

Validitas dan Reliabilitas Tes Kemampuan Diferensial Analogi Verbal

Prita Indah Lestari dan Angela Oktavia Suryani

Fakultas Psikologi
Universitas Katolik Atma Jaya Jakarta

Abstract: *this study is aimed at examining the validity and reliability of TKD 3, which measures verbal analogy. Data (N = 606) analysed were taken from Service Bureau of Faculty of Psychology Atma Jaya Catholic University. The data consisted of scores from high schoolers and employees. The results showed that 55% items in TKD 3 were invalid measuring verbal analogy, and it was unable to discriminate the upper and lower groups. Predictive validity analysis revealed that TKD 3 was unqualified as differential test. Reliability analysis showed that TKD 3 was inconsistent measuring verbal analogy. Thus, a new test was needed to be developed as its alternative.*

Key words: *differential ability tests, verbal analogy, validity and reliability.*

Abstrak: *penelitian ini bertujuan untuk mengetahui validitas dan reabilitas dari TKD 3 yang digunakan untuk mengukur analogi verbal. Data yang dianalisis (N=606) diambil dari biro pelayanan Fakultas Psikologi Universitas Katolik Atma Jaya. Data terdiri dari skor dari para peserta tes: cendekia dan karyawan. Hasil penelitian menunjukkan bahwa 55% soal di TKD 3 tidak valid untuk mengukur analogi verbal, dan ini menunjukkan bahwa tes ini tidak mampu menggolongkan kelompok atas dan bawah. Analisis validitas prediktif menunjukkan bahwa TKD 3 tidak berkualifikasi sebagai tes pembeda. Analisis Reliabilitas menunjukkan bahwa TKD 3 tidak konsisten dalam mengukur analogi verbal. Untuk itu, sebuah tes baru sangat diperlukan untuk dikembangkan sebagai alternatif.*

Kata kunci: *uji kemampuan diferensial, analogi verbal, validitas dan reliabilitas.*

Tes Kemampuan Diferensial yang dikenal dengan sebutan TKD merupakan sebuah baterai tes yang disusun oleh Sudirgo Wibowo (1976) berdasarkan teori *Primary Mental Abilities* yang disusun oleh Thurstone. Tes ini bukan tergolong tes inteligensi yang antar subtesnya berkorelasi tinggi sehingga dapat menghasilkan skor tunggal IQ. Alat tes ini merupakan sebuah baterai tes yang masing-masing subtesnya bisa berdiri sendiri dan bahkan bisa digunakan bersama dengan tes lain diluar TKD. Ketika alat tes ini disusun, tes ini ditujukan untuk mengidentifikasi kekuatan dan kelemahan seseorang sebagai pertimbangan untuk seleksi masuk perguruan tinggi. Dengan kata lain, alat ini digunakan untuk melihat kemampuan dasar seseorang apakah ia mampu mengikuti kegiatan

akademik di perguruan tinggi. Pada perkembangan selanjutnya, tes ini digunakan sebagai instrumen untuk penjurusan di SMA ataupun seleksi karyawan di perusahaan oleh biro-biro konsultasi psikologi.

Sejak disusun pada tahun 1976 di Universitas Indonesia, tes ini telah mengalami beberapa revisi bahkan beberapa bagian mengalami perubahan disesuaikan dengan berbagai situasi dan kebutuhan. Tim konsultasi profesional penelitian dan diagnostik alat tes LPSP3 (Lembaga Pengembangan Sarana Pengukuran dan Pendidikan Psikologi) Universitas Indonesia mengemukakan pertimbangan merivisi antara lain adalah untuk *up-to-date* kemuktahiran, karena lingkup pengetahuan selalu berkembang, lalu perlunya penyesuaian dengan konteks budaya yang ada pada masyarakat, dan perlunya alat tes bersifat *recent*, artinya sesuai dengan informasi terkini. Istilah-istilah dan pengetahuan-pengetahuan pada jaman dahulu berubah, terutama peristiwa sejarah, situasi sosial, dan sebagainya (wawancara pribadi dengan dua orang tim LPSP3-UI). Dikemukakan pula bahwa revisi dan perubahan-perubahan yang terjadi pada TKD dilakukan oleh beberapa pihak, namun hingga saat ini belum ada hasil laporan ataupun publikasi mengenai uji validitas dan reliabilitas maupun revisi ataupun perubahan yang dilakukan. Selain itu, hingga saat ini, tes masih menggunakan norma mahasiswa karena ketika disusun tes ini digunakan untuk keperluan seleksi perguruan tinggi. Jikapun ada yang menyusun norma untuk kelompok normatif lainnya seperti siswa sekolah ataupun calon karyawan, hingga saat ini laporan mengenai hasil uji validitas dan reliabilitas TKD pada kelompok tersebut belum ada laporan ataupun publikasinya. Hal ini sangat riskan untuk kepentingan penelitian maupun praktek layanan psikologis yang menggunakan tes ini apabila ternyata tes tersebut tidak layak untuk digunakan mengingat usianya yang tua dan meskipun ada revisi, laporan hasil revisi tidak pernah diketahui isinya.

