

BEBERAPA SUMBER KESALAHAN PENGUKURAN DALAM UJIAN BENTUK OBJEKTIF DAN UJIAN BENTUK SUBJEKTIF

Oleh: Anas Sudijono

Masalah pengukuran dan penilaian dalam pendidikan adalah masalah yang selalu terkandung dalam pekerjaan dan pendidikan keguruan, sehingga karenanya sudah seharusnya menjadi salah satu bagian yang penting dalam kelengkapan keahlian seorang guru. Bahkan tidak hanya sekedar menjadi salah satu bagian saja, akan tetapi sebaliknya ia merupakan bagian integral, yang tak terpisahkan dari proses mengajar dan belajar. Tanpa titik tolak dasar pikiran yang serupa ini maka pengukuran dan penilaian pendidikan tidak akan dapat menunaikan fungsinya sebagaimana mestinya.

Berbicara tentang pengukuran dan penilaian dalam dunia pendidikan, maka istilah "pengukuran" (*measurement*) sering kali diberi arti yang sama dengan "penilaian" (*evaluation*). Dalam hal ini acapkali orang beranggapan bahwa apabila ia mengukur sesuatu hal -- misalnya "mengukur" kecerdasan seseorang atau sekelompok orang dengan menggunakan alat, misalnya menggunakan tes intelligensi --, maka ia merasa telah melakukan "penilaian"¹⁾. Memang, sekalipun dua pengertian itu mempunyai pertalian yang sangat erat, namun sebenarnya kedua istilah tersebut mempunyai arti yang berbeda.

"Penilaian" -- setidaknya-tidaknya dalam bidang psikologi dan pendidikan -- mempunyai arti yang lebih luas ketimbang "pengukuran". "Pengukuran hanyalah merupakan suatu langkah atau tindakan yang "kiranya perlu diambil" dalam rangka "penilaian". Dikatakan "kiranya perlu diambil", sebab tidak semua perbuatan "penilaian" harus didahului oleh tindakan "pengukuran" secara nyata. "Pengukuran" pada dasarnya bersifat deskriptif-kuantitatif; kegiatan "pengukuran" akan menghasilkan data dan informasi yang pada umumnya berbentuk angka-angka.

¹⁾Lihat : Prof. Dr. Masrun, M.A., "Pengukuran dalam Pendidikan", dalam: *Pengukuran Pendidikan*, Yogyakarta: Universitas Gadjah Mada, 1989, hal.1.

Sedangkan "penilaian" pada dasarnya merupakan penafsiran atau interpretasi yang "sering" bersumber pada data yang bersifat kuantitatif. Dikatakan "sering", sebab tidak semua penafsiran itu bersumber pada data kuantitatif. Sekalipun demikian perlu ditegaskan, bahwa "penilaian" atau evaluasi dalam bidang pendidikan -- khususnya penilaian terhadap prestasi belajar -- sebagian besar bersumber pada data yang bersifat kuantitatif yang asalnya dari hasil pengukuran. Jadi proses "penilaian" dalam bidang pendidikan pada umumnya didahului oleh proses "pengukuran"²⁾.

"Pengukuran" adalah suatu tindakan atau proses untuk menentukan luas atau kuantitas dari sesuatu; "pengukuran" akan memberikan jawab atas pertanyaan: "how much?", sedangkan "penilaian" akan memberikan jawaban terhadap pertanyaan: "what value?"³⁾.

Para ahli di bidang evaluasi pendidikan pada umumnya berpendapat bahwa penilaian mengenai prestasi belajar tidak akan dapat dilaksanakan secara baik apabila penilaian itu tidak didasarkan atas data kuantitatif hasil pengukuran. Ini berarti bahwa penilaian harus didasari oleh hasil-hasil pengukuran yang cermat⁴⁾. Berdasar atas hasil pengukuran, maka penilaian dapat dilakukan dengan lebih mudah, sebab dengan melalui pengukuran, maka pelukisan data akan menjadi lebih *accurate, objective* dan *communicable*⁵⁾.

Itulah sebabnya mengapa di dalam praktek, masalah pengukuran mempunyai kedudukan yang sangat penting dalam proses penilaian. Baik-buruknya hasil penilaian akan banyak bergantung pada hasil pengukuran; hasil pengukuran yang kurang cermat akan memberikan hasil penilaian yang kurang cermat pula. Sebaliknya, dengan mempergunakan teknik pengukuran yang tepat, diharapkan akan dapat menjadi landasan yang kuat untuk melakukan penilaian yang tepat. Kenyataan ini pulalah yang mendorong para pengajar untuk berusaha mendapatkan alat pengukur prestasi belajar yang tepat, atau setidaknya-tidaknya mendekati ketepatan.

Usaha-usaha perbaikan atau penyempurnaan alat pengukuran dalam bidang pendidikan (terutama dalam menilai hasil belajar), terutama ditujukan untuk mendapatkan alat pengukur yang dapat memenuhi

²⁾ Lihat: *ibid.*, hal. 1-2.

³⁾ Drs. Wayan Nurkencana dan Drs. PPN Sumartana, *Evaluasi Pendidikan*, Surabaya: Usaha Nasional, 1981, hal. 2.

⁴⁾ Lihat: Prof. Dr. Masrun, M.A., *op.cit.*, hal. 2.

