

# PENINGKATAN KINERJA GURU DALAM MENYUSUN SOAL OBJEKTIF BENTUK PILIHAN GANDA MELALUI UMPAN BALIK PADA GURU-GURU PAI KECAMATAN PANCUR BATU KABUPATEN DELI SERDANG

Ahmad Suwarno<sup>1</sup>

Pengawas PAI Tingkat Dasar Pada Kantor Kementerian Agama Kabupaten Deli Serdang

---

## Abstract

Received:  
Revised:  
Accepted:

Hasil penelitian dapat dijelaskan sebagai berikut. Secara proses terjadi peningkatan kinerja guru peserta dari masing-masing siklus. Pada siklus pertama skor kinerja guru adalah 40%, kemudian meningkat menjadi 91% pada siklus kedua. Skor kinerja guru yang diperoleh pada siklus ketiga ini dapat melebihi skor yang ditargetkan (90%). Dari aspek produk ditemukan bahwa: (1) pada siklus pertama ketercapaian validitas butir rata-rata adalah 11 (31,43%), taraf kesukaran cenderung berada pada  $P \geq 0,67$ , yaitu 17 (48,57%), indeks daya pembeda yang baik  $D \geq 0,30$  adalah 37,14%, dan pengecoh berfungsi 23 (69%); (2) pada siklus kedua, ketercapaian validitas butir soal rata-rata 15 (42,86%), butir soal cenderung berada pada rentangan  $P=0,33-0,67$ , yaitu 23 (65,71%), indeks daya pembeda yang baik  $D \geq 0,30$  menjadi 21 (60%), dan pengecoh berfungsi 83%; (3) pada siklus ketiga, ketercapaian validitas butir rata-rata 23 (66%), soal cenderung pada rentangan  $P=0,33-0,67$  yaitu 24 (68,57%), soal yang memiliki daya beda yang baik  $D \geq 0,30$  meningkat 26 (74,28%). Data di atas menggambarkan bahwa secara umum terjadi peningkatan mutu produk soal dari siklus pertama, kedua, dan ketiga. Secara khusus, untuk validitas butir soal, peningkatan yang diharapkan terjadi pada siklus ketiga, karena tiga di antara soal buatan guru memiliki jumlah butir valid di atas 70%. Taraf kesukaran soal yang diharapkan di mana  $P=0,33-0,67$  terjadi pada siklus kedua dan ketiga. Indeks daya pembeda soal yang baik juga diperoleh pada siklus ketiga, di mana soal yang memiliki  $D \geq 0,30$  mencapai 74,28%, lebih besar dari kriteria 70% yang ditetapkan.

**Keywords:** Kinerja Guru, Pilihan Ganda, Umpan Balik

(\*) Corresponding Author: Ahmad Suwarno

**How to Cite:** Peningkatan Kinerja Guru ... (2021) ANSIRU PAI: Pengembangan Profesi PAI.

---

## PENDAHULUAN

Guru sebagai tenaga profesional di bidang pendidikan mempunyai peran yang penting dalam membimbing peserta didiknya untuk mencapai tujuan pembelajaran. Salah satu kegiatan pokok dalam melaksanakan tugas profesinya itu adalah mengevaluasi hasil pembelajaran. Evaluasi merupakan proses mengambil keputusan apakah seorang siswa dikategorikan berhasil atau gagal dalam mengikuti kegiatan pembelajaran.

Pengambilan keputusan yang salah, selain merugikan pihak siswa, juga merugikan orang tua dan sekolah tempat mereka belajar. Bagi siswa, hasil belajar atau nilai yang diperoleh merupakan gambaran tentang kemampuan mereka selama mengikuti proses pembelajaran, yang dapat mempengaruhi keberhasilannya untuk melanjutkan pendidikan ke tingkat yang lebih tinggi dan dalam meniti karirnya di masyarakat. Oleh karena itu, guru dituntut melaksanakan

profesinya ini secara benar dan bertanggung jawab. Keputusan yang dibuat oleh guru hendaklah menggambarkan kemampuan siswa yang sesungguhnya.

Salah satu kegiatan pokok yang dilakukan guru dalam mengevaluasi hasil belajar siswa adalah membuat soal. Kegiatan ini cukup kompleks karena terkait dengan kegiatan pembelajaran lainnya. Kegiatan ini mulai dari menetapkan kompetensi dasar, memilih materi ajar, membuat tujuan pembelajaran, membuat kisi-kisi soal, membuat dan menganalisis soal sebelum diujikan kepada siswa. Ketika menganalisis soal guru mempertimbangkan banyak aspek, antara lain: validitas soal (baik validitas isi maupun validitas empiris), taraf kesukaran soal, daya pembeda soal, dan keberfungsian pengecohnya. Disamping itu agar diperoleh hasil dan gambaran yang tepat tentang peserta tes, guru juga perlu memelihara keutuhan soal dan mengawasi jalannya tes dengan baik.

