

## **PENGUJIAN PROPERTI PSIKOMETRIK *INTELLIGENZ STRUKTUR TEST* SUBTES KEMAMPUAN SPASIAL DUA DIMENSI (*FORM AUSWAHL*): STUDI PADA DUA SMA SWASTA DI JAKARTA**

Indro Adinugroho  
Fakultas Psikologi, Unika Atma Jaya  
[indro.adinugroho@atmajaya.ac.id](mailto:indro.adinugroho@atmajaya.ac.id)

### **Abstrak**

*Intelligenz Struktur Test* (IST) adalah salah satu jenis alat ukur yang mengukur konstruk inteligensi manusia. Sebagai alat ukur inteligensi, IST menghasilkan skor mean dalam konteks inteligensi yang disebut dengan *Intelligence Quotient* (IQ). IST dikembangkan oleh Rudolf Amthauer pada tahun 1953 di Jerman dengan teori dasar *Primary Mental Abilities* (PMA). PMA adalah teori inteligensi dasar yang dikembangkan oleh Lois Leon Thurstone dimana teori ini mendeskripsikan tujuh kemampuan dasar dalam tataran kognitif manusia. Hingga saat ini, IST masih digunakan oleh Biro Layanan Psikologi, Fakultas Psikologi Unika Atma Jaya sebagai instrumen utama guna keperluan asesmen dan evaluasi. Studi ini bertujuan untuk menguji properti psikometrik salah satu subtes dalam IST yang mengukur kemampuan spasial dua dimensi, yaitu subtes *Form Auswahl* (FA). Metode uji properti psikometrik akan melibatkan analisis item; validasi *convergent – discriminant* dan uji reliabilitas dengan *split-half*. Hasil analisis item menunjukkan bukti empiris untuk mempertahankan seluruh item. Akan tetapi, analisis distraktor menunjukkan hasil yang berbeda. Ada kebutuhan untuk merevisi pilihan jawaban dalam soal karena ketidakmampuan distraktor untuk mengecoh respon individu. Secara umum, uji validitas dan reliabilitas menunjukkan bahwa subtes FA masih sesuai untuk mengukur kemampuan spasial terhadap orang Indonesia.

Kata kunci: spasial, IST, *Form Auswahl*, properti psikometrik

### **Abstract**

*Intelligenz Struktur Test* (IST) is one-kind psychological measurement that has primary function measure human intelligence. As an intelligence test, IST produces mean score in the context of human ability called intelligence quotient (IQ). IST was developed by Rudolf Amthauer on 1953 in Germany, also with *Primary Mental Abilities* (PMA) as its theoretical framework. PMA is a basic concept of intelligence developed by Lois Leon Thurstone that describes seven primary abilities exists in human cognition. Until now, IST is still used in Psychological Services Bureau (PSB), Faculty of Psychology Atma Jaya Catholic University as their primary tool for assessment and evaluation purposes. This study aims to examine the psychometric properties of one of the subtest in IST that measures two-dimension spatial intelligence, *Form Auswahl* (FA). Techniques related to psychometric examination will include item analysis, convergent and discriminant validation and also split-half method for reliability testing. Item analysis shows the empirical support to retain all the items in FA. However, distractor-analysis shows different result. There is a need to revise some of the choices due to lack of capability to flam the answer of lower group. In conclusion, validity and reliability examinations show that FA is still appropriate for assessing two-dimension spatial ability in Indonesian context.

Keywords: Spatial, IST, *Form Auswahl*, psychometric properties

Dalam psikologi, inteligensi dipahami sebagai sebuah variabel yang menggambarkan kemampuan dasar manusia (*ability*) dan menyangkut berbagai kemampuan khusus (Thurstone, 1938). Setiap kemampuan khusus mampu berperan sebagai instrumen yang digunakan oleh manusia ketika menghadapi persoalan yang membutuhkan kapasitas kognitif. Misalnya, ketika seorang tentara harus mengatur strategi yang paling efektif untuk melakukan penyerangan musuh di hutan belantara. Strategi yang disusun akan berbeda ketika musuh berada di daerah perairan. Penyusunan strategi yang berbeda tentu akan membutuhkan kemampuan khusus yang berbeda pula, bahkan kombinasi dari lebih satu kemampuan khusus. Dalam ranah pengukuran psikologi, kemampuan khusus disebut sebagai *aptitude* (Anastasi & Urbina, 1997). Guna mengidentifikasi inteligensi yang merupakan gabungan dari berbagai *aptitude*, psikologi memiliki konsep yang disebut dengan *intelligence quotient* (IQ). IQ merujuk pada pemaknaan mengenai sejauh mana kemampuan dasar individu yang digambarkan melalui skor rata-rata dari berbagai aspek kemampuan khusus yang diperoleh individu melalui alat ukur inteligensi yang terstandar (Anastasi & Urbina, 1997).

Salah satu alat ukur inteligensi yang telah diadaptasi dalam konteks Indonesia adalah *Intelligenz Struktur Test* (IST). IST dibuat dan dikembangkan oleh Rudolf Amthauer di Jerman pada tahun 1953 (Polhaupessy, 2002). IST lahir dan dikembangkan mengacu kepada teori inteligensi yang dikemukakan L.L Thurstone yaitu *Primary Mental Abilities* (PMA). PMA adalah teori inteligensi yang memaparkan bahwa manusia memiliki tujuh kemampuan dasar yang saling terkait satu sama lain (Thurstone, 1938; Thurstone & Thurstone, 1941). Tujuh kemampuan dasar tersebut lahir dari

proses psikometrik bernama *factor analysis*, yaitu metode psikometrik dengan tujuan mengidentifikasi keterkaitan antar variabel yang disebut dengan faktor baik yang sudah ada (*confirmatory factor analysis*) maupun yang belum ada (*exploratory factor analysis*; Chadha, 2009) Metode analisis faktor dapat memberikan analisis mengenai kontribusi item secara statistik dan psikometrik terhadap faktor-faktor yang terbentuk. Teori PMA adalah teori yang lahir dari kritik Thurstone terhadap teori inteligensi *g factor* yang mendeskripsikan inteligensi sebagai satu kemampuan umum yang dimiliki oleh manusia yang bernama *general ability* atau *general factor* (*g*) yang dikemukakan Charles Spearman (1904;1927).

IST memiliki sembilan subtes dengan karakteristik item yang berbeda-beda. Setiap subtes dibuat untuk mengukur aspek kemampuan khusus pada manusia atau yang disebut dengan *aptitude*. Secara psikometrik, IST tergolong ke dalam *multiple aptitude batteries test*, yaitu tes yang disusun oleh serangkaian subtes dimana setiap subtes mengukur kemampuan khusus yang berbeda (Anastasi & Urbina, 1997). Nilai IQ dalam IST diperoleh dari nilai total rata-rata kemampuan khusus yang diperoleh oleh individu. Namun, interpretasi paling utama dari IST bukanlah merujuk kepada nilai IQ total, namun pada profil inteligensi yang dihasilkan dari 9 subtes pada IST. Misalnya, ketika individu memiliki skor yang tinggi pada subtes kemampuan verbal (*Gemeinsamkeiten* atau GE) dibandingkan dengan kemampuan spasial (*Form Auswahl* atau FA), bukan berarti bahwa selamanya individu tersebut tidak dapat melakukan pekerjaan berbasis kemampuan keruangan. Melalui pengalaman dan pembelajaran kemampuan khusus yang konsisten, maka skor kemampuan khusus tersebut pun dapat berubah.