Berdasarkan permasalahan tersebut, penelitian berikut ini bertujuan untuk menguji validitas konstruk (termasuk di dalamnya uji tingkat kesulitan soal), validitas prediktif dan reliabilitas dari TKD. Selain itu penelitian ini hendak melihat apakah TKD mampu memenuhi fungsinya sebagai tes diferensial. Tulisan berikut ini akan memaparkan hasil uji psikometri subtes TKD 3 yaitu tes analogi verbal.

Alat tes ini terdiri dari 40 soal dengan waktu administrasi 5 menit. Bentuk soal berupa pilihan ganda dengan 4 pilihan jawaban

dan metode skoring dikotomi (1 – 0) yaitu skor 1 untuk jawaban benar dan skor 0 untuk jawaban salah.

Analogi verbal sendiri terdiri atas dua fungsi kognitif yaitu analogi dan verbal. Secara sederhana, analogi dapat berarti "sebanding dengan". Persoalan-persoalan yang bersifat analogi biasanya berbentuk perbandingan $A : B = C : D$ atau modifikasi-modifikasinya. Hubungan yang terjadi antara A dan B bersifat analogi atau sebanding dengan hubungan yang terjadi antara C dan D, atau pola hubungan yang terjadi pada A dan B diterapkan pada pola yang terjadi pada C dan D. Pola-pola analogi bisa digunakan untuk memecahkan masalah yang stimulus-stimulusnya berupa figur geometris, numerik, verbal, dll. Verbal berarti berkaitan dengan kata atau kumpulan kata. Melalui pengertian-pengertian tersebut, analogi verbal dapat dipahami sebagai proses bernalar analogis yang melibatkan kata-kata untuk memecahkan suatu masalah (Duran, Enright, dan Peirce, 1987). Steward, Barnes-Holmes, dan Roche (2004) menyatakan bahwa dalam menyelesaikan persoalan analogi verbal, fungsi-fungsi yang terlibat antara lain adalah fungsi penghubungan atau penyusunan kombinasi, penggunaan logika, fleksibilitas dalam berpikir, dan juga kemauan untuk berpikir eksploratif.

Sebagai tes yang mengukur inteligensi berdasarkan teori *Primary Mental Ability* dari Thurstone, TKD analogi verbal tergolong pada tes penalaran induktif. Penalaran induktif adalah proses bernalar suatu persoalan dari khusus ke umum. Pada analogi verbal, proses yang dilakukan adalah berusaha mencari pola dari suatu persoalan, lalu menciptakan hipotesa atau formula untuk menjelaskan pola dari suatu contoh atau kejadian yang ada tersebut supaya dapat menerapkannya pada persoalan yang baru dengan pola yang sama (Sternberg, 1994).

Pola-pola analogi sangat beragam. Beberapa pola tersebut dijelaskan oleh Bandurski dan Galkowski (2004), Nippold (1994), dan Sternberg (1994). Pola-pola yang dikemukakan oleh ketiga ahli ini saling tumpang tindih, meskipun mereka menggunakan istilah yang berbeda namun sebenarnya maknanya sama. Contohnya dapat dilihat pada pola *opposite* yang dikemukakan oleh Barduski dan Galkowski (2004) yang sebetulnya sama dengan pola *antonymous* pada Nippold (1994) atau pola *causality* pada Barduski dan Galkowski (2004) dengan pola *cause/effect* pada Sternberg (1994). Selain itu pada Nippold (1994) ditemukan adanya pola *categorical*

yang mirip dengan pola *part-whole* pada Barduski dan Galkowski (2004) dan *part-to-whole* pada Sternberg (1994).

