

BAB II

TINJAUAN PUSTAKA

Secara umum ada dua teori pengukuran yaitu teori tes klasik dan teori tes modern. Teori tes klasik merupakan pendekatan pertama yang dikembangkan dalam pengukuran. Teori pengukuran kedua yang merupakan teori tes modern dikenal dengan teori respon aitem atau teori sifat laten (Azwar, 2009).

Teori tes klasik dengan pendekatan teori skor murni berasal dari Charles Spearman dengan konsep korelasinya. Dari tahun 1904 sampai 1913 Spearman mempublikasikan argumen logis dan matematisnya yang menyatakan bahwa skor tes merupakan hasil pengukuran yang bisa keliru dalam mengukur trait manusia. Artinya, skor yang didapat dari hasil pengukuran mengandung eror dan disebut dengan skor tampak. Kemudian Spearman mengatakan bahwa korelasi antara skor tampak yang mengandung eror lebih rendah dari korelasi antara nilai murni dari suatu pengukuran (Spearman dalam Crocker & Algina, 2005).

Teori respon aitem yang selanjutnya disebut IRT merupakan teori yang dikembangkan untuk menjawab kelemahan-kelemahan dari teori tes klasik. Analisis statistik dalam teori tes klasik tidak memberikan informasi mengenai bagaimana performansi subjek yang memiliki level kemampuan berbeda pada suatu trait terlihat dari aitem yang dijawabnya. IRT merupakan teori pengukuran yang dikembangkan untuk menjelaskan secara lebih lengkap mengenai bagaimana gambaran fungsi dari sebuah aitem. Oleh karena itu teori ini juga disebut dengan teori sifat laten (Crocker & Algina, 2005).

IRT merupakan prosedur yang didasarkan pada model matematik yang dikembangkan pada tahun 1960 dan modelnya dikenal dengan model sifat laten. Sifat laten merefleksikan bahwa model ini mencoba mengestimasi level dari berbagai kemampuan yang tidak teramati secara langsung. Hal ini dilihat dari sifat atau konstruk psikologis yang mendasari perilaku seseorang yang ditunjukkan dari respon mereka terhadap aitem-aitem dalam tes. Berbeda dengan pendekatan teori tes klasik, IRT membutuhkan jumlah data yang besar dan sampel yang beragam dalam menerapkan model matematikanya untuk mengestimasi aitem (Embretson & Reise dalam Anastasi & Urbina, 2007). Hal ini menunjukkan kelemahan dari IRT yang tidak praktis dan model matematiknya rumit dibandingkan dengan teori tes klasik.

Penelitian ini akan menganalisis karakteristik psikometri subtes *analogien* (AN) pada *intelligenz struktur test* (IST) versi 70 hasil adaptasi Universitas Padjajaran pada tahun 1973. Untuk penjelasan mengenai teori IST akan dibahas selanjutnya.

Teori pengukuran yang akan dipakai untuk menganalisa karakteristik Psikometri subtes AN pada IST adalah teori skor murni klasik. Teori skor murni klasik digunakan karena teori ini memiliki nilai praktis yang tinggi dalam menerangkan masalah reliabilitas dan validitas, disamping pemahamannya yang tidak menuntut pengetahuan yang dalam sekali mengenai beberapa fungsi distribusi statistik dan mengenai model-model matematiknya. Untuk penjelasan lebih lanjut mengenai teori tes klasik beserta asumsi-asumsinya akan dipaparkan kemudian.

A. Teori Tes Klasik

1. Definisi

Teori tes klasik atau yang dalam bahasa inggrisnya adalah *Classical Test Theory* (CTT) disebut demikian karena unsur-unsur teori itu sudah dikembangkan dan diaplikasikan sejak lama, namun tetap bertahan. Oleh karena itu teori ini dianggap sebagai hasil karya klasik. Teori tes klasik tidak tersusun sekali jadi, melainkan berkembang sedikit demi sedikit melalui unsur-unsur yang kemudian secara akumulatif merupakan bangunan teori yang utuh. Inti teori tes klasik itu berupa asumsi-asumsi yang dirumuskan secara matematis. Modelnya disebut model skor murni (*true score model*). Ada tujuh macam asumsi dalam teori tes klasik, seperti yang akan disajikan secara ringkas berikut ini.

2. Asumsi-asumsi dalam Teori Tes Klasik

$$\text{Asumsi 1: } X_t = X_* + X_e \dots\dots\dots (1)$$

Skor perolehan atau skor tampak (X_t) terdiri dari skor murni (X_*) dan skor kesalahan (X_e). Jadi skor yang diperoleh dari sesuatu pengukuran pada umumnya tidak menunjukkan keadaan yang sebenarnya. Skor perolehan pada umumnya meleset dari menampilkan secara tepat besaran atribut yang diukur. Melesetnya skor perolehan dari keadaan yang sebenarnya yaitu skor murni merupakan kesalahan pengukuran (*error of measurement*). Skor murni dianggap tetap, sedang skor kesalahan terjadi secara rambang, baik ke atas maupun ke bawah. Sebenarnya yang diperlukan adalah skor murni, yaitu skor yang mencerminkan secara tepat besaran atribut yang diukur, namun pada umumnya tidak ada jalan

untuk mengetahui skor murni itu secara langsung. Skor murni hanya dapat diketahui secara tidak langsung melalui galat (kesalahan) baku pengukuran (*standard error of measurement*).

$$\text{Asumsi 2: } \varepsilon(X_i) = X_* \dots\dots\dots (2)$$

Nilai harapan skor perolehan ($\varepsilon(X_i)$) sama dengan skor murni. Asumsi 2 ini merupakan defenisi skor murni (X_*). Skor murni itu adalah nilai rata-rata skor perolehan teoretis sekiranya dilakukan pengukuran berulang-ulang sampai tak terhingga terhadap seseorang dengan menggunakan alat ukur yang sama. Syarat pokok dalam pengukuran ulang itu adalah hasil pengukuran yang satu harus bebas dari hasil pengukuran yang lain.

$$\text{Asumsi 3: } \rho_{x_* x_e} = 0 \dots\dots\dots (3)$$

Skor murni dan skor kesalahan yang dicapai oleh suatu populasi subjek pada suatu tes tidak berkorelasi satu sama lain. Jadi tidak ada hubungan sistematis antara skor murni dan skor kesalahan. Subjek yang tinggi skor murninya tidak mesti mempunyai skor kesalahan baik positif maupun negatif yang lebih tinggi dibanding subjek yang rendah skor murninya.

$$\text{Asumsi 4: } \rho_{x_{e1} x_{e2}} = 0 \dots\dots\dots (4)$$

Skor-skor kesalahan pada dua tes yang dimaksud untuk mengukur dua hal yang sama, tidak saling berkorelasi. Asumsi ini akan tidak terpenuhi sekiranya skor perolehan dipengaruhi oleh kondisi testing, seperti misalnya kelelahan, efek latihan, suasana hati, atau faktor-faktor dari lingkungan.

Asumsi 5: $\rho_{X_{e1}X_{*2}} = 0$ (5)

Jika ada dua tes yang dimaksudkan untuk mengukur atribut yang sama, maka skor-skor kesalahan pada tes 1 (X_{e1}) tidak berkorelasi dengan skor-skor murni pada tes 2 (X_{*2}).

Asumsi 6:

Jika dua perangkat tes yang dimaksudkan mengukur atribut yang sama mempunyai skor perolehan X_t dan X_t' yang memenuhi asumsi-asumsi 1 sampai 5, dan jika untuk setiap populasi subjek $X_{*} = X_{*}'$ dan $\sigma_e^2 = \sigma_e'^2$, maka kedua tes itu disebut tes paralel. Jadi dua perangkat tes akan merupakan tes paralel jika skor-skor suatu populasi yang menempuh kedua tes itu skor murninya sama ($X_{*} = X_{*}'$) dan varians skor-skor kesalahannya sama ($\sigma_e^2 = \sigma_e'^2$).

Asumsi 7:

Jika dua perangkat tes yang dimaksudkan untuk mengukur atribut yang sama mempunyai skor-skor perolehan X_{t1} dan X_{t2} yang memenuhi asumsi-asumsi 1 sampai 5, dan apabila untuk setiap populasi subjek $X_{*1} = X_{*2} + C_{12}$, dimana C_{12} adalah suatu konstanta, maka kedua perangkat tes itu disebut tes-tes yang setara (*equivalent test*).

Asumsi 1 dari teori tes klasik akan menjadi dasar kerangka kerja untuk memahami hasil estimasi terhadap karakteristik psikometri subtes AN pada IST yang masih digunakan di Pusat Penelitian dan pengabdian pada Masyarakat (P3M) Fakultas Psikologi Universitas Sumatera Utara.

B. Analisis Karakteristik Psikometri

Tes disusun dari sejumlah aitem yang bertujuan untuk mengukur atribut atau konstruk Psikologi tertentu. Oleh karena itu, kualitas tes akan sangat bergantung pada kualitas aitem yang menyusunnya (Murphy & Davidshofer, 2003). Sebagai alat ukur, suatu tes baru dapat dikatakan berhasil menjalankan fungsi ukurnya apabila ia mampu memberikan hasil ukur yang cermat dan akurat (Azwar, 2005). Sebagaimana yang telah dikemukakan, isi tes tidak lain daripada sekumpulan aitem berupa pertanyaan mengenai sesuatu hal yang hendak diukur atau diungkap. Jadi tidaklah mengherankan bila kualitas tes secara keseluruhan sangat bergantung pada karakteristik aitem-aitem yang menyusunnya (Murphy & Davidshofer, 2003).

Penilaian kualitas tes dengan menganalisis karakteristik aitem dilakukan untuk melihat validitas dan reliabilitas tes (Crocker & Algina, 2005). Hal ini didukung juga oleh Murphy & Davidshofer (2003) yang menyatakan bahwa aitem-aitem tes dapat mempengaruhi reliabilitas dan validitas tes. Dalam penelitian ini akan dilihat apakah aitem-aitem dalam subtes AN memiliki karakteristik Psikometri yang baik untuk mengetahui kesesuaian tes tersebut dengan tujuan ukurnya. Subtes AN pada IST merupakan tes tipe *multiple choice items* atau tipe objektif dengan 5 pilihan jawaban, maka analisis terhadap kualitas aitemnya dilihat dari karakteristik indeks kesukaran aitem, indeks daya beda aitem dan efektifitas distraktor.

1. Indeks Kesukaran Aitem (p)

a. Definisi

Kesukaran aitem dapat dilihat dari jumlah subjek yang menjawab dengan benar setiap aitem dalam tes. Jika setiap orang dalam suatu tes memilih jawaban yang benar, maka aitem didefinisikan sebagai aitem yang mudah. Jika hanya satu dari 100 subjek yang menjawab suatu aitem dengan benar, maka aitem didefinisikan sebagai aitem yang sulit (Murphy & Davidshofer, 2003).

