

Uji Validitas dan Reliabilitas Alat Tes Penalaran Numerik untuk Penjurusan Siswa SMA di Jakarta

YUSIN LIAUW¹, CHRISTIANY SUWARTONO²

Fakultas Psikologi, Unika Atma Jaya
Jl. Jendral Sudirman 51, Jakarta 12130
²E-mail: christiany.suwartono@atmajaya.ac.id

Diterima 8 Juli 2011, Disetujui 14 November 2011

Abstract: The aim of this study was to examine the validity and reliability of Numerical Reasoning Test which measures numerical reasoning. Data (N= 422) was taken from 13 high schools in Jakarta. Data consist of scores from high school students. The results showed that Numerical Reasoning Test had significant correlations with IST RA and TKD 1-R which measures numerical reasoning and reading comprehension. Criterion validity analysis showed that Numerical Reasoning Test was unable to predict summary report for high school students. Reliability analysis showed that Numerical Reasoning Test was consistent measuring numerical reasoning.

Key words: tes penalaran numerik, penalaran numerik, validity, reliability

PENDAHULUAN

Penjurusan Ilmu Pengetahuan Alam (IPA) dan Ilmu Pengetahuan Sosial (IPS) merupakan hal yang krusial dalam menentukan masa depan siswa (Irawati, 2008). Dari penjurusan ini, siswa lebih fokus belajar pada pelajaran-pelajaran yang sesuai dengan minatnya. Sayangnya, seringkali penentuan penjurusan siswa ini hanya berdasarkan pada nilai rapor siswa, sehingga seringkali siswa merasa kesulitan pada jurusan yang sebenarnya ia tidak merasa cocok dengan pelajarannya. Hal ini membuat siswa mendapatkan nilai yang tidak maksimal dalam jurusan tersebut. Untuk dapat menempatkan siswa tepat di jurusan yang tepat maka harus dilihat kemampuan dari siswa tersebut dalam memahami mata pelajaran. Dari hasil wawancara dengan salah seorang guru BP SMA Swasta di daerah Jakarta Pusat, guru tersebut mengungkapkan bahwa siswa diminta untuk dapat mengerti persoalan yang diberikan oleh guru dan mengembangkannya tidak hanya apa yang diajarkan oleh guru sehingga siswa dituntut untuk memiliki penalaran abstrak yang lebih untuk mata pelajaran tertentu seperti Matematika, Fisika, Kimia, Ekonomi dan Akuntansi (Komunikasi pribadi, 5 Februari 2011). Dalam hal ini, pelajaran-pelajaran tersebut membutuhkan penalaran

terhadap angka-angka sebagai stimulusnya.

Pada penelitian sebelumnya (Margareth, 2011), telah dibuat alat tes untuk melihat kemampuan penalaran numerik siswa yang dapat membantu menentukan penjurusan siswa. Namun alat tes ini masih belum terbukti *valid* dan *reliable* pada penelitian tersebut (Margareth, 2011). Selain itu, peneliti berpendapat bahwa alat tes pembandingan yang digunakan untuk uji validitas konvergen dan diskriminan tidak dapat membuktikan alat tes ini *valid* karena alat tes yang digunakan untuk pembandingan masih mengukur hal yang sama, dalam hal ini, penalaran numerik dibandingkan dengan penalaran verbal sehingga hasil yang didapat signifikan dikarenakan kedua alat tes tersebut masih mengukur hal yang sama, yaitu penalaran.

Pada penelitian ini, peneliti mengambil alat tes penalaran numerik sebagai topik yang diuji validitasnya. Peneliti memilih alat tes penalaran numerik dikarenakan hal yang diukur dalam penalaran numerik merupakan hal yang penting dalam menentukan penjurusan siswa seperti yang telah dijelaskan sebelumnya. Penalaran numerik dapat dilihat dari pengerjaan pola-pola pada suatu deret angka, dimana siswa diberi sebuah deret angka dan harus mencari pola yang ada pada deret angka tersebut kemudian memilih

