

MODEL EVALUASI PROGRAM PENDIDIKAN

Darodjat dan Wahyudhiana M

Dosen Universitas Muhammadiyah Purwokerto

Abstrak

Pembelajaran sebagai suatu sistem tersusun dari unsur-unsur manusiawi, material, fasilitas, perlengkapan, dan prosedur. Menurut Marzano, Pickering, & Tighe (1993: 1-5) ada lima dimensi dalam pembelajaran agar menghasilkan outcomes yang efektif, yaitu: (a) positive attitude & perceptions about learning, (b) acquiring & integrating knowledge, (c) extending & refining knowledge, (d) using knowledge meaningfully, dan (e) productive habits of mind. Pada tahap kelima ini, jika peserta didik yang sudah merasakan bahwa belajar merupakan suatu kebutuhan, mengembangkan apa yang penting bagi kehidupannya, selalu mencari dalil dan bukti terhadap sesuatu, selalu mengevaluasi agar aktivitas menjadi semakin efektif, dan tidak pernah menyerah terhadap problem yang belum dapat diatasinya, maka dia telah menjadi pembelajar yang berhasil. Untuk menentukan tingkat ketercapaian pembelajaran yang telah dicapai oleh peserta didik, maka guru harus melakukan evaluasi pembelajaran.

Secara teknis, ada tiga istilah yang terkait dengan evaluasi pembelajaran, yaitu: pengukuran (measurement), penilaian (assessment), dan evaluasi (evaluation). Kegiatan evaluasi didahului oleh penilaian, kegiatan penilaian didahului oleh pengukuran (measurement). Pengukuran merupakan kegiatan membandingkan hasil pengamatan dengan kriteria, sedangkan penilaian (assessment) merupakan kegiatan menafsirkan dan mendeskripsikan hasil pengukuran, dan evaluasi merupakan penetapan nilai atau implikasi perilaku. Pemahaman terhadap model evaluasi, sangat membantu bagi guru dan evaluator pendidikan, sehingga proses evaluasi dapat dilakukan secara komprehensif, baik menyangkut input, proses, output dan outcomes.

Kata kunci: *model evaluasi, measurement, assessment, evaluation*

Abstract

Learning as a system composed of the elements of human, material, facilities, equipment, and procedures. According to Marzano, Pickering, & Tighe (1993: 1-5) there are five dimensions of learning in order to produce effective outcomes, namely: (a) positive attitude and perceptions about learning, (b) acquiring and integrating knowledge, (c) extending & refining knowledge, (d) using knowledge meaningfully, and (e) productive habits of mind. At this fifth stage, if the learners who already feel that learning is a necessity, developed what is essential for life, always looking for the proposition and proof against something, always evaluating so that the activities become more effective, and never gave up against the problem that can not be overcome, then he has become a successful learner. To determine the level of achievement of learning that has been achieved by learners, the teachers have to do an evaluation of learning.

Technically, there are three terms related to the evaluation of learning, namely: measurements, assessment and evaluation. The evaluation is preceded by assessment, assessment activities preceded by measurements. Measurement is an activity comparing the observations with the criteria, while the assessment is an activity interpret and describe the results of the measurement, and evaluation is the determination of the value or behavioral implications. Understanding of the evaluation model, is very helpful for teachers and educational evaluators, so the evaluation process can be carried out comprehensively, either in relation to inputs, processes, outputs and outcomes.

Keywords: *model of evaluation, measurement, assessment, evaluation*

A. Pendahuluan

Definisi evaluasi yang diajukan para pakar sangat bervariasi, misalnya definisi yang dikemukakan oleh Fitzpatrick, Sanders, & Worthen (2011: 7) evaluasi adalah: "*identification, clarification, and application of defensible criteria to determine an evaluation object's value (worth or merit) in relation to those criteria*". Artinya evaluasi adalah proses identifikasi, klarifikasi, dan penerapan kriteria untuk

menentukan nilai suatu objek evaluasi (nilai/manfaat) berkaitan dengan kriteria tersebut. Sedangkan evaluasi program menurut *Joint Commite*, seperti yang dikutip oleh Brinkerhof (1983: xv) adalah aktivitas investigasi yang sistematis tentang suatu yang berharga dan bernilai dari suatu objek. Gronlund & Linn (1990: 5) menyatakan bahwa evaluasi adalah “*the systematic process of collecting, analyzing, and interpreting information to determine the extent to which pupils are achieving instructional objectives*”. Artinya suatu proses yang sistematis dari pengumpulan, analisis, dan penafsiran data atau informasi untuk menentukan tingkat ketercapaian tujuan pelajaran yang diterima oleh peserta didik.

Berdasarkan pada beberapa definisi tersebut, dapat dipahami bahwa kegiatan evaluasi adalah membandingkan apa yang telah dicapai dari suatu program dengan apa yang seharusnya dicapai berdasarkan standar/kriteria yang telah ditetapkan. Dalam konteks pelaksanaan program, kriteria yang dimaksud adalah kriteria keberhasilan pelaksanaannya, sedangkan hal yang dinilai adalah proses dan hasilnya untuk diambil suatu keputusan. Evaluasi dapat digunakan untuk melihat tingkat keberhasilan program, kemudian diambil suatu keputusan apakah program diteruskan, ditunda, ditingkatkan, dikembangkan, diterima, atau ditolak.

Harlen (2007:12) menjelaskan bahwa istilah yang sering dipakai dalam kegiatan evaluasi pendidikan adalah *assessment* dan *evaluation*, keduanya memiliki arti yang berbeda. Dikatakan: *The terms ‘evaluation’ and assessment in education are sometimes used with different meanings, but also interchangeably. In some countries, including the USA, the term*

'evaluation' is often used to refer to individual student achievement, which in other countries including the UK is described as 'assessment'... 'assessment' refers to the process of collecting evidence and making judgments relating to outcomes, such as students' achievement of particular goals of learning or teacher' and others' understanding.

