

## **BAB V PEMBAHASAN**

### **A. Pembahasan Tentang Proses Pengembangan Perangkat Evaluasi dengan Memperhatikan Aspek Kognitif, Afektif, dan Psikomotor Siswa**

Rangkaian proses pengembangan perangkat evaluasi pembelajaran matematika dengan memperhatikan aspek kognitif, afektif, dan psikomotor mulai dilakukan tanggal 07 Mei 2011 s/d 25 Juli 2011. Proses pengembangan perangkat evaluasi ini diawali dengan pengembangan spesifikasi tes. Pengembangan spesifikasi tes merupakan langkah paling awal yang paling menentukan dalam pengembangan suatu perangkat tes, karena apa yang dilakukan pada langkah-langkah berikutnya sudah direncanakan dalam spesifikasi tes tersebut. Tabel spesifikasi tes merupakan tabel yang memuat sekaligus uraian isi tes dan tingkat kompetensi yang akan diungkap pada setiap bagian isi. Tabel semacam ini seringkali disebut sebagai *test blue print*. Disamping uraian mengenai materi dan batasan perilaku, dalam spesifikasi tes memuat juga informasi mengenai tipe soal, taraf kesukaran soal, banyaknya item soal, lamanya penyajian tes dan cara pemberian skor. Oleh karena itu spesifikasi tes akan menjadi pegangan yang sangat membantu peneliti sewaktu penulisan item berlangsung sebagai pedoman yang akan menjaga agar item soal yang ditulis tetap terarah pada tujuan pengukuran tes dan tidak keluar dari batasan isi.

Berdasarkan atas spesifikasi tes yang telah disusun, peneliti kemudian menulis butir soal. Pada langkah ini, peneliti dituntut memiliki pengetahuan teoritis mengenai teknik penulisan soal agar dihasilkan soal-soal yang baik. Soal-soal yang telah ditulis dengan hati-hati berdasarkan pada pertimbangan tidak begitu saja dianggap soal-soal yang sudah baik. Soal-soal yang ditulis perlu ditelaah oleh suatu panel ahli. Penilaian para ahli meliputi penilaian tentang kevalidan dan kepraktisan perangkat evaluasi oleh para validator. Penilaian kevalidan perangkat didasarkan pada tiga aspek yaitu a) Segi materi yang diuji; b) Segi format dan pertimbangan teknis penulisan soal; c) Segi penerjemahan gagasan kedalam bahasa. Sedangkan penilaian kepraktisan perangkat didasarkan pada tingkat kemudahan dan kepraktisan penggunaan dan pelaksanaan suatu tes dalam hubungannya dengan biaya dan waktu untuk melaksanakan tes, serta pengolahan dan penafsiran hasilnya.

Proses validasi dilaksanakan selama 7 hari oleh para validator yang dipilih, yakni sebagai berikut:

**tabel 5.1**  
**Daftar Nama Validator Perangkat Evaluasi**

No	Nama Validator	Keterangan
1	Drs.Abdullah Sani, M. Pd	Dosen Pendidikan Matematika IAIN Sunan Ampel Surabaya
2	Agus Prasetyo K., M.Pd	Dosen Pendidikan Matematika IAIN Sunan Ampel Surabaya
3	Musriani, S. Pd	Guru matematika MTs Tribakti Kunjang Kediri

Peneliti memilih 2 dosen matematika dari IAIN Sunan Ampel dan 1 guru matematika yang mengajar di MTs Tribakti Kunjang Kediri sebagai validator. Hal ini dilakukan karena seorang guru matematika tentunya mengetahui sedikit banyak kondisi dan permasalahan dalam proses pembelajaran matematika khususnya dalam hal evaluasi pembelajaran matematika. Karena alasan ini peneliti memilih seorang guru matematika yang mengajar di MTs Tribakti, yang peneliti anggap berkompeten dalam bidang matematika untuk menjadi validator dalam penelitian ini.