Menurut Azwar (2005) indeks kesukaran aitem merupakan rasio antara penjawab aitem dengan benar dan banyaknya penjawab aitem. Taraf kesukaran suatu aitem dinyatakan oleh suatu indeks yang dinamakan indeks kesukaran aitem dan disimbolkan oleh huruf p . Secara teoretik dikatakan bahwa p sebenarnya merupakan probabilitas empirik untuk lulus aitem tertentu bagi kelompok subjek tertentu. Formulasi indeks kesukaran aitem adalah:

$$p_i = n_i / N \dots\dots\dots (6)$$

Keterangan,

p_i = Indeks kesukaran untuk aitem i

n_i = Banyaknya subjek yang menjawab aitem i dengan benar

N = Banyaknya subjek yang menjawab aitem

Tate (dalam Kumar, 2009) menyatakan hal yang sama bahwa indeks kesukaran aitem dapat dihitung dengan menentukan proporsi subjek yang menjawab aitem dengan benar. Untuk aitem yang memiliki skor dikotomi seperti aitem-aitem pada subtes AN pada IST, nilai rata-rata skor aitem akan sama dengan jumlah indeks kesukaran aitemnya (Crocker & Algina, 2005).

b. Analisis Indeks Kesukaran Aitem

Kisaran harga p bergerak mulai dari 0 sampai dengan 1. Secara logis dari persamaan tersebut dapat dijelaskan bahwa semakin besar harga p mengindikasikan bahwa aitem tersebut semakin mudah dijawab oleh satu kelompok subjek, dan semakin kecil harga p mengindikasikan semakin sulit suatu aitem untuk dijawab oleh suatu kelompok subjek (Azwar, 2005).

Indeks kesukaran aitem memiliki pengaruh besar terhadap variabilitas skor tes dimana skor tes membedakan antara kelompok-kelompok subjek yang dites. Nilai p ekstrim secara langsung membatasi variabilitas skor tes (Murphy & Davidshofer, 2003). Variabilitas skor tes yang rendah tidak akan menunjukkan perbedaan kelompok-kelompok subjek yang dikenai tes. Nilai p terbaik adalah 0,5 karena nilai ini memungkinkan variasi skor tes yang beragam. Allen & Yen (dalam Lababa, 2008), mengkategorikan nilai p sebagai berikut:

Tabel 1. Kategorisasi Nilai p

No	p	Kategori
1	$p < 0,3$	Sulit
2	$0,3 < p < 0,7$	Sedang
3	$p > 0,7$	Mudah

Umumnya p yang berada di sekitar 0,50 dianggap yang terbaik meskipun hal ini tidak selalu benar. Kadang-kadang dikehendaki harga p yang lebih kecil daripada 0,50 yaitu aitemnya lebih sulit apabila aitem itu dimaksudkan sebagai bagian dari suatu tes yang digunakan dalam prosedur seleksi guna memilih sebagian kecil saja di antara para pelamar. Tidak jarang pula suatu tes dibidang pendidikan seperti tes prestasi perlu disusun dengan memasukkan banyak aitem yang taraf kesukarannya rendah (p tinggi), misalnya dalam tes yang bertujuan

untuk evaluasi formatif. Hal ini menunjukkan bahwa taraf kesukaran yang terbaik bergantung pada tujuan dari tes tersebut (Azwar, 2005).

Berdasarkan tujuannya, IST dipakai untuk seleksi dan penempatan karyawan. Untuk memenuhi tujuan IST, maka pada penelitian ini nilai p akan dikatakan baik jika $p < 0.3$ atau dikategorikan sulit. Hal ini didukung oleh pernyataan Lord (dalam Murphy & Davidshofer, 2003) bahwa untuk tes seleksi karyawan sebaiknya tes berisi aitem dengan nilai p mendekati 0,2 dan jika tujuan tes adalah untuk menyaring kelompok pelamar yang paling baik tes harus berisi aitem-aitem yang sangat sulit. Pengkategorian nilai p pada penelitian ini didasarkan pada kategorisasi nilai p menurut Allen & Yen (dalam Lababa, 2008).

Selain dinyatakan dalam bentuk proporsi, indeks kesukaran aitem dapat dinyatakan dalam bentuk persentase dengan cara mengalikan saja harga p dengan 100%. Aitem yang mempunyai $p = 0,75$ adalah sama pengertiannya dengan mempunyai indeks kesukaran aitem $0,75\% * 100 = 75\%$ (Azwar, 2005). Artinya hanya 75% saja yang mampu menjawab aitem dengan benar dari jumlah keseluruhan subjek yang dikenai tes.

Suatu pengertian yang harus dipahami lebih lanjut, yaitu harga p bagi suatu aitem menunjukkan taraf kesukaran aitem itu bagi kelompok yang bersangkutan, yaitu kelompok yang menjadi dasar dalam menghitung p itu sendiri. Harga p suatu aitem akan berbeda apabila dihitung pada kelompok subjek yang lain. Dengan kata lain, indeks kesukaran suatu aitem tidak selalu sama antara satu kelompok dengan kelompok subjek yang lain karena memang suatu aitem

yang dirasakan sulit bagi satu kelompok subjek mungkin saja dirasa mudah bagi kelompok subjek yang lain (Azwar, 2005).

Harus diingat juga bahwa besarnya harga p yang dihitung merupakan indeks kesukaran aitem bagi seluruh kelompok, bukan indeks kesukaran bagi masing-masing subjek secara individual. Taraf kesukaran suatu aitem bagi setiap subjek adalah berbeda-beda dan tidak diketahui seberapa sulit atau seberapa mudah suatu aitem bagi seorang subjek. Harga p yang dihitung dalam kelompok hanya merupakan rata-rata indeks kesukaran bagi seluruh subjek dalam kelompok itu. Apabila subjek dapat menjawab benar suatu aitem berarti taraf kesukaran aitem itu lebih rendah daripada taraf kemampuannya dalam menjawab. Sebaliknya, apabila subjek salah dalam menjawab suatu aitem berarti bahwa tingkat kemampuannya lebih rendah daripada taraf kesukaran aitem yang bersangkutan (Azwar, 2005).

2. Indeks Daya Beda Aitem (d)

a. Definisi

Daya beda aitem adalah kemampuan aitem dalam membedakan antara subjek yang mempunyai kemampuan tinggi dalam hal ini diwakili oleh mereka yang termasuk kelompok tinggi dan subjek yang mempunyai kemampuan rendah diwakili oleh mereka yang termasuk dalam kelompok rendah (Azwar, 2005). Hal yang sama dikemukakan Murphy & Davidshofer (2003) yang menyatakan bahwa aitem yang baik dapat membedakan antara mereka yang baik dalam tes

ditunjukkan menjawab aitem dengan benar dengan mereka yang buruk karena menjawab aitem dengan salah.

Suatu aitem yang dikatakan mempunyai daya beda tinggi haruslah dijawab dengan benar oleh semua atau sebagian besar subjek kelompok tinggi dan tidak dapat dijawab dengan benar oleh semua atau sebagian besar subjek kelompok rendah. Semakin besar perbedaan antara proporsi penjawab benar dari kelompok tinggi dan dari kelompok rendah, semakin besarlah daya beda suatu aitem. Jika proporsi penjawab benar dari kedua kelompok tersebut sama, berarti aitem yang bersangkutan tidak mampu membedakan antara mereka yang berkecakapan tinggi dan mereka yang berkecakapan rendah. Apalagi bila suatu aitem ternyata dapat dijawab dengan benar oleh sebagian besar subjek kelompok rendah sedangkan sebagian besar kelompok tinggi tidak banyak yang dapat menjawab dengan benar, maka jelas aitem yang seperti itu akan menyesatkan karena daya bedanya terbalik (Azwar, 2005).

Indeks daya beda aitem disimbolkan dengan d . Langkah pertama menghitung indeks daya beda adalah dengan memilih kelompok tinggi dan kelompok rendah. Langkah selanjutnya adalah menghitung persentase subjek yang lulus di setiap aitem dalam kelompok tinggi dan rendah. Indeks daya beda aitem secara sederhana adalah selisih dari dua persentase ini (Murphy & Davidshofer, 2003). Formula daya beda aitemnya adalah:

$$d = n_{iT} / N_T - n_{iR} / N_R \dots\dots\dots (7)$$

Keterangan:

d = Indeks daya beda aitem

n_{iT} = Banyaknya penjawab aitem dengan benar dari kelompok tinggi

N_T = Banyaknya penjawab dari kelompok tinggi

n_{iR} = Banyaknya penjawab aitem dengan benar dari kelompok rendah

N_R = Banyaknya penjawab dari kelompok rendah

Banyak kasus, kelompok tinggi dan kelompok rendah dibagi dengan proporsi yang sama, maka untuk kondisi ini dimana $N_T = N_R$, formula (2) diturunkan menjadi:

$$d = \frac{n_{iT} - n_{iR}}{N} \dots\dots\dots (8)$$

dimana, $N = N_T = N_R$

Keterangan:

d = Indeks daya beda aitem

n_{iT} = Banyaknya penjawab aitem dengan benar dari kelompok tinggi

n_{iR} = Banyaknya penjawab aitem dengan benar dari kelompok rendah

N = Banyaknya penjawab

Hal ini sejalan dengan pendapat Kelly (dalam Kumar, 2009), yang menyatakan bahwa indeks daya beda dilihat dari rasio perbedaan antara mean dari dua kelompok yaitu kelompok tinggi dan kelompok rendah yang ditentukan dari respon mereka. Untuk subjek dalam jumlah besar kelompok tinggi diambil sebesar 27% dari jumlah yang mendapat nilai tinggi dan 27% dari kelompok yang mendapat nilai rendah. Mengingat bahwa $\frac{n_i}{N}$ merupakan rumusan indeks kesukaran aitem, maka dapat dijelaskan pula bahwa indeks daya beda merupakan perbedaan

indeks kesukaran aitem pada kelompok tinggi dan kelompok rendah, sehingga dapat ditulis sebagai:

$$d = p_t - p_r \dots\dots\dots (9)$$

Keterangan:

d = Indeks daya beda aitem

p_t = Indeks kesukaran aitem kelompok tinggi

p_r = Indeks kesukaran aitem kelompok rendah

b. Analisis Indeks Daya Beda Aitem

Daya beda maksimal tercapai apabila seluruh subjek kelompok tinggi dapat menjawab dengan benar ($n_{iT} = N_T$) sedangkan seluruh subjek kelompok rendah tidak seorang pun yang dapat menjawab dengan benar ($n_{iR} = 0$). Dalam hal ini harga $d = 1 - 0 = 1$. Indeks daya beda sebesar 0 akan terjadi apabila proporsi penjawab benar dari kelompok tinggi dan dari kelompok rendah sama besarnya, yaitu ketika indeks kesukaran bagi kelompok tinggi sama besar dengan indeks kesukaran bagi kelompok rendah. Secara matematik, indeks daya beda besarnya akan berkisar mulai dari -1 sampai dengan +1, namun hanya harga d yang positif sajalah yang memiliki arti dalam analisis aitem (Azwar, 2005).