Dari gambaran di atas, diketahui bahwa membuat soal adalah suatu proses yang memerlukan langkah-langkah dan prosedur tertentu. Oleh sebab itu, untuk mendapatkan soal yang bermutu, guru perlu mengetahui proses, langkah-langkah, dan prosedur menyusun soal yang benar. Soal yang dibuat melalui proses dan prosedur yang benar akan memberikan gambaran yang benar pula tentang kemampuan belajar siswa.

Dalam rangka mengetahui sejauhmana kompetensi guru dalam membuat soal di lapangan, peneliti melakukan observasi terhadap beberapa orang guru PAI di kecamatan Pancur Batu yang menjadi binaan penulis di wilayah Pancur Batu Kabupaten Deli Serdang. Selain penulis melakukan wawancara awal dengan guru mitra, penulis juga mengobservasi proses pelaksanaan ulangan harian dan melakukan analisis terhadap butir soal yang diujikan kepada siswa.

Berdasarkan studi awal yang penulis lakukan, diperoleh gambaran bahwa guru belum sepenuhnya mampu menghasilkan soal yang bermutu. Hasil analisis terhadap soal buatan guru menunjukkan bahwa: (1) secara isi, masih ada soal yang tidak sesuai dengan program pembelajaran sebagaimana tercantum dalam indikator hasil belajar; (2) secara empiris, dari 35 butir soal yang diujikan, hanya 17,14 % atau 6 butir soal yang memenuhi syarat validitas, 25,71 % atau 9 butir soal berada pada taraf kesukaran yang diharapkan ( $P = 0,33 - 0,67$ ), 31,42 % atau 11 butir soal memiliki indek daya pembeda dalam taraf penerimaan ( $D = \geq 0,30$ ).

Melihat hasil di atas, diperlukan bimbingan lebih lanjut kepada para guru PAI dalam bentuk umpan balik untuk meningkatkan kinerja guru dalam menyusun soal yang lebih baik. Hal ini membuat penulis tertarik untuk mengadakan penelitian seputar pengaruh umpan balik terhadap peningkatan kinerja guru dalam menyusun soal khususnya dalam bentuk objektif.

## **LANDASAN TEORI**

### **Kinerja Guru dalam Membuat Soal.**

Semenjak tahun 1996 kata kinerja (*performance*) menjadi *trend* dan diterima oleh banyak kalangan, termasuk dalam dunia pendidikan. Kinerja menurut Kubiszyn dan Borich (2003) adalah suatu bentuk demonstrasi yang dilakukan seseorang sebagai hasil pemahamannya setelah melalui proses belajar. Nitko (1996) mengemukakan bahwa penilaian kinerja (*performance assessment*) adalah suatu prosedur untuk memperoleh informasi tentang seberapa baik tugas-tugas pekerjaan dilakukan seseorang. Dijelaskan bahwa penilaian kinerja itu melihat proses (*process*), yakni bagaimana cara melakukan dan melihat hasil

(*product*), yakni apa yang dihasilkan. Lebih lanjut Nitko (1996) menganjurkan untuk melihat kinerja seseorang hendaknya memperhitungkan kedua aspek, karena antara proses dengan hasil adalah sama pentingnya (*equally important*).

Linn dan Gronlund (1995) dan Nitko (1996) memberikan kesimpulan bahwa terdapat tujuh keuntungan dalam menilai kinerja seseorang dalam pendidikan. Tujuh keuntungan tersebut adalah: (1) memperjelas arti dari target belajar yang kompleks, (2) mengetahui kemampuan membuat (*ability to do*), (3) konsisten dengan teori belajar modern, (4) memperlihatkan kombinasi kemampuan, pengetahuan, keahlian, dan kecakapan bekerja seseorang, (5) dapat melihat proses yang dikerjakan dan produk yang dihasilkan seseorang, (6) merupakan pendekatan penilaian yang berkembang, dan (7) lebih menampakkan hasil yang sesungguhnya.

Di dalam pendidikan, peningkatan kinerja guru pada dasarnya adalah usaha “modifikasi diri” (*self modification*) bagi guru itu sendiri (Gordon, 1997). Dalam menilai kinerja guru (*teacher performance*), yang dilihat tidak hanya apa yang dilakukan oleh guru, tetapi juga dilihat seberapa baik dan bagaimana kompetensinya dalam melakukan pekerjaan itu (Medley dan Soar, 1984). Artinya bahwa kinerja itu tidak sekedar melihat aspek kuantitas saja, tetapi juga melihat bagaimana kualitasnya. Lebih Lanjut Medley dan Soar (1984) mengemukakan empat hal untuk melihat kinerja seseorang, termasuk guru, yaitu: (1) tentukan tugas-tugas apa yang dilakukan, (2) buat catatan tingkah laku yang permanen, (3) buat skor kinerja, (4) bandingkan skor yang diperoleh dengan kriteria yang telah ditetapkan.

