

## **PENGARUH PEMBOBOTAN DAN JENIS PENILAI TERHADAP FUNGSI INFORMASI TES PERFORMANSI**

*Wakbinuddin S*

Fakultas Teknik Universitas Padang  
wakhid\_nuddin@yahoo.com

### **Abstrak**

Tujuan penelitian adalah untuk mengetahui pengaruh skala penilaian dengan pembobotan (SPPP), tanpa pembobotan (SPTP), penilai internal dan eksternal terhadap fungsi informasi tes performansi. Penelitian dilakukan dengan pendekatan eksperimen pada tujuh SMK di Sumatera Barat; melibatkan sampel 864 subjek dan 46 performansi. Pengambilan sampel dengan metode random sampel sederhana, penempatan subjek dengan metode random matriks sampel. Data dikumpulkan dengan tes performansi dan dianalisis dengan metode Rasch. Analisis data penelitian menggunakan Anava (desain faktorial 2x2). Hasil penelitian menunjukkan bahwa: 1) Fungsi informasi tes SPPP lebih tinggi daripada fungsi informasi tes SPTP; 2) Fungsi informasi tes oleh penilai eksternal lebih tinggi daripada fungsi informasi tes oleh penilai internal; 3) Faktor interaksi antara ragam tes performansi dan kelompok penilai mempengaruhi fungsi informasi tes.

**Kata kunci:** *tes performansi, fungsi informasi tes, skala penilaian dengan pembobotan, skala penilaian tanpa pembobotan, penilai internal dan penilai eksternal*

## **THE INFLUENCE OF PERFORMANCE TESTS WITH WEIGHTING AND ASSESSOR TYPE ON INFORMATION FUNCTION OF PERFORMANCE TEST**

*Wakhinuddin S*

Fakultas Teknik Universitas Padang  
wakhid\_nuddin@yahoo.com

### **Abstract**

The objective of this research was to investigate the influence of weighted and unweighted rating scales, and internal and external assessors on information function of performance test. The study was conducted on the even semester of 2004 at seven vocational schools in West Sumatera, involving 864 students and 46 performances. The sample was drawn by using simple random sampling technique, and placement was done by means of random matrix method. Factorial 2x2 design was adopted in this study. Data were collected through a performance test, and analyzed by using Rasch and two-way analysis of variance. The results of the research are: (1) the information function of the weighted rating scale test was more effective than the unweighted rating scale test, (2) the information function of the student's test assessed by external assessors was higher than that assessed by internal assessors, (3) The interaction factor between types of performance test (weighted and unweighted) and assessor groups (internal and external) affected the information function of test, (4) the information function of unweighted rating scale used by internal assessors was higher than that used by external assessors, (5) the information function of unweighted rating scale used by internal assessors was higher than the information function of weighted rating scale used by internal assessors, (6) the information function of weighted rating scale used by external assessors was higher than that of weighted rating scale used by internal assessors, (7) the information function of weighted rating scale used by external assessors was higher than that of unweighted rating scale used by internal assessors.

**Keywords:** *Performance test, informations function, weighted rating scale, unweighted rating scale, internal assessor and eksternal assessor*

## **Pendahuluan**

Undang-Undang Nomor 20 tahun 2003 tentang Sistem Pendidikan Nasional, Pasal 57, ayat (1) menyatakan evaluasi dilakukan dalam rangka pengendalian mutu pendidikan secara nasional sebagai bentuk akuntabilitas penyelenggaraan pendidikan kepada pihak-pihak yang berkepentingan. Sekolah menengah kejuruan (SMK) dengan Kurikulum edisi 1999 mencerminkan kurikulum berbasis kompetensi, di antaranya pada akhir program ada uji kompetensi dan uji produktif. Evaluasi kompetensi siswa tidak hanya dinilai pihak sekolah (internal), melibatkan pihak luar (eksternal), seperti pada ujian promosi kompetensi dan ujian produktif. Perpaduan kedua kelompok penilai diharapkan menghasilkan suatu hasil evaluasi yang bermutu. Banyak persoalan yang timbul dalam mengkombinasikan hasil penilai internal dan penilai eksternal, di antaranya persoalan perbedaan pandangan penilaian. Penilai eksternal cenderung fokus pada hasil pembelajaran (*outcomes*), analisis tugas, dan pembelajaran tuntas. Berbeda dari pandangan penilai eksternal, penilai internal fokus pada aktivitas kelas, holistik, dan perkembangan siswa. Penilai eksternal lebih pasti memberi nilai sebab dia memakai pendekatan tuntas, sedangkan penilai internal mengambang karena dia memakai pendekatan holistik, ketidakpastian nilai siswa akan lebih banyak pada penilai internal. Dikmenjur memadu penilai internal dan penilai eksternal dalam menilai keberhasilan siswa di akhir suatu program.

Pengukuran kompetensi siswa pada ujian kompetensi program keahlian otomotif SMK menggunakan tes performansi dengan skala penilaian (*rating scale*) sebagai alat ukur. Penyusunan skala penilaian berkaitan dengan penyekoran yang dirancang terlebih dulu. Penyekoran tidak lepas dari pembobotan, dan umumnya besar skor suatu butir (aktivitas) ditetapkan berdasarkan tingkat kesulitan, tingkat kepentingan dan lama kerja. Materi tes dalam hal ini adalah tugas otentik (*authentic task*) siswa, yaitu tugas yang didemonstrasikan siswa, setara dengan tugas pada bengkel mobil.

Penelitian eksperimen tes ini dilakukan dengan cara memanipulasi bobot, yaitu membelah butir (*split item*), mengempiskan kategori butir

(*collapse item categories*), mengerumunkan/mengelompokkan butir (*cluster item*), dan membuang butir (*drop item*). Dalam penelitian ini, perlakuan pembobotan dilakukan dengan mengempiskan atau menggelembungkan bobot butir. SPTP berarti bobot dikempiskan, sedangkan SPPP berarti bobot digelembungkan. Perubahan pembobotan pada tes jenis skala penilaian diduga dapat mengubah fungsi informasi.

Fungsi informasi kebalikan dari ketidakpastian, dapat diketahui dari besarnya variansi pengukuran walau dilakukan berulang-ulang. Terkait dengan penilai dan skala penilaian, ada kemungkinan penilai tertentu menilai bahwa seorang siswa memperoleh nilai optimal dengan menggunakan SPPP, tetapi tidak memperoleh nilai optimal bila memakai SPTP. Mungkin terjadi penilai tidak memberikan nilai optimal memakai SPPP, tetapi memberikan nilai optimal dengan SPTP. Kecocokan penilai dengan skala penilaian diduga melahirkan fungsi informasi maksimum.

