

Sejarah Penilaian Hasil Belajar; Urgensi dan Perkembangan paradigma

Oleh:Raikhan

Institut Agama Islam Tarbiyatut Tholabah,
Lamongan Indonesia
Email: reihan.lmg@gmail.com

Abstract : *This paper tries to describe the problems of assessment of learning outcomes; Among the complicated problems in education are outlining the paradigm of assessment, assessment still considered a social formality of the relationship between teachers and institutions, as manager that assessment formative is only for the sake of accountability, so that the assessment process does not meet the assessment standards. Curriculum changes are also not widely understood by the community as educational dynamics, but more as mere political projects, even changes in terms in evaluation are still not understood as paradigm shifts, but are understood only to make a difference. This paper uses an explanatory descriptive approach, this paper belongs to the type of library research. Library research is research conducted by collecting data, information, and various other data contained in the literature related to the science of evaluation and its history. From this research, it shows that globally there are seven periods and this is the same as what happened in Indonesia which experienced seven changes in assessment policies and many changes were also influenced by changes in the assessment paradigm, until in the end it can be concluded that the Assessment as Learning paradigm is a paradigm that is in accordance with the concept of muhasabah in Islam.*

Keywords: *History of evaluation, assessment, paradigm, muhasabah*

Abstrak: Tulisan ini mencoba mendeskripsikan permasalahan penilaian hasil belajar; Di antara peliknya problematika pendidikan adalah penjabaran paradigma penilaian, penilaian masih dianggap formalitas sosial hubungan antara guru dan lembaga, sebagai pengelola penilaian formatif hanya untuk kepentingan akuntabilitas, sehingga proses penilaian tidak memenuhi standar penilaian. Perubahan kurikulum juga tidak dipahami secara luas oleh masyarakat sebagai dinamika pendidikan, tetapi lebih sebagai proyek politik belaka, bahkan perubahan dalam hal evaluasi masih belum dipahami sebagai pergeseran paradigma, tetapi dipahami hanya untuk membuat perbedaan. Makalah ini menggunakan pendekatan deskriptif eksplanatori, tulisan ini termasuk dalam jenis penelitian kepustakaan. Penelitian kepustakaan adalah penelitian yang dilakukan dengan mengumpulkan data, informasi, dan berbagai data lain yang terdapat dalam kepustakaan yang berkaitan dengan ilmu evaluasi dan sejarahnya. Dari penelitian ini menunjukkan bahwa secara global terdapat tujuh periode dan hal ini sama dengan yang terjadi di Indonesia yang mengalami tujuh kali perubahan kebijakan penilaian dan banyak perubahan juga dipengaruhi oleh perubahan paradigma penilaian, hingga pada akhirnya dapat disimpulkan bahwa paradigma *Assessment as Learning* merupakan paradigma yang sesuai dengan konsep muhasabah dalam Islam.

Kata kunci: *Sejarah evaluasi, penilaian, paradigma, muhasabah*

Pendahuluan

Pelaksanaan penilaian hasil belajar dalam pendidikan nasional termasuk subsistem kurikulum sekolah sering mendapat kritik tajam dari masyarakat dan lembaga-lembaga lain karena kebijakan dan

pelaksanaannya sering berubah-ubah tanpa di dukung oleh data yang akurat. Salah satu factor penyebabnya adalah kurangnya evaluasi program yang dilakukan secara teratur dan berkelanjutan. Oleh karena itu perubahan yang terjadi dalam kebijakan dan pelaksanaan penilaian hasil belajar perlu didasarkan atas data yang akurat, handal, dan relevan.

Penilaian merupakan komponen yang tak kalah penting dibanding dengan komponen pembelajaran yang lain. Penilaian bertujuan untuk menentukan tingkat ketercapaian peserta didik dalam pembelajaran. Upaya meningkatkan kualitas pendidikan dapat juga ditempuh melalui peningkatan kualitas sistem penilaiannya. Sistem penilaian yang baik akan mendorong pendidik untuk menentukan strategi mengajar yang baik dan memotivasi peserta didik untuk belajar lebih baik. Oleh karena itu, dalam upaya peningkatan kualitas pendidikan diperlukan perbaikan sistem penilaian yang diterapkan. Untuk melaksanakan perbaikan tersebut pendidik perlu memahami betul konsep dasar penilaian itu sendiri karena perkembangan konsep penilaian yang ada pada saat ini menunjuk arah yang lebih luas. Pengetahuan dan pemahaman yang mendalam terhadap konsep dasar penilain akan membantu pendidik lebih mudah dalam melakukan perbaikan terhadap sistem penilaian di sekolah. Pengetahuan dan pemahaman ini juga akan meningkatkan kesadaran pendidik terhadap pentingnya peranan sistem penilain yang baik dalam pembelajaran di sekolah.

Penilaian hasil belajar seringkali didekati dari sudut pandang guru atau pendidik dan tidak jarang pula didekati dari sudut pandang sosial (sekedar memenuhi tugas bahkan terkadang biar *pantes*). Penilaian pada umumnya menekankan evaluasi setelah program berlangsung dan

terpisah dari pengelolaan program pembelajaran. Kebalikannya, pengelolaan menekankan pemenuhan syarat laporan program berlangsung dan sebagai bagian yang tak terpisahkan dari pengelolaan itu sendiri (perencanaan, pelaksanaan dan evaluasi). Pada sisi lain, penilaian di anggap sebagai prestasi akhir dari sebah program oleh pendidik, peserta didik, wali bahkan dan juga pengelola lembaga pendidikan. Terlebih lagi hasil penilaian dibuat untuk menentukan studi keberhasilan, kelayakan untuk program dan identifikasi kebutuhan dan sumber belajar untuk program-program lanjutan.

Dari sudut manajemen atau pengelolaan, sebagian besar evaluasi setelah program berlangsung untuk kepentingan pertanggungjawaban, sementara sebagaimana diatas bahwa peniliaian dalam prosesnya kurang memenuhi standar penilaian. Perbedaan kedua sudut pandang ini menjadi semakin tidak jelas pada waktu menentukan aspek apa dari program yang harus dievaluasi dan mana yang akan dijdikan sebagai bentuk prestasi.

Pembelajaran menurut Ralph Tyler memiliki empat komponen utama, yaitu tujuan, materi, metode/ media, dan penilaian¹. Keempat komponen tersebut saling berkaitan dan mendukung. Keempat komponen ini juga menjadi sorotan besar masyarakat dalam menilai pendidikan terutama terhadap penilaian. Tak jarang masyarakat menganggap penilaian belum dilaksanakan sebagai mestinya. Sistem penilaian yang dilakukan di sekolah sering kali mendapat keraguan masyarakat sehingga memunculkan berbagai

¹ Anitah, Sri dkk. 2008. Strategi Pembelajaran di SD. Jakarta : Universitas Terbuka, hal.126

permasalahan dalam penyelenggaraan pendidikan di sekolah.