Berdasarkan uraian di atas dapat disimpulkan bahwa kinerja adalah kegiatan atau tindakan yang dilakukan atau praktik pengalaman seseorang untuk mendapatkan suatu hasil yang diharapkan. Kinerja dapat dilihat melalui bagaimana kegiatan itu dilakukan (prosesnya) dan apa yang dihasilkan (produknya) yang didasarkan kepada suatu standar atau kriteria tertentu. Kinerja guru membuat soal merupakan kegiatan yang dilakukan oleh guru mulai dari merencanakan program pembelajaran, menetapkan tujuan, memilih materi soal, menyusun kisi-kisi, membuat tes (menganalisis dari aspek bahasa dan ketepatan pilihan), mengujikan soal, mengawasi pelaksanaan ujian, melaksanakan pengukuran, menganalisis data hasil belajar, memberi penilaian, sampai menetapkan keputusan tentang hasil belajar siswa.

### **Umpan Balik.**

Banyak faktor yang dapat menyebabkan seseorang berkembang dalam melaksanakan tugas profesionalnya. Faktor itu ada yang berasal dari dalam dirinya sendiri dan ada pula yang berasal dari luar dirinya. Faktor yang berasal dari dalam diri seseorang disebut sebagai faktor internal, sedangkan yang berasal dari luar dirinya disebut faktor eksternal. Umpan balik yang diberikan seseorang terhadap yang lainnya dengan tujuan orang yang diberi umpan balik itu mau berubah dari kebiasaan-kebiasaan lamanya yang jelek menjadi lebih baik, menurut kategori ini termasuk orang yang berkembang melalui faktor eksternal.

Menurut Rosset dan Arwady (1987), umpan balik merupakan salah satu solusi bagi pemecahan masalah kinerja (*performance problems*) yang dialami seseorang. Umpan balik merupakan bentuk intervensi positif yang dilakukan

terhadap orang lain, misalnya dengan mengajar seseorang untuk melakukan sesuatu (*teach someone to do something*).

Heinich *et al.* (1996) mengemukakan bahwa umpan balik (*feedback*) adalah suatu bentuk kritis dari proses komunikasi, terutama dalam komunikasi instruksional di mana penerima pesan merespon pesan yang disampaikan. Johnson and Johnson dalam Heinich *et al.* (1996) mengemukakan bahwa umpan balik itu bertujuan untuk memberi respon apakah pekerjaan yang dilakukan seseorang atau sekelompok orang sudah dilakukan dengan baik atau belum. Lebih jauh dijelaskan, di dalam pendidikan umpan balik merupakan cara yang praktis untuk membantu individu dalam mencapai tujuannya.

Umpan balik dapat menentukan apakah informasi yang disampaikan telah dimengerti oleh si penerima pesan, yang biasanya diperoleh dengan cara memberikan sejumlah pertanyaan dan mengobservasi informasi yang diterima (Gentry, 1994). Dijelaskan pula bahwa dalam memberikan umpan balik terdapat pesan tentang apa yang telah dan belum dicapai selama suatu kegiatan berlangsung. Melalui umpan balik perbaikan terhadap suatu pekerjaan dapat dilakukan.

Selain itu umpan balik adalah salah satu cara orang memberikan bantuan satu sama lainnya. Melalui bantuan itu seseorang dapat mengatasi kelemahan-kelemahannya. Dengan kata lain, umpan balik merupakan upaya seseorang, *observer*, dan peneliti untuk mengarahkan perilaku individu atau sekelompok orang pada suatu tujuan tertentu (Utomo dan Ruijter, 1985). Lebih lanjut Utomo dan Ruijter (1985) mengemukakan bahwa ada tiga tahap yang perlu dilakukan oleh *observer* dalam memberikan umpan balik, yaitu: (1) pengamatan (*observation*), (2) diagnosis, dan (3) tindakan (*treatment*).

Dapat disimpulkan bahwa umpan balik adalah memberikan bantuan (saran maupun ide) kepada seseorang atas kekurangan yang masih ia lakukan dalam mengerjakan suatu pekerjaan. Umpan balik terhadap guru mitra maksudnya memberikan masukan dan koreksi terhadap cara-cara membuat soal yang dilakukannya selama ini, kemudian menentukan langkah-langkah perbaikan sehingga kinerja guru maupun peneliti sendiri dalam membuat soal semakin meningkat.

### **Mutu Soal Buatan Guru**

Soal yang bermutu adalah soal yang dapat memenuhi persyaratan mutu soal, yaitu: (1) valid, (2) reliabel, (3) memiliki taraf kesukaran yang baik, (4) mempunyai daya pembeda yang baik, dan (5) pengecohnya berfungsi. Mutu soal dapat diketahui dari hasil analisis soal (Wiersma dan Jurs, 1990).

Validitas soal adalah kecocokan atau ketepatan suatu tes dalam mengukur sesuatu yang hendak diukur (Gronlund, 1990). Ada tiga bentuk validitas soal, yaitu: (1) validitas isi (*content validity*), (2) validitas kriteria (*criterion-related validation*), dan (3) validitas konstruk (*construct validation*) (Crocker dan Algina, 1986).