⁵⁾ Lihat: Prof. Dr. T. Raka Joni, M.Sc., *Pengukuran dan Penilaian Pendidikan*, Malang: YP2LPM, 1984, hal. 8.

prinsip-prinsip dasar dari pengukuran pada umumnya. Dengan memenuhi prinsip-prinsip dasar pengukuran ini, diharapkan akan mengurangi kesalahan-kesalahan pengukuran. Bila kesalahan itu ditekan sampai sekecil mungkin, maka diharapkan skor atau nilai yang berasal dari hasil pengukuran akan sedikit sekali bedanya dengan skor atau nilai yang sebenarnya, yaitu skor atau nilai yang benar-benar mencerminkan prestasi belajar anak didik yang kita ukur.

Dalam pengukuran di bidang pendidikan, biasanya kita mengenal tiga macam prinsip dasar pengukuran, yaitu: objektivitas, reliabilitas dan validitas⁶⁾.

Prinsip objektivitas menghendaki adanya hasil yang objektif dari setiap tindakan pengukuran. Menurut Guilford⁷⁾, istilah objektivitas diberi arti sebagai *inter personal agreement*. Bila kebanyakan individu menunjukkan kesamaan pendapat mengenai observasi terhadap sesuatu objek atau kesamaan kongklusi mengenai sesuatu masalah, maka dapat dikatakan bahwa pendapat-pendapat tersebut mendekati kebenaran dan lebih bebas dari penyimpangan-penyimpangan individual. Prinsip objektivitas dalam pengukuran menghendaki adanya personal agreement ini. Hal tersebut berarti bahwa apabila dua orang atau lebih memberi skor terhadap hasil suatu pengukuran, maka ada "kecocokan" skor yang mereka berikan. Bila skor itu diberikan dalam bentuk angka-angka, maka angka-angka yang mereka berikan harus sama.

Prinsip objektivitas ini mempunyai arti yang penting bagi perkembangan ilmu pengukuran dan penilaian, sebab dengan adanya hasil-hasil pengukuran yang objektif, dapatlah diadakan analisis secara objektif pula. Dengan menggunakan teknik-teknik statistik atau matematika, hasil-hasil pengukuran akan memberikan pula informasi-informasi yang lebih banyak dan lebih lengkap. Selain dari itu, prinsip objektivitas ini selaras dengan perkembangan ilmu pengetahuan, dimana setiap ilmu menghendaki adanya sifat yang objektif.

Prinsip kedua yang kiranya perlu diperhatikan dalam setiap usaha pengukuran ialah: *reliabilitas* dari pengukuran itu. Prinsip ini menyangkut pertanyaan: sampai seberapa jauhkah pengukuran yang dilakukan berulang kali terhadap subjek atau sekelompok subjek yang sama, dapat memberikan hasil-hasil yang relatif tidak mengalami

⁶⁾ Prof. Dr. Masrun, M.A., *op.cit.*, hal. 4; Bandingkan dengan: Dr. Ny. Suharsimi Arikunto, *op.cit.*, hal.46; Bandingkan pula dengan: Drs.Wayan Nurkencana dan Drs. PPN Sumartana, *op.cit.*, hal. 122.

⁷⁾ Prof. Dr. Masrun, M.A., *Ibid.*, hal. 4, Lihat juga: J.P. Guilford, *Psychometric Methods*, New York: McGraw Hill Book Company Inc., 1954, p. 1-5.

perubahan. Bila hasil yang kita peroleh selalu sama, maka dapat kita katakan bahwa alat pengukur yang kita gunakan itu mempunyai reliabilitas yang tinggi.⁸⁾

Dengan demikian, prinsip reliabilitas menghendaki adanya keajegan (*consistency*) dari hasil pengukuran yang berulang-ulang terhadap seorang subjek maupun sekelompok subjek yang sama, dengan catatan bahwa subjek-subjek yang diukur itu tidak mengalami perubahan-perubahan.

Dalam praktek, kita tidak dapat memperoleh reliabilitas yang sempurna di dalam pengukuran hasil belajar. Hal ini disebabkan karena adanya penyimpangan-penyimpangan yang sumbernya tidak dapat kita kuasai seluruhnya. Penyimpangan-penyimpangan ini terutama berasal dari kondisi subjek yang kita ukur. Karena yang kita ukur adalah makhluk hidup, maka sesuai dengan sifatnya yang hidup, mereka selalu mengalami perubahan-perubahan. Adanya perubahan ini akan mengurangi "keajegan" dari hasil pengukuran. Namun demikian, apabila prinsip objektivitas telah dapat kita capai, maka dengan menggunakan perhitungan-perhitungan statistik, kita dapat mengestimasi nilai reliabilitas dari suatu pengukuran.⁹⁾

Objektivitas dan reliabilitas adalah merupakan syarat yang harus dipenuhi untuk pengukuran yang baik; akan tetapi dengan hanya memenuhi kedua prinsip itu saja belumlah cukup. Suatu prinsip yang lebih fundamental sifatnya daripada kedua prinsip tadi ialah *prinsip validitas*. Prinsip ini menyangkut pertanyaan: Seberapa jauh pengukuran-pengukuran yang kita lakukan itu benar-benar dapat mengukur secara tepat apa yang ingin kita ukur. Bila prinsip objektivitas menghendaki "kecocokan" hasil penilaian (score) dari para penilai (scorers), dan prinsip reliabilitas menuntut adanya "kecocokan" antara hasil pengukuran yang berulang-ulang terhadap kelompok yang sama, maka prinsip validitas menuntut adanya "kecocokan" antara alat pengukur dengan apa yang ingin diukur.¹⁰⁾

Memang harus diakui, bahwa tidak ada satu teknik penilaian pun yang akan pernah mencapai sifat yang 100% objektif dan eksak. Akan tetapi bahwa keobjektifan, ketetapan dan ketepatan yang setinggi itu, yang tidak mungkin dapat dicapai, tidaklah boleh diartikan bahwa kita harus menghentikan usaha kita untuk mencapai efisiensi yang tinggi.