Setelah proses validasi selesai dilakukan, perangkat pembelajaran direvisi sesuai dengan saran dan masukan dari dosen pembimbing dan para validator. Soal-soal yang telah lulus dari pengujian melalui penelaahan soal secara teori sudah baik. Langkah selanjutnya perlu dilakukan uji-coba instrumen secara empiris. Uji-coba instrumen ini dilakukan di MTs Tribhakti Kunjang Kediri selama 5 hari, mulai tanggal 18 s/d 22 Juli 2011. Satu hari sebelum pelaksanaan uji-coba instrumen peneliti melakukan kegiatan *review* materi yang ditekankan pada penilaian kognitif. Kerena ada beberapa materi yang ditekankan merupakan materi di semester ganjil, anggapan peneliti banyak siswa yang sudah lupa mengenai materi tersebut. Dalam proses uji-coba penilaian afektif dan psikomotor, peneliti bertindak sebagai pengamat dan dibantu oleh teman peneliti yaitu Liana Nur Fadlilah sebagai pengajar di kelas. Ada beberapa kendala dalam proses uji-coba

diantaranya 1) Materi yang diujikan pada penilaian kognitif sangat banyak, sehingga *testee* banyak yang mengalami kesulitan dan kebingungan dalam menjawab soal; 2) waktu uji-coba tes kurang efektif, karena jarak antara penerimaan materi dengan uji-coba tes terlampau jauh sehingga *testee* banyak yang sudah lupa mengenai materi tersebut; 3) kurang ada persiapan secara matang dari *testee*, sehingga hasil yang diperoleh kurang optimal.

Dari Hasil uji-coba ini akan diperoleh data empirik, yaitu skor sekelompok subyek yang dikenai tes tersebut. Data empirik tersebut akan digunakan untuk menguji validitas dan reliabilitas eksternal instrumen yang dikembangkan. Untuk mengetahui validitas instrumen digunakan teknik analisis korelasi antara skor item dengan skor total. Dalam teknik korelasi ini digunakan formula korelasi produk moment dengan angka kasar yang dikemukakan oleh Pearson. Karena instrumen yang disusun memiliki jumlah item yang sedikit (kurang dari 30 item), maka harus dilakukan koreksi terhadap angka korelasi yang diperoleh karena sedikitnya jumlah item yang ada dalam tes akan mengakibatkan terjadi *overestimasi* terhadap korelasi yang sebenarnya. *Overestimasi* ini disebabkan terlalu besarnya kontribusi masing-masing item dalam ikut menentukan skor tes. Oleh karena itu, skor setiap item menjadi bagian atau porsi dari skor tes.

Dalam pengambilan keputusan untuk menentukan item yang valid digunakan  $r_{hitung}$  dibandingkan  $r_{tabel}$  dengan dk jumlah sampel dikurangi variabel.

Jika  $r_{hitung} \geq r_{tabel}$  maka item dikatakan valid, akan tetapi jika  $r_{hitung} < r_{tabel}$  maka item tersebut disimpulkan tidak valid.

Karena instrumen yang dibuat dalam penelitian ini berbentuk soal uraian (penilaian kognitif) dan instrumen yang skornya berbentuk rentangan nilai (penilaian afektif dan psikomotor) untuk mencari reliabilitas instrumen digunakan rumus Alfa Cronbach.

Dalam pengambilan keputusan untuk menentukan reliabilitas tes digunakan  $r_{hitung}$  dibandingkan  $r_{tabel}$  dengan dk jumlah sampel dikurangi variabel. Jika  $r_{hitung} \geq r_{tabel}$  maka item dikatakan reliabel, akan tetapi jika  $r_{hitung} < r_{tabel}$  maka item tersebut disimpulkan tidak reliabel.

