

PENGEMBANGAN INSTRUMEN EVALUASI CIPP PADA PROGRAM PEMBELAJARAN *TAHFIZ* AL-QUR'AN DI PONDOK PESANTREN

Muyasaroh, Sutrisno

Sekolah Tinggi Ilmu Tarbiyah Al-Qur'an Al-Ittifaqiah, UIN Sunan Kalijaga Yogyakarta
muyasnasir@yahoo.com, trisno_63@yahoo.com

Abstrak

Studi ini bertujuan untuk: (1) menghasilkan model pengembangan evaluasi program pembelajar'an *tahfiz* al-Quran diberi nama Coni P2, (2) menghasilkan teknik pelaksanaan evaluasi program pembelajaran *tahfiz* al-Qur'an, dan (3) menghasilkan struktur komponen dan indikator model evaluasi. Studi ini merupakan penelitian dan pengembangan (R&D) dengan menggunakan sembilan langkah dari 10 langkah model Borg dan Gall. Jumlah subjek uji coba pertama 33 orang, uji coba kedua 49 orang, dan uji coba ketiga 224 orang. Komponen model evaluasi yang digunakan adalah model evaluasi Stufflebeam (CIPP). Langkah-langkah evaluasi yang digunakan adalah langkah Malcolm Provus. Teknik pengumpul data yang digunakan adalah Delphi, FGD, kuesioner, observasi, wawancara, dan studi dokumentasi. Validitas konstruk dianalisis menggunakan CFA dan Reliabilitas menggunakan Cronbach Alpha. Hasil penelitian: (1) model evaluasi program Coni P2 dikembangkan dengan cara kajian teori, temuan di lapangan, Delphi, FGD, uji coba sebanyak tiga kali; (2) evaluasi di tiga pondok pesantren: Al-Ittifaqiah, Raudhatul Ulum, dan Raudhatul Qur'an ditemukan kesenjangan sarana belajar, kinerja guru, dan motivasi belajar santri; (3) komponen konstruk model evaluasi Coni P2 terdiri atas konteks, input, proses, dan produk, yang terbagi menjadi 13 indikator. Hasil analisis CFA: (1) *Chi Square* (χ^2) = kecil; (2) *p-value* > 0,05; (3) *Root Mean Square Error of Approximation* (RMSEA) < 0,08; dan (4) *Goodness of Fit Index* (GFI) < 0,90.

Kata kunci: *pengembangan, evaluasi, tahfiz al-Qur'an*

DEVELOPING CIPP EVALUATION INSTRUMENT FOR *TAHFIZ* AL-QUR'AN IN PONDOK PESANTREN

Muyasaroh, Sutrisno

Sekolah Tinggi Ilmu Tarbiyah Al-Qur'an Al-Ittifaqiah, UIN Sunan Kalijaga Yogyakarta
muyasnasir@yahoo.com, trisno_63@yahoo.com

Abstract

The study aimed to: (1) generate an evaluation development model of *tahfiz* Al-Qur'an learning program entitled Coni P2; (2) generate a technique of *tahfiz* Al-Qur'an learning program evaluation implementation; and (3) generate component structures and an indicators of evaluation model. The study was a research and development (R&D) type by implementing 9 of 10 steps in Borg and Gall's model. The subject for the first trial the 33 people, for the second trial were 49 people, and for the third trial were 224 people. The implemented component of evaluation model was Stufflebeam Evaluation Model (CIPP). The evaluation steps that the researchers implemented were the ones taken from Malcolm Provus. The data gathering techniques that the researchers implemented were Delphi, FGD, questionnaires, observation, interview and study of documentation. The construct validity was analyzed by implementing CFA and the construct reliability was analyzed by implementing Cronbach Alpha. The results of the research were as follows: (1) the model of Coni P2 evaluation program was developed by implementing theoretical review, field findings, Delphi, FGD and three-time experiments; (2) from the evaluations performed in three pondok pesantren, namely Al-Ittifaqiah, Raudhatul Ulum and Raudhatul Qur'an, the researchers found discrepancy in learning facilities, teacher performance and santri's learning motivation; and (3) the construct components of Coni P2 evaluation model consisted of context, input, process and product that were divided into 13 indicators. The results of CFA analysis were as follows: (1) *Chi Square* (χ^2) = kecil; (2) *p-value* > 0.05; (3) *Root Mean Square Error of Approximation* (RMSEA) < 0.08; dan (4) *Goodness of Fit Index* (GFI) < 0.90.

Keywords: *development, evaluation, tahfiz Al-Qur'an*

Pendahuluan

Al-Qur'an merupakan sebuah kitab suci bagi umat Islam. Keberadaan al-Qur'an tidak hanya sekedar sebuah kitab suci namun juga memiliki kemukjizatan/keistimewaan bahasa, menjadi sumber kekuatan, dan menjadi sebuah sumber pengajaran bagi pendidikan manusia (Zaid, 2011, p.78). Al-Qur'an menjadi sumber ajaran tangguh, menjadi inspirasi sumber metode pengajaran yang baik tentang berpikir dan kreativitas. Demikian mulia kedudukan al-Qur'an sehingga tetap terjaga kemurnian dan keuciannya.

Menghafal adalah proses mengingat informasi yang telah lalu dan dijadikan sebuah informasi baru. Winkel mengemukakan pendapatnya bahwa: "mengingat adalah suatu aktifitas kognitif, dimana orang menyadari bahwa pengetahuan berasal dari informasi atau kesan-kesan yang diperoleh dari masa lampau" (Winkel, 2009, p.73). Hal tersebut senada dengan pendapat Joyce, Weil, & Chaloun, E, (2009, p.223) bahwa kapasitas dalam memperoleh informasi mendukung secara lebih bermakna yang artinya mendapatkan kembali dengan baik hasil pembelajaran yang telah lalu.

Berbagai metode menjaga kemurnian al-Qur'an yang telah dijelaskan, metode yang paling tepat dan baik adalah dengan menghafal al-Qur'an, karena dengan dihafal tidak akan rusak kecuali karena kematian, namun menjaga kemurnian menggunakan media akan dapat rusak pada masa tertentu.

