

Pengaruh Ragam Tes Performansi dan Kelompok Penilai Terhadap Fungsi Informasi Tugas Praktek Siswa SMK

Wakhinuddin S.

Fakultas Teknik-UNP Padang

e-mail : wakhid_nuddin@yahoo.com; Weblog: Wakhinuddin.wordpress

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

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

Abstract: The objective of the experimental research was to investigate the effect of weighted and unweighted performance test with rating scales, and internal and external assessors to information function of student's tasks. The study was conducted at seven vocational schools in West Sumatera, involving 864 students and 46 tasks. The sample was drawn by using simple random sampling technique, and placement was done by means of random matrix method. Data was collected through a performance test and Rasch, and analyzed using Anova by Factorial 2x2 design. The results of the research are: 1) the information function of the weighted rating scale was more effective than the unweighted rating scale; 2) the information function of the student's task assessed by external assessors was higher than that assessed by internal assessors; and 3) The interaction factor between types of performance test and assessor groups affected the information function of student's task.

Key words: *performance test, information function, weighted rating scale, unweighted rating scale, internal assessors, external assessors*

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, telah mencerminkan kurikulum berbasis kompetensi, di antaranya pada akhir program ada uji kompetensi dan uji produktif. Evaluasi kompetensi siswa tidak hanya dinilai pihak sekolah (internal), juga 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 acuan tuntas, sedangkan penilai internal mengambang

karena dia memakai acuan holistik. Oleh karena itu, ketidakpastian nilai siswa akan lebih banyak pada penilai internal.

Dikmenjur memadu penilai internal dan penilai eksternal dalam menilai keberhasilan siswa di akhir suatu program. Alasan kenapa memakai kedua jenis penilai, karena pemakaian Penilai internal untuk menjaga mutu, dan pemakaian Sedangkan Penilai eksternal untuk mengendalikan mutu pendidikan kejuruan.

Pengukuran kompetensi siswa pada ujian kompetensi program keahlian otomotif SMK menggunakan tes performansi dengan skala penilaian 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 ditampilkan (didemonstrasikan) siswa, setara dengan tugas-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 mempunyai kerja optimal dengan menggunakan SPPP, tetapi tidak dapat kerja optimal bila memakai SPTP. Mungkin pula terjadi penilai tidak dapat bekerja optimal memakai SPPP, tetapi dia dapat bekerja optimal dengan SPTP. Kecocokan penilai dengan skala penilaian tertentu diduga melahirkan fungsi informasi maksimum tinggi.

Fungsi informasi 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 skala penilaian tanpa pembobotan (SPTP) dan skala penilaian pakai pembobotan (SPPP) akan bersinergi, dan dapat menghasilkan skor fungsi informasi tertentu. Berhubungan dengan itu, ada dua pertanyaan. Penilai yang mana yang cocok dengan alat ukur skala penilaian yang mana?

Pada konteks penilaian tidak langsung (*indirect*), seperti pada ujian kompetensi, pengukuran dilakukan pada banyak faktor (*facet*), paling tidak ada tiga faktor yang dominan, yaitu: kemampuan siswa, kesukaran tugas, dan kepelikan penilai. Bila penilai 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.

Ada beberapa masalah yang dapat dirumuskan, 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. Selain itu, manfaat praktis

hasil penelitian ini adalah dapat digunakan sebagai dasar mengambil keputusan tentang kualitas keterandalan suatu butir tes performansi. Secara teoretis penelitian ini memberikan gambaran pengembangan tes performansi mekanik otomotif SMK, dan peranan penilai internal dan eksternal dalam menguji kompetensi siswa program mekanik otomotif SMK.

Kajian Literatur

Teori Respon Butir

Teori Respon Butir (TRB) banyak dipakai dalam evaluasi pembelajaran terutama membahas kaitan butir dengan respon 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 Θ). Kaitan kemampuan siswa dengan taraf sukar butir dapat dirumuskan sebagai berikut, $P_i(\Theta) = f(\Theta - b_i)$. Persamaan tersebut disebut model 1 Parameter. Dalam konteks ini, TRB dipandang sebagai teori tes probabilistik (*probabilistic test theory*), karena TRB mengadopsi model probabilitas untuk melacak parameter-parameter variabel terpendam.

