

## BAB I

### PENDAHULUAN

#### A. Latar Belakang Masalah

Evaluasi hasil belajar merupakan kegiatan yang sangat penting dalam proses pendidikan. Evaluasi ini selalu dilaksanakan dalam proses pendidikan dan berdampak besar terhadap siswa, guru, orangtua siswa, dan masyarakat. Tujuan evaluasi ini adalah untuk menentukan sejauhmana ketercapaian tujuan pendidikan yang tercantum dalam kurikulum atau tingkat penguasaan kemampuan yang diperoleh siswa dalam suatu program pendidikan (Tyler, 1979; Anastasi, 1988, Mardapi, 2000). Kendati fungsi utamanya adalah untuk menentukan tingkat kemampuan siswa dan kelulusan (Arikunto, 1986), namun dalam prakteknya evaluasi ini juga sering difungsikan sebagai alat seleksi. (misal NEM SD untuk seleksi masuk SLTP dan NEM SLTP untuk seleksi masuk SMU), sebagai dasar pertimbangan pengambilan kebijakan (Rutman, 1980; Hasan, 1988) misalnya hasil EBTANAS digunakan sebagai salah satu dasar pertimbangan pengambilan kebijakan pembubaran Proyek Perintis Sekolah Pembangunan (PPSP) pada tahun 1986, di samping sebagai salah satu bentuk pertanggungjawaban/ akuntabilitas sekolah terhadap orangtua siswa. Pada hakikatnya dasar/ landasan dilaksanakannya evaluasi ini adalah untuk meningkatkan kualitas pendidikan khususnya peningkatan proses belajar-mengajar (McCormick dan James (1983). Bagi siswa yang tingkat kemampuannya belum mencapai kriteria lulus agar

belajar lebih giat, bagi siswa yang telah lulus juga agar belajar lebih giat supaya dapat berhasil pada tingkatan pendidikan berikutnya, sedang bagi guru agar meningkatkan kualitas proses pembelajaran yang masih belum baik dan mempertahankan proses pembelajaran yang sudah baik.

Berdasarkan kurikulum 1994 (Depdikbud, 1995 : 2), tujuan pendidikan IPA (Fisika) di SMU adalah agar siswa mampu menguasai konsep-konsep Fisika dan saling keterkaitannya serta mampu menggunakan metode ilmiah yang dilandasi sikap ilmiah untuk memecahkan masalah-masalah yang dihadapinya sehingga lebih menyadari keagungan Tuhan Yang Maha Esa. Berdasarkan ketentuan ini, maka tujuan pembelajaran IPA (Fisika) di SMU dapat diklasifikasikan ke dalam tiga aspek yakni produk, proses, dan sikap IPA. Produk IPA berupa penguasaan konsep-konsep Fisika dan saling keterkaitannya. Proses IPA berupa penerapan metode ilmiah, yang pada hakikatnya merupakan pengejawantahan dari proses-proses seperti merencanakan percobaan, mengajukan hipotesis, mengamati, mengklasifikasi, mengukur, mengontrol variabel, menghitung, mencari hubungan, menginterpretasi data, menyimpulkan, meramalkan, menerapkan, dan mengkomunikasikan hasil. Sikap IPA berupa sikap ilmiah yang melandasi penerapan metode ilmiah. Tujuan tersebut juga dapat diklasifikasikan sesuai taksonomi Bloom (1956) yakni ranah kognitif, afektif, dan psikomotor. Ranah kognitif meliputi penguasaan konsep-konsep Fisika dan saling keterkaitannya, dan penerapan metode ilmiah khususnya yang berkaitan dengan keterampilan proses intelektual. Ranah afektif berupa sikap ilmiah yang melandasi

metode ilmiah. Ranah psikomotor meliputi penerapan metode ilmiah khususnya yang berkaitan dengan keterampilan proses manual.