Guna mengungkap permasalahan dalam pembelajaran *tahfiz* al-Qur'an ini sangat dibutuhkan model evaluasi. Model evaluasi program pembelajaran *tahfiz* al-Qur'an sangat diperlukan guna dapat mengungkap seluruh hambatan dalam pembelajaran *tahfiz* al-Qur'an dan guna meningkatkan kualitas pembelajaran *tahfiz* al-Qur'an. Kualitas pembelajaran dipengaruhi oleh tiga hal seperti ungkapan Harto & Abdurrahmansyah, (2011, pp.73-74), kualitas pembelajaran terdiri atas beberapa faktor yakni: (1) ukuran kelas, (2) suasana belajar secara demokratis, dan (3) ketersediaan fasilitas dan sumber belajar.

Belajar akan mencapai kualitas tinggi jika memiliki beberapa faktor belajar. Faktor-faktor yang mempengaruhi belajar siswa menurut pendapat Cruickshank (1990, pp.10-11) ada empat variabel yaitu (1) guru, (2) konteks: siswa, sekolah dan kelas, (3) proses: kinerja guru, semangat dalam mengajar, sikap yang menyenangkan, strategi mengajar, dan perilaku siswa (motivasi), dan (4) produk: jangka pendek (kecakapan/kepandaian) dan jangka panjang (kecakapan profesional atau kecakapan dalam bidang tertentu).

Tujuan Evaluasi program adalah "untuk meninjau kembali atas pencapaian tujuan dan untuk membantu memberikan alternatif berikutnya dalam pengambilan keputusan" (Darussalam, 2010, p.58). Dengan melakukan evaluasi maka teridentifikasi semua hambatan, hasil evaluasi dijadikan alat rekomendasi untuk melakukan perbaikan, setelah perbaikan dari berbagai sektor maka hambatan telah dapat diselesaikan, jika hambatan telah dapat diselesaikan, maka tujuan dari program pembelajaran *tahfiz* al-Qur'an dapat diwujudkan.

Penelitian ini lebih menitikberatkan pada evaluasi program melalui aplikasi model evaluasi dan model pengembangan R&D (*Research and Development*). Dasar pemilihan model ini adalah melihat konsep pembelajaran bahwa kualitas pembelajaran setidaknya dipengaruhi oleh input, proses, dan produk (Issac, Stephen & Michel, & William 198, pp.2-3). Implementasi penelitian dan pengembangan model ini terletak pada konsep CIPP dikombinasikan dengan Provus sehingga menjadi bentuk modifikasi. Belum ditemukan penelitian yang mengevaluasi program pembelajaran *tahfiz* al-Qur'an melalui hasil MTQ/STQ di tingkat propinsi, kabupaten, maupun kecamatan untuk mengetahui kelemahan kualitas *tahfiz* al-Qur'an di pondok pesantren.

Permasalahan yang dihadapi oleh pondok pesantren sangat kompleks. Rasionalnya Stufflebeam & Shinkfield mengacu pendapat Tyler mengungkapkan bahwa: "*evaluation should determine the congruence between performance and objectives*" (Stufflebeam & Shinkfield,

1985, p.70). Evaluasi harus menghasilkan pendapat yang sama antara kinerja dan kondisi secara nyata tidak boleh direayasa dan hindari subjektifitas.

Program pembelajaran seharusnya dievaluasi untuk mengetahui kelemahan dan kekurangan pada proses pembelajaran, kenyataan di lapangan jarang ditemukan evaluasi program yang mengevaluasi tentang pengawasan, kontrol, dan evaluasi *tahfiz* al-Qur'an dari pihak pimpinan pondok pesantren atau kepala lembaga *tahfiz* al-Qur'an terhadap proses pembelajaran *tahfiz* al-Qur'an.

Belum ditemukan pengembangan model evaluasi yang tepat dan dapat digunakan untuk mengevaluasi program pembelajaran *tahfiz* al-Qur'an di pondok pesantren seperti metode menghafal al-Qur'an, standar baku *tahfiz* al-Qur'an, kurikulum yang tersurat, target minimal *tahfiz* al-Qur'an, minat santri untuk menghafal al-Qur'an dan produk *tahfiz* al-Qur'an.

Belum ditemukan pengembangan model evaluasi menggunakan teknik evaluasi secara sistematis dengan melihat kesenjangan yang dapat digunakan untuk mengevaluasi pelaksanaan program pembelajaran *tahfiz* al-Qur'an di pondok pesantren.

Belum ditemukan komponen dan indikator model evaluasi program yang terdiri atas empat komponen (konteks, input, proses, dan produk) dan terbagi menjadi 13 indikator yang dapat digunakan sebagai acuan penyusunan instrumen model evaluasi program pembelajaran *tahfiz* al-Qur'an di pondok pesantren.

Berangkat dari berbagai permasalahan maka dapat diformulasikan rumusan masalah sebagai berikut. Bagaimanakah model evaluasi program pembelajaran *tahfiz* al-Qur'an Coni P2 dikembangkan agar dapat meningkatkan kualitas pembelajaran *tahfiz* al-Qur'an di pondok pesantren?. Bagaimana teknik pelaksanaan evaluasi program pembelajaran *tahfiz* al-Qur'an secara tepat menggunakan model evaluasi program pembelajaran *tahfiz* al-Qur'an Coni P2 di pondok pesantren?. Bagaimanakah komponen-komponen dan indikator-indikator model eva-

luasi program pembelajaran agar dapat digunakan sebagai acuan penyusunan instrumen evaluasi program pembelajaran *tahfiz* al-Qur'an di pondok pesantren?.

Tujuan Penelitian adalah menghasilkan instrumen model evaluasi program pembelajaran *tahfiz* al-Qur'an yang diberi nama Coni P2 agar dapat meningkatkan kualitas program *tahfiz* al-Qur'an di pondok pesantren, menghasilkan model teknik pelaksanaan evaluasi program pembelajaran *tahfiz* al-Qur'an secara tepat menggunakan model evaluasi Coni P2, menghasilkan struktur komponen-komponen dan indikator-indikator model evaluasi sebagai acuan penyusunan instrumen evaluasi program pembelajaran *tahfiz* al-Qur'an di pondok pesantren.