⁸⁾ Lihat: Prof. Dr. Masrun, M.A., *ibid.*, hal. 5-6.

⁹⁾ *Ibid.*

¹⁰⁾ *Ibid.*, hal. 7.

UJIAN SEBAGAI ALAT PENGUKUR HASIL BELAJAR

Sebagaimana telah kita maklumi, fungsi utama dari ujian ialah untuk mengetahui seberapa jauh peserta didik telah dapat memahami tujuan dan isi pelajaran yang telah diberikan kepada mereka. Pengukuran hasil belajar disini dilakukan dengan menggunakan butir-butir soal (*items*) dalam jumlah tertentu, sesuai dengan kebutuhan. Sudah barang tentu agar ujian itu dapat memenuhi fungsinya dengan baik, maka ujian tersebut harus direncanakan dengan baik pula. Ujian yang disusun tergesa-gesa dan tanpa perencanaan yang cermat, kecil kemungkinannya untuk dapat memenuhi fungsinya secara efektif; bahkan ada kemungkinan ujian yang demikian itu "akan membawa peserta didik untuk memiliki *habit* belajar yang tidak kita inginkan".¹¹⁾

Ditilik dari segi itemsnya, ujian dapat dibagi menjadi dua golongan besar, yaitu ujian bentuk uraian (*essay test*) dan ujian bentuk objektif (*objective test*). Ujian Essay juga sering disebut Ujian Bentuk Subjektif, yaitu ujian yang berwujud pertanyaan-pertanyaan, dimana pihak yang diuji (*testee*) bebas menjawab dengan kalimat serta bahasanya sendiri.¹²⁾ Ujian Bentuk Subjektif adalah ujian yang memungkinkan testee untuk menjawab items secara bebas, jauh lebih bebas daripada jawaban yang dituntut oleh items pada Ujian Bentuk Objektif. Dalam menjawab butir-butir soal Ujian Bentuk Subjektif, testee harus memilih dari khazanah pengetahuan dan pengertian materi yang menurut pendapatnya relevan dengan soal, dan selanjutnya dengan caranya sendiri menyusun jawabannya.¹³⁾ Sedangkan Ujian Bentuk Objektif berwujud pertanyaan-pertanyaan yang jawabannya telah pasti dan pihak yang diuji (*testee*) tidak perlu menulis jawaban dalam bentuk kalimat-kalimat atau uraian-uraian yang panjang.¹⁴⁾

Ditilik dari segi cara testee mengerjakan soal-soalnya, Ujian Bentuk Objektif dapat dibedakan menjadi dua golongan, yaitu: (1) Variasi di mana testee harus mensuply jawabannya, sehingga dalam hal ini tidak jauh berbeda dengan Ujian Bentuk Subjektif. Inilah yang dikenal dengan istilah *Completion Item*, yaitu apabila testee tinggal meneruskan kalimat yang "belum jadi" atau mengisi bagian-bagian kalimat yang dihilangkan; (2) Variasi kedua, menuntut testee untuk memilih alternatif-alternatif jawaban yang telah disediakan. Tiga di antara yang termasuk dalam

¹¹⁾ Lihat: *ibid.*, hal. 10.

¹²⁾ Lihat: *ibid.*, hal. 12.

¹³⁾ Lihat: Prof. Dr. T. Raka Joni M.Sc., *op.cit.*, hal. 77.

¹⁴⁾ *Ibid.*

variasi kedua adalah soal-soal di mana testee harus menyatakan, apakah suatu *statement* itu benar atau salah (*True-False Item*), memasangkan-dua rentetan kata-kata (*Matching Item*) dan memilih di antara alternatif-alternatif jawaban yang telah disediakan untuk setiap butir soal (*Multiple Choice Item*).¹⁵⁾

Dua jenis ujian ini masing-masing memiliki kelemahan-kelemahan dan keunggulan-keunggulan. Di antara keunggulan yang dimiliki Ujian Bentuk Subjektif ialah bahwa Ujian Bentuk Subjektif itu jauh lebih mudah dalam penyusunan butir-butir soalnya, dan waktu yang diperlukan untuk menyusun soal-soalnya relatif lebih singkat daripada Ujian Bentuk Objektif. Kecuali itu, lebih singkat daripada Ujian Bentuk Objektif. Kecuali itu, untuk menyelidiki kemampuan mengorganisir pikiran anak didik lebih mudah diungkap dengan Ujian Bentuk Subjektif ini.¹⁶⁾ Berbeda dengan Ujian Bentuk Objektif maka Ujian Bentuk Subjektif dipandang lebih unggul dalam hal memberikan kelonggaran-kelonggaran bagi testee untuk menunjukkan *originalitas* dan *kreativitas* testee di dalam mendekati persoalan, dalam mengorganisir pikiran dan dalam mengekspresikannya secara teratur dan tepat.¹⁷⁾

Dalam pada itu, Ujian Bentuk Objektif pun banyak memiliki keunggulan jika dibandingkan dengan Ujian Bentuk Subjektif.

