

PENGEMBANGAN MODEL *ASSESSMENT FOR LEARNING* PADA PEMBELAJARAN MATEMATIKA DI SMP

Mansyur

Fakultas Teknik Universitas Negeri Makassar
Kampus Gunungsari Baru Jl. A.P. Pettarani Makassar 90222
aura_fz@yahoo.co.id

Abstrak

Tujuan penelitian ini adalah untuk menemukan cara mengembangkan model *assessment for learning* pada pembelajaran matematika. Penelitian ini merupakan penelitian pengembangan model Hopkins & Clark. Subjek uji coba adalah 20 orang guru matematika dan 199 orang siswa kelas VII SMPN 13 dan SMPN 27 Makasar. Analisis data dilakukan dua tahap dan secara kualitatif dan kuantitatif. Ujicoba model AfL menggunakan *repeated measures analysis*. Hasil penelitian menunjukkan bahwa model *assessment for learning* yang dikembangkan memiliki lima karakteristik utama. pembelajaran lebih memberikan rasa keadilan bagi semua siswa, dan cocok diterapkan untuk semua mata pelajaran. Beberapa temuan dalam penelitian ini yaitu: (1) informasi yang diperoleh melalui penggunaan Model-AfL akurat dan sesuai dengan kebutuhan nyata siswa dan (2) penerapan Model-AfL dalam pembelajaran matematika meningkatkan motivasi, kepercayaan diri, kesadaran diri siswa, perilaku siswa selama pembelajaran, dan kemampuan siswa terhadap matematika; dan (3) kemajuan belajar siswa ditampilkan melalui profil individu dan profil kelas.

Kata kunci: *asesmen, asesmen formatif, assessment for learning, pendidikan matematika, pembelajaran*

DEVELOPMENT OF AN ASSESSMENT-FOR-LEARNING MODEL IN MATHEMATICS LEARNING IN JUNIOR HIGH SCHOOLS

Mansyur

Fakultas Teknik Universitas Negeri Makassar
Kampus Gunungsari Baru Jl. A.P. Pettarani Makassar 90222
aura_fz@yahoo.co.id

Abstract

This study aimed to develop an assessment-for-learning (AfL) model in mathematics learning. This was a research and development study using the Hopkins & Clark model. The tryout subjects consisted of 20 mathematics teachers and 199 Year VII students of SMPN 13 and SMPN 27 Makassar. The data were analyzed in two stages by means of quantitative and qualitative techniques. The tryout of the AfL model used a repeated measures analysis. The results show that the developed AfL model has five main characteristics. The model provides all students with fairness and is appropriate for all subjects. Some research findings show that: (1) information obtained through the application of the AfL model is accurate and relevant to the students' real needs; (2) the application of the AfL model in mathematics learning improves students' motivation, self-confidence, self-awareness, behaviors during the learning process, and competence in mathematics; and (3) students' learning progress is presented through the individual and class profiles.

Keywords: *assessment, formative assessment, assessment for learning, mathematics education, and learning*

Pendahuluan

Penilaian dalam pembelajaran matematika merupakan bagian tidak terpisahkan dari proses pendidikan matematika. Upaya peningkatan kualitas pendidikan matematika dapat ditempuh melalui peningkatan kualitas pembelajaran dan kualitas sistem penilaian. Kualitas pembelajaran ini dapat dilihat dari hasil penilaiannya. Sebaliknya, sistem penilaian yang baik, akan mendorong guru untuk menentukan strategi yang tepat dan memotivasi siswa untuk belajar lebih baik. Salah satu bentuk penilaian yang diharapkan untuk hal tersebut adalah penilaian formatif. Namun demikian, kenyataan di lapangan, di sekolah-sekolah, baik dalam konteks internasional maupun domestik, penilaian formatif masih belum berfungsi.

Dalam konteks internasional, bukti terlihat dari hasil penelitian Fair Test Examiner (1999) menunjukkan bahwa penilaian formatif relatif jarang dilaksanakan di kelas, dan sebagian besar guru tidak mengetahui bagaimana menggunakan penilaian tersebut. Hasil yang sama dikemukakan oleh Black dan William (1998) bahwa sebagian besar pengujian di kelas mendorong belajar superfisial dan hafalan. Para guru tidak dapat menolong guru yang lainnya menjadi penilai yang baik, dan mereka sering menekankan kuantitas dari pada kualitas pekerjaan. Para guru, biasanya mereplikasi tes terstandarisasi dalam praktik penilaian mereka sendiri, sehingga kurang memberi informasi tentang siswa mereka.

Kemudian, dalam konteks Indonesia, pelaksanaan penilaian formatif pada pembelajaran matematika tidak jauh berbeda dengan internasional. Setidaknya hal tersebut dikuatkan dari hasil penelitian Zulkardi (2002) yang menyimpulkan bahwa dalam melakukan penilaian pembelajaran matematika, guru masih menggunakan format penilaian yang lemah. Alat penilaian yang belum dirancang dengan baik dan ditekankan pada hasil dari pada proses, masih mendominasi penilaian. Selanjutnya, Kusnanto (2006) melaporkan hasil survei terhadap siswa sekolah menengah di Semarang bahwa, siswa cenderung tidak memiliki sikap yang benar saat belajar matematika. Salah satu penyebabnya karena sistem penilaian buruk.