Penelitian ini memiliki kegunaan untuk membangun sebuah model instrumen untuk mengevaluasi program pembelajaran *tahfiz* al-Qur'an melalui model modifikasi evaluasi program antara model CIPP dan Provus disebut model evaluasi "Coni P2". Penelitian ini juga berguna untuk mengevaluasi program *tahfiz* al-Qur'an untuk membantu memberi bantuan pengambilan keputusan kepada pengelola program pembelajaran *tahfiz* al-Qur'an di pondok pesantren apakah program ini harus dihentikan, dilanjutkan atau diperbaiki. Worthen & Sanders menjelaskan bahwa: "*evaluation is the process of delineating, obtaining, and providing useful information for judging decision alternatives*" (Worthen & Sanders, 1981, p.129).

Metode Penelitian

Jenis penelitian ini adalah penelitian menggunakan pendekatan prosedur R&D ini menghasilkan: (1) instrumen untuk mengevaluasi program pembelajaran *tahfiz* al-Qur'an; (2) model evaluasi program *tahfiz* al-Qur'an yang diberi nama "Coni P2"; dan (3) buku panduan penggunaan model evaluasi hasil pengembangan model. Tahap penelitian dan pengembangan dalam penelitian ini menggunakan model pengembangan Borg & Gall (1989, pp.772-781). Langkah *develop* teknik pelaksanaan dalam

penelitian ini menggunakan tahap 1-9 dari 10 langkah R & D Borg & Gall. Alasan menggunakan model pengembangan adalah tahap *develop* Borg & Gall lebih rinci dan lebih sistematis.

Siklus perencanaan, tahap ini terdiri atas tiga langkah yang harus dilakukan, pertama penelitian dan pengumpulan informasi. Pada langkah ini peneliti melakukan survei awal di empat pondok pesantren yang mengelola pembelajaran *tafizi* al-Qur'an di Kabupaten Ogan Ilir, Sumatera Selatan.

Siklus pengujian ada empat langkah yang harus dilalui. Langkah pertama, uji coba pendahuluan menggunakan subjek uji coba 33 orang, langkah kedua, melakukan revisi berdasarkan masukan-masukan dari lapangan, langkah ketiga, uji coba utama dengan subjek uji coba 49 orang, langkah keempat, dilakukan revisi sesuai masukan dari lapangan.

Model evaluasi dan perangkat evaluasi diimplementasikan di pondok pesantren dalam skala yang lebih besar. Menurut teori Borg & Gall pada tahap ini ada tiga langkah yang harus dilalui, pertama uji coba lapangan operasional melibatkan tiga pondok pesantren terdiri 224 subjek uji, kedua melakukan revisi.

Observasi dilakukan untuk melihat kondisi pembelajaran. Kaplan dalam (Keeves & Lakomski, 1992, p.84) menjelaskan bahwa, data hasil observasi sendiri harus valid artinya apa yang telah diamati semuanya harus dilaporkan secara benar jangan sampai tidak melaporkan hasil pengamatan.

Penelitian ini dimulai dengan pra-survei tanggal 2-8 Juli 2011- 23 Oktober 2013. Penelitian terbagi menjadi dua bagian: (1) tahap observasi bertujuan untuk melihat kondisi objektif lokasi penelitian dan permasalahan yang dihadapi dilaksanakan pada tanggal 2-8 Juli 2011 dan (2) tahap penelitian selanjutnya pengambilan data lapangan pada uji coba pertama (10-20 Juli 2013), uji coba kedua (28-4 Agustus 2013), dan uji coba ketiga (13-23 Oktober 2013). Rekap nilai dan pengolahan data 24 Oktober-31 Desember 2014. Selama kurun waktu bulan

Juli 2011 hingga 22 Oktober 2013 peneliti selalu menjaga hubungan baik dengan selalu berkomunikasi dengan keempat pondok pesantren tersebut untuk memantau perkembangan yang ada di masing-masing pondok pesantren.

Penelitian ini dilakukan di empat pondok pesantren di kabupaten Ogan Ilir (OI), Sumatera Selatan. Ogan Ilir (OI) merupakan kabupaten/kota pemekaran dari kabupaten Ogan Komering Ilir (OKI). Usia yang masih muda ini (10 tahun) kabupaten ini memiliki 16 kecamatan termasuk kecamatan induk. Kabupaten OI memiliki 21 pondok pesantren yang memiliki program *tafizi* al-Qur'an hanya ada 4 pondok pesantren antara yaitu: (a) Pondok Pesantren Al-Ittifaqiah kecamatan Idralaya Induk, (2) Pondok Pesantren Raudhatul Ulum Indralaya Induk, (3) pondok pesantren Raudhatul-Qur'an, Payaraman Barat, (4) Pondok Pesantren Darul Funun Kecamatan Rambang Kuang.

Subjek penelitian menggunakan beberapa subjek penelitian. Validasi instrumen menggunakan empat orang pakar yaitu pakar evaluasi, pakar pengukuran, pakar metodologi penelitian, dan pakar *tafizi* al-Qur'an. *Focus Group Discussion* (FGD) menggunakan subjek uji 19 orang terdiri atas 2 orang pimpinan pondok pesantren, 7 orang guru *tafizi* al-Qur'an, dan 10 orang santri *tafizi* al-Qur'an. Uji keterbacaan instrumen menggunakan 15 orang validator guna memeriksa isi dari instrumen, susunan bahasa dan komponen dari instrumen. Uji coba pertama menggunakan subjek penelitian sebanyak 33 orang uji subjek yang terdiri atas 2 orang pakar evaluasi, 4 orang pakar al-Qur'an, 2 orang pakar metodologi, 2 orang pengurus forum pondok pesantren sebagai pengawas pondok pesantren, 10 orang guru *tafizi* al-Qur'an, dan 13 orang santri *tafizi* al-Qur'an. Uji coba kedua menggunakan subjek uji 49 orang terdiri atas 1 orang pimpinan pondok pesantren, 12 orang guru *tafizi* al-Qur'an, dan 36 orang santri *tafizi*, dan uji coba ketiga dilakukan di 3 pondok pesantren yaitu pondok pesantren Al-Ittifaqiah, Raudhatul Ulum, dan Raudhatul Qur'an. Subjek uji coba terdiri atas pim-

pinan pondok pesantren 3 orang, guru *tabfiẓ* al-Qur'an 20 orang, dan santri 201 orang, jumlah seluruh 224 Orang.