Validitas isi melihat kesesuaian dan keterwakilan butir soal dengan keseluruhan materi yang diberikan dalam proses belajar mengajar. Langkah-langkah yang dilakukan untuk mendapatkan validitas isi suatu tes di antaranya: (1) mendefinisikan kemampuan yang hendak diukur secara jelas, (2) menyeleksi soal bersama pakar (*expert review*) yang jumlahnya menurut Tessmer (1995)

cukup satu atau dua orang saja. Validitas tes diketahui melalui rumus Point Biserial (*r-bis*), yaitu:

$$P_{pbis} = \frac{\mu_i - \mu_t}{\sigma_x} \sqrt{\frac{p}{q}}$$

Dimana :

$P_{pbis}$  = koefisien korelasi biserial antara skor butir soal nomor ke *i* dengan skor total;  $\mu_i$  = rata-rata skor total responden yang menjawab benar butir ke *i*;

$\mu_t$  = rata-rata skor total semua responden;

$\sigma_x$  = standar deviasi;

*p* = populasi jawaban benar untuk butir soal nomor ke-*i*;

*q* = proporsi jawaban salah untuk butir soal nomor ke *i* (1-*p*).

Pembandingan adalah *r*-tabel pada alpha 0,05. Bila skor koefisien korelasi butir soal dengan skor total lebih besar dari *r*-tabel, maka butir soal tersebut dianggap valid (Crocker & Algina, 1986).

Taraf kesukaran butir soal (*item difficulty*) adalah persentase siswa yang menjawab benar suatu butir soal yang diukur melalui skor butir soal yang dihasilkan oleh jawaban sejumlah peserta ujian (Linn & Gronlund 1995). Makin banyak peserta ujian yang berhasil dengan benar mengerjakan suatu butir soal, makin tidak sukar atau makin mudah butir soal tersebut (McBeath, 1992). Butir soal yang baik adalah butir soal yang tidak terlalu sukar dan tidak terlalu mudah, yaitu  $P = 0,5$ . Tuckman dalam Naga (1992) menganjurkan bahwa sebaiknya taraf kesukaran butir soal itu berada antara  $0,33 \leq P_{ij} \leq 0,67$ . Butir soal dengan nilai  $P = < 0,33$  adalah butir soal yang sukar, sedangkan butir soal dengan nilai  $P = > 0,67$  adalah butir soal yang mudah. Hal ini menunjukkan bahwa semakin mendekati angka 1 suatu butir soal, semakin mudah soal tersebut, sebaliknya semakin mendekati angka 0 suatu butir soal, semakin sulit soal tersebut.

Indek daya pembeda adalah suatu indek yang menyatakan seberapa efisien suatu butir soal dapat membedakan peserta kelompok tinggi dengan peserta kelompok rendah (Gregory, 2000). Daya pembeda butir soal dapat dilihat melalui rumus  $D = P_u - P_i$  ( Crocker dan Algina, 1996), dimana adalah proporsi kelompok atas yang menjawab butir soal dengan benar dan adalah proporsi kelompok bawah yang menjawab butir soal secara benar. Sedangkan *D* adalah nilai indek diskriminasi pada rentangan antara -1,00 - 1,00 dengan ketentuan:  $D = \geq 0,40$  berarti cukup memuaskan;  $0,30 \leq D \leq 0,39$  berarti sedikit atau tanpa revisi;  $0,20 \leq D \leq 0,29$  berarti perbatasan atau perlu direvisi;  $D \leq 0,19$  berarti soalnya dibuang.

Sehubungan dengan batas penerimaan butir soal dari aspek daya pembeda, Crocker dan Algina (1996) mengemukakan bahwa suatu butir soal dapat dipertahankan apabila memiliki indek daya pembeda  $D \leq 0,30$  (Linn dan Gronlund, 1995). Pengelompokan nilai butir soal di atas selanjutnya dijadikan sebagai tolok ukur membuat keputusan apakah soal yang dibuat guru sudah memenuhi target yang ditetapkan atau belum.

Dalam setiap tes objektif selalu digunakan alternative jawaban yang mengandung dua unsure sekaligus, yaitu jawaban tepat dan jawaban yang salah sebagai pengecoh (*distractor*). Tujuan pemakaian pengecoh ini adalah mengecohkan mereka yang kurang mampu (tidak tahu) untuk dapat membedakan

dengan yang mampu. Oleh karena itu pengecoh yang baik adalah yang dapat dihindari oleh anak-anak yang pandai dan terpilih oleh anak-anak yang kurang pandai, jangan sampai terjadi sebaliknya.

Efektivitas pengecoh-pengecoh yang ada pada suatu item dianalisis dari distribusi jawaban terhadap item yang bersangkutan pada setiap alternative yang disediakan. Efektivitas pengecoh diperiksa untuk melihat apakah semua pengecoh atau semua pilihan jawaban yang bukan kunci telah berfungsi sebagaimana mestinya, yaitu apakah pengecoh-pengecoh tersebut telah dipilih oleh lebih banyak (atau semua) siswa kelompok rendah sedangkan siswa dari kelompok tinggi hanya sedikit (atau tidak ada) yang memilihnya. Menurut Toha (1996), pengecoh yang baik paling tidak harus terpilih oleh sedikitnya 2 % .

