

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Dalam praktik pendidikan, guru senantiasa dihadapkan pada keputusan-keputusan dalam memberikan label pada setiap karakteristik atribut siswa. Pemberian atribut tersebut dilakukan sebagai hasil dari pengukuran yang dilakukan dalam kegiatan proses pembelajaran di kelas. Keputusan-keputusan tersebut, tidak hanya akan berdampak terhadap pengelolaan pembelajaran di kelas, tetapi juga terhadap efektifitas program-program untuk meningkatkan kemampuan siswa melalui program pengayaan (*enrichment*), program perbaikan (*remedial teaching*), kebutuhan siswa terhadap program bimbingan dan penyuluhan, dan lebih jauhnya lagi terhadap pengambilan kebijakan di tataran tingkat sekolah.

Pengukuran terhadap berbagai karakteristik atribut siswa dilakukan melalui kegiatan penilaian kelas (*class assessment*). Penilaian kelas merupakan bagian yang integral dari keseluruhan proses pembelajaran yang dilakukan oleh guru sepanjang rentang waktu berlangsungnya proses pembelajaran. Penilaian kelas harus memiliki tujuan yang sejalan dengan tujuan pembelajaran yang telah ditetapkan dalam rencana program pengajaran. Hasil dari kegiatan penilaian tersebut digunakan untuk memperoleh pemahaman terhadap pelaksanaan dan umpan balik (*feedback*) terhadap hasil pembelajaran yang telah dilaksanakan, membuat keputusan berkenaan dengan pelaksanaan dan hasil pembelajaran, dan meningkatkan kualitas proses dan hasil pembelajaran. Mengingat urgensi penilaian (*assessment*) seperti yang dijelaskan di atas, maka perencanaan dan pengadministrasian terhadap kegiatan penilaian (*assesment*) tersebut harus dilakukan secara cermat, tepat, dan terarah.

Hasil akhir dari penilaian adalah berupa skor. Skor inilah yang kemudian menjadi nilai atau atribut yang akan diberikan oleh guru terhadap keberhasilan belajar siswa. Terkait dengan aspek skor tersebut, dalam kurikulum tingkat satuan

Engkos Kosasih, 2014

**KOMPARASI METODA NEDELSKY DAN ANGOFF DALAM PENETAPAN STANDARD SETTING
KELULUSAN UJIAN NASIONAL MATA PELAJARAN BAHASA INGGRIS SMA**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

pendidikan (KTSP), maupun kurikulum 2013, guru telah mengenal istilah kriteria ketuntasan minimal (KKM). Sebagaimana yang dinyatakan oleh Badan Standar Penilaian Nasional (BSNP) dalam panduan penyusunan kurikulum tingkat satuan pendidikan jenjang pendidikan dasar dan menengah dijelaskan bahwa penetapan kriteria minimal ketuntasan belajar tersebut merupakan tahapan awal pelaksanaan penilaian hasil belajar sebagai bagian dari langkah pengembangan KTSP. Dengan demikian Kurikulum KTSP yang berbasis kompetensi dan kurikulum 2013 dalam implementasinya harus menggunakan acuan kriteria dalam penilaian dan mengharuskan pendidik dan satuan pendidikan menetapkan kriteria minimal yang menjadi tolok ukur pencapaian kompetensi.

Penetapan KKM tidak hanya dilakukan untuk jenis-jenis tes seperti ulangan harian (*formative test*), ulangan semester (*summative test*) tetapi juga untuk tes yang berskala nasional misalnya ujian nasional (UN). Dalam UN kriteria ketuntasan minimal dijadikan sebagai standar dan acuan dalam penentuan kelulusan siswa, dimana kriteria kelulusan siswa tersebut dinyatakan dalam indikator butir-butir pencapaian pada standar kompetensi kelulusan (SKL). SKL tersebutlah yang kemudian menjadi patokan terhadap standar penguasaan atau standar kompetensi yang harus dicapai oleh peserta tes untuk memperoleh kriteria kelulusan.