Fungsi informasi dalam teori respons butir merupakan gambaran ketelitian, konsistensi atau stabilitas suatu alat ukur. Yang menjadi persoalan adalah skala penilaian mana dari kedua jenis penskalaan (SPTP dan SPPP) yang memiliki ketelitian yang tinggi? Penilai internal dan penilai eksternal mempunyai karakter tersendiri, bila mereka diberikan alat ukur dengan SPTP dan SPPP, dan dapat menghasilkan skor fungsi informasi tertentu. Berhubungan dengan itu, muncul pertanyaan: pasangan jenis penilai dan skala penilaian manakah yang memiliki kesesuaian sehingga menghasilkan fungsi informasi tes performansi yang maksimum?

Bila penilai (asesor) bekerja secara ideal dan profesional seharusnya menghasilkan penilaian yang tidak jauh berbeda, baik penilai internal maupun eksternal. Kenyataan pada saat uji kompetensi siswa SMK, pada umumnya ada perbedaan hasil dari penilai berbeda. Perbedaan ini menjadi penting diungkapkan agar dapat menjawab, bagaimana penilaian berbeda, bagaimana perbedaan ini dapat dihitung, dan bagaimana mengendalikannya dalam suatu model pengukuran.

Dari uraian di atas dirumuskan beberapa masalah, yaitu a) apakah ada perbedaan fungsi informasi tugas skala penilaian pakai pembobotan (SPPP) dengan fungsi informasi tugas skala penilaian tanpa pembobotan (SPTP)?; b) apakah ada perbedaan fungsi informasi tugas yang dinilai oleh penilai

eksternal dengan penilai internal? c) Apakah ada interaksi antara ragam tes performansi dan kelompok penilai dalam pengaruhnya terhadap fungsi informasi ujian? Secara praktis, penelitian ini dimaksudkan untuk: membuat tes performansi skala penilaian tanpa pembobotan (SPTP) dan skala penilaian pakai pembobotan (SPPP), memanfaatkan fungsi informasi tugas dari kedua jenis tes, mengetahui besar pengaruh tes performansi cara penilaian SPTP dan SPPP pada fungsi informasi ujian, mengetahui besar pengaruh kelompok penilai internal dan kelompok penilai eksternal pada fungsi informasi ujian. Manfaat praktis hasil penelitian adalah dapat digunakan sebagai dasar mengambil keputusan tentang kualitas keterandalan tes performansi. Secara teoretis memberikan gambaran pengembangan tes performansi, dan peranan penilai internal dan eksternal dalam menguji kompetensi siswa program mekanik otomotif SMK.

Teori Respons Butir (TRB) banyak dipakai dalam evaluasi pembelajaran terutama membahas kaitan butir dengan respons siswa, dan TRB dapat mengukur kombinasi kemampuan siswa dengan kesukaran aktivitas (butir) melalui parameter siswa dan parameter aktivitas. Parameter siswa adalah skor nilai siswa, cerminan kemampuan siswa, makin tinggi kemampuan siswa, maka semakin pula tinggi skor nilai siswa. Parameter kemampuan siswa (dituliskan dengan  $\Theta$ ). Kaitan kemampuan siswa dengan taraf sukar butir dapat dirumuskan sebagai berikut,  $P_i(\Theta) = f(\Theta - b_i)$ . Persamaan disebut model 1 (satu) Parameter. Dalam perkembangannya ternyata telah digunakan pula pada penskoran polikotomi (politomi). TRB dapat dipakai pada skala penilaian tes performansi. Pada penelitian dipakai TRB dengan pendekatan model perluasan Rasch, model ini interpretasinya sederhana karena melibatkan satu parameter, karakteristik butir dan responsden stabil dengan sampel kecil. Model Rasch hanya memakai taraf kesukaran butir ( $b_i$ ) yang dimasukkan ke dalam model. Model Rasch berkembang pada jenis tes bersifat multiaspek (*multifacets*), seperti interaksi responsden (siswa) dengan butir (tugas), misalnya pada proses penjurian, aspek berinteraksi adalah aspek penilai (juri). Untuk suatu situasi penjurian, dipakai alat ukur politomi, seperti skala penilaian (*rating scale*).

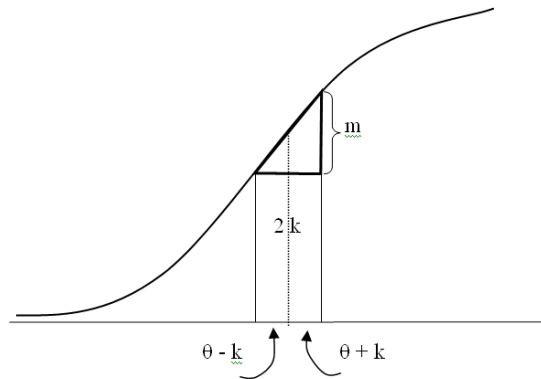
Rumus model perluasan Rasch untuk multi-aspek skala penilaian adalah:  $\text{Log}(P_{nij(k)}/P_{nij(k-1)}) = B_n - D_i - C_j - F_k$  (Linacre., Wright., dan Lunz,

1990:2). Rumus ini digunakan untuk menilai tampilan kompetensi seseorang yang dinilai Penilai pada butir (aktivitas). Formula penilaian dengan butir sama namun penilai berbeda, menjadi :  $\text{Log} (P_{nij}/P_{nij(k-1)}) = B_n - D_i - C_j - F_{jk}$ . Model perluasan Rasch (*rasch extended*) dapat mengolah data politomi, seperti data tes kompetensi mekanik otomotif. Penggunaan model Rasch dapat meningkatkan kualitas pengukuran.

TRB dalam membahas ketepatan suatu alat ukur menggunakan konsep fungsi informasi butir (ujian). Definisi fungsi informasi butir adalah rasio kemiringan (*slope*) kuadrat dengan varian (Thorndike Robert L,1982: 81).

$$\text{Fungsi informasi} = \frac{(\text{slope}(\theta))^2}{p(\theta)q(\theta)} = \frac{m}{2k}$$

Suatu fungsi informasi dapat dijelaskan sebagai berikut, misalkan beberapa siswa yang mempunyai tingkat kemampuan ( $\theta$ ) berbeda mengerjakan satu butir akan menghasilkan kurva karakteristik butir seperti pada Gambar 1.