Lebih jauh Griffin & Nix (1991: 3) menyatakan bahwa: *Measurement, assessment and evaluation are hierarchical. The comparison of observation with the criteria is a measurement, the interpretation and description of the evidence is an assessment and the judgment of the value or implication of the behavior is an evaluation.*

Berdasarkan pada pendapat Griffin & Nix di atas, pengukuran, penilaian dan evaluasi adalah hirarkhis. Kegiatan evaluasi didahului oleh penilaian (*assessment*), sedangkan penilaian didahului oleh pengukuran (*measurement*). Pengukuran merupakan kegiatan membandingkan hasil pengamatan dengan kriteria, sedangkan penilaian (*assessment*) merupakan kegiatan menafsirkan dan mendeskripsikan hasil pengukuran, dan evaluasi merupakan penetapan nilai atau implikasi perilaku.

B. Model-model Evaluasi Program

Terdapat model-model evaluasi program yang dikembangkan oleh para ahli yang dapat dipakai untuk mengevaluasi sebuah program. Model evaluasi merupakan desain evaluasi yang dikembangkan oleh para ahli evaluasi, yang biasanya dinamakan sama dengan pembuatnya atau tahap evaluasinya. Menurut Arikunto & Jabar (2008: 40) meskipun terdapat perbedaan pendapat tentang model-model evaluasi, namun maksudnya

sama yaitu kegiatan pengumpulan data yang berkaitan dengan objek yang dievaluasi sebagai bahan bagi pengambilan keputusan dalam menentukan tindak lanjut suatu program.

Beberapa model yang banyak dipakai untuk mengevaluasi program pendidikan antara lain:

a. Evaluasi Model CIPP

Model evaluasi ini banyak dikenal dan diterapkan oleh para evaluator. Konsep evaluasi model *CIPP (Context, Input, Process and Product)* pertama kali dikenalkan oleh Stufflebeam (1985:153) pada 1965 sebagai hasil usahanya mengevaluasi ESEA (*the Elementary and Secondary Education Act*). Menurut Madaus, Scriven, Stufflebeam (1993: 118), tujuan penting evaluasi model ini adalah untuk memperbaiki, dikatakan: "*the CIPP approach is based on the view that the most important purpose of evaluation is not to prove but to improve*". Evaluasi model Stufflebeam terdiri dari empat dimensi, yaitu: *context, input, process, dan product*, sehingga model evaluasinya diberi nama CIPP. Keempat kata yang disebutkan dalam singkatan CIPP tersebut merupakan sasaran evaluasi, yaitu komponen dan proses sebuah program kegiatan.

1) Evaluasi Konteks (*Context Evaluation*)

Banyak rumusan evaluasi konteks yang dinyatakan oleh para ahli evaluasi, di antaranya adalah Sax (1980: 595). Ia menjelaskan bahwa evaluasi konteks adalah: *Context evaluation is the delineation and specification of project's environment, its unmet needs, the population and sample of individuals to be served, and the project objectives. Context evaluation provides a rationale for justifying a*

particular type of program intervention. Inti dari kutipan di atas yaitu evaluasi konteks adalah kegiatan pengumpulan informasi untuk menentukan tujuan, mendefinisikan lingkungan yang relevan.

Sejalan dengan Sax, Stufflebeam & Shinkfield (1985:169-172) lebih lanjut menjelaskan bahwa evaluasi konteks: *To assess the object's overall status, to identify its deficiencies, to identify the strengths at hand that could be used to remedy the deficiencies, to diagnose problems whose solution would improve the object's well-being, and, in general, to characterize the program's environment. A context evaluation also is aimed at examining whether existing goals and priorities are attuned to the needs of whoever is supposed to be served.*

Inti dari kutipan Stufflebeam & Shinkfield di atas dapat dipahami bahwa evaluasi konteks berusaha mengevaluasi status objek secara keseluruhan, mengidentifikasi kekurangan, kekuatan, mendiagnosa problem, dan memberikan solusinya, menguji apakah tujuan dan prioritas disesuaikan dengan kebutuhan yang akan dilaksanakan.

2) Evaluasi Masukan (*Input Evaluation*)

Menurut Stufflebeam & Shinkfield (1985: 173) orientasi utama evaluasi input adalah menentukan cara bagaimana tujuan program dicapai. Evaluasi masukan dapat membantu mengatur keputusan, menentukan sumber-sumber yang ada, alternatif apa yang diambil, apa rencana dan strategi untuk mencapai tujuan, bagaimana prosedur kerja untuk mencapainya. Komponen evaluasi masukan meliputi: (a) sumber daya manusia (b) sarana dan peralatan pendukung, (c)

dana/anggaran, dan (d) berbagai prosedur dan aturan yang diperlukan.

3) Evaluasi Proses (*Process Evaluation*)

Menurut Stufflebeam & Shinkfield (1985: 173), esensi dari evaluasi proses adalah: mengecek pelaksanaan suatu rencana/program. Tujuannya adalah untuk memberikan *feedback* bagi manajer dan staf tentang seberapa aktivitas program yang berjalan sesuai dengan jadwal, dan menggunakan sumber-sumber yang tersedia secara efisien, memberikan bimbingan untuk memodifikasi rencana agar sesuai dengan yang dibutuhkan, mengevaluasi secara berkala seberapa besar yang terlibat dalam aktifitas program dapat menerima dan melaksanakan peran atau tugasnya. Senada dengan Stufflebeam & Shinkfield, Worthen & Sanders (1981: 137), menjelaskan bahwa evaluasi proses menekankan pada tiga tujuan (1) *do detect or predict in procedural design or its implementation during implementation stage*, (2) *to provide information for programmed decisions*, and (3) *to maintain a record of the procedure as it occurs*.