Dalam perkembangannya ternyata telah digunakan pula pada penskoran polikotomi (politomi). Model TRB dengan format respons butir politomi, mencakup: model bertingkat, model nominal, model kredit parsial, dan model skala penilaian. Penggunaan TRB pada penskoran politomi menunjukkan TRB dapat dipakai pada skala penilaian tes performansi. Pada penelitian ini dipakai TRB dengan pendekatan model perluasan Rasch, model ini interpretasinya sederhana karena hanya melibatkan satu parameter, karakteristik butir dan responden lebih stabil dengan sampel kecil.

Model Rasch hanya memakai taraf kesukaran butir (b_i) yang dimasukkan ke dalam model. Model Rasch adalah suatu metode pengukuran untuk mendapatkan hal yang fundamental dari pengukuran linear dan pengamatan respons kategori. Model Rasch berkembang pada jenis tes bersifat multiaspek (*multifacets*), seperti interaksi responden (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*).

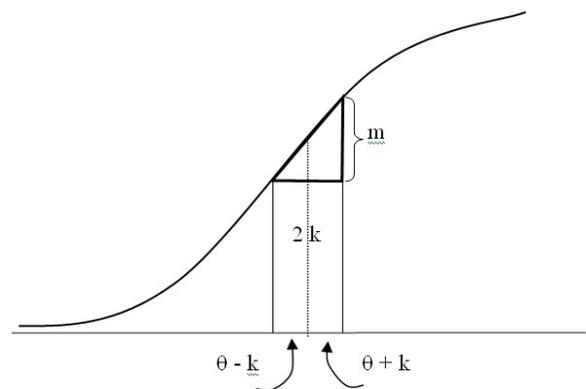
Rumus model perluasan Rasch untuk multiaspek dan pengguna skala penilaian adalah: $\text{Log}(P_{nij(k)}/P_{nij(k-1)}) = B_n - D_i - C_j - F_k$ (Linacre., Wright., dan Lunz, 1990:2). Rumus ini digunakan untuk menilai tampilan kompetensi seseorang yang dinilai semua Penilai (juri) pada semua butir (aktivitas). Formula penilaian dengan butir sama namun penilai berbeda, menjadi : $\text{Log}(P_{nij(k)}/P_{nij(k-1)}) = B_n - D_i - C_j - F_{jk}$. Model perluasan Rasch (*rasch extended*) dapat mengolah data politomi, seperti data yang berasal dari tes kompetensi mekanik otomotif. Penggunaan model Rasch dapat meningkatkan kualitas pengukuran, di antaranya analisis konsistensi respons responden terhadap butir.

Fungsi Informasi Ujian

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 (θ) berbeda mengerjakan satu butir akan menghasilkan kurva karakteristik butir seperti pada Gambar 1.



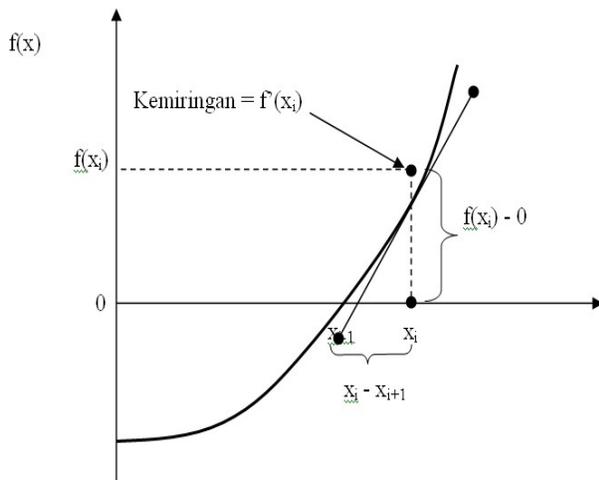
Gambar 1. Slope kurva karakteristik butir

Fungsi informasi dalam disiplin Metode Numerik dapat disebut fungsi kemiringan $f'(i)$, di mana: $R_o = f'(i)$ h. Variansi pada skor tes dikotomi adalah $P_i(\theta).Q_i(\theta)$. Slope dan variansi tergantung pada parameter butir. Metode Newton-Rapshon dapat menjelaskan berdasarkan tafsiran geometris pada Gambar 2. Garis singgung pada fungsi x_i adalah $f'(x_i)$ diekstrapolasi ke bawah sumbu x untuk memberikan taksiran akar di x_{i+1} .

