

BAB III

METODE PENELITIAN

A. Fokus Penelitian

Seperti lazimnya tiap-tiap penelitian, arah penelitian ditentukan oleh tujuannya. Demikian pula dengan penelitian yang berjudul “Pengembangan Model Evaluasi Terpadu dalam Penilaian Hasil Belajar IPA” ini. Arah penelitian ditentukan oleh tiga tujuan yaitu :

1. Menyelidiki validitas dan reliabilitas model EVATER 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.

Dari ketiga tujuan tersebut, tujuan utama atau fokus penelitian diarahkan pada tujuan nomor dua yakni menentukan efektivitas model EVATER dalam meningkatkan pemahaman terhadap karakteristik IPA, sikap terhadap IPA, serta cara belajar IPA siswa. Tujuan nomor satu dan tiga dapat dipandang sebagai prasyarat dan kelanjutan dari pencapaian tujuan nomor dua, artinya untuk menyelidiki efektivitas model EVATER dalam meningkatkan pemahaman terhadap karakteristik IPA, sikap terhadap

IPA, dan cara belajar IPA diprasyarkan agar model tersebut memenuhi syarat validitas dan reliabilitas terlebih dahulu, dan dari pelaksanaan penelitian terhadap efektivitas model dapat dilanjutkan penyelidikan kendala-kendala pelaksanaan dan keterlaksanaan model tersebut.

B. Rancangan Penelitian

Penelitian ini menggunakan pendekatan *Research and Development (R & D)*, dengan tujuan umum untuk mengembangkan model evaluasi terpadu (model EVATER) seperti disebutkan di muka. Penelitian meliputi dua tahap pokok yakni tahap pengembangan dan validasi model, serta tahap pengujian efektivitas dan keterlaksanaan model.

Model yang dikembangkan dirancang dalam bentuk evaluasi (tes) subsumatif bagi siswa SMU kelas I cawu I. Prosedur pengembangan mengikuti prinsip dan langkah-langkah yang tertantum dalam Gambar 2.4. Dalam penelitian ini dikembangkan tiga perangkat tes subsumatif model EVATER, yakni Tes Subsumatif I, Tes Subsumatif II, dan Tes Subsumatif III. Validasi tes dilakukan melalui penilaian ahli (tiga ahli) dan ujicoba di lapangan yang melibatkan 30-45 siswa. Uraian tentang validasi dan ujicoba di lapangan ini dapat dilihat pada bagian III.E.5, III.F, dan IV.A.1.

Selanjutnya perangkat yang telah divalidasi dan memenuhi syarat validitas-reliabilitas diuji dalam eksperimen. Tujuan eksperimen adalah untuk menguji efektivitas model EVATER dalam meningkatkan pemahaman karakteristik IPA, sikap

terhadap IPA, dan cara belajar IPA siswa, di samping mengetahui kendala-kendala serta keterlaksanaan model. Sebagai pembandingan efektivitas model EVATER ini adalah model evaluasi yang konvensional yakni model evaluasi produk (model EVAPROD). Dengan dasar pemikiran metode pembelajaran mempengaruhi peningkatan pemahaman terhadap karakteristik IPA, sikap terhadap IPA, dan cara belajar IPA, maka dalam eksperimen terdapat dua faktor yakni faktor model evaluasi dan faktor metode pembelajaran, karena itu digunakan rancangan eksperimen faktorial seperti disarankan oleh beberapa ahli (Walpole & Myers, 1978; Gaspersz, 1991; Isaac & Michael, 1980; Sudjana, 1991). Dalam eksperimen ini ada dua model evaluasi yang diselidiki yakni model EVATER dan model EVAPROD, serta dua metode pembelajaran yakni metode eksperimen dan metode ceramah, sehingga rancangan eksperimennya adalah faktorial 2×2 seperti disajikan pada Tabel 3.1.

Dalam pelaksanaan eksperimen, mula-mula dilakukan pengukuran kemampuan awal yang meliputi pemahaman karakteristik IPA, sikap terhadap IPA, dan cara belajar IPA dari tiap kelompok eksperimen. Selanjutnya terhadap tiap kelompok eksperimen dilakukan pembelajaran materi I dengan metode pembelajaran dan model evaluasi subsumatif I yang sesuai dengan rancangan eksperimen yang telah ditetapkan. Dengan cara yang sama, langkah terakhir ini diteruskan untuk materi pembelajaran II dan III. Materi pembelajaran I adalah besaran-satuan dan kinematika gerak lurus, materi pembelajaran II adalah dinamika gerak lurus dan memadu gerak, sedangkan materi pembelajaran III adalah gerak melingkar beraturan dan gravitasi. Akhirnya kepada tiap-

tiap kelompok eksperimen dilakukan pengukuran kemampuan akhir, diteruskan wawancara kepada guru & siswa yang terlibat. Perangkat yang digunakan untuk pengukuran kemampuan akhir sama dengan untuk pengukuran kemampuan awal.

TABEL 3.1
RANCANGAN EKSPERIMEN YANG DIGUNAKAN

MODEL EVALUASI METODE PEMBELAJARAN	EVATER (1)	EVAPROD (2)
EKSPERIMEN (1)	$Y_{X.11.1}$ $Y_{X.11.2}$. $Y_{X.11.N}$	$Y_{X.21.1}$ $Y_{X.21.2}$. $Y_{X.21.N}$
KONVENSIONAL (CERAMAH) (2)	$Y_{X.12.1}$ $Y_{X.12.2}$. $Y_{X.12.N}$	$Y_{X.22.1}$ $Y_{X.22.2}$. $Y_{X.22.N}$

Keterangan :

1. $Y_{X.11.N}$ = peningkatan kemampuan X akibat faktor model evaluasi EVATER dan faktor metode pembelajaran eksperimen dari subyek ke N.
2. $Y_{X.21.N}$ = peningkatan kemampuan X akibat faktor model evaluasi EVAPROD dan faktor metode pembelajaran eksperimen dari subyek ke N.
3. $Y_{X.12.N}$ = peningkatan kemampuan X akibat faktor model evaluasi EVATER dan faktor metode pembelajaran konvensional (ceramah) dari subyek ke N.
4. $Y_{X.22.N}$ = peningkatan kemampuan X akibat faktor model evaluasi EVAPROD dan faktor metode pembelajaran konvensional (ceramah) dari subyek ke N.
5. Peningkatan kemampuan X = skor akhir kemampuan X - skor awal kemampuan X
6. Kemampuan X = pemahaman hakikat IPA, sikap terhadap IPA, atau cara belajar IPA

C. Teknik Pengambilan Sampel

Penelitian ini dilaksanakan pada cawu I tahun ajaran 1999/2000, tepatnya pada bulan Juli sampai dengan Oktober 1999. Populasi penelitian adalah siswa kelas I SMU Negeri Rayon I di wilayah Kotamadya Yogyakarta. Ada empat ciri yang melekat pada

populasi tersebut yakni siswa kelas I, SMU negeri, rayon I, dan wilayah Kotamadya Yogyakarta. Penentuan masing-masing ciri didasarkan atas beberapa pertimbangan.

