

PENDEKATAN STATISTIK DALAM PENILAIAN HASIL BELAJAR

Oleh
G. Suharto

Abstrak

Pengukuran dan penilaian merupakan dua kegiatan yang selalu ada dalam kegiatan pendidikan dan merupakan salah satu cara mengevaluasi efektivitas proses belajar mengajar. Berdasarkan kenyataan yang ada, setelah diamatidi sana sini ada beberapa hal yang perlu dipertanyakan. Permasalahan tersebut mencakup antara lain: distribusi sekor hasil belajar, validitas pengukuran, sekala penilaian, dan lain-lain faktor yang terkait dengan ketiga masalah tersebut.

Dalam artikel ini ketiga permasalahan tersebut dibahas dan dikaitkan dengan tujuan pendidikan nasional pada umumnya, teori dan prinsip dasar pengukuran, teori dan prinsip dasar penilaian hasil belajar. Sebagai jalan keluar terhadap pemecahannya, disarankan agar mengacu kepada konsep-konsep dasar statistika. Dalam hal ini ada lima saran yang diajukan yang meliputi: validitas pengukuran, cara pencatatan sekor hasil belajar, cara melaporkan sekor hasil belajar, cara membuat konversi nilai hasil belajar, dan cara menentukan efektivitas pendidikan berdasarkan distribusi sekor hasil belajar.

I. PENDAHULUAN

Dalam dunia pendidikan, kegiatan penilaian hasil belajar merupakan suatu bagian yang sangat penting dan selalu ada. Hal ini tidak hanya karena penilaian merupakan bagian integral dalam proses pendidikan, tetapi terlebih-lebih juga karena terkait dengan proses pengukuran. Pengukuran ini merupakan tahap akhir dalam proses pendidikan sekaligus merupakan usaha untuk mengetahui efektivitas pendidikan, dalam hal ini melalui hasil belajar peserta didik. Penilaian merupakan tindak lanjut dari pengukuran. Pengukuran adalah proses menginterpretasikan atau membandingkan hasil-hasil pengukuran. Kedua istilah ini begitu erat kaitannya sampai kadang-kadang dikacaukan pemakaiannya, atau dipakai untuk mengacu kepada hal yang sama. Apabila dikatakan bahwa A menjawab betul 16 dari 20 soal dalam suatu tes kemudian diartikan bahwa A menjawab 80 persen betul, angka ini belum dapat dikatakan sebagai nilai. Tetapi apabila angka 80 persen diartikan sebagai memuaskan, tinggi, ataupun rendah, ketiga rumusan arti ini merupakan nilai yang merupakan hasil dari penilaian.

Hal tersebut merupakan salah satu contoh penyalahgunaan pengukuran dan penilaian. Tetapi apabila kita kaji lebih lanjut, ada beberapa masalah yang terkait dengan kesalahan ukur (error), masalah yang terkait dengan penyalahgunaan (misuse), dan masalah yang terkait dengan interpretasi (misinterpretation) (Hopkin & Stanley, 1981: 16). Masalah-masalah tersebut akan menjadi semakin kompleks apabila kita kaitkan dengan tujuan pendidikan pada umumnya, pendekatan alat ukur, dan norma penilaian. Dalam artikel ini akan dibahas berbagai masalah tersebut berdasarkan kerangka acuan dan konsep-konsep berpikir statistik dan dicoba menemukan pemecahan yang paling mungkin dapat dilaksanakan oleh pendidik pada umumnya.

II. BEBERAPA MASALAH DALAM PENILAIAN HASIL BELAJAR

Secara garis besar dapat dikatakan bahwa masalah penilaian hasil belajar mencakup tiga hal utama, yaitu distribusi sekor hasil belajar, pendekatan penilaian hasil belajar, dan penilaian hasil belajar ditinjau dari tujuan pengukuran hasil belajar. Untuk jelasnya marilah kita ikuti satu demi satu.