Rumus Newton-Rapshon sebagai metode terbuka. Dari penurunan rumus Newton-Rapshon, diketahui bahwa kesalahan (galat) sama dengan diskrepansi (ketidakcocokan) antara x_{i+1} dan nilai sejati x_r , yaitu $E_{t,i+1} = x_r - x_{i+1}$. Pemakaian Metode Newton-Rapshon berguna untuk mengungkapkan kesalahan pengukuran, tetapi mempunyai kelemahan, di antaranya adalah kekonvergenan lambat karena sifat alami dari fungsi tersebut dan adanya titik balik (*inflection point*).

Gambar 2 menjelaskan turunan pertama di x_i setara dengan kemiringan:

$$f'(x_i) = \frac{f(x_i) - 0}{x_i - x_{i+1}}$$



Gambar 2. Pelukisan grafis metode Newton-Rapshon

Fungsi informasi ditinjau dari distribusi probabilitas penyampelan normal, kejadian kebetulan ketidakpastian dari sejumlah pengukuran tertentu cenderung berdistribusi normal (Dali S Naga, 2001: XIV-10). Pengukuran berulang-ulang tidak selalu sama besar, 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. Bentangan penyebaran nilai hasil pengukuran disebut kekeliruan baku. Dengan demikian, dapat dikatakan bahwa variansi menggambarkan ketepatan suatu alat ukur.

Fungsi informasi merupakan hubungan terbalik variansi, sedangkan variansi merupakan wujud lain daripada ketidakpastian (Dali S Naga, 2001:309). Rumus fungsi informasi butir untuk butir ke-i ditulis:

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

Rumus di atas menunjukkan kekeliruan baku pada estimasi parameter dihitung melalui kebolehjadian maksimum. 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 dapat 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.

Fungsi Informasi Ujian Skala Penilaian Tanpa Pembobotan

Pembahasan fungsi informasi ujian harus diawali dari pembahasan informasi butir. Rumus fungsi informasi butir satu parameter dituliskan,

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

Fungsi informasi tugas merupakan penjumlahan fungsi informasi aktivitas. Pada SPTP rumus penskoran ujian adalah

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

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, 2001:310).

Fungsi Informasi Ujian Skala Penilaian Pakai Pembobotan

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), berupa penjumlahan dari skor setiap aktivitas dengan memperhitungkan besar bobot masing-masing aktivitas. Rumus penskoran pada SPPP memakai ragam kedua, yaitu penggabungan skor butir dengan pembobotan atau butir-butir dengan bobot yang berbeda-beda (Dali S Naga, 2001:329). Rumus penskoran ujian dengan pembobotan (Lord dan Norvick, 1968:96), yaitu:

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

Sejalan dengan rumus penskoran, maka rumus untuk fungsi informasi ujian dengan pembobotan (Dali S. Naga, 2001:332) adalah

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

Dari rumus di atas jelas 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.

Ada beberapa cara untuk mengkonversi skala penilaian tertentu ke bentuk skala penilaian lain, seperti: perhitungan manual, penyetaraan, dan skor baku. Pada penelitian ini, sejalan dengan prosedur model Rasch, konversi yang dipakai adalah skor baku (*z*). 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

Penilai dalam istilah bahasa Inggris lainnya dikenal dengan sebutan *assessor, rater, marker* dan *judge*. Kesemua istilah dapat saling dipertukarkan pemakaiannya. Dalam Penilaian (*assessment*) sistem pendidikan, ada istilah yang sama artinya, tapi berbeda peran dan otoritasnya, yaitu *verifier* dan *assessor*. 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 (*outsider*) 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). Secara teoritis, konfigurasi dari kombinasi kedua jenis penilai dapat berupa: internal untuk internal, internal untuk eksternal, eksternal untuk internal, dan eksternal untuk eksternal. Untuk ujian praktek, konfigurasi yang dipakai adalah 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 (dunia usaha/dunia industri), yang secara umum berfungsi sebagai pengendali kualitas (*quality control*) 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 (sistematika dan cara kerja), sikap dan waktu yang digunakan untuk menyelesaikan pekerjaan (tugas).

