

THE QUALITY OF THE *TES POTENSI AKADEMIK (TPA) 07A*

Saijuddin Azwar

Abstract

Characteristics of the *Tes Potensi Akademik (TPA)* version 07A items and subtests were examined based on the responses of 198 applicants to the *Magister Program Psikologi* UGM 2007 while the predictive validity of the test was analyzed based on the responses of 80 students admitted to the program in 2007. Findings suggested that the items in the Verbal, Quantitative and Reasoning subtests had the medium level of difficulty, satisfying discriminating index, and good distracters. All subtest scores and the TPA test score had very high reliability coefficients. However, findings showed that the analysis of the validity of TPA scores for predicting the first semester Grade Point Average (GPA) was inconclusive.

Key words: *item analysis, predictive validity*

KUALITAS TES POTENSI AKADEMIK VERSI 07A

Saijuddin Azwar

Abstrak

Karakteristik butir dan subtes dari Tes Potensi Akademik (TPA) versi 07A diteliti berdasarkan data jawaban 198 pelamar Program Magister Psikologi UGM tahun 2007, sementara validitas prediksi tes dianalisis berdasarkan data jawaban 80 mahasiswa yang diterima. Hasil penelitian menunjukkan bahwa butir-butir subtes verbal, kuantitatif, dan penalaran memiliki tingkat kesulitan sedang, indeks daya beda sangat baik, dan distraktor berfungsi dengan baik. Semua skor subtes memiliki koefisien reliabilitas yang sangat tinggi, demikian juga skor TPA. Namun, analisis validitas skor TPA kurang dapat memprediksi IPK semester pertama.

Kata kunci: TPA, analisis butir, validitas prediksi

Pendahuluan

Secara tradisional, tes psikologi dikelompokkan menjadi dua macam menurut tujuan ukurnya. Pertama adalah tes yang mengukur aspek kemampuan atau abilitas kognitif yang dalam istilah Cronbach disebut performansi maksimal dan yang ke dua adalah tes yang mengukur aspek bukan kemampuan yang dalam istilah Cronbach disebut sebagai performansi tipikal (Cronbach, 1970).

Tes potensi merupakan salah-satu bentuk pengukuran terhadap kemampuan abilitas kognitif potensial umum (pengukuran performansi maksimal) yang dirancang khusus guna memprediksi peluang keberhasilan belajar di perguruan tinggi, karena itulah tes seperti ini biasanya dinamai Tes Potensi Akademik. Gagasan dasar dalam konstruksi Tes Potensi Akademik sedikit-banyak mengikuti konsep pengembangan *Graduate Record Examinations* (GRE) yang terdiri atas seksi *Verbal Reasoning* (V), *Quantitative Reasoning* (Q), dan *Analytical Writing* (AW) (GRE Bulletin, 2008), dengan beberapa perubahan. Pada umumnya, Tes Potensi Akademik di Indonesia terdiri atas tiga subtes yaitu subtes Verbal, subtes Kuantitatif, dan subtes Penalaran.

Berbeda dari isi tes prestasi yang disusun berdasar silabus mata pelajaran pada suatu jenjang pendidikan atau pelatihan yang lebih merupakan pengungkapan hasil pembelajaran, Tes Potensi Akademik tidak disusun berdasar silabus mata pelajaran dan karenanya keberhasilan menjawab soal dalam tes ini adalah minimal kaitannya dengan penguasaan isi pelajaran tertentu. Hal itu disebabkan konten soal-soal dalam tes potensi dikembangkan sedemikian rupa sehingga peluang keberhasilan untuk menjawab dengan benar lebih tergantung pada penggunaan daya penalaran (*reasoning*) baik logis (*logical*) maupun analitis (*analytical*). Sebagai contoh, soal-soal Geometrika dalam Tes Potensi Akademik dapat dijawab tanpa mengandalkan penguasaan rumus-rumus geometrika yang rumit. Soal Aritmetika dalam Tes Potensi Akademik juga tidak memerlukan penggunaan rumus matematika namun lebih mengandalkan pada penalaran dan strategi pemecahan masalah kuantitatif yang bersifat umum sedangkan

soal Konsep Ajabar mengungkap pemahaman akan konsep-konsep dasar aljabar bukan kemahiran dalam menggunakan rumus-rumus komputasinya.