Gambar 1. Slope kurva karakteristik butir

Pengukuran berulang-ulang tidak selalu sama hasilnya, ada varian. Ketidakpastian merupakan kejadian kombinasi kemampuan responden dengan taraf kesukaran butir berkaitan dengan jawaban berpeluang betul

atau salah. Dalam ketidakpastian kebalikkan fungsi informasi berlaku semakin besar ketidakpastian, fungsi informasi semakin kecil, sebaliknya semakin kecil ketidakpastian semakin besar fungsi informasi. Fungsi informasi merupakan hubungan terbalik variansi, sedangkan variansi merupakan wujud lain daripada ketidakpastian (Dali S Naga,1992:309). Rumus fungsi informasi butir untuk butir ke-i ditulis di bawah ini.

$$I_j(\theta) = \frac{1}{KB_j(\theta)^2}$$

Dengan demikian, jika faktor kekeliruan baku diketahui, maka fungsi informasi dapat diketahui skornya. Fungsi informasi biasanya digambarkan sebagai sebuah grafik fungsi ( $I_o$ ) terhadap kesukaran butir. Puncak grafik merupakan fungsi informasi maksimum, dalam penelitian ini yang dijadikan sebagai referensi nilai fungsi informasi adalah skor fungsi informasi maksimum. Fungsi informasi dipakai untuk: a) konstruksi perangkat ujian (tes); b) pemilihan butir; c) penilaian presisi pengukuran; d) komparasi sejumlah tes; e) penentuan bobot dalam penskoran; dan f) komparasi berbagai metode penskoran (Hambleton, Swaminathan, dan Rogers: 1991: 91-96). Penggunaan fungsi informasi pada TRB mempunyai tujuan yang sama dari analisis butir teori skor klasik. Pembahasan fungsi informasi ujian harus diawali dari pembahasan informasi butir. Rumus fungsi informasi butir satu parameter dituliskan di bawah ini.

$$I_j(\theta) = \frac{1}{P_j(\theta)Q_j(\theta)} \left[ \frac{\partial P_j(\theta)}{\partial \theta} \right]^2 \quad (1)$$

Fungsi informasi tes merupakan penjumlahan fungsi informasi butir aktivitas. Pada SPTP rumus penskoran ujian sebagai berikut.

$$X_{it} = \sum_{j=1}^N X_{ij} \quad (2)$$

Rumus fungsi informasi ujian skala penilaian tanpa pembobotan (SPTP) seperti pada persamaan (1), yaitu sama dengan fungsi informasi

butir. Sifat fungsi informasi ujian diantaranya, keterikatan satu butir dengan butir lainnya tidak ada, efek butir hanya terjadi pada total nilai fungsi informasi ujian. Sifat lain pada fungsi informasi ujian berkaitan dengan penskoran, dan penskoran mempengaruhi fungsi informasi ujian (Dali S. Naga, 1992:310). Bobot relatif butir ditentukan berdasarkan tingkat kepentingan dan kerumitan butir, secara teoritis dapat diketahui melalui analisis tugas dan tanggapan panelis instrumen. Instrumen pakai bobot disebut skala penilaian pakai pembobotan (SPPP), bobot diberikan ada tiga jenis, yaitu: 5, 10, dan 20. Skor akhir tugas yang memakai SPPP merupakan skor komposit (gabungan). Rumus penskoran pada SPPP memakai ragam kedua, yaitu penggabungan skor butir dengan pembobotan atau butir-butir dengan bobot yang berbeda-beda. Rumus penskoran ujian dengan pembobotan (Lord dan Norvick, 1968:96) adalah sebagai berikut.

$$X_{it} = \sum_{j=1}^N W_j X_{ij} \quad (3)$$

Sejalan dengan rumus penskoran, maka rumus untuk fungsi informasi ujian dengan pembobotan (Dali S. Naga, 1992:332) adalah sebagai berikut.

$$I(\theta, X) = \frac{\sum_{j=1}^N w_j \left[ \frac{\partial P_j(\theta)}{\partial \theta} \right]^2}{\sum_{j=1}^N w_j^2 P_j(\theta) Q_j(\theta)} \quad (4)$$

Dari rumus terlihat besarnya peranan pembobotan ( $w$ ) terhadap fungsi informasi tugas. Dengan demikian, dapat diduga fungsi informasi tugas yang memakai SPPP akan lebih tinggi daripada SPTP. Penskalaan (*scaling*) adalah suatu pengukuran kontinum pada suatu objek, person, atau peristiwa. Skala merupakan fasilitas yang sengaja dibuat untuk menghasilkan angka pada garis kontinum yang dijadikan skor yang mempunyai informasi.

Ragam pertama tes performansi adalah SPTP, keistimewaannya antara lain: mudah memakainya, mudah menghitungnya, mengurangi kelelahan, kebosanan penilai, dan mengurangi salah pengukuran. Ragam



kedua tes performansi adalah SPPP. Pembobotan pada butir tes bukan sesuatu yang baru pada program kejuruan. Keistimewaannya, setiap satu aspek penilaian memiliki beberapa kriteria (tidak selalu sama jumlah kriterianya), kriteria terbaik nilainya maksimum (ada skor 10 atau skor 20), sedangkan kriteria terjelek mendapat nilai kosong (skor 0).

Penilai dapat dibagi atas penilai internal dan penilai eksternal. Seorang penilai internal (*insider*) adalah individu atau kelompok yang merupakan kesatuan dari pelaksanaan program, sedangkan penilai eksternal individu atau kelompok di luar program yang dengan suatu alasan tertentu ikut memberi perhatian pada pelaksanaan program (Owen, 1993:33-41). Pada ujian praktek kompetensi di SMK penilai eksternal, adalah orang yang memiliki kepakaran di bidangnya, dapat berasal dari asosiasi profesi, majelis sekolah, DU/DI, BLK, PPPG kejuruan, perguruan tinggi, dan pihak lain yang relevan, sedangkan unsur internal adalah guru kejuruan yang memiliki bidang keahlian yang sesuai (Owen, 1993:33-41). Untuk ujian praktek, konfigurasi adalah penilai eksternal untuk internal, yaitu penilai eksternal diminta menilai suatu program dengan siswa sekolah setempat. Keberadaan penilai eksternal pada sistem pendidikan tidak lepas dari peran dan fungsinya, yaitu untuk menjamin objektivitas materi ajar dan penilaian (Piper, 1994:28).

Posisi penilai eksternal dapat disamakan sebagai pengabsahan proses penilaian, penilai eksternal merupakan jembatan antara sekolah dengan DU/DI, secara umum berfungsi pengendali kualitas di SMK. Penilai eksternal hendaklah: a) memahami tujuan pembelajaran/kriteria unjuk kerja yang harus dikuasai siswa; b) memverifikasi topik-topik tugas yang akan diuji kepada siswa; c) memverifikasi kesediaan mesin dan peralatan; d) menguji kompetensi siswa; e) mengkonfirmasi hasil penilaian kepada penilai internal yang satu tim; f) mengkomunikasikan hasil pengujian kepada tim penilai dan pihak sekolah. Tim penilai untuk satu paket kompetensi diuji oleh minimal tiga orang, satu tim penilai merupakan 2:1 atau 1:1 untuk tiap kelompok ujian dalam satu program keahlian. Artinya, 2 orang dari eksternal dan 1 orang dari internal. Aspek yang dinilai meliputi proses kerja, sikap dan waktu yang digunakan untuk menyelesaikan pekerjaan (tugas).