Jika evaluasi bertujuan untuk menentukan sejauhmana tujuan pendidikan yang tercantum dalam kurikulum telah tercapai, maka evaluasi dalam pendidikan IPA (Fisika) seharusnya mencakup tiga aspek di atas yakni proses, sikap, dan produk IPA. Hal ini diperkuat pula oleh dasar filosofis bahwa pada hakikatnya IPA berdimensi tiga yakni proses, sikap dan produk (Suriasumantri, 1984; Carin & Sund, 1989). Belajar IPA harus meliputi proses, sikap, dan produk, demikian pula evaluasi hasil belajarnya.

Di samping dasar filosofis dan tuntutan kurikulum yang berlaku seperti telah diuraikan, isu-isu yang berkembang dalam Pendidikan IPA seperti literasi Sains (*Science literacy*), Sains untuk Semua (*Science for All*), Sains-Teknologi-Masyarakat (*Science-Technology-Society*), Standar Pendidikan Sains Nasional (*National Science Education Standard*), dan Kurikulum Berbasis Kompetensi (*Competency-Based Curriculum*) juga menuntut agar evaluasi hasil belajar IPA tidak hanya menyangkut aspek produk saja, namun minimal menyangkut aspek produk, proses, dan sikap (Rutherford & Ahlgren, 1990; Bybee & DeBoer, 1994; Poedjiadi, 1996; Yager, 1996; NCSESA, 1996; Direktorat Dikmenum, 2002; Puskur, 2001).

Kenyataan yang ada di lapangan, baik evaluasi yang bersifat lokal maupun nasional seperti ulangan harian, ulangan umum, ulangan umum bersama (UUB), tes hasil belajar (THB), evaluasi belajar tahap akhir (EBTA), evaluasi belajar tahap akhir

nasional (EBTANAS), ujian masuk perguruan tinggi negeri (UMPTN)<sup>1</sup>, masih lebih menekankan pada segi kognitif (Rustaman, 1992; Subali, 1997; Sumaji, 1998; Budi, 1998; Wilardjo, 1998; Padri, 1999; Mardapi, 1999; Subali, 2000; Sukirman, 2000; Hakim, 2000), hanya berorientasi pada produk IPA, menggunakan alat evaluasi jenis *paper-and-pencil test* dengan bentuk pilihan jamak dan sedikit butir uraian.

Sebagai ilustrasi pada Tabel 1.1 disajikan hasil analisis terhadap soal EBTANAS th. 1999 dan soal UMPTN Rayon B th. 1999 sebagai studi pendahuluan dari disertasi ini. Analisis yang dilakukan meliputi tinjauan bentuk soal (pilihan jamak, uraian), kemampuan yang dievaluasi menurut klasifikasi Bloom (kognitif, afektif, psikomotor), dan kemampuan yang dievaluasi menurut aspek/ dimensi IPA (produk, proses, sikap).

TABEL 1.1  
HASIL ANALISIS SOAL IPA (FISIKA) PADA EBTANAS TH 1999  
DAN UMPTN RAYON B TH 1999

Jenis Tes	Bentuk Soal		Kemampuan yang Dievaluasi					
	Pilihan Jamak	Uraian	Taksonomi Bloom			Aspek IPA		
			Kognitif	Afektif	Psiko-motor	Proses	Sikap	Produk
EBTANAS TH 1999	100 %	0 %	100 %	0 %	0 %	27 %	0 %	73 %
UMPTN RAYON B TH 1999	100 %	0 %	100 %	0 %	0 %	47 %	0 %	53 %

<sup>1</sup> mulai tahun 2002 istilah diganti menjadi seleksi penerimaan mahasiswa baru (SPMB)

Tampak pada tabel, baik untuk EBTANAS maupun UMPTN bentuk soal semuanya pilihan jamak, tidak ada yang uraian. Kemampuan yang dievaluasi baik untuk EBTANAS maupun UMPTN hanya meliputi ranah kognitif, tidak menyentuh ranah afektif dan psikomotor, lebih menekankan produk dari pada proses IPA.