1. Distribusi Hasil Pengukuran Hasil Belajar.

Dari pengalaman-pengalaman selama ini para ahli menemukan dan menyimpulkan bahwa sekor tes baku akan cenderung mempunyai distribusi kurva normal (Hopkin & Stanley, 1981: 47, 57). Dari penemuan ini kemudian banyak orang membuat norma penilaian berdasarkan konsep-konsep atau properti yang dimiliki oleh suatu distribusi kurva normal. sebagai contoh, sekor deviasi standar (selanjutnya disingkat DS), rangking persentil, sekor-z, sekor-t, dan yang paling sering dipakai adalah norma penilaian bersekala sembilan atau sekala sebelas. Disamping norma-norma ini, masih kita kenal adanya norma lain yang kurang jelas acuannya, misalnya sekala seratus satu (0 — 100), sekala huruf (ABCDE) dengan plus-minusnya; namun demikian dua sekala ini dibuat sebagai ekuivalensi dari sekala sembilan (1 — 9) atau sekala sebelas (0 — 10).

Apabila penilaian hasil belajar benar-benar menggunakan konsep-konsep di atas, ada satu masalah penting yang dapat kita renungkan, yaitu masalah jumlah kelulusan, karena dari berbagai macam norma di atas, jumlah kelulusannya paling banyak adalah 50 persen. Agar hal ini menjadi lebih jelas marilah kita ikuti contoh berikut.

Penilaian berdasarkan sekor deviasi standar yang banyak dipakai di tingkat SD, SMP, dan SMA. Dalam ketiga macam sekolah ini posisi dan kelulusan atau kenaikan seorang peserta didik ditentukan berdasarkan posisi hasil belajarnya terhadap nilai rerata kelasnya. Dalam hal ini seorang

peserta didik dianggap baik atau lulus apabila nilainya berada di atas nilai rerata kelas atau minimal sama. Apabila sekor hasil belajar benar-benar mempunyai distribusi kurva normal, ini berarti bahwa jumlah kelulusan atau siswa yang berhasil adalah 50 persen. Dalam kenyataannya, jumlah kelulusan selalu lebih banyak bahkan tidak jarang mencapai 90-an persen. Apabila demikian halnya, ini berarti bahwa dalam prakteknya sekor hasil belajar peserta didik tidaklah berbentuk kurva normal, tetapi cenderung mempunyai distribusi kurva juling negatif. Dengan kata lain dapat disimpulkan bahwa masalah pertama adalah adanya kesenjangan antara penemuan para ahli dengan kenyataan yang terjadi di sekitar kita.

Kenyataan tersebut di atas menimbulkan permasalahan kedua yang terkait dengan masalah pengukuran. Kurva juling negatif dapat terjadi apabila soal-soal dalam alat ukur sangat mudah sehingga sebagian besar peserta didik dapat mengerjakannya, atau apabila sebagian besar siswa pandai sehingga walaupun soal-soalnya sukar mereka tetap dapat mengerjakannya. Permasalahan ini akan dibahas dalam bagian berikut.

2. Masalah Pendekatan Dalam Pengukuran Hasil Belajar.

Tujuan pendidikan nasional pada umumnya dapat kita temukan dalam Pembukaan UUD 1945, yaitu mencerdaskan kehidupan bangsa. Hal ini dapat diartikan bahwa suatu tujuan pendidikan dapat dikatakan tercapai apabila semakin banyak jumlah peserta didik yang berhasil dalam belajarnya. Dengan bahasa statistika dapat dikatakan bahwa distribusi sekor hasil belajar mestinya berbentuk kurva juling negatif. Sebelum terjadinya proses pembelajaran distribusinya mungkin berbentuk kurva normal, tetapi setelah proses belajar-mengajar distribusinya berbentuk kurva juling negatif, karena jurang perbedaan antara peserta didik yang kurang pandai dengan yang pandai diperkecil. Apabila harapan ini kita kaitkan dengan dua kemungkinan terjadinya kurva juling negatif seperti di atas, ada dua pertanyaan mendasar yang harus terjawab, yaitu masalah norma penilaian dan masalah pengukuran hasil belajar.

