity of scores from the GRE® revised general test for forecasting performance in business schools: Phase one. *ETS Research Reports Series*, (GRE-14-01, RR-14-17), 1–10. doi: [10.1002/ets2.12019](https://doi.org/10.1002/ets2.12019)
- Zwick, R. (2004). Part I: Standardized tests and American education: What Is the past and future of college admissions testing in the United States? In R. Zwick (Ed.), *SAT Rethinking the future of standardized testing in university admissions*. New York, NY: Routledge.