#### **B. Pembahasan Tentang Kevalidan Hasil Pengembangan Perangkat Evaluasi dengan Memperhatikan Aspek Kognitif, Afektif, dan Psikomotor Siswa**

Berdasarkan perhitungan uji validitas yang telah dilakukan pada Bab IV, diperoleh koefisien validitas dari uji-coba penilaian kognitif pada 32 siswa sebagai berikut:

**Tabel 5.2**  
**Koefisien Validitas Penilaian Kognitif**

<b>Nomor Soal</b>	<b><math>r_{hitung}</math></b>	<b><math>r_{tabel}</math></b>	<b>Keterangan</b>
1	0,362	0,361	VALID
2	-0,66	0,361	INVALID
3	0,26	0,361	INVALID
4	0,42	0,361	VALID
5	0,4	0,361	VALID
6	-0,504	0,361	INVALID
7	-0,92	0,361	INVALID
8	-0,14	0,361	INVALID

9	0,19	0,361	INVALID
10	0,04	0,361	INVALID

Dari tabel diatas, semakin tinggi korelasi positif antara skor item dengan skor total berarti semakin tinggi konsistensi antara item tersebut dengan item tes keseluruhan yang berarti semakin tinggi daya bedanya. Bila koefisien korelasinya rendah mendekati nol berarti fungsi item tersebut tidak cocok dengan fungsi ukur tes dan daya bedanya tidak baik. Bila koefisien korelasi berharga negatif, artinya terdapat cacat serius pada item yang bersangkutan.

Berdasarkan  $r_{\text{tabel}}$  untuk dk 30 dan taraf nyata ( $\alpha$ ): 0,05, didapatkan skornya  $r_{\text{tabel}:0,05;30} = 0,361$ . Dengan mencocokkan harga koefisien korelasi diatas pada tabel r produk moment, dari 10 item soal uraian yang diuji validitasnya, 3 item diantaranya telah dinyatakan valid yaitu nomor 1, 4, dan 5. Sedangkan 7 item lainnya yaitu nomor 2, 3, 6, 7, 8, 9, dan 10 dinyatakan tidak valid.

Banyaknya item soal pada penilaian kognitif yang dinyatakan tidak valid, menurut penulis disebabkan oleh beberapa hal diantaranya 1) Materi yang diujikan sangat banyak, sehingga *testee* banyak yang mengalami kesulitan dalam menjawab soal; 2) waktu uji-coba tes kurang efektif, karena jarak antara penerimaan materi dengan uji-coba tes terlampau jauh sehingga banyak materi yang telah terlupakan ; 3) kurang ada persiapan secara matang dari *testee*, sehingga hasil yang diperoleh kurang optimal.

Berdasarkan hasil perhitungan koefisien validitas pada Bab IV, diperoleh koefisien validitas dari uji-coba instrumen penilaian afektif pada 30 siswa sebagai berikut :

**Tabel 5.3**  
**Koefisien Validitas Penilaian Afektif**

Nomor Soal	$r_{hitung}$	$r_{tabel}$	Keterangan
1	-0,34	0,374	INVALID
2	-0,56	0,374	INVALID
3	-0,33	0,374	INVALID
4	0,758	0,374	VALID
5	-0,64	0,374	INVALID
6	0,219	0,374	INVALID
7	-1,672	0,374	INVALID

Dari tabel diatas, semakin tinggi korelasi positif antara skor item dengan skor total berarti semakin tinggi konsistensi antara item tersebut dengan item tes keseluruhan yang berarti semakin tinggi daya bedanya. Bila koefisien korelasinya rendah mendekati nol berarti fungsi item tersebut tidak cocok dengan fungsi tes dan daya bedanya tidak baik. Bila koefisien korelasi berharga negatif, artinya terdapat cacat serius pada item yang bersangkutan.