Apabila pola analogi yang ada pada TKD 3 dianalisis berdasarkan pola-pola yang dipaparkan oleh ketiga ahli tersebut, ditemukan bahwa pola analogi yang dominan adalah pola *categorical* (8 soal), berikutnya adalah pola *functional* (7 soal) dan pola perbandingan sifat atau *characteristics* (7 soal). Area topik yang dianalogikan pada TKD sangat beragam, antara lain topik mengenai hewan, pakaian daerah, proses rantai makanan, tumbuhan, kendaraan, dsb. Hasil analisis mengenai pola-pola analogi yang ada pada TKD 3 dipaparkan pada Tabel 1.

Tabel 1.
Pola-Pola Analogi Pada Tes-Tes Analogi Verbal

Barduski & Galkowski (2004)	Nippold (1994)	Sternberg (1994)	TKD (Wibowo, 1976)
1. <i>opposite</i> barat : timur = siang : ...	1. <i>functional</i> pilot : terbang = joki : ...	1. <i>part-to whole</i> halaman : buku = kamar : ...	1. <i>functional</i> = 7 soal 2. <i>antonymous</i> <i>lopposite</i> = 4 soal
2. <i>part-whole</i> kepiting : ikan = bangau : ...	2. <i>antonymous</i> tua : muda = sakit : ...	2. <i>subordination</i> kepala sekolah : guru = manajer : ...	3. <i>categorical</i> = 8 soal 4. <i>synonymous</i> = 3 soal
3. <i>causality</i> pengetahuan : belajar = lelah : ...	3. <i>categorical</i> blus : pakaian = sandal :	3. <i>coordination</i> 4. <i>cause/effect</i>	5. <i>sequential</i> = 3 soal 6. <i>part-whole</i> = 2 soal 7. <i>causality</i> = 5 soal
	4. <i>synonymous</i> permadani : karpet = tirai : ...		8. <i>characteristics</i> = 7 soal
	5. <i>sequential</i> musim dingin : musim semi = musim semi : ...		9. <i>aritmatika</i> (deret angka) = 1 soal

Metode

Penelitian ini menggunakan data hasil tes TKD 3 yang ada pada Biro Layanan Fakultas Psikologi Unika Atma Jaya tahun 1996 hingga 2007. Terdapat 8 institusi ($N = 414$) yang menggunakan tes tersebut yaitu 5 perusahaan ($N = 192$) dan 3 sekolah menengah tingkat atas ($N = 422$). Seluruh institusi tersebut berasal dari Jakarta. Schwab (2003) menjelaskan bahwa pada data dengan $N > 80$, data yang berada pada posisi z-score dibawah -3 dan diatas 3 tergolong data ekstrim. Data ekstrim dapat menyebabkan hasil analisis terkontaminasi, sehingga distribusi menjadi tidak normal. Untuk menjamin normalitas distribusi skor, data-data yang tergolong data ekstrim tidak diikutsertakan dalam penelitian sehingga total data yang digunakan adalah sebanyak 606 data.

Pada uji validitas prediktif, dibutuhkan data kriteria untuk dikorelasikan dengan data hasil tes TKD. Namun, dari delapan institusi yang menggunakan tes ini, hanya ada satu sekolah yang memiliki data kriteria secara lengkap. Data tersebut berupa nilai rapor semester 5 dan 6 sebuah SMK swasta di Jakarta, yaitu SMK TR pada 3 jurusan, yakni akuntansi ($N = 22$), penjualan ($N = 24$), dan sekretaris ($N = 47$). Data inilah yang digunakan sebagai data kriteria pada uji validitas prediktif.

Uji validitas konstruk bertujuan untuk mengukur sejauh mana suatu alat tepat mengukur apa yang hendak diukur (Anastasi & Urbina, 1997). Metode yang digunakan pada penelitian ini adalah *item-total correlation*. Metode ini digunakan untuk menguji validitas konstruk pada level soal dan bukan pada level alat tes. Metode *item-total correlation* ini juga mengindikasikan daya diskriminasi soal, yaitu kemampuan soal dalam membedakan individu yang memiliki analogi verbal tinggi (*upper group*) dengan individu yang memiliki analogi verbal rendah (*lower group*). Apabila sebuah soal berkorelasi secara signifikan dengan skor total, hal ini menunjukkan bahwa jawaban benar pada soal tersebut berkorelasi dengan skor tinggi pada skor total, dan jawaban salah pada soal tersebut berkorelasi dengan skor rendah pada skor total.