Evaluasi proses digunakan untuk mendeteksi atau memprediksi rancangan prosedur atau rancangan implementasi selama tahap implementasi, menyediakan informasi untuk keputusan program, dan sebagai rekaman atau arsip prosedur yang telah terjadi. Evaluasi proses meliputi koleksi data penilaian yang telah ditentukan dan diterapkan dalam praktik pelaksanaan program.

4) Evaluasi Hasil (*Product Evaluation*)

Stufflebeam & Shinkfield (1985: 176) menjelaskan bahwa tujuan dari *Product Evaluation* adalah: untuk mengukur, menafsirkan, dan menetapkan pencapaian hasil dari suatu program, memastikan seberapa besar program telah memenuhi kebutuhan suatu kelompok program yang dilayani. Sedangkan menurut Sax (1980: 598), fungsi evaluasi hasil adalah “...to make decision regarding continuation, termination, or modification of program”. Jadi, fungsi evaluasi hasil adalah membantu untuk membuat keputusan yang berkenaan dengan kelanjutan, akhir dan modifikasi program, apa hasil yang telah dicapai, serta apa yang dilakukan setelah program itu berjalan.

Berdasarkan beberapa pendapat di atas, dapat diketahui bahwa evaluasi produk merupakan penilaian yang dilakukan untuk mengukur keberhasilan dalam pencapaian tujuan yang telah ditetapkan. Data yang dihasilkan akan sangat menentukan apakah program diteruskan, dimodifikasi atau dihentikan. Model CIPP saat ini disempurnakan dengan satu komponen O, singkatan dari *outcome*, sehingga menjadi model CIPPO. Bila model CIPP berhenti pada mengukur *output*, sedangkan CIPPO sampai pada implementasi dari *output*.

Dibandingkan dengan model-model evaluasi yang lain, model CIPP memiliki beberapa kelebihan antara lain: lebih komprehensif, karena objek evaluasi tidak hanya pada hasil semata tetapi juga mencakup konteks, masukan (*input*), proses, maupun hasil. Selain memiliki kelebihan, model CIPP juga memiliki keterbatasan, antara lain penerapan model ini dalam bidang program pembelajaran di

kelas perlu disesuaikan atau modifikasi agar dapat terlaksana dengan baik. Sebab untuk mengukur konteks, masukan maupun hasil dalam arti yang luas banyak melibatkan pihak, membutuhkan dana yang banyak dan waktu yang lama.

b. Evaluasi Model Provus (*Discrepancy Model*)

Kata *discrepancy* berarti kesenjangan, model ini menurut Madaus, Sriven & Stufflebeam (1993: 79-99) berangkat dari asumsi bahwa untuk mengetahui kelayakan suatu program, evaluator dapat membandingkan antara apa yang seharusnya diharapkan terjadi (*standard*) dengan apa yang sebenarnya terjadi (*performance*). Dengan membandingkan kedua hal tersebut, maka dapat diketahui ada tidaknya kesenjangan (*discrepancy*), yaitu standar yang ditetapkan dengan kinerja yang sesungguhnya. Model ini dikembangkan oleh Malcolm Provus, bertujuan untuk menganalisis suatu program apakah program tersebut layak diteruskan, ditingkatkan, atau dihentikan.

Model ini menekankan pada terumuskannya *standard*, *performance*, dan *discrepancy* secara rinci dan terukur. Evaluasi program yang dilaksanakan oleh evaluator mengukur besarnya kesenjangan yang ada di setiap komponen program. Dengan adanya penjabaran kesenjangan pada setiap komponen program, maka langkah-langkah perbaikan dapat dilakukan secara jelas.

c. Evaluasi Model Stake (*Countenance Model*)

Model ini dikembangkan oleh Robert E. Stake dari University of Illinois. Menurut Worthen & Sanders (1981: 113), Stake menekankan

adanya dua dasar kegiatan dalam evaluasi, yaitu *description* dan *judgment*, dan membedakan adanya tiga tahap, yaitu: *antecedent (context)*, *transaction/process*, dan *outcomes*. Deskripsi menyangkut dua hal yang menunjukkan posisi sesuatu yang menjadi sasaran evaluasi, yaitu: apa tujuan yang diharapkan oleh program, dan apa yang sesungguhnya terjadi. Evaluator menunjukkan langkah pertimbangan yang mengacu pada standar.

Stufflebeam & Shinkfield (1985: 217-219) menjelaskan tiga tahap evaluasi program model Stake, yaitu: *antecedents*, *transaction*, dan *outcomes*. *Antecedents* mengacu pada informasi dasar yang terkait, kondisi/kejadian apa yang ada sebelum implementasi program. Menurut Stake, informasi pada tipe ini misalnya, terkait dengan kegiatan belajar mengajar sebelumnya, dan terkait dengan *outcome*, seperti: apakah siswa telah makan pagi sebelum datang ke sekolah, apakah siswa telah menyelesaikan pekerjaan rumahnya, apakah siswa tidur malam dengan cukup. Untuk mendeskripsikan secara lengkap dan menetapkan sebuah program atau pembelajaran pada suatu waktu. Stake mengusulkan bahwa evaluator harus mengidentifikasi dan menganalisis kondisi yang berhubungan dengan *antecedent*.

Pada tahap *transactions*, apakah yang sebenarnya terjadi selama program dilaksanakan, apakah program yang sedang dilaksanakan itu sesuai dengan rencana program. Termasuk tahap ini adalah informasi yang dialami oleh peserta didik berkaitan dengan guru, orang tua, konselor, tutor, dan peserta didik lainnya. Stake menganjurkan kepada evaluator agar bertindak secara bijak dalam proses pelaksanaan evaluasi, sehingga dapat melihat aktualisasi

program. Sedangkan *outcomes*, berkaitan dengan apa yang dicapai dengan program tersebut, apakah program itu dilaksanakan sesuai dengan yang diharapkan termasuk di dalamnya: kemampuan, prestasi, sikap dan tujuan.

d. Evaluasi Model Kirkpatrick

Model evaluasi yang dikembangkan oleh Kirkpatrick ini telah mengalami beberapa penyempurnaan, terakhir diperbarui tahun 1998 yang dikenal dengan *Evaluating Training Programs: the Four Levels* atau *Kirkpatrick's evaluation model*. Evaluasi terhadap program pelatihan mencakup empat level evaluasi, yaitu: (a) *reaction*, (b) *learning*, (c) *behavior*, dan (d) *result*.