Berdasarkan  $r_{tabel}$  untuk dk 28 dan taraf nyata ( $\alpha$ ): 0,05, didapatkan skornya  $r_{tabel:0,05;28} = 0,374$ . Dengan mencocokkan harga koefisien korelasi diatas pada tabel r produk moment, dari tujuh aspek yang diukur pada penilaian afektif hanya ada satu aspek yaitu aspek nomor 4 yang dinyatakan valid. Sedangkan enam aspek lainnya yaitu aspek nomor 1, 2, 3, 5, 6, dan 7 dinyatakan tidak valid. Karena instrumen yang digunakan untuk mengukur penilaian afektif

menggunakan lembar pengamatan dengan teknik observasi sehingga banyak kendala yang dialami oleh pengamat dalam pengumpulan data, kendala tersebut antara lain 1) Data yang diperoleh dari hasil observasi tidak dapat memberikan gambaran (*insight*) yang sama tentang struktur kepribadian individu. Untuk itu seharusnya perlu dilakukan data yang diperoleh dengan teknik lain seperti wawancara atau penyebaran angket; 2) unsur subjektivitas dari pengamat sulit untuk dipisahkan. Sehingga mengurangi objektivitas analisis hasil atau data observasi; 3) untuk mengamati struktur kepribadian individu tidak cukup dilakukan hanya beberapa kali pertemuan, perlu pengamatan secara berkala sehingga objektivitas hasil observasi lebih terjamin. Dari beberapa kendala dalam pengamatan tersebut sedikit banyak mempengaruhi hasil pengamatan. Sehingga enam aspek dari tujuh aspek yang dinilai dalam penilaian afektif dinyatakan tidak valid.

Berdasarkan hasil perhitungan koefisien validitas pada Bab IV, diperoleh koefisien validitas dari uji-coba instrumen penilaian psikomotor pada 30 siswa sebagai berikut:

**Tabel 5.4**  
**Koefisien Validitas Penilaian Psikomotor**

Nomor Soal	$r_{hitung}$	$r_{tabel}$	Keterangan
1	0,45	0,374	VALID
2	-0,157	0,374	INVALID
3	0,76	0,374	VALID
4	-0,21	0,374	INVALID
5	0,649	0,374	VALID



Dari tabel diatas, semakin tinggi korelasi positif antara skor item dengan skor total berarti semakin tinggi konsistensi antara item tersebut dengan item tes keseluruhan yang berarti semakin tinggi daya bedanya. Bila koefisien korelasinya rendah mendekati nol berarti fungsi item tersebut tidak cocok dengan fungsi ukur tes dan daya bedanya tidak baik. Bila koefisien korelasi berharga negatif, artinya terdapat cacat serius pada item yang bersangkutan.

Berdasarkan  $r_{\text{tabel}}$  untuk dk 28 dan taraf nyata ( $\alpha$ ): 0,05, didapatkan skornya  $r_{\text{tabel}:0,05;28} = 0,374$ . Dengan mencocokkan harga koefisien korelasi diatas pada tabel r produk moment. Dari lima aspek yang diukur pada penilaian psikomotor ada tiga aspek yaitu aspek nomor 1, 3, dan 5 dinyatakan valid. Sedangkan dua aspek lainnya yaitu aspek nomor 2 dan 4 dinyatakan tidak valid.

### **C. Pembahasan Tentang Kerelibilitas Hasil Pengembangan Perangkat Evaluasi Pembelajaran Matematika dengan Memperhatikan Aspek Kognitif, Afektif, dan Psikomotor Siswa**

Berdasarkan hasil perhitungan estimasi reliabilitas melalui pendekatan konsistensi internal dengan menggunakan teknik formula Alfa Cronbach, angka koefisien reliabilitas yang diperoleh untuk penilaian kognitif masuk pada kategori sedang, karena koefisien reliabilitas 0,6. Dan koefisien reliabilitas penilaian psikomotor masuk pada kategori tinggi, karena koefisien reliabilitas 0,875. Dan

koefisien reliabilitas penilaian afektif masuk pada kategori tidak reliabel, karena koefisien reliabilitas  $-0,875$ .