Harga d yang berada di sekitar 0 menunjukkan bahwa aitem yang bersangkutan mempunyai daya beda yang rendah sedangkan harga d yang negatif menunjukkan bahwa aitem yang bersangkutan tidak ada gunanya sama sekali, bahkan memberikan informasi yang menyesatkan. Indeks daya beda aitem yang ideal adalah yang sebesar mungkin mendekati 1. Semakin besar indeks daya beda aitem (semakin mendekati 1) berarti aitem tersebut semakin mampu membedakan

antara mereka yang menguasai bahan yang diujikan dan mereka yang tidak. Semakin kecil indeks daya beda aitem (semakin mendekati 0) berarti semakin tidak jelaslah fungsi aitem bersangkutan dalam membedakan mana subjek yang menguasai kemampuan tertentu dan mana subjek yang tidak tahu apa-apa (Azwar, 2005).

Ebel (dalam Crocker & Algina, 2005), menyarankan kriteria evaluasi indeks daya beda aitem dalam empat kategori yang dapat dilihat pada Tabel 2.

Tabel 2. Evaluasi Indeks Daya Beda Aitem

Indeks Daya Beda	Evaluasi
$\geq 0,40$	Bagus sekali
0,30 – 0,39	Lumayan bagus, tidak membutuhkan revisi
0,20 – 0,29	Belum memuaskan, perlu direvisi
$\leq 0,19$	Buruk dan harus dibuang

Kriteria evaluasi Ebel didukung oleh Thorndike (dalam Azwar, 2005) yang menyatakan bahwa setiap aitem yang memiliki indeks d lebih besar daripada 0,50 dapat langsung dianggap sebagai aitem yang berdaya beda baik, aitem yang memiliki indeks d kurang dari 0,20 dapat langsung dibuang, sedangkan aitem lainnya dapat ditelaah lebih lanjut untuk direvisi.

Menurut Murphy dan Davidshofer (2003) ada tiga cara statistik yang dapat digunakan untuk mengestimasi daya beda aitem, yaitu:

1) Metode Kelompok Ekstrim

Metode kelompok ekstrim dapat digunakan untuk menghitung secara sederhana kekuatan daya beda dari suatu aitem tes jika tes diberikan kepada kelompok subjek berukuran besar. Kekuatan daya beda aitem dapat diukur dengan membandingkan jumlah subjek yang mendapat skor tes tinggi (misalnya, mereka yang berada diatas 25% dari kelas) yang menjawab aitem dengan benar dengan

jumlah subjek dengan skor rendah (misalnya, dibawah 25%) yang menjawab aitem yang sama dengan benar. Jika aitem menjalankan fungsinya dengan baik dalam membedakan antara kelompok skor tinggi dan skor rendah, kebanyakan subjek yang berada pada kelompok skor tinggi menjawab aitem dengan benar.

Cureton (dalam Murphy dan Davidshofer, 2003), menambahkan bahwa kelompok ekstrim terdiri dari kelompok yang skornya berada diatas dan dibawah 27% atau 33% dari subjek dalam kelompok yang dikenai tes. Demi tujuan praktis, metode kelompok ekstrim ini dapat terdiri dari 25% sampai 35% dari jumlah seluruh subjek. Hal ini juga sesuai dengan pernyataan Azwar (2005), jika subjek yang mengikuti suatu tes banyak sekali, misalnya sampai melebihi 500 orang maka pembagian kelompok tinggi dan kelompok rendah dapat dilakukan dengan mengambil hanya sebagian saja dari mereka yang skor totalnya tertinggi dan sebagian dari mereka yang skor totalnya terendah. Dengan cara 27% kelompok tinggi dan 27% kelompok rendah.

2) Korelasi Aitem-Total

Korelasi aitem total adalah cara berikutnya yang digunakan untuk mengetahui indeks daya beda aitem. Indeks daya beda memberikan informasi mengenai derajat dimana suatu aitem dan tes mengukur hal yang sama. Korelasi aitem total mewakili hubungan sederhana antara skor pada sebuah aitem (respon benar diberi skor 1, respon salah diberi skor 0) dengan total skor tes.

Korelasi aitem total dapat dihitung menggunakan korelasi point biserial. Korelasi point biserial digunakan jika variabel kontinu dihubungkan dengan variabel dikotomi. Variabel kontinu dalam hal ini adalah nilai yang berada

diantara variabel dikotomi yaitu nilai yang berada diantara angka 0 dan 1. Korelasi point biserial yang bernilai positif menunjukkan bahwa aitem dan tes mengukur hal yang sama, nilai mendekati nol menunjukkan bahwa aitem tidak memiliki indeks daya beda yang baik sehingga kelompok tinggi menjawab pertanyaan dengan salah dan kelompok rendah menjawab pertanyaan dengan benar. Nilai negatif menunjukkan bahwa aitem tidak mengukur hal yang sama dengan tesnya.

3) Korelasi Inter-Aitem

Korelasi inter aitem memberikan beberapa informasi. Pertama, memungkinkan untuk menghitung reliabilitas tes dengan memberikan rata-rata korelasi inter aitem dan jumlah aitem dalam tes. Hal yang paling penting dari penggunaan korelasi inter aitem adalah untuk memahami pengukuran indeks daya beda aitem. Secara jelas aitem yang menunjukkan korelasi aitem total yang positif juga akan menunjukkan korelasi positif dengan kebanyakan aitem lainnya dalam tes. Pengujian terhadap korelasi inter aitem dapat membantu dalam memahami mengapa beberapa aitem gagal untuk membedakan antara mereka yang mampu mengerjakan tes dan yang tidak.

Jika korelasi total aitem rendah, ada dua kemungkinan yang bisa dijelaskan. Pertama, ada kemungkinan bahwa suatu aitem dalam tes tidak berkorelasi dengan aitem lainnya dalam tes yang sama. Kedua, kemungkinan bahwa aitem menunjukkan korelasi yang positif dengan beberapa aitem dalam tes tetapi korelasinya mendekati nol, atau bahkan mungkin berkorelasi negatif dengan aitem lainnya dalam tes.

Penelitian ini akan menggunakan pendekatan korelasi aitem total dengan formula point biserial untuk mengestimasi besarnya indeks daya beda aitem yang dimiliki subtes AN pada IST. Formula koefisien korelasi poin biserial digunakan untuk aitem-aitem berkategori dikotomi sebagaimana subtes AN pada IST, dimana jawaban benar diberi angka 1 dan jawaban salah diberi angka 0. Berikut adalah formula point biserial:

$$r_{pb} = [(M_i - M_x) / S_x] / \sqrt{[p / (1-p)]} \dots\dots\dots(10)$$

Keterangan:

- r_{pb} = Korelasi poin biserial
- M_i = Mean skor X dari seluruh subjek yang mendapat angka 1 pada variabel dikotomi i
- M_x = Mean skor dari seluruh subjek
- S_x = Standard deviasi skor X
- p = Proporsi subjek yang mendapat angka 1 pada variabel dikotomi

Nilai d yang didapat dari hasil analisis korelasi aitem total terhadap aitem-aitem subtes AN pada IST selanjutnya dikategorikan berdasarkan panduan evaluasi indeks daya beda aitem oleh Ebel yang dapat dilihat pada Tabel 2.

3. Efektifitas Distraktor

a. Definisi

Terdapat hanya satu jawaban benar untuk masing-masing aitem dalam suatu tes tipe objektif. Efektifitas distraktor dapat diketahui dengan menghitung frekuensi dari setiap pilihan jawaban yang dipilih oleh kelompok yang dikenai tes

(Murphy & Davidshofer, 2003). Subtes AN merupakan tipe objektif, maka dalam penelitian ini akan dilihat bagaimana efektifitas distraktor yang dimiliki setiap aitem subtes AN pada IST.

Efektifitas distraktor dilihat dari dua kriteria, yaitu 1) distraktor dipilih oleh subjek dari kelompok rendah, dan 2) permilih distraktor tersebar relatif proporsional pada masing-masing distraktor yang ada (Azwar, 2005). Hal ini sejalan dengan ungkapan Murphy & Davidshofer (2003), yang menyatakan bahwa aitem tes yang baik seharusnya memiliki dua karakteristik. Pertama, orang-orang yang mengetahui jawaban terhadap suatu pertanyaan akan selalu menjawab pilihan yang benar. Kedua, orang-orang yang tidak tahu jawaban dari suatu pertanyaan dalam tes akan memilih secara random antara semua pilihan jawaban yang ada. Ini artinya bahwa akan ada beberapa orang yang tidak mengetahui jawabannya menjawab dengan benar. Dan ini juga berarti bahwa setiap kemungkinan pilihan jawaban yang salah setara tingkat populernya. Hal ini akan terlihat dari frekuensi subjek yang menjawab pilihan jawaban yang salah akan menyebar sama banyak.

Murphy & Davidshofer (2003), menuliskan formula sederhana dalam bukunya untuk menghitung jumlah subjek yang diharapkan memilih setiap distraktor dalam suatu aitem (ϵ_{di}), yaitu membagi jumlah orang yang menjawab aitem i dengan salah (n_{si}) dengan jumlah distraktor (\sum_d).

$$\epsilon_{di} = n_{si} / \sum_d \dots\dots\dots(11)$$

Keterangan:

ϵ_{di} = Jumlah subjek yang diharapkan memilih setiap distraktor suatu aitem

n_{si} = Jumlah orang yang menjawab aitem i salah

\sum_d = Jumlah distraktor

Efektifitas distraktor-distraktor yang ada pada suatu aitem dianalisis dari distribusi jawaban terhadap aitem yang bersangkutan pada setiap pilihan yang disediakan. Efektifitas distraktor diperiksa untuk melihat apakah semua distraktor atau semua pilihan jawaban yang bukan kunci telah berfungsi sebagaimana mestinya. Apakah distraktor-distraktor tersebut telah dipilih oleh lebih banyak atau semua subjek kelompok rendah sedangkan subjek dari kelompok tinggi hanya sedikit atau tidak ada yang memilihnya.

b. Analisis Efektifitas Distraktor

Indeks kesukaran (p) dan indeks daya beda (d) saja tidaklah cukup memberikan informasi mengenai kualitas aitem. Bagaimana jawaban-jawaban subjek tersebar pada pilihan jawaban yang tersedia juga harus diperhatikan agar fungsi suatu aitem dapat dipenuhi secara maksimal. Tentu saja kombinasi penyebaran jawaban akan sangat banyak variasinya, namun dasar terpenting dalam melakukan penilaian terhadap kualitas aitem adalah dengan melihat fungsi pilihan jawaban, terutama distraktor-distraktor yang harus tampak sebagai jawaban benar bagi subjek dari kelompok rendah (Azwar, 2005).