Ragam tes performansi adalah cara penilaian yang menuntut siswa melakukan tugas dalam bentuk perbuatan yang dapat diamati guru atau *penilai* lainnya dengan menggunakan skala penilaian. Kemampuan performansi dalam membedakan keadaan siswa menunjukkan bahwa performansi memiliki sifat mengukur dan menilai. Sifat tersebut membuatnya menjadi suatu instrumen penilaian yang disebut tes performansi, dan membuat tes performansi dikategori sebagai Tes acuan-kriteria.

Tes performansi mempunyai beberapa keistimewaan, antara lain: a) mengatasi beberapa hal (terutama keterampilan) yang tidak dapat dinilai dengan ‘tes kertas-pensil’; b) lebih alami, langsung, lebih tuntas menilai keterampilan (*skill*); c) bermanfaat bagi siswa yang kurang ingin membaca; d) mendorong aplikasi pembelajaran kepada situasi kehidupan nyata. Namun, ada juga beberapa kelemahan penilaian performansi, yaitu: a) pemakaian waktu dan usaha yang banyak; b) penskoran dan penilaian cenderung subjektif, tidak adil, rendah reliabilitas; c) penilaian individual daripada kelompok (Gronlund, 1993: 115). Ada tiga elemen utama tujuan tes performansi, yaitu: 1) kondisi tugas yang akan ditampilkan; 2) deskripsi tugas; dan 3) kriteria kerja. Elemen ini menunjukkan, tes performansi tidak dapat dilepaskan dari suatu perbuatan (kerja). Beberapa penulis menggandeng kata performansi dengan kata tugas, dan dibaca tugas performansi (*performance task*).

Penelitian ini bertujuan untuk mengetahui pengaruh skala penilaian dengan pembobotan (SPPP), tanpa pembobotan (SPTP), penilai internal dan eksternal terhadap fungsi informasi tes performansi. Pada penelitian ini, ada dua ragam tes performansi yang berbeda konsep dan teknis pelaksanaan pengukuran kompetensi siswa, yaitu: 1) SPTP; skala penilaian mempunyai skor tetap 0 sampai 5, karena setiap jarak aktivitas mempunyai skor sama dan, 2) SPPP; pada SPPP skala penilai pada aktivitas mempunyai bobot berbeda, mulai dari skor 0-5, skor 0-10, dan skor 0-20.

Dalam banyak hal, prosedur dan produk adalah aspek penting performansi, umpamanya keterampilan memeriksa kerusakan dan memperbaiki mobil. Prosedur ditekankan di langkah awal dan produk belakangan, sesudah prosedur dikuasai. Dalam keterampilan khusus, seperti

kecepatan dievaluasi awal kegiatan, belakangan difokuskan pada kebersihan dan akurasi jenis bahan dan kecepatan prosedur.

Fungsi tes performansi adalah untuk mengetahui sejauh mana suatu program pendidikan berhasil diterapkan. Dalam pendidikan kejuruan performansi dikaitkan dengan tujuan pembelajaran (program), sehingga tujuan performansi diartikan sebagai tujuan belajar (Larson, 1972:126).

Pembuatan cara menilai performansi dengan: 1) memperhatikan tujuan pembelajaran yang adanya tampilan siswa; 2) menyeleksi topik yang akan dinilai, dan yang dinilai adalah proses kerja; 3) menyeleksi tingkat otentik tugas, pada langkah ini perhatikan relevansi antara topik pelajaran di kelas dengan tugas di lapangan kerja, 4) menyeleksi kepantasan melaksanakan tugas; 5) menetapkan kriteria tingkat keberhasilan tampilan siswa, seperti kualitas pekerjaan dan waktu kerja; 6) menetapkan metode observasi, di sini dipakai cara penilaian skala penilaian (*rating scale*) pakai bobot dan tanpa pembobotan. Pada SMK batas kelulusan ditetapkan berdasarkan normatif, Depdiknas menetapkan batas lulus untuk matapelajaran produktif/keahlian adalah tujuh (7).

### **Metodologi Penelitian**

Penelitian ini merupakan penelitian eksperimen, pada butir tes, yaitu dengan perlakuan mengempiskan atau menggelembungkan bobot butir. Penelitian eksperimen pada butir telah pernah dilakukan seperti de Gruijter dan Van der Kamp (1991), penggunaan pada teori generalibilitas telah dilakukan Cronbach (1971) (de Gruitjter, 2002: 50-55). Berdasarkan teori generalisabilitas dan teori pensekoran, maka desain penelitian yang cocok adalah Anava faktorial 2 x 2. Desain tersebut banyak dipakai pada psikometri, karena mampu membedakan perbedaan personal dan perbedaan butir. Perbedaan personal membicarakan kemampuan (*ability*) siswa dan perbedaan butir membahas kesukaran butir (aktivitas).

Pemilihan dan penetapan penelitian ini sebagai penelitian eksperimen didasari pandangan bahwa butir mempunyai karakter tersendiri, dengan demikian butir dapat dimanipulasi. Perlakuan dengan memanipulasi bobot butir, yaitu membelah butir (*split item*), mengempiskan

kategori butir (*collapse item categories*), mengelompokkan butir (*cluster item*), dan membuang butir (*drop item*).

Dalam melakukan eksperimen, ada beberapa tahapan yang akan dilalui: Proses pra-eksperimen, (1) melakukan *assesment* terhadap pelaksanaan program sesuai dengan kurikulum SMK edisi 1999; (2) memeriksa ketercapaian materi ajar di kelas responden; (3) mencari beberapa tes format skala penilaian yang pernah dipakai dalam dua tahun belakangan sesuai materi kurikulum; (4) menetapkan topik tugas yang akan diuji ke siswa; (5) mengembangkan tes cara penilaian SPPP dan skor SPTP sesuai topik tugas yang telah ditetapkan; (6) melakukan studi pendahuluan (*pilot studies*); (7) untuk verifikasi dicari koefisien reliabilitas dan validitas; (8) sesuai dengan studi pendahuluan, selanjutnya melakukan *upgrade* proposal penelitian, seandainya ditemukan kelemahan di lapangan; (9) membuat laporan.