Dalam Tabel 1 dimuat perbandingan karakteristik antara tes potensi dan tes prestasi belajar. Sekalipun secara kontrak keterkaitan isi tes potensi dengan kurikulum dan silabus mata pelajaran adalah minimal, namun asumsi bahwa tes potensi mengukur kemampuan penalaran kognitif umum yang diperlukan para mahasiswa dalam belajar di jenjang pendidikan tinggi menjadi dasar pemikiran bahwa tes ini dapat digunakan sebagai salah-satu pertimbangan dalam seleksi masuk dan karena itu skor tes potensi diharapkan berkorelasi tinggi dengan indikator-indikator keberhasilan belajar di perguruan tinggi.

Tabel 1. Perbandingan Karakteristik Tes Potensi dan Tes Hasil Belajar

Tes Potensi	Tes Hasil Belajar
1. Dirancang untuk mengungkap kemampuan kognitif potensial.	1. Dirancang untuk mengukur hasil pembelajaran.
2. Disusun berdasar konsep abilitas dasar yang hendak diukur.	2. Disusun berdasar domain materi pembelajaran pada bidang tertentu.
3. Keterkaitan minimal dengan silabus/kurikulum.	3. Mengacu pada isi silabus.
4. Skor tes = <i>probability of future success</i> .	4. Skor tes = <i>indication of present success</i> .
5. Skor tinggi diperoleh berdasar strategi umum penyelesaian masalah.	5. Skor tinggi diperoleh berdasar penguasaan bahan ajar.
6. Penekanan pada validitas prediktif.	6. Penekanan pada validitas isi.

Berkaitan dengan penggunaan Tes Potensi Akademik untuk tujuan seleksi, aspek validitas (khususnya validitas prediktif) menjadi penting demi akurasi prediksi sedangkan masalah bebas bias menjadi penting untuk tercapainya *fairness* dalam keputusan seleksi tersebut. Kedua isu tersebut penting untuk diperhatikan sebagaimana dikatakan oleh para ahli bahwa untuk berfungsi secara efektif tes haruslah memiliki minimal tiga kualitas yaitu reliabel, valid, dan *unbiased* (Zucker, 2003).

Dari beberapa versi tes potensi yang telah disusun, salah satunya adalah Tes Potensi Akademik versi 07A (Azwar, 2007) yang diperuntukkan bagi calon mahasiswa pascasarjana. Sebagaimana umumnya Tes Potensi Akademik, TPA 07A terdiri atas tiga subtes; yaitu subtes Verbal (V), subtes Kuantitatif (K), dan subtes Penalaran (P); dan merupakan penyempurnaan dari versi TPA sebelumnya dengan penambahan komponen Penalaran Simbol dalam subtes Penalaran. Masing-masing subtes berisi beberapa komponen tes yang rinciannya dapat dilihat pada Tabel 2.

Tabel 2. Subtes dan Komponen TPA 07A

Subtes	Komponen	Banyaknya Item	Waktu Pengerjaan
Verbal	Padanan Kata	15	7 menit
	Lawan Kata	15	7 menit
	Analogi Kata	10	6 menit
	Total Verbal 40		20 menit
Kuantitatif	Deretan Angka	10	10 menit
	Aritmetika	10	10 menit
	Konsep Aljabar	10	10 menit
	Geometrika	10	10 menit
	Total Kuantitatif 40		40 menit
Penalaran	Penalaran Logis	10	10 menit
	Penalaran Diagram	10	10 menit
	Penalaran Simbol	10	10 menit
	Penalaran Analitis	10	10 menit
	Total Penalaran 40		40 menit
Total TPA 120		100 menit	

Semenjak awal diterbitkan di Tahun 2007, TPA 07A telah digunakan dalam seleksi calon mahasiswa pascasarjana di beberapa Fakultas/Jurusan Studi di Universitas Gadjah Mada namun belum pernah dilakukan analisis untuk menguji kualitas item dan konstraknya secara empiris. Penggunaan tes abilitas, terutama untuk tujuan diagnosis individual dan seleksi, tanpa melalui pengujian empiris terlebih dahulu terhadap kualitas item-itemnya merupakan praktek yang lazim dilakukan mengingat sifat kerahasiaan item dan terutama belum tersedianya data kriteria apabila diinginkan untuk melakukan validasi prediktif tes.