Pertama, dipilih siswa kelas I karena kelas ini merupakan kelas dasar dari kelas-kelas di atasnya pada jenjang persekolahan (dalam hal ini SMU). Materi-materi dasar pada umumnya dibahas pada kelas ini, sehingga jika materi-materi ini tidak difahami pada umurnya siswa akan mengalami kesulitan pada kelas-kelas di atasnya. Di samping itu ada alasan teknis dipilihnya kelas I ini yakni karena kelas II disediakan untuk pelaksanaan PPL sedang kelas III menghadapi EBTANAS sehingga yang disediakan untuk penelitian adalah kelas I saja.

Kedua, dipilih SMU negeri karena dewasa ini semua SMU negeri (di wilayah Kotamadya Yogyakarta) sudah memiliki alat-alat dan laboratorium IPA. SMU swasta belum semuanya memiliki alat-alat dan laboratorium IPA sendiri, sehingga untuk melakukan eksperimen ada yang menggabung pada SMU-SMU negeri. Seperti telah dijelaskan, dalam penelitian ini melibatkan metode eksperimen yang tentunya memerlukan sarana dan prasarana alat-alat/ laboratorium IPA.

Ketiga, dipilih wilayah Kotamadya Yogyakarta karena kota ini merupakan tempat domisili peneliti sehingga dapat mengurangi kendala keterbatasan biaya, tenaga, dan waktu dari dan ke lokasi penelitian dibandingkan dengan jika dipilih wilayah lain. Dengan dikembangkannya otonomi daerah, pengembangan daerah termasuk bidang pendidikan mendapat perhatian tiap-tiap daerah. Karena itu pemilihan Kotamadya Yogyakarta sebagai tempat penelitian, di samping untuk mengatasi kendala-kendala

keterbatasan biaya, tenaga, dan waktu tersebut, juga dimaksudkan untuk ikut mengembangkan pendidikan di daerah domisili.

Dewasa ini di wilayah Kotamadya Yogyakarta terdapat 11 SMU negeri yang terbagi dalam 3 rayon. Ke 11 SMU negeri tersebut beserta peringkatnya menurut rata-rata jumlah NEM EBTANAS th 1998/1999 disajikan pada Tabel 3.2.

TABEL 3.2
RAYON DAN PERINGKAT NEM EBTANAS 1998/1999
SMU NEGERI DI KOTAMADYA YOGYAKARTA^{*)}

Rayon	SMU	Peringkat
I	SMU Negeri 3	2
	SMU Negeri 6	4
	SMU Negeri 11	6
	SMU negeri 9	10
II	SMU Negeri 1	1
	SMU Negeri 2	5
	SMU Negeri 4	8
	SMU Negeri 10	11
III	SMU Negeri 8	3
	SMU Negeri 7	7
	SMU Negeri 5	9

^{*)} Sumber : Proyek Perluasan dan Peningkatan SLTP DIY 1999/2000. Peringkat SLTP/SMU Negeri & Swasta Prop. DIY Berdasarkan Rata-Rata NEM th. 1998/1999. Yogyakarta : Kanwil Depdikbud.

Dengan menggunakan uji statistik non parametrik ranking Kruskal-Wallis melalui program komputer SPSS (lihat Lampiran 3-1) dan taraf signifikansi 0,05 atau 5 %, dapat disimpulkan bahwa ketiga rayon tersebut tidak berbeda signifikan peringkatnya ($\chi^2 = 0.1439$; $p > 0.05$). Uji beda ini dilakukan dengan statistik non parametrik karena data yang ada tidak memenuhi syarat untuk dianalisis dengan statistik parametrik

(jenis data hanya ordinal, jumlah sampel tiap kelompok hanya kecil yakni tiga/ empat SMU sehingga asumsi normalitas sebaran sulit dipenuhi). Diantara uji statistik non parametrik yang ada yakni uji Chi Kuadrat, uji median, dan uji Kruskal-Wallis, dipilih uji Kruskal-Wallis karena uji ini paling ketat dalam mempertahankan besaran data dibandingkan kedua uji yang lain (dalam analisisnya uji Chi Kuadrat mengubah data menjadi berskala nominal/ kategorik, uji median mengubah data menjadi tanda tambah atau kurang dari median, sementara uji Kruskal-Wallis memperlakukan data tetap sebagai ranking).

Keempat, atas dasar pertimbangan karakteristik rayon yang identik ditinjau dari peringkat NEM lulusan tersebut serta pertimbangan keterbatasan biaya, tenaga, dan waktu maka diputuskan pelaksanaan penelitian hanya pada satu rayon saja. Adapun penentuan Rayon I dilakukan secara random (acak). Dengan karakteristik ketiga rayon yang identik tersebut wilayah generalisasi tidak hanya meliputi siswa kelas I SMU negeri Rayon I di wilayah Kotamadya Yogyakarta, tetapi dapat diperluas meliputi seluruh siswa kelas I SMU negeri di wilayah Kotamadya Yogyakarta.