Pertanyaan pertama terkait dengan norma penilaian yang dipergunakan dalam menentukan kelulusan. Seperti kita ketahui, ada dua macam pendekatan dalam penilaian, yaitu pendekatan acuan norma (selanjutnya disingkat PAN) dan pendekatan acuan patokan (selanjutnya disingkat PAP). Dalam PAN kelulusan seorang peserta didik ditentukan berdasarkan posisinya terhadap anggota kelompoknya; biasanya batas kelulusan minimalnya adalah nilai rerata dengan nilai DS sebagai lebar kelas intervalnya. Norma penilaian berdasarkan PAN ini akan menguntungkan peserta didik apabila nilai reratanya rendah dan DS-nya kecil, tetapi hal ini mengandung resiko karena kualitas lulusannya juga rendah. Hal ini mestinya tidak kita harapkan apabila kita kaitkan dengan tujuan pendidikan nasional pada

umumnya. Norma penilaian berdasarkan PAN akan sangat merugikan peserta didik apabila nilai reratanya tinggi, karena seorang yang sebetulnya telah mencapai kompetensi minimal terpaksa tidak dapat diluluskan karena banyak anggota kelompoknya yang mempunyai skor yang lebih tinggi. Akan tetapi apabila hal ini yang terjadi, kualitas lulusan akan lebih terjamin. Namun demikian kemungkinan terjadinya hal ini cenderung kecil, karena PAN cenderung berisi butir-butir soal yang sukar untuk memperoleh variabilitas yang tinggi. Dalam PAP, sebaliknya, kelulusan seseorang ditentukan berdasarkan posisinya terhadap kompetensi minimal yang telah ditentukan sebelumnya. Di sini batas kelulusan minimal mestinya merupakan kumpulan dari sejumlah tujuan instruksional khusus (TIK) untuk suatu satuan pengajaran. Dengan kata lain kelulusan seseorang ditentukan berdasarkan seberapa jauh dia mencapai atau melampaui batas kelulusan minimal yang telah ditentukan sebelumnya.

Apabila kedua pendekatan tersebut kita kaitkan dengan tujuan pendidikan nasional pada umumnya, kita boleh menaruh harapan pada norma penilaian PAP, karena PAP mempersyaratkan adanya butir-butir yang mudah, di samping butir-butir soal yang sukar. Hanya saja, dalam PAP yang berorientasi ke konsep belajar tuntas, batas kelulusan minimal biasanya dinyatakan dalam persen, yaitu antara 75 — 90 persen. Kriteria ini menimbulkan pertanyaan tentang jumlah butir soal, karena antara 75 — 90 persen tidak selalu sama dengan 85 — 90 persen yang lain. Dalam praktek kriteria tersebut sering dipergunakan secara a-priori. Dengan demikian dapat disimpulkan bahwa masalah yang terkait dengan pendekatan dalam penilaian hasil belajar adalah adanya kesenjangan antara tujuan pendidikan nasional dengan norma penilaian yang sering dipakai. Secara operasional masalah tersebut dapat dirumuskan menjadi seberapa jauh nilai yang diperoleh seorang peserta didik mencerminkan kualitas hasil belajarnya. Masalah ini menyangkut validitas alat ukur yang akan dibicarakan dalam bagian berikut.

3. Masalah Penilaian Ditinjau Dari Tujuan Pengukuran.

Dalam dunia pendidikan, tujuan pengukuran hasil belajar adalah mengetahui seberapa jauh efektivitas pendidikan. Dengan demikian setiap usaha pengukuran tentunya mengacu ke tujuan instruksional khusus dan validitas pengukurannya mestinya berorientasi kepada validitas isi atau tepatnya validitas kurikuler. Tetapi apabila kita amati dalam prakteknya banyak pengukuran yang berorientasi kepada PAN. Salah satu karakteristik PAN adalah bahwa butir-butir soalnya dibuat sedemikian rupa untuk mencapai variabilitas yang tinggi. Ini berarti bahwa apabila peserta didik dapat mengerjakan yang sukar, dengan sendirinya dapat juga mengerjakan yang mudah. Menurut hemat penulis, kelemahannya justru di sini karena

bersifat kurang *fair* dan baru sebatas asumsi. Sehingga ditinjau dari tujuan pengukuran masalah penilaian adalah adanya kesenjangan antara tujuan pengukuran dengan validitas alat ukurnya.