Keunggulan *pertama* ialah, bahwa Ujian Bentuk Objektif hasilnya dapat dianalisis secara objektif pula. Lebih dari itu, dengan menggunakan teknik statistik, Ujian Bentuk Objektif akan dapat memberikan berbagai macam informasi tentang kemampuan anak didik yang kita uji.¹⁸⁾

Kedua, bahwa apabila dibandingkan dengan Ujian Bentuk Subjektif, maka Ujian Bentuk Objektif pada umumnya dapat lebih banyak memenuhi prinsip-prinsip dasar pengukuran, terutama prinsip objektivitas. Ujian Bentuk Subjektif biasanya memiliki kadar objektivitas yang amat rendah. Hasil-hasil penelitian menunjukkan bahwa angka indeks korelasi antara dua orang pemeriksa hasil ujian Bentuk Subjektif adalah sekitar 0,50. Ini merupakan petunjuk bahwa "*agreement*" yang terdapat antara *scorers* (para pemberi skor = korektor) pada umumnya sangat rendah. Ketidakcocokan penilaian ini bukan hanya terjadi pada dua orang *scorers* saja, bahkan seorang *scorer* untuk memeriksa hasil ujian yang sama dalam waktu yang berbeda, seringkali memberikan hasil

15) *Ibid.*, hal. 85.

16) Lihat: Prof. Dr. Masrun, M.A., *op.cit.*, hal. 12.

17) Lihat: Prof. Dr. T. Raka Joni M.Sc., *op.cit.*, hal. 77.

18) Lihat: Prof. Dr. Masrun, M.A., *op.cit.*, hal. 13.

yang berbeda pula. Kelemahan lain yang terdapat dalam Ujian Bentuk Subjektif dalam hubungannya dengan penyekoran ialah adanya kecenderungan pada diri *scorer* untuk memberi nilai mendekati rata-rata. Dengan kata lain, disini dapat terjadi adanya *Central Tendency Effects*. Sebagai contoh, misalnya apabila rentangan nilai bergerak antara 0 sampai dengan 10, maka seorang *scorer* pada umumnya lebih cenderung untuk memberi nilai yang bergerak sekitar 5 dan 6.¹⁹⁾

Ketiga, Ujian Bentuk Objektif lebih banyak dapat memuat bahan atau materi ujian yang lebih luas dan lebih representatif ketimbang Ujian Bentuk Subjektif. Hal ini disebabkan karena items dalam Ujian Bentuk Objektif jauh lebih banyak ketimbang items pada Ujian Bentuk Subjektif. Dengan demikian Ujian Bentuk Objektif dapat mengurangi atau memperkecil adanya kesalahan-kesalahan sampel yang tidak representatif. Dengan adanya items yang cukup banyak itu, maka Ujian Bentuk Objektif dapat mendorong peserta didik untuk membangun suatu latar belakang pengetahuan dan kemampuan yang luas.²⁰⁾

Keempat, pada Ujian Bentuk Objektif, penyekoran (penghitungan jumlah jawaban betul dengan memperhitungkan bobot jawabannya), jauh lebih mudah ketimbang Ujian Bentuk Subjektif. Ujian Bentuk Objektif dapat diskor oleh setiap orang, asalkan ia telah mengerti cara menggunakan kunci jawabannya. Bahkan dengan perkembangan teknologi sekarang ini, hasil Ujian Bentuk Objektif telah dapat diskor dengan alat-alat elektronik; hal ini mempunyai arti yang sangat penting bila ditinjau dari segi pendidikan.²¹⁾

Demikianlah, baik Ujian Bentuk Subjektif maupun Ujian Bentuk Objektif sama-sama memiliki keunggulan dan kekurangan-kekurangan. Yang penting bagi kita adalah, bagaimana kita dapat memanfaatkannya dalam keadaan-keadaan yang sesuai dan bagaimana usaha yang perlu kita tempuh untuk memperkecil kelemahan-kelemahan yang disandang oleh kedua jenis ujian tersebut.

KEMUNGKINAN-KEMUNGKINAN TIMBULNYA KESALAHAN DALAM PENGUKURAN HASIL BELAJAR

Ada dua macam arah pandangan yang sangat merugikan efisiensi penunaian fungsi pengukuran dan penilaian hasil belajar, yaitu: (1) pandangan yang menganggap bahwa untuk melaksanakan pengukuran dan penilaian hasil belajar tidak diperlukan persiapan yang disengaja dan

¹⁹⁾ Lihat: Prof. Dr. Masrun, M.A., *op. cit.*, hal. 12.

²⁰⁾ *Ibid.*, hal. 13-14.