Penentuan besar sampel dilakukan dengan cara: Pertama, membagi SMU negeri rayon terpilih (Rayon I) ke dalam dua tingkat strata yakni peringkat atas dan peringkat bawah. Untuk SMU negeri peringkat 1 sampai dengan 5 dikategorikan peringkat atas (diberi kode 1) sedang untuk peringkat 6 sampai dengan 11 dikategorikan ke dalam SMU negeri peringkat bawah (diberi kode 2). Kedua, memeriksa jumlah subyek (siswa kelas I) pada masing-masing SMU negeri pada tiap-tiap peringkat seperti disajikan

pada Tabel 3.3. Menggunakan rumus $V_p = (1 - f)/n \sum W_h P_h Q_h$ dan $V_p = (d/t)^2$ (Cochran, 1991: 123) dengan f = fraksi sampel terhadap populasi; n = besar sampel; W_h = fraksi subyek dalam strata ke h terhadap total subyek dalam populasi; P_h = proporsi dalam populasi; $Q_h = 1 - P_h$; d = kesalahan estimasi; t = nilai z untuk α yang dipilih, maka dengan mengambil sampel sebesar 236 orang diperoleh peluang kesalahan estimasi $\geq 0,05$ akibat kesalahan dalam pengambilan sampel secara proporsional sebesar 0,05 atau 5 %.

TABEL 3.3
SEBARAN SISWA KELAS I SMU NEGERI RAYON I
KOTAMADYA YOGYAKARTA

SMU	Strata	Jumlah Siswa	
		Tiap SMU	Tiap Strata
SMU Negeri 3	Peringkat atas (1)	7 X 40	520
SMU Negeri 6		6 X 40	
SMU Negeri 9	Peringkat bawah (2)	5 X 40	360
SMU Negeri 11		4 X 40	
		Total	880

Sampel sebesar 236 orang tersebut untuk mewakili 880 orang siswa kelas I SMU negeri rayon I Kotamadya Yogyakarta. Namun karena 880 siswa tersebut terbagi dalam dua strata yakni siswa SMU negeri peringkat atas (520 orang) dan SMU negeri peringkat bawah (360 orang) maka agar proporsi anggota dalam tiap strata tercermin dalam proporsi anggota sampel, penentuan besar sampel dilakukan dengan perhitungan sebagai berikut.

Besar sampel yang mewakili SMUN peringkat tinggi = $520/880 \times 236 = 140$ orang

Besar sampel yang mewakili SMUN peringkat rendah = $360/880 \times 236 = 96$ orang

Idealnya sampel sebesar 140 orang tersebut harus tersebar di dua SMU Negeri peringkat atas yakni SMU Negeri 3 dan SMU Negeri 6, demikian pula sampel sebesar 96 orang tersebut harus tersebar di dua SMU negeri peringkat rendah yakni SMU Negeri 9 dan SMU Negeri 11. Tetapi untuk efisiensi biaya, tenaga, dan waktu, diambil praktisnya dengan menentukan secara random satu SMU negeri yang mewakili SMU negeri peringkat tinggi dan satu SMU negeri yang mewakili SMU negeri peringkat rendah. Setelah langkah ini dilakukan, diperoleh hasil SMU Negeri 6 yang mewakili SMU negeri peringkat tinggi dan SMU Negeri 9 yang mewakili SMU negeri peringkat rendah. Dengan demikian sampel sebesar 140 orang tersebut diambil secara random dari SMU Negeri 6 sedang sampel sebesar 96 orang tersebut diambil secara random dari SMU Negeri 9.

Selanjutnya sampel sebesar 140 orang dan 96 orang tersebut dikelompokkan dalam empat kelompok sesuai dengan rancangan eksperimen yang telah ditentukan di muka, sehingga tiap kelompok terdiri dari 59 orang. Agar terjadi kesepadanan antar kelompok-kelompok tersebut, maka sampel sebesar 140 orang tersebut didistribusikan ke dalam 4 kelompok tersebut, demikian pula untuk sampel yang sebesar 96 orang tersebut. Dalam prakteknya, dipilih secara random empat kelas dari 6 kelas I yang ada di SMU Negeri 6 untuk mengisi empat kelompok eksperimen yang telah disediakan, selanjutnya dari tiap tiap kelas dipilih 35 siswa secara random. Demikian pula

peneliti memilih secara random empat kelas dari lima kelas I yang ada di SMU Negeri 9 untuk mengisi empat kelompok eksperimen tersebut, dan memilih 24 siswa secara random dari tiap-tiap kelas terpilih tersebut. Dengan langkah pemasangan ini diharapkan diperoleh empat kelompok eksperimen yang seimbang sepadan dalam hal kemampuan yang mempengaruhi variabel terikat yakni pemahaman terhadap karakteristik IPA, sikap terhadap IPA, dan cara belajar IPA. Karena diduga NEM SLTP mempengaruhi variabel bergantung itu, maka NEM SLTP empat kelompok eksperimen tersebut harus seimbang atau jika tidak seimbang harus diperlakukan sebagai kovarian. Dengan asumsi NEM keempat kelompok tersebut berdistribusi normal, dilakukan uji anava dengan program komputer SPSS (lihat Lampiran 3-2), untuk mengetahui apakah ada perbedaan rata-rata NEM dari empat kelompok tersebut. Hasilnya diperoleh rata-rata NEM SLTP tiap kelompok adalah sebagai berikut : 45.82 ; 45.83; 45.72; 45.87. Pengujian selanjutnya menyatakan bahwa tidak ada perbedaan rata-rata NEM SLTP antara empat kelompok tersebut ($F = 0,117$; $p > 0,05$), demikian pula varian NEM SLTP keempat kelompok tersebut homogen (uji Levene $F = 1, 575$; $p > 0,05$). Jadi keempat kelompok tersebut seimbang dalam hal NEM SLTP nya.

D. Variabel Penelitian

Dari uraian yang telah dikemukakan di muka, maka dapat diidentifikasi variabel-variabel dalam penelitian ini.

1. Variabel Bebas

Dalam penelitian ini ada dua variabel bebas yakni model evaluasi (X1) dan metode pembelajaran (X2). Model evaluasi yang diselidiki ada dua yakni model EVATER dan model EVAPROD, sedang metode pembelajaran yang diselidiki juga ada dua yakni metode eksperimen dan ceramah/ konvensional. Dengan demikian kedua variabel bebas tersebut berskala diskrit (nominal).

2. Variabel Terikat

Dalam penelitian ini ada tiga variabel terikat yakni peningkatan pemahaman hakikat IPA (Y7), peningkatan sikap terhadap IPA (Y8), dan peningkatan cara belajar IPA (Y9). Karena butir-butir penyusun alat ukur ketiga variabel terikat ini cukup banyak (sekitar 30 butir), dan penggunaan alat ukur telah melalui ujicoba, maka ketiga variabel dapat dipandang mempunyai skala interval. Hal ini sesuai dengan pendapat Kerlinger (2000:706) yang menyatakan bahwa tes dan skala kecerdasan, prestasi, dan sikap dapat dipandang berskala interval.