Bagaimana jika pendidikan tanpa penilaian, Robert L. Ebel dalam *Practical Problems in Educational Measurement*² menyebutkan beberapa konsekuensi yang muncul jika ujian negara tidak dilakukan, yaitu:

1. Dorongan dan penghargaan atas usaha seseorang untuk belajar akan menjadi lebih sulit.
2. Kesuksesan program pendidikan kurang dapat dinyatakan sebagai tujuan dan pencapaian kurang dapat dibuktikan.
3. Keputusan-keputusan penting terkait dengan masalah kurikulum dan metode tidak diambil berdasarkan bukti-bukti yang kuat melainkan lebih berdasarkan pada perkiraan dan cenderung plin-plan.
4. Kesempatan menempuh pendidikan tidak berdasarkan bakat dan prestasi namun lebih berdasarkan keturunan dan pengaruh yang dimiliki.
5. Hambatan kelas sosial kurang dapat ditembus.

Perkembangan ilmu pengetahuan dan teknologi (Iptek) serta globalisasi menimbulkan perubahan dalam berbagai aspek kehidupan masyarakat. Perubahan tersebut memberi pengaruh yang besar terhadap berbagai kegiatan dan kebutuhan masyarakat, termasuk kebutuhan akan pendidikan. Pandangan, kritikan, sumbangan, dan tuntunan terhadap pendidikan kini terus mengalami pergeseran. Pendidikan kini telah menarik perhatian masyarakat dalam pelaksanaan maupun penilaiannya serta hasil yang diperoleh.

Belum banyak juga dari sekian praktisi pendidikan yang memahami akan sejarah dan perkembangan penilaian pendidikan

baik global maupun di Indonesia. Perubahan kebijakan kurikulum di Indonesia yang telah mengalami beberapa kali perubahan sejak kemerdekaan (1945) sampai dengan saat ini, hal ini berimplikasi pada paradigma para praktisi pendidikan dalam memaknai penilaian dalam pendidikan, utamanya dalam praktek pelaksanaan penilaian.

Perubahan paradigm penilaian seiring dengan perubahan kurikulum, hampir disetiap perubahan kurikulum disertai dengan perubahan model penilaian, hal ini dikarenakan perubahan dasar filosofis, psikologis, ilmu pengetahuan dan teknologi, dan budaya yang berkembang. Dari *teacher center* ke *Student center*, dari *behavioristik* ke *konstruktivistik* dan dari era 1.0 sampai dengan era 5.0 yang terus menuntut bagi dunia pendidikan menyesuaikan dengan kondisi dan kebutuhan masyarakat, termasuk dalam pendidikan.

Perubahan istilah dari tes (Tes Hasil Belajar), ujian UAN/UUB, bahkan penilaian (PAS/PAT) dan sebagainya, disadari ataupun tidak sebenarnya para ahli pendidikan yang telah merumuskan mempunyai harapan besar dalam memaknai penilaian, akan tetapi dalam masyarakat (praktisi pendidikan utamayan) di anggap sebagai sebuah istilah baru yang tidak bernilai, sekedar hanya difahami perubahan *pure* teknis semata, paradigma perubahannya bukan menjadi bagian substansi yang juga harus dirubah. Tulisan ini mengurai problematika diatas dengan menjelaskan secara deskriptif tentang sejarah penilaian secara global dan nasional dan perkembangan pemaknaan atau paradigm yang digunakan utamanya dalam menghadapi abad 21.

Dengan metode penelitian deskriptif eksplanatif, tulisan ini termasuk jenis penelitian kepustakaan (*library research*). Penelitian kepustakaan adalah penelitian

² Janesick, V.J. 2012, *The Assessment Debate; A Reference Handbook*, Califronia, ABC-CLIO.inc., hal. 186

yang dilakukan dengan cara mengumpulkan data, informasi, dan berbagai macam data-data lainnya yang terdapat dalam kepustakaan³. Dengan mengutarakan jenis penelitian ini, fokus dan langkah-langkah yang akan dalam penelitian ini menjadi semakin jelas. Mengenai sumber data, karena tulisan ini sifatnya adalah kajian pustaka, maka obyek yang dapat dijadikan sumber dibagi menjadi dua, yaitu primer dan sekunder. Data primer adalah adalah buku, jurnal, dan bulletin, Sedangkan data sekunder adalah buku buku masih dianggap relevan dengan kajian penelitian⁴. Untuk mengarahkan keakuratan dan ketepatan terhadap data yang diteliti, metode analisa yang digunakan yaitu *content analysis*⁵.

Pengolahan data dalam penelitian ini menggunakan metode deskriptif-analitis, yaitu model penelitian yang berupaya mendeskripsikan, mencatat, menganalisa dan menginterpretasikan kondisi-kondisi yang ada sehingga diharapkan bisa memberikan gambaran kepada generasi pendidik di Indonesia dalam menyikapi masyarakat digital terutama hadirnya generasi alfa sebagai generasi yang lahir di era revolusi industri ke empat. Upaya mengumpulkan informasi dimaksud dapat diperoleh dari buku-buku, laporan penelitian, jurnal-jurnal, ensiklopedia, dan sumber-sumber tertulis baik cetak maupun elektronik lainnya. Adapun langkah kongkritnya adalah menentukan hubungan antar katageri

dengan yang lain, melakukan analisis dan interpretasi sesuai dengan peta penelitian yang dibimbing oleh masalah dan tujuan penelitian. Proses analisis data ini dilakukan untuk mewujudkan kontruksi teoritis sesuai dengan masalah.

Pembahasan

Sejarah Penilaian

Griffin & Nix 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*⁶. Berdasarkan pada pendapat Griffin & Nix di atas, pengukuran, penilaian dan evaluasi adalah sebuah hirarkis. 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

Dalam tinjauan sejarahnya, evaluasi secara umum dan evaluasi kurikulum secara lebih spesifik telah melampaui sejumlah periode. Madaus dan Kellaghan (dalam Stufflebeam, Madaus dan Kellaghan, Ed., 2000) dalam *Program Evaluation: A Historical Overview* mengelompokkan sejarah evaluasi menjadi tujuh periode: *Age of Reform* (1792-1900), *Age of Effeciency* (1900-1930), *Age of Tylerian* (1930-1945), *Age of of*

³ PJ Subagyo. 1991, *Metode Penelitian: dalam teori dan praktek*. Jakarta : Rineka Cipta,hal. 109

⁴ Arikunto, S. 2012. *Prosedur Penelitian: Suatu Pendekatan Praktek*. Jakarta: Rineka Cipta, hal. 131.

⁵ *Ibid.*

⁶ Griffin, et.all. (2091), *Assesment and Teaching of 21th Century Skill*. Victoria Springer, p.3

*Innocence (1946-1957), Age of Development (1958-1972), Age of Professionalization (1973-1983), dan Age of Expansion and Integration (1983-2000)*⁷.

Pada masing-masing periode, terdapat karakteristik evaluasi yang membedakannya dengan periode sebelum atau sesudahnya

1. Masa Pembaharuan (*Age of Reform* 1792-1900)

Pada masa revolusi industri disertai dengan perubahan ekonomi dan teknologi yang kemudian tersusun struktur masyarakat modern dengan segenap akibatnya. Perubahan masa ini telah menjadikan beberapa perubahan baik ekonomi, sosial, maupun politik, khususnya di Amerika Serikat dan di Inggris.