1) Evaluasi Reaksi (*Reaction Evaluation*)

Catalanello & Kirkpatrick (1968: 2-9) menjelaskan bahwa evaluasi terhadap reaksi peserta pelatihan berarti mengukur kepuasan peserta. Program pelatihan dianggap efektif apabila proses pelatihan dirasa menyenangkan peserta, sehingga mereka tertarik dan termotivasi untuk belajar dan berlatih. Sebaliknya, apabila peserta tidak merasa puas terhadap proses pelatihan yang diikutinya, maka mereka tidak akan termotivasi untuk mengikuti pelatihan lebih lanjut. Keberhasilan proses kegiatan pelatihan tidak terlepas dari minat, perhatian, dan motivasi peserta dalam mengikuti jalannya kegiatan ini. Orang akan belajar lebih baik manakala mereka memberi reaksi positif terhadap lingkungan belajar. Kepuasan peserta dapat dikaji dari beberapa aspek, yaitu materi yang diberikan; fasilitas yang tersedia; strategi penyampaian materi yang digunakan,

media pembelajaran; jadwal kegiatan, sampai menu dan penyajian konsumsi yang disediakan.

Instrumen untuk mengukur reaksi antara lain dengan *reaction sheet* dalam bentuk angket. Menurut Kirkpatrick (2008: 26) dalam menentukan instrumen tersebut dapat digunakan prinsip mampu mengungkap informasi sebanyak mungkin; tetapi dalam pengisiannya seefisien mungkin. Evaluasi pada level ini difokuskan pada reaksi peserta yang terjadi pada saat kegiatan dilakukan, disebut juga sebagai evaluasi proses pelatihan.

2) Evaluasi Belajar (*Learning Evaluation*)

Menurut Kirkpatrick & Kirkpatrick (2008: 42) evaluasi hasil belajar dapat dilihat pada perubahan sikap, perbaikan pengetahuan, dan atau peningkatan keterampilan peserta setelah selesai mengikuti program. Peserta program dikatakan telah belajar apabila pada dirinya telah mengalami perubahan sikap, perbaikan pengetahuan maupun peningkatan keterampilan. Untuk mengukur efektivitas program maka ketiga aspek tersebut perlu untuk diukur. Tanpa adanya perubahan sikap, peningkatan pengetahuan maupun perbaikan keterampilan pada peserta *training* maka program dapat dikatakan gagal. Penilaian ini ada yang menyebut dengan penilaian hasil (*output*) belajar. Oleh karena itu, dalam pengukuran hasil belajar harus menentukan: (a) pengetahuan apa yang telah dipelajari; b) perubahan sikap apa yang telah dilakukan; c) keterampilan apa yang telah dikembangkan atau diperbaiki.

Mengukur hasil belajar membutuhkan waktu yang lama jika dibandingkan dengan mengukur reaksi. Mengukur reaksi dapat

dilakukan dengan *reaction sheet* dalam bentuk angket. Menurut Kirkpatrick & Kirkpatrick (2008: 42) penilaian terhadap hasil belajar dapat dilakukan dengan dengan kelompok pembandingan. Kelompok yang ikut pelatihan dan kelompok yang tidak ikut pelatihan perkembangannya diperbandingkan dalam periode waktu tertentu. Di samping itu, penilaian terhadap hasil belajar dapat juga dilakukan dengan membandingkan hasil *pre test* dengan *post test*, tes tertulis maupun tes kinerja.

3) Evaluasi Perilaku (*Behavior Evaluation*)

Penilaian difokuskan pada perubahan tingkah laku setelah peserta kembali ke tempat kerja, disebut juga evaluasi terhadap *outcomes* dan kegiatan pelatihan. Perubahan apa yang terjadi di tempat kerja setelah peserta mengikuti program tersebut, baik menyangkut pengetahuan, sikap maupun keterampilannya. Menurut Kirkpatrick & Kirkpatrick (2008: 53), evaluasi perilaku dapat dilakukan dengan: (1) membandingkan perilaku kelompok kontrol dengan perilaku peserta program, (2) membandingkan perilaku sebelum dan sesudah mengikuti program maupun, (3) survei/interviu dengan pelatih, atasan maupun bawahan peserta program setelah kembali ke tempat kerja.

4) Evaluasi Hasil (*Result Evaluation*)

Evaluasi pada tahap ini difokuskan pada hasil akhir yang terjadi karena peserta telah mengikuti suatu program. Beberapa contoh dari hasil akhir dalam konteks perusahaan antara lain: kenaikan produksi, peningkatan kualitas, penurunan biaya, penurunan kecelakaan kerja, kenaikan keuntungan. Cara melakukan evaluasi hasil akhir menurut

Kirkpatrick & Kirkpatrick (2008: 63) adalah dengan: (1) membandingkan kelompok kontrol dengan kelompok peserta program, (2) mengukur kinerja sebelum dan setelah mengikuti pelatihan, (3) membandingkan biaya yang digunakan dengan keuntungan yang didapat setelah dilakukan pelatihan, dan bagaimana peningkatannya.

Evaluasi program model Kirkpatrick dapat diterapkan dalam program pembelajaran di sekolah, karena: (1) fokusnya sama, yaitu diarahkan pada proses dan hasil belajar dengan mengikuti suatu program, (b) perubahan pembelajaran pada empat level sama-sama diarahkan pada aspek pengetahuan, sikap, dan kecakapan. Namun demikian, penerapan evaluasi model ini dalam program pembelajaran perlu dimodifikasi dengan *setting* sekolah.