Langkah-langkah pengembangan model Evaluasi program pembelajaran *tabfiẓ* al-Qur'an menggunakan 9 langkah dari 10 langkah model Borg & Gall (1989, pp.772-7781); (1) penelitian dan pengumpulan informasi. Pengumpulan informasi dalam penelitian ini dituntut untuk mulai mengumpulkan informasi secara rinci dan mengumpulkan sejumlah permasalahan yang ditemukan dalam penelitian di lapangan (Creswell (2009, p.97). Pengumpulan informasi dengan melakukan tinjauan literatur, mengkaji beberapa konsep penelitian dan pengembangan (R&D), melakukan penelitian pendahuluan di empat pondok pesantren; (2) perencanaan meliputi merencanakan produk yang akan dihasilkan, waktu penelitian, penulisan draf awal; (3) pengembangan bentuk produk awal berupa bentuk penyusunan instrumen, bentuk model evaluasi, dan panduan penggunaan model evaluasi; (4) uji lapangan pendahuluan dengan skala terbatas yaitu melakukan uji coba lapangan dengan subjek uji coba sebanyak 33 orang uji subjek yang terdiri atas 2 orang pakar evaluasi, 4 rang pakar al-Qur'an, 2 orang pakar metodologi, 2 orang pengurus forum pondok pesantren sebagai pengawas pondok pesantren, 10 orang guru *tabfiẓ* al-Qur'an, dan 13 orang santri *tabfiẓ* al-Qur'an; (5) revisi produk, dilakukan perbaikan sebagaimana yang telah diusulkan oleh hasil uji lapangan pendahuluan; (6) uji coba lapangan utama/diperluas dilakukan di satu pondok pesantren yaitu pondok pesantren Darul Funun yang terdiri atas 1 orang pimpinan pondok pesantren, 12 orang guru *tabfiẓ* al-Qur'an, dan 36 orang santri *tabfiẓ*, jumlah seluruh 49 orang; (7) revisi produk, melakukan perbaikan instrumen berdasarkan masukan dan saran sebagaimana diusulkan oleh hasil uji lapangan utama. (8) Uji coba lapangan operasional dilakukan di 3 pondok pesantren yaitu pondok pesantren Al-Ittifaqiah, Raudhatul Ulum, dan Raudhatul Qur'an. Subjek uji coba terdiri atas pimpinan pondok pesantren 3 orang, guru

tabfiẓ al-Qur'an 20 orang, dan santri 201 orang, jumlah seluruh 224 Orang. (9) Revisi produk akhir, melakukan perbaikan sebagaimana yang diusulkan dalam uji lapangan operasional, yaitu melakukan perbaikan dan penyempurnaan akhir pada instrumen model yang dikembangkan sehingga menjadi produk akhir hasil dari pengembangan (*final product*).

Data yang diperoleh dalam penelitian ini adalah data kuantitatif dan kualitatif. Data tersebut memberi gambaran tentang program pembelajaran *tabfiẓ* al-Qur'an meliputi konteks, input, proses, dan produk pelaksanaan program yang dicapai. Data yang diperoleh dijelaskan di bawah ini.

Data kuantitatif diperoleh dari instrumen angket/kuesioner teknik Delphi merupakan pendapat dari para pakar dan para praktisi untuk membuat *judgement* berupa hasil kesepakatan tentang konseptualisasi program pembelajaran *tabfiẓ* al-Qur'an di pondok pesantren. Kuesioner juga digunakan untuk menjaring data tentang perencanaan program, pelaksanaan program, dan hasil dari pelaksanaan program pembelajaran *tabfiẓ* al-Qur'an melalui pimpinan pondok pesantren, guru, dan santri.

Data kualitatif diperoleh dari hasil data dokumentasi pondok pesantren, studi observasi, wawancara, penilaian pengamatan langsung di kelas saat guru mengajar di kelas. Seluruh data yang berhasil dihimpun akan diolah guna untuk membuat keputusan tentang status program apakah akan diputuskan program tetap berjalan, dihentikan atau dimodifikasi dan juga untuk menentukan rekomendasi untuk pondok pesantren.

Teknik pengumpul data melalui beberapa teknik diantaranya: (1) Teknik Delphi, teknik ini merupakan sebuah peramalan secara interaktif dan sistematis berdasarkan *feedback* secara individu dari nilai-nilai yang dipilih sebanyak satu putaran atau lebih. Bukti validitas isi tes atau instrumen dilakukan oleh panel pakar sesuai bidang yang diukur dan pakar bidang pengukuran (Mardapi, 2012, pp.17-19). Teknik ini dilakukan dengan mengedarkan draf awal instrumen menggunakan skala likert dengan

pilihan jawaban sangat cocok, cocok, cukup cocok, kurang cocok, dan tidak cocok. Selain menggunakan pilihan jawaban tersebut para pakar juga diminta usulan, saran, dan pendapat pada lembar asulan, saran, dan pendapat; (2) Teknik FGD, teknik ini merupakan metode partisipasi dalam pengumpulan informasi mengenai suatu permasalahan dan kebutuhan tertentu yang sangat spesifik melalui diskusi secara kelompok dan untuk mendapatkan kesepakatan bersama. Mertens (2010, p.240) menjelaskan bahwa FGD dapat digunakan sebagai alat pengumpul data atau merupakan strategi penelitian; (3) uji keterbacaan instrumen hasil pembelajaran *tahfīz* al-Qur'an difokuskan pada penilaian: (a) kejelasan indikator kecakapan *tahfīz* al-Qur'an, (b) kejelasan indikator kecakapan individu, (c) kejelasan indikator sikap sosial, (d) penggunaan bahasa Indonesia secara baku, (e) perumusan pertanyaan, dan (f) penggunaan kata dan kalimat yang mudah dipahami serta metode tata tulis yang tepat dan benar; (4) uji coba pertama dengan melibatkan skala kecil yang mewakili dari seluruh pondok pesantren untuk menguji perangkat evaluasi, model evaluasi, dan panduan evaluasi; (5) uji coba kedua dalam skala lebih besar melibatkan satu pondok pesantren untuk menguji perangkat evaluasi, model evaluasi, dan panduan evaluasi; (6) uji coba ke-3 dalam skala paling besar melibatkan 3 pondok pesantren untuk menguji perangkat evaluasi, model evaluasi, dan panduan evaluasi.