Hasil penelitian pendahuluan dan wawancara yang peneliti lakukan terhadap guru matematika SMP dan Madrasah Tsanawiah di Sulawesi Selatan mendukung kesimpulan yang dikemukakan di atas bahwa: (1) penilaian formatif belum berfungsi untuk mendiagnosis kesulitan belajar matematika siswa dan (2) guru belum memiliki kebiasaan untuk melakukan analisis diagnosis kesulitan belajar siswa. Hasil temuan awal tersebut, menguatkan kesimpulan hasil penelitian Kumaidi (2005) terhadap guru-guru di Sumatera Barat bahwa guru kurang memiliki pemahaman dan keterampilan pengujian, mulai dari pengembangan dokumen pengembangan tes, penulisan butir soal, telaah silang dan perbaikan soal, *editing* dan penyuntingan akhir, juga keterampilan melakukan analisis diagnostik. Selanjutnya, Kumaidi mengatakan bahwa penilaian untuk tujuan diagnostik (pemanfaatan diagnostik kesulitan belajar siswa) pada dasarnya jarang atau bahkan tidak dilakukan guru karena (1) kebanyakan guru tidak dapat melaksanakan analisis diagnostik kesulitan belajar siswa dan (2) tidak menjadi kebiasaan guru, guru lebih sering melaksanakan tujuan *grading*.

Akibat dari belum berfungsinya penilaian formatif pada level kelas, maka kualitas pembelajaran matematika masih belum tercapai. Setidaknya hal tersebut terlihat dari hasil studi TIMSS tahun 2006 yang menempatkan siswa-siswa Indonesia berada pada urutan 34 dari 38 negara yang disurvei. Untuk mengatasi kelemahan praktik penilaian formatif yang telah dikemukakan di atas dan untuk peningkatan kualitas kemampuan matematika, peneliti menawarkan suatu model penilaian yang terintegrasi dengan pembelajaran, yaitu model *assessment for learning* (Model-AfL) (Mansyur, 2009). Model ini merupakan hasil dari penelitian pengembangan yang dilakukan dengan menggunakan model pengembangan hasil penyesuaian dari the R, D, & D model yang dikembangkan oleh Hopkins & Clark (Havelock, 1976) dengan tujuan untuk menguji apakah model-AfL dapat meningkatkan kemampuan matematika siswa.

Tujuan penelitian ini adalah untuk (1) menemukan cara mengembangkan model *assessment for learning*, (2) mengetahui informasi apa saja yang dapat dimunculkan jika menggunakan model *assessment for learning*, (3) mengetahui bentuk pemanfaatan informasi hasil penilaian dengan menggunakan model *assessment for learning*, (4) mengetahui apakah model

assessment for learning dapat berfungsi untuk meningkatkan pemahaman, perilaku, dan kemampuan siswa, dan (5) mengetahui cara pelaporan hasil penilaian dengan menggunakan model *assessment for learning* pada pembelajaran matematika. Pada tulisan ini, peneliti hanya menyajikan tujuan penelitian yang ketiga yaitu khusus pada peningkatan kemampuan matematika melalui penerapan model *assessment for learning*.

Konsep *assessment for learning* pada dasarnya bukanlah hal baru dalam penilaian pendidikan, tetapi wujud penerapannya dalam konteks perbaikan dan peningkatan kualitas pembelajaran, *assessment for learning* lebih baik, terencana, terarah, dan terfokus. Setidaknya hal tersebut tercermin dari pengertian *assessment for learning* yang dikemukakan dalam *Assessment Reform Group* (2002), yang mengatakan bahwa:

Assessment for learning is the process of seeking and interpreting evidence for use by learners and their teachers to decide where the learners are in their learning, where they need to go and how best to get there.

Penekanan dari pengertian *assessment for learning* terletak pada proses perolehan informasi dan pemanfaatan informasi. Informasi atau keterangan diperoleh melalui kerjasama antara guru dengan siswa dan informasi tersebut dimanfaatkan oleh mereka (guru dan siswa) untuk perbaikan dan peningkatan kualitas pembelajaran berikutnya. Bagi guru, informasi digunakan untuk perbaikan dan penyempurnaan strategi pengajaran sesuai dengan kebutuhan nyata para siswanya. Sementara bagi siswa, dapat digunakan sebagai dasar dalam mengubah strategi belajar yang lebih baik. Pengertian yang hampir sama dikemukakan juga dalam *Pearson Education* (2006), yaitu:

Assessment for learning is a collaborative process between teacher and pupil, and with pupils engaging with each other in structuring their own learning. It is built on a foundation of shared learning objectives and shared criteria for success. Pupils are given the criteria for success and the support they need to achieve that success. Feedback, either during or on completion of the task, is essential if pupils are to know what else is must be done to ensure further learning. Pupils are provided with opportunities to participate in self- or peer-assessment as this develops an understanding of personal responsibility in learning.

Tampak pada pengertian terakhir, bahwa penekanannya juga terletak pada kolaborasi antara guru dengan siswa maupun antara sesama siswa. Kolaborasi mereka dalam kaitan dengan kegiatan pembelajaran dalam upaya menjadikan semua siswa sukses. Untuk mencapai kesuksesan tersebut, tugas guru seperti *sharing* tujuan pembelajaran dan kriteria sukses di awal pelajaran juga menjadi penekanan definisi tersebut. Pemberian masalah secara terstruktur untuk dikerjakan oleh siswa, kemudian guru memberikan umpan balik atas pekerjaan mereka, sehingga diperoleh informasi tentang kekuatan dan kelemahan siswa.