Kelemahan evaluasi hasil belajar yang bersifat nasional ini tidak hanya terjadi di Indonesia. Nitko (1996) dalam studinya di berbagai negara menemukan kritik-kritik terhadap evaluasi yang berlingkup nasional sebagai berikut : (1) hasilnya tidak peka terhadap program perbaikan, (2) laporan hasil evaluasi tidak menjelaskan pengetahuan dan keterampilan yang telah dipelajari siswa, (3) hasil evaluasi memberi dasar yang rapuh dalam membimbing karir siswa, (4) sering kurang sesuai dengan tujuan yang dinyatakan dalam kurikulum resmi, (5) beresiko tinggi dan mengabaikan kinerja siswa di kelas bertahun-tahun, (6) menimbulkan spekulasi guru untuk mengajar tidak berdasarkan kurikulum melainkan perkiraan soal yang muncul dalam evaluasi.

Di negara kita, tidak sejalannya evaluasi yang bersifat nasional seperti EBTANAS dan UMPTN dengan kurikulum yang berlaku ini menimbulkan dampak negatif terhadap proses pembelajaran siswa di kelas oleh guru. Pembelajaran cenderung berorientasi pada ranah kognitif aspek produk IPA melalui pendekatan informatif dengan metode ceramah (Subekti, 1997; Wilardjo, 1998; Budi, 1998; Corebima, 1999; Padri, 1999; Sukanto, 2000). Aspek keterampilan proses dan sikap IPA melalui pendekatan *inquiry* dengan metode eksperimen yang dituntut dalam kurikulum kurang mendapatkan perhatian yang memadai dalam pembelajaran. Lebih

spesifik lagi menurut Belen (2000), pembelajaran mengarah pada mempersiapkan siswa mengikuti ujian akhir (EBTANAS dan UMPTN). Langkah yang diambil guru dalam pembelajaran siswa ini logis dan wajar, mengingat saat ini yang dipandang sebagai barometer mutu suatu sekolah oleh masyarakat adalah NEM dan banyaknya siswa yang lulus seleksi UMPTN, bukan pada tercapainya tujuan-tujuan pendidikan yang tercantum dalam kurikulum. Dalam hal ini guru tidak dapat dipersalahkan, selama sistem evaluasi tersebut belum diubah sehingga relevan dengan kurikulum yang berlaku.

Resep-resep klasik untuk meningkatkan mutu pendidikan seperti (1) menambah anggaran pendidikan, (2) membenahi kurikulum, (3) meningkatkan pelatihan bagi guru, (4) melengkapi alat-alat laboratorium sekolah, (5) memprioritaskan alokasi dana dan tenaga pada daerah miskin dan terpencil, (6) menerapkan manajemen berbasis sekolah dan sebagainya, dapat saja dilakukan/ diulangi dan tidak ada yang membantah pentingnya, namun jika tidak dapat membenahi sistem evaluasi (lebih-lebih yang berlingkup nasional seperti EBTANAS) sebagai mesin pendorong peningkatan mutu proses belajar-mengajar sesuai kurikulum, maka resep-resep tersebut tidak akan manjur (Belen, 2000).

Idealnya harus ada kesesuaian/ sinergi antara kurikulum, proses belajar-mengajar, dan sistem evaluasi (Belen, 2000). Untuk mengetahui posisi negara kita di antara negara-negara lain, pada Tabel 1.2 disajikan hubungan antara pendekatan kurikulum, ujian akhir, dan proses belajar-mengajar di 13 negara.

TABEL 1.2  
HUBUNGAN ANTARA PENDEKATAN KURIKULUM, UJIAN AKHIR,  
DAN PROSES BELAJAR MENGAJAR (PBM) DI 13 NEGARA<sup>\*)</sup>

Negara	Pendekatan Kurikulum	PBM-Kurikulum		Ujian Akhir	
		Reduksi	Sinergi	Kognitif	<i>Basic Skills</i>
1. Malaysia	kombinasi	V		V	
2. Singapura	<i>outcome</i>		V		V
3. Filipina	kombinasi		V		V
4. Jepang	<i>content</i>		V		V
5. Korea Selatan	kombinasi	V		V	
6. India	<i>content</i> → <i>outcome</i>	V		V	
7. RRC	<i>content</i>	V		V	
8. Australia	<i>outcome</i>		V		V
9. New Zealand	<i>outcome</i>		V		V
10. Inggris <sup>**)</sup>	<i>outcome</i>		V		V
11. Belanda	<i>outcome</i>		V		V
12. Jerman <sup>***)</sup>	kombinasi		V		V
13. AS <sup>****)</sup>	kombinasi		V		V