Namun demikian, kondisi empiris yang terjadi di Indonesia menunjukkan bahwa standar minimum kelulusan UN masih merupakan gabungan antara nilai UN murni dengan nilai US, dengan proporsi 60% untuk nilai UN murni dan 40% untuk nilai US. Penggabungan nilai tersebut dalam berbagai sisi masih memiliki banyak kelemahan. Kelemahan yang pertama adalah penggabungan nilai UN dan US mengandung makna bahwa semua peserta didik akan lulus. Kalau semua peserta didik akan lulus artinya pendidikan tersebut tidak bermutu. Dengan demikian penentuan standar kelulusan UN yang mendasarkan kepada nilai gabungan tersebut menjadi kontraproduktif dengan tujuan dilaksanakannya UN tersebut.

Sisi kelemahan yang kedua adalah bahwa interferensi guru/sekolah melalui nilai US tersebut adalah bahwa nilai akhir UN tidak mencerminkan tujuan

penyelenggaraan tes yang sebenarnya. Salah satu tujuan diadakan tes adalah untuk mendapatkan gambaran sejauh mana peserta didik telah menguasai materi pembelajaran yang sudah disampaikan. Dengan makna lain adalah bahwa berapapun skor yang diperoleh oleh peserta didik akan selalu merefleksikan kemampuan peserta didik yang sebenarnya. Dengan memberikan nilai US yang relatif besar, walaupun tujuannya sebagai langkah antisipatif agar peserta tes mendapat peluang lulus lebih besar, dari sudut pandang teori pengukuran dan evaluasi pembelajaran tetap saja bertolak belakang dan tidak bisa dibenarkan.

Sisi kelemahan berikutnya adalah bahwa sekolah-sekolah tertentu yang berakreditasi A dan sekolah lainnya yang berakreditasi B atau C memiliki input siswa, kualitas sumber daya manusia (guru), fasilitas pendukung, dll yang berbeda. Perbedaan-perbedaan tersebut juga akan berpengaruh terhadap besarnya nilai US yang diberikan oleh sekolah/guru. Dengan demikian, penetapan kelulusan UN selayaknya tidak mengacu kepada hasil penggabungan nilai tetapi berdasarkan kepada skor minimum yang telah ditetapkan.

Sejalan dengan pemikiran yang dideskripsikan dalam kurikulum tingkat satuan pendidikan (KTSP) maupun kurikulum 2013, tentang penetapan KKM, peneliti telah mengenal istilah *standard setting*. *Standard setting* merupakan suatu penetapan yang telah ditafsirkan oleh panelist dalam bentuk skor minimal (*cut off score*) yang harus dicapai oleh peserta didik tentang aspek pengetahuan dan keterampilan yang memadai atau kompetensi minimum (*minimum competency*) yang harus dicapai. Kaitan antara *standard setting* dan UN adalah bahwa melalui *standard setting* akan ditetapkan skor minimum (*cut off score*) yang harus diperoleh peserta didik pada mata pelajaran yang diujikan nasional sebagai prasyarat untuk mendapatkan kelulusan. Penetapan *cut off score* tersebut lebih umum digunakan untuk penilaian yang menggunakan pendekatan acuan atau kriteria.

Secara garis besar, ada dua jenis pendekatan penilaian yaitu 1) pendekatan dengan menggunakan acuan norma (*norm referenced*), dimana dalam acuan norma interpretasi skor peserta tes dibandingkan dengan kemampuan individu yang lain dalam kelompok acuan, dan 2) acuan kriteria (*criterion referenced*),

dimana dalam acuan kriteria interpretasi skor dibandingkan dengan level kemampuan tertentu berdasarkan indikator-indikator pembelajaran.