Model-AFL dalam penelitian ini mengadopsi kedua pengertian *assessment for learning* yang dikemukakan dalam *Assessment Reform Group* dan *Pearson Education*, yang memberi penekanan pada kolaborasi antara guru dan siswa serta pemanfaatan informasi hasil penilaian untuk memodifikasi strategi dan teknik mengajar dan belajar.

Operasionalisasi pengertian *assessment for learning* dalam kerangka tujuan, dinyatakan dalam CEA (2003) yang mengatakan bahwa tujuan *assessment for learning* untuk:

- a. memberi wawasan pembelajaran kepada guru dan siswa dalam upaya meningkatkan kesuksesan untuk semua,
- b. membantu proses penetapan tujuan,
- c. memungkinkan refleksi secara kontinyu terhadap apa yang siswa ketahui sekarang dan apa yang mereka butuhkan untuk diketahui berikutnya,
- d. mengukur apa yang dinilai,
- e. menetapkan intervensi secara cepat dan tepat sesuai dengan tujuan pembelajaran, dan
- f. meningkatkan standar yang diperoleh siswa.

Sehubungan dengan tujuan tersebut, kesuksesan dalam pembelajaran diperuntukkan bagi guru maupun siswa. Guru dituntut memiliki wawasan dan kemampuan profesional tentang pembelajaran, misalnya penguasaan materi, perencanaan, penetapan tujuan pembelajaran, dan membuat keputusan yang tepat, sehingga siswa termotivasi untuk memperbaiki dan meningkatkan belajarnya.

Untuk merealisasikan tujuan yang telah dikemukakan di atas, *Assessment Reform Group* (2002) memberikan sepuluh prinsip utama, yaitu *assessment for learning*:

- a. *should be part of effective planning of teaching and learning*
- b. *should focus on how students learn*
- c. *should be recognized as central to classroom practice*
- d. *should be recognized as a key professional skill for teachers*
- e. *should be sensitive and constructive because any assessment has an emotional impact*
- f. *should take account of the importance of learner motivation*
- g. *should promote commitment to learning goals and a shared understanding of the criteria by which they are assessed*
- h. *learners receive constructive guidance about how to improve*
- i. *develops learners capacity for self-assessment so that they can become reflective and self-managing*
- j. *should recognize the full range of achievements of all learners*

Aplikasi dari pengertian, tujuan, dan prinsip *assessment for learning* dalam pelaksanaan penilaian di kelas, maka disusun sintaks model-AfL. Sintak ini merupakan panduan dan petunjuk praktis bagi praktisi pendidikan, guru, dosen dalam penerapannya di dalam kelas. Sebagai panduan dalam pelaksanaan uji model secara empirik, maka hipotesis penelitian ini dirumuskan sebagai "Penerapan model-AfL dapat meningkatkan kemampuan matematika siswa".

Metode Penelitian

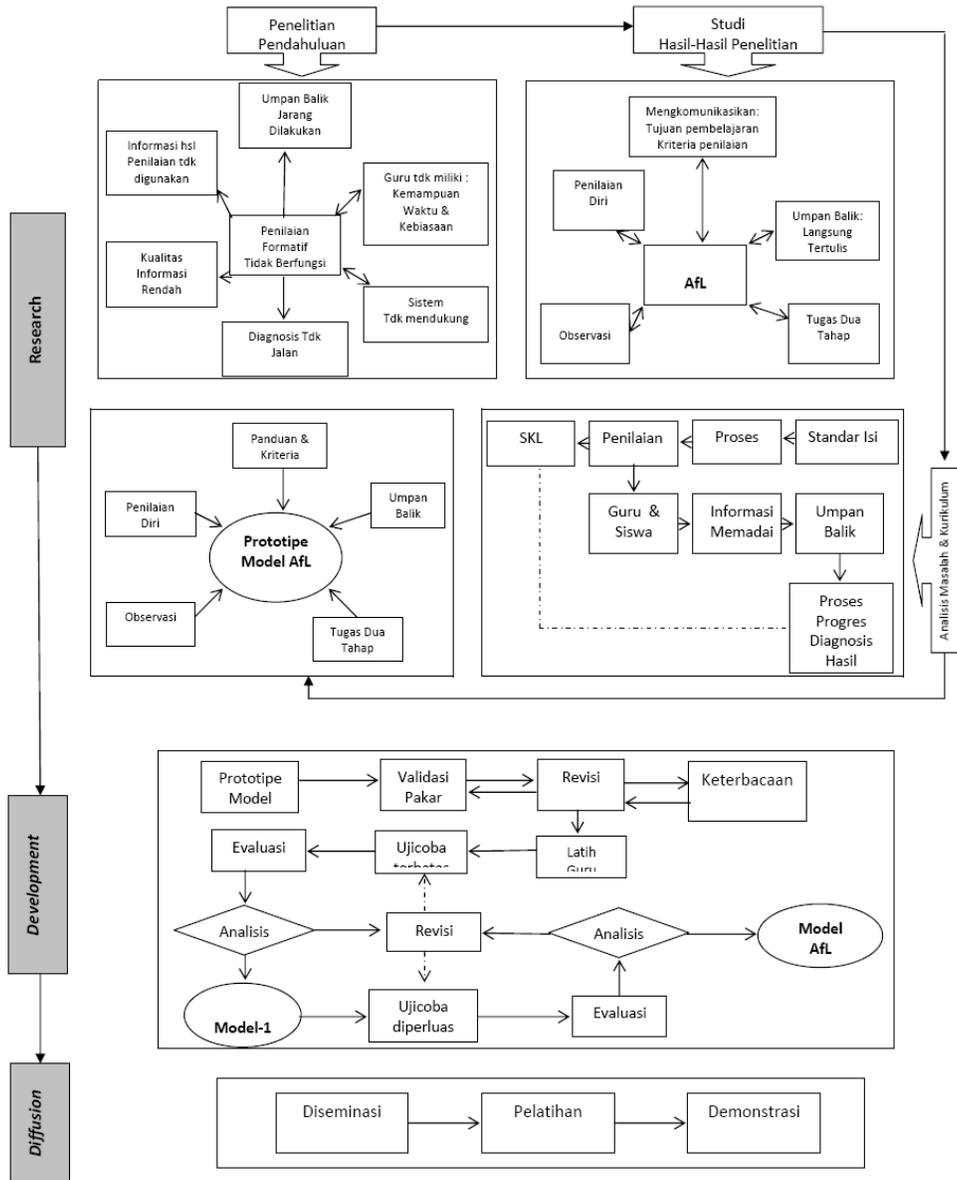
Penelitian ini merupakan penelitian pengembangan dengan mengadopsi model yang dikembangkan oleh Hopkins & Clark yaitu *the R, D & D model*' (Havelock, 1976). Pada tahap *research*, ada 4 tahap kegiatan yang dilakukan, yaitu penelitian pendahuluan, studi hasil-hasil penelitian, analisis kurikulum, dan penyusunan *prototype* model. Pada tahap *development*, ada 5 tahap kegiatan yang dilakukan, yaitu: validasi pakar, uji keterbacaan, latih guru, ujicoba terbatas, dan ujicoba diperluas. Pada tahap *diffusion*, ada