Metode instrumen pengumpulan data menggunakan beberapa macam; (1) metode tes, metode ini digunakan untuk mengumpulkan data tentang *output* pembelajaran *tahfīz* al-Qur'an kompetensi psikomotorik (keterampilan) dalam menghafal al-Qur'an. Bentuk instrumen berupa test *tahfīz* al-Qur'an digunakan untuk menjangkau kemampuan pemahaman praktik keterampilan menghafal materi hafalan, penerapan hukum tajwid *adab* dan *faṣāḥah*, metode ini menggunakan nilai raport *tahfīz* al-Qur'an. (2) angket/kuesioner, teknik ini dapat memberikan informasi penting dan jelas tentang kualitas, nilai atau keefektifan program

(Davidson, 2005, p.228). Dalam penelitian ini metode angket/kuisisioner digunakan untuk menjangkau sikap responden tentang komponen proses pembelajaran *tahfīz* al-Qur'an yang meliputi, minat, latar belakang, kesulitan yang dialami, dan prestasi yang dicapai. Cosby menyatakan, dalam kuisisioner, pertanyaan-pertanyaan dihadirkan dalam format tertulis dan para responden menuliskan jawaban-jawaban mereka (Cosby, 2009, pp.214-215). Kuisisioner diisi oleh pimpinan, guru, dan santri untuk memudahkan dalam menjawab karena sudah disiapkan jawaban tinggal memilih yang dirasa tepat; (3) observasi, observasi bertujuan untuk melakukan pengamatan dan proses pembelajaran *tahfīz* al-Qur'an serta melakukan pencatatan secara sistematis dan logis, objektif dan rasional (Arifin, 2009, p.153); (4) wawancara, wawancara adalah pengajuan pertanyaan secara lisan oleh *interviewer* kepada responden dan dijawab secara lisan pula oleh responden. Gall, Gall & Borg (2007, p.228) menjelaskan bahwa "*interviewer consist of oral question asked by the interviewer and oral responses by the research participants*". Wawancara digunakan untuk mewawancarai pimpinan, guru, dan santri. Wawancara diperlukan untuk dapat mengungkap hal yang masih tersembunyi, yang masih tertutup pada saat melakukan observasi. Tujuan wawancara adalah memberi kebebasan kepada pimpinan, guru dan santri untuk mengungkap informasi hingga informasi tersulit akan dapat dijelaskan; (5) dokumentasi, metode ini digunakan untuk menggali informasi melalui dokumen lembaga *tahfīz* al-Qur'an seperti nilai rapor, jumlah perkembangan santri tiap tahun, jumlah wisudawan setiap tahun, dan jumlah santri berprestasi setiap tahun.

Data hasil penelitian dianalisis dengan cara mengorganisasikan data ke dalam kategori, menjabarkan ke dalam unit-unit, melakukan sintesis, menyusun ke dalam pola. Data dibagi menjadi dua yaitu kualitatif dan kuantitatif. Data kualitatif adalah data yang dinyatakan dalam bentuk kata, kalimat, dan gambar. Data kuantitatif adalah data yang berbentuk angka atau data kualitatif yang diangkakan (Sugiyono, 2010, pp.3-4). Ada-

pun tahapan-tahapan dapat dilihat sebagai berikut.

Analisis data secara kuantitatif, data yang terkumpul dianalisis dengan *Exploratory Factor Analysis* (EFA) dengan menggunakan bantuan program SPSS 16.0. Variabel laten dibentuk berdasarkan konsep teoretis dengan beberapa indikator/manifest (Ghozali 2008, p.121). Tujuan penggunaan EFA ini adalah untuk menguji apakah konstruk yang telah dibentuk dapat dijelaskan oleh indikator-indikatornya atau tidak. Apabila indikator-indikator tersebut dapat menjelaskan konstruk atau variabel yang ada, maka kebenaran tersebut ditunjukkan pada nilai *loading factor* yang tinggi. Setiap butir harus memiliki muatan lebih besar dari 0,3 (>0,3), demikian juga sebaliknya jika butir tersebut memiliki muatan lebih kecil dari 0,3 (<0,3) maka butir tersebut dinyatakan tidak valid, maka harus dibuang atau gugur. Nurosis (1986, p 123). Konstruk diuji dengan melihat nilai *Kaiser Mayer Olkin* (KMO) diharapkan 0,5 (> 0,5). Analisis menggunakan EFA untuk menguji validitas pada uji coba kedua dan digunakan untuk uji normalitas konstruk sebelum dianalisis menggunakan CFA pada uji coba ketiga.

Reliabilitas dilihat dari nilai alphanya jika nilai alpha > 0,7 maka instrumen dianggap reliabel dan jika kurang dari 0,7 dianggap tidak reliabel. Pengujian menggunakan Cronbach Alpha digunakan untuk menguji reliabilitas instrumen pada uji coba kedua dan ketiga.

Hasil uji coba ketiga data dianalisis menggunakan metode CFA dengan tujuan untuk melihat validitas dari masing-masing instrumen yaitu konteks, input, proses, dan produk. Analisis menggunakan CFA validitas dilihat dari nilai t (*t-value*), muatan faktor dilihat dari nilai alpha, jika nilai alpha > 0,03 (tarap signifikan 5 %) dibanding dengan nilai t_{tabel} 1,96 maka butir dinyatakan valid. Valid dapat dibuktikan pada diagram *path* ditunjukkan dengan tulisan angka hitam, semakin besar nilai t hitung maka semakin valid. Jika nilai alpha tertulis merah

maka dinyatakan bahwa butir tersebut tidak valid sehingga harus di-drop.