Hasil studi ini akan memberikan informasi empiris guna melakukan revisi terhadap aitem bilamana diperlukan dan guna meningkatkan kualitas tes yang pada gilirannya akan meningkatkan validitas hasil pengukuran dan interpretasi TPA 07A. Sebagai arah kajian kualitas, pertanyaan yang diajukan dalam penelitian ini adalah:

- a. Apakah item-item TPA 07A memiliki daya beda yang memuaskan?
- b. Bagaimanakah variasi tingkat kesukaran item TPA 07A?
- c. Apakah distraktor item TPA 07A berfungsi sebagairnana mestinya?
- d. Apakah subtes TPA 07A berfungsi sebagaimana mestinya?
- e. Apakah TPA 07A memiliki validitas prediktif yang memuaskan?
- f. Apakah skor subtes TPA 07A memiliki reliabilitas yang tinggi?

Penelitian ini tidak melaporkan secara terperinci analisis mendalam terhadap isi dan redaksional pilihan jawaban item demi item dalam TPA 07A dikarenakan sifat kerahasiaan isi tes. Laporan semacam itu dijadikan bahan acuan internal dalam revisi dan peningkatan kualitas TPA 07A hingga diterbitkannya edisi revisi yang akan datang.

Metode Penelitian

1. Data

Data kajian berupa skor item dan skor tes TPA 07A dan 198 orang calon mahasiswa Program Magister Profesi Psikolog Fakultas Psikologi UGM Angkatan Tahun 2007. Data indeks prestasi belajar (IP) semester pertama diperoleh dari 80 orang mahasiswa Angkatan 2007.

2. Variabel

Variabel dalam analisis item adalah skor item TPA 07A pada masing-masing ketiga subtes, skor subtes, dan skor total atau skor TPA. Skor item merupakan skor dikotomi, yaitu 1 untuk jawaban yang benar dan 0 untuk jawaban yang salah. Skor subtes adalah penjumlahan skor item dan skor TPA adalah penjumlahan skor subtes. Kesemua skor tersebut dianalisis sebagai skor mentah (*raw scores*) tanpa derivasi apapun. Dalam analisis validitas prediktif, skor TPA menjadi variabel prediktor terhadap prestasi belajar.

Prestasi belajar sebagai variabel kriteria validasi TPA dioperasionalkan sebagai indeks prestasi belajar semester pertama (IP-1) yang diperoleh dari nilai tiga mata kuliah yang merupakan mata kuliah kemagisteran pokok dalam Program Magister Profesi Psikolog yaitu Metode Penelitian Kualitatif (MPK), Eksperimen Kuasi (EK), Analisis Multivariat (AM), dan Penyusunan Alat Ukur (PAU). Nilai mata kuliah Filsafat Manusia tidak diikutsertakan dalam analisis validasi karena merupakan mata kuliah yang bersifat umum, bukan merupakan kompetensi spesifik Program Magister Profesi. Sebagai kriteria validasi, IP Semester dihitung dengan formula:

$$IP = \sum (\text{sks} \times \text{nilai}) / \sum \text{sks}$$

3. Analisis

Parameter daya beda item ditunjukkan oleh koefisien *r*-point biserial (r_{pbis}) antara skor item dengan skor subtesnya masing-masing. Dengan adanya 40 item di setiap subtes, tidak diperlukan koreksi terhadap *spurious overlap* (Guilford, 1953; Guilford, 1956; Wolf, 1967; Azwar, 1997). Daya beda yang ideal diperoleh bila r_{pbis} mendekati angka 1,00 (Scorepak, 2005). Dalam analisis ini daya beda dianggap memuaskan bila mencapai angka 0,25. Angka ini lebih tinggi dibanding rekomendasi Thorndike sebesar 0,20 (Thorndike et. al., 1991) dan rekomendasi ahli lain (Crocker & Algina, 1986) dan masih jauh lebih tinggi daripada yang disarankan oleh Kehoe yaitu 0,15 (Kehoe, 1997).

Parameter tingkat kesukaran item (p) ditampakkan oleh proporsi subjek yang menjawab item dengan benar. Nilai p yang berada di kisaran 0,30 sampai dengan 0,70 merupakan indikasi tingkat kesukaran item yang sedang (Scorepak, 2005).