dari pilihan jawaban yang ada pola yang sejenis dengan deret angka pada soal. Menurut Gregory (2004), penalaran numerik merupakan salah satu bagian dari *Primary Mental Abilities (PMA)* yang dikemukakan oleh Thurstone. PMA sendiri terdiri dari kemampuan berpikir (R), kemampuan menggunakan angka (N), kemampuan ingatan (M), kemampuan persepsi (P), kemampuan ruang (S), kemampuan verbal (V), dan kemampuan kelancaran menggunakan perkataan (W) (dalam Gregory, 2004). Penalaran numerik ini merupakan bagian dari penalaran induktif. Penalaran induktif merupakan proses berpikir dari khusus ke umum, yaitu melihat hal-hal detail atau pola yang ada kemudian menyimpulkan atau mengkaitkan detail atau pola yang sebelumnya dengan kesimpulan yang baru atau yang terkait (Sternberg, 1999). Tujuan dari penalaran numerik ini adalah melihat cara seseorang memproses hal-hal khusus kemudian disimpulkan menjadi hal umum, yang mana siswa mengkaitkan pola sebelumnya dengan pola yang baru. Pada alat tes yang diuji peneliti memiliki stimulus angka (numerik). Hal ini tentu berhubungan dengan mata pelajaran siswa, pada jurusan IPA dan IPS. Ada pun mata pelajaran yang terkait dengan penalaran numerik di jurusan IPA adalah pelajaran Matematika dan Fisika. Sedangkan pada jurusan IPS, mata pelajaran yang terkait adalah Matematika dan Ekonomi yang dikarenakan konsep dalam pengerjaan soal-soal di mata pelajaran ini adalah siswa memahami setiap bagian dari persoalan kemudian menemukan hubungan dari setiap bagian persoalan untuk dihubungkan dengan materi yang diajarkan oleh guru (Margareth, 2011).

METODE

Penelitian ini menggunakan data dari hasil tes Penalaran Numerik pada 422 subjek siswa Sekolah Menengah Atas (SMA), yang terdiri dari 13 SMA di Jakarta. Subjek terdiri laki-laki sejumlah 223 orang dan perempuan sejumlah 199 orang. Sejumlah 227 orang berasal dari jurusan IPS dan sejumlah 134 orang dari jurusan IPA. Ada pun *range* usia subjek antara 17 tahun sampai dengan 21 tahun.

Untuk pengujian validitas konstruk, subjek diharuskan mengerjakan tes PN dan juga tes

pembanding, yaitu IST RA dan TKD 1-R. Namun ada subjek (N=109) yang tidak dapat mengerjakan alat tes pembanding dikarenakan permasalahan teknis di lapangan. Sehingga untuk validitas konstruk menggunakan subjek yang tersedia (N=313).

Pada validitas prediktif, data kriteria yang dikorelasikan dengan hasil tes PN adalah nilai rapor. Namun, dari tiga belas sekolah hanya dua sekolah yang dapat memberikan nilai rapor. Sehingga untuk sampel pada validitas prediktif menggunakan jurusan IPA (N=26) dan jurusan IPS (N=66). Data inilah yang digunakan untuk validasi prediktif.

Metode validitas dan reliabilitas yang digunakan. Uji validitas konstruk bertujuan untuk mengukur sejauh mana alat tes mengukur hal yang ingin diukur (Anastasi & Urbina, 1997). Metode yang digunakan pada penelitian ini adalah *internal consistency* dan *convergent-discriminant*. *Internal consistency* digunakan untuk melihat sejauh mana item-item pada alat tes mengukur hal yang sama. Hal ini dilihat dari median dari korelasi setiap item melalui teknik *item-total correlation*. *Convergent-discriminant* digunakan untuk melihat apakah konstruk alat tes memiliki korelasi yang signifikan dengan konstruk yang serupa dan yang tidak serupa. Pada penelitian ini peneliti menggunakan IST RA untuk *convergent validity* dan TKD 1-R untuk *discriminant validity*.

Uji validitas kriteria digunakan untuk menguji sejauh mana alat tes mampu memprediksi kriteria/performa seseorang (Anastasi & Urbina, 1997). Pada penelitian ini kriteria yang diprediksi adalah performa belajar siswa di sekolah yang dilihat dari nilai rapor, sehingga nilai dari tes PN dikorelasikan dengan nilai rapor siswa. Uji reliabilitas bertujuan untuk mengukur konsistensi skor tes, dilihat dari *error* yang muncul (Anastasi & Urbina, 1997). Metode reliabilitas yang digunakan adalah metode satu kali pengambilan, yaitu metode *split-half*. Metode ini digunakan dengan membagi dua belahan secara *random* dan sistematis berdasarkan derajat kesulitan item.

Proses pengambilan data. Pengambilan data dilakukan di sekolah yang telah memberikan ijin. Data diambil saat jam pelajaran sekolah dan memberikan tes PN kemudian IST RA dan TKD 1-R. Pengambilan data nilai rapor dilakukan

dengan meminta ijin kepada pihak sekolah untuk data nilai rapor dari siswanya. Dari 13 sekolah hanya 2 sekolah yang memberikan data nilai rapor.

HASIL

Uji validitas konstruk pada *internal consistency* menunjukkan hasil yang signifikan, yaitu $r = 0.54$, maka artinya 29.16% item-item tersebut mengukur hal yang sama. Untuk bisa mengetahui lebih lanjut apakah benar alat tes ini mengukur konstruk penalaran numerik maka digunakan metode validitas *convergent-discriminant*.