Pembaharuan pada sistem pendidikan, pada agen-agen Evaluasi sosial dan pendidikan ini seringkali bersifat informal dan mengandalkan kepada kesan dan perasaan yang dilaksanakan oleh Komisi yang ditunjuk oleh Pemerintah. Misal '*Royal Commission of Inquiry*' yang terjadi di Irlandia, setelah menerima kesaksian dan bukti-bukti, dan menghasilkan rekomendasi untuk mengadopsi sistem 'bayar berdasarkan hasil' (*payment by results*) yang telah diterapkan di Inggris dan guru-guru digaji tergantung sebagian dari hasil ujian siswanya dalam hal membaca, menulis dan matematika. Saat program ditetapkan di Inggris, evaluasi periodik dilaksanakan melalui laporan tahunan yang disampaikan oleh Inspektur. Umpamanya, dalam bidang pendidikan terhadap inspektur-inspektur yang mengunjungi sekolah-sekolah

⁷ Stufflebeam, D.F., Madaus, G.F. 2000. *Program Evaluation: A Historical Overview* dalam Stufflebeam, D.F., Madaus, G.F., Kellaghan, T. Ed. *Evaluation Models: Viewpoints on Educational and Human Services Evaluation*. Second Edition. Boston: Kluwer Academic Publishers, p. 3-18.

setiap tahun dan membuat laporan tentang kondisi sekolah-sekolah dan hasil belajar siswa-siswanya.

Diantara hal menarik adalah terdapat agen kebijakan tersembunyi di samping gerakan ujian lisan; yaitu pengumpulan data untuk melakukan perbandingan yang dapat digunakan dalam pengambilan keputusan pada waktu pemilihan kepala sekolah. Akan tetapi kemajuan evaluasi program pendidikan nonformal ternyata tidak sepesat evaluasi pendidikan formal. Program pendidikan nonformal di Amerika yang disajikan untuk petani dan buruh-buruh tambang tidak menjadi

bagian dari evaluasi formal⁸.

2. Masa Efisiensi dan Testing (*Age of Efficiency* 1900-1930)

Pada saat tersebut gagasan pengelolaan yang berlandaskan ilmiah menjadi pendorong teori administrasi di bidang pendidikan dan industri. Penekanan gagasan ini terletak pada sistemik, standarisasi dan yang terpenting adalah efisiensi. Sesuatu program harus dievaluasi efisiensinya dan untuk ini tes menjadi alat utama evaluasi.

Seiring standarisasi *tes achievement* pasca dunia kesatu, sekolah-sekolah negara bagian di Amerika menggunakan tes untuk menentukan efektifitas, Universitas juga tidak ketinggalan membentuk lembaga khusus untuk melakukan evaluasi yang diperlukan oleh negara bagian.

Pada awal masa ini (1900 - 1930), evaluasi seringkali disalahgunakan untuk mengekspose kelemahan-kelemahan lembaga-lembaga pendidikan dan pengembangan masyarakat dan mengusulkan saran-saran perbaikan.

⁸ Achmad Rasyad, sejarah-evaluasi, online dalam <http://kuliahevaluasiprogrampls.blogspot.com/2013/09/pertemuan-ke-2-sejarah-evaluasi.html> di akses 7 Juli 2020

Sangat disayangkan evaluator pada saat ini tidak menyadari bahwa survey untuk melakukan evaluasi dapat dan harus menghindari penyalahgunaan dan dilakukan untuk hal yang konstruktif, dikerjakan bekerja sama dengan advisor setempat dan dirancang untuk menciptakan dukungan publik terhadap perubahan yang tidak dikenal tetapi diperlukan.

3. Masa Tyler (*Age of Tylerian* 1930-1945)

Dinamakan masa Tyler, karena pada masa itu pengaruh Tyler sangat besar dalam bidang evaluasi pendidikan khususnya dalam bidang testing. Dia dianggap sebagai “nabinya” evaluasi. Istilah evaluasi pendidikan diartikan mengukur seberapa jauh objektif program pengajaran telah dicapai. Kemudian konsep ini pada tahun 1930-an digunakan oleh Tyler untuk menolong instruktur-instruktur di negara bagian OHIO Amerika Serikat untuk meningkatkan kursus-kursus yang mereka organisir dan tes yang dipergunakan mereka.

Tyler bergabung dan memimpin gerakan pendidikan progresif studi evaluasi terhadap pembaharuan yang dikenal dengan istilah ‘studi delapan tahun’ yang disponsori oleh *Carnegie Corporation*. Studi evaluasi yang dipimpin oleh Tyler adalah studi yang terbesar dan pertama kali tentang perbedaan efektivitas bermacam-macam tipe persekolahan.

Evaluasi adalah perbandingan antara hasil yang diharapkan dan hasil aktual. Evaluasi meliputi perbandingan antara hasil dengan objektif program dan tidak perlu harus menggunakan design penelitian dengan biaya tinggi dan membagi menjadi kelompok eksperimen dan kontrol. Aspek yang dievaluasi adalah hasil belajar bukan input pengajaran dan organisasi sehingga tidak memerlukan jastifikasi dari evaluator profesional. Pada masa ini evaluasi untuk pendidikan nonformal sudah mulai dikembangkan.

4. Masa Innocence (*Age of of Innocence* 1946-1957)

Masa ‘innocence’ atau disebut juga tahun masa bodoh, khususnya di Amerika Serikat. Pada masa ini keputusan, prasangka dan permusuhan sosial terjadi di daerah pedesaan dan perkotaan. Tetapi tidak ada orang kulit putih yang memperhatikan keadaan ini sebagai masalah sosial. Dalam bidang pendidikan, perluasan pemberian kesempatan belajar, penambahan personal dan fasilitas dilaksanakan di Amerika Serikat dan Inggris secara besar-besaran. Pembangunan dan peningkatan pendidikan mempunyai dampak terhadap evaluasi program dan hasilnya. Pada masa ini pengembangan teknik-teknik evaluasi telah tumbuh dengan pesat sejalan dengan perkembangan dalam bidang teknologi. Dua macam design evaluasi telah berkembang dengan pesat untuk eksperimen yang menggunakan kelompok kontrol dan design yang hanya membandingkan antara hasil dan objektif. Walaupun teknik evaluasi telah berkembang pesat, sedikit sekali terdapat bukti bahwa evaluasi program telah dilaksanakan baik di lembaga-lembaga pendidikan maupun pengembangan masyarakat.

5. Masa Ekspansi (*Age of Development* 1958-1972)

Di awal masa ini Amerika Serikat mengalami kekecewaan dengan kemampuan Uni Soviet yang mampu meluncurkan Sputnik tahun 1959. Seperti yang dikatakan Popham (1974) Amerika termasuk bangsa yang ingin terbaik dalam segala hal. Karena itu mereka perlu meninjau kembali sistem pendidikan yang menghasilkan tenaga-tenaga ahli dan mengevaluasi mengapa sampai tertinggal dengan Uni Soviet. Atas dasar ini program matematika dan sains dimasukkan ke dalam kurikulum dan bahkan menjadi kurikulum utama dengan

harapan bukan hanya sekedar mengejar ketinggalan, tetapi ingin melampaui Uni Soviet yang menjadi saingan utama mereka.