Kegiatan pendahuluan (1) menentukan kriteria penilai internal dan eksternal; seperti: pengalaman 5 tahun; (2) *choacing*, seperti: cara menggunakan kedua jenis tes dan lainnya; (3) menentukan jadwal eksperimen; (4) penggantian bahan tes; (5) mempersiapkan fasilitas bengkel; (6) menentukan dan mengundang responden; (6) melakukan experiment sebanyak topik pengujian praktik selama ini ; (7) menjaga validitas internal dan eksternal penelitian. Pelaksanaan eksperimen mencakup: (1) menyediakan meja kerja penilai internal dan eksternal secara terpisah; (2) menyediakan ruang tunggu bagi siswa responden; (3) menata letak mesin sehingga dapat terpantau kedua penilai; (4) set *stopwatch* sesuai lama pengujian; (5) penilai diperbolehkan menanya suatu hal kepada responden berkaitan dengan perbuatan responden; (6) sesama penilai tidak diperbolehkan berinteraksi selama ujian; (7) pengujian selesai tepat waktu; (8) Ujian bentuk sumatif, yaitu diakhir kelas II. Pengukuran dilakukan berulang-ulang, yaitu dalam bentuk beberapa tugas yang dipraktekkan siswa. Proses pasca-eksperimen: (1) lembar tes diparaf penilai; (2) lembar tes dikumpulkan segera; (3) lembar tes yang cacat disortir; (4) data pada lembar tes dipindahkan ke kerangka data mentah.

Ada beberapa metode yang dipakai memecahkan permasalahan penelitian adalah: 1) penyekoran : yaitu penyekoran tanpa pembobotan

(SPTP) dan penyekoran pakai pembobotan (SPPP); 2) acuan kriteria dijadikan dasar penilaian, yaitu skor nilai  $\geq 7,0$ ; dengan penetapan ini sebagai batas lulus maka ada pengkategorian siswa, yaitu: 1) tanpa kompetensi, 2) kompetensi kurang, 3) kompetensi minimal, 4) kompetensi bagus. Pengkategorian ini dijadikan dasar kode pengolahan data untuk dapat dipakai pada program *Facet* dari Linacre. Selanjutnya metode ke 3) yaitu penentuan unidimesi (ketunggalanmatra) digunakan metode analisis faktor; 4) penentuan independensi lokal digunakan statistik kai-kuadrat; 5) skor baku ( $z$ ) digunakan untuk mentransformasi skala penilaian yang berbeda ke bentuk skala logit; dan 6) desain generalizabiliti, untuk menganalisis banyak faktor (*facet*), seperti: penilai, tugas, dan kemampuan siswa, dikenal dengan sebutan *multifacet*.

Sesuai dengan desain penelitian, ada dua sumber varian, yaitu tugas dan siswa. Populasi siswa adalah siswa kelas II SMK di Provinsi Sumatera Barat. Adapun jumlah keseluruhan SMK di Sumatera Barat, yang negeri sebanyak 14 dan Swasta sebanyak 39. Jumlah populasi siswa kelas II sekitar 16.552 orang, diambil dari sekolah SMK N I Pariaman, BLPT Padang (SMKN I dan V Padang), SMK N II Payakumbuh, SMK N II Solok, SMK Muhammadiyah I Padang, SMK Mitra Payakumbuh, SMK Muhammadiyah Solok.

Tabel 1. Desain penelitian anava faktorial dua jalur

Kelompok Penilai (B)	Ragam tes Performansi (A)	
	SPTP (A <sub>1</sub> )	SPP (A <sub>2</sub> )
Internal (B <sub>1</sub> )	A <sub>1</sub> B <sub>1</sub>	A <sub>2</sub> B <sub>1</sub>
Eksternal (B <sub>2</sub> )	A <sub>1</sub> B <sub>2</sub>	A <sub>2</sub> B <sub>2</sub>

Pengambilan sampel pada penelitian ini dilakukan secara bertahap: Tahap pertama, penentuan sekolah dilakukan teknik random sampling sederhana, yaitu dengan mengundi 7 dari 17 sekolah. Ada dua

pertimbangan utama dalam menentukan sekolah populasi, (1) legimitas pelaksanaan ujian, ke-17 sekolah telah mendapat izin dari Dinas Diknas Sumbar melaksanakan Ujian kompetensi secara mandiri; (2) kelengkapan fasilitas praktek memenuhi persyaratan pelaksanaan ujian kompetensi.

Tabel 2. Sebaran sampel penelitian dan asal sekolah

No.	Nama lembaga pendidikan	Status	Jumlah siswa
1	BLPT	Negeri	157
2	SMK N 2 Payakumbuh	Negeri	78
3	SMK N 2 Solok	Negeri	116
4	SMK N 1 Pariaman	Negeri	112
5	SMK Muhammadiyah 1	Swasta	147
6	Padang	Swasta	112
7	SMK Muhammadiyah Solok	Swasta	142
	SMK Mitra Payakumbuh		
Jumlah			864

Tahap ke dua pengambilan sampel, diawali pembuatan penomoran siswa, kemudian mengambil subyek dengan teknik sampel random sederhana. Dari proses ini didapat dua kerangka sampel sementara, yaitu sub kelompok dinilai dengan SPTP ( $A_1$ ) sebanyak 432 siswa dan sub kelompok dinilai SPPT ( $A_2$ ) sebanyak 432 siswa. Tahap ketiga penempatan siswa pada sub-sub matriks sampel.

Desain matriks sampel dibuat berdasarkan prinsip distribusi seimbang jumlah tugas (aproksimasi). Alasan pembuatan desain matriks sampel disebabkan banyak tugas yang akan diujikan kepada siswa, namun keterbatasan fisik siswa dan waktu pelaksanaan tidak memungkinkan ke-46 tugas dapat dilakukan satu siswa, maka setiap kompetensi dibagi tiga kelompok. Dengan demikian satu siswa mendapatkan tujuh tugas dan paling sedikit empat tugas.