Uji validitas prediktif digunakan untuk menguji sejauh mana alat tes mampu memprediksi keberhasilan seseorang pada suatu kriteria/performa (Anastasi & Urbina, 1997). Pada penelitian ini kriteria yang diprediksi adalah keberhasilan belajar siswa di sekolah yang direpresentasikan oleh nilai rapor. Uji reliabilitas bertujuan untuk mengukur konsistensi skor tes, yaitu apakah ketika tes diberikan pada kesempatan yang berbeda skor tes akan tetap stabil

(Anastasi & Urbina, 1997). Metode reliabilitas yang digunakan adalah metode yang digunakan untuk tes yang diujikan sebanyak 1 kali, hal ini terkait dengan data yang dimiliki oleh peneliti, yaitu hanya satu kali tes. Metode yang dipilih adalah *split-half* dengan dua teknik pembagian belahan, yaitu secara random dan secara sistematis berdasarkan urutan tingkat kesulitan dan pembelahan ganjil genap.

Selain itu dalam penelitian ini juga dianalisis tingkat kesulitan soal. Metode pengukuran tingkat kesulitan soal menggunakan pendekatan *Classical Test Theory* yaitu dengan menggunakan rumus *proporsi (p)* dan *corrected proportion (cp)* (Crocker & Algina, 1986). Rumus *p* menunjukkan proporsi jumlah orang yang menjawab benar suatu soal berbanding dengan jumlah orang yang menjawab soal tersebut. Soal dengan nilai *p* yang besar (mendekati 1) berarti memiliki tingkat kesulitan yang rendah, artinya soal tersebut tergolong mudah, di mana pada soal tersebut banyak orang yang menjawab benar. Rumus *cp* menunjukkan tingkat kesulitan yang sudah dikoreksi, yaitu dengan memperhitungkan faktor menebak pada soal-soal pilihan ganda.

Data yang digunakan merupakan data hasil tes TKD yang tersimpan di Biro Layanan Fakultas Psikologi Unika Atma Jaya. Data tersebut merupakan data yang tidak pernah digunakan untuk kepentingan apapun selain pengetesan dan penyusunan laporan hasil tes untuk klien biro. Peneliti menghubungi kepala biro untuk meminta ijin menggunakan data dan mencatat setiap respon dari tiap peserta tes. Seluruh kegiatan dilakukan di dalam ruangan biro dan dalam pengawasan kepala biro untuk menghindari kebocoran soal dan hasil tes. Sedangkan data berupa nilai rapor semester 5 dan 6 yang tersimpan di bagian administrasi SMK TR. Peneliti mencatat data tersebut setelah mendapat ijin dari kepala sekolah. Pencatatan data dilakukan di ruang administrasi SMK TR dibawah pengawasan staf administrasi SMK TR.

Hasil dan Pembahasan

Uji validitas konstruk pada level soal menunjukkan koefisien korelasi antara skor soal dengan skor total mulai dari -0.08 hingga 0.61. Dengan menggunakan standar validitas $r_{bis} \geq 0.30$, maka pada tes ini terdapat hanya 18 soal (45%) yang valid mengukur konstruk analogi verbal (lihat Tabel 2). Dari soal-soal yang tidak valid tersebut terdapat soal yang bukan tergolong verbal seperti misalnya

soal no. 19 yang mengukur hitungan aritmatika dan soal no. 27 dan no. 40 yang mengukur bidang keruangan 2 dimensi dan 3 dimensi.

Tabel 2.
Hasil Uji Item-Total Correlation

No. soal	r_{bis}	No. soal	r_{bis}
1.	0.26	21.	0.48
2.	0.02	22.	0.19
3.	0.29	23.	0.40
4.	0.51	24.	0.61
5.	0.25	25.	0.41
6.	0.25	26.	0.40
7.	0.05	27.	0.22
8.	-0.08	28.	0.13
9.	0.29	29.	0.21
10.	0.33	30.	0.47
11.	0.10	31.	0.50
12.	0.20	32.	0.30
13.	0.10	33.	0.53
14.	0.01	34.	0.18
15.	0.42	35.	0.41
16.	0.38	36.	0.33
17.	0.32	37.	0.39
18.	0.24	38.	0.45
19.	0.05	39.	0.47
20.	0.16	40.	0.23

Soal no. 8 menunjukkan korelasi yang negatif. Dalam konteks daya diskriminasi soal korelasi tersebut menunjukkan bahwa soal tersebut gagal memenuhi fungsinya. Pada soal tersebut justru individu dengan kemampuan rendah berhasil menjawab soal dengan benar.