Pusat Penilaian Pendidikan Balitbang Departemen Pendidikan Nasional (2007) telah menentukan jumlah alternative jawaban pada soal pilihan ganda. Untuk soal tingkat sekolah dasar (SD) dan sekolah menengah pertama (SMP) jumlah alternative jawaban soal pilihan ganda adalah 4 (empat), sedangkan untuk soal tingkat sekolah menengah atas (SMA) jumlah alternative jawaban yang ditentukan adalah berjumlah 5 (lima).

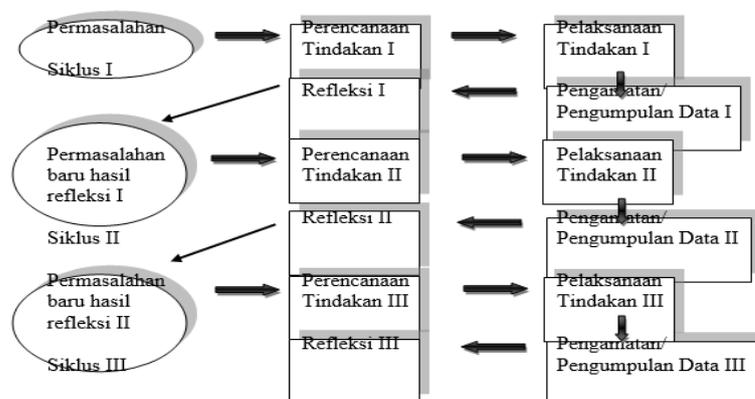
## METODE PENELITIAN

### Jenis Penelitian

Penelitian ini adalah penelitian tindakan (*Action Research*). Carr dan Kemmis dalam McNiff (1992) mengemukakan bahwa penelitian tindakan adalah suatu bentuk refleksi alamiah yang dilakukan oleh para partisipan, guru, dan peserta didik untuk meningkatkan aspek-aspek praktis. Gay (1996) mengemukakan bahwa tujuan penelitian tindakan adalah untuk memecahkan masalah praktis melalui aplikasi metode ilmiah (*the purpose of action research is to solve practical problems through the application of scientific method*). Hall & Hall (1996) mengemukakan bahwa salah satu ciri penelitian tindakan adalah peneliti itu sendiri yang bertindak sebagai aktivis (*the researcher becomes an activist*).

### Desain Penelitian

Desain penelitian mengacu pada model yang dikembangkan oleh Kurt Lewin (McNiff, 1992), yakni: adanya perencanaan (*planning*), tindakan (*acting*), observasi (*observing*), dan refleksi (*reflecting*) yang dilakukan secara bersiklus.



Gambar 1: Desain penelitian tindakan dalam rangka meningkatkan kinerja penyusunan soal objektif.

## HASIL PENELITIAN

### Siklus I.

#### 1. Perencanaan.

Dalam rangka memecahkan masalah penyusunan soal objektif bentuk pilihan ganda, kegiatan yang dilakukan pada perencanaan di siklus pertama adalah:

- a. memberikan informasi umum tentang evaluasi dan cara-cara menyusun soal yang baik;
- b. meminta guru mempedomani prosedur membuat soal yang sudah disiapkan;
- c. meminta guru menyiapkan SK-KD, Silabus serta membuat kisi-kisi
- d. meminta guru menyusun soal objektif bentuk pilihan ganda (PG);

#### 2. Pelaksanaan Tindakan.

Pelaksanaan tindakan pada siklus I dilaksanakan pada hari Senin tanggal 9 bulan September 2019, peneliti melakukan kegiatan sesuai dengan apa yang telah direncanakan. Adapun kegiatan pada tahap pelaksanaan ini adalah:

- a. Peneliti meminta guru menganalisis soal sebelum diujikan kepada siswa dengan tiga aspek pokok, yaitu apakah soal sesuai materi, konstruk, memenuhi kaidah bahasa, dan di tambah dengan opsi pilihan yang sudah tepat;
- b. melaksanakan ulangan harian (UH) pertama di setiap madrasah penelitian;
- c. memberikan umpan balik terhadap hasil kerja guru, baik proses maupun produknya.

#### 3. Hasil Pengamatan.

Dari aspek proses, fenomena yang teramati selama guru menyusun soal antara lain:

- a. belum sepenuhnya melakukan analisis materi berdasarkan buku teks,
- b. belum mempedomani panduan menyusun soal,
- c. belum mempertimbangkan taraf kesukaran soal (karena soal dibuat asal jadi),
- d. soal langsung diujikan kepada siswa tanpa dianalisis terlebih dulu, dan butir soal cenderung dibuat dalam bentuk pertanyaan kognisi tingkat rendah.

Berdasarkan hasil kesepakatan dan analisis terhadap soal buatan guru, diperoleh hasil sebagai berikut:

- a. Jumlah butir soal yang valid masih rendah. Hanya 31,43% butir soal yang memenuhi syarat validitas soal, sedangkan target minimal adalah 70%.
- b. Soal terlalu mudah karena cenderung berada dalam rentang  $P > 0,67$ , yaitu sebesar 48,57%, sedangkan kecenderungan yang diharapkan adalah pada  $P = 0,33 - 0,67$ .
- c. Soal yang memiliki indek daya pembeda yang baik ( $D \geq 0,30$ ) hanya 37,14%, sedangkan yang diharapkan adalah 70%.
- d. Kinerja guru pada aspek proses memperoleh skor 40%.