Ragam Tes Performansi

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-elemen ini menunjukkan, tes performansi tidak dapat dilepaskan dari suatu perbuatan (kerja). Beberapa penulis menggandeng kata performansi dengan kata tugas, sehingga dibaca tugas performansi (*performance task*).

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,

Penilaian performansi berfokus pada prosedur bila a) bukan produk, atau evaluasi produk tidak memungkinkan, b) prosedur rapi dan dapat diobservasi langsung, c) ketepatan prosedur penting untuk sukses kemudian hari, d) analisis langkah prosedural dapat membantu dalam memperbaiki suatu produk. Dalam beberapa hal evaluasi produk menjadi fokus penilaian performansi bila a) prosedur berbeda namun kualitas sama, b) prosedur tidak layak diobservasi, c) langkah prosedural telah dikuasai, d) produk mempunyai kualitas yang dapat diidentifikasi dan dinilai (Gronlund,1993: 117-118).

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

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

Pembuatan cara penilaian 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).

Penelitian Yang Relevan

Penelitian yang dilakukan oleh Sherron (2000) menemukan kemampuan siswa sesuai dengan model pengukuran skala penilaian ALCP (*Adaptive Leadership Competency Profile*). Rentang logit terlebar yang dimiliki skala penilaian ALCP adalah pada dimensi inisiatif, dengan skor terendah - 4,76 dan tertinggi 7,11 yang dinilai oleh 34 juri.

Penelitian dilakukan Tay Lim, Tang, Davis dan Tang (24 April 2003) tentang pengaruh perlakuan butir pada skor skala laporan NAEP, ditemukan bahwa fungsi informasi strategi moderat lebih tinggi daripada strategi pemurah dan strategi keras. Rochaya (2001) menemukan bahwa fungsi informasi butir soal bentuk pilihan ganda dengan

lima pilihan jawaban/option lebih tinggi daripada nilai fungsi informasi butir soal bentuk pilihan ganda dengan tiga pilihan jawaban.

Studi dilakukan Webb (de Gruijter, Dato N.M. dan Kamp, L.J.T.V. 2002: 45-46), dimana penelitian tersebut tentang pengukuran performansi pekerjaan dengan mengaplikasikan teori generalizabiliti. Pada penelitian tersebut, penilaian performansi pekerja dinilai tiga kelompok penilai, yaitu supervisor, peer, dan diri sendiri). Pada penelitian ini penilai menilai masinis pada tiga jenis pekerjaan, yaitu ujian tampilan kecakapan tangan, tes pengetahuan kerja, dan penilaian umum. Temuan mereka terdapat perbedaan koefisien reliabilitas peserta, penilai, tugas (*task*), dan faktor interaksi peserta dengan penilai, peserta dengan tugas, dan penilai dengan tugas.

Metodologi Penelitian dan Desain Penelitian

Penelitian ini merupakan penelitian eksperimen, eksperimen dilakukan pada butir tes, yaitu dengan perlakuan mengempiskan dan menggelembungkan bobot butir. Penelitian eksperimen pada butir telah pernah dilakukan seperti de Gruijter.

Berdasarkan pada teori generalizabiliti 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*).

Prosedur Penelitian

Dalam melakukan eksperimen, ada beberapa tahapan yang akan dilalui: Proses pra-eksperimen, meliputi: 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 sekor 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, senadainya ditemukan kelemahan di lapangan; 9) pada langkah 1 - 7 dibuat laporan ke promotor.

Proses eksperimen dilakukan dengan kegiatan pendahuluan. Kegiatan pendahuluan ini mencakup: 1) menentukan kriteria penilai internal dan eksternal; seperti: pengalaman paling tidak 5 tahun dalam profesinya, berpenglihatan bagus; 2) pelatihan penilai, seperti: cara menggunakan kedua jenis tes dan lainnya; 3) menentukan jadwal eksperimen; 4) penggandaan 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.