3. Variabel Kontrol

Seperti telah diuraikan di muka, bahwa dalam eksperimen dilakukan penyepadanan sehingga nilai rata-rata NEM SLTP kelompok-kelompok eksperimen sama, demikian pula varians NEM SLTP kelompok-kelompok tersebut homogen. Dengan demikian dapat dikatakan bahwa NEM SLTP merupakan variabel kontrol (karena harga rata-ratanya “dibuat” sama untuk empat kelompok yang ada). Variabel

kontrol yang lain adalah materi pembelajaran, serta guru yang melaksanakan proses pembelajaran (masing-masing kelompok diajar oleh guru yang sama).

4. Variabel Sertaan/ Pengiring (Kovariabel)

Pada umumnya kemampuan awal, mempengaruhi variabel terikat dan tidak dapat dikendalikan, sehingga diperhitungkan dalam analisis sebagai variabel sertaan (kovariabel). Oleh karena itu dalam penelitian ini kemampuan awal yang ditunjukkan oleh skor pretes pemahaman terhadap karakteristik IPA (Y1), sikap terhadap IPA (Y2), dan cara belajar IPA (Y3) diklasifikasikan sebagai variabel sertaan (kovariabel).

E. Instrumen Penelitian

Dalam penelitian ini digunakan sembilan instrumen penelitian yang semuanya dibuat (dikembangkan) sendiri oleh peneliti. Tujuh di antara instrumen-instrumen tersebut berbentuk tes, satu berbentuk skala sikap dan satu lagi berbentuk *inventory*. Sebenarnya ketiganya hampir sama yakni merupakan instrumen pengukur, namun tes mempunyai jawaban benar/ salah atau gagal/ berhasil sementara skala dan *inventory* tidak mempunyainya (Mardapi, 1996; Kerlinger, 2000). Skala pada umumnya digunakan untuk pengukuran sikap sedang *inventory* lebih luas lagi. Berikut akan dibahas masing-masing instrumen secara lebih rinci meliputi konsep yang mendasari, kisi-kisi, serta validitas dan reliabilitasnya berdasarkan hasil ujicoba yang telah dilakukan.

1. Tes Pemahaman Karakteristik IPA (TPK IPA)

TPK IPA dimaksudkan untuk mengukur pemahaman karakteristik IPA oleh siswa, berupa tes tertulis bentuk pilihan jamak dengan empat alternatif jawaban. Tes dikembangkan dari dua konsep yakni konsep pemahaman dan konsep karakteristik IPA.

Seperti telah dikemukakan dalam Bab II, bahwa pemahaman merupakan kemampuan menangkap arti/makna yang terkandung dalam suatu bahan. Pemahaman mencakup tiga kemampuan yakni kemampuan translasi, interpretasi, dan ekstrapolasi. Kemampuan translasi merupakan kemampuan memindahkan (menterjemahkan) bahan dari bentuk konkrit ke bentuk abstrak, dari bentuk verbal ke bentuk simbol, dari bentuk verbal ke bentuk verbal lainnya dan sebagainya. Kemampuan interpretasi merupakan kemampuan menjelaskan atau merangkum suatu bahan, berupa penjelasan ke arah lebih dalam atau luas, penjelasan ke arah kualifikasi yang sesuai, atau rangkuman dari beberapa variasi. Kemampuan ekstrapolasi merupakan kemampuan memperkirakan kecenderungan atau meramalkan akibat. Dalam ekstrapolasi individu mampu mengadakan perluasan-perluasan untuk menentukan kesimpulan tentang hal-hal yang akan terjadi maupun yang telah terjadi.

Seperti telah dikemukakan juga dalam Bab II bahwa pada hakikatnya IPA merupakan produk, proses dan sikap. Pemahaman terhadap karakteristik IPA dapat ditinjau dari pemahaman karakteristik IPA sebagai produk, karakteristik IPA sebagai proses, dan karakteristik IPA sebagai sikap.

Berangkat dari konsep dan pemikiran di atas, dikembangkan kisi-kisi tes dan selanjutnya disusun butir-butir tes pemahaman karakteristik IPA tersebut. Tes berbentuk obyektif dengan maksud agar semua konsep yang terkandung dalam isi/materi serta semua kemampuan yang digariskan dalam kisi-kisi dapat diungkap (terwakili). Jumlah butir tes mula-mula (sebelum ujicoba) 35 butir. Kisi-kisi tes berikut nomor butirnya dapat dilihat pada Tabel 3.4, sedang butir-butir tesnya dapat dilihat pada TPK IPA yang disajikan pada Lampiran 3-3.

Ujicoba dimaksudkan untuk menyisihkan butir-butir yang tidak memenuhi syarat dan untuk menentukan keterandalan (reliabilitas) tes. Syarat yang dimaksud adalah indeks kemudahan butir dan validitas butir. Syarat daya pembeda butir tidak dimasukkan, dengan pertimbangan ada hubungan daya pembeda butir dengan indeks kemudahan butir dan validitas butir seperti diuraikan pada Bab II, sehingga informasi tentang indeks kemudahan butir dan validitas butir sudah dapat memberikan gambaran besarnya indeks daya pembeda butir.

Syarat indeks kemudahan butir dimaksudkan untuk menyisihkan butir-butir yang terlalu mudah dan terlalu sukar. Dalam penelitian ini indeks kemudahan butir dicari dengan teknik membagi jumlah siswa yang menjawab butir dengan benar (sesuai kunci jawaban) dengan jumlah siswa seluruhnya sehingga juga sering disebut proporsi benar. Penghitungan dilakukan dengan menggunakan program komputer *Iteman* dari *Assessment Systems Corporation* (1984).