#### 4. Refleksi.

Setelah melalui proses refleksi bersama peserta penelitian, diperoleh kesepakatan bahwa untuk meningkatkan mutu soal perlu tindakan yang lebih efektif. *Pertama* memberikan bantuan praktis kepada peserta penelitian, dimana semua peserta (peneliti, guru, dan pengamat) bersama membuat soal untuk UH-2. *Kedua*, membuat soal lebih awal dari waktu pelaksanaan ulangan

harian. Hal ini dimaksudkan supaya tidak tergesa-gesa dalam membuat dan menganalisis soal. *Ketiga*, membuat soal minimal 10 butir soal setiap kali proses pembelajaran berakhir. Setelah empat kali pertemuan, sekurang-kurangnya peserta penelitian telah memiliki sebanyak 20 butir soal untuk ulangan harian kedua.

## **Siklus II.**

### **1. Perencanaan.**

Kegiatan yang dilakukan pada perencanaan di siklus kedua adalah:

- a. memberikan informasi tambahan tentang evaluasi dan cara-cara menyusun soal yang baik;
- b. meminta guru aktif bertanya untuk memperoleh hasil kerja penyusunan soal objektif bentuk pilihan ganda yang semakin baik;
- c. meminta guru memeriksa kembali kisi-kisi yang telah dibuat, apakah sudah tepat dalam menggunakan kata kerja operasional

### **2. Pelaksanaan Tindakan.**

Pada prinsipnya tindakan yang dilakukan untuk meningkatkan mutu soal buatan guru pada siklus kedua sama dengan tindakan siklus pertama. Perubahan yang terjadi pada siklus ini adalah memfokuskan perhatian pada guru bermasalah, dengan cara memberikan umpan balik yang intensif.

### **3. Hasil Pengamatan**

Selain menyiapkan butir soal setiap menyelesaikan proses pembelajaran, kegiatan lain adalah menyesuaikan butir soal dengan materi yang baru saja diberikan di kelas dan mendiskusikan bahasa soal dan opsi pilihan. Pada kegiatan ini butir soal yang dianggap tidak relevan dengan materi dan kurikulum disingkirkan, pilihan yang tidak tepat diganti, dan soal yang tidak mengacu kepada level kognitif yang hendak diukur diperbaiki.

Setelah melalui proses refleksi, diperoleh hasil sebagai berikut:

- 1) Jumlah butir valid meningkat dari 42,86% menjadi 66%;
- 2) Taraf kesukaran soal makin lebih baik di mana  $P=0,33-0,67$  meningkat menjadi 68,57% (untuk lebih jelasnya dapat dilihat pada Gambar 4);
- 3) Soal yang memiliki indek daya pembeda yang baik ( $D \geq 0,30$ ) meningkat dari 51,43% menjadi 60%;
- 4) (Skor kinerja guru membuat soal dari aspek proses meningkat menjadi 91%.

### **4. Refleksi**

Berdasarkan data di atas, dapat disimpulkan bahwa pada siklus II terjadi peningkatan mutu soal yang lebih baik dibandingkan siklus I. Hal ini dapat dilihat dari meningkatnya skor kinerja guru dalam membuat soal menjadi 91. Skor ini lebih besar dari kriteria yang ditargetkan, yaitu 90. Meningkatnya jumlah soal yang valid menjadi 66%. Soal memiliki taraf kesukaran sedang pada rentang  $P=0,33-0,67$  sebesar 68,57%; (4) daya pembeda soal yang baik meningkat menjadi 60%.

### Deskripsi Antar Siklus.

Berdasarkan hasil pelaksanaan tindakan mulai pemantauan keadaan awal hingga pelaksanaan tindakan pada siklus II maka dapat digambarkan seperti dibawah :

No	Indikator	Prosentase yang dicapai		
		Awal	Siklus I	Siklus II
1	Kemampuan dalam menyusun Soal	0 %	40 %	91%
2	Kemampuan mengerjakan lembar kerja		40 %	91 %
3	Keaktifan dalam pembahasan lembar kerja			91 %

### PEMBAHASAN

Dari tabel antar siklus di atas tampak adanya hasil dari masing – masing indikator yang harus dikuasai guru setelah diberi tindakan mengalami peningkatan yang sangat luar biasa.

Pada siklus I peneliti cenderung membantu dalam bentuk teoretis, guru pengamat pasif, karena hampir semua guru belum melaksanakan penyusunan soal berdasarkan kriteria pengembangan tes yang berlaku, bagi guru yang telah membuat soal cenderung dibuat apa adanya. Sedangkan pada siklus II peneliti dan guru benar-benar ikut bersama berlatih menyusun soal objektif bentuk pilihan ganda. Guru diminta untuk lebih aktif dan serius (bukan asal jadi). Setiap selesai proses pembelajaran, masing-masing kelompok guru mempresentasikan hasil kerjanya dan kelompok guru yang lain mengkritisi serta memberi masukan terhadap hasil kerja guru tersebut.