Dengan menggunakan standar tingkat kesulitan soal dari Hayat, Pranata, dan Suparnanto (1992), hasil perhitungan tingkat kesulitan soal menunjukkan bahwa soal-soal TKD cenderung mudah. Hal ini dapat dilihat dari nilai rata-rata cp yaitu 0.73 dan proporsi jumlah soal mudah sedang dan sulit. Soal yang tergolong mudah sebanyak 26 soal (65%), soal yang tergolong sedang sebanyak

10 soal (25%) dan soal tergolong sulit sebanyak 4 soal (10%) (lihat Tabel 3).

Tabel 3.
Hasil Perhitungan Tingkat Kesulitan Soal

No. Soal	Nilai cp	Kategori	No. Soal	Nilai cp	Kategori
1.	0.90	M	21.	0.81	M
2.	0.97	M	22.	0.78	M
3.	0.99	M	23.	0.98	M
4.	0.95	M	24.	0.98	M
5.	0.95	M	25.	0.94	M
6.	0.82	M	26.	0.43	Sd
7.	0.76	M	27.	0.78	M
8.	0.84	M	28.	0.74	M
9.	0.93	M	29.	0.90	M
10.	0.90	M	30.	0.66	Sd
11.	0.55	Sd	31.	0.38	Sl
12.	0.78	M	32.	0.48	Sd
13.	0.78	M	33.	0.92	M
14.	0.82	M	34.	0.38	Sl
15.	0.93	M	35.	0.90	M
16.	0.81	M	36.	0.57	Sd
17.	0.58	Sd	37.	0.50	Sd
18.	0.50	Sd	38.	0.48	Sd
19.	0.69	Sd	39.	-0.05	Sl
20.	0.82	M	40.	0.26	Sl

Catatan: M = Mudah; Sd = Sedang; Sl = Sulit

Distribusi tersebut kurang baik karena dengan demikian soal-soal kurang dapat melakukan fungsinya untuk mendiskriminasi individu-individu dengan kemampuan tinggi dan rendah. Tingkat kesulitan yang ideal untuk tes berjenis *power-test* adalah tersebar mulai dari mudah hingga sulit dengan rata-rata tingkat kesulitan 0.5. Nilai rata-rata p atau cp sebesar 0.5 menunjukkan bahwa alat tes mendiskriminasi kelompok *upper* dan *lower* dalam tingkat maksimal, yaitu 50% kelompok *upper* dan 50% kelompok *lower* (Anastasi & Urbina, 1997). Dalam penelitian ini uji validitas prediktif bertujuan menguji apakah skor TKD 3 dapat memprediksi keberhasilan siswa SMK pada tiga jurusan, yakni akuntansi, penjualan, dan sekretaris.

Hasil uji korelasi skor tes dengan nilai rapor semester 5 dan 6 pada masing-masing jurusan ditampilkan pada Table 4 di bawah ini.

Tabel 4.
Hasil Uji Korelasi Antara Skor TKD dengan Nilai Rapor di Semester 5 pada Jurusan Akuntansi, Penjualan dan Sekretaris

Data statistik	Jurusan Akuntansi	Jurusan Penjualan	Jurusan Sekretaris
Nilai rapor semester 5			
<i>N</i>	22	24	48
<i>M</i>	6.38	6.48	6.58
<i>SD</i>	0.83	0.61	0.53
<i>r</i>	0.42*	0.25	0.51*
<i>r_{critical}</i>	0.36	0.34	0.24
<i>p</i>	< 0.05	> 0.05	< 0.05
Nilai rapor Semester 6			
<i>N</i>	22	24	48
<i>M</i>	6.57	7.02	6.71
<i>SD</i>	0.90	0.65	0.36
<i>r</i>	0.45*	0.24	0.45*
<i>r_{critical}</i>	0.36	0.34	0.24
<i>p</i>	< 0.05	> 0.05	< 0.05