Dalam PAP sebaliknya, validitas alat ukur dapat dijamin justru karena PAP harus mengutamakan validitas kurikuler. Yang sering terjadi adalah bahwa soal-soalnya berorientasi kepada PAN, tetapi penilaiannya mempergunakan PAP. Peserta didik sering berkeluh-kesah karena dari sekian jumlah pengikut tes semuanya dapat nilai C, atau bahkan hanya beberapa orang saja yang lulus. Di sini peserta didik juga merasakan diperlakukan dengan tidak wajar sampai-sampai terjadi pemeo bahwa: nilai A hanya untuk Tuhan, nilai B untuk guru, nilai C untuk siswa. Kesimpulan-nya ialah bahwa baik dalam PAN maupun dalam PAP terdapat kesenjangan antara norma penilaian dengan pendekatan pengukurannya.

Disamping masalah-masalah yang telah dikemukakan di atas, ada satu masalah lagi yang bersifat lebih umum dan erat kaitannya dengan masalah validitas kurikuler, yaitu masalah konversi nilai. Dalam hal ini kita mengenal dua macam sekala, yaitu sekala sebelas dan sekala seratus satu dengan berbagai variasi plus-minusnya. Ada beberapa pertanyaan menggelitik apabila ketiga macam sekala penilaian ini kita kaitkan dengan konsep-konsep dasar pengukuran dan penilaian hasil belajar. Konsep apakah yang dipakai dalam pembuatan konversi nilai tersebut? Kurva normal? Rerata dan DS empirik? Ataukah rerata dan simpang baku atau DS teoretik? Kalau ketiga macam sekala tersebut kita cermati, kita tidak dapat menemukan siratan dari berbagai konsep tersebut di atas, baik dalam hal sekor kelulusan minimal maupun dalam hal lebar kelas intervalnya. Dalam kedua sekala tersebut, lebar kelas untuk nilai A (80—100) adalah 21; untuk nilai B (66—79) adalah 14; untuk nilai C (56—65) adalah 10; untuk nilai D (40—55) adalah 16; untuk nilai E (0—39) adalah 40. Di lain pihak ternyata bahwa pembobotannya mempunyai jarak yang sama: nilai A mempunyai bobot 4, nilai B mempunyai bobot 3, nilai C mempunyai bobot 2, nilai D dan E masing-masing 1 dan 0 (nol). Pembobotan seperti ini mencerminkan data interval, sementara isinya tidak mencerminkan bobot yang sama. Dari contoh kasus ini dapat disimpulkan bahwa dalam hal konversi nilai terdapat kesenjangan antara sistem konversi dengan kerangka acuan yang mendasarinya.

Konversi nilai berdasarkan angka ini juga mengandung kelemahan dalam artian bahwa sekor hasil belajar harus ditransfer terlebih dulu ke dalam sekala yang dipakai sebelum ditentukan nilainya. Hal seperti ini cenderung tidak efisien dan ada kecenderungan bahwa soal dalam tes hasil belajar jumlahnya disesuaikan dengan sekala yang akan dipakai dalam penilaian. Padahal akan lebih tepat apabila jumlah butir soal lebih didasarkan pada validitas isi atau lebih tepatnya TIK-TIK dalam kurikulum. Dengan demikian masalah penilaian adalah adanya kesenjangan antara sistem konversi

nilai dengan pelaksanaan penyiapan alat ukur dan penyekoran hasil belajar.

Berangkat dari berbagai permasalahan di atas, dalam bagian berikut akan ditawarkan suatu pemecahan yang mencakup prinsip dasar pengukuran dan penilaian hasil belajar dengan pendekatan statistik.