TABEL 3.4
KISI-KISI TPK IPA

No.	Jenis Kemampuan Isi/ Materi	Translasi	Interpretasi	Ekstrapolasi
1.	Karakteristik IPA sebagai produk	1,2,3,4,6,8,10,11	5, 12, 13	7,9, 14
2.	Karakteristik IPA sebagai proses	15,20,22,24,26,28	16,17, 19,23,27	18,21,25
3.	Karakteristik IPA sebagai sikap	29,32,35	30,33	31,34

Validitas butir dicari dengan cara mengkorelasikan jawaban benar pada butir dengan jumlah jawaban benar pada kelompok (faktor) tempat butir tersebut ada, serta mengkorelasikan jawaban benar pada butir dengan total jawaban benar. Konsep yang mendasari kedua teknik ini adalah butir yang valid harus konsisten (selaras) dengan tes. Teknik yang pertama menggunakan korelasi *biserial* sedang teknik yang kedua menggunakan korelasi *point biserial*, keduanya juga dihitung dengan program komputer *Iteman*.

Reliabilitas dicari melalui pendekatan *single trial* dengan teknik Anava menggunakan rumus Alfa. Penghitungan juga dilakukan dengan program komputer *Iteman*.

Dengan menggunakan kriteria indeks kemudahan yang diterima antara 0.10-0.90 dan validitas butir yang diterima ≥ 0.20 maka dengan perputaran penghitungan dua kali, diperoleh 20 butir yang memenuhi syarat yakni butir no. 1, 2, 3, 4, 9, 10, 12, 14, 15, 17, 20, 21, 24, 25, 26, 29, 30, 32, 34, 35. Karena butir-butir ini masih memenuhi kisi-kisi yang dibuat, maka butir-butir ini digunakan sebagai TPK IPA yang akan digunakan

dalam penelitian. Dari hasil-hasil *print out* komputer dan perhitungan manual diperoleh karakteristik tes sebagai berikut : indeks kemudahan butir rata-rata 0.59, validitas butir rata-rata biserial 0.62 point biserial 0.45, dan koefisien reliabilitas alfa 0.30 (signifikan). *Print out* komputer analisis butir tes ini dapat dilihat pada Lampiran 3-4.

2. Skala Sikap terhadap IPA/ Fisika (SST IPA/F)

SST IPA/F dimaksudkan untuk mengukur sikap siswa terhadap IPA/ Fisika. Skala dikembangkan atas dasar definisi operasional sikap menurut Mar'at (1980) yakni sebagai derajat atau tingkat kesesuaian seseorang terhadap obyek tertentu. Kisi-kisi instrumen dikembangkan menurut dua dimensi sikap, yakni struktur dan target (sasaran/ obyek) sikap.

Dalam instrumen ini struktur sikap meliputi komponen kognitif, afektif, dan konatif (Mar'at, 1980 ; Zuchdi, 1995). Komponen kognitif berhubungan dengan persepsi, kepercayaan dan stereotipe yang dimiliki individu mengenai sesuatu. Persepsi dan kepercayaan mengenai obyek sikap berwujud pandangan (opini) dan seringkali merupakan stereotipe atau sesuatu yang telah terpolakan dalam pikirannya. Komponen afektif berkaitan dengan perasaan atau emosi. Reaksi emosional terhadap suatu obyek akan membentuk sikap positif atau negatif terhadap obyek tersebut. Komponen konatif adalah kecenderungan bertindak (berperilaku) berkaitan dengan obyek sikap.

Target sikap yakni obyek, kegiatan atau gagasan yang menjadi sasaran sikap dalam instrumen ini meliputi isi-materi IPA, kegiatan IPA, peralatan IPA, sumber informasi IPA, dan perkembangan IPA. Sesuai dengan pembatasan masalah pada Bab I, IPA yang dimaksud di sini adalah bagian (sub) Fisika saja.

Butir-butir instrumen dikembangkan dari instrumen yang dibuat oleh Nur (1982) yang digunakan untuk mengukur sikap terhadap IPA dari mahasiswa FPMIPA IKIP program S1 tingkat Sm1 sampai dengan Sm10, dimodifikasi berdasarkan kisi-kisi tersebut dan disesuaikan untuk siswa SMU. Skala yang digunakan adalah skala model Likert (SS, S, R, TS, STS). Kisi-kisi berikut nomor butir-butirnya dapat dilihat pada Tabel 3.5, sedang butir-butirnya disajikan pada SST IPA/F yang disajikan pada Lampiran 3-5.

TABEL 3.5
KISI-KISI SST IPA/F

No.	Obyek	Struktur	Kognitif	Afektif	Konatif
1.	Isi-materi IPA		3,4,9,10	1,2,5	6,7,8
2.	Kegiatan IPA		12,14,16	11,15,19	13,17,18,20
3.	Peralatan IPA		24,26	21,22	23,25
4.	Sumber Informasi IPA		28,31	27,32	29,30
5.	Perkembangan IPA		34,37	33,38	35,36

Jumlah butir mula-mula (sebelum ujicoba) adalah 38. Ujicoba dimaksudkan untuk menentukan validitas dan reliabilitas instrumen. Perhitungan validitas dan reliabilitas menggunakan program komputer *Iteman* seperti halnya pada TPK IPA,

demikian pula kriteria yang digunakan. Pada putaran pertama ada empat butir yang tidak valid yakni butir nomor 16, 20, 29 dan 31. Validitas butir rata-rata (korelasi butir-total) 0.56 dan reliabilitas Alfa rata-rata 0.61. Pada putaran kedua, butir-butir yang tidak valid disisihkan, meningkatkan validitas butir rata-rata menjadi 0.62 dan reliabilitas Alfa rata-rata menjadi 0.68. *Print out* komputer untuk analisis butir ini dapat dilihat pada Lampiran 3-6.

3. Inventory Cara Belajar IPA/ Fisika (ICB IPA/F)

ICB IPA/F dimaksudkan untuk mengetahui cara belajar IPA bagian Fisika, ditinjau dari segi tingkat/kadar keaktifan siswa. Seperti telah diketahui, tingkat keaktifan siswa dalam belajar sangat menentukan hasil belajar yang dicapai, sehingga para ahli berusaha untuk mengembangkan cara-cara belajar agar siswa aktif, yang dikenal dengan sebutan cara belajar siswa aktif (CBSA).

Instrumen berbentuk pilihan dengan dua alternatif jawaban "Ya/Tidak". Jumlah butir mula-mula (sebelum ujicoba) adalah 32. Kisi-kisi angket disajikan pada Tabel 3.6, dikembangkan dari prinsip-prinsip pengaktifan dalam belajar (Semiawan dkk, 1986:9-13) yang meliputi prinsip motivasi, latar/ konteks, keterarahan pada fokus tertentu, hubungan sosial/ sosialisasi, belajar sambil bekerja/ melakukan aktivitas, perbedaan perorangan/ individualisasi, prinsip penemuan, dan prinsip pemecahan masalah. Butir-butir instrumen dapat dilihat pada Lampiran 3.7.

