

Keakuratan Hasil Analisis Butir Menurut Teori Tes Klasik dan Teori Respons Butir Ditinjau dari Ukuran Sampel

Kana Hidayati
Jurusan Pendidikan Matematika FMIPA UNY

ABSTRAK

Penelitian ini bertujuan untuk mengungkapkan: (1) hasil analisis butir menurut teori tes klasik dan teori respons butir ditinjau dari berbagai ukuran sampel, (2) besarnya kesalahan pengukuran (SEM), dan (3) ukuran sampel yang akurat untuk analisis butir menurut teori tes klasik dan teori respons butir.

Populasi penelitian ini adalah respons siswa peserta Ujian Akhir Nasional mata pelajaran Matematika SLTP di propinsi DIY tahun pelajaran 2001/2002 yang tersebar dalam 436 SLTP dengan keseluruhan peserta berjumlah 42.574 siswa. Sampel penelitian diambil dengan teknik *random systematic sampling*. Ukuran sampel dibedakan dalam dua kategori yakni besar (>500) dan kecil (≤ 500). Analisis butir untuk ukuran sampel kategori besar dilakukan pada ukuran sampel 600, 750, dan 1000, sedangkan untuk ukuran sampel kategori kecil dilakukan pada ukuran sampel 100, 250, dan 500. Kegiatan analisis butir menurut teori tes klasik dan teori respons butir dilakukan dengan program *Bilog* versi 3.07. Besarnya harga indeks reliabilitas dihitung dengan program SPSS versi 10.0 dan harga fungsi informasi tes dihitung dengan menggunakan program EXCEL 2000.

Hasil analisis butir menurut teori tes klasik untuk kategori ukuran sampel besar menunjukkan bahwa pada ukuran sampel 600, 31 butir baik, indeks reliabilitasnya 0,7732, dan SEM 2,8672; ukuran sampel 750, 31 butir baik, indeks reliabilitasnya 0,6341, dan SEM 2,9342; ukuran sampel 1000, 29 butir baik, indeks reliabilitasnya 0,6945, dan SEM 3,2951. Untuk kategori ukuran sampel kecil menunjukkan bahwa pada ukuran sampel 100, 33 butir baik, indeks reliabilitasnya 0,8729, dan SEM 2,6670; ukuran sampel 250, 34 butir baik, indeks reliabilitasnya 0,8532, dan SEM 2,1342; ukuran sampel 500, 32 butir baik, indeks reliabilitasnya 0,8921, dan SEM 2,3278. Menurut teori respons butir, untuk kategori ukuran sampel besar menunjukkan bahwa pada ukuran sampel 600, 33 butir baik, fungsi informasinya 0,4237, dan SEM 1,6479; ukuran sampel 750, 36 butir baik, fungsi informasinya 0,5496, dan SEM 1,1489; ukuran sampel 1000, 35 butir baik, fungsi informasinya 0,5642, dan SEM 1,3423. Untuk kategori ukuran sampel kecil menunjukkan bahwa pada ukuran sampel 100, 31 butir baik, fungsi informasinya 0,5496, dan SEM 1,6731; ukuran sampel 250, 34 butir baik, fungsi informasinya 0,5496, dan SEM 1,7349; ukuran sampel 500, 34 butir baik, fungsi informasinya 0,4315, dan SEM 1,9241. Dengan demikian dapat disimpulkan bahwa analisis butir menurut teori tes klasik akan akurat apabila menggunakan ukuran sampel kecil dan menurut teori respons butir akan akurat apabila menggunakan ukuran sampel besar. Secara keseluruhan SEM yang paling kecil diperoleh jika menggunakan teori respons butir.

Kata kunci: analisis butir, teori tes klasik, teori respons butir, ukuran sampel.

Pendahuluan

Pelaksanaan pengukuran di bidang pendidikan pada prinsipnya bertujuan untuk mengetahui karakteristik suatu objek seperti kemampuan, keberhasilan belajar, sikap, minat atau ciri terpendam lainnya yang terdapat pada peserta didik namun tidak kelihatan dan tidak dapat diukur langsung. Untuk mengukur berbagai karakteristik yang terpendam itu sangat diperlukan alat ukur yang baik sehingga mampu mengungkap secara benar ciri terpendam pada peserta didik. Alat ukur yang baik adalah alat ukur yang memenuhi persyaratan dan mampu menghasilkan informasi yang mengandung kesalahan sekecil mungkin.