²¹⁾ *Ibid.*, hal. 14.

sistematik, sehingga siapa saja akan bisa melakukannya; (2) pandangan yang menganggap bahwa pengukuran dan penilaian hasil belajar merupakan kegiatan yang lepas dari kegiatan mengajar dan belajar, atau setidaknya merupakan kegiatan pengiring saja yang dilakukan setelah kegiatan mengajar dan belajar selesai dilaksanakan. Kedua pandangan ini tentu saja dapat menjadi penyebab tidak dilaksanakannya pengukuran dan penilaian hasil belajar sesuai dengan dasar-dasar fikiran yang seharusnya, bahkan secara sengaja atau tidak disengaja, kemungkinan dapat terjadi cara-cara pelaksanaan yang menyimpang dari prinsip-prinsip dasar pengukuran dan penilaian hasil belajar.²²⁾

Dalam usaha menilai hasil perbuatan belajar dari para peserta didik yang kita asuh, biasanya kita mengadakan pengukuran yang tertuang dalam bentuk ujian-ujian, baik ujian tertulis maupun ujian lisan. Dari ujian-ujian itu kita memberikan skor, yang -- walaupun adakalanya kurang tepat -- kita beri nama "nilai", atau nilai mana diwujudkan dalam bentuk angka-angka.

Mungkin dalam memberikan nilai tersebut timbul pertanyaan dalam diri kita sendiri: Apakah nilai yang kita berikan kepada peserta didik itu betul-betul sama atau dapat dianggap sama dengan nilai yang sebenarnya -- yaitu nilai yang benar-benar mencerminkan prestasi belajar dari peserta didik yang bersangkutan -- atukah tidak. Bila pertanyaan itu muncul, besar kemungkinannya kita akan memberikan jawaban: "Tidak". Memang pada umumnya kita mengetahui bahwa nilai yang kita berikan tidak selalu merupakan pencerminan yang tepat dari prestasi belajar yang sebenarnya dari peserta didik yang bersangkutan.

Sebagai seorang pengajar kita selalu menyadari tentang kemungkinan adanya perbedaan antara nilai dari hasil pengukuran dengan nilai yang sebenarnya. Ada dua macam kemungkinan yang dapat menyebabkan timbulnya perbedaan tersebut.

Pertama, mungkin nilai yang kita berikan kepada peserta didik, yang bersumber dari hasil pengukuran (ujian) adalah *lebih rendah* ketimbang nilai yang sebenarnya.

Kedua, mungkin nilai yang kita berikan itu *lebih tinggi* ketimbang nilai yang sebenarnya.

Bila salah satu dari kedua kemungkinan ini terjadi, maka timbullah *kesalahan pengukuran* yang sering disebut *error*.²³⁾

Dari uraian di atas dapatlah dipahami bahwa kesalahan pengukuran timbul apabila ada selisih antara nilai yang kita berikan kepada peserta

²²⁾ Lihat: Prof. Dr. T. Raka Joni M.Sc., *op.cit.*, hal.1.

²³⁾ Lihat: Prof. Dr. Masrun, M.A., *op.cit.*, hal. 2-3.

didik dengan nilai sebenarnya yang ada pada peserta anak didik itu. Jika selisih itu timbul karena nilai yang kita berikan *lebih tinggi* ketimbang nilai sebenarnya, maka selisih itu dinamakan selisih lebih atau selisih positif; sebaliknya jika nilai yang kita berikan itu adalah lebih rendah ketimbang nilai sebenarnya, maka selisih itu dinamakan selisih kurang atau selisih negatif.²⁴⁾

Demikianlah, kesalahan pengukuran sebenarnya dapat saja terjadi dalam segala bentuk pengukuran dalam semua bidang. Bahkan dalam bidang ilmu eksakta sekalipun sering terjadi "ketidak-eksakan" dari hasil pengukuran. Kesalahan itu mungkin disebabkan karena alat-alatnya yang "kurang eksak" dan mungkin pula karena pengamatan yang kurang teliti. Sudah barang tentu pengukuran dalam bidang pendidikan, khususnya pengukuran mengenai hasil belajar, di mana subjek yang diukur adalah makhluk hidup, kesalahan pengukuran itu adalah jauh lebih kompleks sifatnya dan jauh lebih banyak sumber-sumber penyebabnya.

BEBERAPA SUMBER KESALAHAN PENGUKURAN DALAM UJIAN BENTUK OBJEKTIF DAN UJIAN BENTUK SUBJEKTIF

Pada pembicaraan terdahulu telah dikemukakan, bahwa kesalahan pengukuran bisa terjadi apabila ternyata antara skor hasil pengukuran dengan skor yang sebenarnya terdapat perbedaan atau selisih, baik selisih positif maupun selisih negatif. Jadi adanya perbedaan antara skor hasil pengukuran dengan skor yang seharusnya diberikan oleh *tester* kepada *testee* adalah merupakan indikator bahwa dalam proses pengukuran telah terjadi kesalahan atau *error*.

Yang menjadi persoalan sekarang ialah, bagaimana kesalahan pengukuran itu bisa terjadi dan sumber kesalahan itu terletak di mana. Salah satu pendapat menyatakan, bahwa banyak sekali sumber-sumber kesalahan itu; namun empat di antaranya yang dipandang paling erat hubungannya dengan kesalahan pengukuran itu, yakni:²⁵⁾ (1) kesalahan pengukuran yang bersumber dari kesalahan *sampling*, (2) kesalahan pengukuran yang bersumber dari kesalahan *scoring*, (3) kesalahan pengukuran yang bersumber dari kesalahan *ranking*, dan (4) kesalahan pengukuran yang bersumber dari kesalahan *guessing*.

Yang dimaksud dengan *kesalahan sampling* di sini ialah, kesalahan yang dilakukan oleh *tester* (fihak penguji) di dalam memilih dan menentukan butir-butir item sebagai sampel atau wakil dari materi

²⁴⁾ Lihat: *Ibid.*, hal.3.