Sebagai acuan bagi pemerintah dan sekolah/guru dalam menentukan kelulusan UN, dipandang perlu untuk menerapkan pendekatan penilaian acuan kriteria. Hal tersebut sejalan dengan prinsip penetapan KKM yang salah satunya adalah bahwa penetapan KKM merupakan kegiatan pengambilan keputusan yang dapat dilakukan melalui metode kualitatif dan atau kuantitatif. Metode kualitatif dapat dilakukan melalui *professional judgment* oleh pendidik yang kompeten di bidangnya dengan mempertimbangkan kemampuan akademik dan pengalaman pendidik dalam mengampu mata pelajaran tertentu di sekolahnya. Sedangkan metode kuantitatif dilakukan dengan rentang angka yang disepakati sesuai dengan penetapan kriteria yang ditentukan.

Secara teoritis, prinsip pedoman kriteria dalam PAP sendiri telah berkembang. Beberapa referensi menyebutkan bahwa ada beberapa metoda PAP yang dapat digunakan sebagai acuan dalam penetapan kelulusan suatu tes. Penetapan standar kelulusan yang selanjutnya disebut dengan *standard setting* terdiri atas dua kelompok. Kelompok yang pertama adalah penetapan *standard setting* yang berbasis kepada pertanyaan atau item tes (*judgment about test questions*) yang mendasarkan kepada *professional judgment* dalam penetapan *cut off score* nya. Metoda yang berbasis *judgment about test questions* tersebut adalah Metoda Angoff, Modified Angoff's (*Yes/No*), Nedelsky, dan Ebel. Sedangkan metoda yang mendasarkan kepada peserta didik (*test taker*) dalam menetapkan *cut off score* nya adalah metoda Bookmark dan Constrasting Group.

Metoda *standard setting* yang berbasis kepada pertanyaan tes (*judgment about test questions*) terdiri atas metoda Nedelsky, Angoff, Modified Angoff, dan Ebel. Metoda Ebel memiliki dua tahapan prosedur. Pada tahap pertama, tiap-tiap panelis mengklasifikasikan pertanyaan kedalam kelompok-kelompok dan kemudian membuat *judgment* dalam bentuk bilangan numerik tunggal untuk masing-masing kelompok pertanyaan tersebut. Pengklasifikasian pertanyaan tes kedalam kelompok tersebut berdasarkan pada dua jenis *judgment*, yaitu tingkat kesulitan soal (*difficulty level*), dan tingkat relevansi atau kepentingannya

(*relevance/importance*). Kemudian Ebel menganjurkan tiga tingkat kesulitan soal dengan label mudah (*easy*), sedang (*medium*), dan sukar (*hard*) dan empat kategori relevansi, yaitu; sangat penting (*essential*), penting (*important*), dapat diterima (*acceptable*), dan dipertanyakan (*questionable*).

Metoda Nedelsky dan Angoff sangat menarik dalam banyak konteks karena prosedur yang digunakan lebih mudah dipahami oleh panelis dan pengguna tes itu sendiri. Prosedur dalam kedua metoda tersebut mengharuskan panelis memberikan pertimbangan yang detail terhadap isi tes (*test content*) secara spesifik.

Peneliti telah mengkaji dan mengidentifikasi persamaan-persamaan prosedur yang dimiliki oleh kedua metoda tersebut. Namun demikian, walaupun terdapat beberapa persamaan, dalam beberapa hal kedua metoda tersebut memiliki perbedaan dalam menentukan *cut off score* nya. Perbedaan yang dianggap sangat penting dan mendasar tersebut yang mendorong peneliti untuk melakukan penelitian dan membandingkan kedua metoda tersebut sebagai bahan kajian dan referensi yang bermanfaat apabila kedua metoda *standard setting* tersebut dijadikan kriteria dalam menentukan kelulusan peserta tes.