Suatu tes yang menghasilkan koefisien reliabilitas  $0,6$ , berarti perbedaan (variasi) yang tampak pada skor tersebut mampu mencerminkan  $60\%$  dari variasi yang terjadi pada skor murni subjek yang bersangkutan. Dapat pula dikatakan  $40\%$  dari perbedaan skor yang tampak disebabkan oleh variasi eror pengukuran tersebut. Sedangkan suatu tes yang menghasilkan koefisien reliabilitas  $0,875$ , berarti perbedaan (variasi) yang tampak pada skor tersebut mampu mencerminkan  $87,5\%$  dari variasi yang terjadi pada skor murni subjek yang bersangkutan. Dapat pula dikatakan  $12,5\%$  dari perbedaan skor yang tampak disebabkan oleh variasi eror pengukuran tersebut. Dan bila suatu tes menghasilkan koefisien reliabilitas  $< 0$ , berarti pengukuran yang dilakukan oleh tes yang bersangkutan terkandung sejumlah eror. Besar kecilnya eror dicerminkan oleh seberapa jauh jarak angka koefisien reliabilitas yang dihasilkan tes tersebut dari angka  $1$ . Sehingga tes yang menghasilkan koefisien reliabilitas  $-0,875$  berarti eror yang dihasilkan oleh tes tersebut sebesar  $1,875$ . Dari interpretasi tersebut dapat dikatakan bahwa besarnya varians eror akan mempengaruhi tinggi rendahnya reliabilitas. Bila varians eror sangat kecil maka tes akan menghasilkan reliabilitas yang tinggi.

Dalam penelitian ini, ada beberapa faktor yang mempengaruhi besar kecilnya koefisien reliabilitas tes diantaranya: 1) jumlah item soal yang kecil pada penilaian kognitif, afektif dan psikomotor dapat mempengaruhi tinggi-rendahnya

validitas tes dan tinggi rendahnya validitas dapat mempengaruhi reliabilitas tes; 2) instrumen yang dibuat untuk mengukur gejala non fisik seperti kemampuan, kecakapan, sikap dan lain sebagainya sering berubah-ubah dari waktu ke waktu. Dan dalam pengukuran gejala sosial boleh dikatakan tidak pernah terjadi hasil pengukuran ulang yang persis sama dengan hasil pengukuran sebelumnya sehingga hasil pengukuran koefisien reliabilitas penilaian afektif pada penelitian ini menghasilkan koefisien reabilitas  $< 0$

#### **D. Pembahasan Tentang Kepraktisan Pengembangan Perangkat Evaluasi Pembelajaran Matematika dengan Memperhatikan Aspek Kognitif, Afektif, dan Psikomotor Siswa**

Penilaian kepraktisan perangkat yang dikembangkan pada penelitian ini diperoleh dari hasil penilaian validator yang didasarkan pada tingkat kemudahan dan kepraktisan penggunaan dan pelaksanaan suatu tes, dalam hubungannya dengan biaya dan waktu untuk melaksanakan tes tersebut, serta pengolahan dan penafsiran hasilnya. Ada tiga macam perangkat yang dikembangkan dalam penelitian ini yaitu perangkat penilaian kognitif, afektif, dan psikomotor. Berdasarkan penilaian validator terhadap perangkat penilaian kognitif yang dikembangkan, perangkat penilaian kognitif juga memenuhi kriteria praktis yang ditetapkan pada Bab III. Dua dari ketiga validator memberikan nilai “B”, yang berarti perangkat penilaian kognitif yang dikembangkan dapat digunakan dengan sedikit revisi. Sedangkan satu Validator memberikan nilai “C” , yang berarti