Catatan: * = signifikan

Pada Tabel 4 terlihat bahwa skor TKD 3 berkorelasi secara signifikan dengan nilai rapor pada jurusan akuntansi dan sekretaris semester 5 ($r_{akun} = 0.42, p < 0.05; r_{sekre} = 0.51, p < 0.05$) dan semester 6 ($r_{akun} = 0.45, p < 0.05; r_{sekre} = 0.45, p < 0.05$) dan tidak berkorelasi dengan nilai rapor pada jurusan penjualan. Hal ini menunjukkan bahwa analogi verbal valid memprediksi keberhasilan siswa SMK TR di jurusan akuntansi dan sekretaris, dan tidak valid memprediksi keberhasilan siswa di jurusan penjualan.

Analisis validitas prediksi ini dilanjutkan dengan menguji perbedaan korelasi antara skor TKD 3 dengan keberhasilan akademik di masing-masing jurusan. Analisis ini dilakukan untuk melihat apakah hubungan yang terjadi antara TKD 3 dengan keberhasilan akademik berbeda pada masing-masing jurusan. Kegiatan analisis tersebut dilakukan dengan menggunakan metode *statistic differences between coefficient of correlation when coefficient correlation are not correlated* (Guilford & Fruchter, 1981). Hasil

pengujian ini menunjukkan tidak adanya perbedaan korelasi skor analogi verbal dengan keberhasilan akademik di masing-masing jurusan. Hal ini ditunjukkan oleh nilai-nilai selisih koefisien korelasi antar jurusan yang lebih kecil dari $Z_{critical}$ ($Z_{critical} = 1.96; LOS = 0.05$) (lihat Tabel 5). Hal ini menunjukkan TKD 3 tidak dapat melakukan diferensiasi keberhasilan akademik siswa pada 3 jurusan tersebut. Idealnya, dapat ditentukan korelasi mana yang paling kuat atau lemah dari korelasi-korelasi tersebut, sehingga dapat ditentukan TKD 3 dapat dijadikan sebagai instrument seleksi siswa untuk jurusan tertentu.

Tabel 5.
Perbandingan Korelasi Antar Jurusan

Selisih korelasi antar jurusan	Semester 5	Semester 6
Z akuntansi – penjualan	0.61	0.76
Z akuntansi – sekretaris	0.42	0.00
Z penjualan – sekretaris	1.17	0.91

Pada pengujian reliabilitas dengan metode *split-half*, peneliti membagi soal-soal ke dalam dua belahan dengan dua teknik, yakni pembelahan dengan cara random dan sistematis. Pada pembelahan secara random, peneliti menggunakan bantuan program SPSS untuk menghitung reliabilitas. Hasil pengujian reliabilitas dengan program SPSS menunjukkan bahwa alat tes terbagi atas belahan 1 yang terdiri dari soal no. 1 hingga no. 20 dan belahan 2 yang terdiri dari soal no. 21 hingga no. 40. Koefisien reliabilitas satu alat tes utuh yang dihitung dengan rumus *Spearman-Brown* menunjukkan hasil sebesar 0.73 (lihat Tabel 6). Berdasarkan standar reliabilitas Anastasi & Urbina (1997), hasil tersebut menunjukkan bahwa subtes TKD 3 tidak reliabel.

Pada pembelahan secara sistematis, peneliti melakukan beberapa langkah. Pertama, peneliti mengurutkan soal berdasarkan tingkat kesulitan mulai dari mudah hingga sulit. Kedua, peneliti memberi nomor urut baru pada daftar soal-soal tersebut. Ketiga,

peneliti memisahkan soal-soal ke dalam dua belahan berdasarkan urutan nomor ganjil-genap. Keempat, peneliti menghitung rata-rata skor dan standar deviasi tiap belahan untuk kemudian menguji kesetaraan kedua belahan tersebut dengan *t-test*. Hasil *t-test* menunjukkan adanya perbedaan antara belahan 1 dan belahan 2 (*t-value* = -2.79; *p* < 0.05), yang berarti kedua belahan tersebut tidak paralel. Dengan kondisi tersebut, peneliti menghitung reliabilitas satu alat tes utuh dengan rumus *Rulon* (persyaratan penggunaan rumus Spearman-Brown dan *Rulon* lihat Crocker & Algina, hal 137). Hasil perhitungan menunjukkan koefisien reliabilitas sebesar 0.89. Berdasarkan standar reliabilitas Anastasi & Urbina (1997) hasil ini menunjukkan bahwa subtes TKD 3 reliabel, yang berarti konsisten mengukur analogi verbal pada tiap belahan.