Dengan menggunakan model-model dan teknik evaluasi yang sudah ada ternyata tidak menolong pengembang-pengembang kurikulum, program pendidikan dan pengembangan masyarakat. Selain itu, hasil evaluasi bukan sebagai respon terhadap pertanyaan-pertanyaan yang dikemukakan oleh orang-orang yang ingin mengetahui efektifitas program. Para evaluator menyadari masalah ini setelah mereka dengan dana yang cukup besar dan dengan cermat menerangkan teknik evaluasi yang sudah ada pada saat itu dan menghasilkan sesuatu yang tidak berguna untuk pengembangan kurikulum dan program-program lagi.

Evaluasi meliputi output proses, input dan konteks dari program itu sendiri. Evaluasi output termasuk evaluasi terhadap tujuan, hasil yang diharapkan dan tidak diharapkan. Evaluasi program tidak hanya terbatas untuk pendidikan formal, tetapi sudah mulai dikembangkan teori-teori evaluasi yang cocok untuk pendidikan nonformal seperti yang dikembangkan oleh Paulo Freire dan Meziro W. Untuk evaluasi program pendidikan nonformal banyak dikembangkan teori-teori partisipator yang berasal dari penelitian partisipatori dan evaluasi demokratis yang semula diterapkan untuk pendidikan formal. Lebih jauh lagi, teori-teori evaluasi dikembangkan dengan pesat tidak hanya di Amerika Serikat tetapi juga di Inggris seperti *illuminative evaluation* yang dikembangkan oleh Parlett dan Hamilton.

6. Masa Profesionalisasi (*Age of Professionalization* 1973-1983)

Evaluasi sudah tumbuh menjadi profesi tersendiri yang tidak sama dengan penelitian dan testing. Walaupun teknik pengumpulan data dan penentuan sampel

mungkin saja penelitian tetapi tingkat generalisasinya tidak sejauh penelitian, khususnya dalam hal waktu dan instansinya. Dalam evaluasi program, tes mungkin digunakan, tetapi penggunaannya hanya sebagai bagian dari evaluasi yang lazim disebut dengan istilah produk.

Pada mulanya evaluator merasakan adanya krisis identitas. Mereka tidak yakin apakah mereka harus masuk kubu peneliti, tester atau administrator dan kualifikasi apa yang harus dimiliki. Tidak banyak majalah evaluasi program yang dapat digunakan oleh evaluator untuk bertukar hasil evaluasi dan gagasan. Belum ada organisasi profesional yang berdedikasi terhadap evaluasi sebagai disiplin tersendiri. Pada dasarnya belum ada buku literatur evaluasi program kecuali sekedar masalah yang tidak diterbitkan. Lebih buruk lagi, evaluasi banyak dilaksanakan oleh orang-orang yang bukan dipersiapkan untuk itu (*amatiran*). Untuk mengatasi krisis ini, evaluasi program perlu diprofesionalisasikan. Majalah-majalah diterbitkan seperti *Educational Evaluation and Policy Analysis*, *Hindies in Evaluation*, *CEDRQuarterly*, *Evaluation Review*, *New Direction for Program Evaluation* dan lain-lain. Majalah majalah ini menjadi alat efektif untuk merekam dan menyebarluaskan informasi tentang evaluasi program.

7. Zaman Ekspansi dan Integrasi (*Age of Expansion and Integration* 1983-2000).

Profesionalisasi evaluasi telah melahirkan beberapa permasalahan; *Pertama*, tidak diragukan adanya komunikasi yang lebih baik dan intensif dalam bidang evaluasi, di samping juga menimbulkan banyak gunjingan. *Kedua*, pada waktu terjadi peningkatan training dan kewenangan evaluator untuk meyakinkan bahwa lembaga-lembaga pemesan mendapat pelayanan dari orang-orang

yang qualified, beberapa pengamat cemas bahwa pengembangan ini akan menimbulkan klub yang eksklusif dan berpandangan sempit. *Ketiga*, kerjasama di antara organisasi-organisasi profesional yang bergerak dalam bidang evaluasi program pendidikan adalah suatu hal yang positif, tetapi juga gampang buyar, untuk mempromosikan dan melaksanakan kerja evaluasi yang berkualitas tinggi. *Keempat*, pembentukan organisasi-organisasi profesional baru telah meningkatkan komunikasi antar evaluator dan mengurangi fragmentasi

dalam bidang evaluasi⁹.

Masalah lain yang mendasar adalah perbedaan antara pendekatan kualitatif dan kuantitatif dalam evaluasi program. Perbedaan bukan hanya sekedar teknik, tetapi landasan filosofis. Kualitatif berangkat dari *subjectivist epistemology* yang menempatkan kebenaran tidak terlepas dari konteks dan pengetahuan adalah persepsi dan pemahaman individu terhadap realita. Kuantitatif berangkat dari *objectivist epistemology* yang menempatkan kebenaran terlepas dari konteks. Karena itu pengetahuan adalah realita yang terlepas dari persepsi individu. Evaluator yang kurang awas seringkali menggabungkan dua teknik dari pendekatan yang berbeda ini dan akhirnya mungkin akan membuahkan hasil evaluasi yang menyedihkan.

Yunus Abidin secara umum juga membuat empat generasi dalam sejarah perkembangan penilaian pembelajaran, analisa yang didasarkan atas buku-buku pokok penilaian diantaranya karya Popham, Janesick, dan Griffin *et.all*, Tidak disebutkan secara spesifik masa

yang dimasukkan tetapi didasarkan atas tujuan dari penilaian pembelajaran¹; Generasi pertama memaknai tujuan penilaian dalam makna yang cukup sempit dan ini yang sampai sekarang masih banyak dilaksanakan oleh sebagian praktisi pendidikan, pada generasi kedua penilaian dimaknai dalam cara pandang yang pragmatis, penilaian tidak lebih dari sekedar alat ukur semata, sementara pada generasi ketiga dan keempat menempatkan penilaian sebagai bagian inti dari pendidikan.

Tujuan penilaian generasi pertama memiliki tiga tujuan utama yakni mengukur hasil belajar, mengukur efektivitas pembelajaran, dan memberikan umpan balik kepada guru Popham. Dari tujuan diatas banyak dikritisi karena sempitnya makna penilaian sebagai bentuk ukuran keberhasilan belajar yang berarti bersifat sumatif atau hanya berorientasi pada hasil, adapun tujuan yang kedua dan ketiga lebih kepada *teaching oriented* bukan pada pendidikan secara utuh.

