



EFEK PEMUSATAN DATA TERHADAP PARAMETER ITEM BERBASIS *CLASICAL TEST THEORY (CTT)*

Nonoh Siti Aminah¹, Didik Eko Saputro², Dyah Fitriana Masithoh³

^{1, 2, 3} Program Studi Pendidikan Fisika Universitas Sebelas Maret, Surakarta, 57126

Email Korespondensi: nonoh_nst@yahoo.com

Abstrak

Penelitian yang dilakukan bertujuan untuk mencermati efek pemusatan data hasil analisis tes terhadap parameter item berbasis Clascal Test Theory (CTT). Metode penelitian yang digunakan metode eksplorasi, model penelitian yang digunakan penelitian psikometri. Data yang digunakan, pola respon siswa pada Ulangan Akhir Semester (UAS) Genap mata pelajaran Fisika kelas XI SMA di Kabupaten Boyolali Tahun Pelajaran 2012/2013. Banyak sampel 1036. Pengambilan data menggunakan teknik dokumentasi. Analisis kuantitatif menggunakan program olah data komputer dari ITEMAN versi 3.00. Hasil penelitian, efek pemusatan data terhadap parameter item berbasis CTT dari data 1 ke data 2 yaitu nilai reliabilitas dari 0,879 menjadi 0,678 (menurun walaupun masih pada kriteria reliabel), nilai mean p dari 0,526 menjadi 0,681 (nilai rerata taraf kesukaran meningkat), nilai mean biserial dari 0,542 menjadi 0,374 (nilai rerata daya beda menurun, walaupun masih dalam kriteria baik), nilai mean dari 21,035 menjadi 27,230 (meningkat), median dari 21 menjadi 27, modus dari 15 menjadi 28. varians dari 61, 310 menjadi 21,706, simpangan baku bergeser dari 7,830 menjadi 4,659 (menurun, menunjukkan nilai variansnya menurun, atau standardeviasinya menurun, atau errornya mengecil). Kesimpulan dari hasil penelitian, jika ingin meningkatkan parameter reliabilitas tes, maka banyak sampel harus ditambah (bukan menambah item tes). Parameter taraf kesukaran akan meningkat jika skor hasil tes lebih memusat artinya variansnya atau simpangan bakunya kecil, walaupun akan berpengaruh pada nilai daya beda yang mengecil.

Kata Kunci: Pemusatan, Clascal Test Theory (CTT), Parameter Item

Pendahuluan

Hasil pembelajaran di kelas berkaitan dengan pengetahuan, keterampilan dan sikap siswa. Pembelajaran yang berkualitas akan meningkatkan kualitas lulusan. Kualitas lulusan menjadi indikator dari kualitas pendidikan. Ditinjau dari cakupannya, penilaian terdiri dari penilaian yang bersifat makro dan penilaian yang bersifat mikro. Penilaian yang bersifat makro cenderung menggunakan sampel untuk menelaah suatu program dan dampaknya, program yang dimaksud yaitu program pendidikan. Program pendidikan adalah program yang direncanakan untuk memperbaiki kualitas dalam bidang pendidikan. Penilaian yang bersifat mikro digunakan di tingkat kelas, yang bertujuan untuk mengetahui capaian hasil belajar siswa. Capaian hasil belajar tidak hanya bersifat kognitif, melainkan juga mencakup semua potensi yang ada pada siswa. Sasarannya yaitu program

pembelajaran di kelas dengan penanggungjawab pengajar.

Penilaian pembelajaran diklasifikasikan menjadi dua, yaitu penilaian formatif dan penilaian sumatif. Penilaian formatif bertujuan untuk memperbaiki proses pembelajaran. Hasil penilaian formatif dianalisis untuk mengetahui konsep yang belum dipahami sebagian siswa, kemudian diikuti kegiatan remedial. Kegiatan remedial yaitu kegiatan pembelajaran untuk mengatasi kesulitan belajar siswa. Kesulitan belajar siswa diidentifikasi dari kegiatan penilaian formatif. Penilaian sumatif bertujuan menetapkan tingkat keberhasilan siswa. Teknik penilaian antara lain observasi, penugasan, inventori, jurnal, penilaian diri dan penilaian yang sesuai dengan karakteristik kompetensi dan perkembangan siswa. Teknik penilaian dapat dilakukan dengan menggunakan teknik tes dan non tes.