III. STATISTIKA SEBAGAI ALAT KOMUNIKASI INFORMASI

Dalam bagian ini akan dibahas prinsip-prinsip pengukuran dan penilaian hasil belajar yang mencakup bagaimana mendeskripsikan skor hasil belajar, bagaimana membuat skala atau norma penilaian, dan bagaimana mencatat dan melaporkan hasil penilaian hasil belajar.

1. Statistika Deskriptif Sekor Hasil Belajar.

Informasi apakah yang dapat diperoleh dari statistika? Untuk menjawab hal ini perlu terlebih dulu kita bicarakan serba sedikit tentang apa dan bagaimana statistika. Statistika sebagai ilmu dapat dibedakan menjadi dua, yaitu statistika deskriptif dan statistika inferensial. Berdasarkan fungsi praktisnya, statistika deskriptif merupakan alat bantu komunikasi dan interpretasi, sementara statistika inferensial merupakan alat bantu prediksi. Yang pertama bekerja berdasarkan data yang ada dalam suatu kelompok distribusi atau sampel; statistika inferensial membuat prediksi atau prakiraan terhadap populasi atau kesemestaan berdasarkan informasi yang ada dalam suatu sampel. Keduanya merupakan alat bantu generalisasi atau pembuatan teori yang mantap. Namun demikian, dalam hal ini pembahasan akan dipusatkan pada statistika deskriptif saja.

Ada dua masalah penting yang menjadi pusat perhatian statistika deskriptif, yaitu tendensi sentral dan variabilitas. Tendensi sentral atau kecenderungan memusatnya suatu sekor dalam suatu kelompok dapat diartikan sebagai suatu sekor atau angka yang menjadi titik anjak sebaran menaik ataupun menurun dengan satuan jarak yang sama. Satuan jarak yang sama inilah yang menjadi ukuran variabilitas atau kecenderungan menebar yang diartikan sebagai besar-kecilnya angka atau sekor dalam suatu distribusi sekor yang merupakan tolok ukur pembeda antar sub-kelompok dalam distribusi tersebut. Dalam konsep kurva normal, titik anjak sebaran ini diwujudkan dengan simbol angka nol dan angka +1 untuk setiap anjakan naik, sedangkan angka -1 untuk setiap anjakan turun.

Titik anjak sebaran ini dapat berupa: mode (disingkat M_o), median (disingkat M_d), dan mean (disingkat M_n). Mode adalah suatu sekor dalam distribusi yang mempunyai pemunculan atau frekuensi terbanyak. Median

atau titik tengah adalah suatu skor atau angka dalam distribusi yang membagi anggota distribusi dua, masing-masing dengan jumlah anggota yang sama. Mean atau rerata adalah suatu skor atau angka dalam distribusi yang merupakan hasil bagi dari jumlah seluruh skor dengan jumlah anggota dalam distribusi tersebut. Ketiga macam tendensi sentral ini dapat dipakai sebagai alat komunikasi tentang tinggi-rendahnya pencapaian rata-rata dari kelompok tersebut, juling-tidaknya suatu distribusi. Apabila ketiganya mempunyai nilai atau berupa angka yang sama, distribusi tersebut berbentuk kurva normal dan atau simetrik. Apabila nilai M_o lebih besar dari nilai M_d atau M_n , distribusi tersebut berbentuk kurva juling negatif; sebaliknya, apabila nilai M_o lebih kecil dari M_d atau M_n , distribusi tersebut berbentuk kurva juling positif. Kurva juling negatif dapat diartikan bahwa sebagian besar anggota kelompok atau distribusi tersebut mempunyai skor atau pencapaian yang tinggi atau di atas rata-rata. Kurva juling positif dapat diartikan bahwa sebagian besar anggota kelompok atau distribusi mempunyai skor atau pencapaian yang rendah atau di bawah nilai rerata. Sebagai alat pengukur tendensi sentral M_n merupakan alat ukur yang paling dapat diandalkan. Informasi tentang kecenderungan memusat ini jelas sangat membantu guru atau pendidik dalam proses penilaian atau pengambilan keputusan secara cepat dan tepat tentang sedikit-banyaknya jumlah kelulusan.