Efektivitas distraktor ditampakkan oleh koefisien r -point biserial bagi masing-masing distraktor. Suatu distraktor yang efektif adalah yang memiliki koefisien r_{pbis} negatif. Semakin besar harga negatif r_{pbis} menunjukkan bahwa fungsi distraktor semakin efektif sedangkan r_{pbis} yang berada di sekitar nol berarti distraktor tidak berfungsi sebagaimana mestinya. Pada item-item yang sulit, yaitu yang persentase subjek menjawab benar sangat kecil, interpretasi efektivitas distraktor tidak dapat semata-mata disandarkan pada angka statistik namun harus disertai dengan pertimbangan mengenai distribusi peluang subjek yang menjawab salah pada item yang bersangkutan.

Konkurensi tujuan ukur subtes dan kesamaan potensi dasar yang diukur oleh ketiga subtes diuji dengan analisis korelasional. Koefisien korelasi yang tinggi antarsubtes merupakan indikasi bahwa masing-masing subtes memang mengungkap aspek dasar yang sama sekalipun diungkap melalui komponen performansi yang berbeda, sedangkan koefisien korelasi yang tinggi di antara masing-masing subtes dan skor TPA merupakan indikasi bahwa tujuan ukur subtes konsisten dengan tujuan ukur TPA. Validitas TPA untuk memprediksi keberhasilan belajar diketahui dengan komputasi koefisien korelasi linier antara skor TPA dengan IP semester pertama.

Hasil Penelitian

1. Parameter Item

Hasil analisis item meliputi statistik tingkat kesukaran item (p), daya beda item (r_{pbis}), dan efektivitas distraktor. Untuk masing-masing komponen dalam setiap subtes, indeks p telah diurutkan dari item yang paling mudah sampai dengan item yang paling sukar. Pada angka r_{pbis} diberikan tanda * bila koefisien korelasinya kurang dari 0,250. Mengingat ukuran sampel yang tidak begitu besar ($n = 198$) maka analisis efektivitas

distraktor hanya dilakukan pada item yang tingkat kesukarannya berada di sekitar taraf sedang ($0,20 < p < 0,80$). Distraktor dianggap efektif bila berharga negatif atau lebih kecil daripada 0,10.

a. Subtes Verbal

Tabel 3. Hasil Analisis Item untuk Subtes Verbal

Komponen	No. Item	p	r _{pbis}	dte
Padanan Kata	14	0,848	0,808	
	15	0,783	0,762	-
	11	0,652	0,579	-
	8	0,621	0,504	-
	9	0,616	0,582	b, c
	4	0,591	0,541	-
	12	0,485	0,480	-
	6	0,429	0,339	e
	10	0,283	0,358	e
	2	0,187	0,315	
	13	0,167	0,256	
	7	0,162	0,196*	
	5	0,116	0,150*	
	1	0,101	0,212*	
	3	0,101	0,241	
Lawan Kata	18	0,808	0,755	
	16	0,788	0,747	a

	22	0,788	0,728	-
	21	0,707	0,652	b
	28	0,682	0,614	-
	23	0,707	0,604	-
	30	0,530	0,562	-
	24	0,667	0,537	c, e
	17	0,606	0,532	e
	27	0,515	0,516	a
	19	0,591	0,458	e
	25	0,419	0,451	a, c
	26	0,485	0,419	b, e
	20	0,167	0,247*	
	29	0,152	0,098	
Analogi Kata	36	0,742	0,640	-
	40	0,737	0,651	-
	33	0,712	0,658	-
	37	0,621	0,517	-
	35	0,596	0,532	a
	38	0,571	0,588	-
	34	0,556	0,540	-
	31	0,535	0,473	b
	32	0,247	0,275	a, c
	39	0,025	0,133*	

Keterangan: * = item dengan daya beda tidak memuaskan
 dte = distraktor tidak efektif

Hasil analisis item untuk subtes Verbal disajikan pada Tabel 3. Dengan menggunakan batas penerimaan daya beda item $r_{pbis} \geq 0,250$ tampak jelas dalam Tabel 3, bahwa item-item yang terlalu sukar cenderung tidak memiliki daya beda yang memuaskan. Bahkan tampak bahwa naiknya tingkat kesukaran item hampir selalu diikuti oleh turunnya indeks daya beda item.