3 kegiatan yang dilakukan, yaitu diseminasi, pelatihan, dan demonstrasi. Disain penelitian pengembangan ini disajikan pada Gambar 1.

Prosedur pengembangan model *assessment for learning* mengacu pada disain yang telah dikemukakan di atas. Secara operasional, prosedur pengembangan yang dilaksanakan yaitu:

- a. Menyusun perangkat model, yaitu panduan praktis penggunaan model, instrumen penilaian efektivitas model, instrumen penilaian diri siswa, instrumen tugas dua tahap, kriteria dan rubrik penskoran, dan rencana pelaksanaan pembelajaran.
- b. Validasi pakar.
- c. Revisi.
- d. Keterbacaan perangkat pada guru-guru matematika SMP.
- e. Validasi pakar, keterbacaan, dan revisi prosesnya berulang.
- f. Latih guru.
- g. Ujicoba terbatas.
- h. Revisi model.
- i. Ujicoba diperluas.
- j. Ditemukan model *assessment for learning* yang fit antara kerangka teoritis dengan data empirik.

Pembuktian efektivitas model secara empirik dalam kerangka penelitian pengembangan menggunakan pendekatan kuasi eksperimen dengan disain *single-group interrupted time-series design* (Creswell, 1994: 133). Rancangan disain tersebut disajikan pada Tabel 1.



Gambar 1. Disain Pengembangan Model-AfL

Tabel 1. *Single-Group Interrupted Time-Series Design* untuk Uji Coba Model

Pertemuan ke-	1	2	3	4	5
Kelas A	X/O ₁	X/O ₂	X/O ₃	X/O ₄	X/O ₅

Keterangan:

X = perlakuan

O_i = pengukuran kemampuan matematika siswa, $i = 1, 2, \dots 5$.

Pelaksanaan ujicoba terbatas dan ujicoba diperluas masing-masing 2 bulan (setara dengan 16 kali pertemuan), yaitu dimulai pada bulan November 2007 sampai dengan bulan Februari 2008, dengan banyaknya pengukuran yang dilakukan 5 kali. Perlakuan dan observasi dilaksanakan bersamaan dalam suatu pertemuan. Sebelum pelaksanaan ujicoba peneliti terlebih dahulu mengontrol validitas internal dan validitas eksternal.

Responden penelitian terdiri atas guru dan siswa, keterlibatan kedua responden ini tergantung pada kebutuhan pada fase pengembangan. Untuk responden ujicoba terbatas, dipilih 1 kelas sebagai kelas perlakuan. Responden dari unsur guru dipilih 4 orang yang mengajar mata pelajaran matematika pada sekolah yang dijadikan sebagai tempat ujicoba terbatas (SMP negeri 27 Makassar). Guru tersebut, 1 orang untuk mengajar, dan 2 orang untuk dijadikan sebagai pengamat. Subjek coba untuk ujicoba diperluas, dipilih 2 sekolah, masing-masing terdiri atas 2 kelas. Setiap kelas diajar oleh masing-masing seorang guru dan diamati oleh seorang pengamat (dari unsur guru). Semua guru yang dijadikan sebagai responden penelitian berjumlah 20 orang, sebelum terjun ke kelas dilatih oleh peneliti tentang penggunaan Model-AfL. Subjek penelitian yang terlibat dalam tahap pengembangan model disajikan pada Tabel 2.