Pertama, evaluasi terhadap *outcome* maupun *impact* kegiatan pembelajaran di kelas sulit untuk dilakukan, karena sekolah sulit memonitor sejauhmana peserta didik mampu mengaplikasikan pengetahuan maupun kecakapan yang diperolehnya dalam kegiatan pembelajaran di sekolah, maupun di tengah masyarakat dalam waktu tertentu. Sebab untuk menjangkau pada level ini membutuhkan waktu yang lama, tenaga dan biaya yang besar, terlebih lagi dilanjutkan pada evaluasi dampak. *Kedua*, fokus program pembelajaran pada *setting* sekolah dapat diarahkan pada kompetensi yang telah ditentukan.

Menurut Holton (1996: 5), Praslova (2010: 215), kekuatan dari model ini adalah kesederhanaan model, kemampuannya membantu memperjelas kriteria, dan membuat indikator penilaian. Dengan

adanya kejelasan kriteria dan indikator yang sudah ditetapkan, maka capaian suatu program akan dapat diukur dengan baik. Model ini dapat diterapkan untuk mengevaluasi program pembelajaran di sekolah, bahkan pada level yang lebih kecil, misalnya kelas dan suatu program tertentu.

Model ini juga memiliki beberapa kelemahan, jika diterapkan dalam *setting* sekolah. Oleh karena itu, harus ada penyesuaian dan modifikasi, sehingga tujuan evaluasi program suatu sekolah dapat tercapai dengan penggunaan model ini. Menurut Bates (2004:341-347), Alliger & Janak (1989: 331-333), model ini terlalu menyederhanakan efektivitas pelatihan, karena tidak mempertimbangkan individu atau pengaruh kontekstual dalam evaluasi program. Padahal karakteristik organisasi, lingkungan kerja/sekolah, dan karakteristik individu peserta pelatihan--sebagai masukan penting dari *input*--turut mempengaruhi efektivitas proses dan hasil pelatihan. Sedangkan model Kirkpatrick ini secara implisit mengasumsikan bahwa pemeriksaan faktor-faktor ini tidak penting bagi evaluasi program yang efektif.

e. Evaluasi Model Brinkerhoff

Brinkerhoff, et.al., (1983: 37) mengemukakan tiga pendekatan evaluasi yang disusun berdasarkan penggabungan elemen-elemen yang sama, yaitu:

(1) Fixed vs Emergent Evaluation Design.

Desain evaluasi yang baik ditentukan dan direncanakan secara sistematis sebelum implementasi dikerjakan. Desain dikembangkan

berdasarkan tujuan program disertai seperangkat pertanyaan yang akan dijawab dengan informasi yang akan diperoleh dari sumber-sumber tertentu. Rencana analisis dibuat sebelumnya yang pemakainya akan menerima informasi seperti yang telah ditentukan dalam tujuan. Desain ini dapat disesuaikan dengan kebutuhan yang mungkin berubah.

(2) *Formative vs Sumative Evaluation*

Evaluasi formatif digunakan untuk memperoleh informasi yang dapat membantu memperbaiki program, dilaksanakan pada saat implementasi program sedang berjalan. Fokus evaluasi berkisar pada kebutuhan yang telah dirumuskan oleh evaluator. Evaluasi sumatif dilaksanakan untuk menilai manfaat suatu program, dari hasil evaluasi ini dapat ditentukan apakah suatu program tertentu akan diteruskan atau dihentikan. Pada evaluasi sumatif difokuskan pada variabel yang dianggap penting bagi pembuat keputusan. Waktu pelaksanaan evaluasi sumatif pada akhir pelaksanaan program.

(3) *Experimental & Quasi-Experimental Designs vs. Unobtrusive Inquiry.*

Beberapa evaluasi memakai metodologi penelitian klasik. Dalam hal seperti ini subjek penelitian diacak, perlakuan diberikan dan pengukuran dampak dilakukan. Tujuan dari penelitian untuk menilai manfaat suatu program yang dicobakan. Apabila siswa atau program dipilih secara acak, maka generalisasi dibuat pada populasi yang agak lebih luas. Dalam beberapa hal intervensi tidak mungkin dilakukan atau tidak dikehendaki. Apabila proses sudah diperbaiki,

evaluators harus melihat dokumen-dokumen, seperti mempelajari nilai tes atau menganalisis penelitian yang dilakukan dan sebagainya. Strategi pengumpulan data terutama menggunakan instrumen formal seperti tes, survey, kuesioner serta memakai metode penelitian yang terstandar.

Selain berbagai model di atas, Nana Sudjana & Ibrahim (2004: 234) mengelompokkan model-model evaluasi pendidikan menjadi empat kelompok berdasarkan perkembangannya, yaitu: *measurement model*, *congruence model*, *educational system evaluation model* dan *illuminative model*.

f. Measurement Model

Model ini dapat dipandang sebagai model yang tertua di dalam sejarah penilaian dan lebih banyak dikenal di dalam proses penilaian pendidikan. Tokoh-tokoh penilaian yang dipandang sebagai pengembang model ini adalah R. Thorndike dan R.I. Ebel. Sesuai dengan namanya, model ini sangat menitikberatkan peranan kegiatan pengukuran di dalam melaksanakan proses evaluasi. Pengukuran dipandang sebagai suatu kegiatan yang ilmiah dan dapat diterapkan dalam berbagai bidang persoalan termasuk ke dalam bidang pendidikan. Pengukuran, menurut model ini tidak dapat dilepaskan dari pengertian kuantitas atau jumlah. Jumlah ini akan menunjukkan besarnya (*magnitude*) objek, orang ataupun peristiwa sehingga dengan demikian hasil pengukuran itu selalu dinyatakan dalam bentuk bilangan.