Alat ukur yang digunakan dalam bidang pendidikan biasanya berupa tes dan non tes. Selain itu, pengukuran juga bisa ditempuh melalui observasi atau wawancara. Berkaitan dengan alat ukur berupa tes yang terdiri dari sejumlah butir yang sering digunakan dalam pengukuran pendidikan harus dianalisis untuk mengetahui kualitas alat ukur tersebut. Kegiatan analisis butir dilakukan berdasarkan suatu teori dengan asumsi-asumsi yang mendasarinya. Secara umum teori yang kuat adalah teori yang menggunakan asumsi yang banyak sehingga menghasilkan informasi yang lebih teliti.

Ada dua teori pengukuran pendidikan yang saat ini berkembang dan banyak digunakan dalam merancang dan menganalisis suatu tes. Teori yang pertama adalah teori tes klasik yang dikembangkan sejak tahun 1940 dan telah digunakan secara luas, kedua adalah teori respons butir yang menggunakan lebih banyak asumsi dibandingkan teori pertama. Teori tes klasik telah banyak berjasa dalam bidang pengukuran dan bahkan masih digunakan sampai sekarang. Namun demikian, oleh karena teori tes klasik memiliki berbagai keterbatasan dengan adanya sifat *group dependent* dan *item dependent* maka munculnya teori respons butir menjadi sangat berguna dan terus dikembangkan karena mampu mengatasi keterbatasan tersebut (Hambleton, Swaminathan & Rogers, 1991: 2-5).

Ditinjau dari banyaknya asumsi, teori respons butir memang diharapkan akan memberikan informasi yang lebih teliti dibandingkan teori tes klasik. Namun demikian, berkaitan dengan ukuran sampel, baik teori tes klasik maupun teori respons butir tidak mensyaratkan ukuran sampel secara pasti (Crocker L. &

Algina J., 1986:322). Dengan demikian penggunaan kedua teori ini tergantung kepada pemakainya. Setiap teori tentu memiliki kelemahan dan kelebihan yang diantaranya dapat dilihat dari sudut kemudahan penggunaan dan pemahaman, ketepatan analisis, dan biaya yang digunakan. Oleh karena itu penelitian tentang keakuratan hasil analisis butir menurut teori tes klasik dan teori respons butir dengan menggunakan ukuran sampel yang berbeda ini sangat diperlukan untuk menunjang efektifitas kegiatan analisis butir.

Parameter yang digunakan pada analisis butir berdasarkan teori tes klasik dan teori respons butir pada dasarnya adalah sama yaitu tingkat kesukaran, daya pembeda, tebakan semu (*pseudo guessing*), dan kemampuan. Perbedaannya terletak pada formula, skala, dan satuan yang digunakan. Selain itu, analisis butir suatu tes dengan teori tes klasik dan teori respons butir pada prinsipnya juga dilakukan untuk menaksir kemampuan seseorang yang diharapkan memiliki kesalahan sekecil mungkin. Kesalahan pengukuran menurut teori tes klasik dinyatakan dengan kesalahan baku pengukuran (*Standar Error of Measurement/SEM*) yang besarnya tergantung pada indeks kehandalan tes. Untuk teori respons butir kesalahan pengukuran dinyatakan dengan kesalahan baku pengukuran (*Standar Error of Measurement/SEM*) yang besarnya tergantung pada tingkat kemampuan seseorang dan fungsi informasi tes. Adanya kesalahan yang melekat pada data hasil pengukuran ini disebabkan oleh banyak faktor diantaranya adalah alat ukur itu sendiri, pelaksanaan pengukuran, objek pengukuran, dan teknik analisis yang digunakan.