Tabel 2. Sebaran Subjek Penelitian

Jenis Subjek	Jumlah Subjek Ujicoba			Jumlah
	Keterbacaan	Terbatas	Diperluas	
Siswa	10 orang	39 orang	150 orang	199 orang
Guru	7 orang	4 orang	9 orang	20 orang

Metode pengumpulan data dan instrumen yang akan digunakan dalam pengembangan ini terdiri atas beragam teknik yang penggunaannya disesuaikan dengan kebutuhan dan jenis data yang dikumpulkan sesuai dengan tahap-tahap kegiatan pengembangan. Pada dasarnya ada dua kelompok instrumen, yaitu instrumen pengumpul data dan instrumen perlakuan. Instrumen perlakuan terdiri atas penilaian diri siswa, dan tugas dua tahap. Instrumen perlakuan ini sekaligus sebagai instrumen pengumpul data. Jadi instrumen pengumpul data terdiri atas lembar pengamatan perilaku, tugas dua tahap (untuk mengukur kemampuan dan miskonsepsi siswa terhadap materi pelajaran yang diberikan), dan lembar penilaian diri siswa.

Validitas dan reliabilitas hasil pengukuran instrumen-instrumen yang digunakan akan dilakukan tindakan validasi (Nitko & Brookhard, 2007: 38-44). Instrumen tugas dua tahap akan dilihat beberapa dimensi persyaratan alat ukur tes antara lain tingkat kesulitan, reliabilitas, validitas, dan daya pembeda soal (Cohen & Swerdlik, 2005: 212). Kemudian, untuk mengukur tingkat kesepakatan antar penilai (*inter-rater reliability*) terhadap instrumen lembar penilaian diri, lembar pengamatan perilaku, angket efektifitas model, keterlaksanaan model dalam kelas, dan lembar penilaian model hasil validasi digunakan koefisien Cohen's Kappa (Cohen, 2001: 657) dan *percentages of agreements* (Grinnell, 1988: 160). Untuk menghitung koefisien Cohen's Kappa (κ) digunakan formula yang dikemukakan oleh Cohen (2001: 657) sebagai berikut:

$$\kappa = \frac{\sum f_0 - \sum f_e}{N - \sum f_e}$$

dimana:

κ = tingkat kesepakatan penilai (koefisien reliabilitas antar penilai)

f_0 = frekuensi hasil pengamatan

f_e = frekuensi yang diharapkan

N = banyaknya butir soal yang dinilai (diklasifikasi)

Kemudian untuk menghitung tingkat *percentages of agreements* antara kedua penilai yang datanya hanya ya atau tidak digunakan rumus yang dikemukakan oleh Grinnell (1988: 160) sebagai berikut,

$$\text{Percentages of agreements} = \frac{\text{Agreements}}{(\text{Disagreements} + \text{Agreements})} \times 100$$

Batas bawah koefisien reliabilitas yang digunakan untuk suatu tes yang baik yaitu sebesar 0,70 (Linn, 1989: 106).

Analisis data hasil penelitian pendahuluan dilakukan dengan menggunakan pendekatan deskriptif disertai dengan narasi yang sesuai dengan kepentingan penelitian. Analisis data pada saat pengembangan model dilakukan dengan pendekatan kualitatif dan kuantitatif. Analisis kualitatif dilakukan untuk menganalisis data hasil validasi model oleh para ahli yang memberikan masukan dalam rangka perbaikan. Analisis kualitatif juga digunakan untuk menganalisis data hasil penilaian diri siswa terhadap pertanyaan-pertanyaan terbuka yang diberikan. Analisis untuk mengetahui perkembangan kemampuan, penilaian diri, dan perilaku siswa selama/ setelah ujicoba model AfL digunakan *repeated measures analysis* (Field, 2000: 326) dengan desain analisis data disajikan pada Tabel 3.

Tabel 3. *Repeated Measures Design* untuk Analisis Data

		Pertemuan				
		1	2	3	4	5
Subjek	1	Y11	Y12	Y13	Y14	Y15
	2	Y21	Y22	Y23	Y24	Y25
	3	Y31	Y32	Y33	Y34	Y35

	N	Yn1	Yn2	Yn3	Yn4	Yn5

Hasil Penelitian dan Pembahasan

Hasil validasi model oleh para pakar menunjukkan bahwa perangkat model seperti panduan dan instrumen model telah memenuhi syarat validitas dan reliabilitas. Semua hasil validasi pakar memberikan penilaian bahwa model layak digunakan tanpa revisi. Hasil uji efektivitas model secara empiris merupakan jawaban dari hipotesis penelitian. Uji efektivitas model-AfL secara empiris dilakukan dengan menggunakan analisis *repeated measures*. Dalam analisis, variabel yang diperhatikan adalah kemampuan siswa pada pengukuran yang berulang (*repeated measures*), dalam hal ini pengukuran dilakukan sebanyak 5 kali pada 2 sekolah masing-masing sekolah untuk 2 kelas. Untuk melakukan analisis dengan *repeated measures*, terlebih dahulu dilakukan uji asumsi, yaitu asumsi *sphericity* (Field, 2000: 333, Stevens, 1996: 459). Untuk keperluan pengujian asumsi tersebut, digunakan *Mauchly's Test of Sphericity*. Asumsi yang diuji adalah:

Ho : Matriks kovarians kesalahan hasil transformasi ortonormal dari variabel yang diperhatikan = matriks identitas

lawan

H1 : Matriks kovarians kesalahan hasil transformasi ortonormal dari variabel yang diperhatikan \neq matriks identitas