Metoda Nedelsky, yang diusulkan oleh Leo Nedelsky pada tahun 1954, hanya dapat digunakan untuk tes dengan bentuk pilihan ganda (*multiple choice*), dimana tugas panelis adalah memberikan *judgment* terhadap tiap opsi butir soal yang salah atau yang tidak akan dijawab oleh siswa. Jumlah opsi butir soal yang dieliminasi tersebut akan menentukan proporsi peserta tes menjawab benar. Hasil akumulasi *judgment* terhadap keseluruhan opsi butir soal dijadikan sebagai *cut off score* ketuntasan/kelulusan dari peserta tes. Kalau dikaitkan dengan teori tes klasik, *judgment* yang dilakukan dengan metoda Nedelsky adalah *judgment* terhadap daya pengecoh (*distractor*) dan keberfungsian opsi butir soal (*item test functioning*). Dalam hal ini, panelis harus mampu menganalisis setiap opsi butir soal yang dianggap salah dan tidak akan dipilih oleh peserta tes.

Metoda Angoff, yang diusulkan oleh William H. Angoff pada tahun 1971, sama halnya dengan prosedur metoda Nedelsky tetapi metoda Angoff bisa diaplikasikan tidak hanya pada tes berbentuk pilihan ganda (*multiple choice*) tapi

juga pada bentuk tes lainnya. Dalam metoda Angoff, *cut off score* dihitung berdasarkan skor yang diharapkan dari tiap butir pertanyaan, tetapi tidak berdasarkan pada daya pengecoh (*distractor*). Panelis mempertimbangkan tiap-tiap butir pertanyaan secara keseluruhan dan membuat *judgment* terhadap proporsi peserta tes yang akan menjawab benar. Kalau dikaitkan dengan teori tes klasik, penentuan *judgment* dengan metoda Angoff lebih kepada taraf kesulitan (*difficulty level*) dari tiap butir soal. Semakin mudah butir soal, proporsi peserta tes menjawab benar akan semakin tinggi dan semakin sukar butir soalnya, proporsi peserta tes menjawab benar akan semakin rendah.

Penulis beranggapan bahwa kedua metoda tersebut bisa dijadikan acuan dalam penentuan kriteria kelulusan UN. Pertimbangan rasionalnya adalah bahwa jenis tes seperti UN merupakan jenis tes yang terstandarisasi dengan kualitas soal yang sangat baik. Perangkat soal dikonstruksi sedemikian rupa sehingga memenuhi kelayakan suatu tes. Namun demikian, untuk memastikan metoda mana yang paling tepat diaplikasikan dalam penetapan *standard setting* kelulusan, prosedur pertama yang harus ditempuh adalah melakukan perbandingan estimasi panelis terhadap perangkat tes tersebut. Hal tersebut dimaksudkan bahwa apabila dikaji dari tingkat kesulitan soal, berapa banyak probabilitas peserta tes yang diprediksi dapat lulus, dan apabila dikaji dari daya pengecoh (*distractor*) nya, berapa banyak probabilitas bagi peserta tes untuk dinyatakan lulus. Walaupun tingkat kesukaran soal (*difficulty level*) dengan proporsi 50% soal dengan kategori mudah, hal tersebut belum bisa dipastikan semua peserta tes akan mampu menjawab soal tes sebanyak 50% benar. Sebaliknya, dengan daya pengecoh (*distractor*) opsi pada butir soal, kemampuan peserta tes yang minimum pun belum bisa dipastikan dapat menjawab setiap butir soal dengan benar. Dengan demikian kedua metoda tersebut harus dibandingkan terlebih dahulu, sebelum sampai kepada keputusan metoda mana yang dianggap paling tepat untuk dijadikan acuan standar kelulusan dan memberikan proporsi paling banyak bagi peserta tes untuk dapat lulus.