IST pertama kali diadaptasi dalam konteks Indonesia oleh Fakultas Psikologi Universitas Padjajaran Bandung. Penggunaan IST sebagai alat ukur inteligensi yang menghasilkan profil kecerdasan individu telah berlangsung lama dan cukup luas, seperti seleksi calon peserta didik di institusi pendidikan dan calon karyawan di perusahaan. Salah satu institusi pendidikan yang menggunakan IST hingga saat ini adalah Biro Layanan Fakultas Psikologi Unika Atma Jaya. Sebagai alat tes yang diadaptasi pada tahun 1953 dan masih digunakan hingga saat ini, IST harus mendapat sentuhan pengujian properti psikometrik guna mengidentifikasi apakah item dan subtes dalam IST masih layak digunakan atau tidak. Kajian mengenai properti psikometrik sebuah alat tes menjadi sangat esensial karena kunci utama dari alat ukur dalam ilmu psikologi adalah soal presisi (validitas) dan konsistensi hasil pengukuran (reliabilitas) (Crocker & Algina, 1986; Anastasi & Urbina, 1997). Sebenarnya, studi terkait properti psikometrik terhadap IST pernah dilakukan oleh Santosa, Bonang, dan Panggabean (1998) terhadap seluruh subtes IST. Studi ini membahas mengenai kekuatan IST dalam memberikan prediksi terhadap skor IPK mahasiswa di Fakultas Keguruan dan Ilmu Pendidikan (FKIP); Fakultas Hukum (FH) dan Fakultas Psikologi (FP) Universitas Katolik Indonesia Atma Jaya. Hasil studi menunjukkan beberapa subtes mampu memberikan prediksi yang signifikan terhadap nilai IPK mahasiswa di FKIP, FH dan FP, yaitu subtes *Satzerganzung* (SE); *Analogien* (AN); *Merk Aufgaben* (ME), dan *Rachen Aufgaben* (RA). Namun, studi ini sudah berlangsung cukup lama dan hingga saat ini dunia telah mengalami perkembangan ilmu pengetahuan yang dirasa menjadi sebuah pemicu perlunya dilakukan kembali studi terkait properti psikometrik secara komprehensif terhadap IST dan subtes-subtes di dalamnya.

Studi ini merupakan salah satu rangkaian dari studi besar guna mengidentifikasi properti psikometrik keseluruhan subtes dalam IST secara detil. Secara khusus, studi ini hanya akan berfokus pada satu subtes dalam IST, yaitu subtes *Form Auswahl* (FA). Subtes FA dalam IST memiliki fungsi mengukur kemampuan spasial dua dimensi. Secara khusus, kemampuan spasial dua dimensi pada IST-FA dipahami sebagai kemampuan manusia untuk membayangkan objek yang memiliki bentuk konkret dua dimensi, seperti bangun datar seperti persegi, lingkaran, dan trapezium. Salah satu hal yang merupakan ciri utama dari objek dua dimensi adalah tidak memiliki *volume* (isi). Sebagai sebuah kemampuan khusus, manfaat dari kemampuan spasial terhadap kehidupan manusia tidak dapat dipandang sebelah mata. Studi yang dilakukan oleh Wanzel, Hamstra, Anastakis, Matsumoto, dan Cusimano (2002) mendeskripsikan bahwa individu dengan kemampuan spasial yang tinggi cenderung memiliki kemampuan bedah medis yang baik. Studi ini merekrut 37 mahasiswa kedokteran yang sedang mempelajari ilmu pembedahan. Setiap partisipan diminta untuk mengisi enam jenis alat ukur kemampuan spasial dari tingkat kesulitan sederhana hingga kompleks. Guna mengidentifikasi kemampuan mereka dalam melakukan operasi bedah, Wanzel et al. (2002) menggunakan *role play* operasi pembedahan yang telah terstandarisasi untuk menguji kemampuan operasi bedah partisipan bernama *two-flap Z-plasty* dan *four-flap Z-plasty*. *Role play* merujuk kepada sebuah aktivitas yang berperan sebagai ilustrasi dari aktivitas yang sebenarnya. Hasil studi menunjukkan bahwa partisipan dengan kemampuan spasial yang tinggi cenderung memiliki kemampuan pembedahan medis yang lebih baik dibandingkan dengan partisipan yang memiliki kemampuan spasial rendah.

Studi yang mengidentifikasi properti psikometrik terhadap subtes FA akan dilakukan dalam tiga tahap penting, yaitu: (1) analisis item, (2) validitas, dan (3) reliabilitas (Cohen & Swerdlik, 2009). Tahap pertama adalah analisis item guna mengidentifikasi kualitas setiap item dalam membedakan partisipan dengan kemampuan spasial tinggi dan rendah. Guna mencapainya, penulis menggunakan metode analisis *item difficulty*, *item discrimination*, dan analisis distraktor. Ketiganya dilakukan sebagai masukan empiris mengenai item mana saja yang masih berkualitas dan yang tidak berkualitas. Tahap kedua adalah melakukan uji validitas dengan metode *convergent-discriminant* guna mengidentifikasi kemampuan subtes FA secara keseluruhan (kombinasi item-item) untuk membedakan partisipan dengan kemampuan spasial dua dimensi yang rendah dan tinggi. Teknik ini mengharuskan penulis melakukan uji korelasi antara subtes FA dengan subtes lain yang memiliki konstruk serupa (*convergent*) dan alat tes lain yang memiliki konstruk berbeda (*discriminant*) (Cohen & Swerdlik, 2009). Tahap ketiga adalah melakukan uji reliabilitas dengan membagi alat tes yang telah melalui proses analisis item ke dalam dua belahan yang setara atau disebut dengan *split-half*. Kesimpulan apakah subtes FA masih layak digunakan atau tidak akan diperoleh melalui analisis yang mendalam dari tiga tahapan tersebut.

### **Primary Mental Abilities (PMA) sebagai Landasan Teori IST**

Teori PMA pertama kali dikembangkan oleh Lois Leon Thurstone, seorang tokoh psikologi yang memiliki berbagai kontribusi terhadap lahir dan berkembangnya teori inteligensi, statistik serta analisis faktor dalam psikologi. PMA adalah sebuah teori yang mengkritik teori inteligensi *g-factor* yang dibuat oleh Charles Spearman (1904; 1927). Teori *g-factor* mengidentifikasi inteligensi sebagai

komponen yang terdiri dari beragam aktivitas intelektual yang merujuk kepada satu kemampuan dasar yang disebut dengan *general ability (g)*. *General ability (g)* mengidentifikasi eksistensi kemampuan-kemampuan khusus yang dimiliki oleh manusia dan setiap kemampuan khusus memiliki kontribusi terhadap kemampuan umum (*g*). Terdapat empat kemampuan khusus yang diidentifikasi oleh teori *g-factor*, yaitu kemampuan mekanis (*mechanical*); kemampuan berpikir logis (*logical*); kemampuan keruangan (*spatial*) dan kemampuan numerik (*arithmetic*). Salah satu alat ukur inteligensi dengan kerangka teori *g-factor* adalah *Raven Progressive Matrices (RPM; Raven, 2000)*. RPM memandang inteligensi manusia sebagai kombinasi (agregat) skor total alat ukur dengan menyesuaikan pada usia biologis partisipan. Artinya, ada standar baku yang ditetapkan dalam norma RPM di setiap jenjang usia biologis terkait dengan nilai minimum inteligensi yang seharusnya diperoleh oleh partisipan.