Uji kecocokan model antara lain dilihat RMSEA harus dibawah 0,08 dan GFI harus lebih besar dari 0,90, (Wijanto, 2008, pp.154 & 193). Uji kecocokan antara model teoretis dan model empiris didasarkan pada empat kategori; (1) *Chi-Square*, (2) *Signifikan Probability*, (3) *Root Mean Square of Error Approximation* (RMSEA), dan *Goodness of Fit Index* (GFI), Standar yang digunakan dapat dilihat pada tabel nomor 1 berikut ini.

Tabel 1. Goodness of Fit Index

| No | Indeks | Cut of value | Keterangan |
|----|------------------------------------|--------------|----------------------------------|
| 1 | <i>Chi-Square</i> (X^2) | Kecil | Semakin kecil X^2 semakin baik |
| 2 | <i>Probability</i> (p) | ≥ 0,05 | Harus lebih besar dari 0,05 |
| 3 | RMSE | ≤ 0,08 | Rata-rata perbedaan/df |
| 4 | <i>Goodness of Fit Index</i> (GFI) | ≥ 0,90 | Nilai berkisar 0-1 |

Sumber Wijanto, 2008, p.61

Data hasil observasi terdiri atas data penilaian proses pembelajaran *tahfiz* al-Qur'an yang terdiri atas data hasil observasi santri putra dan putri secara terpisah dari tiga pondok pesantren: Pondok Pesantren Al-Ittifaqiah, Raudhatul Ulum, dan Raudhatul Qur'an. Data observasi masing-masing diambil dari dua orang penilai. Untuk menjaga reliabilitas antarkedua penilai maka data dianalisis menggunakan metode Cohen's Kappa menggunakan bantuan komputer dengan program SPSS *for Window* 16.0. Reliabilitas didasarkan pada hasil hitungan Kappa, jika hasil hitungan memiliki nilai k < 0,40 (jelek), 0,40 ≤ k ≤ 0,75 (bagus), dan k > 0,75 (sempurna) (Eye & Mun, 2005, p.6). Standar inilah yang digunakan untuk melihat kesepakatan antardua penilai hasil observasi.

Analisis secara deskriptif analisis dilakukan dengan melihat kelengkapan model, kejelasan panduan model evaluasi, kelengkapan instrumen dan analisis keefektifan model. Analisis model terdiri dari empat

kriteria yang harus dipenuhi, yaitu komprehensif, praktis, ekonomis dan didukung oleh instrumen yang valid dan reliable. Keempat kriteria syarat model tersebut hanya validitas dan reliabilitas yang dianalisis secara kuantitatif sedangkan komprehensif, praktis, dan ekonomis akan dianalisis secara kualitatif.

Cara menganalisis data secara deskriptif, data kuantitatif yang berasal dari instrumen dicari skor reratanya, kemudian dikonfirmasi ke dalam data kualitatif dengan menggunakan skala 5, kemudian dideskripsikan. Hasil diskripsi dijadikan dasar untuk penilaian apakah instrumen, model evaluasi yang dikembangkan, dan panduan telah memenuhi syarat pengembangan model yang digunakan untuk mengevaluasi program pembelajaran *tafizi* al-Qur'an di pondok pesantren. Bentuk konferensi data kuantitatif menjadi kualitatif menggunakan skala 5 ini mengacu terhadap model modifikasi Sudijono (2003, pp.329-339). Ketentuan Sudijono dijelaskan Tabel 2.

Analisis data secara kualitatif adalah menganalisis data hasil validasi (penilaian) dari para ahli (*expert*) dan pengguna model evaluasi (pimpinan pondok pesantren) serta praktisi yang memberi masukan-masukan dalam rangka perbaikan model evaluasi dan perangkatnya. Penelitian ini menggunakan pendekatan deskriptif kualitatif yaitu mendeskripsikan atau memberi gambaran tentang objek yang diteliti melalui data sampel atau populasi sebagaimana adanya tanpa membuat analisis ataupun kesimpulan yang berlaku untuk umum (Sugiyono, 2006, p.21) melalui wawancara, pengolahan angket, dan dokumentasi.

Data yang telah dikumpulkan dianalisis melalui tiga jenis kegiatan yang terjadi secara bersamaan yaitu: reduksi data, penyajian data, dan penarikan kesimpulan/verifikasi (Miles & Huberman 1985, pp.21-22). Analisis data secara kualitatif menggunakan tampilan berupa narasi, tabel, dan grafik meliputi data perkembangan jumlah santri selama tiga tahun, perkembangan nilai nilai raport selama tiga semester, perkembangan jumlah wisudawan-wisudawati sela-

ma tiga tahun, dan perkembangan hasil prestasi MTQ/STQ santri selama tiga tahun.

Tabel 2. Konferensi Data Kuantitatif ke Kualitatif

| Rumus | Rerata skor | Klasifikasi |
|--|-------------|---------------|
| $X > x_i + 1,8 \times sb_i$ | > 4,2 | Sangat baik |
| $X_i + 0,6 \times sb_i < X \leq i + 1,8 \times sb_i$ | > 3,4-4,2 | Baik |
| $X_i - 0,6 \times sb_i < X \leq i + 0,6 \times sb_i$ | >2,6-3,4 | Cukup |
| $X_i - 1,8 \times sb_i < X \leq i - 0,6 \times sb_i$ | >1,8-2,6 | Kurang |
| $X \leq X_i - 1,8 \times sb_i$ | $\leq 1,8$ | Sangat kurang |

Hasil Penelitian dan Pembahasan

Hasil uji coba pertama instrumen diujikan kepada 33 orang pakar dan praktisi yaitu 2 orang pakar evaluasi, 4 orang pakar al-Qur'an, 2 orang pakar metodologi, 2 orang pengawas forum pondok pesantren tingkat kabupaten, 10 orang guru *tafizi* al-Qur'an, dan 13 orang santri *tafizi* al-Qur'an. Penilaian yang digunakan adalah skala likert dengan lima pilihan jawaban. Skor minimal 1 dan skor maksimal 5.

Berdasarkan penilaian para pakar dan praktisi terhadap instrumen menunjukkan bahwa instrumen yang telah diujicobakan adalah instrumen yang baik. Kriteria baik tersebut ditunjukkan pada nilai rerata dari seluruh komponen penilaian instrumen menunjukkan hasil nilai 4,0 posisi ini menunjukkan pada rentang skor di antara 3,4-4,2 artinya instrumen sudah baik dan layak digunakan untuk melakukan uji coba kedua pada penilaian selanjutnya.