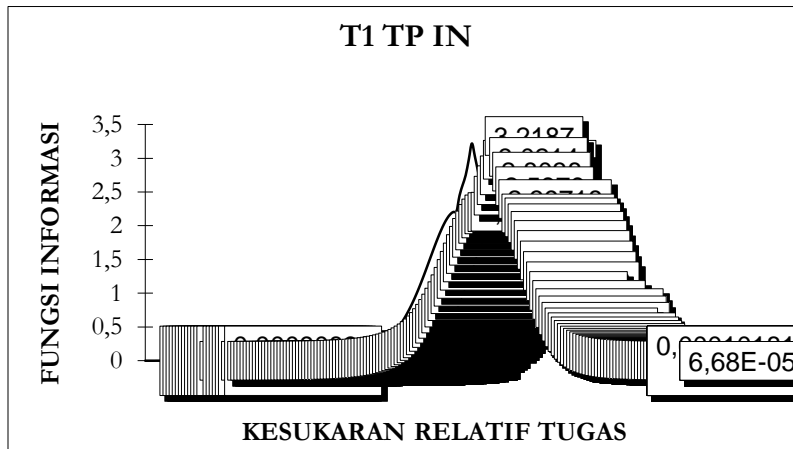
Teknik penempatan subyek pada sub-matrik juga menggunakan random acak sederhana. Tiga digit (angka) terakhir dijadikan referensi nomor defenitif subyek. Dari pemerosesan ini, matriks sampel terisi nama-nama subyek ke-1 sampai subyek ke-432. Setiap siswa diuji penilai internal dan penilai eksternal. Pengujian hipotesis penelitian digunakan teknik analisis varian (Anava) faktorial 2 x 2. Sebelum menggunakan Anava terlebih dahulu dilakukan uji persyaratan analisis, yaitu uji normalitas dan homogenitas data.

### **Hasil Penelitian dan Pembahasan**

Suatu data yang akan dimasukkan ke dalam model Rasch harus memenuhi persyaratan analisis. Namun, ada dua hal pokok persyaratan suatu tugas (tes), yaitu independensi lokal dan unidimensi. Uji independensi menggunakan kai-kuadrat dan untuk menguji unidimensi tugas menggunakan analisis faktor. Ditemukan semua tugas memiliki indepedensi lokal, karena t hitung kurang dari t tabel ( $\chi^2_{\text{hit.tgs}} < \chi^2_{(0,05)(5)}$ ). Dengan demikian dapat dikatakan, antara tugas satu dengan tugas lain tidak tergantung satu sama lainnya. Dari hasil pengolahan data juga ditemukan bahwa semua tugas adalah mengukur satu dimensi, ini teruji pada KMO = 0,859.

Hasil uji normalitas terhadap data fungsi informasi dari setiap perlakuan diketahui seluruh kelompok data normal. Pengujian homogenitas data sampel dilakukan uji Levene's, didapat nilai probabilitas uji Levene (0,792), di atas 0,05, sehingga  $H_0$  dapat diterima; varians sampel populasi adalah homogen.

Fungsi Informasi ujian Tugas 1 (T1) dari skala penilaian tanpa pembobotan (SPTP) oleh penilai internal.



Gambar 2. Fungsi informasi tugas 1 dinilai oleh penilai internal dengan skala penilaian tanpa pembobotan (SPTP).

Dari gambar di atas ditemukan fungsi informasi tugas sebesar 3,21 memakai skala penilaian tanpa pembobotan (SPTP) oleh penilai internal. Grafik tanpa meruncing ke atas, dan grafik melandai seimbang antara kiri dan kanan. Ini menunjukkan bahwa tingkat kesukaran tugas satu cenderung berdistribusi seimbang, artinya semua siswa yang melaksanakan tugas satu relatif normal, dan keterandalan tugas satu baik karena mencapai 3,21 untuk skala lima.

Ada empat kategori dari setiap tugas dapat terjadi saat seorang penilai memberi nilai pada siswa. Pada gambar 2, dimulai dari hadap kiri, grafik paling kiri (1) adalah kategori kompetensi tak ada, grafik (2) adalah kategori kompetensi kurang, grafik (3) adalah kategori kompetensi minimal, dan grafik (4) adalah kategori kompetensi bagus. Dari hasil penilaian penilai, probabilitas siswa berada pada salah satu kategori dapat saja terjadi. Gambar 2 adalah tugas 38 dengan SPTP dengan penilai eksternal ditemukan probabilitas terbesar terjadi pada kategori (1) yaitu kompetensi tak ada, dan probabilitas terkecil terjadi pada kategori (3), yaitu kompetensi minimal. Ada empat kategori pada semua tugas (46) yang diujikan pada siswa. Statistik deskriptif fungsi informasi ujian disampaikan pada Tabel 3.



Tabel 3. Deskripsi Data Fungsi Informasi

Sumber Statistik		A <sub>1</sub>	A <sub>2</sub>	Σ
B <sub>1</sub>	n	46	46	92
	$\bar{x}$	3,44	3,20	3,32
	s	0,74	0,57	0,62
B <sub>2</sub>	n	46	46	92
	$\bar{x}$	2,68	4,53	3,60
	s	0,60	0,67	0,63
Σ	n	92	92	184
	$\bar{x}$	3,06	3,86	3,46
	s	0,67	0,62	1,25

Rata-rata terendah ada pada sel A<sub>1</sub>B<sub>1</sub>, sebesar 2,68, dan rata-rata tertinggi ada pada sel A<sub>2</sub>B<sub>2</sub>, sebesar 4,53. Ini menunjukkan bahwa peningkatan fungsi informasi sebagai akibat perlakuan meningkat jauh pada kelompok penilai eksternal. Berbeda pada kelompok penilai internal terdapat penurunan rata-rata pada sel A<sub>1</sub>B<sub>1</sub>, yang semula sebesar 3,44 menjadi sebesar 3,20. Kesemua perubahan rata-rata fungsi informasi pada sel-sel disebabkan oleh perlakuan dari ragam tes SPTP ataupun SPPP.

Pengujian hipotesis dilakukan dengan menggunakan analisis varian (Anava) dua jalan yang dilanjutkan dengan uji Tukey. Anava dua jalan digunakan untuk menguji pengaruh utama (*main effect*) dan interaksi (*interaction effect*) variabel bebas ragam tes dan kelompok penilai terhadap variabel terikat fungsi informasi. Selanjutnya hasil analisis data dengan anava disajikan dalam bentuk Tabel 4.

Tabel 4. Tabel 4. Rangkuman Hasil Anava Data Fungsi Informasi

Sumber Varian	JK	dk	RJK	F <sub>0</sub>	F <sub>t</sub>	
					0,05	0,10
Antar Kolom	3,775	1	3,775	8,892**	3,84	6,63
Antar Baris	29,764	1	29,764	70,118**	3,84	6,63
Interaksi	50,085	1	50,085	117,989**	3,84	6,63
Dalam kelompok	76,408	180	0,424			
Jumlah	2366,712	183				

Pada perbedaan antarkolom  $F_0 = 8,892 > F_t = 3,84$ , berarti hipotesis nol ( $H_0$ ) yang menyatakan bahwa tidak ada perbedaan fungsi informasi tugas yang dinilai dengan SPTP dengan fungsi informasi tugas SPPP ditolak, atau hipotesis alternatif ( $H_a$ ) yang menyatakan fungsi informasi tugas SPPP lebih tinggi daripada fungsi informasi tugas SPTP diterima. Dengan kata lain, bahwa fungsi informasi tugas SPPP lebih tinggi daripada fungsi informasi tugas SPTP. Siswa yang dinilai SPTP awalnya lebih tinggi dari fungsi informasi yang dinilai dengan SPPP atau  $x_{A1} = 3,44 > x_{A2} = 2,68$ , namun karena adanya peran karakteristik skala penilaian dan karakteristik penilai maka pada saat tertentu fungsi informasi SPPP lebih tinggi daripada SPTP.