Tabel 1. Hasil uji validitas *convergent-discriminant*

n = 296	<i>Convergen (IST RA)</i>	<i>Discriminant (TKD 1-R)</i>
Penalaran Numerik	0.74*	0.58*

*. Korelasi signifikan pada l.o.s 0.05

Hasil korelasi antara tes PN dengan IST RA menunjukkan korelasi yang signifikan ($r = 0.74$, $p < 0.05$) dan korelasi yang signifikan antara PN dengan TKD 1-R ($r = 0.58$, $p < 0.05$). Hal ini menunjukkan bahwa terdapat korelasi yang signifikan antara tes PN dengan IST RA dan TKD 1-R.

Uji validitas kriteria yang dilakukan dengan melihat korelasi antara nilai tes PN dengan nilai rapor siswa menunjukkan hasil yang tidak

signifikan baik pada mata pelajaran jurusan IPA dan mata pelajaran jurusan IPS (lihat Tabel 2). Dari hasil korelasi, tidak ada yang memiliki korelasi signifikan dengan hasil tes PN. Sehingga tes PN tidak dapat memprediksi performa siswa yang dilihat dari nilai rapor.

Pada pengujian reliabilitas dengan metode *split-half*, peneliti membagi soal-soal ke dalam dua belahan dengan membagi sesuai derajat kesulitan. Setelah dibagi menjadi dua, peneliti melihat apakah kedua belahan setara atau tidak dengan menggunakan teknik *t-test*. Setelah diketahui bahwa kedua belahan setara maka peneliti menghitung koefisien reliabilitas alat tes dengan menggunakan rumus *Spearman-Brown*. Hasil menunjukkan bahwa koefisien korelasi alat tes sebesar 0.89 (lihat Tabel 3). Dapat dikatakan bahwa tes PN reliabel menurut standar reliabilitas Anastasi & Urbina (1997).

Perhitungan hasil reliabilitas menunjukkan bahwa alat tes PN terpengaruh oleh *content sampling error* sebesar 0.11 (lihat Tabel 3). *Error* pada alat tes PN bersumber pada homogenitas dari soal-soal tes, yang seharusnya bila dilakukan belahan apapun maka koefisien reliabilitas yang dihitung akan menghasilkan hasil yang relatif sama (Crocker & Algina, 1986).

SIMPULAN

Penelitian ini bertujuan untuk menguji validitas konstruk, validitas kriteria, dan reliabilitas dari alat

Tabel 2. Hasil uji validitas kriteria

n=26	Mata Pelajaran IPA					
	BI	Mat	Fisika	Biologi	Kimia	Sejarah
PN	0.165	-0.053	-0.002	0.145	0.060	0.038
n=66	Mata Pelajaran IPS					
	Geografi	Ekonomi	Sosiologi	BI	Matematika	Sejarah
PN	0.005	0.010	-0.166	-0.032	-0.092	-0.008

Tabel 3. Hasil uji reliabilitas dengan teknik *Split-half*

Hasil uji <i>Split-Half</i>	
<i>Split-Half</i>	0.89
<i>Content-sampling error</i>	0.11
Mean	17.4
SD	5.42
SEM	1.79

tes PN yang mengukur penalaran numerik. Hasil uji validitas konstruk menunjukkan bahwa secara *internal* dan *external* terbukti valid. Walaupun secara *discriminant validity* menunjukkan hasil korelasi yang signifikan namun hal tersebut sudah seharusnya karena TKD 1-R dan PN masih dalam ranah yang sama, yaitu penalaran. Sehingga korelasi antar alat tes yang memiliki penalaran memiliki korelasi antara 0.5-0.75 (Gregory, 2004).

Dari hasil pengujian validitas kriteria dengan metode *concurrent validity* dapat disimpulkan bahwa alat tes ini tidak dapat memprediksi nilai rapor mata pelajaran baik untuk jurusan IPA maupun jurusan IPS karena tidak ada nilai rapor mata pelajaran yang memiliki korelasi yang signifikan dengan nilai alat tes. Menurut asumsi peneliti, hal ini bisa disebabkan karena korelasi yang dilakukan adalah dengan nilai rapor semester sebelumnya, yang dapat dipengaruhi oleh berbagai macam faktor untuk mencapai nilai rapor tersebut, seperti standar penilaian sekolah maupun tenaga pengajar. Jika dilihat antara nilai alat tes dengan nilai rapor, ada siswa yang memiliki nilai tes tinggi sedangkan nilai rapornya rendah, begitu juga sebaliknya pada mereka yang mendapatkan nilai tes rendah tetapi nilai rapornya cukup tinggi. Dari penjelasan ini menyebabkan alat tes tidak dapat memprediksi secara akurat nilai mata pelajaran yang seharusnya.