Diperoleh pula rata-rata indeks kesukaran item 0,502 dan rata-rata indeks daya beda sebesar 0,481. Dengan menyertakan semua item subtes Verbal yang berjumlah 20, diperoleh koefisien reliabilitas subtes Verbal sebesar $\alpha = 0,919$ dengan eror standar dalam pengukuran sebesar $S_e = 2,458$. Koefisien reliabilitas ini akan meningkat bila item-item yang daya bedanya tidak cukup memuaskan tidak diikutsertakan dalam penghitungan.

b. Subtes Kuantitatif

Pada Tabel 4 terlihat dari hasil analisis item bahwa hanya ada dua item dalam subtes Kuantitatif yang daya bedanya tidak memuaskan, masing-masing terdapat pada komponen Aritmetika dan komponen Geometrika.

Sebagaimana juga kecenderungan yang terjadi pada subtes Verbal, kedua item yang daya bedanya tidak memuaskan tersebut adalah item-item yang tingkat kesukarannya tinggi, yaitu kurang dari 0,200 sedangkan pada item-item yang tingkat kesukarannya sedang atau rendah pada umumnya diperoleh daya beda yang tinggi. Secara keseluruhan, rata-rata tingkat kesukaran item untuk subtes Kuantitatif adalah 0,495 dan rata-rata indeks daya beda adalah 0,503.

Tabel 4. Hasil Analisis Item untuk Subtes Kuantitatif

Komponen	No. Item	p	r_{pbis}	dte
Deretan Angkat	49	0,872	0,706	
	45	0,867	0,765	
	41	0,810	0,657	
	42	0,810	0,702	
	47	0,810	0,619	
	44	0,785	0,598	-
	43	0,759	0,608	a
	50	0,738	0,580	-
	46	0,728	0,549	-
	48	0,277	0,421	e
Aritmetika	54	0,744	0,584	-
	55	0,718	0,579	-
	56	0,697	0,519	-
	53	0,615	0,657	-
	51	0,574	0,641	-
	58	0,574	0,552	-

	60	0,497	0,559	-
	57	0,487	0,548	-
	52	0,210	0,332	a
	59	0,123	0,249*	
Konsep Aljabar	65	0,595	0,547	-
	67	0,477	0,513	-
	68	0,441	0,579	-
	64	0,405	0,440	a
	62	0,379	0,429	-
	69	0,359	0,496	-
	63	0,282	0,294	e
	61	0,277	0,315	b
	66	0,256	0,359	a, c
	70	0,226	0,428	b
	Geometrika	76	0,503	0,493
75		0,497	0,529	-
73		0,467	0,537	-
72		0,451	0,520	-
78		0,364	0,422	b
77		0,318	0,480	b
79		0,308	0,533	-
80		0,241	0,323	-
74		0,179	0,327	
71		0,082	0,137*	

Keterangan: * = item dengan daya beda tidak memuaskan
 dte = distraktor tidak efektif

Koefisien reliabilitas subtes Kuantitatif adalah $\alpha = 0,924$ dengan eror standar dalam pengukuran sebesar $S_e = 2,478$. Koefisien reliabilitas ini sedikit lebih tinggi dibanding koefisien reliabilitas subtes Verbal namun memiliki S_e yang relatif sama.

c. Subtes Penalaran

Tabel 5. Hasil Analisis Item untuk Subtes Penalaran

Komponen	No. Item	p	r_{pbis}	dte
Penalaran Logis	89	0,831	0,746	
	83	0,713	0,615	-
	84	0,703	0,585	-
	90	0,692	0,651	-
	86	0,569	0,590	c
	85	0,559	0,430	a, b
	87	0,549	0,478	-
	82	0,400	0,424	c
	88	0,215	0,297	e
	81	0,015	-0,014*	

Penalaran Diagram	94	0,549	0,519	-
	91	0,518	0,532	-
	96	0,415	0,491	e
	95	0,354	0,404	a
	100	0,344	0,467	b
	99	0,308	0,278	c, d
	92	0,303	0,388	a
	97	0,256	0,446	a
	93	0,241	0,329	a
	98	0,179	0,204*	
Penalaran Simbol	103	0,882	0,792	
	101	0,856	0,720	
	104	0,821	0,685	
	108	0,815	0,673	
	102	0,759	0,655	d
	105	0,667	0,587	-
	109	0,600	0,536	-
	107	0,508	0,475	-
	106	0,287	0,297	a
	110	0,231	0,384	a
Penalaran Analitis	112	0,779	0,641	-
	114	0,733	0,577	d
	119	0,615	0,470	-
	115	0,400	0,219*	b, c, d
	116	0,349	0,390	b, e
	111	0,333	0,292	e
	118	0,318	0,479	c
	113	0,241	0,282	a
	117	0,210	0,255	a, b
	120	0,108	0,066	