Distraktor yang baik adalah yang tampak jelas salahnya bagi subjek kelompok tinggi akan tetapi sukar dibedakan dari kunci jawaban lain oleh subjek kelompok rendah. Tentu saja akan sangat ideal apabila aitem dapat dijawab dengan benar oleh semua subjek kelompok tinggi sedangkan subjek kelompok

rendah semua memilih distraktor secara merata. Akan tetapi pola penyebaran jawaban seideal itu dapat dikatakan tidak akan pernah ada. Model-model penyebaran jawaban yang mendekati ideal seperti itu pun sudah sangat memuaskan (Azwar, 2005).

Memeriksa pola penyebaran jawaban untuk menilai efektifitas distraktor, seringkali lebih baik menggunakan proporsi atau persentase jawaban bagi setiap pilihan sebagai ganti frekuensi jawaban dikarenakan proporsi dapat dibandingkan langsung antara satu aitem dengan aitem yang lain tanpa tergantung pada besarnya N (Azwar, 2005).

Proporsi bagi setiap pilihan jawaban diperoleh dari membagi frekuensi jawaban pada setiap pilihan dengan jumlah penjawab (N) dari kelompok masing-masing subjek. Sehingga proporsi, dalam hal ini, merupakan proporsi dalam kelompok masing-masing. Penyajian dalam bentuk proporsi akan memberikan gambaran yang lebih cermat terutama apabila jumlah penjawab dari kelompok tinggi tidak sama dengan jumlah penjawab dari kelompok rendah (Azwar, 2005).

Ketika jumlah orang yang menjawab suatu distraktor melebihi jumlah yang diharapkan, terdapat dua kemungkinan. Pertama, ada kemungkinan bahwa pilihan subjek tersebut menunjukkan pengetahuan parsial. Artinya subjek mengetahui bahwa distraktor yang dipilihnya tersebut juga berkaitan dengan pengetahuan yang dipertanyakan. Kedua, kemungkinan yang ditakutkan adalah aitem tersebut merupakan aitem buruk yang menjebak. Artinya, jika salah satu distraktor lebih dikenal oleh subjek yang memiliki pengetahuan baik mengenai domain ukur dan jika identifikasi dari respon benar merupakan jawaban yang

kurang dikenal atau tidak jelas, maka aitem ini tidak valid mengukur kawasan ukurnya. Untuk kebanyakan tes, kehadiran aitem dengan distraktor yang sangat tidak asing bagi subjek memiliki reliabilitas dan validitas tes yang rendah (Murphy & Davidshofer, 2003).

Beberapa distraktor ada yang sangat asing bagi subjek. Jumlah subjek yang memilih respon ini lebih kecil dari jumlah yang diharapkan. Kehadiran distraktor yang sangat asing akan mempengaruhi jumlah karakteristik aitem yang penting. Pengaruh yang paling nyata adalah akan menurunkan indeks kesukaran aitem. Jika seseorang tidak mengetahui apapun tentang pengetahuan yang sedang dites, kemungkinannya untuk menjawab respon yang benar pada suatu aitem dengan empat pilihan jawaban adalah satu banding empat. Tetapi jika satu dari keempat pilihan jawaban tersebut jelas tidak benar, maka subjek akan langsung mengeliminasi distraktor ini. Artinya, kemungkinan subjek untuk menjawab respon benar akan meningkat dari satu banding empat menjadi satu banding tiga (Murphy & Davidshofer, 2003).

Efektifitas distraktor pada penelitian ini dilihat dari pola penyebaran respon jawaban pada masing-masing pilihan jawaban di setiap aitem dalam subtes AN pada IST untuk mengetahui apakah semua distraktor berfungsi sebagaimana mestinya.

4. Reliabilitas Alat Ukur

a. Definisi

Reliabilitas diterjemahkan dari kata *reliability* yang mempunyai asal kata *rely* dan *ability*. Pengukuran yang memiliki reliabilitas tinggi maksudnya adalah pengukuran yang dapat menghasilkan data yang konsisten. Walaupun reliabilitas mempunyai berbagai nama lain seperti keterpercayaan, keterandalan, keajegan, konsistensi, kestabilan, dan sebagainya namun ide pokok dalam konsep reliabilitas adalah sejauh mana hasil suatu pengukuran dapat dipercaya (Azwar, 2005).

Reliabilitas dilihat dari skor yang tidak jauh berbeda sekiranya dilakukan dua kali pengukuran terhadap suatu atribut Psikologi (Murphy & Davidshofer, 2003). Hal yang sama juga diungkapkan Azwar (2005), bahwa hasil ukur dapat dipercaya apabila dalam beberapa kali pengukuran terhadap kelompok subjek yang sama diperoleh hasil yang relatif sama, jika aspek yang diukur dalam diri subjek memang belum berubah. Pengertian relatif menunjukkan bahwa ada toleransi terhadap perbedaan-perbedaan kecil diantara hasil pengukuran. Bila perbedaan itu besar dari waktu ke waktu, maka hasil pengukuran itu tidak dapat dipercaya atau tidak reliabel.

Pengertian reliabilitas alat ukur dan reliabilitas hasil ukur bisa saja dianggap sama. Namun penggunaannya masing-masing perlu diperhatikan. Konsep reliabilitas dalam arti reliabilitas alat ukur erat berkaitan dengan masalah eror pengukuran. Eror pengukuran sendiri menunjuk pada sejauhmana inkonsistensi hasil pengukuran terjadi apabila pengukuran dilakukan ulang pada

kelompok subjek yang sama. Reliabilitas dalam arti reliabilitas hasil ukur erat berkaitan dengan eror dalam pengambilan sampel yang mengacu kepada inkonsistensi hasil ukur apabila pengukuran dilakukan ulang pada kelompok individu yang berbeda (Azwar, 2005).

b. Koefisien Reliabilitas

Tinggi rendahnya reliabilitas, secara empirik ditunjukkan oleh suatu angka yang disebut koefisien reliabilitas. Pada awalnya, tinggi rendahnya reliabilitas dicerminkan oleh tinggi rendahnya korelasi antara dua distribusi skor dari dua alat ukur yang paralel yang dikenakan pada sekelompok individu yang sama. Semakin tinggi koefisien korelasi antara hasil ukur dari dua tes yang paralel, berarti konsistensi diantara keduanya semakin baik dan kedua alat ukur itu disebut sebagai alat ukur yang reliabel. Sebaliknya, apabila korelasi antara hasil dari dua alat ukur yang paralel ternyata tidak tinggi, maka disimpulkan bahwa reliabilitasnya rendah (Azwar, 2005).

Koefisien korelasi antara dua variabel dilambangkan oleh huruf r . Apabila skor pada alat ukur yang pertama diberi lambang X dan skor pada alat ukur paralelnya diberi lambang X' , maka koefisien korelasi antara keduanya diberi lambang $r_{XX'}$. Simbol inilah yang kemudian diadopsi sebagai simbol koefisien reliabilitas (Azwar, 2005).

Secara teoretik, meskipun besar koefisien reliabilitas berkisar 0 sampai dengan 1 akan tetapi pada kenyataannya koefisien sebesar 1 tidak pernah dijumpai. Disamping itu, walaupun koefisien korelasi dapat saja bertanda (+) ataupun (-) akan tetapi dalam hal reliabilitas, koefisien yang besarnya kurang dari

0 tidak ada artinya karena interpretasi reliabilitas selalu mengacu kepada koefisien yang positif. Jangankan koefisien yang bertanda negatif, koefisien reliabilitas yang positif tetapi mendekati harga 0 pun tidak ada artinya dalam interpretasi reliabilitas (Azwar, 2005).

c. Bentuk Estimasi Reliabilitas

Estimasi reliabilitas dapat dilakukan melalui salah satu pendekatan umum, yaitu pendekatan tes ulang, pendekatan tes sejajar dan pendekatan konsistensi internal (Azwar, 2005). Selain tiga pendekatan tersebut Murphy & Davidshofer (2003) menambahkan estimasi reliabilitas dengan pendekatan belah dua.

1) Pendekatan tes ulang

Pendekatan tes ulang secara langsung mengukur derajat konsistensi skor dari satu penyajian tes ke penyajian berikutnya (Murphy & Davidshofer, 2003).

Pendekatan tes ulang meliputi:

- a) Menyajikan tes kepada suatu kelompok subjek,
- b) Menyajikan ulang tes yang sama kepada kelompok subjek yang sama untuk beberapa waktu kemudian, dan
- c) Mengkorelasikan skor dari hasil pengukuran pertama dengan skor pengukuran kedua.

Korelasi antara skor tes pertama dengan kedua digunakan untuk mengestimasi reliabilitas tes. Korelasi tes ulang sering diartikan sebagai koefisien stabilitas daripada koefisien reliabilitas (Murphy & Davidshofer, 2003).

Kelemahan pendekatan tes ulang adalah kurang praktisnya penerapan tes dua kali dan besarnya kemungkinan terjadi efek bawaan (*carry over effects*) dari satu

pengenaan tes ke pengenaan yang kedua. Artinya, kondisi subjek pada tes kedua tidak lagi sama dengan kondisi subjek pada tes pertama baik karena proses belajar, perubahan motivasi, pengalaman, sehingga pendekatan ini lebih baik digunakan bila objek ukur berupa keterampilan, terutama keterampilan fisik.

2) Pendekatan tes sejajar atau *alternate form method*

Pendekatan tes sejajar hanya dapat dilakukan apabila tersedia dua bentuk tes yang dapat dianggap memenuhi asumsi paralel. Salah satu indikator terpenuhinya asumsi paralel adalah setaranya korelasi antara skor kedua tes tersebut dengan skor suatu ukuran lain (Azwar, 2005).

Asumsi paralel tidak akan pernah terpenuhi, maka pendekatan bentuk alternatif pun dibentuk. Inti dari pendekatan ini adalah terdapat dua tes yang setara dalam hal konten, proses meresponnya, dan karakteristik statistiknya (Murphy & Davidshofer, 2003). Dalam mengestimasi reliabilitas pendekatan tes bentuk sejajar meliputi:

- a) Menyajikan satu bentuk tes (misalnya tes A) kepada satu kelompok subjek,
- b) Beberapa waktu berikutnya menyajikan bentuk tes lain yang setara dengan tes pertama (tes B) kepada kelompok subjek yang sama,
- c) Mengkorelasikan skor tes A dengan tes B. Korelasi antara kedua tes ini digunakan untuk mengestimasi reliabilitas tes.