Jika *g-factor theory* memandang bahwa inteligensi adalah satu kemampuan utama yang terbentuk dari empat kemampuan khusus, maka berbeda halnya dengan PMA sebagai landasan teori alat ukur IST. Teori PMA tidak mengidentifikasi bahwa manusia hanya memiliki satu kemampuan dasar, melainkan tujuh kemampuan dasar. Kemampuan-kemampuan dasar ini disebut sebagai *Primary Mental Abilities (Thurstone, 1938; Guilford, 1972)*. Tujuh kemampuan dasar tersebut, adalah *verbal comprehension (V)*; *word fluency (W)*; *number (N)*; *space (S)*; *associative memory (M)*; *perceptual speed (P)*; *general reasoning (I)*. Setiap kemampuan dasar berkaitan dengan beberapa subtes yang mengukur konstruk serupa. Tujuh kemampuan dasar ini diperoleh melalui analisis faktor yang bertujuan mengidentifikasi kontribusi dari dimensi alat ukur terhadap faktor yang terbentuk.

Tabel 1 menunjukkan skema dasar PMA yang terbagi ke dalam tujuh kemampuan dasar atau yang disebut dengan faktor utama. Setiap faktor dapat diukur dengan dimensi perilaku berbeda yang dapat direpresentasikan dengan subtes. Subtes merupakan representasi yang lebih konkret dari faktor utama atau yang disebut dengan *primary factor*. Kombinasi dari tujuh faktor utama dalam teori PMA tetap mampu menghasilkan nilai IQ, namun terlepas dari itu, profil kemampuan dasar yang dihasilkan oleh teori PMA dari tujuh faktor utama memiliki peran yang lebih penting dalam mendeskripsikan kemampuan manusia.

**Tabel 1: Struktur Primary Mental Abilities (Thurstone, 1938)**

Faktor utama	Dimensi
<i>Verbal comprehension (V)</i>	Pemahaman dalam membaca Analogi verbal Memahami kalimat yang tidak terstruktur Penalaran verbal
<i>Word fluency (W)</i>	Mengenai nama objek dalam kategori tertentu Mengenali rima kata
<i>Number (N)</i>	Kecepatan dan akurasi dalam problem matematika
<i>Space (S)</i>	Kemampuan membayangkan objek (2 & 3 dimensi) Membayangkan objek secara visual dengan rotasi tertentu
<i>Associative memory (M)</i>	Mengingat objek Mengingat kata (verbal)
<i>Perceptual speed (P)</i>	Mengenai detail visual terkait persamaan dan perbedaan dari sebuah objek
<i>General reasoning (I)</i>	Berpikir secara induktif Berpikir secara deduktif

Sebagai teori besar yang menjelaskan inteligensi, teori PMA telah digunakan sebagai kerangka teori berbagai

alat ukur inteligensi dan salah satunya adalah *Intelligenz Struktur Test (IST)* yang dikembangkan oleh Rudolf Amthauer di Jerman pada tahun 1953 (Polhaupessy, 2002). IST memiliki sembilan subtes yang merepresentasikan teori PMA. Kombinasi skor dari beberapa subtes mampu merepresentasikan tujuh faktor utama PMA. Misalnya, faktor *space (PMA)* dapat diukur dengan subtes *Form Auswahl (FA)* yang mengukur kemampuan spasial dua dimensi dan subtes *Wurfel Aufgaben (WU)* yang mengukur kemampuan spasial tiga dimensi. Artinya, partisipan dikatakan memiliki kemampuan keruangan yang baik (*space*) diidentifikasi dari skor total subtes FA dan WU.

### ***Intelligenz Struktur Test* yang Digunakan di Indonesia**

Di Indonesia, IST pertama kali dibawa dan diadaptasi oleh Fakultas Psikologi, Universitas Padjajaran Bandung. Penggunaan IST sebagai alat ukur inteligensi yang dapat menghasilkan profil inteligensi telah diaplikasikan dalam berbagai konteks seperti seleksi karyawan di dunia kerja hingga penentuan peminatan di sekolah menengah atas (SMA). Salah satu institusi yang masih menggunakan IST hingga saat ini adalah Biro Layanan Psikologi (BLP) Fakultas Psikologi (FPSI) Unika Atma Jaya. Namun, sebelum IST digunakan oleh BLP FPSI Unika Atma Jaya, terdapat tiga fakultas di universitas Atma Jaya yang telah menggunakan IST sebagai instrumen seleksi calon mahasiswa tingkat S1 pada tahun 1997 yaitu Fakultas Psikologi (FPSI); Fakultas Hukum (FH) dan Fakultas Ilmu Keguruan dan Ilmu Pendidikan pada tahun 1998 (FKIP; Santosa et. al, 1998). Ketiga fakultas tersebut menggunakan IST sebagai instrumen seleksi mahasiswa baru program S1.

Sebagai instrumen utama yang digunakan oleh Unika Atma Jaya, Fakultas Psikologi pernah melakukan studi terkait uji psikometrik terhadap keseluruhan subtes dalam IST (Santosa et. al, 1998).

Pengujian properti psikometrik melibatkan 461 mahasiswa FPSI, FH dan FKIP Unika Atma Jaya sebagai partisipan. Pengujian properti psikometrik yang dilakukan melibatkan uji validitas prediktif terhadap seluruh subtes IST. Uji validitas prediktif merujuk pada pengujian validitas yang terkait dengan kemampuan suatu alat ukur memberikan prediksi terhadap perilaku atau konstruk lain yang terkait dengan alat ukur (Cohen & Swerdlik, 2009). Ketika melakukan uji validitas prediktif, akan ada dua elemen utama, yaitu skor alat ukur (skor mentah) dan kriteria eksternal. Kriteria eksternal merujuk kepada dimensi perilaku yang terkait dengan konstruk yang diukur dalam sebuah alat ukur psikologi (Anastasi & Urbina, 1997; Cohen & Swerdlik, 2009). Misalnya, alat ukur yang mengukur kemampuan spasial dapat dikaitkan dengan kriteria eksternal berupa catatan dari atasan mengenai performa supir taksi dalam mengidentifikasi rute jalan yang paling efisien. Pengujian validitas prediktif yang dilakukan oleh Santosa, et. al (1998) menggunakan nilai Indeks Prestasi Semester I (IPS) mahasiswa FPSI, FH dan FKIP. Secara teknis, uji validitas akan dilakukan dengan teknik statistik korelasi *Pearson* antara skor yang dihasilkan setiap subtes dalam IST dengan skor IPS partisipan. Korelasi *Pearson* merujuk pada teknik statistik yang dapat digunakan untuk mengidentifikasi nilai korelasi atau keterkaitan antar dua variabel yang dideskripsikan dengan nilai  $r$  dan arah korelasi (positif atau negatif; Howell, 2010). Sebanyak 27 nilai korelasi dihasilkan melalui korelasi setiap subtes IST dengan nilai IPS partisipan dari tiga fakultas. Hasilnya, sebanyak tujuh nilai  $r$  memiliki korelasi yang positif dan signifikan, yaitu antara subtes SE dengan IPS FKIP ( $r=.181$ ,  $p<.05$ ); subtes AN dengan IPS FKIP ( $r=.220$ ,  $p<.05$ ); subtes SE dengan IPS FH ( $r=.248$ ,  $p<.05$ ); subtes AN dengan IPS FH ( $r=.213$ ,  $p<.05$ ); subtes ME dengan IPS FH ( $r=.322$ ,  $p<.05$ ); subtes RA dengan IPS FH ( $r=.193$ ,  $p<.05$ ) dan

subtes RA dengan IPS FPSI ( $r=.199$ ,  $p<.05$ ). Studi ini menunjukkan bahwa beberapa subtes memiliki kemampuan prediksi yang tepat. Misalnya subtes ME yang mengukur kemampuan mengingat kata memiliki korelasi yang positif dan signifikan dengan nilai IPS FH. Hal ini sejalan dengan argumen bahwa mata kuliah yang dijalani oleh mahasiswa FH erat kaitannya dengan menghafal undang-undang dan berbagai aturan lain yang berlaku di Republik Indonesia.