Atas dasar kritikan diatas maka tujuan penilaian pada generasi kedua lebih ditekankan pada tujuan penilaian sebagai ukuran perkembangan siswa, perkembangan efektivitas pembelajaran dalam pandangan masyarakat, dan pemberian umpan balik bagi kegagalan belajar siswa sesuai dengan tahapan autentik belajarnya Popham. Demikian pun dalam generasi kedua penilaian masih mendapatkan kritikan karena penilaian masih sebatas ukuran untuk menentukan perkembangan keberhasilan siswa yang kemudian disikapi dengan remedial atau pengayaan, dan penilaian

⁹ Achmad Rasyad, sejarah-evaluasi, online <http://kuliahevaluasiprogrampls.blogspot.com/2013/09/pertemuan-ke-2-sejarah-evaluasi.html> di akses 7 Juli 2020

¹ Yunus Abidin, 2016. *Revitalisasi Penilaian Pembelajaran dalam konteks Pendidikan Multiliterasi abad ke-21*, Bandung: Refika Aditama, hal. 8-14

dijadikan sebagai ukuran pragmatis penentuan kualitas sumber daya manusia bagi masyarakat. Dalam konteks ini penilaian dipercaya sebagai alat yang valid untuk mengukur kemampuan siswa dalam berkiprah baik dalam dunia usaha maupun industri.

Dalam rangka revitalisasi munculah era generasi ketiga yang menentukan tujuan penilaian sebagai:

1. Pengarah pengembangan kurikulum yang berbasis keilmuan dan sinyal pasar
2. Pemandu perencanaan proses pembelajaran yang efektif
3. Pembimbing pada setiap tahapan perkembangan belajar siswa
4. Alat ukur yang akuntabel dan *high stakes*¹

Dari paradigma generasi ketiga inilah muncul istilah *Curriculum based assesment* (CBA) yakni kurikulum yang memiliki ciri bahwa tujuan kurikuler bertindak sebagai kriteria untuk mengidentifikasi target pembelajaran dan untuk menilai status serta kemajuan pembelajaran yang dilakukan dilingkungan alamiah praktisi pendidikan. Asumsi dasar dari pengembangan CBA bahwa penilaian sebagai; 1). Bentuk identifikasi entry point program, 2). Perencanaan program, 3). Monitoring program, dan 4). Evaluasi program. Konsep akuntabilitas dan *high stakes* dengan implementasi validitas instrument yang tidak hanya sekedar dari sisi konstruk, isi dan butir tetapi lebih jauh pada tataran pragmatis(*use*), hal ini dilandaskan atas dimensi-dimensi social yang berkembang dimasyarakat.

Pada generasi keempat, penilaian memiliki beberapa tujuan diantaranya:

1. Membantu siswa lebih termotivasi belajar

2. Membantu guru untuk memperbaiki dan mengembangkan kemampuan dan ketrampilan praktis dalam mengajar
3. Meningkatkan kualitas system pendidikan secara menyeluruh¹

Asumsi dasar dalam penilaian generasi keempat menempatkan penilaian sebagai motivasi eksternal siswa dalam belajar, bukan sekedar alat ukur tetapi alat paku untuk belajar dan belajar. Jika pada generasi ketiga penilaian sebagai pemandu proses belajar tetapi selanjutnya generasi keempat menempatkan penilaian sebagai bentuk *muhasabah*, alat bantu guru dalam memperbaiki dan mengembangkan kemampuan guru dalam pembelajaran. Termasuk juga dalam generasi ini, penilaian sebagai alat pengembang dalam kurikulum, proses KBM, control kompetensi lulusan, dasar pengembangan media pembelajaran, melalui tujuan ini maka pendidikan akan dipandu oleh sistem penilaian yang baik. Sehingga cakupan dan urgensi penilaian dalam pendidikan secara komperhensif memiliki makna yang luas, pendidikan memiliki kriteria dan standart yang jelas melalui pendidikan berbasis penilaian.

Sejarah Penilaian Di Indonesia

Ujian nasional diberlakukan di Indonesia mulai tahun 1950 dan telah beberapa kali berganti sistem dan format ujian. Berikut sejarah perjalanan ujian

nasional Indonesia¹ :

1. Ujian Penghabisan (Periode 1950 – 1965)
Materi ujian dibuat oleh Departemen Pendidikan, Pengajaran, Dan Kebudayaan menerapkan format ujian yaitu seluruh

¹ Yunus Abidin, 2016. *Revitalisasi Penilaian Pembelajaran dalam konteks Pendidikan Multiliterasi abad ke-21*, Bandung: Refika Aditama, hal. 11

¹ Yunus Abidin, 2016. *Revitalisasi Penilaiana*, hal.13 .

¹ Admin, sejarah-ujian-nasional-indonesia dalam online <https://www.padamu.net/sejarah-ujian-nasional-indonesia> di akses pada juni 2020, data asal dari kemendikbud.go.id tetapi saat ini dalam website tersebut sudah tidak ada

soal dikerjakan dalam bentuk esai atau isian. Hasil ujian diperiksa di pusat rayon yang telah ditentukan.

2. Ujian Negara (Periode 1965 – 1971)

Waktu dan materi ujian ditentukan oleh pemerintah pusat. Bahan ujian adalah seluruh mata pelajaran. Artinya materi yang diuji adalah semua mata pelajaran yang diajarkan kepada siswa didik.

Tujuannya adalah untuk menentukan kelulusan, sehingga siswa dapat melanjutkan ke sekolah negeri atau perguruan tinggi negeri apabila telah lulus Ujian Negara. Sedangkan bagi yang tidak lulus Ujian Negara tetap memperoleh ijazah dan dapat melanjutkan ke sekolah atau perguruan tinggi swasta. Ujian Negara kala itu memiliki kelebihan dan kekurangan. Untuk kelebihannya, yakni dapat mendorong siswa giat belajar dan guru mengajar dengan baik, Kelebihan dari Ujian Negara pada saat itu adalah dapat mendorong siswa giat belajar dan guru mengajar dengan baik, nilai ujian setiap siswa/sekolah/daerah memiliki makna yang sama dan dapat diperbandingkan (comparable). Kekurangannya, yakni biaya distribusi bahan ujian cukup tinggi, risiko kebocoran soal cukup tinggi, dan tingkat ketidakkelulusan peserta didik tinggi.

3. Ujian Sekolah (Periode 1972 – 1979)

Pada periode ini, pemerintah pusat hanya menyusun pedoman dan panduan ujian nasional yang bersifat umum. Penyelenggaraan ujian nasional dilakukan oleh masing-masing sekolah atau sekelompok sekolah dengan waktu dan materi yang dapat disesuaikan. Dengan demikian materi dan hasil ujian dilakukan masing-masing sekolah atau kelompok.

Kriteria tamat ditentukan oleh masing-masing sekolah dengan tidak mengenal Lulus atau Tidak Lulus, tetapi menggunakan istilah TAMAT. Biaya ujian sepenuhnya ditanggung oleh peserta didik. Persentase kelulusan

sangat tinggi bahkan dapat dikatakan semua peserta didik lulus (100%), namun mutu lulusan tidak dapat diperbandingkan.

Ujian Sekolah juga memiliki kelebihan dan kekurangan. Untuk kelebihannya, yakni menurunkan tingkat drop out peserta didik, tidak ada tekanan bagi sekolah dalam hal kelulusan, dan sekolah memiliki otoritas yang tinggi dalam penentuan kelulusan.