Tabel 6.

Hasil Uji Reliabilitas dengan Teknik Pembagian Belahan Random

Reliability methods		Values	
Cronbach Alpha	Part 1	Value	0.67
		N item	20 ^a
	Part 2	Value	0.86
		N item	20 ^b
	Total N of items	40	
Correlation between forms		0.58	
Spearman-Brown Coefficient	Equal length	0.73	
	Unequal length	0.73	
Guttman Split-Half Coefficient		0.66	

Catatan: a = soal 1- soal 20. b = soal 21 - soal 40.

Meskipun menggunakan jenis metode yang sama, dua pengujian reliabilitas dengan dua cara pembelahan yang berbeda menunjukkan bahwa TKD 3 sangat dipengaruhi oleh *content sampling error*, artinya ada masalah pada kesetaraan dan homogenitas soal-soal pada TKD 3. Idealnya, bila soal-soal tersebut setara dan homogen, pembelahan dengan cara apapun akan menghasilkan belahan yang paralel sehingga koefisien reliabilitas

yang dihitung dengan rumus *Rulon* maupun *Spearman-Brown* akan relatif sama (Crocker & Algina, 1986).

Berdasarkan hasil uji validitas konstruk pada level soal menunjukkan bahwa soal-soal TKD 3 banyak yang tidak valid (55%), bahkan ditemukan adanya soal-soal yang mengukur konstruk non-verbal, yaitu konstruk keruangan 2 dan 3 dimensi, dan konstruk numerik. Soal-soal yang tidak valid ini menunjukkan bahwa soal-soal tersebut tidak mampu memenuhi tujuan dasar pengukuran yaitu membedakan individu-individu (*individual differences*) yang memiliki kemampuan tinggi dan rendah. Hal ini didukung pula oleh hasil uji tingkat kesulitan soal yang menunjukkan bahwa soal tergolong mudah.

Hasil uji validitas prediktif menunjukkan bahwa skor TKD valid memprediksi keberhasilan belajar siswa SMK TR jurusan akuntansi dan sekretaris pada semester 5 dan 6 dan tidak valid memprediksi keberhasilan belajar siswa SMK TR jurusan penjualan pada semester 5 dan 6. Skor TKD 3 mampu memprediksi keberhasilan akademik siswa pada jurusan akuntansi karena pada jurusan tersebut siswa belajar bagaimana menerapkan rumus dan pola-pola hitungan pada kasus-kasus akuntansi. Kemampuan analogi sangat dibutuhkan pada jurusan tersebut. Pada jurusan sekretaris, pelajaran yang ada di dalamnya seluruhnya bersifat aktifitas praktis seperti misalnya menulis surat, menyimpan surat (*filing*), membuat agenda/jadwal rapat agar tidak bertabrakan, mengetik, dan membuat notulensi. Kegiatan-kegiatan tersebut tidak berkaitan dengan analogi, namun membutuhkan pemahaman verbal yang baik. Jadi, diperkirakan korelasi yang terjadi antara skor TKD 3 dengan keberhasilan pada jurusan sekretaris adalah karena adanya konstruk verbal yang kuat di jurusan ini. Sementara itu, pada jurusan penjualan, para siswa belajar mengenai pengenalan alat-alat elektronik dan bisnis. Para siswa belajar menetapkan harga yang pas, cara memasarkan produk, strategi pemasaran dan promosi produk. Kegiatan-kegiatan tersebut sedikit melibatkan analogi verbal, sehingga korelasi yang terjadi antara skor TKD dan keberhasilan akademik siswa pada jurusan ini tidak signifikan.

Meskipun hasil uji korelasi pada uji prediksi ini menunjukkan adanya korelasi yang signifikan pada beberapa jurusan, namun pada uji lebih lanjut yaitu uji perbedaan koefisien korelasi, terlihat tidak adanya perbedaan koefisien korelasi antara skor TKD dengan keberhasilan akademik siswa antar jurusan. Hasil ini menunjukkan alat tes tidak berfungsi sebagai tes diferensial.