Kelemahan utama pendekatan ini terletak pada sulitnya menyusun dua alat ukur yang memenuhi persyaratan paralel atau sejajar di samping pendekatan ini juga tidak menghilangkan sama sekali kemungkinan terjadinya efek bawaan.

3) Pendekatan belah dua

Pendekatan belah dua memberikan solusi terhadap permasalahan pada pendekatan tes sejajar atau setara yaitu sulitnya mengembangkan dua tes yang sejajar atau setara dan membutuhkan dua kali penyajian (Murphy & Davidshofer, 2003). Pendekatan belah dua ini meliputi:

- a) Menyajikan sebuah tes kepada suatu kelompok subjek,
- b) Membagi tes menjadi dua sama besar,
- c) Mengkorelasikan skor tes antara setengah belahan pertama dan setengah belahan kedua. Korelasi antara dua skor belahan ini digunakan untuk mengestimasi reliabilitas tes.

Kelemahan dari pendekatan ini adalah respon pada belahan pertama secara sistematis bisa berbeda dengan belahan kedua karena urutan kesukaran aitem yang dimulai dari paling mudah sampai paling sulit. Korelasi antara belahan pertama dan kedua bisa bervariasi bergantung pada cara pembelahan yang dilakukan (misalnya pembelahan random, gasal genap dan *matched random subtes*). Hal ini menimbulkan pertanyaan mana yang merupakan reliabilitas tes.

4) Pendekatan konsistensi internal

Estimasi reliabilitas dengan pendekatan konsistensi internal didasarkan pada data dari sekali pengenaaan satu bentuk alat ukur pada sekelompok subjek (*single trial administration*). Pendekatan konsistensi internal mengestimasi reliabilitas tes berdasarkan jumlah aitem dalam tes (k) dan rata-rata interkorelasi antar aitem tesnya (r_{ij}) (Murphy & Davidshofer, 2003). Dua faktor ini dapat dikombinasikan dalam formula berikut:

$$r_{xx'} = \frac{k (r_{ij})}{1+(k-1)r_{ij}} \dots\dots\dots (12)$$

Keterangan:

$r_{xx'}$ = Koefisien reliabilitas

k = Jumlah aitem dalam tes

r_{ij} = Interkorelasi antar aitem dalam tes

Terdapat hubungan antara pendekatan belah dua dengan konsistensi dalam mengestimasi reliabilitas dalam estimasi matematikanya. Perbedaan antara kedua pendekatan ini adalah pada analisis unitnya dimana pendekatan belah dua membandingkan setengah belahan tes pertama dengan lainnya sementara pendekatan konsistensi internal membandingkan setiap aitem dengan setiap aitem lainnya yang ada pada tes.

Keuntungan menggunakan pendekatan konsistensi internal adalah, dapat menghindari masalah-masalah yang biasanya ditimbulkan oleh pendekatan tes ulang dan pendekatan tes paralel.

Bentuk dan sifat alat ukur serta banyaknya belahan yang dibuat akan menentukan teknik estimasi koefisien reliabilitasnya. Di antara teknik-teknik komputasi reliabilitas, konsistensi internal menggunakan formula spearman-brown, formula rulon, formula Alpha, formula-formula kuder-richardson, formula kristof, formula analisis varians, dan sebagainya.

Uraian berikut akan disajikan beberapa di antara formula tersebut yang akan dijelaskan secara praktis.

d. Formula Estimasi Koefisien Reliabilitas

1) Formula Spearman Brown

Formula spearman brown dapat digunakan apabila jumlah aitem dalam tes adalah genap sehingga dapat dibelah menjadi dua bagian yang seimbang. Pembelahan ini dapat dilakukan dengan mengelompokkan aitem-aitem bernomor ganjil menjadi satu dan aitem-aitem bernomor genap menjadi satu pula. Formula spearman brown adalah:

$$r_{XX'} = 2 (r_{Y_1Y_2}) / (1+r_{Y_1Y_2}) \dots\dots\dots(13)$$

Keterangan:

$r_{XX'}$ = Koefisien korelasi

$r_{Y_1Y_2}$ = Koefisien korelasi antara skor belahan Y1 dan skor belahan Y2

Formula Sperman Brown digunakan jika aitem dikotomi ataupun politomi. Pembelahan tes dilakukan dengan cara gasal-genap dan *matched-random subtes* dan menghasilkan dua bagian yang paralel satu sama lain dan korelasi antara kedua belahan paralel tersebut cukup tinggi.

2) Formula Alpha

Untuk melakukan estimasi reliabilitas Alpha, tes dapat dibelah menjadi beberapa bagian. Dalam pembelahan ini, sangat penting untuk menjadikan banyaknya aitem dalam setiap belahan sama sehingga diharapkan belahan-belahan itu seimbang. Bila formula Alpha dikenakan pada tes yang dibelah tidak seimbang, maka koefisien yang diperoleh akan rendah dan merupakan

underestimasi terhadap reliabilitas yang sebenarnya. Formula Alpha untuk tes yang dibelah dua adalah:

$$\alpha = 2 \left(1 - \frac{s_1^2 + s_2^2}{s_x^2} \right) \dots\dots\dots (14)$$

Keterangan:

α = Koefisien reliabilitas Alpha

s_1^2 = Varians pada belahan 1

s_2^2 = Varians pada belahan 2

s_x^2 = Varians skor tes (X)

Tes yang dibelah menjadi lebih dari dua belahan yang masing-masing berisi aitem yang berjumlah sama banyak, dapat menggunakan formula Alpha dengan rumus:

$$\alpha = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum s_j^2}{s_x^2} \right] \dots\dots\dots (15)$$

Keterangan:

k = Banyaknya belahan tes

s_j^2 = Varians belahan j ; $j = 1, 2, \dots, k$

s_x^2 = Varians skor tes

Formula Alpha dapat digunakan jika aitem dikotomi ataupun politomi. Belahan tes tidak harus paralel, namun harus memenuhi asumsi τ -*equivalent*. Untuk menggunakan formula Alpha disyaratkan homogenitas isi belahan agar estimasi yang diperoleh mendekati reliabilitas yang sebenarnya.

3) Formula Kuder Richardson 20 (KR-20)

Kuder dan Richardson merumuskan banyak formula reliabilitas. Di antara yang paling terkenal adalah formula KR-20 dan KR-21. Formula KR-20 sebetulnya sama dengan formula Alpha, akan tetapi formula KR-20 hanya dapat dikenakan pada data skor dikotomi dari tes yang seolah-olah dibagi menjadi belahan sebanyak aitemnya. Formula KR-20 adalah:

$$KR - 20 = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum p(1-p)}{s_x^2} \right] \dots\dots\dots (16)$$

Keterangan:

- k = Banyaknya aitem
- p = Indeks kesukaran aitem
- s_x^2 = Varians skor tes (X)

4) Kuder-Richardson 21 (KR-21)

Estimasi KR-21 menggunakan rata-rata harga p dari keseluruhan aitem, hal inilah yang membedakan antara KR-20 dengan KR-21. Rumusan formula KR-21 adalah:

$$KR - 21 = \left[\frac{k}{k-1} \right] \left[1 - \frac{k\bar{p}(1-\bar{p})}{s_x^2} \right] \dots\dots\dots(17)$$

Keterangan:

- k = Banyaknya aitem dalam tes
- \bar{p} = Rata-rata p yaitu, $(\sum p) / k$
- s_x^2 = Varians skor tes

Formula KR-21 dapat pula dinyatakan sebagai:

$$KR - 21 = \left[\frac{k}{k-1} \right] \left[1 - \frac{M_x - M_x/k}{s_x^2} \right] \dots\dots\dots(18)$$

Keterangan:

M_x = Harga rata-rata *means* skor tes

k = Banyaknya aitem dalam tes

s_x^2 = Varians skor tes

Formula KR-21 dapat digunakan jika aitem dikotomi, jumlah aitem sedikit dan membelahan tes sebanyak jumlah aitem.

Estimasi koefisien reliabilitas subtes AN pada IST menggunakan pendekatan konsistensi internal berdasarkan metode sekali penyajian dan diestimasi dengan formula KR-20.

e. Interpretasi Koefisien Reliabilitas

Interpretasi terhadap koefisien reliabilitas bersifat relatif. Tidak ada batasan mutlak yang menunjukkan angka koefisien terendah yang harus dicapai agar suatu pengukuran dapat disebut reliabel. Kesepakatan informal menghendaki bahwa koefisien reliabilitas haruslah setinggi mungkin, biasanya suatu koefisien reliabilitas di sekitar 0,900 dapat dianggap memuaskan (Azwar, 2005).

Murphy dan Davidshofer (2003) menyatakan bahwa reliabilitas yang tinggi sangat diperlukan untuk tes yang digunakan untuk menetapkan keputusan akhir bagi seseorang dan untuk menentukan kategori individu berdasarkan perbedaan yang dimilikinya. Tabel 3 akan menunjukkan level reliabilitas berdasarkan tipe pengukuran Psikologi yang berbeda.

Tabel 3. Tingkat Reliabilitas untuk Berbagai Tipe Tes

Nilai Estimasi Reliabilitas	Jenis Tes	Interpretasi
0.95		Eror pengukuran memiliki efek yang rendah
0.90	Tes Intelegensi	Tingkat reliabilitas tinggi sampai menengah
0.85 0.80	Tes Prestasi	Tingkat reliabilitas tinggi sampai menengah
0.75	Tes Pilihan Berganda	Tingkat reliabilitas menengah sampai rendah
0.70	Skala	Tingkat reliabilitas menengah sampai rendah
0.65 0.60 0.55 0.50	Tes Proyektif	Reliabilitas rendah Skor murni dan eror memiliki efek yang sama terhadap hasil skor pengukuran

Berdasarkan Tabel 3 tersebut maka IST seharusnya memiliki koefisien reliabilitas sebesar 0.9 karena IST merupakan tes intelegensi yang digunakan untuk tujuan seleksi karyawan. Dimana hasil tes dijadikan dasar untuk menentukan apakah seseorang sesuai untuk suatu pekerjaan tertentu atau tidak berdasarkan kemampuan yang dimilikinya dalam hal ini intelegensi.

Saat membandingkan koefisien reliabilitas, interpretasi tidak dapat lepas dari besarnya varians skor (s_x^2). Dari sini kemudian dapat dihitung suatu statistik yang disebut eror standar dalam pengukuran (s_e). Eror standar dalam pengukuran merupakan ukuran variabilitas eror yang terjadi dalam pengukuran, yaitu:

$$s_e = s_x \sqrt{(1-r_{xx'})} \dots\dots\dots (19)$$

Keterangan:

- s_e = Eror standar pengukuran
- s_x = Deviasi standar skor tes
- $r_{xx'}$ = Koefisien reliabilitas tes

Besar kecilnya s_e merupakan indikator keterpercayaan pengukuran yang komparabel. Semakin kecil harga s_e berarti pengukuran tersebut semakin terpercaya dikarenakan variasi erornya semakin kecil.