Saat ini, banyak peneliti menguji instrumen yang berupa tes untuk mengumpulkan data menggunakan teori tes klasik. Hal ini karena teori ini sudah lama berkembang sehingga banyak peneliti mengetahui teori ini walau tidak secara mendalam. Penggunaan formula hanya berdasarkan kemudahan tanpa memperhatikan asumsi yang mendasarinya. Teori respons butir yang menghasilkan informasi yang lebih teliti semestinya juga menjadi acuan bagi para peneliti dalam menganalisis suatu tes. Dengan demikian estimasi karakteristik suatu objek akan lebih teliti serta kesimpulan yang dibuat dan kebijakan yang dipilih akan semakin tepat.

Untuk mengetahui keakuratan hasil analisis butir menurut teori tes klasik dan teori respons butir ditinjau dari ukuran sampel dibutuhkan data. Data yang digunakan dalam penelitian ini adalah respons siswa terhadap perangkat soal UAN SLTP mata pelajaran Matematika yang disiapkan oleh Pusat Penilaian Pendidikan mengingat berbagai ukuran sampel yang digunakan dalam penelitian ini. Permasalahan yang diajukan dalam penelitian ini adalah sebagai berikut: (1) bagaimanakah hasil analisis butir menurut teori tes klasik dan teori respons butir ditinjau dari berbagai ukuran sampel?, (2) berapakah besarnya kesalahan pengukuran (SEM)?, dan (3) bagaimanakah keakuratan hasil analisis butir menurut teori tes klasik dan teori respons butir ditinjau dari ukuran sampel?.

Penelitian ini bertujuan untuk mengungkapkan: (1) bukti empiris tentang hasil analisis butir menurut teori tes klasik dan teori respons butir ditinjau dari berbagai ukuran sampel, (2) besarnya kesalahan pengukuran (SEM), dan (3) keakuratan hasil analisis butir menurut teori tes klasik dan teori respons butir ditinjau dari ukuran sampel.

Hasil penelitian ini diharapkan memberikan wacana baru dalam kegiatan peningkatan kualitas suatu tes dan secara akademis dapat digunakan sebagai dasar atau pendukung kesimpulan awal bagi para peneliti lain untuk melakukan penelitian lanjutan (pengembangan) maupun penelitian sejenis (perluasan). Secara praktis penelitian ini diharapkan bermanfaat untuk mengetahui ketepatan teori dalam melakukan analisis butir pada ukuran sampel tertentu dan sebagai sumber informasi dalam pemilihan butir-butir soal yang layak untuk bank soal Matematika yang akan dikembangkan baik di pusat maupun di daerah.

Metode Penelitian

Penelitian ini bersifat *ex-post facto* sehingga dalam penelitian ini tidak dilakukan perlakuan apapun terhadap variabel penelitian. Penelitian dilakukan di Kantor Dinas Pendidikan propinsi DIY. Data utama dalam penelitian ini adalah respons siswa peserta UAN SLTP mata pelajaran Matematika paket utama 1 di Propinsi DIY tahun pelajaran 2001/2002. Pengumpulan data dilakukan dengan

menggunakan teknik dokumentasi yaitu dengan mengutip respons siswa peserta Ujian Akhir Nasional.

Populasi penelitian ini adalah respons siswa peserta Ujian Akhir Nasional SLTP mata pelajaran Matematika di propinsi DIY tahun pelajaran 2001/2002 yang tersebar dalam 436 SLTP dengan keseluruhan peserta berjumlah 42.574 siswa. Sampel penelitian diambil dengan teknik *random systematic sampling*. Ukuran sampel dibedakan dalam dua kategori yakni besar (>500) dan kecil (≤ 500). Analisis butir untuk ukuran sampel kategori besar dilakukan pada ukuran sampel 600, 750, dan 1000, sedangkan untuk ukuran sampel kategori kecil dilakukan pada ukuran sampel 100, 250, dan 500.