Tes merupakan alat pengumpul data untuk mengetahui kemampuan individu atau kelompok dalam menyelesaikan suatu

persoalan atau memperlihatkan ketrampilan tertentu yang menunjukkan hasil belajar, atau dalam menggunakan kemampuan psikologis untuk memecahkan suatu persoalan (Djemari Mardapi, 2004). Tes yang berkualitas berkaitan dengan soal yang berkualitas. Soal yang berkualitas memenuhi kriteria sebagai alat ukur yang baik. Salah satu cara untuk menentukan kriteria alat ukur yang baik dapat ditinjau dari teori tes klasik (Classical Test Theory: CTT). Parameter tes yang diukur pada CTT yaitu reliabilitas tes (reliability), indeks daya beda item (discrimination index), indeks kesukaran item (difficulty index), validitas isi (content validity), validitas konstruk (construct validity), dan validitas berdasar kriteria (criterion related validity).

CTT digunakan secara luas pada pengukuran pendidikan di Indonesia, walaupun memiliki keterbatasan. Salah satu keterbatasan dari CTT yaitu, ketika tes yang sama diberikan pada siswa yang berbeda, dan hasilnya dinyatakan dengan skor total. Tingkat kemampuan siswa tidak dapat dibandingkan berdasarkan skor total yang diperoleh siswa tersebut, karena skor total yang diperoleh siswa tidak menunjukkan tingkat kesulitan tes yang dikerjakan. Seorang siswa yang kemampuannya lebih rendah mungkin mendapatkan skor lebih tinggi pada suatu tes yang mudah dibandingkan dengan siswa lain yang lebih tinggi kemampuannya tetapi mendapatkan tes yang lebih sulit. Keadaan tersebut menggambarkan bahwa skor total tidak dapat dibandingkan, karena kedua siswa tersebut mengerjakan item yang mungkin memiliki tingkat kesulitan berbeda, walaupun skor total yang dicapainya sama. Pada kegiatan pembelajaran di kelas, guru menggunakan teori tes klasik untuk mengidentifikasi karakteristik soal yang digunakan, serta mengidentifikasi tingkat kemampuan siswa berdasarkan skor total yang diperoleh siswanya.

Analisis butir tes secara klasik adalah proses penelaahan butir tes melalui informasi dari jawaban siswa untuk mengetahui karakteristik tes yang digunakan. Aspek yang diperhatikan pada analisis butir tes secara klasik yaitu validitas, reliabilitas tes, tingkat kesukaran butir, daya beda butir dan

penyebaran pilihan jawaban (Depdiknas, 2008:8). Validitas tes pada teori tes klasik ditentukan oleh ketepatan dan kecermatan hasil pengukuran. Suatu alat ukur yang tinggi validitasnya akan memiliki eror pengukuran yang kecil, artinya skor setiap subjek yang diperoleh oleh alat ukur yang digunakan tidak jauh berbeda dari skor sesungguhnya. Dengan demikian secara keseluruhan alat tes yang digunakan akan menghasilkan varians error yang kecil, hal ini disebut dengan validitas intrinsik, yang dirumuskan sebagai akar kuadrat dari perbandingan antara varians skor murni dengan varians skor tampak, yaitu $r_{xy} = \sqrt{(s_t^2) / s_x^2}$ (Syaifuddin Azwar, 2006. p: 43).