Setelah melalui proses refleksi, sebagian besar guru telah berhasil meningkatkan kompetensinya dalam menyusun soal objektif bentuk pilihan ganda sesuai dengan mata pelajaran yang diampu. Guru dengan teliti dan cekatan memilih memilih cara menetapkan kata kerja operasional untuk menyusun soal objektif bentuk pilihan ganda. Secara umum, pencapaian keberhasilan guru pada siklus keduanya telah mencapai nilai 91 atau berhasil.

### SIMPULAN

Penelitian tindakan ini dilakukan dalam tiga siklus. Masing-masing siklus terdiri dari empat kegiatan, yaitu perencanaan (*planning*), tindakan (*acting*), observasi (*observing*) dan refleksi (*reflecting*). Tujuan utama penelitian adalah untuk memberikan umpan balik (*feed back*) terhadap soal buatan guru, diikuti oleh beberapa kegiatan yang bertujuan untuk meningkatkan kinerja guru membuat soal, baik dari aspek proses maupun produk.

Pada siklus pertama, umpan balik ditindaklanjuti dengan informasi teoritis tentang evaluasi pendidikan pada umumnya dan prosedur menyusun soal objektif bentuk pilihan ganda (*multiple choice*) pada khususnya. Kegiatan selanjutnya meminta peserta menyusun soal yang diujikan kepada siswa dalam pelaksanaan

ulangan harian pertama (UH-1). Untuk membantu guru dalam merancang soal peneliti menyusun panduan membuat soal.

Pada siklus kedua, umpan balik tentang kinerja guru siklus pertama ditindaklanjuti dengan memberikan bantuan praktis, di mana peneliti dan peserta secara kolaboratif menyusun soal untuk diujikan kepada siswa dalam pelaksanaan ulangan harian kedua (UH-2). Model peningkatan kompetensi guru dalam menyusun soal dikembangkan dengan membuat contoh-contoh praktis, model Tabel distribusi, dan bahasa yang lebih efektif dan mudah dipahami.

Hasil penelitian dapat dijelaskan sebagai berikut. Secara proses terjadi peningkatan kinerja guru peserta dari masing-masing siklus. Pada siklus pertama skor kinerja guru adalah 40%, kemudian meningkat menjadi 91% pada siklus kedua. Skor kinerja guru yang diperoleh pada siklus ketiga ini dapat melebihi skor yang ditargetkan (90%).

Dari aspek produk ditemukan bahwa: (1) pada siklus pertama ketercapaian validitas butir rata-rata adalah 11 (31,43%), taraf kesukaran cenderung berada pada  $P \geq 0,67$ , yaitu 17 (48,57%), indeks daya pembeda yang baik  $D \geq 0,30$  adalah 37,14%, dan pengecoh berfungsi 23 (69%); (2) pada siklus kedua, ketercapaian validitas butir soal rata-rata 15 (42,86%), butir soal cenderung berada pada rentangan  $P = 0,33 - 0,67$ , yaitu 23 (65,71%), indeks daya pembeda yang baik  $D \geq 0,30$  menjadi 21 (60%), dan pengecoh berfungsi 83%; (3) pada siklus ketiga, ketercapaian validitas butir rata-rata 23 (66%), soal cenderung pada rentangan  $P = 0,33 - 0,67$  yaitu 24 (68,57%), soal yang memiliki daya beda yang baik  $D \geq 0,30$  meningkat 26 (74,28%).

Data di atas menggambarkan bahwa secara umum terjadi peningkatan mutu produk soal dari siklus pertama, kedua, dan ketiga. Secara khusus, untuk validitas butir soal, peningkatan yang diharapkan terjadi pada siklus ketiga, karena tiga di antara soal buatan guru memiliki jumlah butir valid di atas 70%. Taraf kesukaran soal yang diharapkan di mana  $P = 0,33 - 0,67$  terjadi pada siklus kedua dan ketiga. Indeks daya pembeda soal yang baik juga diperoleh pada siklus ketiga, di mana soal yang memiliki  $D \geq 0,30$  mencapai 74,28%, lebih besar dari kriteria 70% yang ditetapkan.

## SARAN

Saran penelitian antara lain: (1) Dalam membuat soal, guru sebagai evaluator hendaknya selalu menggunakan pedoman penyusunan soal; (2) Diharapkan guru mendiskusikan soal dan memanfaatkan umpan balik dari berbagai pihak, misalnya kepala sekolah dan teman sejawat; (3) Guru hendaknya menganalisis soal sebelum dan sesudah soal diujikan, mengkaji ulang kesalahan (*correction for possible deviation*) sebagai tanda bahwa guru lebih berhati-hati memberikan penilaian terhadap jawaban siswa; (4) Guru yang sudah mampu mengolah dan menganalisis soal berdasarkan kriteria mutu soal yang baik disarankan mempunyai bank soal milik sendiri yang sewaktu-waktu dapat digunakan untuk kepentingan penilaian hasil belajar siswa; (5) Pembuat kebijakan (*decision makers*), seperti kepala sekolah dan kepala dinas pendidikan di daerah, hendaknya memberi kesempatan yang seluas-luasnya kepada guru mata pelajaran untuk mengembangkan kemampuan guru dalam bidang evaluasi, khususnya evaluasi pendidikan.