Variabilitas dapat berupa rentangan atau cakupan (range, disingkat R) dan deviasi standar atau simpang baku (disingkat DS). Rentangan adalah selisih antara skor terbesar dengan skor terkecil dalam suatu distribusi skor. Dengan membagi R dengan angka tertentu (biasanya 5 atau 6 - sesuai asumsi kurva normal) dapat diperoleh satuan jarak sebagai dasar pembagian suatu distribusi menjadi beberapa kategori atau kelas interval. DS adalah suatu angka atau besaran yang merupakan akar dari varians atau moment sentral kedua. Momen sentral kedua ini adalah suatu besaran yang merupakan hasil bagi dari jumlah kuadrat simpang rerata dengan jumlah seluruh anggota dalam distribusi. DS dianggap sebagai alat ukur variabilitas yang paling stabil sehingga seperti halnya R sering dipakai sebagai satuan jarak dalam membagi suatu distribusi menjadi beberapa kelas interval. Sebagai satu jarak sebaran, DS bahkan mampu menjamin signifikansi perbedaan antar kelas interval. Besar-kecilnya besaran alat ukur variabilitas memberikan informasi tentang homogenitas/heterogenitas suatu kelompok atau distribusi. Semakin kecil besarnya, semakin homogen; semakin besar, semakin heterogen. Apabila digabung dengan alat ukur tendensi sentral, informasinya menjadi heterogen tinggi atautkah heterogen rendah, homogen tinggi atautkah homogen rendah. Dengan kata lain, kurva juling negatif cenderung homogen rendah apabila DS-nya kecil, kurva juling positif cenderung homogen rendah apabila DS-nya kecil.

2. Pendekatan Statistik Dalam Penilaian Hasil Belajar.

Pendekatan statistik dalam sistem penilaian hasil belajar dimaksudkan bahwa sistem konversi nilai ditentukan berdasarkan nilai rerata sebagai nilai minimal kelulusan, dan nilai DS sebagai lebar kelasnya. Dalam PAN nilai rerata ini masih dipertanyakan, yaitu seberapa jauh nilai rerata yang dicapai dapat dipertanggungjawabkan; atau seberapa jauh hal tersebut dapat mencerminkan kualitas hasil belajar peserta didik. Kualitas hasil belajar dapat dikatakan rendah apabila nilai rerata rendah, kualitas hasil belajar tinggi apabila nilai rerata tinggi. Begitu pula halnya dengan DS sebagai satuan pembeda akan menjamin kebermaknaan perbedaan antara sub-kelompok yang ada, apabila lebar kelasnya berdasarkan satu DS. Yang sering terjadi ialah bahwa lebar kelas tidak menjamin adanya perbedaan yang bermakna, justru karena kurang dari satu DS, sementara bobot dalam skala nilai (biasanya skala ABCDE) jelas menunjukkan perbedaan yang signifikan. Hal ini dapat menjadi jelas apabila kita telusuri, misalnya, skala sembilan ataupun skala sebelas.

Dari konsep tendensi sentral dan variabilitas ini muncul konsep rerata teoretik dan DS teoretik atau sering pula disebut rerata ideal dan DS ideal. Dua konsep ini merupakan konsep dasar yang penting apabila kita kaitkan dengan masalah pengukuran dan penilaian hasil belajar. Rerata teoretik ini dapat diperoleh dengan cara: jumlah skor maksimal dengan skor kebetulan dibagi dengan bilangan dua. Skor kebetulan ini diperoleh dengan membagi skor maksimal dengan jumlah jawaban alternatif pada masing-masing butir soal. Apabila suatu tes objektif dengan jumlah soal 100, dengan jawaban alternatif 5 misalnya, maka skor kebetulan adalah: $100 : 5 = 20$. Dengan demikian untuk tes ini rerata teoretiknya adalah: $(100 + 20) : 2 = 60$. Dengan asumsi kurva normal, dapat dicari nilai DS teoretiknya, yaitu: jumlah skor maksimal dikurangi dengan nilai rerata teoretik, yang selanjutnya dibagi 3. Dengan demikian DS teoretiknya adalah: $(100 - 60) : 3 = 13,3$. Jadi nilai rerata teoretik dan DS teoretik diperoleh melalui tiga acuan, yaitu jumlah skor maksimal, jumlah jawaban alternatif, dan konsep DS dalam kurva normal. Apabila soal berbentuk isian, atau lain-lain yang sejenis, nilai rerata teoretik adalah setengah dari jumlah skor maksimal. Konsep ini menurut hemat penulis dapat diterapkan pada soal-soal bentuk uraian, dan sesuai pula dengan prinsip-prinsip pengukuran dan penilaian hasil belajar.