Sayangnya, pengujian properti psikometrik yang telah dilakukan hanya menyorot aspek validitas prediktif. Uji properti psikometrik terkait validitas prediktif juga sudah terlampaui lama, yaitu tahun 1998. Dalam uji properti psikometrik yang melibatkan validitas prediktif tersebut tidak disertai dengan pengujian kualitas item dan konsistensi subtes dalam mengukur konstruk. Hal ini mengisyaratkan untuk dilakukannya uji validitas dan reliabilitas yang lebih komprehensif dan kontemporer guna menyorot item hingga kemampuan keseluruhan subtes dalam mengukur konstruk. Studi ini bermaksud untuk melakukan uji properti psikometrik yang lebih komprehensif terhadap salah satu subtes dalam IST, yaitu subtes FA. Studi subtes FA merupakan bagian dari studi besar pengujian properti psikometrik terhadap seluruh subtes IST yang dilakukan oleh Bagian Pengukuran Psikologi, FPSI Unika Atma Jaya.

### **Subtes *Form Auswahl* (FA) dan Kemampuan Spasial Dua Dimensi**

Sebagai alat ukur inteligensi yang menggunakan kerangka teori PMA, setiap subtes dalam IST mewakili tujuh kemampuan dasar manusia. Secara khusus, dalam konteks subtes FA, kemampuan dasar dalam kerangka PMA yang terukur adalah faktor *space* atau keruangan. Sebagai salah satu subtes yang mengukur kemampuan inteligensi, FA memiliki kemampuan untuk mengidentifikasi kemampuan spasial yang sifatnya dua

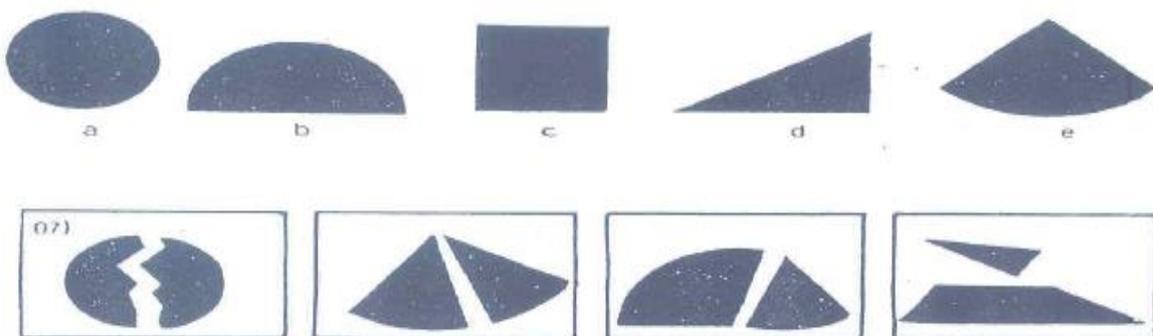
dimensi. Dua dimensi yang dimaksudkan dalam subtes FA adalah kemampuan individu untuk mengidentifikasi apakah sebuah objek dua dimensi memiliki bentuk yang serupa meskipun diletakkan, dirotasi dalam berbagai arah hingga dibelah (Polhaupessy, 2002). Artinya, kemampuan spasial pada individu bekerja ketika individu mampu memutuskan apakah dua atau lebih dari dua objek yang ditampilkan memiliki bentuk yang serupa atau tidak.

Sebagai salah satu dimensi kemampuan dasar (*primary ability*), kemampuan spasial tidak dapat berperan secara independen. Selalu ada irisan kemampuan-kemampuan lain yang memiliki keterkaitan dengan kemampuan spasial. Misalnya, ketika individu ingin pergi ke *mall* di daerah Pondok Indah, Jakarta Pusat, ada dua kemampuan dasar yang berperan disini. Kemampuan pertama adalah kemampuan terkait dengan mengingat objek (*memory*) dan kemampuan kedua adalah kemampuan spasial yang terkait dengan membayangkan dan mengidentifikasi rute jalan yang akan dilalui oleh individu untuk menuju ke tempat tersebut. Ilustrasi ini menggambarkan bahwa kemampuan spasial adalah kemampuan yang tidak hanya berkaitan dengan kemampuan matematis semata, misalnya mengerjakan soal matematika untuk mengidentifikasi panjang, lebar, dan tinggi sebuah bangun datar. Namun, besaran kemampuan yang

digunakan oleh individu dalam memahami dan menyelesaikan permasalahan sosial berbeda tergantung dengan permasalahannya.

Sebagai salah satu subtes yang bertujuan mengukur kemampuan spasial, subtes FA menetapkan persoalan mengenai bangun datar sebagai persoalan yang harus diselesaikan. Item dengan jenis ini diharapkan mampu memberikan stimulasi kepada partisipan untuk menggunakan proporsi kemampuan spasial yang lebih besar dibandingkan dengan kemampuan lainnya. Gambar 1 menunjukkan contoh item pada subtes FA. Pada setiap item, partisipan diminta untuk mengidentifikasi satu bangun datar (5 pilihan) yang tidak dapat dibentuk dengan bangun datar yang terbelah (4 pilihan).

Sebagai bagian dari alat ukur IST, subtes FA dapat digunakan sebagai skor yang membentuk kemampuan dasar space, maupun nilai agregat yang digunakan sebagai dasar kalkulasi nilai IQ. Ketika ingin mengidentifikasi hanya aspek kemampuan dasar space saja, maka skor subtes FA akan digabungkan dengan skor subtes WU yang mengukur kemampuan tiga dimensi. Nilai rata-rata dari penjumlahan skor di dua subtes tersebut nantinya mampu merepresentasikan kemampuan dasar space atau keruangan. Kesimpulannya, nilai subtes FA akan menjadi tidak bermakna ketika skor subtes FA diinterpretasikan secara tunggal.



**Gambar 1. Contoh Item pada Subtes FA**

## METODE

### Metode Pengumpulan Data

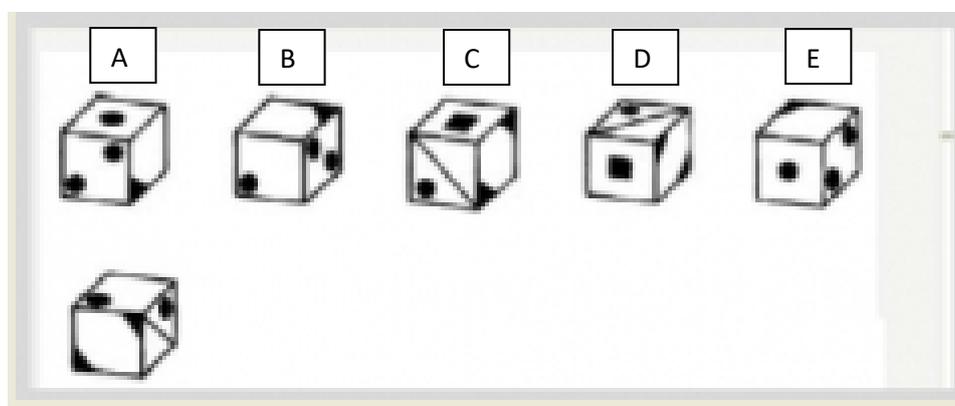
Data yang digunakan penulis dalam studi ini merupakan data sekunder (data telah tersedia) dari BLP FPSI Unika Atma Jaya. Data subtes FA dikumpulkan bersamaan dengan administrasi terstandar IST yang dilakukan oleh tim asesor dari BLP FPSI Unika Atma Jaya. Artinya, kontrol yang terstandarisasi oleh tim asesor BLP FPSI Unika Atma Jaya saat administrasi IST telah dilakukan sesuai prosedur standar. Pengumpulan data dilakukan di dua Sekolah Menengah Atas (SMA) swasta di Jakarta, yaitu SMA Kolese Kanisius dan SMA Gonzaga pada tahun 2004 hingga 2006. Saat itu, kedua sekolah ini adalah pengguna (*user*) dari layanan tes psikologi yang disediakan oleh BLP.