Pengertian validitas berkaitan dengan hasil ukur bukan alat ukurnya. Validitas tes diartikan validitas hasil pengukuran yang diperoleh dari tes yang digunakan, hal ini ditekankan oleh Cronbach yang ditulis ulang oleh Syaifuddin Azwar, bahwa dalam proses validasi sebenarnya tidak bertujuan untuk melakukan validasi tes, akan tetapi melakukan validasi terhadap interpretasi data yang diperoleh menggunakan prosedur tertentu. Tipe validitas digolongkan dalam tiga katagori yaitu validitas isi (content validity), validitas konstruk (construct validity) dan validitas berdasar kriteria (criterion-related validity).

Peningkatan validitas yang agak berarti hanya terjadi jika tes yang semula pendek dengan koefisien validitas yang rendah. Oleh sebab itu untuk meningkatkan validitas tes, dalam prosedur penyusunan dan pengembangan tes, lebih baik meningkatkan kualitas item yang ada daripada sekedar menambah banyak item tanpa disertai usaha untuk memperbaiki item yang kualitasnya masih rendah (Syaifuddin Azwar 2006: 157)

Reliabilitas merupakan kriteria untuk menetapkan taraf ketelitian teknik atau alat penelitian, bila digunakan untuk mengukur hasil belajar seorang siswa. Suatu alat ukur dikatakan memiliki reliabilitas tinggi jika digunakan berulang-ulang untuk mengulang objek yang dilakukan oleh orang yang sama maupun berlainan maka akan menghasilkan hasil yang sama. Saifudin Azwar (2002:181) ada tiga cara untuk mengestimasi reliabilitas, yaitu melalui: pendekatan tes ulang (test-

retest), pendekatan tes sejajar (alternate-forms), dan pendekatan konsistensi internal (internal consistency).

Pendekatan tes ulang menunjukkan konsistensi pengukuran dari waktu ke waktu dan menghasilkan koefisien reliabilitas yang sering disebut sebagai koefisien stabilitas. Prosedur yang digunakan yaitu, memberikan tes yang sama dua kali untuk kelompok yang sama dengan interval waktu antara tes, dari beberapa menit sampai beberapa tahun. Pendekatan tes sejajar dapat dilakukan apabila tersedia dua bentuk instrumen pengukur yang dapat dianggap memenuhi asumsi sejajar. Estimasi reliabilitas dilakukan setelah kedua instrumen dikenakan berturut-turut pada kelompok subjek kemudian dikomputasi koefisien korelasi antara distribusi skor dari pengamatan kedua instrumen tersebut., Estimasi reliabilitas dengan pendekatan konsistensi internal di dasarkan pada data dari sekali penguasaan suatu bentuk alat ukur pada sekelompok subjek (single trial administration). Secara umum persamaan untuk mengestimasi reliabilitas dapat digunakan rumus Koefisien Alpha atau untuk data dikotomi dapat digunakan persamaan KR-20 (Badrun Kartowirangan, 2009:8).

Secara teoritik besar koefisien reliabilitas berkisar antar 0.0 sampai dengan 1.0, pada kenyataannya koefisien sebesar 1.0 dan 0.0, tidak pernah dijumpai. Koefisien reliabilitas dapat bertanda negatif (-), namun koefisien reliabilitas selalu mengacu pada angka positif (+). Hal ini disebabkan angka yang negatif tidak ada artinya bagi interpretasi reliabilitas hasil ukur. Koefisien reliabilitas sama dengan 1, berarti ada konsistensi yang sempurna pada hasil ukur yang bersangkutan. Pada pengukuran aspek-aspek psikologis dan sosial yang menggunakan manusia sebagai subyek hal ini tidak pernah terjadi. Manusia sebagai subyek memiliki berbagai sumber error dalam dirinya, dan dalam pelaksanaan pengukuran sumber error tersebut sangat mudah mempengaruhi kecermatan hasil pengukuran.