Dari berbagai pengalaman dan pengamatan penulis selama ini terhadap skor hasil belajar mahasiswa, rerata teoretik ini jarang tercapai dalam suatu kelompok peserta didik. Namun demikian beberapa kasus menunjukkan bahwa apabila nilai rerata teoretik ini tercapai, ternyata kelompok tersebut benar-benar kelompok yang homogen pandai berdasarkan pengamatan kesehariannya. Dengan demikian dapat disimpulkan bahwa kedua

konsep ini nampaknya dapat memecahkan masalah validitas sekor hasil belajar dan norma penilaiannya. Atau dengan singkat dapat dikatakan bahwa penilaian hasil belajar berdasarkan nilai rerata teoretik dan DS teoretik dapat menjamin kualitas kelulusan peserta didik.

3. Menentukan Sekala Penilaian Sesuai Tujuan Pengukuran.

Seperti diuraikan di bagian depan, tujuan pengukuran hasil belajar antara lain adalah mengukur perolehan belajar yang selanjutnya dipakai untuk mengevaluasi tercapai tidaknya tujuan pendidikan nasional pada umumnya. Tujuan ini dapat dinyatakan tercapai dengan memuaskan apabila hasil belajar menunjukkan kualitas yang tinggi atau setidaknya memenuhi kriteria kelulusan minimal. Dalam PAN kriteria ini kurang meyakinkan karena bersifat relatif, dalam arti sangat tergantung pada nilai rerata. Dalam PAP kriteria ini juga menimbulkan masalah karena dinyatakan dalam bentuk persen yang bersifat relatif. Untuk memecahkan dua masalah ini, kita dapat kembali kedua konsep dasar tentang pengukuran dan penilaian.

Pengukuran hasil belajar hendaknya mengutamakan validitas isi atau tepatnya validitas kurikuler dalam hal menentukan butir-butir soalnya. Dalam hal ini acuannya adalah TIK-TIK yang ada dalam satuan program pengajaran (semester, caturwulan, dan sebagainya). Kadang-kadang terjadi bahwa jumlah butir soal akan menjadi sangat besar apabila setiap TIK harus diwakili secara pantas. Dengan menggunakan pendekatan statistik jumlah butir soal besarpun tidak menimbulkan masalah. Dengan cara demikian, validitas isi dari suatu sekor hasil belajar dapat dicapai dengan cara mengakumulasi seluruh sekor hasil pengukuran dalam suatu program atau periode pengajaran. Jumlah kumulatif inilah yang selanjutnya dipergunakan sebagai dasar penilaian akhir.

Apabila pengukuran hasil belajar dikerjakan seperti tersebut di atas, para pendidik dapat menggunakan konsep rerata teoretik dan DS teoretik. Dua konsep ini rasanya cukup dapat diandalkan, baik ditinjau dari berbagai masalah yang telah dipecahkan di bagian depan maupun dari kepraktisan teknisnya. Rerata teoretik kumulatif dapat diperoleh dengan menjumlahkan seluruh rerata teoretik dari setiap sub-tes yang ada (baik yang objektif maupun yang berbentuk uraian). Apabila jumlah ini untuk mengurangi jumlah kumulatif sekor maksimal akan diperoleh juga nilai DS teoretiknya, setelah selisih tersebut dibagi tiga. Dengan memakai rerata teoretik kumulatif dan DS teoretik kumulatif ini pendidik dapat dengan mudah membuat konversi nilainya. Apabila rerata kelompok berada jauh di bawah rerata teoretik, pendidikan dapat mengubah konversi nilai dengan batas toleransi antara 0,5—1,0 DS teoretiknya dihitung ke bawah dari nilai rerata

teoretik. Batas toleransi seperti ini rasanya masih dapat dipertanggungjawabkan dalam hal kualitas kelulusannya.