²⁵⁾ Lihat: Prof. Dr. H. M. Ismadi, "Multiple Choice Items", dalam: *Pengukuran Pendidikan*, Yogyakarta: Universitas Gadjah Mada, 1989, hal. 50-53.

pelajaran yang seharusnya diujikan. Seperti diketahui dalam ujian adalah tidak mungkin bagi penguji untuk membuat butir-butir soal yang materinya terdiri dari keseluruhan bahan yang telah diajarkan; sebab apabila semua bahan ditanyakan dalam ujian, maka butir-butir soal akan menjadi sangat banyak, sehingga tidak mungkin akan dapat dijawab atau diselesaikan oleh testee. Itulah sebabnya mengapa pada setiap kegiatan ujian, pihak penguji hanya akan mengambil beberapa saja dari sekian banyak bahan yang seharusnya diujikan, atau dengan kata lain pihak penguji melakukan *sampling* terhadap keseluruhan materi yang harus diujikan (sebagai populasi). Di sini pihak penguji membuat perkiraan bahwa dengan hanya menggunakan sebagian saja dari keseluruhan materi yang seharusnya diujikan itu, ia menyusun butir-butir soal yang menurut keyakinannya cukup representatif untuk mewakili keseluruhan materi itu. "Oleh karena butir-butir soal ujian itu dijabarkan dari suatu perkiraan berdasarkan suatu sampel, maka kesalahan pengukuran akan segera terjadi apabila pemilihan dan penentuan sampel tidak dilaksanakan secara sistematik dan representatif."²⁶⁾

Ditilik dari segi ini, maka Ujian Bentuk Subjektif mengandung resiko kesalahan pengukuran yang lebih besar daripada Ujian Bentuk Objektif. Alasannya kiranya cukup jelas, yaitu bahwa pada Ujian Bentuk Subjektif, di mana butir-butir itemsnya amat terbatas, maka keterbatasan luasnya sampel ini adalah merupakan sumber kesalahan pengukuran. Jika dikehendaki sampel yang luas, maka butir-butir items yang diajukan dalam ujian akan menjadi banyak sekali; suatu hal yang tidak mungkin dilakukan atau ditempuh pada Ujian Bentuk Subjektif. Pada Ujian Bentuk Objektif, kesalahan pengukuran yang bersumber dari kesalahan *sampling* dapat lebih diperkecil. Alasannya juga cukup jelas, yaitu bahwa pada Ujian Bentuk Objektif itu dimungkinkan bagi pihak penguji untuk mengajukan butir-butir items yang cukup banyak, sehingga karenanya pengambilan sampel dapat dilakukan dengan secara lebih luas ketimbang pada Ujian Bentuk Subjektif.

Sumber kesalahan yang *kedua* adalah *kesalahan scoring*. Yang dimaksud dengan kesalahan *scoring* ialah kesalahan pengukuran yang bersumber dari kesalahan pihak penguji dalam memberikan skor terhadap jawaban-jawaban betul yang diberikan oleh testee, atas butir-butir items yang diajukan dalam suatu ujian.

Ditilik dari segi ini, maka resiko kesalahan pengukuran sebagai akibat dari kesalahan *scoring* pada Ujian Bentuk Objektif jauh lebih kecil jika dibandingkan dengan Ujian Bentuk Subjektif. Hal ini disebabkan

²⁶⁾ Lihat: *ibid.*, hal. 50.

karena pada Ujian Bentuk Objektif, scoring dapat dilakukan dengan lebih cepat (dapat dilakukan menggunakan alat-alat bantu elektronik semacam komputer dan lain-lain), lebih Objektif (karena setiap butir item telah disediakan kunci jawabannya), disamping hasil scoring itu akan selalu sama walaupun scoring dilakukan oleh orang yang berlainan dan dalam waktu yang berbeda. Tidak demikian halnya dengan Ujian Bentuk Subjektif; scoring hasil Ujian Bentuk Subjektif yang dilakukan oleh scorer tidak selalu sama antara testee yang satu dengan testee yang lain. Di sini, pemberian skor di samping memakan waktu yang cukup lama, juga hasil scoring kurang dapat dijamin akan selalu sama; sebab di sini pemberian skor dapat dipengaruhi oleh atau tergantung pada keadaan para scorers pada saat memeriksa atau mengoreksi hasil Ujian Bentuk Subjektif tersebut. Misalnya, skor yang diberikan untuk jawaban butir soal yang sama kepada testee "A" yang kebetulan tulisannya baik, berbeda dengan skor yang diberikan kepada testee "B" yang kebetulan tulisannya jelek, padahal jawaban yang diberikan oleh kedua orang testee tersebut tidak berbeda. Faktor kelelahan, rasa susah, gembira, hallo effect dan sebagainya yang melekat pada diri scorer pun dapat mengakibatkan pemberian skor yang tidak selalu sama.

Sumber potensial *ketiga* dari kesalahan pengukuran hasil belajar ialah *kesalahan ranking*, yaitu kesalahan yang diperbuat oleh scorer dalam menentukan urutan kedudukan skor yang diberikan kepada testee terhadap butir-butir items yang dikemukakan dalam ujian.