Prosedur pengambilan sampel penelitian untuk setiap ukuran sampel adalah sebagai berikut: (1) membuat daftar peserta UAN berdasarkan skor peserta yang diperoleh dengan urutan dari yang tertinggi hingga terendah, (2) menentukan ukuran interval dengan menggunakan rumus N/n , dimana N adalah jumlah seluruh siswa dan n menyatakan ukuran sampel yang diinginkan, (3) mengambil sampel dari interval pertama secara acak sederhana yaitu dengan menggunakan bilangan acak dari kalkulator, (4) setelah sampel dari interval pertama diperoleh, pengambilan sampel berikutnya dilakukan secara sistematis dengan jarak satu interval dari sampel pertama sehingga diperoleh ukuran sampel yang telah ditentukan dan apabila sampai dengan interval terakhir belum memenuhi ukuran sampel yang dikehendaki pengambilan diteruskan secara berputar ke interval pertama lagi sampai diperoleh ukuran sampel yang dikehendaki.

Kegiatan analisis data dalam penelitian ini meliputi dua hal pokok yaitu pertama, analisis butir menurut teori tes klasik dan teori respons butir dengan ukuran sampel yang telah ditentukan dan kedua, menentukan besarnya kesalahan pengukuran (SEM). Analisis butir meliputi kegiatan estimasi parameter butir dan parameter kemampuan yang dilakukan dengan menggunakan program *Bilog* versi 3.07 dengan menggunakan model logistik tiga parameter. Analisis butir menurut teori tes klasik dilakukan dengan program *Bilog* versi 3.07 fase pertama dan menurut teori respons butir dilakukan dengan program *Bilog* versi 3.07 fase kedua. Estimasi parameter kemampuan diperoleh dengan program *Bilog* versi

3.07 fase ketiga. Besarnya harga indeks reliabilitas dihitung dengan program SPSS versi 10.0 dan harga fungsi informasi tes dihitung dengan menggunakan program EXCEL 2000.

Hasil Penelitian dan Pembahasan

Berdasarkan hasil analisis butir menurut teori tes klasik yang dilakukan dengan bantuan program *Bilog* versi 3.07 fase pertama menunjukkan bahwa dari 40 butir soal yang dianalisis, untuk kategori ukuran sampel besar dan kecil masing-masing memberikan hasil yang bervariasi. Untuk ukuran sampel besar menunjukkan bahwa pada ukuran sampel 600, ada 31 butir baik dan indeks reliabilitasnya 0,7732, pada ukuran sampel 750, ada 31 butir baik dan indeks reliabilitasnya 0,6341, dan pada ukuran sampel 1000, ada 29 butir baik dan indeks reliabilitasnya 0,6945. Untuk kategori ukuran sampel kecil menunjukkan bahwa pada ukuran sampel 100, ada 33 butir baik dan indeks reliabilitasnya 0,8729, pada ukuran sampel 250, ada 34 butir baik dan indeks reliabilitasnya 0,8532, pada ukuran sampel 500, ada 32 butir baik dan indeks reliabilitasnya 0,8921.

Berdasarkan hasil analisis butir menurut teori respons butir yang dilakukan dengan bantuan program *Bilog* versi 3.07 fase kedua menunjukkan bahwa dari 40 butir soal yang dianalisis, untuk kategori ukuran sampel besar dan kecil masing-masing juga memberikan hasil yang bervariasi. Untuk kategori ukuran sampel besar menunjukkan bahwa pada ukuran sampel 600, ada 33 butir baik dan fungsi informasinya 0,4237, pada ukuran sampel 750, ada 36 butir baik dan fungsi informasinya 0,5496, pada ukuran sampel 1000, ada 35 butir baik dan fungsi informasinya 0,5642. Untuk kategori ukuran sampel kecil menunjukkan bahwa pada ukuran sampel 100, ada 31 butir baik dan fungsi informasinya 0,5496, pada ukuran sampel 250, ada 34 butir baik dan fungsi informasinya 0,5496, pada ukuran sampel 500, ada 34 butir baik dan fungsi informasinya 0,4315.

Besarnya kesalahan pengukuran menurut teori tes klasik menunjukkan bahwa untuk kategori ukuran sampel besar pada ukuran sampel 600 memiliki SEM 2,8672, pada ukuran sampel 750 memiliki SEM 2,9342, pada ukuran sampel

1000 memiliki SEM 3,2951. Untuk kategori ukuran sampel kecil menunjukkan bahwa pada ukuran sampel 100 memiliki SEM 2,6670, pada ukuran sampel 250 memiliki SEM 2,1342, dan pada ukuran sampel 500 memiliki SEM 2,3278.