Hasil pengujian reliabilitas dengan metode *split-half* menunjukkan hasil yang tidak konsisten. Pengujian dengan cara pembelahan random menunjukkan bahwa TKD 3 tidak reliabel, sementara pengujian dengan cara pembelahan sistematis menunjukkan bahwa TKD 3 reliabel. Hal ini menunjukkan bahwa reliabilitas TKD 3 sangat sensitif terhadap cara pembelahan, yang berarti soal-soal yang ada di dalamnya tidak setara dan kurang mewakili analogi verbal. Hasil uji reliabilitas ini sejalan dengan hasil uji validitas konstruk bahwa item-item tidak homogen mengukur analogi verbal, kelihatannya ada kemampuan lain yang diukur oleh soal-soal tersebut.

Kesimpulan

Dalam penelitian ini data yang digunakan adalah data yang sudah tersedia dan belum pernah dianalisis. Dengan demikian pengolahan data hanya terbatas pada pengolahan validitas konstruk pada tingkat soal dan bukan pada tingkat tes secara keseluruhan. Sehingga kesimpulan mengenai validitas konstruk hanya pada tingkat soal. Pertanyaan apakah Tes TKD 3 tepat mengukur analogi verbal masih belum bisa dijawab, karena uji validitas konstruk pada level tes idealnya dilakukan dengan cara mengkorelasikan skor total individu pada TKD 3 dengan skor total individu pada tes lain yang mengukur konstruk yang sama dan konstruk yang berbeda. Kemudian, meskipun jumlah data tersebut besar dan berasal dari kelompok sampel yang bervariasi, namun tidak semua sampel memiliki data kriteria sehingga kesimpulan mengenai validitas prediktif TKD 3 hanya berlaku pada kelompok sampel tersebut. Diharapkan pada penelitian selanjutnya dapat dilakukan pengambilan data primer yang juga melibatkan penggunaan alat tes lain dan data kriteria yang lebih lengkap, sehingga penelitian mengenai validitas TKD 3 dapat utuh. Meskipun analisis-analisis dalam penelitian ini masih belum lengkap, hasil uji-uji di atas perlu menjadi perhatian di mana soal-soal TKD 3 kurang dapat memenuhi fungsinya untuk mengukur analogi verbal dan mendiskriminasi kemampuan individu. Hasil penelitian ini menunjukkan dibutuhkan penyusunan soal-soal analogi verbal yang baru sebagai alternatif.

Daftar Pustaka

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. (7th Ed). Indiana: Prentice-Hall.

- Bandurski, M., & Galkowski, T. (2004). The development of analogical reasoning in deaf children and their parents communication mode. *Journal of Deaf Studies and Deaf Education*, 9, 153-175.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Florida: Harcourt Brace Jovanovich Publisher.
- Duran, R.P., Enright, M.K., & Peirce, L.P. (1987). *GRE verbal analogy items: Examinee reasoning on items*. New Jersey: Educational Testing Service. Diakses pada 20 Juni 2011 dari http://www.ets.org/research/policy_research_reports/rr-87-05
- Guilford, J.P., & Fruchter, B. (1981). *Fundamental statistics in psychology and education*. (6th Ed). New York: McGraw-Hill.
- Hayat, B., Pranata, S., Suparnanto, S. (1992). *Manual item and test analysis*. Jakarta: Pusat Penelitian dan Pengembangan Sistem Pengujian (PUSSIJIAN), Balitbang Dikbud.
- Nippold, M.A. (1994). *Brief research report, third-order verbal analogical reasoning: A development study of children and adolescents*. University of Oregon: Academic Press.
- Scwab, J. (2003). Detecting outliers. *Data analysis and computer II: Modul of learning*. Diakses pada 20 Juni 2011 dari www.utexas.edu/courses/schwab/sw388r7/.../DetectingOutliers.ppt
- Sternberg, R.J. (1994). *Encyclopedia of human intelligence*. New York: Macmillan Publishing Company.
- Steward, I., Barnes-Holmes, D., & Roche, B. (2004). A functional-analytical model of analogy using the relational evaluation procedure. *The Psychological Record*, 54, 531-552.
- Wibowo, S. (1976). *Penyusunan tes kemampuan diferensial sebagai tes untuk seleksi calon mahasiswa*. Jakarta: Universitas Indonesia.