<sup>\*)</sup>Sumber: Belen (2000: 5)

<sup>\*\*)</sup>England & Wales

<sup>\*\*\*)</sup>Bavaria

<sup>\*\*\*\*)</sup>New York State

Pada tabel tersebut pendekatan kurikulum dibagi tiga yakni *content*, *outcome*, dan kombinasi. Pendekatan *content* menekankan materi/ bahan ajar, pendekatan *outcome* menekankan kompetensi dasar, dan pendekatan kombinasi merupakan kombinasi antara keduanya. Hubungan PBM-kurikulum dibagi dua yakni reduksi (PBM merupakan reduksi dari tuntutan kurikulum) dan sinergi (PBM sejalan dengan tuntutan kurikulum). Ujian akhir dibagi dua yakni kognitif (menekankan aspek kognitif) dan *basic skills* (menekankan kompetensi dasar). Tampak pada tabel untuk negara-negara Barat pada umumnya menggunakan pendekatan kurikulum *outcome* atau kombinasi, hubungan PBM- kurikulum sinergi, ujian akhir mengevaluasi *basic skills*.

## B. Perumusan Masalah

Adanya kesenjangan antara pelaksanaan evaluasi hasil belajar di lapangan dengan tuntutan kurikulum yang berlaku, dasar filosofis, dan isu-isu dalam Pendidikan IPA seperti tersebut di muka mengisyaratkan perlunya dikembangkan model evaluasi yang sesuai dengan tuntutan-tuntutan tersebut. Karena tuntutan-tuntutan tersebut menggariskan pencapaian kompetensi keterampilan proses, sikap, dan pemahaman produk IPA untuk pemecahan masalah IPA, maka perlu dikembangkan suatu model evaluasi hasil belajar yang dapat mengevaluasi keterampilan proses, sikap dan produk IPA yang digunakan secara terpadu dalam pemecahan masalah IPA.

Melaksanakan evaluasi hasil belajar siswa yang meliputi tiga aspek IPA yakni proses, sikap dan produk secara terpadu tidaklah mudah, apalagi untuk yang berlingkup nasional seperti EBTANAS dan UMPTN. Banyak kendala-kendala yang harus dihadapi. Namun jika ini dapat dilaksanakan, maka dapat merupakan *shock therapy* terhadap proses pembelajaran yang ada, sehingga diperoleh mutu pendidikan IPA yang tinggi sesuai tuntutan masyarakat yang telah dituangkan dalam kurikulum.

Menurut Doran, *et al.* (1994: 28) beberapa negara maju seperti Inggris, Amerika Serikat, dan Jepang telah melaksanakan evaluasi hasil belajar IPA yang tidak hanya menyangkut produk IPA, tetapi juga proses IPA. Bahkan Inggris telah melaksanakannya secara nasional sejak tahun 1986. Namun negara-negara tersebut belum pernah melaksanakan evaluasi terpadu yang meliputi proses, sikap dan produk IPA.



Penelitian ini dimaksudkan untuk mengembangkan suatu model evaluasi hasil belajar IPA yang dapat digunakan untuk mengukur kemampuan/ keterampilan proses, sikap, dan produk IPA yang digunakan secara terpadu dalam kerja ilmiah/ pemecahan masalah ilmiah bidang IPA, dalam rangka mengatasi kelemahan atau keterbatasan model evaluasi yang ada, yang cenderung menekankan pada pengetahuan/ produk IPA saja. Model evaluasi yang dikembangkan ini disebut model evaluasi terpadu (model EVATER), sedangkan model evaluasi yang hanya menekankan pada pengetahuan/ produk IPA tersebut disebut model evaluasi produk (model EVAPROD).