Kriteria pengujian: terima hipotesis H_0 jika nilai signifikansi hasil perhitungan $p \geq \alpha$, sebaliknya H_0 ditolak (Miller & Miller, 2004: 402). Hasil perhitungan uji asumsi tersebut dirangkum pada Tabel 4. Informasi yang disajikan dalam Tabel 4 tersebut, yaitu nilai statistik untuk *Mauchly's test* sebesar 0,235 dan signifikansi hasil perhitungan p sebesar 0,000001. Jika nilai signifikansi hasil perhitungan dibandingkan dengan signifikansi $\alpha = 0,05$, maka $\alpha > p$. Dengan demikian asumsi *sphericity* tidak terpenuhi. Field (2000: 333) mengatakan bahwa jika asumsi *sphericity* tidak terpenuhi, maka dapat dilakukan koreksi dengan melihat salah satu dari Epsilon (*Greenhous e-Geisser*, *Huynh-Feldt*, atau *Lower-bound*). Pada analisis ini, peneliti memilih *Epsilon Greenhous e-Geisser* yang nilai signifikansi hasil perhitungan ε

sebesar 0,863. Jika nilai ϵ ini dikaitkan dengan kriteria, yaitu $\frac{1}{k-1} \leq \epsilon \leq 1$ (k banyak pengukuran, dalam hal ini $k = 5$), maka ϵ yang diperoleh terletak dalam interval tersebut. Dengan demikian, dapat disimpulkan bahwa asumsi *sphericity* terpenuhi, sehingga uji *repeated measures* dapat dilanjutkan.

Tabel 4. Rangkuman Hasil Perhitungan *Mauchly's Test of Sphericity*

Measure: MEASURE_1							
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Pertemuan	.235	210.890	9	.000	.863	.765	.250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

Persyaratan analisis *repeated measures* terpenuhi sehingga dapat dilakukan analisis selanjutnya. Hasil perhitungan analisis *repeated measures* uji multivariat dirangkum pada Tabel 5. Perlakuan (penerapan model-AfL) untuk setiap pertemuan maupun interaksi antara pertemuan dengan kelas menunjukkan pengaruh yang signifikan. Setidaknya hal tersebut diperlihatkan oleh nilai signifikansi hasil perhitungan (p) masing-masing pertemuan dan interaksi pertemuan dengan kelas lebih kecil dari nilai signifikansi $\alpha = 0,05$ yang dipilih. Dengan demikian, dapat disimpulkan bahwa penerapan model-AfL dalam pembelajaran matematika memberikan pengaruh yang signifikan terhadap peningkatan kemampuan matematika siswa.

Tabel 5. Rangkuman Hasil Perhitungan Uji *Multivariate* pada Analisis *Repeated Measures*

Effect		Value	F	Hypothesis df	Error df	Sig.
Pertemuan	Pillai's Trace	.401	24.086 ^a	4.000	144.000	.000
	Wilks' Lambda	.599	24.086 ^a	4.000	144.000	.000
	Hotelling's Trace	.669	24.086 ^a	4.000	144.000	.000
	Roy's Largest Root	.669	24.086 ^a	4.000	144.000	.000
Pertemuan * Kelas	Pillai's Trace	.327	4.471	12.000	438.000	.000
	Wilks' Lambda	.700	4.579	12.000	381.280	.000
	Hotelling's Trace	.389	4.628	12.000	428.000	.000
	Roy's Largest Root	.255	9.323 ^b	4.000	146.000	.000

a. Exact statistic

b. The statistic is an upper bound on F that yields a lower bound on the significance level.

Hasil uji *multivariate* tersebut dikuatkan dari hasil *tests of within-subjects effects* seperti yang disajikan pada Tabel 6.

Tabel 6. Rangkuman Hasil Perhitungan *Tests Of Within-Subjects Effects* pada Analisis *Repeated Measures*

Measure: MEASURE_1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	
Pertemuan	Sphericity Assumed	478.079	4	119.520	74.611	.000
	Greenhouse-Geisser	478.079	2.503	191.025	74.611	.000
	Huynh-Feldt	478.079	2.601	183.772	74.611	.000
	Lower-bound	478.079	1.000	478.079	74.611	.000
Pertemuan * Kelas	Sphericity Assumed	80.261	12	6.688	4.175	.000
	Greenhouse-Geisser	80.261	7.508	10.690	4.175	.000
	Huynh-Feldt	80.261	7.804	10.284	4.175	.000
	Lower-bound	80.261	3.000	26.754	4.175	.007
Error(Pertemuan)	Sphericity Assumed	941.925	588	1.602		
	Greenhouse-Geisser	941.925	367.896	2.560		
	Huynh-Feldt	941.925	382.418	2.463		
	Lower-bound	941.925	147.000	6.408		

a. Computed using alpha = .05

Mencermati Tabel 6, tampak bahwa perlakuan (penerapan model-AfL) untuk setiap pertemuan maupun interaksi antara pertemuan dengan kelas menunjukkan pengaruh yang signifikan. Hal tersebut diperlihatkan oleh nilai signifikansi hasil perhitungan (*p*) masing-masing pada baris pertemuan (*p* = 0,0001) dan pertemuan* kelas (*p* = 0,0001) lebih kecil dari nilai signifikansi $\alpha = 0,05$ yang dipilih. Dengan demikian, dapat disimpulkan bahwa penerapan model-AfL dalam pembelajaran matematika memberikan pengaruh signifikan terhadap peningkatan kemampuan matematika siswa.