5. Validitas Alat Ukur

a. Definisi

Validitas berasal dari kata *validity* yang mempunyai arti sejauh mana ketepatan dan kecermatan suatu instrumen pengukuran (tes) dalam melakukan fungsi ukurnya (Azwar, 2005).

Suatu tes dapat dikatakan mempunyai validitas yang tinggi apabila tes tersebut menjalankan fungsi ukurnya, atau memberikan hasil ukur yang tepat dan akurat sesuai dengan tujuan tes tersebut. Suatu tes yang menghasilkan data yang tidak relevan dengan tujuan diadakannya pengukuran dikatakan sebagai tes yang memiliki validitas rendah (Azwar, 2005).

Sisi lain yang sangat penting dalam konsep validitas adalah kecermatan pengukuran. Suatu tes yang validitasnya tinggi tidak saja akan menjalankan fungsi ukurnya dengan tepat akan tetapi juga dengan kecermatan tinggi, yaitu kecermatan dalam mendeteksi perbedaan-perbedaan kecil yang ada pada atribut yang diukurnya (Azwar, 2005).

Pengertian validitas sangat erat kaitannya dengan tujuan pengukuran. Tidak ada validitas yang berlaku secara umum untuk semua tujuan pengukuran. Suatu tes hanya menghasilkan ukuran valid untuk satu tujuan pengukuran saja yang spesifik. Oleh karena itu, suatu tes yang valid guna pengambilan keputusan

dapat saja tidak valid sama sekali guna pengambilan keputusan lain dan bagi kelompok lain (Azwar, 2005).

b. Koefisien Validitas

Hasil estimasi validitas suatu pengukuran pada umumnya dinyatakan secara empirik oleh suatu koefisien yang disebut koefisien validitas. Koefisien validitas dinyatakan oleh korelasi antara distribusi skor tes yang bersangkutan dengan distribusi skor suatu kriteria. Kriteria ini dapat berupa skor tes lain yang mempunyai fungsi ukur sama dan dapat pula berupa ukuran-ukuran lain yang relevan (Azwar, 2005).

Bila skor tes diberi simbol X dan skor kriteria mempunyai simbol Y, maka koefisien korelasi antara tes dan kriteria itu merupakan koefisien validitas, yaitu r_{xy} . Koefisien validitas hanya punya makna apabila mempunyai harga yang positif. Walaupun semakin tinggi mendekati angka 1 berarti suatu tes semakin valid hasil ukurnya namun pada kenyataannya suatu koefisien validitas tidak pernah mencapai angka 1. Bahkan memperoleh koefisien validitas yang tinggi adalah lebih sulit daripada memperoleh koefisien reliabilitas yang tinggi (Azwar, 2005).

c. Jenis dan Estimasi Validitas

Pada dasarnya, estimasi validitas dilakukan dengan menggunakan teknik analisis korelasional. Namun tidak semua pendekatan validitas memerlukan analisis statistika. Tipe validitas yang berbeda menghendaki cara analisis yang berbeda pula (Azwar, 2005).

Tipe validitas terbagi atas validitas isi, validitas konstruk, dan validitas berdasarkan kriteria. Validitas kriteria terbagi lagi atas tipe validitas konkuren dan validitas prediktif (Azwar, 2005).

1) Validitas isi

Validitas isi merupakan validitas yang diestimasi lewat pengujian terhadap isi tes dengan analisis rasional atau lewat *professional judgment*. Validitas isi menunjukkan sejauhmana aitem-aitem dalam tes mencakup keseluruhan kawasan isi yang hendak diukur oleh tes itu (Azwar, 2005). Pengertian mencakup keseluruhan kawasan isi tidak saja berarti tes itu harus komprehensif akan tetapi isinya harus pula tetap relevan dan tidak keluar dari batasan tujuan pengukuran.

Salah satu cara yang praktis untuk melihat apakah validitas isi telah terpenuhi adalah dengan melihat apakah aitem-aitem dalam tes telah ditulis sesuatu dengan *blue printnya* yaitu telah sesuai dengan batasan domain ukur yang telah ditetapkan semula dan memeriksa apakah masing-masing aitem telah sesuai dengan indikator perilaku yang hendak diungkapkannya (Azwar, 2005).

2) Validitas konstruk

Validitas konstruk adalah validitas yang menunjukkan sejauh mana suatu tes mengukur trait atau konstruk teoretik yang hendak diukurnya. Untuk pengujian validitas konstruk, diperlukan analisis statistika yang kompleks seperti prosedur analisis faktor (Azwar, 2005).

Salah satu prosedur pengujian validitas konstruk yang lebih sederhana adalah dengan melalui pendekatan *multitrait multimethod*. Pendekatan ini akan menguji serentak dua atau lebih trait yang diukur melalui dua atau lebih metode. Dari

prosedur ini akan diperoleh adanya bukti validitas diskriminan dan validitas konvergen (Azwar, 2005).

Validitas diskriminan diperlihatkan oleh rendahnya korelasi antara skor skala atau tes yang mengukur trait yang berbeda terutama bila digunakan metode yang sama, sedangkan validitas konvergen ditunjukkan oleh tingginya korelasi skor tes-tes yang mengukur trait yang sama dengan menggunakan metode yang berbeda (Azwar, 2005).

3) Validitas berdasar kriteria

Pengujian validitas berdasarkan kriteria, bukti validitas suatu tes diperlihatkan oleh adanya hubungan skor pada tes yang bersangkutan dengan skor suatu kriteria. Untuk melihat hubungan termaksud dilakukan analisis korelasional (Azwar, 2005).

Apabila suatu tes disusun untuk memprediksikan performansi di masa yang akan datang, maka akan lebih menekankan pentingnya validitas prediktif. Validitas prediktif dapat dilihat dari hasil analisis korelasional antara skor tes tersebut dengan skor performansi yang hendak diprediksikan. Skor performansi sebagai kriteria validasi tentu baru akan diperoleh setelah tenggang waktu tertentu (Azwar, 2005).

Setelah tenggang waktu tertentu, yaitu setelah diperolehnya skor performansi yang ingin diprediksikan, koefisien validitas tes dapat dihitung dengan menghitung korelasi antara skor tes dan skor performansi tersebut. Dalam hal ini, tes yang diuji validitasnya disebut sebagai prediktor dan performansi yang akan diprediksikan disebut sebagai kriteria (Azwar, 2005).

Ada kalanya kriteria untuk pengujian validitas tes telah tersedia. Artinya skor kriteria tersebut dapat diperoleh bersamaan dengan diperolehnya skor tes. Hal ini terjadi dikarenakan tujuan yang memang hendak melihat konkordansi atau kesesuaian hasil ukur tes dengan hasil suatu pengukuran yang relevan. Dengan demikian tentu saja komputasi korelasi antara tes dan kriteria dapat langsung dilakukan. Komputasi koefisien korelasi ini akan menghasilkan koefisien validitas konkuren (Azwar, 2005).

d. Interpretasi Koefisien Validitas

Sebagaimana halnya dengan koefisien reliabilitas, interpretasi koefisien validitas juga bersifat relatif, artinya tidak ada batasan pasti mengenai koefisien terendah yang harus dipenuhi agar validitas dinyatakan memuaskan. Dalam memberikan arti terhadap koefisien validitas, haruslah diingat bahwa walaupun dikehendaki adanya koefisien yang setinggi mungkin mendekati $r_{xy} = 1$ akan tetapi memperoleh koefisien validitas yang tinggi seperti itu adalah lebih sulit daripada memperoleh koefisien reliabilitas yang tinggi (Azwar, 2005).

Hal tersebut terjadi karena korelasi antara skor suatu tes dengan skor tes lainnya (validitas) tidak dapat lebih tinggi daripada korelasi antara skor tes itu dengan skor murninya sendiri (reliabilitas). Artinya, tingginya koefisien validitas akan dibatasi oleh tingginya koefisien reliabilitas (Azwar, 2005).

Koefisien validitas yang dianggap memuaskan akan dikembalikan kepada para penguji validitas dan pemakai tes itu sendiri, terutama pemakai tes yang akan memanfaatkan keputusan yang didasari hasil pengukuran yang bersangkutan.

Koefisien validitas dapat dianggap memuaskan apabila melebihi 0,30 (Azwar, 2005).

6. Analisis Karakteristik Psikometri Alat Ukur

Tes disusun dari sejumlah aitem yang bertujuan untuk mengukur atribut atau konstruk Psikologi tertentu. Oleh karena itu, kualitas tes akan sangat bergantung pada kualitas aitem yang menyusunnya (Murphy & Davidshofer, 2003).

Untuk melihat apakah aitem-aitem yang menyusun suatu tes itu memiliki kualitas yang baik dapat dilihat dari karakteristik-karakteristik aitem yang terdiri dari indeks kesukaran aitem dan indeks daya beda aitem. Selain dua karakteristik tersebut penilaian kualitas tes berdasarkan karakteristik Psikometri juga dapat dilihat dari efektifitas distraktor-distraktornya (Azwar, 2005).

Meskipun tiga karakteristik Psikometri ini dihitung secara terpisah, namun dalam evaluasinya terhadap aitem, karakteristik aitem tersebut tidak bisa berdiri sendiri melainkan dilihat sebagai kesatuan komponen yang akan menentukan apakah suatu aitem itu dapat dianggap baik atau tidak (Azwar, 2005).

Tingkat kesukaran aitem sangat dipengaruhi oleh tingkat keterpercayaan distraktor, jika semua distraktor tidak masuk akal maka subjek akan dengan mudah untuk memilih jawaban yang benar tanpa harus memiliki pengetahuan tentang hal yang ditanyakan, tentu hal ini mempengaruhi tingkat kesukaran aitem, sehingga tingkat kesukaran aitem menjadi rendah. Hal yang sama terjadi pada daya beda aitem. Efektifitas distraktor yang buruk akan mempengaruhi daya beda

aitem sehingga aitem-aitem dalam tes tidak mampu membedakan mana subjek yang memiliki pengetahuan dan mana yang tidak.

Tingkat kesukaran aitem secara langsung mempengaruhi daya beda aitem. Suatu aitem yang terlalu sukar sehingga tidak seorang pun dari kelompok subjek dapat menjawab dengan benar tidak akan mempunyai daya beda. Demikian juga apabila suatu aitem begitu mudahnya sehingga semua subjek dapat menjawab dengan benar. Aitem yang terlalu sulit atau terlalu mudah tidak akan bisa membedakan antara mereka yang memiliki atribut yang diukur atau tidak sehingga indeks daya beda bernilai rendah (Azwar, 2005).