Pengukuran dengan demikian dipandang sebagai kegiatan menentukan besarnya suatu sifat (*attribute*) tertentu yang dimiliki objek,

orang, dan peristiwa dalam bentuk unit ukuran tertentu. Dalam bidang pendidikan, model ini telah diterapkan dalam proses penilaian untuk melihat dan mengungkapkan perbedaan-perbedaan individual maupun perbedaan-perbedaan kelompok dalam hal kemampuan serta minat dan sikap. Hasil pengukuran mengenai aspek-aspek tingkah laku di atas digunakan untuk keperluan seleksi siswa, bimbingan, dan perencanaan pendidikan bagi siswa itu sendiri.

g. Congruence Model

Model yang kedua ini dipandang sebagai reaksi terhadap model yang pertama, sekalipun dalam beberapa hal masih menunjukkan adanya persamaan dengan model yang pertama. Tokoh-tokoh evaluasi yang merupakan pengembangan model ini antara lain W. Tyler, John B. Carrol, dan Lee J. Cronbach. Tyler menggambarkan pendidikan sebagai suatu proses yang di dalamnya terdapat tiga hal yaitu: tujuan pendidikan, pengalaman belajar, dan penilaian terhadap hasil belajar. Kegiatan evaluasi dimaksudkan sebagai kegiatan untuk melihat sejauh mana tujuan-tujuan pendidikan telah dapat dicapai siswa dalam bentuk hasil belajar yang mereka perlihatkan pada akhir kegiatan pendidikan. Mengingat tujuan pendidikan mencerminkan perubahan-perubahan tingkah laku yang diinginkan pada anak didik, maka yang penting dalam proses evaluasi adalah memeriksa sejauhmana perubahan-perubahan tingkah laku yang diinginkan itu telah terjadi pada anak didik.

Dengan diperolehnya informasi tentang pencapaian tujuan pendidikan yang telah dicapai oleh siswa, baik secara individual

maupun secara kelompok, dapat diambil keputusan tentang tindakan-tindakan apa yang perlu diambil. Tindak lanjut hasil evaluasi yang menyangkut kepentingan siswa tersebut, misalnya: memberikan layanan atau bimbingan untuk memperbaiki hasil yang telah dicapai, memberikan pengayaan materi, dan merencanakan program lain bagi masing-masing siswa. Ditinjau dari kepentingan sistem pendidikan, hasil evaluasi dimaksudkan sebagai umpan balik untuk kebutuhan memperbaiki bagian-bagian sistem yang masih lemah.

Selain untuk kepentingan bimbingan siswa dan perbaikan sistem, evaluasi ini dimaksudkan pula untuk memberikan informasi kepada pihak-pihak di luar pendidikan tentang sejauh mana tujuan-tujuan yang diinginkan itu telah dapat dicapai oleh sistem pendidikan yang ada (Nana Sudjana & Ibrahim, 2004: 239). Secara singkat dapat dikatakan bahwa model evaluasi berusaha memeriksa persesuaian (*congruence*) antara tujuan-tujuan pendidikan yang diinginkan dengan hasil belajar yang telah dicapai.

h. *Illuminative Model*

Model illuminatif ini lebih menekankan pada penilaian kualitatif. Tujuan evaluasi model ini adalah mengadakan studi yang cermat terhadap sistem maupun program yang bersangkutan, yang meliputi: (1) bagaimana implementasi program di lapangan, (2) bagaimana implementasi dipengaruhi oleh situasi sekolah tempat program yang bersangkutan dikembangkan, (3) apa kebaikan-kebaikan dan kelemahan-kelemahannya dan bagaimana program tersebut mempengaruhi pengalaman-pengalaman belajar para siswa. Hasil

evaluasi yang dilaporkan bersifat deskripsi dan interpretasi, bukan pengukuran dan prediksi. Oleh karena itu, dalam pelaksanaan evaluasi model yang keempat ini lebih banyak menekankan pada penggunaan *judgement*.

Perbedaan penelitian dengan penelitian evaluasi adalah adanya kriteria pada penelitian evaluasi. Berdasarkan kriteria, peneliti pada penelitian evaluasi memberikan nilai terhadap objek yang ditelitinya. Menilai kriteria keefektifan suatu model evaluasi program tidak dapat dilepaskan dari tujuan/fungsi evaluasi program. Evaluasi program mempunyai fungsi menyediakan informasi yang digunakan untuk membantu pembuatan keputusan/penyusunan kebijakan maupun penyusunan program selanjutnya. Agar keputusan yang dihasilkan merupakan keputusan yang baik, maka dibutuhkan informasi yang lengkap, akurat, dan dapat dipercaya (*valid*, dan *reliable*) serta tepat waktu (*timely*). Informasi yang lengkap mempunyai makna bahwa informasi yang dihasilkan dari evaluasi mencakup komponen-komponen program secara lengkap. Informasi yang akurat mempunyai makna bahwa informasi yang dihasilkan dari evaluasi merupakan informasi yang tepat menggambarkan keadaan yang sebenarnya dari objek evaluasi dan dapat dipercaya. Untuk mendapatkan informasi yang akurat dibutuhkan instrumen pengumpulan data yang *valid* dan *reliable*.

Informasi yang tepat waktu mempunyai makna bahwa informasi yang diperoleh dari hasil evaluasi dapat disampaikan kepada pihak-pihak yang membutuhkan, untuk mengambil keputusan, menyusun kebijakan maupun menyusun program selanjutnya. Syarat

ketepatan waktu ini berkaitan dengan kepraktisan dalam pengumpulan, pengolahan, dan penyajian/pelaporan informasi. Hal ini membutuhkan panduan evaluasi yang sederhana dan mudah dipahami oleh pemakai model evaluasi, sehingga proses evaluasi dapat berlangsung lebih cepat tanpa mengabaikan kelengkapan dan keakuratan informasi.