Keterangan: * = item dengan daya beda tidak memuaskan
 dte = distraktor tidak efektif

Hasil analisis item dalam subtes Penalaran disajikan pada Tabel 5. Sebagaimana halnya untuk subtes Verbal dan subtes Kuantitatif, pada subtes Penalaranpun tampak bahwa item-item yang lebih sulit cenderung memiliki daya beda yang lebih rendah, kecuali kasus item nomor 115. Rata-rata indeks kesukaran item untuk subtes Penalaran adalah 0,481 dan rata-rata indeks daya bedanya adalah 0,458.

Koefisien reliabilitas subtes Penalaran adalah $\alpha = 0,904$ dengan eror standar dalam pengukuran sebesar $S_e = 2,516$. Koefisien reliabilitas ini adalah yang terendah di antara ketiga subtes dalam TPA 07A.

2. Interkorelasi antar Subtes

Ketiga subtes TPA mengukur komponen performansi yang berbeda namun mengungkap abilitas dasar yang sama yaitu penalaran. Oleh karena itu, diharapkan antara satu subtes dengan yang lainnya saling berkorelasi. Tabel 6 memuat hasil komputasi interkorelasi antar subtes.

Tabel 6. Interkorelasi Subtes TPA 07A (n = 174)

	K	P
V	0,494	0,419
K		0,651

Sebagaimana diharapkan, terdapat korelasi yang signifikan di antara ketiga subtes dengan korelasi tertinggi adalah antara subtes Kuantitatif dengan subtes Penalaran.

3. Korelasi Subtes-total

Sebagai bagian dari skor keseluruhan TPA, skor masing-masing ketiga subtes diharapkan berkorelasi tinggi dengan skor total. Untuk itu, dihitung korelasi setiap subtes dengan skor TPA yang telah dikoreksi terhadap *spurious overlap*. Dari hasil komputasi untuk subtes Verbal dihasilkan koefisien korelasi $r_{vt} = 0,506$ ($r_{vt}^2 = 0,256$); untuk subtes Kuantitatif diperoleh koefisien korelasi $r_{kt} = 0,686$ ($r_{kt}^2 = 0,471$); dan untuk subtes Penalaran diperoleh koefisien korelasi $r_{pt} = 0,640$ ($r_{pt}^2 = 0,410$).

Tingginya koefisien korelasi tersebut mengindikasikan tingginya konkordansi tujuan ukur antara masing-masing subtes dengan ukuran potensi yang diperlihatkan oleh skor TPA.

4. Validitas Prediktif

Skor prediktor TPA adalah skor dari kelompok subjek mahasiswa yang lolos seleksi dan merupakan kelompok yang lebih terbatas (*restricted sample*). Karena subjek terpilih berdasarkan, antara lain, tingginya skor TPA maka kelompok ini memiliki statistik deskriptif yang berbeda dari seluruh

calon mahasiswa semula. Tabel 7 memuat perbandingan deskriptif skor TPA dari kedua kelompok tersebut.

Tabel 7. Deskriptif Skor TPA

Statistik	Calon Mahasiswa	Mahasiswa
N	174	80
Mean	67,351	74,587
Median	69,000	74,500
Modus	72,000	72,000
Deviasi Std.	12,912	9,131
Varians	166,715	83,385
Sebaran	59,000	47,000
Minimum	38,000	50,000
Maximum	97,000	97,000

Kriteria validasi prediktif skor TPA dan subtesnya adalah IP Tahun pertama (IP-1) yang dihitung dari hasil ujian mata kuliah Metode Penelitian Kuaitatif (MPK), Eksperimen Kuasi (EK), Analisis Multivariat (AM), dan Penyusunan Aat Ukur (PAU). Tabel 8 memuat deskriptif skor masing-masing mata kuliah dan skor IP-1.