Metode pengambilan sampel yang digunakan adalah *non-purposive sampling* dimana partisipan yang berpartisipasi dalam studi tidak memiliki kesempatan yang sama untuk dipilih menjadi partisipan (Cozby & Bates, 2010). Karakteristik partisipan dalam studi adalah siswa/i SMA swasta yang akan masuk ke dalam kelas penjurusan Ilmu Pengetahuan Alam (IPA) dan Ilmu Pengetahuan Sosial (IPS). Jurusan IPA merujuk kepada mata pelajaran yang sebagian besar berisi ilmu-ilmu objektif dan membutuhkan perhitungan numerik yang akurat seperti

Fisika, Kimia, dan Matematika IPA; sedangkan jurusan IPS merujuk kepada mata pelajaran yang sebagian besar berisi pemahaman konseptual seperti sosiologi, antropologi dan ekonomi. Seluruh data yang digunakan dalam studi tidak disebarluaskan ke publik sehingga anonimitas partisipan sangat dijaga dalam *output* analisis yang dihasilkan.

### Pengujian Validitas *Convergent* dan *Discriminant* pada Subtes FA

Uji *convergent* merujuk kepada usaha psikometrik untuk mengidentifikasi apakah suatu alat ukur yang diuji memiliki korelasi dengan alat ukur lain yang mengukur konstruk serupa, sedangkan uji *discriminant* merujuk kepada mengidentifikasi bagaimana kaitan suatu alat ukur dengan alat ukur lain yang secara teori tidak memiliki keterkaitan (Crocker & Algina, 1986; Anastasi & Urbina, 1997; Cohen & Swerdlik, 2009). Pengambilan keputusan akan validitas sebuah alat ukur melalui metode *convergent-discriminant* sangat dipengaruhi oleh tiga hal. Pertama adalah kerangka teori yang digunakan. Kedua merujuk kepada besaran nilai korelasi ( $r$ ) yang dihasilkan. Ketiga terkait dengan arah korelasi yang dihasilkan oleh metode ini. Secara khusus, kerangka teori yang ditetapkan sebagai panduan interpretasi *convergent-discriminant* memiliki peran yang krusial.



Gambar 2. Contoh Item pada Subtes WU

Dalam konteks pengujian terhadap subtes FA, metode *convergent* akan dilakukan dengan melakukan korelasi statistik antara subtes FA dan subtes WU. Subtes WU adalah subtes dalam IST yang mengukur kemampuan spasial tiga dimensi (Polhaupessy, 2002). Artinya, dalam subtes WU partisipan harus mengenali dan mengidentifikasi objek bangun ruang mana saja yang memiliki persamaan dan perbedaan. Gambar 2 menunjukkan contoh item pada subtes WU dimana partisipan harus memilih bangun ruang yang memiliki persamaan dari lima pilihan jawaban yang disediakan (dengan gambar di bawahnya). Karakteristik item pada subtes WU cenderung sedikit berbeda dengan subtes FA, namun jika dilihat dengan kerangka teori PMA, keduanya sama-sama mengukur kemampuan dasar *space*. Terkait dengan metode *discriminant*, penulis akan melakukan korelasi antara subtes FA dengan subtes AN. Subtes AN adalah subtes dalam IST yang mengukur kemampuan analogi verbal yaitu terkait dengan bagaimana individu mengkombinasikan dan menghubungkan kata dengan makna yang sama (Polhaupessy, 2002). Item pada subtes AN terdiri dari tiga kata dan lima pilihan jawaban. Partisipan diminta untuk menempatkan kata yang paling sesuai dengan kata ketiga sesuai dengan pola hubungan antara pertama dan kedua. Tabel 2 menunjukkan salah satu contoh item pada subtes AN. Jika dikaitkan dengan teori PMA, aspek kemampuan dasar yang diukur oleh AN adalah aspek kemampuan dasar *general reasoning (I)*.

**Tabel 2: Contoh Item pada Subtes AN**

Item	Pilihan jawaban
Sikut : tangan = lutut : .....	a. Berjalan b. Jari tangan c. Kaki d. Hidung e. Mata kaki

Terkait dengan item ini, jawaban yang benar adalah pilihan kaki (c). Jika sikut merupakan bagian dari tangan, maka lutut adalah bagian

---

dari kaki. Kata kunci yang harus diidentifikasi adalah “bagian”.

---

### **Pengujian Reliabilitas terhadap IST-FA**

Selain validitas, salah satu syarat utama dari alat ukur psikologi yang mampu memberikan bukti empiris yang terpercaya adalah konsistensi hasil pengukuran yang dihasilkan. Konsep ini disebut dengan reliabilitas (Cohen & Swerdlik, 2009). Guna mengidentifikasi apakah subtes FA konsisten dalam mengukur kemampuan spasial dua dimensi, penulis menggunakan uji reliabilitas *split-half: Spearman-brown* guna mengidentifikasinya. Metode *split half* adalah metode uji reliabilitas dengan cara melakukan pembelahan menjadi dua bagian berdasarkan jumlah item yang proporsional, sedangkan *spearman brown* adalah teknik korelasi untuk mengidentifikasi koefisien reliabilitas alat ukur yang mengalami pembelahan *split-half*. Pembelahan alat ukur dilakukan dengan mempertimbangkan jumlah yang proporsional antar belahan hingga derajat kesulitan setiap item. Belahan pertama dan belahan kedua harus memiliki nilai rata-rata derajat kesulitan yang tidak berbeda jauh atau signifikan. Hal ini penting dilakukan karena perhitungan korelasi *spearman brown* (koefisien reliabilitas) sangat ditentukan oleh varians dari kedua belahan.

## **HASIL DAN PEMBAHASAN**

### **Karakteristik Partisipan**

Tiga karakteristik utama yang dicatat dalam studi terkait partisipan adalah asal sekolah, jenis kelamin dan tahun pengambilan data. Jumlah partisipan yang terlibat sebanyak 319 partisipan. Pengambilan data subtes FA dilakukan dalam metode administrasi tes IST yang terstruktur oleh tim asesor dari BLP FPSI Unika Atma Jaya. Artinya, seluruh standar, ketentuan dan batasan waktu yang disediakan di setiap subtes mengikuti

panduan standar IST. Tabel 3 menunjukkan demografi partisipan yang terlibat dalam studi.