Selanjutnya, kegiatan 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 dilakukan dengan format sumatif, yaitu diakhir kelas II. Walaupun dalam kondisi sumatif, namun pengukuran dilakukan berulang-ulang, yaitu dalam bentuk beberapa tugas yang dipraktekkan siswa. Proses pasca-eksperimen: 1) lembaran tes diparaf penilai; 2) lembaran tes dikumpulkan segera; 3) lembaran tes yang cacat disortir; 4) data pada lembaran 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 (ketunggalan-matra) 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 (Rasch); dan 6) desain generalizabiliti, untuk menganalisis banyak faktor (*facet*), seperti: penilai, tugas, dan kemampuan siswa, dikenal dengan sebutan *multifacet*.

Tabel 1. Desain penelitian anava faktorial dua jalur

Kelompok Penilai (B)	Ragam tes Performansi (A)	
	SPTP(A ₁)	SPP(A ₂)
Internal (B ₁)	A ₁ B ₁	A ₂ B ₁
Eksternal (B ₂)	A ₁ B ₂	A ₂ B ₂

Populasi dan Sampel Penelitian

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 siswa SMK negeri sebanyak 25.678 dan siswa pada swasta sebanyak 29.496, sehingga total siswa SMK di Sumatera barat sebanyak 55.174. Jumlah populasi siswa kelas II secara pasti tidak dapat diketahui, menurut pejabat dinas Diknas ini terjadi disebabkan otonomi daerah tingkat II, sehingga tidak selalu semua daerah tingkat II melaporkan jumlah siswa di daerahnya, namun dapat diperkirakan sekitar 30% dari jumlah keseluruhan populasi siswa SMK berarti sekitar 16.552 orang. Sampel siswa adalah siswa kelas II program keahlian mekanik otomotif pada 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.

Populasi tugas berupa populasi universal, yaitu seluruh tugas mekanik otomotif. Sampel tugas adalah tugas-tugas (sub-sub kompetensi) yang terdapat pada kurikulum SMK program keahlian mekanik otomotif kelas II (dua).

Teknik Pengambilan Sampel

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 yaitu: 1) legimitas pelaksanaan ujian, ke-17 sekolah telah mendapai izin dari Dinas Diknas Sumbar melaksanakan Ujian kompetensi secara mandiri; (2) kelengkapan mesin dan peralatan memenuhi persyaratan pelaksanaan ujian kompetensi.

banyak mendapatkan tugas adalah tujuh dan paling sedikit empat tugas.

Teknik penempatan subyek pada sub-matrik juga menggunakan random acak sederhana, yaitu secara berurutan setiap subyek diambil dari kerangka sampel sementara dan bersamaan dengan itu diaktifkan nomor random Excel yang menampilkan satu set angka random. Tiga digit (angka) terakhir dijadikan referensi nomor defenitif subyek. Dari pemerosesan ini, matriks sampel terisi nama-nama subyek ke-1 sampai subyek ke-432. Dari matriks tampak setiap subyek mendapatkan beberapa tugas yang diujikan padanya. Setiap siswa diuji penilai internal dan penilai eksternal. Dalam desain matriks sampel tampak bahwa, penilai tersarang dalam ragam skala penilaian.

Uraian di atas menjelaskan ada tiga teknik sampling dipakai: pertama, teknik purposif untuk

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 Padang	Swasta	147
6	SMK Muhammadiyah Solok	Swasta	112
7	SMK Mitra Payakumbuh	Swasta	142
Jumlah			864

Tahap ke dua pengambilan sampel, diawali pembuatan penomoran siswa, kemudian mengambil subyek dengan teknik sampel random sederhana, yaitu mengundi nomor subyek. Dari proses ini didapat dua kerangka sampel sementara, yaitu sub kelompok dinilai dengan SPTP (A_1) sebanyak 432 siswa dan sub kelompok dinilai SPPP (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, lihat lampiran 2. Dengan demikian satu siswa paling

menentukan sekolah; kedua, teknik random sederhana untuk pengambilan subyek dan membentuk dua kelompok, ketiga teknik random matriks sampling untuk penempatan subyek ke kerangka matriks sampling.