TABEL 3.6
KISI KISI ICB IPA/F

No.	Isi/ Materi	Nomor Butir
1.	Motivasi	1,2,3
2.	Latar/ konteks	4,5,6
3.	Keterarahan/ fokus	7,8,9,10
4.	Hubungan sosial/sosialisasi	11,12,13,14,15
5.	Aktivitas	16,17,18
6.	Individualisasi	19,20,21,22
7.	Penemuan	23,24,25,26
8.	Pemecahan masalah	27,28,29,30,31,32

Ujicoba dilakukan untuk menyelidiki validitas dan reliabilitas instrumen. Ujicoba dilakukan terhadap 40 siswa. Hasil ujicoba dianalisis menggunakan program komputer *Iteman*. Hasil ujicoba putaran pertama menghasilkan 27 butir yang valid (homogen). Butir-butir yang tidak valid antara lain butir nomor 3, 9, 11, 22, dan 28. Pada putaran pertama ini dihasilkan validitas rata-rata *biserial* 0.86 dan *point biserial* 0.61, dan reliabilitas Alfa rata-rata 0.36. Pada putaran kedua butir-butir yang tidak valid disisihkan sehingga instrumen tinggal terdiri dari 27 butir. Setelah dianalisis, ternyata dapat meningkatkan indeks validitas rata-rata *biserial* menjadi 0.92 dan *point biserial* menjadi 0.66 dan reliabilitas alfa rata-rata menjadi 0.42. *Print out* komputer untuk analisis butir ini dapat dilihat pada Lampiran 3-8.

4. Instrumen EVAPROD

Instrumen dirancang dalam bentuk tes subsumatif, dimaksudkan untuk mengetahui penguasaan (pemahaman) siswa terhadap produk IPA yang berupa konsep, prinsip, hukum sesuai dengan kurikulum yang berlaku yakni kurikulum 1994. Label EVAPROD (kependekan dari kata evaluasi produk) menunjuk pada maksud tes yakni untuk mengevaluasi pemahaman terhadap produk IPA tersebut.

Seperti halnya pada tes pemahaman terhadap karakteristik IPA, tes ini mengukur kemampuan pemahaman yang meliputi kemampuan translasi, interpretasi, dan ekstrapolasi terhadap produk IPA (Fisika) yang tercantum dalam GBPP kurikulum 1994 untuk kelas I SMU pada cawu 1 yang meliputi enam pokok bahasan yakni: (1). besaran dan satuan, (2). kinematika gerak lurus, (3). dinamika gerak lurus, (4).memadu gerak, (5). gerak melingkar beraturan, (6). gravitasi.

Untuk mengungkap pemahaman konsep, prinsip, dan hukum pada keenam pokok bahasan tersebut dilakukan tes subsumatif. Tes subsumatif adalah tes yang dilakukan setelah sejumlah pokok bahasan diselesaikan (Depdikbud, 1991: 4). Tes subsumatif yang dilakukan sebanyak tiga kali yakni tes subsumatif I yang meliputi pokok bahasan (1) dan (2), tes subsumatif II yang meliputi pokok bahasan (3) dan (4), tes subsumatif III yang meliputi pokok bahasan (5) dan (6). Dengan pemberian tes tiga kali selama satu catur wulan ini diharapkan efeknya dalam eksperimen sudah dapat diketahui. Kisi-kisi ketiga tes subsumatif model EVAPROD.IPA ini dapat dilihat pada Tabel 3.7, dan butir-butir tesnya dapat dilihat pada Lampiran 3-9, 3-11, dan 3-13.

TABEL 3.7
KISI-KISI TES SUBSUMATIF MODEL EVAPROD

Tes	Isi/Materi	Jenis Kemampuan		
		Translasi	Interpretasi	Ekstrapolasi
Subsumatif I	(1). Besaran dan Satuan	6,7,8,9,12, 13,15,20, 25,26	1,2,3,4,5,11, 14,16,17,21,2 2,23,24,27,28	10,18,19,29
	(2). Kinematika Gerak Lurus	30,31,32, 35,39	33,34,37,38	36,40,41,42, 43,44,45
Subsumatif II	(3). Dinamika Gerak Lurus	1,3,4,6,10,1 2,14,15, 17,21,22,	2,5,13,16,18, 20	7,8,9,11,19, 23,24
	(4). Memadu Gerak	28,29,30, 37,39	25,26,27,36, 40	31,32,33,34, 35,38
Subsumatif III	(5). Gerak Melingkar Beraturan	1,2,3,4,5, 16	6,7,8,9,10, 17	11,12,13,14, 15,18
	(6). Gravitasi	19,21,24, 26,	20,22,23,25,2 7,28	29,30

Setelah tes diujicobakan dan dilakukan analisis butir dengan program komputer *Iteman*, didapatkan hasil-hasil seperti tersaji pada Tabel 3.8. *Print out* komputer analisis butir tes di atas dapat dilihat pada Lampiran 3-10 untuk tes subsumatif I, Lampiran 3-12 untuk tes subsumatif II, dan Lampiran 3-14 untuk tes subsumatif III. Karena setelah butir-butir yang memenuhi syarat masih memenuhi kisi-kisi yang

dibuat, maka butir-butir tersebut dapat digunakan sebagai tes subsumatif model EVAPROD.