**Tabel 3: Karakteristik Partisipan dalam Studi (N=319)**

Demografi partisipan	
Asal sekolah	SMA Gonzaga (66,70%) SMA Kolese Kanisius (30,09%)
Jenis kelamin	Laki-laki (69%) Perempuan (31%)
Tahun pengambilan data	Tahun 2004 (30,09%) Tahun 2005 (36,67%) Tahun 2006 (33,22%)

### Analisis Item pada Subtes FA

Kemampuan spasial dua dimensi yang diukur oleh subtes FA tergolong ke dalam *optimal performance test*. *Optimal performance test* (OPT) adalah suatu bentuk alat ukur yang mengidentifikasi kemampuan manusia yang dapat diklasifikasikan secara dikotomi, yaitu benar/salah atau tinggi/rendah (Anastasi & Urbina, 1997). Konsekuensinya, jenis item dalam OPT dinilai dengan penilaian benar atau salah untuk menentukan apakah individu mampu/tidak mampu dalam suatu aspek kemampuan dasar atau khusus tertentu. Selain skoring dan interpretasi skor, salah satu karakteristik utama dalam *optimal performance test* adalah adanya pemberian waktu yang telah dikalkulasi secara terstruktur sehingga mampu berperan sebagai variabel yang menentukan tercapainya kemampuan optimal dari individu saat mengerjakan tes. Dalam konteks subtes FA, batas waktu yang disediakan adalah 9 menit untuk menyelesaikan 20 item pilihan ganda.

Sebagai tahap pertama pengujian properti psikometrik terhadap subtes FA, penulis melakukan analisis item terhadap seluruh item pada subtes FA. Teknik analisis item yang digunakan adalah uji

derajat kesulitan item (*item difficulty*) dan uji diskriminasi item (*D-index*). Uji derajat kesulitan item adalah analisis item yang mampu mengidentifikasi tingkat kesulitan yang dimiliki setiap item (Anastasi & Urbina, 1997; Cohen & Swerdlik, 2009). Identifikasi tingkat kesulitan diperoleh melalui proporsi partisipan yang menjawab benar pada suatu item. Jika proporsi yang menjawab benar lebih banyak dibandingkan dengan proporsi yang menjawab salah, maka suatu item dikatakan mudah, begitu juga sebaliknya. Uji diskriminasi item merujuk kepada pengujian mengenai kemampuan item mendiskriminasi partisipan yang memiliki kemampuan tinggi dan rendah (Cohen & Swerdlik, 2009). Secara khusus *D-index* (D) adalah metode perhitungan diskriminasi item yang berbasis pada perbedaan skor kelompok *upper* dan kelompok *lower*. Uji statistik yang digunakan untuk mengidentifikasi nilai kontribusi item terhadap skor total/dimensi adalah uji korelasi (*r*). Tabel 4 menunjukkan hasil analisis item terhadap setiap subtes FA. Nomor item disesuaikan dengan nomor subtes FA sebagai bagian dari IST.

**Tabel 4: Hasil Pengujian Analisis Item terhadap Subtes FA**

No.	Derajat kesulitan (p)	<i>D-Index</i> (D)
117	0,98	0,047
118	0,55	1
119	0,48	1
120	0,52	1
121	0,84	0,495
122	0,72	0,857
123	0,40	1
124	0,47	1
125	0,60	1
126	0,54	1
127	0,33	1
128	0,33	0,99
129	0,73	0,819
130	0,44	1
131	0,77	0,704
132	0,14	0,438
133	0,59	1
134	0,49	1

135	0,20	0,609
136	0,11	0,323

Catatan: *Semakin tinggi nilai p, maka suatu item dikatakan memiliki tingkat kesulitan yang rendah, sedangkan nilai D yang mendekati nilai 1 mengindikasikan bahwa item memiliki kemampuan yang baik dalam membedakan kelompok partisipan yang memiliki kemampuan spasial dua dimensi tinggi dan rendah.*

Hasil analisis item mengindikasikan adanya ragam nilai yang diperoleh dari nilai D dan p. Tidak ada item yang dieliminasi dari hasil analisis item meskipun ditemukan nilai D yang cukup ekstrim pada item No. 117 ( $D=0,047$ ). Hal ini disebabkan item 117 memang dibuat sebagai item pembuka dengan tingkat kesulitan yang sangat mudah. Konsekuensinya, item ini tidak mampu membedakan kelompok *upper* dan kelompok *lower*. Rentang nilai p berkisar antara 0,98 (sangat mudah) hingga 0,11 (sangat sulit). Hasil ini mengindikasikan bahwa subtes FA telah memiliki variasi tingkat kesulitan yang mampu memotivasi partisipan (*test-taker*) untuk menampilkan kemampuannya yang paling optimal. Secara statistik, proporsi nilai 1 untuk nilai D yang berjumlah 11 ( $>9$ ) pada subtes FA juga menjadi pendukung bahwa mayoritas item dalam subtes FA masih memiliki daya diskriminasi item yang baik. Kesimpulannya, subtes FA telah memiliki prinsip proporsionalitas, yaitu semakin

tinggi nilai p, maka semakin kecil nilai D yang dihasilkan.

### Analisis Distraktor pada Pilihan Jawaban di Subtes FA

Subtes FA mengidentifikasi kemampuan spasial dua dimensi dengan item berjenis pilihan ganda. Item dengan jenis ini mengharuskan responden untuk menganalisis pilihan jawaban yang paling tepat di setiap item. Salah satu syarat kualitas item yang baik pada pilihan ganda adalah kemampuan distraktor yang mampu mengecoh jawaban partisipan (Crocker & Algina, 1986; Osterlind, 2002; Cohen & Swerdlik, 2009). Suatu distraktor dikatakan berkualitas ketika partisipan dengan kemampuan rendah dapat terkecoh dengan pilihan yang disediakan dalam alat ukur. Analisis distraktor dilakukan dengan membandingkan jawaban yang dibuat oleh kelompok *upper* dan kelompok *lower*. Penentuan keputusan mengenai kualitas pilihan jawaban di setiap item dilakukan atas dasar pilihan yang dibuat oleh kelompok *lower*. Seharusnya, kelompok *lower* sebagai kelompok yang mendapatkan skor rendah pada suatu item harus terkecoh dengan pilihan jawaban yang disediakan. Apabila partisipan yang terkecoh adalah kelompok *upper*, menandakan bahwa pilihan jawaban yang disediakan tidak memiliki kemampuan mengecoh yang baik. Tabel 5 mendeskripsikan analisis distraktor terhadap subtes FA.