Analisis Data

Data dianalisis dengan teknik analisis deskriptif untuk mendapatkan sekor rerata, modus, median; ini disebut sekor ukuran kecenderungan memusat. Pencaran (keragaman) data mencakup varian dan simpangan baku (SD). 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
Persyaratan Analisis Varian

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 independensi lokal, karena t hitung kurang dari t tabel ($\div^2_{hit.tgs} < \div^2_{(0,05)(5)}$). 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 disampaikan pada Tabel 3.

Deskripsi Data Fungsi Informasi

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

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.

Selain grafik fungsi informasi, ada juga grafik kurva probabilitas, yaitu kurva probabiliti. Sesuai dengan kategori yang dibuat, yaitu: Kompetensi tak ada, Kompetensi kurang, Kompetensi minimal, Kompetensi bagus. Selanjutnya dapat dilihat

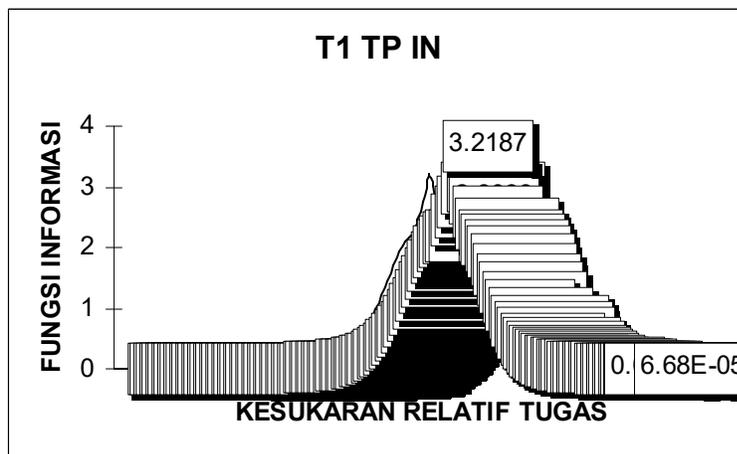
Tabel 3. Hasil Uji Normalitas Data Fungsi Informasi

Kelompok Data	df	Statistik	D _{tabel}	Signifikansi	Simpulan
1. Kelompok A ₁	92	0,087	0,188	0,083	Normal
2. Kelompok A ₂	92	0,061	0,188	0,200	Normal
3. Kelompok B ₁	92	0,082	0,188	0,169	Normal
4. Kelompok B ₂	92	0,090	0,188	0,064	Normal

Dari Tabel 3 dapat 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 H₀ dapat diterima; varians sampel populasi adalah homogen.

probabilitas setiap kategori sebagai berikut.

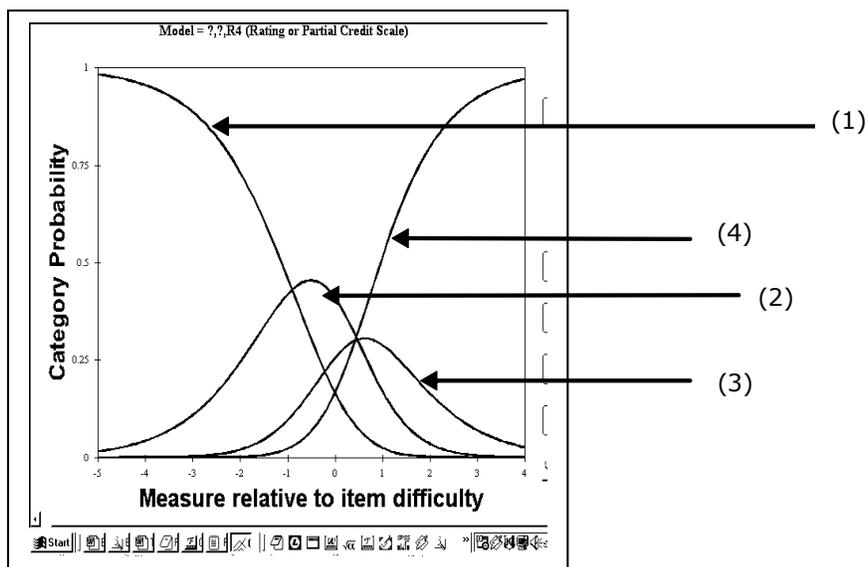
Ada empat kategori dari setiap tugas yang kemungkinan dapat terjadi saat seorang penilai memberi nilai pada siswa. Pada gambar 4.17, dimulai dari hadap kiri, grafik paling kiri, yaitu: 1) adalah kategori kompetensi tak ada, grafik; 2)



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

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 4.17 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 2), yaitu kompetensi minimal. Pengkategorian dalam empat kategori dilakukan pada semua tugas (46) yang diujikan pada siswa.