Kekurangannya, seperti nilai hasil ujian antar sekolah tidak dapat diperbandingkan, hasil ujian sekolah tidak dapat dilakukan pemetaan sekolah pada tingkat daerah dan nasional, dan hasil ujian tidak dapat dijadikan sebagai alat seleksi.

4. EBTA/EBTANAS (Periode 1980 – 2001)

Pada periode ini, ujian nasional disebut dengan EBTANAS (Evaluasi Belajar Tahap Akhir Nasional) dan EBTA (Evaluasi Belajar Tahap Akhir). Kelulusan siswa ditentukan oleh kombinasi nilai EBTANAS yang dikoordinasi oleh pemerintah pusat dan EBTA yang dikoordinasi oleh pemerintah daerah serta ditambah nilai ujian harian yang tertera di buku rapor. Pada periode ini dikenal dengan DANEM (daftar nilai ebtanas murni) yaitu hasil nilai ebtanas apa adanya. Namun kelulusan berdasarkan nilai rata-rata seluruh mata pelajaran yang diujikan meskipun ada siswa yang mendapat nilai dibawah tiga pada mata pelajaran tertentu.

Tujuan dari Ebtanas dan Ebta adalah untuk memperoleh Surat Tanda Tamat Belajar (STTB). Pada awal diberlakukannya mata pelajaran yang diujikan dalam Ebtanas adalah Pendidikan Moral Pancasila (PMP), kemudian pada tahun berikutnya ditambah dengan beberapa mata pelajaran lainnya. Sejumlah mata pelajaran pokok diujikan melalui Ebtanas, sedangkan mata pelajaran lainnya diujikan melalui Ebta.

Bahan Ebtanas yang berupa kumpulan soal disiapkan oleh pusat (Dit. Pendidikan Dasar dan Menengah). Panitia daerah merakit paket tes dan menggandakannya. Sedangkan bahan ujian Ebta disiapkan oleh masing-masing sekolah/daerah/wilayah. Tanggung jawab penyelenggaraan Ebtanas dan Ebta adalah sekolah, pemerintah daerah, dan pemerintah pusat.

Tentu Ebtanas dan Ebta memiliki kelebihan dan kekurangan. Untuk kelebihannya, seperti nilai hasil ujian (khususnya Ebtanas) dapat dibandingkan, nilai Ebtanas Murni (NEM) dapat dijadikan sebagai pertimbangan dalam seleksi masuk ke jenjang yang lebih tinggi, dan dapat dilakukan pemetaan mutu sekolah berdasarkan NEM pada tingkat daerah dan nasional.

Sedangkan kekurangannya, sekolah yang nilai Ebtanasnya rendah cenderung menaikkan (memanipulasi) nilai P dan Q untuk mencapai batas kelulusan, motivasi belajar peserta didik rendah, karena peserta didik beranggapan bahwa semua akan lulus sehingga tidak memotivasi untuk giat belajar

5. UAN (Periode 2002 – 2004)

Pada periode ini, ujian nasional disebut dengan Ujian Akhir Nasional (UAN) menggantikan Ebtanas. Standar kelulusan UAN setiap tahun berbeda-beda:

- a. Pada UAN 2002 kelulusan ditentukan oleh nilai mata pelajaran secara individual.
- b. Pada UAN 2003 standar kelulusan adalah 3.01 pada setiap mata pelajaran dan nilai rata-rata minimal 6.00. Soal ujian dibuat oleh Depdiknas dan pihak sekolah tidak dapat mengangkat nilai UAN. Para siswa yang tidak/belum lulus masih diberi kesempatan mengulang selang satu minggu sesudahnya.

- c. Pada UAN 2004, kelulusan siswa didapat berdasarkan nilai minimal pada setiap mata pelajaran 4.01 dan tidak ada nilai rata-rata minimal. Pada mulanya UAN 2004 ini tidak ada ujian ulang bagi yang tidak/belum lulus. Namun setelah mendapat masukan dari berbagai lapisan masyarakat, akhirnya diadakan ujian ulang.

6. UN (Periode 2005 – 2012)

Pada periode ini, ujian nasional disebut dengan Ujian Nasional (UN) menggantikan Ujian Akhir Nasional (UAN). Standar kelulusan Ujian Nasional (UN) setiap tahun juga berbeda :

- a. Pada UN 2005 minimal nilai untuk setiap mata pelajaran adalah 4.25.
- b. Pada UN 2005 ini para siswa yang belum lulus pada tahap I boleh mengikuti UN tahap II hanya untuk mata pelajaran yang belum lulus.
- c. Pada UN 2006 standar kelulusan minimal adalah 4.25 untuk tiap mata pelajaran yang diujikan dan rata-rata nilai harus lebih dari 4.50 dan tidak ada ujian ulang.
- d. Pada UN 2007 ini tidak ada ujian ulang. Dan bagi yang tidak lulus disarankan untuk mengambil paket C untuk meneruskan pendidikan atau mengulang UN tahun depan.
- e. Pada UN 2008 mata pelajaran yang diujikan lebih banyak dari yang semula tiga, pada tahun ini menjadi enam.
- f. Pada UN 2009 standar untuk mencapai kelulusan, nilai rata-rata minimal 5.50 untuk seluruh mata pelajaran yang di-UN-kan, dengan nilai minimal 4.00 untuk paling banyak dua mata pelajaran dan minimal 4.25 untuk mata pelajaran lainnya.
- g. Pada UN 2010 standar kelulusannya adalah, memiliki nilai rata-rata

minimal 5.50 untuk seluruh mata pelajaran yang diujikan, dengan nilai minimal 4.0 untuk paling banyak dua mata pelajaran dan minimal 4.25 untuk mata pelajaran lainnya. Khusus untuk SMK, nilai mata pelajaran praktek kejuruan minimal 7.00 dan digunakan untuk menghitung rata-rata UN.

- h. Pada UN 2011 dan 2012 Nilai kelulusan siswa masih tetap yaitu 5,5. Begitu juga soal tetap dibagi dalam enam macam paket, yakni lima soal utama dan satu cadangan bila ada soal tak lengkap atau rusak.

Ujian Nasional (UN) diselenggarakan untuk mengukur pencapaian kompetensi lulusan peserta didik pada jenjang satuan pendidikan dasar dan pendidikan menengah sebagai hasil dari proses pembelajaran sesuai dengan Standar Kompetensi Lulusan (SKL).

Selain itu, salah satu kegunaan hasil UN adalah untuk melakukan pemetaan tingkat pencapaian hasil belajar siswa pada satuan pendidikan. Salah satu upaya untuk mewujudkan pendidikan berkualitas diperlukan adanya sistem penilaian yang dapat dipercaya (*credible*), dapat diterima (*acceptable*), dan dapat dipertanggungjawabkan (*accountable*).