Karakteristik tes diketahui setelah analisis dilakukan pada tes yang digunakan. Salah satu program komputer yang dapat membantu kegiatan analisis tes yaitu MicroCat ITEMAN version 3.00. Cara

menggunakan MicroCat ITEMAN version 3.00 yaitu, menuliskan data input dalam bentuk file notepad kemudian menuliskan nama file input dan output untuk memperoleh hasil analisis. Uraian yang telah disebutkan memberikan gambaran betapa pentingnya usaha mengetahui karakteristik tes, sehingga informasi yang didapatkan dari alat ukur yang digunakan telah teruji. Pada penelitian yang dilakukan, dipusatkan pada karakteristik tes secara kuantitatif. Tes yang dianalisis yaitu tes Ulangan Akhir Semester (UAS) genap mata pelajaran Fisika kelas XI SMA di Kabupaten Boyolali Tahun Pelajaran 2012/2013 “.

Metode Penelitian

Metode Penelitian yang digunakan yaitu metode eksploratif, langkah-langkah yang dilakukan yaitu, (1) mengidentifikasi parameter tes Ulangan Akhir Semester (UAS) Genap menggunakan program ITEMAN, (2) mengidentifikasi parameter tes menggunakan tendency central. Hasil analisis menggambarkan karakteristik tes mata pelajaran Fisika pada Ulangan Akhir Semester (UAS) Genap Kelas XI SMA Tahun Pelajaran 2012/2013 di Boyolali.

Analisis data menggunakan analisis kuantitatif. Analisis kuantitatif dilakukan menggunakan teori tes klasik, dengan program MicroCat ITEMAN versi 3.00. ITEMAN merupakan program komputer yang digunakan untuk menganalisis butir soal secara klasik. Program ini termasuk satu paket program dalam MicroCat yang dikembangkan oleh Assessment Systems Corporation pada tahun 1982 dan mengalami revisi pada tahun 1984, 1986, 1988, dan 1993. (<http://www.scribd.com>)

Hasil Penelitian dan Pembahasan

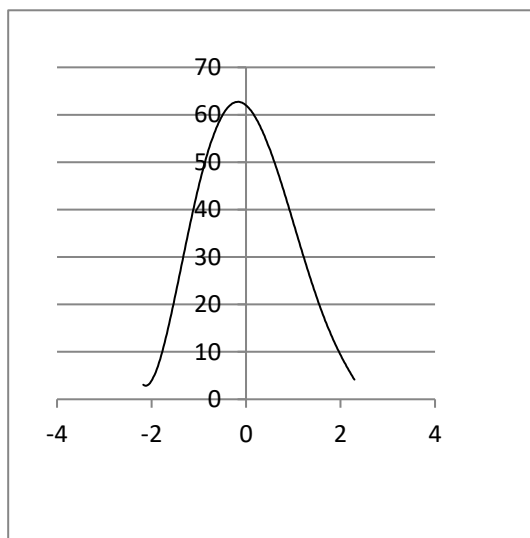
Hasil olah data menggunakan Program Microcat ITEMAN Versi 3.00, dituliskan pada Tabel 1.

Hasil olah data pada Tabel 1 dan Tabel 2, menunjukkan bahwa efek pemusatan berpengaruh pada parameter item. Data semakin halus atau varians semakin kecil, realibitas tes relatif tetap, daya beda item

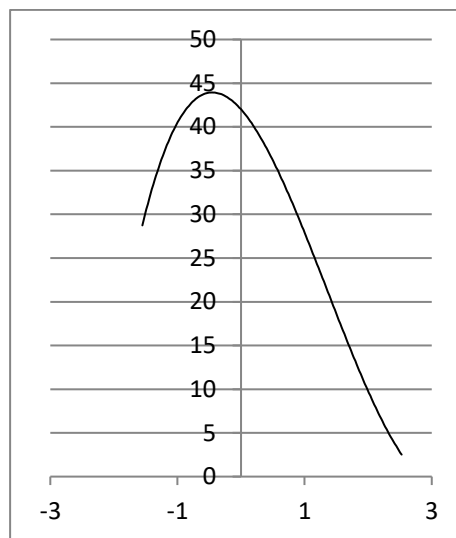
semakin besar, taraf kesukaran item semakin meningkat, dan kesalahan pengukuran semakin kecil.