TABEL 3.8
HASIL UJICOBA TES SUBSUMATIF MODEL EVAPROD

Karakteristik	Tes Subsumatif I	Tes Subsumatif II	Tes Subsumatif III
1. Jumlah Butir yang Memenuhi Syarat	30 (67 %)	29 (73 %)	20 (67 %)
2. Butir yang tidak Memenuhi Syarat	1, 2, 7, 11, 17, 23, 27, 32, 33, 38, 39, 40, 42, 44, 45	4, 6, 9, 12, 13, 14, 17, 19, 20, 21, 25, 30, 31, 33, 36, 39	7, 11, 17, 18, 20, 22, 26, 27, 28, 29
3. Indeks Kemudahan Butir Rata-Rata	0,58	0,44	0,53
4. Validitas butir Rata-Rata <i>Biserial</i>	0,62	0,63	0,65
5. Validitas Butir Rata-Rata <i>Poin Biserial</i>	0,42	0,44	0,50
6. Reliabilitas Alfa Rata-Rata	0,44	0,39	0,46

5. Instrumen EVATER

Instrumen dirancang dalam bentuk tes subsumatif, dimaksudkan untuk mengukur kemampuan/ keterampilan proses, sikap dan produk IPA yang digunakan secara terpadu dalam kerja ilmiah/ pemecahan masalah ilmiah bidang IPA sesuai dengan kurikulum yang berlaku yakni kurikulum 1994. Kemampuan/ keterampilan proses, sikap, dan produk IPA yang akan diukur terkandung dalam enam pokok bahasan yakni (1) besaran dan satuan, (2) kinematika gerak lurus, (3) dinamika gerak lurus, (4) memadu gerak, (5) gerak melingkar beraturan, (6) gravitasi. Untuk mengukur kemampuan-kemampuan ini dikembangkan tiga perangkat tes model EVATER yakni tes subsumatif I, II, dan III. Tes

subsumatif I dimaksudkan untuk mengungkap pencapaian dua tujuan pembelajaran umum (TPU) yakni : 1) siswa mampu melakukan pengukuran-pengukuran dan berdiskusi untuk memahami arti besaran dan satuan; 2) siswa mampu melakukan percobaan dan bernalar untuk memahami kinematika gerak lurus. Konsep (pokok bahasan) yang terkait dengan dua TPU tersebut adalah : 1) Besaran dan satuan; 2) Kinematika Gerak Lurus. Jumlah butir tes mula-mula (sebelum divalidasi) sebanyak 45 butir yang tersebar dalam tiga paket permasalahan yakni : (1) mengukur volume bola, (2) mengukur besaran vektor, dan (3) menyelidiki gerak lurus. Tes Subsumatif II dimaksudkan untuk mengungkap pencapaian dua TPU yakni : 1) siswa mampu melakukan percobaan dan bernalar untuk memahami dinamika gerak lurus; 2) siswa memahami pengertian memadu gerak dan persamaan gerak dalam bidang serta mengembangkan kemampuan berdiskusi dan bernalar. Konsep (pokok bahasan) yang terkait dengan dua TPU tersebut adalah : 1) Dinamika Gerak Lurus; 2) Memadu Gerak. Jumlah butir tes mula-mula (sebelum divalidasi) sebanyak 40 butir yang tersebar dalam tiga paket permasalahan yakni : (1) pengaruh gaya terhadap keadaan gerak benda, (2) dinamika dan kinematika kelereng pada bidang miring, (3) siapa penembak jitu? Tes Subsumatif III dimaksudkan untuk mengungkap pencapaian dua TPU yakni : 1) siswa memahami konsep dan persamaan gerak melingkar beraturan dan mampu melaksanakan percobaan serta mengembangkan kemampuan bernalar; 2) siswa memahami konsep gaya gravitasi dan hukum-hukum gravitasi Newton. Konsep (pokok bahasan) yang terkait dengan dua TPU tersebut adalah: 1) Gerak Melingkar Beraturan

TABEL 3.9
KISI-KISI TES SUBSUMATIF MODEL EVATER

Tes	Isi/Materi	Jenis Kemampuan/Keterampilan		
		Proses	Sikap	Produk
Subsumatif I	(1). Besaran dan Satuan	1,3,4,5,6,7, 8,9,10,11, 12,18,21, 22,23,26, 27,28	13,14,15,17, 19	2,16,20,23,24, 25
	(2). Kinematika Gerak Lurus	30,32,33, 34,35,36, 37,38,39, 40,41,45	29,31,35,43, 44	42
Subsumatif II	(3). Dinamika Gerak Lurus	1,2,3,4,5,6, 7,8,10,11, 12,13,14, 15,16,19, 21,22,23, 24,25,28, 29,30	9,17,18,26, 27	20
	(4). Memadu Gerak	33,35,36, 37,38,39	34,40	31,32
Subsumatif III	(5). Gerak Melingkar Beraturan	3,4,5,6,7,8, 12,13,14, 15,16,17, 19,20,21, 22,23	1,9,10,11,18	2
	(6). Gravitasi	24,26,28, 30,31,32, 34	27,29,33,35	25

dan 2) Gravitasi. Jumlah butir tes mula-mula (sebelum divalidasi) sebanyak 35 butir yang tersebar dalam tiga paket permasalahan yakni (1) Jarum Jam, (2) Gaya Sentripetal, (3) Satelit Buatan.

Kisi-kisi ketiga tes subsumatif ini dapat dilihat pada Tabel 3.9, sedang butir-butir tesnya sebelum divalidasi dapat dilihat pada Lampiran 3-15, 3-17, 3-19 masing-masing untuk tes subsumatif I, II, dan III. Kunci jawaban beserta kriteria penskorannya dapat dilihat pada Lampiran 3-16, 3-18, 3-20.

Karena penyelidikan validitas dan reliabilitas instrumen (tes) ini juga menjadi tujuan penelitian, agar tidak terjadi pembahasan ganda, uraian mengenai cara penentuan validitas dan reliabilitas tes ini dibahas dalam bagian III. F sedang hasil-hasil-hasilnya dilaporkan pada bagian IV.A.

F. Teknik Analisis Data

Sesuai dengan tujuan dan fokus yang telah ditentukan, ada tiga analisis pokok yang dilakukan yaitu analisis butir, analisis komparasi, dan analisis deskriptif kualitatif.