Skala dapat diklasifikasikan berdasarkan orientasi pengukuran yaitu: pendekatan sitimulus, pendekatan respons, dan pendekatan subjek. Penskalaan pada penelitian ini berorientasi pendekatan respons. Karakteristiknya, jawaban yang disusun bertingkat, dimulai dari butir jawaban rendah hingga ke jawaban tinggi. Prosedur yang demikian dikenal dengan penskalaan Guttman, bilamana siswa dapat mengerjakan aktivitas urutan kerja tertinggi, tentunya aktivitas yang lebih rendah dapat juga dikerjakannya. Skala penilaian tanpa pembobotan (SPTP) adalah satuan yang menilai kualitas tampilan siswa dengan jumlah skala konsisten,

kategori 0 sampai kategori 5. Penskoran ini mencerminkan tingkat kualitas performansi. Kategori 5 berarti tingkat tertinggi, dinyatakan sebagai nilai kompetensi maksimum; dan angka kategori 0 berarti terendah, dinyatakan sebagai nilai tanpa kompetensi. Keistimewaan alat ukur SPTP: mudah memakainya, mudah menghitungnya, mengurangi kelelahan, kebosanan penilai, dan mengurangi salah pengukuran.

Pada skala penilaian pakai pembobotan (SPPP) mempunyai keistimewaan, setiap satu aspek penilaian memiliki beberapa kriteria (tidak selalu sama jumlah kriterianya), kriteria terbaik nilainya maksimum (ada sekor 10 dan 20), kriteria terjelek mendapat nilai kosong (sekor 0). Semakin besar pembobotan ( $w$ ) semakin besar fungsi informasi, ini sesuai dengan rumusdi bawah ini.

$$I(\theta, X) = \frac{\sum_{j=1}^N w_j \left[ \frac{\partial P_j(\theta)}{\partial \theta} \right]^2}{\sum_{j=1}^N w_j^2 P_j(\theta) Q_j(\theta)}$$

Dengan demikian temuan yang menyatakan fungsi informasi tugas SPPP lebih tinggi daripada fungsi informasi tugas SPTP dapat diterima. Ini sesuai dengan temuan Lily Rochaya menemukan bahwa fungsi informasi butir soal bentuk pilhan ganda dengan lima pilihan jawaban lebih tinggi daripada nilai fungsi informasi butir soal bentuk pilhan ganda dengan tiga pilihan jawaban. Semakin besar bentangan penyebaran nilai, peluang kecocokan siswa dengan alat ukur akan lebih tinggi.

Pada perbedaan antarbaris harga  $F_0 = 70,118 > F_t = 3,84$ , berarti hipotesis nol ( $H_0$ ) yang menyatakan ‘fungsi infromasi tugas yang dinilai penilai eksternal tidak berbeda daripada fungsi informasi tugas yang dinilai penilai internal’; ditolak, atau hipotesis alternatif ( $H_a$ ) ‘fungsi infromasi tugas yang dinilai penilai eksternal lebih tinggi daripada fungsi informasi tugas yang dinilai penilai internal’; dapat diterima. Dengan kata lain, dapat dinyatakan bahwa terdapat perbedaan fungsi informasi antara siswa yang dinilai oleh penilai internal dan siswa yang dinilai oleh penilai eksternal.

Fungsi informasi yang dinilai oleh penilai internal lebih rendah dari siswa yang dinilai oleh penilai eksternal ( $x_{B1}=3,319 < x_{B2}=3,606$ ).

Ini disebabkan, kebiasaan bekerja di industri (bengkel) yang mengacu pada disiplin, produktivitas tinggi, variatif dan teliti membuat seorang penilai eksternal cenderung banyak gagasan dan independen. Berbeda, dengan penilai internal lebih mementingkan proses pembelajaran sesuai keinginan kurikulum. Penilai internal menekankan adanya usaha siswa, dan secara rinci penilai internal cenderung lebih suka sasaran hasil segera, tidak variatif, terfokus, linear, sistematis, dan dependen.

Hipotesis nol ( $H_0$ ) yang menyatakan bahwa tidak ada interaksi antara fungsi informasi dan siswa yang dinilai oleh penilai internal tidak berbeda dengan siswa yang dinilai oleh penilai eksternal ditolak, atau hipotesis alternatif ( $H_a$ ) yang menyatakan faktor interaksi antara ragam tes performansi dan kelompok penilai mempengaruhi fungsi informasi tugas diterima, dimana harga  $F_0= 117,989 > F_t= 3,84$ . Dengan demikian, pencapaian fungsi informasi dipengaruhi oleh interaksi antara ragam skala penilaian yang dipakai dalam ujian dengan kelompok penilai.

Ini dapat dijelaskan dari karakter pekerjaan yang setiap hari dilaksanakan penilai, cenderung membuat karakter berbeda dalam menilai siswa. Pada skala yang sama, penilai eksternal cenderung memberi nilai sangat variatif mulai dari nilai rendah sampai nilai tertinggi, berbeda daripada penilai internal lebih banyak penilai terfokus pada skor nilai tertentu. Pada penilai eksternal keragaman hasil penilaiannya membuat kecocokan kemampuan siswa dengan tugas tinggi, berbeda dari pada hasil penilaian penilai internal cenderung terfokus membuat fungsi informasi tugas rendah. Dengan demikian dapat dikatakan temuan 'fungsi informasi tugas dengan penilai eksternal lebih tinggi daripada fungsi informasi tugas dengan penilai internal' dapat diterima.

Dari uji hipotesis ketiga terbukti ada interaksi antara ragam skala penilaian dan kelompok penilai, ini menunjukkan bahwa efek utama ada perbedaan, sehingga perlu dilacak sel-sel mana saja yang berbeda. Teknik analisis yang dapat menjelaskan pengujian adalah uji Tukey. Analisis ini digunakan untuk menguji perbedaan nilai rerata absolut dari dua kelompok yang dipasang dengan cara membandingkan nilai tersebut dengan nilai

kritis HSD. Analisis ini merupakan analisis lanjut, yang akan mengungkapkan apakah ada perbedaan antarsel. Hasil uji Tukey pada taraf signifikansi ( $\alpha$ )=0,05. Pada kelompok A1B1 dan A2B1 harga  $q_0 = 11,53 > q_t = 0,18$ , berarti hipotesis nol ( $H_0$ ) ditolak atau hipotesis alternatif keempat diterima, berarti fungsi informasi tugas SPTP dengan penilai internal lebih tinggi daripada fungsi informasi tugas (SPPP) dengan penilai internal.