Tabel 1. Hasil Olah Data Menggunakan Program *Microcat ITEMAN* Versi 3.00, Pada Data Awal dengan Data Setelah Pemusatan

Parameter	Data Awal	Data Setelah Pemusatan
1 N of Items	40	40
2 N of Examinees	1036	556
3 Mean	21.035	27.230
4 Variance	61.310	21.706
5 Std. Dev	7.830	4.659
6 Skew	0.152	0.350
7 Kurtosis	-0.948	-0.699
8 Minimum	4.000	20.000
9 Maximum	39.000	39.000
10 Median	21.000	27.000
11 Alpha	0.879	0.678
12 SEM	2.724	2.643
13 Mean P	0.526	0.681
14 Mean Item-Total	0.416	0.269
15 Mean Biserial	0.542	0.681



Grafik 1. Distribusi Hasil Olah Data Menggunakan Program *Microcat ITEMAN* Versi 3.00 pada Data Awal



Grafik 1. Distribusi Hasil Olah Data Menggunakan Program *Microcat ITEMAN* Versi 3.00 pada Data Awal

Simpulan, Saran, dan Rekomendasi

Hasil olah data pada Tabel 1 dan Tabel 2, menunjukkan bahwa efek pemusatan berpengaruh pada parameter item. Data semakin halus atau varians semakin kecil, realibitas tes relatif tetap, daya beda item semakin besar, taraf kesukaran item semakin meningkat, dan kesalahan pengukuran semakin kecil.

Hasil penelitian yang dilakukan, dapat dimanfaatkan untuk memilih tes yang baik menurut CTT. Tes yang baik tidak hanya dilihat dari koefisien reliabilitasnya tetapi juga dilihat dari standard error of measurement (SE), taraf kesukaran dan daya beda item. Semakin kecil nilai SE, menunjukkan tes tersebut layak untuk dipilih.

Daftar Pustaka

- Allen, M.J. & Yen, W.M. (1979). *Introduction to Measurement Theory*. Monterey, CA: Brooks/ Cole Publishing Company.
- Brennan, R.L. (2006). *Educational Measurement*. Iowa City: United State of America: American Council on Education and Praeger Publisher.
- Brennan, R.L. & Kolen, M.J. (2004). *Test Equating, Scaling, and Linking*. Iowa City: United State of America:

- American Council on Education and Springer Publisher.
- Crocker & Algina. (1986). *Introduction to Classical and Modern Test Theory*. New York: United State of America: CBS College Publishing.
- Djemari Mardapi. (2004). *Penyusunan Tes Hasil Belajar*. Yogyakarta: Program Pascasarjana Universitas Negeri Yogyakarta.
- Djemari Mardapi. (2008). *Teknik Penyusunan Instrumen Tes dan Non Tes*. Yogyakarta: Mitra Cendikia Press.
- Dikdasmen Dikbud (1999). *Pengelolaan Pengujian bagi Guru Mata Pelajaran*. Jakarta: Departemen Pendidikan dan Kebudayaan Direktorat Jenderal Pendidikan Dasar dan Menengah Direktorat Pendidikan Menengah Umum.
- Ebel R.L. & Friesbie. D.A (1986). *Essentials Of Educational Measurement*. New Jersey: Prentice – Hall. Inc.
- Linn, R.L. & Gronlund, N. E. (1976). *Measurement and Assesment In Teaching*. New Jersey: Prentice – Hall. Inc.
- Nitko, J. & Susan, B.M. (2007). *Educational Assesment Of Student*. Ohio: Pearson.

Pertanyaan:

1. Mengenai data sebelum pemusatan dan sesudah pemusatan, bagaimana mendeteksi?

Jawab:

1. Dengan mereduksi soal. Yang memiliki reliabilitas rendah, maka soal dibuang