**Tabel 5: Hasil Analisis Distraktor terhadap Pilihan Jawaban di Subtes FA**

No item	Kelompok	Pilihan Jawaban				
		A	B	C	D	E
117	<i>Upper</i>	107	0	2	0	0
	<i>Lower</i>	102	2	1	0	0
118	<i>Upper</i>	24	9	75	1	0
	<i>Lower</i>	32	17	51	0	0
119	<i>Upper</i>	9	53	1	23	21
	<i>Lower</i>	8	29	1	32	27
120	<i>Upper</i>	50	48	9	2	0
	<i>Lower</i>	32	56	9	2	1
121	<i>Upper</i>	0	6	2	95	3
	<i>Lower</i>	1	12	4	70	5

122	<i>Upper</i>	1	83	2	11	8
	<i>Lower</i>	1	51	6	12	18
123	<i>Upper</i>	17	25	36	14	13
	<i>Lower</i>	9	24	23	24	13
124	<i>Upper</i>	3	5	41	5	48
	<i>Lower</i>	10	13	40	4	25
125	<i>Upper</i>	1	15	3	16	61
	<i>Lower</i>	1	14	8	18	46
126	<i>Upper</i>	5	15	11	50	11
	<i>Lower</i>	4	25	16	31	7
127	<i>Upper</i>	1	40	3	32	30
	<i>Lower</i>	3	25	6	42	17
128	<i>Upper</i>	53	23	3	2	23
	<i>Lower</i>	44	13	5	8	21
129	<i>Upper</i>	1	7	2	81	11
	<i>Lower</i>	3	7	8	58	10
130	<i>Upper</i>	66	0	42	1	0
	<i>Lower</i>	64	2	29	0	1
131	<i>Upper</i>	0	94	0	0	11
	<i>Lower</i>	1	55	2	1	18
132	<i>Upper</i>	15	1	0	1	92
	<i>Lower</i>	4	3	0	6	86
133	<i>Upper</i>	0	72	0	10	13
	<i>Lower</i>	0	32	5	13	17
134	<i>Upper</i>	1	8	13	57	3
	<i>Lower</i>	1	12	14	23	7
135	<i>Upper</i>	12	12	17	7	35
	<i>Lower</i>	14	11	15	8	10
136	<i>Upper</i>	92	2	6	1	3
	<i>Lower</i>	77	5	9	0	2

Hasil analisis distraktor menunjukkan bahwa ada dua pola yang terbentuk dari pilihan jawaban pada item subtes FA. Pola pertama terkait adalah perbedaan frekuensi yang minim antara kelompok *upper* dan *lower* dalam memilih pilihan yang menjadi pengecoh. Hal ini terjadi misalnya pada pilihan jawaban B pada item 123 dan C pada item 124. Prinsip distraktor yang baik seharusnya mampu mengecoh jawaban, terutama kelompok *lower* untuk menjawab pilihan yang keliru. Namun, pada kasus ini terlihat bahwa kelompok *upper* justru menjadi kelompok yang terkecoh lebih banyak dibandingkan dengan kelompok *lower*, meskipun perbedaannya cenderung minim (1 angka). Pola kedua yang terjadi adalah perbedaan yang ekstrim dimana kelompok *upper* cenderung lebih banyak memilih pilihan jawaban yang keliru dibandingkan dengan kelompok *lower*. Hal

ini terjadi pada distraktor A pada item 128; distraktor E pada item 132; distraktor E pada item 135 dan distraktor A pada item 136. Kesimpulan terkait dengan pola kedua adalah perlunya evaluasi dan revisi terhadap distraktor-distraktor yang ternyata banyak dipilih oleh kelompok *upper* (pola kedua).

### Gambaran Persebaran Skor Subtes FA, WU dan AN

Setelah melalui proses analisis item, keputusan yang dihasilkan adalah tidak melakukan eliminasi terhadap item-item pada subtes FA. Namun, sebelum melangkah kepada tahap pengujian validitas dengan mayoritas teknik korelasi statistik, deskripsi mengenai persebaran skor perlu diidentifikasi. Identifikasi dilakukan sebagai usaha melihat persebaran skor apakah membentuk

distribusi normal atau tidak. Konsekuensi lanjutannya terkait dengan teknik korelasi statistik yang akan digunakan. Identifikasi bentuk distribusi skor dilakukan terhadap 3 distribusi skor, yaitu skor yang berasal dari subtes FA; WU dan AN. Guna mengidentifikasi normalitas dari ketiga distribusi frekuensi, digunakan uji *shapiro-wilk*. Uji *shapiro-wilk* adalah uji statistik yang digunakan untuk mengidentifikasi apakah suatu distribusi frekuensi terdistribusi dengan normal atau tidak (Graveter & Wallnau, 2004; Howel, 2010). Pengambilan keputusan uji *shapiro-wilk* dilakukan dengan membandingkan nilai *p-values* (p) dengan nilai *confidence interval* yang ditetapkan. Ketika nilai  $p > .05$  atau  $p > .01$  (tidak signifikan), maka suatu distribusi frekuensi disimpulkan bahwa suatu distribusi frekuensi tidak terdistribusi normal. Tabel 4 menunjukkan nilai *shapiro-wilk* dari tiga distribusi frekuensi, yaitu subtes FA, WU dan AN.

**Tabel 6: Hasil Uji Shapiro-Wilk pada Distribusi Frekuensi Subtes FA, WU dan AN**

Distribusi frekuensi	Uji normalitas (W)
Subtes FA	.982**
Subtes WU	.987**
Subtes AN	.981**

\*Signifikan pada  $p < .05$

\*\*Signifikan pada  $p < .01$

Hasil pengujian normalitas dengan uji *shapiro-wilk* ternyata menghasilkan nilai yang signifikan pada ketiga distribusi frekuensi. Artinya, persebaran skor dari ketiga subtes tidak menghasilkan distribusi normal. Konsekuensi lanjutan dari hasil ini adalah pengujian korelasi pada *convergent-discriminant* serta uji reliabilitas *split half* akan menggunakan teknik korelasi *spearman-rank order* yang dapat digunakan untuk persebaran skor yang tidak membentuk distribusi normal atau disebut juga perhitungan statistik non-parametrik (Corder & Foreman, 2009).

### **Pengujian *Convergent* dan *Discriminant* terhadap Subtes FA**

Uji *convergent* dan *discriminant* dilakukan penulis guna mengidentifikasi kemampuan subtes FA dalam mengukur kemampuan spasial dua dimensi. Uji *convergent* merujuk kepada membandingkan dua alat ukur yang mengukur konstruk serupa, sedangkan uji *discriminant* merujuk pada metode yang membandingkan dua alat ukur yang mengukur konstruk yang berbeda (Cohen & Swerdlik, 2009). Secara teknis, metode statistik yang digunakan untuk mengidentifikasi pengujian ini adalah teknik korelasi *spearman-rank order*. Penggunaan teknik korelasi *spearman* didasarkan atas pengujian normalitas *shapiro-wilk* yang memutuskan bahwa distribusi skor tidak membentuk distribusi normal.

Dalam pengujian *convergent validation*, subtes FA akan dikorelasikan dengan subtes WU yang mengukur kemampuan spasial tiga dimensi. Kedua subtes ini dalam kerangka teori PMA sama-sama mengukur kemampuan dasar *space* atau keruangan. Sedangkan terkait dengan uji *discriminant validation*, subtes FA akan dikorelasikan dengan subtes AN yang mengukur kemampuan analogi verbal. Dalam kerangka teori PMA, subtes AN mengukur kemampuan dasar *general reasoning*. Tabel 5 menunjukkan hasil korelasi antara subtes FA dengan subtes WU dan subtes AN. Hasil ini mengindikasikan bahwa korelasi yang terjadi antara subtes FA dengan subtes WU lebih kuat ( $r = .439$ ,  $p < .01$ ) dibandingkan dengan nilai korelasi yang terjadi antara subtes FA dengan subtes AN ( $r = .294$ ,  $p < .01$ ). Nilai *convergent* diidentifikasi lebih kuat karena subtes FA dan WU mengukur kemampuan yang serupa, yaitu visualisasi objek atau keruangan. Hasil signifikan pada korelasi *discriminant* juga memberikan *insight* baru kepada penulis bahwa kemampuan keruangan juga tidak dapat dilepaskan dari kemampuan analogi verbal. Artinya, ketika partisipan

mengerjakan subtes AN, partisipan juga menggunakan kemampuan membayangkan objek yang direpresentasikan oleh kata.