Selanjutnya, dilakukan uji untuk mengetahui apakah pengaruh tersebut bersifat linier atau tidak. Untuk keperluan tersebut digunakan *tests of within-subjects contrast*, yang hasilnya dirangkum pada Tabel 7. Tampak pada Tabel 7 tersebut, pada baris pertemuan, pengaruh yang sesuai bersifat linear. Hal tersebut diperlihatkan oleh nilai signifikansi hasil perhitungan *p* sebesar 0,00001 lebih kecil dari signifikansi $\alpha = 0,05$ untuk linear. Dengan

demikian, dapat dikatakan bahwa peningkatan jumlah pertemuan pada penerapan model-AfL pada pembelajaran matematika akan meningkatkan kemampuan siswa dalam matematika.

Tabel 7. Rangkuman hasil Perhitungan *Tests Of Within-Subjects Contrast* pada Analisis *Repeated Measures*

Measure: MEASURE_1

Source	Pertemuan	Type III Sum of Squares	df	Mean Square	F	Sig.
Pertemuan	Linear	469.010	1	469.010	138.258	.000
	Quadratic	7.066	1	7.066	5.464	.021
	Cubic	1.740	1	1.740	2.055	.154
	Order 4	.262	1	.262	.300	.585
Pertemuan * Kelas	Linear	44.822	3	14.941	4.404	.005
	Quadratic	23.536	3	7.845	6.066	.001
	Cubic	5.120	3	1.707	2.016	.114
	Order 4	6.783	3	2.261	2.582	.056
Error(Pertemuan)	Linear	498.665	147	3.392		
	Quadratic	190.118	147	1.293		
	Cubic	124.425	147	.846		
	Order 4	128.716	147	.876		

a. Computed using alpha = .05

Langkah berikutnya, yaitu melihat pasangan pertemuan mana yang memberikan rata-rata yang berbeda. Untuk keperluan tersebut diperhatikan hasil uji *post hoc* (dalam hal ini menggunakan metode Bonferroni), yang rangkuman hasil perhitungannya disajikan pada Tabel 8. Tampak pada Tabel 8 tersebut, nilai signifikansi hasil perhitungan p untuk semua pasangan yang diperhatikan adalah sama yaitu sebesar 0,00001. Jika nilai signifikansi hasil perhitungan tersebut dibandingkan dengan signifikansi $\alpha = 0,05$, maka $\alpha > p$. Dengan demikian dapat disimpulkan bahwa setiap pasang pertemuan yang diperhatikan memberikan perbedaan kemampuan siswa yang signifikan.

Tabel 8. Rangkuman Hasil Perhitungan Uji *Post Hoc* pada Analisis *Repeated Measures* Kasus Pertemuan

Measure: MEASURE_1

(I) Pertemuan	(J) Pertemuan	Mean Difference (I-J)	Std. Error	Sig. ^a
1	2	-.809*	.155	.000
	3	-1.407*	.166	.000
	4	-1.790*	.180	.000
	5	-2.302*	.191	.000
2	1	.809*	.155	.000
	3	-.598*	.138	.000
	4	-.981*	.142	.000
	5	-1.492*	.164	.000
3	1	1.407*	.166	.000
	2	.598*	.138	.000
	4	-.383*	.091	.000
	5	-.894*	.115	.000
4	1	1.790*	.180	.000
	2	.981*	.142	.000
	3	.383*	.091	.000
	5	-.512*	.072	.000
5	1	2.302*	.191	.000
	2	1.492*	.164	.000
	3	.894*	.115	.000
	4	.512*	.072	.000

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

Setelah melihat perbedaan kemampuan siswa untuk setiap pertemuan, langkah berikutnya adalah melihat apakah perbedaan kelas memberikan perbedaan kemampuan siswa dalam matematika. Untuk mengetahui hal tersebut, perhatikan Tabel 9. Informasi yang disajikan dalam Tabel 9, nilai signifikansi hasil perhitungan p untuk pasangan kelas VII-5 dan VII-A sebesar 0,021 serta pasangan kelas VII-5 dan VII-B sebesar 0,005. Jika nilai signifikansi hasil perhitungan tersebut dibandingkan dengan signifikansi $\alpha = 0,05$, maka $\alpha > p$. Dengan demikian dapat disimpulkan bahwa dua pasangan kelas tersebut memiliki perbedaan kemampuan matematika. Empat pasangan lainnya tidak memiliki perbedaan atau memiliki kemampuan matematika yang sama.

Tabel 9. Rangkuman Hasil Perhitungan Uji *Post Hoc* pada Analisis *Repeated Measures* Kasus Kelas

Measure: MEASURE_1

Bonf erroni

(I) Kelas VII	(J) Kelas VII	Mean Difference (I-J)	Std. Error	Sig.
Kelas VII-1	Kelas VII-5	-.2875	.34505	1.000
	Kelas VII-A	.7245	.33111	.181
	Kelas VII-B	.8874	.33318	.052
Kelas VII-5	Kelas VII-1	.2875	.34505	1.000
	Kelas VII-A	1.0121*	.34096	.021
	Kelas VII-B	1.1749*	.34296	.005
Kelas VII-A	Kelas VII-1	-.7245	.33111	.181
	Kelas VII-5	-1.0121*	.34096	.021
	Kelas VII-B	.1629	.32893	1.000
Kelas VII-B	Kelas VII-1	-.8874	.33318	.052
	Kelas VII-5	-1.1749*	.34296	.005
	Kelas VII-A	-.1629	.32893	1.000

Based on observed means.

*. The mean difference is significant at the .05 level.