Secara teknis, prosedur di atas pun sangat sederhana. Setelah sekor akhir dari masing-masing peserta didik diperoleh, pendidik cukup mencari nilai M_o dan M_n -nya saja untuk dapat dipakai sebagai langkah pertama dalam mengevaluasi efektivitas pengajaran atau tercapai-tidaknya tujuan pendidikan pada umumnya. Dengan membandingkan besaran nilai M_o dari M_n seorang pendidik dapat dengan segera mengetahui dan menentukan bentuk kurvanya. Prosedur ini juga akan memudahkan seorang pendidik dalam menentukan indeks prestasi akhir seorang peserta didik apabila setiap hasil belajar sudah dinilai berdasarkan prosedur ini. Dengan melihat nilai akhir dengan bobotnya masing-masing, seorang pendidik akan dapat dengan mudah menghitung IP terakhir seorang peserta didik.

IV. KESIMPULAN

Dari uraian di atas dapat ditarik beberapa kesimpulan yang dapat dijadikan prinsip-prinsip atau prosedur dalam pengukuran dan penilaian hasil belajar pada umumnya.

Pertama, bahwa pengukuran hasil belajar hendaknya mengutamakan validitas kurikuler dengan TIK sebagai acuan utamanya. Dalam hal ini jumlah butir soal menjadi tanggung jawab pendidik. Besar kecilnya jumlah soal lebih ditentukan oleh seberapa jauh jumlah tersebut mencerminkan validitas isi.

Kedua, bahwa pencatatan sekor hasil belajar hendaknya bersifat kumulatif. Jumlah ini mencakup segala macam usaha pendidik dalam memonitor perolehan belajar peserta didik. Jumlah sekor inilah yang dipakai dalam menentukan nilai akhir.

Ketiga, bahwa dalam setiap pencatatan sekor hasil belajar hendaknya dicantumkan pula sekor maksimal, nilai rerata teoretik, nilai DS , dan jumlah peserta didik untuk masing-masing daerah DS -nya (distribusi frekuensi). Data semacam ini merupakan data otentik yang sangat berguna untuk kegiatan penelitian dan evaluasi pendidikan, pelayanan alumni, pendokumentasian, dan lain-lain.

Keempat, bahwa penilaian hasil belajar hendaknya menggunakan rerata teoretik sebagai batas kelulusan minimal, dan DS teoretik sebagai lebar kelasnya. Dengan cara ini kualitas kelulusan dapat terjamin dan perbedaan nilai antar kategori dan bobot nilainya dapat dipertanggungjawabkan secara ilmiah.

Kelima, bahwa suatu pendidikan dapat disebut berhasil secara efektif dan efisien apabila setelah dikenakan prosedur penilaian seperti di atas ternyata distribusi kelulusannya mempunyai kurva juling negatif. Semakin juling negatif, semakin berhasil.

V. DAFTAR PUSTAKA

- Departemen Pendidikan dan Kebudayaan, Republik Indonesia, *Bahan Penataran dan Bahan Referensi Penataran P-4*, Jakarta, 1984.
- Departemen Pendidikan dan Kebudayaan Republik Indonesia, *Diagnostik Kesulitan Belajar dan Pengajaran Remedial*, Jakarta, 1985.
- Fernandes, H.J.X., *Comparison of CRM and NRM Tests*, Jakarta, 1983.
- Hopkin, Kenneth D. and Julian C. Stanley, *Educational and Psychological Measurement and Evaluation*, London: Prentice Hall Inc., 1981.
- Institut Keguruan dan Ilmu Pendidikan Yogyakarta, *Peraturan Akademik* 1988, Yogyakarta, 1988.
- Sutrisno Hadi, *Kuliah Statistik Program S2 IKIP Jakarta*, Yogyakarta, 1982.