Mempelajari dan mengkaji begitu banyak penelitian yang terkait dengan penetapan *standar setting* ketuntasan belajar, peneliti hanya sedikit menemukan

artikel atau jurnal penelitian yang mencoba membandingkan metoda Nedelsky dan metoda Angoff dalam penentuan *standard setting*. Muncul suatu gagasan mengapa harus metoda Nedelsky dan metoda Angoff yang harus dibandingkan? Berangkat dari suatu asumsi bahwa metoda nedelsky dan Angoff memiliki prosedur yang lebih mudah diaplikasikan dan lebih mudah dipahami oleh panelis dan pengguna tes, maka penulis memandang sangat perlu untuk melakukan penelitian yang mengkaji secara intensif dan komprehensif terhadap kedua metoda tersebut, dan kemudian membandingkan jumlah proporsi peluang kelulusan dari masing-masing metoda tersebut. Hasil temuan dari penelitian yang dilakukan diharapkan akan sangat bermanfaat dan berdaya guna, tidak hanya bagi pemerintah, tetapi juga bagi sekolah/guru atau praktisi kependidikan dalam menetapkan *standard setting* kelulusan yang tepat. Alasan rasional tersebut di atas yang mendorong peneliti untuk melakukan penelitian dengan membandingkan metoda Nedelsky dan metoda Angoff dalam penentuan *standard setting* kelulusan UN khususnya pada mata pelajaran bahasa Inggris.

Memahami adanya pendekatan, prinsip, dan prosedur yang berbeda antara metoda Nedelsky dan metoda Angoff dalam menentukan *standadr setting* kelulusan maka peneliti mengangkat judul **“Komparasi Metoda Nedelsky dan Angoff dalam Penentuan Standar Setting Kelulusan UN Pada Mata Pelajaran Bahasa Inggris SMA”** sebagai topik utama yang akan diteliti, dikaji, dianalisis, dan diterprestasi secara intensif dan komprehensif. Hasil temuan-temuan dalam penelitian digunakan untuk menginterpretasi dan menggeneralisasi metoda yang mana yang lebih tepat untuk diaplikasikan.

B. Identifikasi Masalah

Perbaikan terhadap penyelenggaraan dan pelaksanaan UN harus terus dilakukan untuk mencapai tujuan yang diharapkan dari penyelenggaraan UN tersebut. Salah satu revisi yang sangat mendasar dan harus segera diimplementasikan dalam pelaksanaan UN tersebut adalah sistem pendekatan penilaian yang digunakan dalam menetapkan *cut off score* kelulusan peserta tes.

Beberapa masalah yang dapat diidentifikasi dari sistem penilaian tersebut adalah sebagai berikut;

1. Terkait dengan nilai minimal pada penilaian akhir untuk semua mata pelajaran, dijelaskan tentang kriteria kelulusan peserta didik untuk UN, dimana nilai akhir (NA) merupakan gabungan nilai sekolah dan UN dengan bobot 40% nilai US dan 60% dari nilai UN. Berdasarkan pengalaman guru/pendidik bahwa pemberian nilai US oleh sekolah/guru tersebut tidak mencerminkan kemampuan peserta tes yang sesungguhnya.
2. Penentuan nilai akhir kelulusan juga ditentukan melalui kebijakan institusi/sekolah, musyawarah guru mata pelajaran (MGMP) yang masih belum mencerminkan tujuan dilaksanakannya tes. Sekolah yang berakreditasi A memiliki sumber daya (guru), input siswa, dan sarana dan prasarana yang lebih memadai dibandingkan dengan sekolah yang berakreditasi B atau C. Dengan demikian penetapan nilai akhir kelulusan masih menyesuaikan dengan kondisi di lapangan.
3. Penentuan nilai akhir kelulusan melalui *standard setting*, baik metoda yang mendasarkan kepada pertanyaan atau item tes (*judgment about test questions*) maupun metoda yang mendasarkan kepada peserta didik (*test taker*), menjadi alternatif terbaik dalam menetapkan nilai minimal kelulusan. Dengan demikian, maka pemerintah sebagai penyelenggara dan sekaligus penanggung jawab pelaksanaan UN dipandang perlu untuk merevisi sistem penilaian yang digunakan dengan cara mengimplementasikan metoda *standard setting* dalam penentuan batas minimal kelulusan.