Tabel 8. Deskriptif Skor Kriterion

	Nilai Mata Pelajaran				IP-1
	MPK	EK	AM	PAU	
N	80	80	79	80	80
Mean	3,612	3,362	3,253	3,391	3,404
Median	3,500	3,500	3,000	3,500	3,375
Modus	3,500	3,000	3,000	3,000	3,250
Deviasi Std.	0,347	0,497	0,629	0,358	0,272
Sebaran	1,000	2,000	3,000	1,000	1,250
Minimum	3,000	2,000	1,000	3,000	2,750
Maximum	4,000	4,000	4,000	4,000	4,000

Persentil 25	3,500	3,000	3,000	3,000	3,250
Persentil 75	4,000	3,500	4,000	3,500	3,625

Dalam deskriptif pada Tabel 8 diperlihatkan dengan jelas bahwa skor IP-1 yang hendak dijadikan kriteria validasi adalah sangat homogen. Dengan modus IP sebesar 3,25 dan hanya 25% dan subjek memiliki skor IP yang lebih kecil daripada 3,25 maka distribusi skor IP sangat tidak cukup variatif sebagai kriteria. Kondisi ini akan mengakibatkan tidak diperolehnya estimasi validitas prediktif yang benar karena komputasi validitas akan menghasilkan underestimasi (Allen & Yen, 1979; Hartigan & Wigdor, 1989; Kaufman, 1972). Namun demikian, dalam laporan ini komputasi koefisien validitas tetap dilakukan sebagai informasi dan data yang ada.

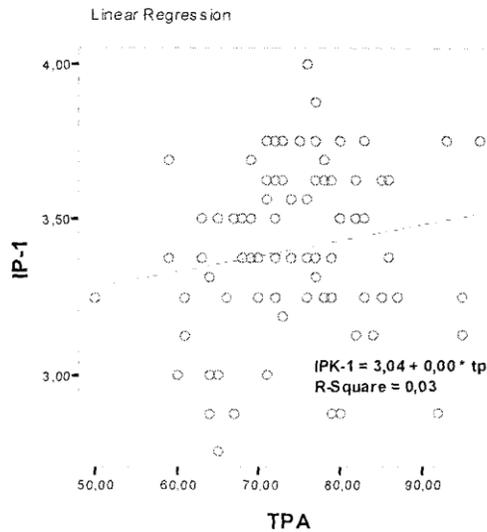
Korelasi antara skor TPA dan masing-masing subtesnya dengan skor IP-1 dimuat dalam Tabel 9, yang menampakkan bahwa semua koefisien korelasi tersebut terbilang sangat rendah. Dengan mengingat efek restriksi sebaran pada sampel dan variasi skor IP-1 sebagai kriteria yang sangat homogen, kuat dugaan bahwa angka koefisien yang diperoleh merupakan underestimasi dan tidak mencerminkan validitas prediktif yang sebenarnya.

Tabel 9. Korelasi Skor TPA dan Skor Subtes dengan IP-1 (n = 80)

	Subtes			TPA
	V	K	P	
IP-1	0,214	0,059	0,130	0,166

Korelasi antara IP-1 dengan skor TPA, yang merupakan koefisien validitas prediktif TPA, dalam kasus ini hanya sebesar 0,166. Dengan koreksi terhadap atenuasi akibat ketidaksempurnaan reliabilitas skor TPA diperoleh koefisien validitas prediktif terkoreksi:

$$r_{Tv} = 0,166/\sqrt{0,876} = 0,177.$$



Gambar 1. Plot Skater IP- 1 dan Skor TPA

Peningkatan koefisien validitas dari 0,166 ke 0,177 ini dapat dikatakan tidak bermakna, yaitu tidak mengubah interpretasi bahwa skor TPA tidak memperlihatkan keterkaitan dengan skor IP-1. Rendahnya koefisien validitas tersebut agaknya terjadi bukan dikarenakan problem pada prediktornya melainkan bersumber dari masalah reliabilitas skor kriteria yang patut dipertanyakan dan variabilitas skor prediktor dan skor kriteria yang sama-sama sangat homogen.

Dengan menggunakan skor TPA yang merupakan penjumlahan dari skor ketiga subtesnya, diperoleh koefisien reliabilitas TPA (120 item) sebesar $\alpha = 0,876$ dengan eror standar dalam pengukuran sebesar $s_e = 4,505$.