7. UN di Hapus

Sejak tahun 2006 timbul berbagai kritik, saran, dan tuntutan masyarakat tentang penyelenggaraan UN. Puncak kritik datang dari lembaga sosial yang menuntut agar UN ditiadakan karena dianggap melanggar Hak Asasi Manusia yaitu hak anak untuk melanjutkan sekolah. Setelah melalui serangkaian persidangan, keputusan Mahkamah Agung atas tuntutan tersebut adalah bahwa UN dapat dilaksanakan apabila pemerintah memperbaiki kualitas guru dan sarana prasarana sekolah. Berdasarkan keputusan tersebut pemerintah telah berusaha memenuhinya, sambil

melaksanakan perbaikan secara terus menerus dan UN tetap dilaksanakan.

Hingga diberlakukan kurikulum 2013 dan hingga kini UN ditiadakan tetapi pada kurikulum dilaksanakan dengan berbagai model, sampai direvisi menjadi kurikulum 2016 atau kurikulum Nasional yang menghapus kebijakan nilai minimal dan akhirnya kelulusan sepenuhnya diserahkan satuan pendidikan masing-masing.

Paradigma Penilaian

Paradigma atau cara pandang dimkasudkan disini adalah pola pikir yang mempengaruhi terhadap perilaku dalam melaksanakan dalam evaluasi atau asesmen pendidikan dalam hubungannya dengan aktivitas belajar (*learning*), dan evaluasi atau asesmen pendidikan di sini dibatasi dalam konteks evaluasi atau asesmen kelas (*classroom assessment*). Peta perkembangannya berarti awal mula tiga paradigma evaluasi atau asesmen pendidikan dan pertumbuhan selanjutnya. Dalam tulisan-tulisan terkait pola pelaksanaan penilaian dan implikasi-implikasi penilaian terdapat tiga paradigma asesmen pendidikan (*assessment of learning* [AoL], *assessment for learning* [AfL], dan *assessment as learning*

[AaL]¹. Perkembangan awal tiga paradigma evaluasi pendidikan ditandai dengan kemunculan ide AaL. AaL ini membedakan dengan jelas antara AoL, AfL dan AaL itu sendiri. Perbedaan ini menunjukkan pergeseran paradigma evaluasi belajar atau evaluasi pendidikan dan juga menunjukkan kebutuhan akan AfL dan AaL atau asesmen formatif. Namun, ini tidak berarti bahwa AoL tidak berguna dan harus dihilangkan, tetapi lebih tepatnya, perlu penerapan ketiganya dalam keseimbangan baru.

¹ Agung Prihantoro, Tiga Paradigma Evaluasi Pendidikan: Sebuah Peta Perkembangan, AoEJ: Academy of Education Journal Volume 12 Nomor 1, Januari 2021, hal. 22-38

AoL bertujuan untuk menilai hasil belajar dan melaporkan penilaian hasil belajar kepada siswa dan orang tua siswa. Penilaian itu umumnya membandingkan hasil belajar antarsiswa. Penilaiannya dilakukan lazimnya di akhir pelajaran, semester atau program, dan penilaiannya berbentuk tes yang berisi pertanyaan-pertanyaan tentang apa yang telah diajarkan. Hasil penilaiannya berupa simbol huru atau angka. Asumsi yang mendasari AoL adalah mengukur hasil belajar (to measure or to learn). Selama ini, AoL mendominasi sistem pendidikan di seluruh dunia.

Adapun Afl bertujuan tidak untuk membandingkan hasil belajar para siswa, tetapi untuk menemukan kelebihan dan kelemahan setiap siswa dan memberi umpan balik (*feedback*) kepada siswa tentang aktivitas-aktivitas belajar mereka selanjutnya. Data tentang kelebihan dan kelemahan setiap siswa digunakan untuk memodifikasi aktivitas-aktivitas belajar dan mengajar berikutnya.

AaL menguatkan dan meluaskan peran asesmen formatif untuk meningkatkan aktivitas belajar. Cara menguatkan dan meluaskannya adalah dengan menekankan peran siswa, bukan hanya sebagai kontributor dalam proses asesmen dan belajar, melainkan juga sebagai konektor kritis (*critical connector*) antara proses asesmen dan proses belajar. Siswa berperan ganda sebagai pelajar atau pemelajar (orang yang belajar) dan sebagai asesor atau evaluator yang aktif dan kritis. Sebagai asesor atau evaluator, siswa menilai kemajuan belajarnya: apa sudah sudah dicapai, apa yang belum dicapai, apa kelebihan dan kelemahan dirinya dalam belajar, apa yang akan dilakukan dalam aktivitas belajar berikutnya. Asumsi yang mendasarinya adalah bahwa swa-evaluasi atau *swa-asesmen* (*self-assessment*) merupakan inti dari aktivitas belajar dan evaluasi.

Menurut Stiggins mendefinisikan penilaian untuk belajar sebagai suatu alternatif penggunaan berbagai metode penilaian yang berbeda secara terus-menerus untuk memperoleh bukti penguasaan pengetahuan dan ketrampilan

siswa terhadap standar¹. Pendapat Stiggins tersebut menekankan pada penggunaan berbagai metode penilaian untuk memperoleh informasi tentang penguasaan konsep pengetahuan dan ketrampilan yang telah diajarkan. Informasi tersebut dapat digunakan sebagai dasar bagi guru, siswa, dan orang tua dalam membuat keputusan yang berkaitan dengan perkembangan belajar siswa. Dari berbagai definisi diatas maka dapat diambil kesimpulan bahwa AFL atau yang bisa juga dikatakan sebagai penilain untuk belajar adalah suatu penilaian yang terjadi saat proses pembelajaran berlangsung yang melibatkan interaktif guru dengan siswa atau siswa dengan siswa sehingga menumbuhkan motivasi siswa untuk mempertinggi aktivitas didalam pembelajaran ketika berlangsung kemudian dijadikan keputusan bersama antara guru dan siswa untuk menciptakan suasana belajar selanjutnya dan dilakukan secara terus-menerus demi mencapai perkembangan belajar siswa baik dari dimensi pengetahuanya, seikap dan ketrampilanya.

Berdasarkan definisi yang telah dikemukakan diatas, CEA menyajikan tujuan penilaian untuk belajar untuk: 1. Memberi wawasan tentang belajar siswa kepada guru dan siswa, 2. Meningkatkan kesuksesan untuk semua, 3. Membantu proses penetapan tujuan, 4. Memungkinkan refleksi secara kontinu terhadap apa yang siswa ketahui sekarang dan apa yang mereka butuhkan untuk diketahui berikutnya, 5. Mengukur apa

¹ R.J. Stiggins, Student-Center: Classroom Assessment, (New York: MacMillan College Publishing Company, 1994), hlm. 375.

yang dinilai, 6. Mempromosikan intervensi secara cepat dan menghubungkan dengan penetapan tujuan pembelajaran, dan 7. Meningkatkan standar yang diperoleh siswa pada *edges of capability*¹.