Pada Ujian Bentuk Objektif, penentuan ranking dapat dilakukan berdasar distribusi frekuensi skor hasil ujian. Dengan menggunakan Tabel Distribusi Frekuensi itu dapat dilakukan pengukuran tendensi sentral skor hasil ujian; misalnya dengan cara mencari nilai rata-rata hitung yang dapat dipandang sebagai pencerminan kemampuan umum testee, yang juga dikenal dengan istilah rata-rata kelas. Juga dapat dilakukan pengukuran variabilitasnya; misalnya dengan mencari Deviasi Standar skor-skor hasil ujian, yang dapat melukiskan variasi rata-rata dari prestasi testee dalam ujian tersebut. Atas dasar itu dapat ditentukan secara objektif, urutan kedudukan atau ranking dari masing-masing testee.

Pada Ujian Bentuk Subjektif, hal serupa itu sangat sulit dilakukan, sebab penentuan ranking harus dilakukan dengan terlebih dahulu membaca dan membandingkan satu demi satu, jawaban testee terhadap butir-butir soal yang sama. Jadi di sini penguji membuat *ranking score* untuk masing-masing testee, dengan membandingkan atau menghubungkannya dengan skor-skor yang dicapai oleh teman-temannya untuk masing-masing butir item yang bersangkutan. Karena jawaban item berbentuk uraian, yang umumnya cukup panjang dan waktu untuk memeriksa jawaban itu

biasanya terbatas, maka dalam keadaan seperti itu kemungkinan terjadinya kesalahan pengukuran menjadi lebih besar.

Kesalahan karena *guessing* (menebak) juga merupakan sumber penyebab terjadinya kesalahan dalam pengukuran hasil belajar. Kesalahan karena menebak itu bisa muncul baik pada Ujian Bentuk Objektif maupun pada Ujian Bentuk Subjektif; hanya saja ada perbedaan diantara keduanya dalam hal "apa yang ditebak".

Pada Ujian Bentuk Objektif -- khususnya yang menggunakan model *Multiple Choice Items* -- yang ditebak oleh testee adalah jawaban yang betul (=kunci jawaban) dari rangkaian kemungkinan jawaban (alternatif) yang telah disediakan untuk masing-masing butir item. Sedangkan pada Ujian Bentuk Subjektif, yang ditebak oleh testee adalah subjek-subjek atau materi-materi yang mungkin akan ditanyakan atau dikeluarkan dalam ujian, dan itulah yang dipelajari dengan sungguh-sungguh. Adapun subjek atau pun materi yang diperkirakan tidak akan ditanyakan dalam ujian, kurang menarik untuk dipelajari. Kalaupun dipelajari, itu hanya secara sepintas lalu saja. Yang dijadikan dasar berpijak atau alasan oleh testee dalam hal ini terutama adalah kesenangan atau kegemaran atau fihak penguji terhadap subjek-subjek atau materi-materi tertentu. Misalnya: Penguji "X" lebih menyukai untuk mengeluarkan jenis items yang mengungkap aspek kognitif saja, sedangkan penguji "Y" lebih sering menyukai jenis items yang mengungkap aspek afektif.

Jika ternyata bahwa penebakan-penebakan yang dilakukan oleh testee itu menghasilkan jawaban betul, padahal betulnya jawaban itu hanya secara kebetulan saja (by chance), maka skor yang dihasilkan dari pengukuran lewat ujian itu sebenarnya bukan merupakan skor yang mencerminkan kemampuan yang dimiliki oleh testee; karena itu dalam skor hasil pengukuran itu telah terkandung adanya *error* atau kesalahan.

Selain yang telah dikemukakan di atas, kesalahan pengukuran juga bisa terjadi karena: (1) situasi dan kondisi fisik dan psikis yang menyelimuti diri testee pada saat mengerjakan pekerjaan ujian (suasana hati yang kalut, sedih, tertekan, menderita sakit) dan (2) situasi ujian yang tidak menguntungkan (kegaduhan di dalam maupun di luar ruang ujian, pengawasan ujian yang terlalu ketat atau terlalu longgar).²⁷⁾ Dengan situasi dan kondisi seperti telah dikemukakan tadi, maka testee yang semestinya mendapat skor yang tinggi, menjadi rendah ("jatuh" dalam ujian).

²⁷⁾ Lihat: Dr. Suharsimi Arikunto, *Dasar-dasar Evaluasi Pendidikan*, Jakarta: Bumi Aksara, Cetakan ke-VII, 1991, hal. 14-16.

PENUTUP

Seperti dapat diikuti melalui lembar-lembar makalah ini maka pengukuran hasil belajar lewat ujian yang dilaksanakan oleh para pengajar (guru, dosen dan lain-lain) dalam rangka penilaian terhadap kemajuan dan perkembangan peserta didik yang menjadi asuhan dan tanggung jawabnya masing-masing, sepintas lalu kelihatannya hanya merupakan hal yang amat sederhana dan bersifat rutin saja. Namun apabila dilacak dan ditelusuri lebih dalam, ternyata bahwa banyak hal dan banyak segi yang perlu mendapatkan perhatian yang layak.

Dari uraian di muka dapat disimpulkan, bahwa kesalahan dalam pengukuran hasil belajar bukanlah semata-mata bersumber dari dalam diri *peserta didik* sendiri yang sebagai makhluk hidup sering mengalami perubahan, melainkan juga bersumber pada diri *pendidik* (pengajar, guru, dosen) yang sebagai manusia biasa juga dapat terpengaruh oleh berbagai faktor di dalam maupun di luar dirinya, *suasana* atau *situasi* dan *kondisi* pada saat berlangsungnya ujian, dan tidak kalah pentingnya adalah bahwa kesalahan pengukuran itu justru bersumber dari *alat pengukurnya* sendiri.