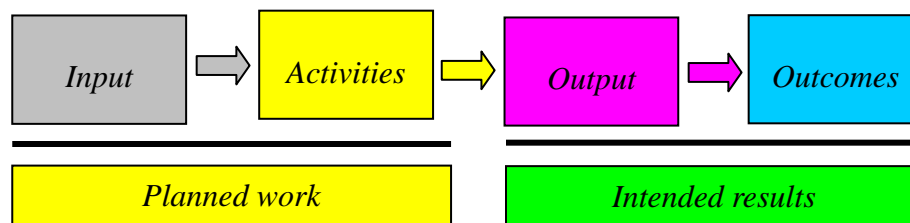
Jadi model evaluasi program yang baik adalah: (1) bersifat komprehensif, menyangkut semua komponen/subkomponen program, baik input, proses, *output*, dan *outcome*, (2) praktis, yaitu mudah dalam penggunaan dan pengelolaan, (3) ekonomis, yaitu membutuhkan biaya relatif sedikit, demikian halnya dengan waktu dan tenaga, (4) instrumen pengumpulan data valid dan reliabel.

i. Model Logik (*Logic Model*)

Model logik adalah suatu penggambaran program yang logis dan tepat menurut kondisi tertentu dalam rangka memecahkan problem. Pada umumnya bentuk penggambaran menggunakan diagram alur yang menjelaskan aktivitas yang direncanakan dan *outcome* yang diharapkan dari model evaluasi ini (Bickman, 1987; Dwyer, 1997; McLaughlin, & Jordan, 1999). Senada dengan pendapat di atas, W.K. Kellogg Foundation (2004: 1) menjelaskan bahwa: *basically, a logic model is a systematic and visual way to present and share your understanding of the relationships among the resources you have to operate program, the activities you plan, and the changes or results you hope to achieve.*

Kekhasan dari model logik adalah penggunaan tabel dan grafik alir yang berisi *input*, aktivitas, dan hasil. Sebagian besar menggunakan teks dan anak panah atau grafik untuk menggambarkan urutan aktivitas

untuk menghasilkan perubahan, dan bagaimana aktivitas tersebut terhubung dengan hasil program yang diharapkan tercapai. Dibutuhkan keputusan yang tepat sebelum menggunakan model logik (W.K. Kellogg Foundation, 2004: 9), karena penyusunan model logik merupakan hal yang kompleks dan menyangkut satu dari tiga pendekatan, yaitu: pendekatan model (*conceptual*), pendekatan *outcomes*, dan pendekatan aktivitas (*applied*) atau merupakan campuran dari beberapa tipe di atas. Secara sederhana, model logik dapat digambarkan sebagai berikut:



Gambar 1

Model Logik Sederhana

Sumber: W.K. Kellogg Foundation (2004: 4)

Model logik telah banyak digunakan dalam berbagai bidang semenjak tahun 1980-an dan awal 1990-an. Sebagai contoh, model logik telah digunakan untuk menggambarkan program di bidang pendidikan, kesehatan, pembangunan internasional, kerja sosial, pelayanan sosial, serta bidang-bidang lain. Elemen model logik yang penting menurut United Way of America (1996) terdiri dari tiga bagian, yaitu: *inputs*, *outputs (activities and participants or methodology)*, and *outcomes*. *Inputs* berkaitan dengan sumber-sumber penting yang akan ditanamkan dalam

program (*what we invest*), *outputs* berkaitan dengan aktivitas apa yang dijalankan (*what we do, and who we reach*) dan *outcome* berkaitan dengan pengaruh atau perubahan yang diinginkan dengan adanya program yang dijalankan. Agar model logik lebih fokus, maka perlu dibuat cakupannya, misalnya menyangkut jangka waktu program dilaksanakan, jangka waktu *output* dan *outcomes* yang dikehendaki, dan bentuk perubahan yang diinginkan

1) Manfaat Penggunaan Evaluasi Model Logik

Para ahli di bidang evaluasi setuju bahwa penerapan model logik merupakan cara yang efektif untuk menjamin agar program bisa sukses. Evaluasi menggunakan model ini memberikan banyak manfaat, sebagaimana dijelaskan oleh James Bell Associates (2007), dan WK Kellogg Foundation (2004: 5), beberapa manfaat tersebut antara lain:

- a. Membantu menyusun rencana kerja/peta evaluasi dan *outcomes* yang diharapkan.
- b. Membantu memastikan pemahaman yang jelas tentang layanan apa yang sedang dilaksanakan, apa harapan yang akan dicapai, dan bagaimana mengukur keberhasilan program.
- c. Membantu menjelaskan mengapa berbagai data dikumpulkan dalam evaluasi dan bagaimana data akan digunakan.
- d. Model Logik membantu membangun konsensus di antara evaluator, pengawas, *stakeholder* terhadap *outcome* yang sesuai, memberikan kesempatan kepada *stakeholder* secara bersama-sama mengevaluasi kemungkinan terjadinya perubahan pengukuran pada *outcomes* program yang terpilih.
- e. Model logik yang digambarkan secara visual menjadi mudah

- dipahami, dapat berfungsi sebagai kunci rencana kerja, dan dapat disebarluaskan kepada pihak ketiga yang berkepentingan untuk memberikan ringkasan tujuan program.
- f. Model logik dapat mengidentifikasi kesenjangan yang logis/inkonsistensi antara kegiatan program dan hasil yang diharapkan, dan untuk menilai kelayakan serta kepraktisan pengukurannya.
 - g. Model logik dapat berfungsi sebagai "titik referensi" untuk memodifikasi program, yaitu membandingkan perubahan yang diusulkan dengan model logika yang asli, menentukan apakah perubahan sedang dilakukan terhadap unsur-unsur inti dari program tersebut.
 - h. Model logik berfungsi sebagai alat pengawas program dan membantu mengidentifikasi pertanyaan kunci: apakah komponen program kunci telah dilaksanakan?, apa *output* program sampai saat ini?, apakah data yang relevan telah dikumpulkan? apa hasil yang telah dicapai sampai saat ini?
 - i. Memudahkan dalam membuat perbandingan penjelasan program dengan mengidentifikasi kesamaan dan perbedaan pada intervensi program, mengidentifikasi *outcomes* secara umum, indikator umum, alat pengukuran, dan sumber data.