perangkat yang dikembangkan dapat digunakan dengan banyak revisi. Hal ini terjadi karena peneliti membuat item soal yang jumlahnya sangat banyak. Sehingga membutuhkan waktu yang cukup lama dalam pelaksanaan, penskoran dan penafsiran hasil tes. Selain itu ada beberapa item soal dalam penilaian kognitif yang merupakan jenis penilaian proses. Hal itu menyebabkan perlu banyak revisi terhadap perangkat penilaian kognitif yang dikembangkan. Akan tetapi karena perangkat telah direvisi sesuai dengan saran Validator maka dapat dikatakan bahwa perangkat penilaian kognitif yang dikembangkan telah dapat digunakan, sehingga telah masuk pada kategori “praktis”. Walaupun demikian masih diperlukan perbaikan dan penyempurnaan lebih lanjut atau penyesuaian-penyempurnaan jika perangkat penilaian kognitif akan diterapkan pada kondisi lain.

Perangkat penilaian afektif yang dikembangkan pada penelitian ini, memenuhi kriteria “praktis” yang ditetapkan pada Bab III. Ketiga validator memberikan nilai “B”, yang berarti perangkat yang dikembangkan dapat digunakan dengan sedikit revisi. Walaupun demikian masih diperlukan perbaikan dan penyempurnaan lebih lanjut atau penyesuaian-penyempurnaan jika perangkat penilaian afektif akan diterapkan pada kondisi lain.

Berdasarkan penilaian validator terhadap perangkat penilaian psikomotor yang dikembangkan, perangkat penilaian psikomotor juga memenuhi kriteria praktis yang ditetapkan pada Bab III. Dua dari ketiga validator memberikan nilai “B”, yang berarti perangkat penilaian psikomotor yang dikembangkan dapat digunakan dengan sedikit revisi. Sedangkan satu Validator memberikan nilai “C”

, yang berarti perangkat yang dikembangkan dapat digunakan dengan banyak revisi. Hal ini terjadi karena peneliti membuat item soal yang jumlahnya sangat banyak. Sehingga membutuhkan waktu yang cukup lama dalam pelaksanaan, penskoran dan penafsiran hasil tes. Hal itu menyebabkan perlu banyak revisi terhadap perangkat penilaian psikomotor yang dikembangkan. Akan tetapi karena perangkat telah direvisi sesuai dengan saran Validator maka dapat dikatakan bahwa perangkat penilaian psikomotor yang dikembangkan telah dapat digunakan, sehingga telah masuk pada kategori “praktis”. Walaupun demikian masih diperlukan perbaikan dan penyempurnaan lebih lanjut atau penyesuaian-penyempurnaan jika perangkat penilaian psikomotor akan diterapkan pada kondisi lain.

#### **E. Kelemahan Penelitian**

Dari pembahasan diatas dapat dilihat bahwa dalam penelitian ini terdapat banyak kendala dan kelemahan, diantaranya sebagai berikut: 1) uji-coba instrumen hanya dilakukan selama satu kali uji-coba. Kegiatan uji-coba instrumen dan kegiatan yang mengikutinya ini perlu dilakukan lebih dari satu siklus. Mungkin perlu dilakukan uji-coba kedua guna menyempurnakan hasil uji-coba yang pertama, uji-coba ketiga guna menyempurnakan hasil uji-coba yang kedua, dan selanjutnya, sampai semua persyaratan yang direncanakan dalam tujuan indikator soal dipenuhi; 2) waktu uji-coba instrumen yang kurang efektif. Karena uji-coba terbatas dilaksanakan di awal tahun ajaran baru sehingga banyak siswa

yang sudah lupa mengenai materi yang ditekankan; 3) interpretasi koefisien reliabilitas pada hasil penelitian ini adalah spesifik bagi situasi dan kelompok subyek dalam penelitian ini saja; 4) pijakan teori yang digunakan oleh peneliti dalam mengembangkan evaluasi pada penelitian ini masih belum relevan. Sehingga perlu teori pengembangan evaluasi yang lebih spesifik.