Dari hasil pengolahan data di atas dapat disampaikan rata-rata terendah ada pada sel A_1B_1 , sebesar 2,68, dan rata-rata tertinggi ada pada sel A_2B_2 , 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_1B_1 , yang semula sebesar 3,44 menjadi sebesar 3,20. Kesemua perubahan rata-rata fungsi informasi pada sel-sel tersebut disebabkan oleh perlakuan dari ragam tes SPTP ataupun SPPP.



Gambar 4. Kurva probabilitas tugas 38 dengan SPTP oleh penilai eksternal.

Secara umum data fungsi informasi ujian disampaikan pada Tabel 4.

Tabel 4. Deskripsi Data Fungsi Informasi

Sumber Statistik		A_1	A_2	Σ
B_1	n	46	46	92
	\bar{X}	3,44	3,20	3,32
	s	0,74	0,57	0,62
B_2	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

Pengujian Hipotesis Statistik

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 5.

Tabel 5. Rangkuman Hasil Anava Data Fungsi Informasi

SUMBER VARIAN	JK	dk	RJK	F ₀	F _t	
					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 jauh lebih tinggi daripada SPTP.

Pada perbedaan antarbaris harga $F_0 = 70,118 > F_t = 3,84$, berarti hipotesis nol (H_0) yang menyatakan fungsi informasi siswa yang dinilai oleh penilai internal tidak berbeda dengan siswa yang dinilai oleh penilai eksternal ditolak, atau hipotesis alternatif (H_a) yang menyatakan fungsi informasi 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 ($x_{B1} = 3,319 < x_{B2} = 3,606$).

Pada interaksi (kolom dan baris) harga $F_0 = 117,989 > F_t = 3,84$, berarti 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 dapat diterima. Dengan demikian, dapat dikatakan bahwa pencapaian fungsi informasi dipengaruhi oleh interaksi antara ragam skala penilaian yang dipakai dalam ujian dengan kelompok penilai.

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 (*honestly significant difference*). Analisis ini merupakan analisis lanjut, yang akan mengungkapkan apakah ada perbedaan antarsel. Hasil uji Tukey pada taraf signifikansi (α)=0,05 dirangkum pada tabel 6.

Tabel 6. Hasil ANAVA Tahap Lanjut dengan Uji Tukey

Kelompok yang diperbandingkan	Harga perbedaan rerata absolut (q_0)	dk	Harga Kritis	Keterangan
$A_1B_1 - A_2B_1$	11,53	180	0,18	Signifikan
$A_1B_2 - A_2B_2$	28,07	180	0,18	Signifikan
$A_1B_1 - A_1B_2$	0,282	180	0,18	Signifikan
$A_2B_1 - A_2B_2$	27,91	180	0,18	Signifikan

Pada kelompok A_1B_1 dan A_2B_1 harga $q_0 = 11,53 > q_t 0,18$, berarti hipotesis nol (H_0) ditolak atau hipotesis alternatif keempat diterima, hal itu berarti pada fungsi informasi tugas SPTP dengan penilai internal lebih tinggi daripada fungsi informasi tugas (SPPP) dengan penilai internal. Pada kelompok A_1B_2 dan A_2B_2 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.

Pada kelompok A_1B_1 dan A_1B_2 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.

Pada kelompok A_2B_1 dan A_2B_2 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.

Simpulan dan Saran

Simpulan

Penelitian ini menggunakan metode eksperimen yang melibatkan variabel bebas ragam tes performansi dan kelompok penilai, dan variabel terikat fungsi informasi ujian. Ragam tes performansi yang dimanipulasi dibagi menjadi SPTP dan SPPP. Kelompok penilai sebagai variabel bebas dibagi menjadi penilai internal dan penilai eksternal, fungsi informasi sebagai variabel terikat didapat setelah melalui proses komputasi data kompetensi mekanik otomotif dengan model perluasan Rasch.