Selain itu, Kantor Wilayah Departemen Agama sudah semestinya memiliki suatu wadah atau organisasi profesi yang anggotanya terdiri dari para

pakar pendidikan dan ahli evaluasi, khususnya ahli di bidang evaluasi pendidikan yang berfungsi sebagai fasilitator dan *problem solver* bidang evaluasi di daerah. Hasil penelitian menunjukkan bahwa untuk meningkatkan kemampuan guru di bidang evaluasi tidak hanya cukup melalui informasi teori saja (seperti seminar, penataran, dan sejenisnya), tetapi perlu upaya kolaborasi yang lebih konkrit antara pakar atau tenaga ahli dengan guru di lapangan. Melalui wadah keorganisasian atau lembaga evaluasi yang dekat dengan masyarakat guru, kolaborasi evaluasi antar pakar, guru, dan pimpinan sekolah akan berjalan lebih efektif. Wadah ini tentu akan terlaksana melalui kerjasama yang baik dengan instansi terkait, misalnya Pusat Penilaian Pendidikan (Puspendik) Jakarta, terutama dalam membahas isu-isu kritis dalam bidang evaluasi pendidikan. Sedangkan secara akademik dapat dijalin kerjasama yang lebih erat dengan Perguruan Tinggi (PT) lain di Indonesia.

## DAFTAR PUSTAKA

- Gay, L.R. (1996). *Educational Research: Competencies for Analysis and Application*. New Jersey: Englewood Cliffs.
- Gentry, Castelle G. (1994). *Introduction to Instructional Development: Process and Technique*. California: Wadsworth Publishing Company.
- Gordon, Thomas. (1997). *Teacher Effectiveness Training (Menjadi Guru Efektif)*. Jakarta: PT. Gramedia Pustaka Utama.
- Gregory, Robert L. (2000). *Psychological Testing: History, Principle, and Applications*. Needham Height.
- Gronlund, Norman E. (1990). *Measurement and Evaluation in Teaching*. New York: MacMillan Publishing Company.
- Hall, David dan Hall, Irene. (1996). *Practical Social Research: Project Work in the Community*. London: Macmillan Press.
- Heinich, Robert, Molenda M, dan Russel J.D. (1996). *Instructional Media and Technologies for Learning*. New Jersey: Prentice-Hall, Inc.
- Kubiszyn, Tom dan Borich, Gary. (2003). *Educational Testing and Measurement: Classroom Application and Practice*. New York: John Wiley & Son Inc.
- Linn, L Robert dan Gronlund, Norman E. (1995). *Measurement and Assessment in Teaching*. New Jersey: Prentice-Hall Inc.
- McBeath, R.J. (1992). *Instructing and Evaluating in High Education: A guide Book For Planning Learning Out Comes*. New Jersey: Educational Technology Publication, Inc. Englewood Cliffs.
- McNiff, Jean. (1992). *Action Research: Principle and Practice*. New York: Macmillan Education. Ltd.
- Medley, DM., Coker, H., dan Soar, R.S. (1984). *Measurement Based Evaluation of Teacher Performance: An Emperical Approach*. New York & London: Longman Inc.

- Mills, Geoffrey E. (2000). *Action Research: A Guide for the Teacher Researcher*. New Jersey: Prentice Hall.Inc.
- Naga, Dali S. (1992). *Pengantar Teori Skor pada Pengukuran Pendidikan*. Jakarta: Gunadarma.
- Nitko, Anthony J. (1996). *Educational Assessment of Studies*. New Jersey: Prentice-Hall, Inc.
- Popham, James W. (1995). *Classroom Assessment: What Teachers Need to Know*. Boston: Allyn & Bacon.
- Pusat Penilaian Pendidikan. (2007). *Panduan Penulisan Soal Pilihan Ganda*. Balitbang Departemen Pendidikan Nasional.
- Rosset, Allison dan Arwady, Joseph. (1987) *Training Needs Assessment*. New Jersey: Englewood Cliffs.
- Stanley, J.C. dan Hopkins, K.D. (1972). *Educational and Psychological Measurement and Evaluation*. New Jersey: Englewood Cliffs.
- Tessmer, Martin. (1995). *Planning and Conducting Formative Evaluation*. Philadelphia: Kogan Page.
- Toha, Chabib. (1996). *Teknik Evaluasi Pendidikan*. Jakarta : Rajawali Pers.
- Utomo, Tjipto dan Ruijter, Kees. (1985). *Peningkatan dan Pengembangan Pendidikan*. Jakarta: PT. Gramedia.
- Wiersma, William dan Jurs, G. Stephen. (1990). *Educational Measurement and Testing*. Boston: Allyn and Bacon.