Analisis butir dilakukan untuk menyelidiki validitas dan reliabilitas model EVATER dalam mengukur kemampuan yang diharapkan dicapai siswa berdasarkan kurikulum/ GBPP 1994. Analisis butir yang dimaksud meliputi penentuan validitas isi, indeks kemudahan butir, validitas butir, dan reliabilitas tes. Penentuan validitas isi dilakukan melalui pertimbangan/ penilaian tiga ahli masing-masing ahli Pendidikan

Fisika, ahli Statistik dan Pengukuran Pendidikan yang berlatar belakang Eksakta/ Teknik, dan praktisi evaluasi Pendidikan Fisika yang berlatar belakang Pendidikan Fisika dan Evaluasi Pendidikan. Dalam pelaksanaannya, kepada masing-masing pakar tersebut disampaikan (1) surat permohonan validasi, (2) konstruk model EVATER, (3) kisi-kisi tes, (4) tes yang akan divalidasi, (5) lembar validasi (lihat Lampiran 3-21). Dalam lembar validasi ada dua aspek yang dimintakan pertimbangan/ penilaian ketiga ahli yakni aspek jenis perilaku/ kemampuan yang hendak diukur (aspek konstruk) dan aspek isi/ materi Fisika yang terkandung (aspek isi/ materi). Untuk menilai aspek konstruk, masing-masing pakar diminta memberikan nilai : 1 jika butir tes tidak cocok/ sesuai dengan jenis kemampuan yang diukur, 2 jika butir tes cocok/ sesuai dengan jenis kemampuan yang diukur, 3 jika butir tes sangat cocok dengan jenis kemampuan yang diukur. Kriteria suatu butir dinyatakan memenuhi aspek konstruk jika nilai rata-rata yang diperoleh ≥ 2 . Untuk menilai aspek isi/ materi, masing-masing pakar diminta memberikan nilai : 1 jika isi butir tes tidak termasuk dalam wilayah isi pokok bahasan/ pokok bahasan yang diungkap, 2 jika isi butir termasuk dalam wilayah isi pokok bahasan/ pokok bahasan yang diungkap. Kriteria suatu butir dinyatakan memenuhi aspek isi/ materi jika jumlah/ cacah nilai $2 \geq 2$. Kriteria suatu butir dinyatakan memenuhi validitas isi jika kedua aspek (konstruk dan isi/ materi) valid.

Indeks kemudahan butir dihitung dengan membagi jumlah skor butir yang diperoleh seluruh siswa dengan jumlah skor butir menurut kunci dari seluruh siswa tersebut. Sesuai dengan fungsi tes yakni sebagai tes subsumatif, kriteria indeks

kemudahan butir yang diterima mempunyai rentang yang luas (Bloom, Hasting & Madame, 1971; Arikunto, 1986), dalam penelitian ini digunakan rentang antara 0,10-0,90. Validitas butir (empiris) dihitung dengan mengkorelasikan skor butir dengan skor total pada kelompok (faktor) tempat butir tersebut ada dengan korelasi Pearson. Konsep yang mendasari teknik ini adalah butir yang valid harus konsisten (selaras) dengan tes. Validitas butir (korelasi skor butir dan skor total faktor) yang diterima $\geq 0,20$ (Crocker & Algina, 1986; Azwar, 2000). Reliabilitas tes ditentukan melalui pendekatan *single trial* dengan teknik anava menggunakan rumus Alfa. Menurut Azwar (2000), tidak ada batasan mutlak yang menunjukkan angka koefisien reliabilitas terendah yang harus dicapai agar suatu pengukuran dapat disebut reliabel, karena itu dalam penelitian ini digunakan kriteria tabel r (Hadi, 1991). Untuk jumlah subyek 30 orang, diperoleh r tabel 0,24 (pada taraf kesalahan 5 %), sehingga koefisien reliabilitas $\geq 0,24$ dapat dinyatakan signifikan (reliabel). Dalam teknis perhitungannya penentuan indeks kemudahan butir, validitas butir (empiris), dan reliabilitas tes dilakukan dengan program komputer *Iteman* dari *Assessment System Corporation* (1986). Khusus untuk penentuan indeks kemudahan, karena program tidak menyediakan perhitungan untuk butir-butir non dikotomi, maka ditambahkan perhitungan secara manual.

Analisis komparasi dilakukan untuk menentukan efektivitas model EVATER dalam meningkatkan pemahaman karakteristik IPA, sikap terhadap IPA, dan cara belajar IPA. Karena dalam penelitian ini terdapat empat kelompok eksperimen yakni :

(1). kelompok yang diberi evaluasi model EVATER dan metode pembelajaran eksperimen; (2) kelompok yang diberi evaluasi model EVAPROD dan metode pembelajaran eksperimen; (3) kelompok yang diberi evaluasi model EVATER dan metode pembelajaran ceramah; (4) kelompok yang diberi evaluasi model EVAPROD dan metode pengajaran ceramah, maka untuk itu dilakukan komparasi terhadap empat kelompok ini. Sedang yang dikomparasikan adalah : (1) peningkatan pemahaman karakteristik IPA; (2) peningkatan sikap terhadap IPA; (3) peningkatan cara belajar IPA. Dalam hal ini yang dimaksud dengan peningkatan adalah selisih antara pengukuran kemampuan akhir (setelah mengalami perlakuan eksperimen) dan pengukuran kemampuan awal (sebelum mengalami perlakuan eksperimen). Seperti telah dijelaskan pada bagian ILLD, bahwa empat kelompok tersebut berpasangan (rata-rata dan varians NEM SLTP sama). Karena itu untuk komparasi terhadap empat kelompok ini dilakukan dengan anava/ anakova dan uji t sampel berpasangan, dalam hal dipenuhi syarat distribusi skor empat kelompok tersebut normal, keempat kelompok mempunyai varians skor yang homogen, skor minimal berderajat interval, dan empat kelompok tersebut dipilih secara random (lihat Hadi, 1982: 388; Nurgiyanto, 2000: 192). Namun dalam hal syarat-syarat tersebut tidak dipenuhi, digunakan analisis non parametrik, menggunakan *Friedman Two-Way Anova* dan *Wilcoxon Matched-Pairs Test*. Sehubungan dengan itu sebelum analisis dilakukan terlebih dahulu dilakukan pengujian persyaratan analisis untuk menentukan teknik mana yang sesuai. Ada dua pengujian persyaratan analisis yang dilakukan yakni uji normalitas Kolmogorov-Smirnov dan uji

homogenitas varians Levene. Semua uji persyaratan analisis dan analisis komparasi ini dilakukan menggunakan program komputer *SPSS (under windows)*. Dalam penelitian ini semua pengujian dan teknik analisis menggunakan tingkat kesalahan 5 % atau taraf kepercayaan 95 %.

Analisis deskriptif kualitatif digunakan untuk mengidentifikasi kendala-kendala pelaksanaan model EVATER, serta kemungkinan keterlaksanaannya, berdasarkan pengalaman peneliti, wawancara dengan guru dan siswa yang terlibat dalam penelitian.