Model EVATER ini berangkat dari pemikiran bahwa dalam kerja ilmiah atau pemecahan masalah ilmiah bidang IPA, pada umumnya terdapat dua pola. Pola pertama dimulai dengan adanya suatu masalah ilmiah, kemudian masalah ini dipecahkan secara rasional dan selanjutnya diuji secara empiris. Pemecahan secara rasional dapat menghasilkan hipotesis, paradigma, atau teori. Untuk dapat melakukan pemecahan secara rasional ini diperlukan keterampilan proses terutama keterampilan intelektual dan sosial, sikap ilmiah, serta produk IPA (konsep, prinsip, teori, hukum yang relevan). Pemecahan secara empiris dapat menghasilkan fakta (bukti empiris), prinsip, atau hukum. Untuk dapat melakukan pemecahan secara empiris diperlukan keterampilan proses terutama keterampilan manual dan sosial, sikap ilmiah, dan produk IPA (pengetahuan prosedural, pengetahuan tentang alat-alat). Selanjutnya hasil pemecahan secara rasional dan empiris ini dipertemukan (dibandingkan) sehingga akhirnya

diperoleh kesimpulan jawaban terhadap permasalahan tersebut yang secara tentatif merupakan produk IPA baru.

Pola kedua merupakan kebalikan pola pertama, yakni dari masalah yang ada mula-mula dipecahkan secara empiris kemudian dipecahkan (dianalisis) secara rasional, dan selanjutnya hasil kedua pemecahan tersebut dipertemukan (dibandingkan) untuk menghasilkan kesimpulan akhir pemecahan masalah tersebut yang dapat dipandang sebagai produk IPA baru.

Dari kedua pola tersebut dapat diambil makna bahwa di dalam IPA kriteria kebenaran adalah kesesuaian antara pemecahan secara rasional dan secara empiris, tidak pandang apakah proses pemecahan rasional yang lebih awal sementara proses pemecahan empiris yang lebih akhir atau sebaliknya.

Dalam model EVATER ini, yang dievaluasi adalah kemampuan/ keterampilan proses, sikap IPA (ilmiah), serta pemahaman produk IPA yang digunakan secara terpadu untuk pemecahan masalah secara rasional dan empiris tersebut, serta kesimpulan akhir pemecahan masalah (produk IPA baru) yang dihasilkannya.

Dalam usaha menghasilkan suatu model yang nantinya diharapkan dapat diterapkan secara luas, maka sebelum diterapkan secara luas terlebih dahulu model tersebut harus diuji efektivitas dan keterlaksanaannya. Demikian pula dalam usaha menghasilkan model sistem EVATER ini, masalah utama yang muncul sebelum diputuskan dapat dan tidaknya model ini untuk diterapkan secara luas, adalah apakah model ini efektif ditinjau dari tujuan dan dasar evaluasi, serta bagaimanakah

keterlaksanaan penerapan model ini ditinjau dari segi tenaga, waktu, peralatan, ruang, dan biaya.

Pada penelitian ini efektivitas model EVATER dibatasi terhadap tujuan dan dasar evaluasi untuk mengetahui pencapaian kemampuan oleh siswa serta untuk meningkatkan belajar siswa. Dengan demikian model EVATER ini dikatakan efektif jika valid dan reliabel untuk mengukur pencapaian kemampuan oleh siswa, serta dapat meningkatkan cara belajar IPA siswa. Namun karena cara belajar IPA berkaitan dengan pemahaman terhadap karakteristik IPA serta sikap terhadap IPA, maka kedua variabel terakhir ini juga dapat dijadikan indikator efektivitas.