2) Kelemahan Evaluasi Model Logik

Penerapan model logik dalam bidang evaluasi mengandung sisi kelemahan/kekurangan. Powel & Heneret (2008: 10) menemukan enam hal kelemahan model logik, yaitu: (1) *a logic model represents intention, it is not reality*, (2) *it focuses on expected outcomes so people may*

overlook unintended outcomes (positive and negative), (3) it focuses on positive change—change isn't always positive, (4) it may simplify the complex nature of causal attribution where many factors influence process and outcomes, (5) it doesn't address whether we are doing the right thing—we may get caught up in creating a logic model and lose track of whether the program is the right thing to do, and (6) may stifle creativity and spontaneity.

Saran-saran berikut penting diperhatikan untuk mengatasi kelemahan yang ada pada model tersebut ketika diterapkan, yaitu: (1) model harus disesuaikan dengan tujuan, kegunaan dan level program yang diinginkan, kemudian diwujudkan dalam bentuk gambar (*flow chart*) secara logis dengan hubungan relasional yang dikenal dengan *if-then relationship*. Misalnya, jika tersedia sumber daya program (*input*), maka kegiatan program dapat dilaksanakan (*activities*), jika kegiatan program dilaksanakan dengan sukses, maka *output* atau hasil dapat diharapkan, (2) model ini harus dimengerti oleh para pengguna secara mendalam (Powel & Heneret (2008) sebelum diterapkan. Porteous, et.al., (2002: 116-117) membuat kata kunci (agar mudah diingat) dalam pemahaman model logik dengan singkatan "CAT SOLO". Secara berturut-turut C adalah singkatan dari *component*, A = *Activities*, T = *Target group*, S = *Short-term Outcomes*, L = *Long-term Outcomes*. Dengan kata kunci ini maka dapat disusun/diketahui komponen, aktivitas, dan cakupan *outcome* yang diharapkan.

Daftar Pustaka

- Alliger, G.M. & Janak, E.A. (1989). Kirkpatrick's levels of training criteria: Thirty Years Later. *Personnel Psychology*, 42, 331-342
- Arikunto, S., & Jabar, C.S.A. (2008). *Evaluasi program pendidikan*. Jakarta: Bumi Aksara.
- Bates, R. (2004). A critical analysis of evaluation practice: The Kirkpatrick model and the principle of beneficence. *Journal of Evaluation and Program Planning*, 27, 341–347.
- Bickman, L. (Ed.). (1987). The functions of program theory. *New Directions in Program Evaluation: Using Program Theory in Evaluation*.
- Brinkerhoff, R.O., et.al, (1983). *Program evaluation: A practitioner's guide for trainers and educators*. Western Michigan: Kluwer-Nijhoff.
- Catalanello, & Kirkpatrick, D.L. (1968). Evaluation training programs the state of the art. *Training and Development Journal*, 22 , 2-9.
- Dwyer, J., & Makin, S. (1997). Using a program logic model that focuses on performance measurement to develop a program. *Canadian Journal of Public Health*, 88, 421–425.
- Griffin, P., & Nix, P. (1991). *Educational assessment and reporting*. Sydney: Harcourt Brace Javanovich, Publisher.
- Gronlund, N. E., & Robert, L. L. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan
- Harlen, W. (2007). *Assessment of learning*. London: Sage Publication.
- Holton, E. F. (1996). The flawed four-level evaluation model. *Human Resource Development Quarterly*, (7), 5-21.
- James Bell Associates. (2007). *Evaluation brief: Developing a logic model*. Arlington, VA. August 2007.

- Kellogg, W.K. Foundation. (2004). *Using logic models to bring planning, evaluation, and action*. Michigan: WK Kellogg Foundation. Diambil pada tanggal 10 Juli 2012, dari <http://www.wkkf.org>.
- Kirkpatrick, D.L. & Kirkpatrick, J.D. (2008). *Evaluating training programs, the four levels (3rd ed)*. San Francisco: Berrett-Koehler Publisher, Inc.
- Madaus, G.F., Scriven, M.S., & Stufflebeam, D.L. (1993). *Evaluation models, viewpoints on educational and human services evaluation*. Boston: Kluwer-Nijhoff Publishing.
- McLaughlin, J. A., and Jordan, G. B. (1999). Logic models: A tool for telling your programs Performance Story. *Evaluation and Program Planning*, 22, 65-72.
- Porteous, N.L., et.al, (2002). Introducing program team to logic models: Facilitating the learning process. *Canadian Journal of Program Evaluation*. Vol. 17, No. 3. Pages. 113-141.
- Powell, T., & Heneret, E. (2008). Enhancing program performance with logic models. Diakses tanggal 15 November 2011 dari Wisconsin Extension Website: <http://www.UWEX/edu.ces/pdande/>
- Praslova, L. (2010). Adaptation of Kirkpatrick's four level model of training criteria to assessment of learning outcomes and program evaluation in higher education. *Educational Assessment, Evaluation and Accountability*, (22), 215-225.
- Sax, G. (1980). *Principles of educational and psychological measurement and evaluation, (2nd ed.)*. California: Wadsworth Publishing Company.
- Stufflebeam, D.L., & Shinfield, A.J. (1985). *Systematic evaluation*. Boston: Kluwer Nijhof Publishing.
- Stufflebeam, D.L., & Shinfield, A.J. (1985). *Systematic evaluation*. Boston: Kluwer Nijhof Publishing.

Sudjana, N., & Ibrahim. (2004). *Penelitian dan penilaian pendidikan*. Bandung: Sinar Baru Algesindo.

United Way of America. (1996). *Measuring program outcomes: A practical approach*. Alexandria, VA: Author.

Worthen, B.R., & Sanders, J.R. (1981). *Educational evaluation: Theory and practice*. Ohio: Charles A. Jones Publishing Company.