Beberapa saran berikut ini kiranya dapat dijadikan pedoman dalam rangka pengukuran hasil belajar peserta didik (khususnya para mahasiswa yang berada dalam asuhan kita):

1. *Soal-soal ujian* yang akan dijadikan sebagai alat pengukur untuk mengetahui tingkat pengetahuan dan keberhasilan belajar peserta didik hendaknya disusun dan dipersiapkan terlebih dahulu dengan perencanaan yang cermat dan matang, sehingga hasil pengukuran yang diperoleh melalui ujian-ujian itu dapat dijadikan bahan yang akurat dalam menentukan nilai hasil belajar mereka.
2. Dengan memperhatikan segi-segi keunggulan maupun segi-segi kelemahan yang dimiliki oleh Ujian Bentuk Objektif dan Ujian Bentuk Subjektif, adalah sangat bijaksana apabila dalam pengukuran hasil belajar mahasiswa tidak hanya digunakan salah satu diantara kedua jenis ujian itu, melainkan bahwa kedua jenis ujian itu hendaknya dapat dipergunakan secara serempak atau berbareng. Dengan cara demikian maka kelemahan-kelemahan yang disandang oleh Ujian Bentuk Objektif dapat dieleminir dengan keunggulan-keunggulan yang dimiliki oleh Ujian Bentuk Subjektif, demikian pula sebaliknya. Dengan cara demikian maka kesalahan-kesalahan dalam pengukuran akan menjadi mengecil.
3. Untuk mengurangi atau memperkecil resiko kesalahan pengukuran yang disebabkan karena *kesalahan sampling*, sebelum butir-butir soal disusun, sebaiknya dibuat Tabel Spesifikasi atau Blue Print atau Kisi-kisi Soal, yang di dalamnya dapat diperhitungkan perimbangan

yang lebih representatif antara aspek-aspek psikologis yang ingin diukur (ranah kognitif, ranah afektif dan ranah psikomotorik (aplikasi), dan materi pengajaran yang akan diwakili oleh butir-butir soal ujian. Berikut ini adalah salah satu contohnya:

Aspek Psiko- logis Materi	Hafalan/ Ingatan (50%)	Pemahaman (30%)	Aplikasi (20%)	Jumlah
Bab I (10%)	3	2	1	6
Bab II (20%)	6	4	2	12
Bab III (25%)	8	4	3	15
Bab IV (30%)	9	5	4	18
Bab V (15%)	4	3	2	9
TOTAL	30	18	12	60

4. Khusus untuk staf pengajar yang hanya menggunakan Ujian Bentuk Subjektif, agar kesalahan pengukuran yang disebabkan karena kesalahan *scoring* dapat dibatasi seminimal mungkin, maka sangat dianjurkan agar bersamaan waktu dengan penyusunan butir-butir soal Ujian Bentuk Subjektif disusun pula pedoman atau ancar-ancar

jawaban untuk masing-masing butir soal yang bersangkutan. Pedoman atau ancar-ancar jawaban itu diharapkan akan dapat berfungsi seolah-olah sebagai kunci jawaban seperti yang ada pada butir-butir soal Ujian Bentuk Objektif. Dengan cara demikian, scoring dapat dilakukan dengan lebih objektif, adil dan perbedaan skor yang diberikan untuk butir soal yang sama dengan testee yang berbeda, akan dapat dipertahankan keajegannya.

5. Dalam hal mengatasi kesalahan pengukuran sebagai akibat dari kesalahan *ranking*, bagi staf pengajar yang dalam pengukuran hasil belajar menggunakan Ujian Bentuk Objektif, sangat dianjurkan untuk dapat menggunakan atau memanfaatkan ukuran-ukuran statistik berupa Arithmetic Mean (Rata-rata Hitung) dan Standard Deviation (Deviasi Standar = Simpangan Baku) yang dalam Ilmu Statistik dikenal sebagai ukuran dengan kadar ketelitian yang tinggi.
6. Untuk mengatasi kesalahan pengukuran yang bersumber dari kesalahan *guessing*, salah satu cara yang dapat ditempuh adalah sebagai berikut:
 - a. Jenis-jenis soal yang akan dikemukakan dalam Ujian Bentuk Subjektif hendaknya tidak bersifat "seragam" atau "yang itu-itu juga", melainkan dibuat atau disusun variatif. Kecuali itu, pada ujian yang berbeda waktu harus dapat diusahakan jenis soal ujian yang berbeda pula, sehingga sulit bagi testee untuk menebak-menebak mana kira-kira soal yang akan dikeluarkan dalam Ujian Bentuk Subjektif itu.
 - b. Untuk jenis soal yang akan dikemukakan dalam Ujian Bentuk Objektif, jika dipandang perlu supaya dibuat suatu ketentuan yang bersifat peringatan, bahwa setiap jawaban yang salah akan dikenai denda berupa pengurangan skor. Dengan cara demikian maka setiap jawaban yang akan diberikan oleh testee adalah merupakan jawaban yang telah difikirkan dan dipertimbangkan secara masak dan cukup berhati-hati; bukan jawaban yang sekedar merupakan tebakan atau bersifat spekulasi belaka.

Demikianlah kajian yang dapat saya kemukakan semoga ada manfaatnya.