Pelaksanaan EVATER ini rumit/ kompleks, banyak kendala-kendala yang harus diatasi antara lain peralatan, tenaga, waktu, ruang, dan biaya. Menurut *National Research Council* (1996 : 100), evaluasi harus menekankan pada hal-hal yang dipandang bernilai daripada hal-hal yang mudah dinilai. Sesuai dengan pandangan tersebut, walaupun EVATER tidak mudah dilaksanakan, namun perlu dilaksanakan karena EVATER menilai hal-hal yang dipandang bernilai yakni proses, sikap dan produk yang merupakan tujuan pembelajaran IPA. Selanjutnya jika ini sudah dicoba dilakukan, maka dapat diidentifikasi kesulitan-kesulitan, yang nantinya dapat dicari cara untuk mengatasinya.

Dari uraian-uraian di muka, maka secara lebih eksplisit masalah yang akan diselidiki dalam penelitian ini adalah bagaimana efektivitas dan keterlaksanaan model EVATER dalam mengukur kemampuan yang diharapkan dicapai siswa berdasarkan

GBPP kurikulum yang berlaku pada saat penelitian ini dilakukan, yakni GBPP kurikulum 1994. Masalah ini dapat dipecah menjadi:

1. Bagaimanakah validitas dan reliabilitas model EVATER dalam mengukur kemampuan yang diharapkan dicapai siswa berdasarkan GBPP kurikulum 1994 ?
2. Bagaimanakah efektivitas model EVATER dalam meningkatkan pemahaman karakteristik IPA, sikap terhadap IPA, serta cara belajar IPA siswa dibandingkan model EVAPROD?
3. Apakah kendala-kendala yang dihadapi dalam pelaksanaan model EVATER ini, dan bagaimanakah keterlaksanaannya ditinjau dari segi waktu, tenaga, peralatan, ruang dan biaya ?

#### C. Tujuan Penelitian

Sesuai dengan rumusan masalah yang telah dikemukakan di atas, maka tujuan penelitian ini adalah untuk :

1. Menyelidiki validitas dan reliabilitas model EVATER ini dalam mengukur kemampuan yang diharapkan dicapai siswa berdasarkan GBPP kurikulum 1994.
2. Menentukan efektivitas model EVATER dalam meningkatkan pemahaman terhadap karakteristik IPA, sikap terhadap IPA, serta cara belajar IPA siswa.
3. Mengidentifikasi kendala-kendala dalam pelaksanaan model EVATER, dan mengetahui keterlaksanaannya ditinjau dari segi waktu, tenaga, peralatan, ruang, dan biaya.

#### **D. Manfaat Penelitian**

Manfaat teoretis dari hasil penelitian ini adalah diperolehnya konsep/ teori/ prinsip evaluasi terpadu beserta karakteristik-karakteristiknya dalam pendidikan IPA.

Manfaat praktis hasil penelitian ini bagi lembaga pendidikan adalah diperolehnya bank soal yang telah diuji validitas dan reliabilitasnya, yang sewaktu-waktu dapat digunakan secara langsung atau diambil bagian-bagiannya untuk menyusun soal-soal tes/ evaluasi sesuai kebutuhan. Manfaat praktis bagi siswa adalah terlatihnya keterampilan memecahkan masalah ilmiah menggunakan proses, sikap, dan produk IPA secara terpadu, sedang bagi guru adalah diperolehnya pengalaman dalam merancang dan melaksanakan model evaluasi terpadu. Bagi pemegang otoritas pengambil kebijakan adalah diperolehnya bahan pertimbangan dalam meninjau kembali sistem evaluasi hasil belajar IPA yang telah dilaksanakan selama ini, dan diperolehnya model evaluasi alternatif yakni model evaluasi terpadu.

#### **E. Batasan Istilah**

Agar diperoleh kesamaan persepsi terhadap istilah-istilah yang digunakan dalam penelitian serta untuk memudahkan dalam pengukurannya, perlu dikemukakan beberapa batasan istilah/ definisi operasional dari istilah-istilah tersebut.

1. Model evaluasi terpadu (disingkat model EVATER) adalah model evaluasi hasil belajar IPA yang dirancang untuk mengevaluasi aspek-aspek kemampuan yang digunakan secara terpadu dalam kerja ilmiah/ pemecahan masalah ilmiah bidang

IPA. Kemampuan-kemampuan yang dimaksud adalah keterampilan proses IPA, sikap IPA (ilmiah), dan pemahaman produk IPA.