Bagaimana kualitas ketiga karakteristik Psikometri tersebut akan sangat mempengaruhi nilai reliabilitas dan validitas alat ukur. Kualitas aitem-aitem yang buruk akan menurunkan bahkan tidak memiliki nilai reliabilitas dan validitas sehingga hasil dari pengukuran tidak akan berguna (Azwar, 2005).

Indeks daya beda aitem dapat dihitung dengan menggunakan korelasi inter aitem atau aitem total. Jika korelasi yang didapat dari dua estimasi tersebut tinggi maka dapat diartikan aitem mampu membedakan antara kelompok yang memiliki kemampuan atribut ukur atau tidak. Selain menunjukkan daya beda korelasi inter aitem dan korelasi aitem total juga menunjukkan apakah aitem-aitem dalam tes mengukur hal yang sama dengan tes nya. Hal ini menunjukkan apakah tes tersebut valid atau tidak. Oleh karena itu, bisa disimpulkan bahwa aitem yang memiliki indeks daya beda aitem yang tinggi juga memiliki nilai validitas yang tinggi (Murphy & Davidshofer, 2003).

Tes yang valid jika dilakukan pengukuran ulang terhadap sampel perilaku yang sama meski dengan subjek yang berbeda tetapi masih dalam kriteria yang setara akan menghasilkan nilai reliabilitas yang tinggi. Tetapi perlu diingat suatu tes yang reliabel belum tentu valid. Hal ini sangat tergantung pada tujuan dan kepada siapa tes tersebut diberikan. Jika korelasi dari dua tes tinggi memang menunjukkan tes tersebut reliabel tetapi belum tentu menunjukkan apakah tes tersebut mengukur atribut ukur yang seharusnya (Azwar, 2009).

Banyaknya jumlah aitem dalam tes juga mempengaruhi reliabilitas dan validitas tes. Semakin banyak jumlah aitem, maka nilai reliabilitas akan meningkat. Tetapi hal ini terjadi jika hanya kualitas aitem baik. Kualitas aitem yang baik walaupun jumlahnya sedikit juga akan memiliki reliabilitas yang tinggi tetapi belum tentu valid karena aitem yang sedikit tidak mewakili luasnya kawasan atribut yang hendak diukur (Azwar, 2005).

C. *Intelligenz Struktur Test (IST)*

1. Sejarah dan Perkembangan IST

Intelligenz structure test (IST) merupakan salah satu tes yang digunakan untuk mengukur intelegensi individu. Tes ini dikembangkan oleh Rudolf Amthauer di Frankfurt, Jerman pada tahun 1953. Amthauer mendefinisikan intelegensi sebagai keseluruhan struktur dari kemampuan jiwa-rohani manusia yang akan tampak jelas dalam hasil tes. Intelegensi hanya akan dapat dikenali atau dilihat melalui manifestasinya misalnya pada hasil atau prestasi suatu tes (Polhaupessy dalam Diktat Kuliah Universitas Padjajaran, 2009).

Amthaeuer (dalam Polhaupessy dalam Diktat Kuliah Universitas Padjajaran, 2009) berpendapat bahwa intelegensi merupakan suatu kesatuan dari seluruh kemampuan yang dimiliki oleh seseorang. Berdasarkan pemikiran ini Amthauer menyusun sebuah tes yang dinamakan IST dengan hipotesis kerja sebagai berikut:

“komponen dalam struktur tersebut tersusun secara hierarkis; maksudnya bidang yang dominan kurang lebih akan berpengaruh pada bidang-bidang yang lain; kemampuan yang dominan dalam struktur intelegensi akan menentukan dan mempengaruhi kemampuan yang lainnya.”

Pandangan Amthaeur pada dasarnya di dasari oleh teori faktor, baik itu teori dua faktor, teori bifaktor, teori multifaktor, model struktur intelek Guilford dan teori hierarki faktor. Menurut Spearman (dalam Azwar, 2004) performa dalam tes intelegensi ditentukan oleh faktor umum (g), ditambah 1 atau lebih faktor spesifik (s). Faktor g merupakan kekuatan atau energi mental yang mendasari aktivitas-aktivitas mental yang rumit. Kemudian Thurstone (dalam Anastasi & Urbina, 2006) menyebutkan ada banyak faktor (*multiple factor theories*) yang mempengaruhi kemampuan seseorang, yang disebut kemampuan mental primer. Faktor-faktor tersebut meliputi pemahaman verbal, *word fluency*, berhitung, pemahaman spasial, asosiasi memori, kecepatan perceptual dan penalaran. Teori tersebut merupakan salah satu dasar pengembangan IST oleh Amthauer dimana intelegensi menurutnya merupakan suatu kesatuan dari seluruh kemampuan yang dimiliki seseorang (Polhaupessy dalam Diktat Kuliah Universitas Padjajaran, 2009). Hal ini didukung oleh Anastasi dan Urbina (2006) yang mengungkapkan bahwa intelegensi tidak hanya berupa satu keseluruhan kemampuan saja, tetapi berupa gabungan dari beberapa fungsi.

Berdasarkan teori faktor tersebut Amthaeuer (dalam Polhaupessy dalam Diktat Kuliah Universitas Padjajaran, 2009) menyatakan bahwa untuk mengukur intelegensi seseorang diperlukan suatu rangkaian baterai tes yang terdiri dari subtes-subtes. Antara subtes satu dengan lainnya, ada yang saling berhubungan karena mengukur faktor yang sama (*general factor* atau *group factor*), tetapi ada juga yang tidak berhubungan karena masing-masing subtes mengukur faktor khusus (*special factor*). Berdasarkan hal ini Amthauer menyusun IST sebagai baterai tes yang terdiri dari 9 subtes yaitu *Satzergänzung* (SE), *Wortauswahl* (WA), *Analogien* (AN), *Gemeinsamkeiten* (GE), *Rechenaufgaben* (RA), *Zahlenreihen* (ZR), *Figurenauswahl* (FA), *Würfelaufgaben* (WU) dan *Merkaufgaben* (ME). Untuk pertama kalinya IST dipublikasikan Rudolff Amthaeuer pada tahun 1970 (Polhaupessy dalam Diktat Kuliah Universitas Padjajaran, 2009).

Karakteristik dari baterai tes Amthauer menunjukkan adanya suatu interkorelasi yang rendah antar subtesnya ($r=0.25$) dan korelasi antara subtes dengan jumlah (keseluruhan subtes) yang rendah pula ($r=0.60$). Hal ini mengandung arti bahwa intelegensi terdiri dari struktur-struktur tertentu, dimana subtes-subtes pada IST disusun untuk mengukur kemampuan yang berbeda dalam penggambaran keseluruhan kemampuan intelegensi seseorang, sehingga setiap subtes dalam IST dapat digunakan untuk mengukur kemampuan umum dan kemampuan khusus (Polhaupessy dalam Diktat Kuliah Universitas Padjajaran, 2009). Berdasarkan karakteristiknya, IST tergolong kelompok *Multiple Aptitude Batteries Test*, yaitu sebuah tes yang tersusun atas serangkaian subtes dimana

masing-masing subtes tersebut mengukur suatu kemampuan (Anastasia & Urbina, 2006).

Semenjak diciptakan, IST terus dikembangkan oleh Amthauer dengan bantuan dari para koleganya. Mulai populer digunakan di Jerman pada tahun 1970, IST versi revisi yang pertama yaitu IST 70 kemudian pada tahun 1999 IST kembali mengalami revisi menjadi IST 2000, dan kemudian pada tahun 2007 mengalami revisi kembali menjadi IST 2000R (http://de.wikipedia.org/wiki/I-S-T_2000R). Berikut adalah perkembangan tes IST dari tahun 1953 hingga tahun 2000-an (Polhaupessy dalam Diktat Kuliah Universitas Padjajaran, 2009).

a. IST 1953

IST yang pertama ini pada awalnya hanya diperuntukan untuk usia 14 sampai dengan 60 tahun. Proses penyusunan norma diambil dari 4000 subjek pada tahun 1953.

b. IST 1955

IST merupakan pengembangan dari IST 1953. Pada IST 1955 *rentang* untuk subjek diperluas menjadi berawal dari umur 13 tahun. Subjek dalam penyusunan norma bertambah menjadi 8642 orang. Pada tes ini sudah ada pengelompokan jenis kelamin dan kelompok usia.

c. IST 70

Berdasarkan permintaan dan tuntutan pengguna yang menyarankan pengkoreksian dengan mesin juga pengembangan tes setelah penggunaan lebih dari 10 tahun, maka disusunlah IST 70. Dalam IST 70 ini tidak terlalu banyak perubahan, tes ini memiliki 6 bentuk, setiap pemeriksaan dilakukan 2 tes sebagai

bentuk paralel; yaitu A1 dan B2, atau C3 dan D4. Dua bentuk lainnya untuk pemerintah dan hanya bagi penggunaan khusus. Pada IST 70, rentang kelompok usia diperluas menjadi berawal dari 12 tahun. Disamping itu telah ditambah Tabel kelompok dan pekerjaan. Namun demikian, pada IST 70 terdapat kekurangan yaitu penyebaran bidang yang tidak merata dan menggunakan kalimat dalam subtes RA sehingga jika subjek gagal dalam subtes ini dapat dimungkinkan karena tidak mampu mengerjakan soal hitungannya atau tidak mengerti kalimatnya.

d. IST 2000

Sebagai koreksi dari IST 70, pada IST 2000 tidak terdapat soal kalimat pada soal hitungan.

e. IST 2000-Revised

IST 2000-R mengalami beberapa perkembangan pada subtesnya dan juga terdapat penambahan subtes. IST ini terdiri dari 3 modul, yaitu sebagai berikut:

1) *Grundmodul-Kurzform* (Modul Dasar-Singkatan)

Terdiri dari subtes *Satzergänzung* (SE), *Analogien* (AN), *Gemeinsamkeiten* (GE), *Rechenaufgaben* (RE), *Zahlenreihen* (ZR), *Rechenzeichen* (RZ), *Figurenauswahl* (FA), *Würfelaufgaben* (WÜ), dan *Matrizen* (MA).

2) Modul *Merkaufgaben*

Terdiri dari subtes *Merkaufgaben Verbal* dan *Merkaufgaben Figural*.

3) *Erweiterungmodul* (Modul "menguji pengetahuan")

Terdiri dari subtes *Wissentest* (tes pengetahuan).