C. Pembatasan Masalah

Sebagai kerangka acuan kajian, analisis, interpretasi, dan upaya memberikan ruang lingkup yang lebih sistematis dan terarah dalam pembahasan penelitian ini, maka masalah-masalah dalam penelitian ini dibatasi pada hal-hal berikut, yaitu;

1. Metoda *standard setting* yang secara khusus akan dibahas secara komprehensif dalam penentuan nilai akhir kelulusan adalah metoda Nedelsky dan Angoff.
2. Pembatasan pembahasan terhadap kedua metoda *standard setting* tersebut bertujuan untuk memudahkan peneliti dalam membuat interpretasi dan generalisasi hasil dari temuan-temuan yang diperoleh.

D. Perumusan Masalah

Rumusan masalah yang dikemukakan dalam penelitian ini adalah sebagai berikut:

1. Metoda *standard setting* yang manakah yang memberikan peluang lebih besar bagi peserta UN untuk dinyatakan lulus pada mata pelajaran bahasa Inggris SMA dengan kompetensi minimum yang dipersyaratkan?
2. Berapakah jumlah proporsi peluang kelulusan berdasarkan kedua metoda *standard setting* Nedelsky dan Angoff tersebut?

E. Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut, yaitu :

1. Untuk membandingkan hasil *judgment* panelis terhadap dua metoda *standard setting* dalam menentukan *cut off score* kelulusan UN mata pelajaran bahasa Inggris SMA.
2. Memperoleh gambaran tentang skor baku (*standardized score*) peserta tes hasil dari respon peserta tes terhadap UN bahasa Inggris SMA dengan *cut off score* hasil *judgement* panelis terhadap dua metoda *standard setting*, yaitu Nedelsky dan Angoff.
3. Menganalisis proporsi banyaknya peserta tes yang lulus berdasarkan masing-masing metoda *standard setting* tersebut, dan kemudian membuat interpretasi dan generalisasi terhadap metoda *standard setting* yang manakah yang memberikan peluang yang lebih besar bagi peserta tes untuk dinyatakan lulus UN pada mata pelajaran bahasa Inggris SMA.

F. Kegunaan Penelitian

Kegunaan hasil penelitian ini dideskripsikan sebagai berikut:

1. Kegunaan Secara Teoritis

Kegunaan penelitian secara teoritis adalah bahwa hasil kajian penelitian ini akan memperkaya khasanah pengetahuan dan perkembangan ilmu pendidikan, terutama dalam pengembangan metoda pendekatan penilaian sehingga hasil temuan penelitian akan menjadi bahan rujukan dalam menentukan *cut off score*, baik untuk kepentingan kelulusan UN atau tes-tes lain yang terstandardisasi, maupun dalam penentuan ketuntasan belajar (*mastery learning*) melalui penetapan kriteria ketuntasan minimal (KKM).

2. Kegunaan Secara Praktis

- a. Pada tataran tingkat sekolah, hasil penelitian ini dapat memberikan kontribusi yang positif bagi pemerintah, sekolah/guru, dan praktisi kependidikan dalam hal mengaplikasikan dan mengimplementasikan metoda *standar setting* yang digunakan dalam penilaian (*assessment*), dan untuk menyelesaikan masalah-masalah praktis dalam penilaian hasil pembelajaran.
- b. Hasil penelitian memungkinkan untuk didesiminasikan oleh guru-guru bahasa Inggris melalui kegiatan MGMP sebagai upaya untuk memberikan input yang bermamfaat dalam membuat keputusan hasil pendidikan dan memperbaiki praktik penilaian hasil pembelajaran supaya lebih efektif dan tepat sasaran.