Konsep penilaian untuk belajar atau disingkat PuB (*assessment for learning atau AfL*) bukanlah hal baru dalam penilaian pendidikan. PuB dikembangkan melalui perpaduan antara hasil penelitian dan praktik penilaian dalam kelas. Prinsip-prinsip yang mendasari lahirnya PuB disusun dari kesimpulan hasil revidi penilaian formatif yang dilakukan oleh Black dan William, yaitu; 1. Keterlibatan aktif dari siswa dalam kegiatan belajarnya; 2. Provisi umpan balik yang efektif untuk siswa; 3. Pengaruh dari penilaian dalam memotivasi rasa percaya diri siswa, yang juga memberikan pengaruh penting dalam pembelajaran; 4. Keterbatasan diri siswa untuk mengasah dirinya sendiri dan mengerti bagaimana untuk meningkatkannya; 5. Penambahan jam mengajar untuk memperhitungkan hasil penilaian¹.

Yunus Abidin memberikan sepuluh prinsip utama dalam penilaian untuk belajar di abad 21, yaitu:

1. Harus disejajarkan dengan tujuan penting abad21;
2. Harus memusatkan bagaimana siswa beradaptasi;
3. Harus menitik beratkan pada aspek performa;
4. Harus memberikan nilai tambah bagi proses pembelajaran;
5. Harus peka dan bersifat membangun untuk menerapkan berbagai konsep atau strategi pemecahan masalah;
6. Harus memperhitungkan aspek keadilan;

7. Harus harus mampu memberikan informasi yang akurat untuk aspek pengambilan keputusan
8. Harus berorientasi pada tujuan jangka panjang pendidikan;
9. Harus mengahasilkan informasi yang dapat ditindaklanjuti dan memberikan umpanbalik secara produktif dan bermanfaat¹

Penutup

Jika melihat pada fase sejarah penilaian baik secara global maupun nasional maka perjalanan penilaian hasil belajar di Indonesia telah mengikuti perkembangan tersebut, meskipun jika dilihat dari penyesuaiannya baru terjadi pada lima tahun terakhir. Penilaian sejak kurikulum 2013 telah mengikuti paradigma Assessment as Learning (AaL), sehingga nilai minimum tidak lagi menjadi kebijakan atau standart dalam menentukan lulusan peserta didik, bahkan pada akhirnya diserahkan sepenuhnya kepada satuan pendidikan.

Pemerintah indoneisa telah menyadari⁷ bahwa penilaian bukan sebagai ukuran atau timbangan pencapaian belajar tetapi penilaian adalah bagian dari belajar itu sendiri. Paradigma lama yang menjadikan penilaian sebagai keberhasilan siswa akan tetapi siswa disamping sebagai pembelajar juga evaluator pada dirinya sendiri swa-asesmen yang dalam islam disebut dengan muhasabah. Jika ditinjau dan digabungkan dengan perkembangan paradigma penilaian, bahwa paradigma AaL melempgkapi dua paradigma seblumnya, maka penilaian disamping sebagai muhasabah kemampuan siswa dalam menyerap materi, juga sebagai muhasabah guru atas kemampuan dirinya dalam memberikan/menyampaikan materi,

¹ Harun, Rasyid & Mansur, 2009, Penilaian Hasil Belajar/ Bandung : Wacana Prima, Hal..88.

¹ Ibid.

⁶ ¹ Yunus Abidin, 2016. *Revitalisasi Penilaian Pembelajaran dalam konteks Pendidikan Multiliterasi abad ke-21*, Bandung: Refika Aditama, hal.19-20

sejauh mana kemampuan guru dalam mentransfer pengetahuan kepada siswa, dan difahami oleh siswa. Sehingga dari sini bias disimpulkan bahwa penilaian adalah kebutuhan guru dan murid.

DAFTAR PUSTAKA

- Anitah, Sri dkk. 2008. Strategi Pembelajaran di SD. Jakarta : Universitas Terbuka
- Arikunto, S. 2012. *Prosedur Penelitian: Suatu Pendekatan Praktek*. Jakarta: Rineka Cipta
- Cronbach, L.J. Ed., 1980. *Toward Reform of Program Evaluation: Aims, Methods, and Institutional Arrangements*. San Francisco: Jossey-Bass.
- Eko Putro Widoyoko, 2004. *Penilaian Hasil Pembelajaran di Sekolah*, Yogyakarta: Pustaka Pelajar.
- Griffin, et.all. (2012), *Assesment and Teaching of 21th Century Skill*. Victoria Springer.
- Guba, E., Lincoln, Y.S. 1981. *Effective Evaluation: Improving the Usefulness of Evaluation Results Through Responsive and Naturalistic Approach*. Thousand Oaks: SAGE Publication.
- Harun, Rasyid & Mansur, 2009, *Penilaian Hasil Belajar*, Bandung : Wacana Prima
- Imadikus.com/evaluasi-program-pendidikan-luar-sekolah/ (Online) diakses pada: 09 September 2014
- Janesick, V.J. 2012, *The Assessment Debate; A Reference Handbook*, California, ABC-CLIO.inc.
- Kemendikbud. Go.id
- Lorin W. Anderson, David R. Krathwohl, 2015. *Kerangka Landasan Untuk Pembelajaran, Pengajaran, dan Asesmen; Revisi Taksonomi Pendidikan Bloom*, (Yogyakarta: Pustaka Pelajar.
- Madaus, G.F., Scriven, M.S., & Stufflebeam, D.L. 1993. *Evaluation models, viewpoints on educational and human services evaluation*. Boston: Kluwer-Nijhoff Publishing.
- Meyers, L.H. 2005. *Time for Tun-Up: Comprehensive Curriculum Evaluation dalam Principal Leadership*, 6 (1), 27-30.
- Nitko, A. J. 2011. *Educational Assesment of Students*. Boston: Pearson.
- Popham, James W. 2011..Boston Pearson *Classrrom assessment; What teachers need to know*
- PJ Subagyo. 1991, *Metode Penelitian: dalam teori dan praktek*. Jakarta : Rineka Cipta,
- Robert L. Ebel, 1980. *Practical Problems in Educational Measurement*, Publisher: D C Heath & Co
- Stake, R. E. 1975. *Program Evaluation: Particularly Responsive Evaluation*. Illionis: Center for Instructional Research and Curriculum Evaluation (CIRCE) University of Illionis.
- Stake, R.E. 2004. *Standard Based Evaluation and Responsive Evaluation*. Thousand Oaks: SAGE Publication.
- Stufflebeam, D.F., Madaus, G.F. 2000. dalam Stufflebeam, D.F., Madaus, G.F., Kellaghan, T. Ed. *Evaluation Program Evaluation: A Historical Overview Models: Viewpoints on Educational and Human Services Evaluation*. Second Edition. Boston: Kluwer Academic Publishers.
- Sudjiono, A. 2008. *Evaluasi Pendidikan*. Jakarta: Rajawali Press.
- Tyler, R.W. 1981. *Specific Approaches to Curriculum Development* dalam Gyroux, H.A., Penna, A.N., Pinar
- Sax, G. 1980. *Principles of educational and psychological measurement and evaluation*, (2nd ed.). California: Wandsworth Publishing Company
- Yunus Abidin, 2016. *Revitalisasi Penilaian Pembelajaran dalam konteks Pendidikan Multiliterasi abad ke-21*, Bandung: Refika Aditama.