Menurut teori respons butir, untuk kategori ukuran sampel besar menunjukkan bahwa pada ukuran sampel 600 memiliki SEM 1,6479, pada ukuran sampel 750 memiliki SEM 1,1489, dan pada ukuran sampel 1000 memiliki SEM 1,3423. Untuk kategori ukuran sampel kecil menunjukkan bahwa pada ukuran sampel 100 memiliki SEM 1,6731, pada ukuran sampel 250 memiliki SEM 1,7349, dan pada ukuran sampel 500 memiliki SEM 1,9241.

Menurut teori tes klasik suatu alat ukur berupa tes dikatakan baik apabila memiliki banyak butir yang termasuk dalam kategori baik, harga indeks reliabilitas tinggi, dan SEM yang kecil. Secara keseluruhan berdasarkan hasil analisis butir menurut teori tes klasik banyaknya butir baik yang lebih banyak, harga indeks reliabilitas yang lebih besar, dan SEM yang lebih kecil akan dicapai apabila menggunakan ukuran sampel kecil walaupun nilainya bervariasi untuk ukuran sampel masing-masing kategori.

Menurut teori respons butir suatu alat ukur berupa tes dikatakan baik apabila memiliki banyak butir yang termasuk dalam kategori baik, harga fungsi informasi tes tinggi, dan SEM yang kecil. Berdasarkan hasil analisis di atas ternyata analisis butir menurut teori respons butir justru memberikan hasil yang berlawanan dengan teori tes klasik. Secara keseluruhan banyaknya butir baik yang lebih banyak, harga indeks reliabilitas yang lebih besar, dan SEM yang lebih kecil diperoleh jika menggunakan ukuran sampel besar walaupun nilainya bervariasi pada ukuran sampel masing-masing kategori.

Dengan demikian dapat disimpulkan bahwa analisis butir menurut teori tes klasik akan akurat apabila menggunakan ukuran sampel kecil dan menurut teori respons butir akan akurat apabila menggunakan ukuran sampel besar. Secara keseluruhan SEM yang paling kecil diperoleh jika menggunakan teori respons butir. Walaupun kriteria untuk menentukan kategori ukuran sampel bersifat relatif namun hasil penelitian ini diharapkan dapat dijadikan acuan sesuai dengan jenis data yang digunakan.

Berkaitan dengan data yang digunakan, secara keseluruhan perangkat soal UAN SLTP mata Pelajaran Matematika ini termasuk dalam kategori perangkat soal yang baik dan dapat dipercaya. Hal ini ditunjukkan dengan banyaknya butir soal yang termasuk kategori baik dan layak dimasukkan ke dalam bank soal. Kualitas perangkat soal UAN SLTP yang baik ini diantaranya dapat disebabkan karena perangkat soal ini memang telah disiapkan secara matang oleh Pusat Penilaian Pendidikan. Perangkat soal yang akan diujikan telah disusun melalui penelaahan baik secara kualitatif maupun kuantitatif (Tim Sisjia, 1997: 2).

Simpulan

Berdasarkan hasil penelitian dan pembahasan dapat diambil simpulan bahwa secara umum untuk analisis butir menurut teori tes klasik memberikan hasil berupa banyaknya butir baik yang lebih banyak, harga indeks reliabilitas yang lebih besar dan SEM yang lebih kecil pada ukuran sampel kategori kecil (≤ 500). Sebaliknya, analisis butir menurut teori respons butir memberikan hasil berupa banyaknya butir baik yang lebih banyak, harga fungsi informasi tes yang lebih tinggi, dan SEM yang lebih kecil pada ukuran sampel kategori besar (> 500). Secara keseluruhan SEM yang paling kecil diperoleh jika menggunakan teori respons butir. Dengan demikian, analisis butir menurut teori tes klasik akan akurat apabila menggunakan ukuran sampel kecil dan menurut teori respons butir akan akurat apabila menggunakan ukuran sampel besar.