Skala penilaian SPTP terbentuk dari penskoran tetap, penggunaan penskoran demikian melahirkan format lembaran observasi lebih mudah digunakan dalam menilai tampilan kerja siswa, format penilaian yang tetap (monoton) dan tidak membutuhkan kriteria macam-macam sesuai dengan karakter penilai internal. Sedangkan SPPP memiliki penskoran menggelembung dan banyak kriteria, dan skala penilaian lebar merupakan fasilitas bagi penilai eksternal memberi tanda nilai. Penilai sebagai subyek evaluasi mempunyai karakteristik masing-masing, karakter penilai internal cenderung dependen, sistematis dan tidak variatif. Sesuai dengan karakter ini, penilai internal lebih cocok dengan SPTP dibandingkan pada SPPP.

Pada kelompok A1B2 dan A2B2 harga  $q_0 = 28,067 > q_t = 0,18$ , berarti hipotesis nol ( $H_0$ ) ditolak atau hipotesis alternatif ke lima diterima, hal ini berarti pada fungsi informasi tugas SPPP dengan penilai eksternal lebih tinggi daripada fungsi informasi tugas SPTP dengan penilai eksternal. Bagi penilai eksternal yang memiliki kecenderungan jelimet dan rinci kriteria kerja, saat melaksanakan pengujian akan memberi respons berbagai pertimbangan untuk menentukan atau kemungkinan lainnya dalam menilai kompetensi siswa yang tampil. Karakter penilai eksternal demikian tersalurkan melalui SPPP, sehingga SPPP dengan kecenderungan karakter penilai eksternal mempunyai kecocokan tinggi.

Pada kelompok A1B1 dan A1B2 harga  $q_0 = 0,282 > q_t = 0,18$ , berarti hipotesis nol ( $H_0$ ) ditolak atau hipotesis alternatif keenam diterima, hal ini berarti pada fungsi informasi tugas SPTP dengan penilai internal lebih tinggi daripada fungsi informasi tugas SPTP dengan penilai eksternal. Berdasarkan karakter guru yang memiliki kecenderungan bertindak tidak variatif (monoton) namun sistematis, dapat dikatakan bahwa skala penilaian tanpa pembobotan (SPTP) lebih cocok untuk Penilai internal ketimbang

Penilai eksternal. Dengan demikian dapat dikatakan, ada kecenderungan skor fungsi informasi yang dinilai oleh penilai internal menggunakan SPTP lebih tinggi daripada Penilai eksternal yang juga menggunakan SPTP.

Pada kelompok A2B1 dan A2B2 harga  $q_0 = 27,91 > q_t = 0,18$ , berarti hipotesis nol ( $H_0$ ) ditolak atau hipotesis alternatif ketujuh diterima, hal ini berarti pada fungsi informasi tugas SPPP dengan penilai eksternal lebih tinggi daripada fungsi informasi tugas SPPP dengan penilai internal. Temuan ini dapat dijelaskan, bahwa penilai eksternal dalam ujian praktek berasal dari kalangan profesional, KADIN, dan perguruan tinggi. Pada penelitian ini, penilai eksternal yang dipakai pihak SMK adalah dari kalangan profesional, dari mekanik mobil. Mekanik dalam bekerja umumnya teliti, banyak alternatif metode kerja, dan kuantitatif (produktivitas) dijadikan ukuran dalam bekerja. Kebiasaan-kebiasaan ini terbawa dalam menilai kompetensi siswa, sehingga mekanik dalam menilai siswa mempunyai banyak kriteria dan cenderung presisi. Berbeda dari penilai internal, kebiasaan kerja logis, sistematis dan monoton membuat penilai internal bekerja kurang variasi. Karakter kerja mekanik (penilai eksternal) bekerja demikian tersalurkan melalui alat ukur SPPP.

## **Simpulan**

Berdasarkan data dapat disimpulkan bahwa: a) Secara keseluruhan fungsi informasi ujian yang diukur dengan SPPP lebih tinggi daripada SPTP; b) Secara keseluruhan fungsi informasi tugas yang dinilai penilai eksternal lebih tinggi daripada fungsi informasi tugas yang dinilai penilai internal; c) Ada interaksi antara ragam tes performansi dan kelompok penilai dalam pengaruhnya terhadap pencapaian fungsi informasi tugas; bagi penilai internal, fungsi informasi tugas dinilai dengan SPTP lebih tinggi daripada menggunakan SPPP; bagi penilai eksternal, fungsi informasi tugas dengan SPPP lebih tinggi dari menggunakan SPTP.

### **Daftar Pustaka**

- Dali S Naga. 2002. Diklat perkuliahan 'Psikometri'. Jakarta: Pascasarjana UNJ.
- de Gruijter, Dato N.M. dan Van der Kamp, Leo J. Th. *Statistical Test Theory for Education and Psychology..* May 2002. [http://www.unt.edu/rss/class/rich/5840/test\\_theory\\_text.pdf](http://www.unt.edu/rss/class/rich/5840/test_theory_text.pdf).
- Gronlund, N. E., 1993. *How to Make achievement Test and Assessment*. Boston: Ally and Bacon.
- Hambleton, R.K., Swaminathan, H., Rogers, H.J. 1991. *Fundamentals of Item Response Theory*. Newbury Park: SAGE Publications.
- Landy, F. J., Trumbo, D. A. 1980. *Psychology of Work Behavior*. Illionis: The Dorsey Press.
- Larson, M. E. 1972. *Teaching Related Subjects in Trade and Industrial and Technical Education*. Ohio: Charles E, Merrill Publishing Co;
- Lunz, M.E., Wright, B.D., Linacre, J.M. 1990. *Measuring The Impact of Judge Severity on Examination Scores*. Pp. 1-2. [www.rasch.org/memo47.htm](http://www.rasch.org/memo47.htm).
- Rochaya. 2001. Fungsi Informasi Test Pilihan Ganda. Thesis. Jakarta: PPs UNJ.
- Tay-Lim, B. S-H. Davis dan Tang. *The Impact Treatments on NAEP Reporting Scale Secore*, makalah pada National Council on Measurement in Education, 22-24 April 2003. Chicago: Educational Testing Service. [www.ets.org/legal/copyright](http://www.ets.org/legal/copyright).
- Undang-Undang Nomor 20 tahun 2003. Sistem Pendidikan Nasional.