2. Subtes-subtes pada IST

Intelligenz struktur test (IST) memiliki 9 subtes yang masing-masing memberikan skor taraf intelegensi seseorang. Subtes-subtes tersebut adalah (Amthauer dalam Polhaupessy dalam Diklat Kuliah Universitas Padjajaran, 2009).

a. *Satzergänzung* (SE): Melengkapi Kalimat

Hal yang diukur dalam Subtes ini adalah pembentukan keputusan, dapatkah seseorang berprestasi, rasa realitas atau menilai yang mendekati realitas, pengetahuan umum dengan memanfaatkan pengalaman masa lalu, dapatkah seseorang berpikir secara berdikari atau mandiri dan berpikir praktis dalam kehidupan sehari-hari.

b. *Wortauswahl* (WA): Melengkapi Kalimat

Hal yang diukur dalam Subtes ini adalah kemampuan bahasa, perasaan empati, berpikir induktif menggunakan bahasa, memahami pengertian.

c. *Analogien* (AN): Persamaan Kata

Hal yang diukur dalam Subtes ini adalah kemampuan fleksibilitas dalam berpikir, kemampuan berpikir logis atau menggunakan pikiran sebagai dasar dalam berpikir (kedalaman berpikir).

d. *Gemeinsamkeiten* (GE): Sifat yang Dimiliki Bersama

Hal yang diukur dalam Subtes ini adalah kemampuan abstraksi, kemampuan untuk menyatakan pengertian akan sesuatu dalam bentuk bahasa, membentuk suatu pengertian atau mencari inti persoalan.

e. *Rechenaufgaben* (RA): Berhitung

Hal yang diukur dalam Subtes ini adalah bagaimana cara berpikir induktif praktis hitungan, kemampuan berhitung, menggunakan bilangan secara praktis dalam masalah hitungan.

f. *Zahlenreihen* (ZR): Deret Angka

Hal yang diukur dalam Subtes ini adalah bagaimana cara berpikir teoritis dengan hitungan.

g. *Figurenauswahl* (FA): Memilih Bentuk

Hal yang diukur dalam Subtes ini adalah kemampuan dalam membayangkan, kemampuan mengkonstruksi (sintesa dan analisa), berpikir konkrit menyeluruh, memasukkan bagian pada suatu keseluruhan.

h. *Würfelaufgaben* (WU): Latihan Balok

Hal yang diukur dalam Subtes ini adalah daya bayang ruang, kemampuan tiga dimensi.

i. *Merkaufgaben* (ME): Latihan Simbol

Subtes ini mengukur daya ingat, konsentrasi yang menetap, dan daya tahan.

3. Fungsi dan Tujuan IST

Tes ini dipandang sebagai gestalt (menyeluruh), yang terdiri dari bagian-bagian yang saling berhubungan secara bermakna (struktur). Dimana struktur intelegensi tertentu menggambarkan pola kerja tertentu, sehingga akan cocok untuk profesi atau pekerjaan tertentu. Berdasarkan hal tersebut IST umumnya

digunakan untuk memahami diri dan pengembangan pribadi, merencanakan pendidikan dan karier serta membantu pengambilan keputusan dalam hidup individu (Polhaupessy dalam Diktat Kuliah Universitas Padjajaran, 2009).

4. Skoring dan Interpretasi

a. Skoring

Tahap skoring yang digunakan untuk setiap subtes adalah dengan memeriksa setiap jawaban dengan menggunakan kunci jawaban yang telah disediakan. Untuk semua subtes (SE, WA, AN, RA, ZR, FA, WU, & ME), kecuali subtes 04-GE, setiap jawaban benar diberi nilai 1 dan untuk jawaban salah atau kosong diberi nilai 0. Khusus untuk subtes 04-GE, tersedia nilai 2, 1, dan 0; karena subtes ini berbentuk isian singkat maka nilai yang akan diberikan tergantung dengan jawaban yang diberikan oleh subjek (Polhaupessy dalam Diktat Kuliah Universitas Padjajaran, 2009).

Total nilai benar yang sesuai dengan kunci jawaban merupakan *Raw Score* (RW); nilai ini belum dapat diinterpretasi sesuai dengan norma yang digunakan. Nilai RW yang sudah dibandingkan dengan norma disebut dengan *Standardized Score* (SW). Nilai SW inilah yang dapat menjadi materi untuk tahap selanjutnya, yaitu interpretasi. Adapun norma yang digunakan adalah sesuai dengan kelompok umur subjek (Polhaupessy dalam Diktat Kuliah Universitas Padjajaran, 2009).

b. Interpretasi

Setelah didapatkan *Standardized Score*, maka tahap interpretasi dapat dilakukan. Kesembilan subtes saling berkaitan, sehingga harus dilakukan

semuanya dan interpretasinya harus dilakukan secara keseluruhan (Amthauer dalam Polhaupessy dalam Diktat Kuliah Universitas Padjajaran, 2009).

Interpretasi yang dapat dilakukan dari tes IST adalah sebagai berikut:

1) Taraf Kecerdasan

Taraf kecerdasan didapat dari total SW. Nilai ini dapat diterjemahkan menjadi *Intelligent Quotient* (IQ). Nilai ini dapat menggambarkan perkembangan individu melalui pendidikan dan pekerjaan. Nilai ini perlu dihubungkan dengan latar belakang sosial serta dibandingkan dengan kelompok seusianya.

2) Dimensi *Festigung-Flexibilitas*

Dimensi *Festigung-Flexibilitas* menggambarkan corak berpikir yang dimiliki oleh subjek. Dimensi *Festigung-Flexibilität* merupakan dua kutub yang ekstrim, keduanya menggambarkan corak berpikir yang ekstrim pula. Kutub *Festigung* memiliki arti corak berpikir yang eksak, sedangkan kutub *Flexibilität* memiliki arti corak berpikir yang non-eksak. Corak berpikir ini merupakan hasil perkembangan dan pengalaman individu yang akan semakin mengarah ke salah satu kutub seiring bertambahnya usia.

Tabel 4. Dimensi *Festigung-Flexibilitas*

GE+RA	>	AN+ZR	Festigung
GE+RA	<	AN+ZR	Flexibilitat

* Catatan: jika selisih ≤ 10 , maka tidak dapat dilakukan interpretasi.

Untuk menentukan seseorang subjek apakah memiliki kecenderungan *Festigung* atau *Flexibilitat* adalah dengan membandingkan nilai GE+RA dengan nilai AN+ZR. Jika nilai GE+RA lebih besar, maka subjek memiliki kecenderungan

Festigung, sebaliknya jika nilai AN+ZR lebih besar, maka subjek memiliki kecenderungan *Flexibilitas*.

3) Profil M-W

Profil M-W menggambarkan cara berpikir, apakah verbal-teoritis atau praktis-konkrit. Untuk mendapatkan profil M atau W ini dapat dilihat dari 4 subtes pertama (SE, WA, AN, GE) yang tampak pada grafik. Jika grafik menunjukkan bentuk M pada empat subtes pertama, maka profilnya adalah M (verbal-teoritis), jika yang tampak adalah bentuk huruf W, maka profilnya adalah W (praktis-konkrit).

4) Struktur Kecerdasan

Struktur kecerdasan menggambarkan kecerdasan subjek berdasarkan masing-masing subtes.

5) Kesesuaian terhadap Jurusan atau Pekerjaan

Interpretasi yang kelima adalah kesesuaian dengan jurusan atau pekerjaan artinya sesuai dengan kepentingan. IST biasanya digunakan dalam proses seleksi, baik seleksi jurusan di SMU, seleksi perguruan tinggi, maupun seleksi pekerjaan. Untuk melihat kesesuaian terhadap jurusan atau pekerjaan, perlu ditinjau norma untuk masing-masing jurusan atau pekerjaan yang berisi nilai SW sebagai batas yang dibutuhkan untuk jurusan atau pekerjaan tersebut.

Selain untuk melihat kesesuaian terhadap jurusan dapat pula dilakukan dengan melihat grafik subjek dan membandingkannya dengan bentuk grafik jurusan atau pekerjaan tertentu. Jika bentuknya kurang lebih sama, maka subjek memiliki kesempatan untuk menempuh jurusan atau pekerjaan tersebut.

5. Subtes *Analogien* (AN)

Penelitian ini berfokus pada analisis karakteristik Psikometri subtes *analogien* (AN) pada *intelligenz struktur test* (IST). Subtes AN merupakan subtes urutan ketiga dari sembilan subtes IST yang mengukur daya mengkombinasikan, fleksibilitas atau kelincahan berfikir, kemampuan mendeteksi dan memindahkan hubungan-hubungan, kejelasan dan kekonsekuenan dalam berfikir, dan perlawanan atau pertentangan terhadap penguraian yang bersifat mengira-ngira (Amthauer, dalam Polhaupessy, dalam Diktat Kuliah Universitas Padjajaran, 2008).

Subtes AN mengharuskan individu untuk mencari terlebih dahulu hubungan antara kata pertama dan kedua, baru kemudian menerapkan hubungan yang sama pada kata yang ketigadan keempat. Secara umum, subtes ini mengukur bagaimana menemukan hubungan antara dua hal atau lebih dan menemukan ide atau hal yang lainnya, diasosiasikan dengan hubungan sebelumnya. Subtes AN menuntut individu untuk dapat berfikir fleksibel, mudah beralih dan dapat melihat suatu hal dari banyak sisi permasalahan, sehingga dapat menyelesaikan suatu persoalan. Hal ini menunjukkan bahwa diperlukan daya analisis dalam pengambilan keputusan di setiap persoalan (Diktat Kuliah Universitas Padjajaran, 2008).

Pola berpikir seperti ini merupakan pola berpikir secara induktif (*inductive reasoning*) yang termasuk kepada kemampuan *general reasoning*, sebagaimana yang diungkapkan oleh Thurstone dalam Anastasi dan Urbina (2006), bahwa *general reasoning* atau disebut *induction* paling baik diukur dengan menggunakan

tes yang menuntut responden untuk menemukan suatu aturan atau pola, seperti misalnya tes melengkapi kata. Kemudian Stenberg (dalam Anastasi & Urbina, 2006) juga menambahkan bahwa kemampuan penalaran induktif berakar pada *fluid intelligence* yang berhubungan dengan kemampuan bernalar. Jadi penalaran induktif digunakan ketika seseorang menghubungkan informasi yang terpisah-pisah, kemudian digeneralisasikan ke dalam suatu kesimpulan yang berkaitan dengan informasi tersebut. Dalam subtes AN ini, informasi yang dimaksud berupa kata-kata (verbal).

Analogien atau analogi verbal berarti penalaran verbal yang terdiri dari dua kata, yaitu penalaran dan verbal. Menurut Chaplin (2006), penalaran bermakna proses berpikir logis atau memecahkan masalah dan verbal berkaitan dengan kata-kata baik dalam bentuk tulisan, cetakan dan seterusnya. Dari arti kata tersebut dapat disimpulkan bahwa penalaran verbal merupakan proses berpikir logis untuk memecahkan masalah yang berkaitan dengan pengertian kata-kata.

Tahap skoring yang digunakan adalah dengan memeriksa setiap jawaban dengan menggunakan kunci jawaban yang telah disediakan. Untuk semua subtes AN, setiap jawaban benar diberi nilai 1 dan untuk jawaban salah atau kosong diberi nilai 0 (Diktat Kuliah Universitas Padjajaran, 2008).