Saran

Berdasarkan hasil penelitian diajukan saran-saran sebagai berikut:

1. Kepada penyusun dan pengembang suatu tes baik dalam skala lokal maupun nasional sebaiknya menggunakan soal-soal yang secara empiris terbukti berkualitas baik.
2. Pembuktian secara empiris terhadap kualitas suatu butir soal perlu lebih dikembangkan dan disosialisasikan oleh berbagai praktisi pendidikan misalnya dengan mengadakan berbagai pelatihan bagi guru-guru baik bidang studi Matematika maupun lainnya untuk dapat melakukan analisis empiris

terhadap perangkat soal yang digunakan baik menurut teori tes klasik maupun teori respons butir.

3. Untuk menganalisis butir dengan ukuran sampel besar seperti UAN sebaiknya menggunakan teori respons butir dan untuk keperluan tes di kelas dengan ukuran sampel kecil sebaiknya menggunakan teori tes klasik.
4. Perlu adanya penelitian lebih lanjut dengan menggunakan data berbeda untuk mengetahui konsistensi keakuratan hasil analisis butir menurut teori tes klasik dan teori respons butir ditinjau dari ukuran sampel.

DAFTAR PUSTAKA

- Allen, M. J., & Yen, W. M. (1979). *Introduction to measurement theory*. Belmont, CA: Wadsworth, Inc.
- Anastasi, A. & Urbina, S. (1997). *Psychological testing*. Upper Saddle River, NJ: Prentice Hall.
- Crocker L. & Algina J. (1986). *Introduction to Classical and Modern test Theory*. New York: CBS college publishing.
- Ebel, Robert L. (1972). *Essential of educational measurement and evaluating in education and psychology*. New York: Holt, Rine hart, and Winston. Inc.
- Gronlund, N. E. (1976). *Measurement and evaluating in teaching*. New York: Macmillan Publishing Co.
- Hambleton, R.K. & Swaminathan H. (1985). *Item response theory: principles and applications*. Boston, MA: Kluwer Inc.
- Hambleton, R.K., Swaminathan H. & Rogers, H.J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc.
- Hullin, C. L., et. al. (1983). *Item response theory: Application to psychological measurement*. Homewood, IL: Dow Jones–Irwin
- Mardapi, Djemari. (1991). Konsep dasar teori respons butir: Perkembangan dalam pengukuran pendidikan. *Cakrawala Pendidikan* 3(X). 1-16..
- Mardapi, Djemari. (1998). Analisis butir soal dengan teori tes klasik dan teori respons butir. *Jurnal kependidikan*. Edisi khusus Dies-tahun XXVIII.

- Mardapi Djemari. (1999). Estimasi kesalahan pengukuran dalam bidang pendidikan dan implikasinya pada ujian nasional. Pidato pengukuhan Guru Besar. Universitas Negeri Yogyakarta.
- Mehrens, W. A. & Lehman, I. J. (1973). *Measurement and evaluation in education and psychology*. New York: Hold, Rinehart and Winston , Inc.
- Mislevy, R. J., & Bock, R. D. (1990). *BILOG 3: Item analysis & test scoring with binary logistic models* (Computer program). Mooresville, IN: Scientific Software, Inc.
- Mislevy, R.J., & Wingersky, M. S., & Sheehan, K. M. (1994). *Dealing with uncertainty about item parameters: Expected response functions* (Research Report RR-94-28-ONR). Princeton, NJ: Educational Testing service.
- Morris, L.L. & Fitz Gibbon. C. T. (1978). *How to Measure Achievement*. Los Angeles, CA: Sage Publication.
- Santoso, Singgih, (2000), *Buku latihan SPSS statistik parametrik*, Jakarta: PT elex Media Komputindo Kelompok Gramedia
- Suherman, E. et al. (2003). *Strategi pembelajaran matematika kontemporer*. Bandung: FMIPA UPI
- Suryabrata, S. (1987). *Pengembangan tes hasil belajar*. Jakarta: CV Rajawali.
- Tim Sisjian. (1997). *Bank soal*. Jakarta: Pusat Penelitian dan Pengembangan Sistem Pengujian.