Data awal kompetensi mekanik otomotif yang diolah dengan *software facet* (model perluasan Rasch) menghasilkan data fungsi informasi. Selanjutnya, data fungsi informasi tugas dianalisis dengan menggunakan Anava dan diteruskan dengan uji Tukey (untuk melihat signifikansi interaksi). Berdasarkan data yang diperoleh, hasil pengujian hipotesis, dan pembahasan hasil penelitian 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 daripada menggunakan SPTP.

Saran

Untuk Kepada Sekolah

Sebagai suatu temuan dalam pengukuran SPPP oleh penilai eksternal diperlukan langkah-langkah dalam menerapkannya. Oleh karena itu, penerapan SPPP oleh penilai eksternal memerlukan persiapan, penjadwalan, materi uji (bank soal), rancangan matriks penilai internal dan eksternal, tata letak, dan fasilitas mesin dan peralatan yang akan dipakai siswa untuk tampil dalam ujian. Penilai eksternal tetap dipakai dalam sistem penilaian sekolah, penilai eksternal tidak hanya sebagai penilai (*assessor*) tetapi ditingkatkan otoritasnya menjadi verifier sekolah, sehingga terlibat dalam perencanaan dan proses pembelajaran terutama dalam penilaian hasil pembelajaran. Dengan terlibatnya secara langsung penilai eksternal dalam proses pendidikan, diharapkan kriteria-kriteria penilaian kerja (*performance assessment*) yang terjadi di lapangan kerja dan terserap dengan cepat oleh pihak sekolah.

Untuk Guru

Guru perlu melaksanakan penilaian dengan lebih banyak kriteria kompetensi mekanik otomotif. Dengan SPPP pencapaian tujuan evaluasi lebih optimum, karena di samping aspek materi uji, juga memberikan kejelasan berupa tingkat ketelitian dan ketepatan dalam mengukur kompetensi mekanik otomotif. Selain itu, guru perlu mengembangkan tes performansi dengan mempertimbangkan karakteristik penilai, aktivitas dan siswa; dengan harapan akan mendapatkan hasil penilaian yang objektif, adil dan jujur.

Untuk Pakar dan Peneliti Pendidikan Kejuruan

Berdasarkan temuan bahwa skala penilaian pakai pembobotan (SPPP) mampu meningkatkan kualitas evaluasi dalam menilai keluaran program keahlian mekanik otomotif SMK. Hal tersebut dapat digunakan sebagai informasi secara empirik yang dapat digunakan sebagai acuan bagi para ahli dan peneliti pendidikan kejuruan khususnya teknik mekanik otomotif di SMK.

Melalui kegiatan penelitian secara terprogram dan terstruktur para ahli dan peneliti bidang kejuruan khususnya teknik mekanik otomotif disarankan untuk merancang berbagai alat ukur yang sesuai dengan karakteristik materi mekanik

otomotif untuk semua tingkat pendidikan. Di samping itu temuan tersebut perlu dipertimbangkan, dalam menentukan rancangan matriks penilai internal dan eksternal. Karena itu, untuk mendapatkan data empirik dan pengetahuan yang lebih luas tentang efektivitas SPPP dalam ujian promosi atau ujian kompetensi, perlu dilakukan penelitian lanjutan. Selain itu, dapat dilakukan penambahan variabel lainnya yang berkaitan dengan usaha peningkatan kualitas evaluasi pendidikan, seperti: variabel pembelajaran tuntas, variabel batas lulus ganda, dan peningkatan pemakaian model parameter teori respon butir.

Pustaka Acuan

- Dali S Naga. 2001. Diktat 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.
- Gronlund, N. E., 1993. *How to Make achievement Test and Assessment*. Boston: Allyn and Bacon.
- Hambleton, R.K., Swaminathan, H., Rogers, H.J. 1991. *Fundamentals of Item Response Theory*. Newbury Park: SAGE Publications.
- 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.
- 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 Score*, makalah pada National Council on Measurement in Education, 22-24 April 2003. Chicago: Educational Testing Service. www.ets.org/legal/copyright.
- Undang-Undang Nomor 20 tahun 2003. Sistem Pendidikan Nasional.