Berdasarkan hasil analisis *repeated measures* tersebut di atas, dapat diketahui bahwa penerapan model-AfL pada pembelajaran matematika memberikan pengaruh yang signifikan terhadap peningkatan kemampuan matematika siswa. Peningkatan kemampuan siswa mengikuti tren linear, yaitu setiap peningkatan pertemuan (tentunya disertai dengan pemberian tugas dua tahap untuk mengukur kemampuan siswa) akan memberikan peningkatan pemahaman siswa terhadap materi matematika. Kenyataan tersebut, mengindikasikan bahwa model-AfL yang diterapkan pada pembelajaran, efektif untuk peningkatan kualitas pembelajaran matematika. Di samping memberikan pengaruh pada peningkatan kemampuan, hasil penelitian juga memberikan pengaruh pada peningkatan kesadaran, motivasi, tanggung jawab, dan perilaku siswa dalam pembelajaran.

Simpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dikemukakan dalam bab sebelumnya, maka kesimpulan yang menjadi temuan penelitian ini yaitu: (1) Model-AfL pada pembelajaran matematika di Sekolah Menengah Pertama dikembangkan dengan metode *research and development*. Tahapan *research* meliputi prasarvai, analisis masalah, analisis kurikulum, studi hasil penelitian, konsultasi pakar, penyusunan *prototype* model. Kemudian, tahapan pengembangan meliputi validasi pakar, uji keterbacaan, pelatihan guru, ujicoba terbatas, dan ujicoba diperluas sehingga model fit secara teortis dan empiris; (2) Informasi yang dimunculkan jika menggunakan Model-AfL dalam pembelajaran adalah informasi yang akurat dan sesuai dengan kebutuhan nyata siswa dalam hal pemahaman siswa terhadap materi pembelajaran, perilaku siswa selama pembelajaran, dan kemampuan matematika siswa; (3) Pemanfaatan informasi hasil penilaian dalam Model-AfL dilakukan melalui umpan balik dan refleksi; (4) Penerapan Model-AfL dalam pembelajaran matematika meningkatkan: pemahaman, perilaku, dan kemampuan matematika siswa; dan (5) Kemajuan belajar siswa ditampilkan melalui profil individu dan profil kelas. Kedua profil tersebut menampilkan tren perkembangan pemahaman individu (atau kelas) terhadap materi pembelajaran, perilaku siswa selama pembelajaran, dan kemampuan siswa terhadap matematika untuk setiap pertemuan.

Daftar Pustaka

- Assessment Reform Group. (2002). *Assessment for learning: 10 principles*. Norwich: DfES Publications. Diakses tanggal 02 Pebruari 2006 dari <http://www.assessment-reform-group.org.uk>.
- Black, P. & Wiliam, D. (1998). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80(2), 139-148.

- CEA. (Juli 2003). *Quality statement on assessment practice (secondary)*. Diakses tanggal 01 Pebruari 2006 dari <http://www.aaia.org.uk>.
- Cohen, B.H. (2001). *Explaining psychological statistics*. (2nd ed.). New York: John Wiley & Sons, Inc.
- Cohen, R.J & Swerdlik, M.E. (2005). *Psychological testing and assessment: An introduction to test and measurement*. (6th ed.). New York: McGraw Hill
- Creswell, J.W. (1994). *Research design: Qualitative & quantitative approaches*. Thousand Oaks: Sage Publications.
- Fair Test Examiner. (1999). *The value of formative assessment*. Diakses tanggal 17Juni2006 dari <http://www.provost.cmich.edu/assessment/toolkit/formativesummative.html>
- Field, A. (2000). *Discovering statistics using SPSS for Windows: Advanced techniques for the beginner*. Thousand Oaks: Sage Publications.
- Grinnell, R.M. Jr. 1988. *Social work research and evaluation*. (3rd ed.). Itasca, Illinois: F.E. Peacock Publisher, Inc.
- Havelock, R.G. (1976). *Planning for innovation: Through dissemination and utilization of knowledge*. Michigan: Institute for Social Research. The University of Michigan.
- Kumaidi. (2005). *Profil siswa dan kelas untuk bantuan peningkatan pembelajaran*. Makalah disampaikan pada Seminar Nasional Hasil Penelitian tentang Evaluasi Hasil Belajar serta Pengelolaannya, Tanggal 14-15 Mei 2005. Yogyakarta.
- Kusnanto, B.A.(01 Pebruari 2006). Tiga metode baru pembelajaran matematika. *Suara Merdeka*, p. 1.

Linn, R.L. (1989). *Educational measurement*. (3rd ed.). New York: Macmillan Publishing Company.

Mansyur. (2009). Pengembangan model assessment for learning pada pembelajaran matematika di SMP. Universitas Negeri Yogyakarta: Disertasi Doktor. Tidak diterbitkan.

Miller, I. & Miller, M. (2004). *Mathematical statistics with applications*. (7th ed.). New York: Pearson Prentice Hall.

Nitko, A.J., & Brookhart, S.M. (2007). *Educational assessment of students*. New Jersey: Pearson Merrill Prentice Hall.

Pearson Education. (2006). *Literacy objective: Assessment for learning*. Diakses tanggal 30 Juni 2006 dari <http://www.tda.gov.uk/teachers/hottopics/assessment.aspx>.

Stevens, J. (1996). *Applied multivariate statistics for the social sciences*. (3rd ed.). New Jersey: Lawrence Erlbaum Associates.

Zulkardi. (2002). RME: *Suatu inovasi dalam pendidikan matematika di Indonesia*. Diakses tanggal 10 Juni 2005 dari <http://www.cascadeimei.com>.