Kesimpulan dan Rekomendasi

1. Kesimpulan

- a. Item-item dalam subtes Verbal, subtes Kuantitatif, dan subtes Penalaran masing-masing secara umum memiliki tingkat kesukaran yang sedang, daya beda yang memuaskan, dan distraktor yang berfungsi baik meskipun terdapat beberapa item masih belum berfungsi sebagaimana mestinya.
- b. Skor subtes Verbal, subtes Kuantitatif, subtes Penalaran, dan skor TPA 07A memiliki reliabilitas yang sangat tinggi.
- c. Subtes Verbal, subtes Kuantitatif, dan subtes Penalaran berhasil mengungkap abilitas dasar yang sama melalui pengukuran performansi yang berbeda dan secara konvergen mendukung skor TPA.
- d. Skor ketiga subtes dan skor TPA memiliki reliabilitas yang sangat memuaskan.
- e. Memperhatikan sebaran skor subjek, waktu pengerjaan yang diberikan bagi setiap subtes terbukti layak (*appropriate*).
- f. Validitas prediktif TPA 07A untuk memprediksi keberhasilan belajar belum konklusif.

2. Rekomendasi

- a. Penggantian dan revisi beberapa item dan distraktor perlu dilakukan.
- b. Urutan nomor item perlu diatur ulang dengan mempertimbangkan tingkat kesukaran empirisnya.
- c. Soal yang tingkat kesukarannya terlalu tinggi tidak disertakan dalam TPA dan diganti dengan item yang memiliki tingkat kesukaran moderat.

Daftar Pustaka

- Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey: Brooks/Cole Publishing Company.
- Azwar, S. (1997). *Reliabilitas dan validitas* (edisi ke 3). Yogyakarta: Pustaka Pelajar.
- _____. (2007). *Tes Potensi Akademik - Versi 07A*. Yogyakarta: Sigma Alpha Test Specialist.

- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Forth Worth: Holt, Rinehart, and Winston, INC.
- Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed.). New York, NY: Harper and Row.
- GRE-bulletin. (1980). *GRE General Tests, 1980*. http://www.ets.org/MediaTests/GRE/pdf/0708_gre_bulletin.pdf
- Guilford, J. P. (1953). The correlation of an item with a composite of the remaining items in a test. *Educational and Psychological Measurement*, 13, 1, 87-93.
- _____.(1956). *Fundamental statistics in psychology and education* (3rd ed.). New York: McGraw-Hill.
- Hartigan, J. A., & Wigdor, A. K. (Eds.) (1989). *Fairness in employment testing: Validity generalization, minority issues, and the general aptitude test battery*. Washington, DC: National Academy Press.
- Kehoe, J. (1997). Basic item analysis for multiple-choice tests. *ERIC Digest*. <http://www.ericdigests.org/1997-1/basic.html>
- Kaufman, A. S. (1972). Restriction of range: Questions and answers. *Test Service Bulletin*. The Psychological Corporation, 59.
- Scorepak. (2005). *Item analysis*. Office of Educational Assessment, University of Washington. [http://www.washington.edu/oea/pdfs/resources/item analysis.pdf](http://www.washington.edu/oea/pdfs/resources/item%20analysis.pdf).
- Thorndike, R. M., Cunningham, G.K., Thorndike, R.L., & Hagen, E. P. (1991). *Measurement and evaluation in psychology and education*. New York. NY: Macmillan Publishing Company.
- Wolf, R. (1967). Evaluation of several formulae for correction of item-total correlations in item analysis. *Journal of Educational Measurement*, 4, 1, 2 1-26.
- Zucker, S. (2003). Fundamentals of standardized testing. *Harcourt Assessment Report*, Harcourt Assessment, Inc.

Biodata

Saifuddin Azwar. Tahun 1972 memperoleh gelar sarjana muda psikologi, dan gelar sarjana psikologi tahun 1976 dari Fakultas Psikologi Universitas Gadjah Mada. Tahun 1982 memperoleh gelar *Master of Arts* dalam bidang statistika dan pengukuran pendidikan dari The University of Iowa, Iowa City, USA. Kemudian memperdalam bidang psikologi belajar di universitas yang sama. Sepanjang musim gugur tahun 1991 mengikuti *internship* di The University of Massachusetts, Amherst, USA untuk memperdalam bidang psikometri khususnya mengenai *item response theory*.