Hasil penilaian model evaluasi uji coba pertama dibagikan kepada 33 orang terdiri atas 2 orang pakar evaluasi, 4 orang pakar al-Qur'an, 2 orang pakar metodologi, 2 orang pengawas forum pondok pesantren tingkat kabupaten, 10 orang guru *tafizi* al-Qur'an, dan 13 orang santri. Penilaian yang digunakan adalah skala likert dengan lima pilihan jawaban. Skor minimal 1 dan skor maksimal 5. Hasil penilaian uji coba per-

tama dianalisis secara deskriptif. Hasil uji coba pertama model evaluasi program pembelajaran *tahfiz* al-Qur'an adalah baik, kriteria baik tersebut ditunjukkan pada nilai rerata total 4,11 posisi ini menunjukkan pada rentang skor di antara 3,4-4,2 artinya instrumen sudah baik dan layak digunakan untuk melakukan uji coba kedua.

Uji coba kedua dilakukan di pondok pesantren Darul Funun dengan melibatkan subjek uji coba terdiri atas 1 pimpinan pondok pesantren, 12 orang guru *tahfiz* al-Qur'an, dan 36 santri *tahfiz* al-Qur'an, jumlah seluruh subjek uji coba berjumlah 49 orang. Berdasarkan penilaian terhadap instrumen mengalami sedikit peningkatan skor total pada uji coba pertama adalah 4,09 dan pada uji coba kedua menjadi 4,11 kriteria baik karena terletak rentang skor antara 3,4 - 4,2 artinya instrumen sudah baik.

Berdasarkan penilaian para pakar terhadap model menunjukkan bahwa model evaluasi yang telah diujicobakan adalah model yang baik, kriteria baik tersebut ditunjukkan pada nilai rerata dari seluruh komponen menunjukkan hasil nilai 4,17, posisi ini menunjukkan pada rentang skor di atas 3,4-4,2 artinya model ini sudah layak digunakan tanpa perbaikan untuk melakukan uji coba ketiga.

Validitas instrumen konteks, input, proses, dan produk semua butir menunjukkan bahwa butir-butir tersebut telah memenuhi kriteria *Kaiser Meiyer Olkin Measure of Sampling Adequacy* (KMO) > 0,5 sehingga sudah memenuhi kriteria. Uji *Barletts Tes of Sphericity* nilai *Chi-Square*, df seperti ditampilkan pada Tabel 3 dan signifikansi dari semua konstruk sudah di bawah 0,00 artinya sudah signifikan. Nilai semua butir berada pada di atas 0,3. Dengan demikian, instrumen konteks dengan faktor-faktornya memiliki butir-butir yang valid semuanya karena telah memiliki nilai muatan faktor di atas 0,3.

Estimasi reliabilitas instrumen konteks dilakukan dengan menggunakan rumus *Linier Combination of Reliability* yaitu masing-masing indikator dilihat nilai koefisiennya kemudian koefisien dari seluruh indikator

tersebut dilihat koefisien rata-ratanya. Rata-rata nilai koefisien muatan faktor pada instrumen konteks, input, proses, dan produk adalah menunjukkan pada posisi lebih dari 0,7. Dengan demikian, instrumen konteks dinyatakan reliabel sehingga dapat digunakan untuk mengevaluasi program pembelajaran *tahfiz* al-Qur'an. Hasil analisis reliabilitas selengkapnya dapat dilihat pada Tabel 4.

Tabel 3. Hasil Analisis Faktor untuk Instrumen pada Uji Coba Kedua

| Komponen | Hasil Hitungan | | |
|----------|----------------|----------------|----------|
| Konteks | KMO | 0,694 | |
| | Barlet's | X ² | 781,037 |
| | | Df | 190 |
| | | Sig | 0,00 |
| Input | KMO | 0,769 | |
| | Barlet's | X ² | 1358,042 |
| | | df | 378 |
| | | Sig | 0,00 |
| Proses | KMO | 0,612 | |
| | Barlet's | X ² | 2260,450 |
| | | df | 780 |
| | | Sig | 0,00 |
| Produk | KMO | 0,655 | |
| | Barlet's | X ² | 2490,850 |
| | | df | 561 |
| | | Sig | 0,00 |

Tabel 4. Reliabilitas Cronbach Alpha Uji coba kedua

| Komponen | Koefisien Alpha | Keterangan |
|----------|-----------------|------------|
| Kontek | 0,994 | Reliabel |
| Input | 0,952 | Reliabel |
| Proses | 0,954 | Reliabel |
| Produk | 0,971 | Reliabel |

Hasil evaluasi Program *tahfiz* al-Qur'an di Pondok Pesantren Darul Funun, rerata total skor evaluasi konteks, input, proses, dan produk menghasilkan nilai 4,08. Nilai tersebut dikonfirmasi pada standar kenfersi penilaian maka nilai tersebut berada pada rentang skor nilai antara 3,4-4,2 dengan klasifikasi "baik".

Hasil Evaluasi menyatakan ter-dapat sedikit kesenjangan pada kinerja guru *tahfiz* al-Qur'an dan motivasi belajar santri keduanya memiliki nilai di bawah angka 4 sekalipun kriteria baik. Demikian juga kompetensi *tahfiz* al-Qur'an kompetensi akhlak mulia masih memiliki nilai rendah masih di bawah angka 4. Dengan demikian, direkomendasikan (1) tingkatkan kinerja guru terutama penguasaan materi oleh guru dan metode mengajar; (2) bangun motivasi agar semangat menghafal al-Qur'an; (3) tingkatkan kompetensi *tahfiz* al-Qur'an santri; dan (4) tingkatkan pembinaan akhlak santri berdasarkan al-Qur'an yang dihafal.