**Tabel 7: Hasil Pengujian Convergent dan Discriminant terhadap Subtes FA**

	Subtes FA
Subtes WU	.439**
Subtes AN	.294**

\*\*Signifikan pada  $p < .01$

### Pengujian Reliabilitas terhadap Subtes FA

Pengujian reliabilitas dilakukan dengan metode psikometri *split half* yang membagi subtes FA menjadi dua belahan (Cohen & Swerdlik, 2009). Pembagian kedua belahan dilakukan atas dasar pertimbangan nilai rata-rata derajat kesulitan item pada belahan pertama ( $M=0,454$ ) dan belahan kedua ( $M=0,51$ ). Kedua belahan sama-sama berjumlah 10 item. Guna mengidentifikasi nilai koefisien korelasi antar belahan, penulis menggunakan metode korelasi *spearman-brown formula* yang digunakan untuk mengidentifikasi reliabilitas keseluruhan subtes yang dibelah menjadi dua belahan (Crocker & Algina, 1986; Cohen & Swerdlik, 2009). Nilai *spearman-brown* yang diperoleh adalah  $r=.678$  yang merupakan nilai koefisien reliabilitas dari subtes FA. Nilai ini mencapai batas nilai minimum koefisien reliabilitas suatu alat ukur yang ditetapkan oleh Tavakol dan Dennick (2011).

Hasil koefisien korelasi ini dapat menunjukkan tiga hal utama dari subtes FA. Pertama, hasil ini mengindikasikan bahwa 68% varians *observed score* pada subtes FA disebabkan oleh varians *true score*. *True score* merujuk kepada nilai murni dari partisipan tanpa dipengaruhi oleh *error*, sedangkan *observed score* adalah nilai *true score* yang dipengaruhi oleh nilai *error* (Crocker & Algina, 1986). Artinya, indikasi *error* yang terjadi dan

mempengaruhi nilai individu di subtes FA berkisar sebesar 32%. Kedua, hasil ini juga mengindikasikan bahwa 46% dari varians *observed score* pada belahan kedua dapat diprediksi oleh belahan pertama. Ketiga, nilai koefisien reliabilitas juga mengindikasikan korelasi antara *true score* dan *observed score* yang terjadi sebesar  $r=.82$ . Kesimpulannya, nilai koefisien reliabilitas subtes FA mampu menggambarkan konsistensi pengukuran terhadap partisipan.

### SIMPULAN DAN SARAN

Hasil analisis item dengan dua metode, baik derajat kesulitan dan *D-index* menghasilkan kesimpulan bahwa tidak ada item yang dieliminasi atau direvisi. Hasil kedua analisis mengindikasikan adanya pola yang baik, yaitu semakin tinggi nilai *p* (item mudah), maka semakin kecil pula nilai *D* yang diperoleh. Sebaliknya, semakin rendah nilai *p* (item sulit), maka semakin besar nilai *D* yang diperoleh. Artinya, item sulit yang ditampilkan pada subtes FA hanya dapat dikerjakan oleh partisipan dengan kemampuan spasial yang tinggi. Namun, terkait dengan analisis distraktor, penulis menemukan bahwa ada beberapa pilihan jawaban yang seharusnya mengecoh kelompok *lower* tidak menjalankan fungsinya dengan baik. Hal ini ditunjukkan dengan proporsi kelompok *upper* yang justru lebih besar menjawab pilihan jawaban yang keliru dibandingkan dengan kelompok *lower*.

Terkait dengan uji properti psikometrik validitas dan reliabilitas, subtes FA berada pada posisi yang cukup baik. Hal ini ditandai dengan dua indikator. Indikator pertama adalah subtes FA memiliki korelasi yang positif dan signifikan lebih tinggi dengan subtes WU dibandingkan dengan subtes AN. Indikator kedua adalah melalui analisis reliabilitas yang dihasilkan. Hasil analisis menunjukkan bahwa *error* yang terjadi pada subtes FA cenderung kecil (32%) yang artinya item-item dengan pilihan

jawaban pada subtes FA mampu memberikan stimulasi kepada partisipan (*test taker*) untuk menampilkan kemampuan spasial dua dimensi. Selain itu, analisis reliabilitas juga memprediksi bahwa kontribusi *true score* terhadap *observed score* sebesar 67% yang diperoleh dari nilai  $R^2$  dari nilai korelasi antara *observed score* dan *true score*. Nilai ini mengindikasikan subtes FA mampu memberikan prediksi *true score* di atas 50%. Kesimpulan utama dari studi ini adalah, tidak ada item yang dieliminasi, namun memerlukan sedikit revisi pada pilihan jawaban yang disediakan oleh subtes FA. Pada akhirnya, subtes FA tetap masih bisa digunakan sebagai kepentingan mengukur kemampuan spasial dua dimensi.

#### DAFTAR PUSTAKA

- Anastasi, A., & Urbina, S. (1997). *Psychological testing 7<sup>th</sup>.ed.* New Jersey: Prentice Hall.
- Chadha, N. K. (2009). *Applied psychometry.* Singapore: Sage Publication.
- Cohen, R. J., & Swerdlik, M. E. (2010). *Psychological testing and assessment: An introduction to tests and measurement (7th.ed).* McGraw-Hill.
- Corder, G. W., & Foreman, D. I. (2009). *Nonparametric statistics for non-statisticians.* New Jersey: Wiley Inc.
- Cozby, P. C., & Bates, S. C. (2010). *Methods in behavioral research (11th.ed).* Singapore: McGraw-Hill.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* New York: Horcourt Brace Jovanovich.
- Gravetter, F. J., & Wallnau, L. B. (2004). *Statistics for the behavioral sciences (6th.ed).* Singapore: Thomson Learning, Inc.
- Guilford, J. P. (1972). Thurstone's primary mental abilities and structure-of-intellect abilities. *Psychological Bulletin*, 77(2), 129-143.
- Howell, D. C. (2010). *Statistical methods for psychology (7th.ed).* Singapore: Wadsworth, Cengage Learning.
- Osterlind, S. J. (2002). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats.* Kluwer Academic Publishers.
- Polhaupessy, L. F. (2002). *Intelligenz struktur test (Terjemahan).* Laporan tidak diterbitkan.
- Raven, J. (2000). The raven's progressive matrices: Change and stability over culture and time. *Cognitive Psychology*, 41, 1-48.
- Santosa, E. M., Bonang, E. T., & Panggabean, H. (1998). *Pengujian validitas prediktif Intelligenz Struktur Test (IST).* Laporan tidak diterbitkan.
- Spearman, C. (1904). "General intelligence" objectively determined and measured. *American Journal of Psychology*, 15, 201-293.
- Spearman, C. (1927). *The abilities of man.* New York: Macmillan.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55.
- Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational*, 16, 433-451.
- Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, No. 1.
- Thurstone, L. L. (1944). A factorial study of perception. *Psychometric Monographs*, No. 4.
- Thurstone, L. L. (1947a). The calibration of test items. *American Psychologist*, 2, 103-104.
- Thurstone, L. L. (1947b). *Multiple factor analysis.* Chicago: University of Chicago Press.

- Thurstone, L. L. (1950). *Some primary abilities in visual thinking* (No.59). Chicago: University of Chicago, Psychometric Laboratory.
- Thurstone, L. L. (1959). *The measurement of values*. Chicago: University of Chicago Press.
- Wanzel, K. R., Hamstra, S. J., Anastakis, D. J., Matsumoto, E. D., & Cusimano, M. D. (2002). Effect of visual-spatial ability on learning of spatially-complex surgical skills. *The Lancet*, 359, 230-231