Hasil perhitungan uji reliabilitas alat tes penalaran numerik dengan metode *Split-Half* didapatkan hasil koefisien reliabilitas sebesar 0.89 yang berarti alat tes ini merupakan alat tes yang reliabel. Sehingga bisa disimpulkan item-item dari alat tes ini konsisten mengukur hal yang sama. Dari hasil analisis sumber *error*, didapatkan *error* terbesar muncul dari *content sampling error*.

DISKUSI

Dari hasil penelitian ini, peneliti melihat bahwa dengan jumlah item yang cukup banyak membuat beberapa subjek sudah kelelahan dalam mengerjakan sebelum mengerjakan seluruh soal, sehingga pada soal-soal akhir cenderung menjawab asal-asalan pada beberapa subjek yang tidak dapat menyelesaikannya. Untuk mengurangi item-item yang ada, menurut peneliti bisa dengan memilih item-item yang lebih sulit karena masih banyak

item dengan derajat kesulitan yang mudah pada alat tes ini.

Selain hal tersebut, untuk pemilihan sampel harus dilakukan dengan proporsi persebaran penjurusan yang lebih merata ataupun proporsi persebaran sekolah unggulan dengan sekolah non-unggulan. Oleh karena itu akan lebih baik bila pengambilan sampel dilakukan dengan *cluster sampling*, yaitu dengan mengambil jumlah sampel dari pengelompokan yang sudah dilakukan sebelumnya berdasarkan ketentuan tertentu seperti sekolah unggulan dengan sekolah non-unggulan ataupun dengan membagi sesuai wilayah Jakarta.

Terakhir, untuk pengujian *criterion validity* dengan metode *concurrent* menunjukkan hasil yang tidak signifikan ketika dikorelasikan dengan nilai rapor. Peneliti memberi saran untuk bisa melakukan prediksi atas nilai mata pelajaran dapat menggunakan nilai ujian, karena menurut peneliti dari nilai rapor memiliki proses yang panjang dan dipengaruhi oleh banyak faktor. Selain itu karena keterbatasan yang dihadapi sehingga peneliti tidak mendapat data nilai rapor yang lebih banyak lagi. Schmidt dkk. (dalam Crocker & Algina, 1986) menyarankan jumlah responden untuk uji *criterion validity* minimal berjumlah 200 subjek.

Walaupun analisis dalam penelitian ini masih belum sempurna, hasil dari pengujian pada penelitian ini perlu menjadi pertimbangan bagi perkembangan tes PN yang belum dapat mencapai tujuannya untuk memprediksi performa dari siswa SMA.

DAFTAR PUSTAKA

- Adinugroho, I. (2010). *Pengujian validitas dan reliabilitas Intelligenz Struktur Test Subtes Form Auswahl*. Skripsi Sarjana S1. Jakarta: Fakultas psikologi Universitas Katolik Indonesia Atma Jaya.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (4th ed.). New Jersey: Prentice Hall.
- Cohen, R. J. & Swerdlik, M. E. (2002). *Psychological testing and assessment: An introduction to tests and measurement* (5th ed.). USA: McGraw Hill.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New

- York: Horcourt Brace Jovanovich.
- Gravetter, F. J. & Wallnau, L. B. (2007). *Statistics for behavioral sciences* (7th ed.). USA : Thomson Wadsworth.
- Gregory, R. J. (2004). *Psychological testing: History, principles and applications*. USA: Pearson.
- Kaplan, R. M., & Saccuzzo, D. P. (2005). *Psychological testing: Principles, applications and issues*. California : Thomson Wadsworth
- Klauer, K. J. (1989). *Training to reason for children*. Germany: Hogrefe.
- Kumar, R. (2005). *Research methodology: A step by step guide for beginners*. London: SAGE Publications.
- Lestary, P. I. (2007). *Uji validitas dan reliabilitas tes kemampuan diferensial subtes analogi verbal*. Skripsi Sarjana S1. Jakarta: Fakultas Psikologi Universitas Katolik Indonesia Atma Jaya.
- Margareth, O. (2011). *Penyusunan alat tes deret angka untuk penjurusan siswa SMA di Jakarta*. Skripsi Sarjana S1. Jakarta: Fakultas Psikologi Universitas Katolik Indonesia Atma Jaya.
- Nunnally, J. C. (1970). *Introduction to psychological measurement*. New York: McGraw-Hill.
- Polhaupessy, L. F. (2002). *Intelligenz struktur test* (Terjemahan). Tidak diterbitkan.
- Robbins, S. P. (2001). *Perilaku organisasi* (Jilid 1). Yogyakarta: Aditya Media.
- Sattler, J. M. (1989). *Assessment of children's intelligence*. Philadelphia: W. B. Saunders Company.