2. Proses IPA adalah proses-proses yang dilakukan dalam kerja ilmiah (pemecahan masalah ilmiah) bidang IPA meliputi proses merencanakan percobaan, mengajukan hipotesis, mengamati, mengklasifikasi, mengukur, mengontrol variabel, menghitung, mencari hubungan, menginterpretasi data, menyimpulkan, meramalkan, mengkomunikasikan, dan menerapkan.
3. Sikap IPA (ilmiah) adalah sikap-sikap yang melandasi kerja ilmiah/ pemecahan masalah ilmiah bidang IPA yang meliputi ingin tahu, teliti, obyektif, kritis, kreatif, dan terbuka.
4. Produk IPA adalah hasil yang diperoleh dari proses IPA yang dilandasi dengan sikap IPA (ilmiah) yang meliputi konsep, prinsip, hukum, teori, dan terminologi.
5. Model evaluasi produk (model EVAPROD) adalah model evaluasi hasil belajar IPA yang hanya menekankan pada aspek pemahaman produk IPA saja.

#### **F. Pembatasan Masalah**

Menurut Arikunto (1986 : 20) alat evaluasi dapat digolongkan menjadi dua jenis yakni tes dan bukan tes (non tes). Yang tergolong non tes adalah : skala bertingkat, kuesioner, daftar cocok, wawancara, pengamatan, riwayat hidup, laporan. Menurut Anastasi (1976) tes pada dasarnya merupakan suatu pengukuran yang obyektif dan standar terhadap sampel perilaku. Sedang menurut Brown (Azwar, 1996) tes adalah

prosedur yang sistematis guna mengukur sampel perilaku seseorang. Definisi yang lebih lengkap diberikan oleh Cronbach (1970) yaitu "... *a systematic procedure for observing a person's behavior and describing it with the aid of a numerical scale or category system*". Keuntungan penggunaan tes dibanding non tes adalah pelaksanaan tes lebih praktis, murah, dan cepat dari pada non tes.

Berdasarkan pelaksanaannya tes dapat digolongkan menjadi tes lisan (*verbal test*), tes tertulis (*written test, pencil-and-paper test*), tes perbuatan/unjuk kerja (*performance test*). Untuk memperoleh data yang akurat dan lengkap dapat saja dalam suatu evaluasi dilakukan kombinasi berbagai tes atau tes dan non tes.

Untuk mengevaluasi penguasaan produk dan sikap IPA dapat dilakukan dengan tes tertulis (*pencil-and-paper test*). Menurut beberapa hasil penelitian (APU, 1985, Lock, 1989; Bartley, *et al*, 1992; Doran & Tamir, 1992) untuk mengevaluasi keterampilan proses IPA dapat dilakukan dengan *pencil-and-paper test* dan tes praktikum (*performance test*). Namun karena kendala waktu, tenaga, biaya, dan peralatan maka *pencil-and-paper test* cenderung lebih dominan dilakukan. Atas dasar hasil penelitian ini dan untuk memudahkan organisasi/ pelaksanaan karena aspek yang dievaluasi meliputi proses, sikap, dan produk maka dalam penelitian ini tes yang dikembangkan dipilih (dibatasi) jenis *pencil-and-paper test* saja. Selanjutnya tes yang dikembangkan ini dirancang untuk tes hasil belajar (tes subsumatif) IPA/ Fisika bagi siswa SMU kelas satu caturwulan satu. Pemilihan kelas satu caturwulan satu ini

dengan pertimbangan materi pembelajaran pada tingkat ini merupakan materi dasar bagi tingkat-tingkat selanjutnya. Jika materi dasar ini tidak dikuasai dengan baik oleh siswa, maka siswa akan mengalami kesulitan pada tingkat-tingkat di atasnya. Selanjutnya agar siswa tidak hanya pandai memilih atau menebak melainkan pandai mengutarakan jawaban, tes tidak berbentuk pilihan jamak melainkan berbentuk uraian terbatas.