Uji Coba Ketiga di Tiga Pondok Pesantren

Penilaian instrumen pada uji coba ketiga mengalami peningkatan. Hasil penilaian instrumen uji coba pertama memiliki rerata total 4,09 artinya instrumen evaluasi program ini sudah baik. Hasil uji coba kedua mengalami peningkatan rerata skor total yaitu 4,11, nilai ini memiliki klasifikasi tetap "baik". Instrumen diperbaiki sesuai dengan saran dan rekomendasi dari para validator dan terbukti hasil uji coba ke tiga instrumen mengalami peningkatan yang signifikan yaitu dengan nilai 4,21. Nilai ini berada pada klasifikasi sangat baik. Dengan demikian, dari uji coba pertama, kedua, dan ketiga menunjukkan bahwa nilai hasil penilaian oleh validator mengalami peningkatan hingga mencapai "sangat baik" sehingga instrumen dinyatakan layak untuk digunakan mengevaluasi program pembelajaran *tahfiz* al-Qur'an di pondok pesantren dimanapun berada.

Penilaian model evaluasi memiliki rerata total skor 4,20 menunjukkan bahwa nilai naik tiga angka dibanding penilaian model kedua, nilai tersebut berada pada kriteria evaluasi yang "baik". Artinya hasil penilaian ini berada pada puncak kriteria baik hampir mencapai kriteria "sangat baik" jika nilai satu angka lagi. Model ini dapat digunakan untuk mengevaluasi program pembelajaran *tahfiz* al-Qur'an di pondok pesantren.

Hasil evaluasi program pembelajaran *tahfiz* al-Qur'an di tiga Pondok Pesantren yaitu PP. Al-Ittifaqiah, PP Radhatul Ulum, dan PP. Raudhatul Qur'an dengan analisis evaluasi kontek memiliki nilai 4,21, input memiliki nilai 4,23, proses memiliki nilai 4,19, dan produk memiliki nilai 4,23. Rerata total dari evaluasi kontek, input, proses, dan produk pada program *tahfiz* al-Qur'an memiliki rerata 4,22 kriteria "sangat baik". Hasil selengkapnya dapat dilihat pada Tabel 5.

Tabel 5. Rangkuman Hasil Evaluasi Program *Tahfiz* Al-Qur'an di 3 Pondok Pesantren

| Komponen Evaluasi | Pondok Pesantren | | | Rerata Komponen Evaluasi | Rerata Total |
|-------------------|------------------|------|------|--------------------------|-----------------------|
| | RQ | RU | PPI | | |
| Konteks | 4,21 | 4,22 | 4,21 | 4,21 | 4,22 (Sangat baik) |
| Input | 4,24 | 4,24 | 4,21 | 4,23 | |
| Proses | 4,16 | 4,22 | 4,21 | 4,19 | |
| Produk | 4,29 | 4,19 | 4,21 | 4,23 | |

Hasil pelaksanaan evaluasi program pembelajaran *tahfiz* al-Qur'an di tiga pondok pesantren yaitu Pondok Pesantren Al-Ittifaqiah, Raudhatul Ulum, dan Raudhatul Qur'an yang terdiri atas komponen konteks, input, proses, dan produk kemudian dibandingkan dengan standar evaluasi kesenjangan. Hasilnya ditemukan 6 macam kesenjangan tiga kesenjangan di Pondok Pesantren Radhatul Qur'an yakni pertama santri tidak dilakukan seleksi kemampuan dasar, kedua, dan ketiga hasil penilaian kinerja guru dinilai oleh pimpinan dan dinilai oleh guru sendiri sama-sama memiliki nilai agak rendah sekalipun kriteria baik, namun perlu diperhatikan agar meningkatkan kualitas kinerja guru melalui metode dan strategi mengajar. Tiga kesenjangan ditemukan di tiga pondok pesantren yaitu sarana belajar menghafal tidak memiliki kelas khusus untuk menghafal al-Qur'an semua tidak memiliki tempat khusus setoran hafalan, namun hal tersebut tidak mengurangi kualitas pembelajaran *tahfiz* al-Qur'an sehingga kesenjangan tersebut tidak terlalu dihiraukan, memang idealnya pon-

dok pesantren menyiapkan lokal secara permanen khusus untuk kegiatan *tahfiz* al-Qur'an.

Pengambilan keputusan tentang program pembelajaran *tahfiz* al-Qur'an di tiga pondok pesantren didasarkan atas keterlaksanaan program sesuai dengan standar evaluasi yang disepakati yaitu standar konteks, input, proses, dan produk. Hasil evaluasi program pembelajaran *tahfiz* al-Qur'an menggunakan model evaluasi Stufflebeam (CIPP) dan model evaluasi Malcolm Provus (*Descripancy*) memutuskan bahwa: (1) program *tahfiz* al-Qur'an di Pondok Pesantren Al-Ittifaqiah dapat dilanjutkan, dengan perbaikan sarana belajar; (2) program *tahfiz* al-Qur'an di Pondok Pesantren Raudhatul Ulum dapat dilanjutkan dengan perbaikan sarana belajar; dan (3) program *tahfiz* al-Qur'an di Pondok Pesantren Raudhatul Qur'an program *tahfiz* al-Qur'an tetap dilanjutkan dengan perbaikan pada rekrut santri hendaknya dilakukan seleksi kemampuan dasar menghafal al-Qur'an, perbaikan kinerja guru, dan perbaikan sarana belajar *tahfiz* al-Qur'an.

Penilaian pada panduan penggunaan model evaluasi uji coba pertama memiliki nilai 4,11 menunjukkan bahwa nilai tersebut berada pada kriteria baik. Pada uji kedua skor total mengalami peningkatan 5 angka namun masih berada pada posisi kriteria baik, dan uji coba ketiga nilai skor total 4,21 penilaian mengalami peningkatan hingga 5 angka kriteria "sangat baik". Dengan demikian, penilaian yang berdasarkan ketujuh kriteria dari 3 kali uji coba menunjukkan bahwa panduan ini "sangat baik" yang bersifat objektif, praktis, dan efisien sehingga cocok dan tepat digunakan untuk mengevaluasi program pembelajaran *tahfiz* al-Qur'an di pondok pesantren.

Guna menjawab pertanyaan bagaimanakah model evaluasi program pembelajaran *tahfiz* al-Qur'an Coni P2 dikembangkan agar dapat digunakan untuk mengevaluasi program pembelajaran *tahfiz* al-Qur'an di pondok pesantren? maka diperlukan deskripsi tahapan-tahapan pengembangan model evaluasi secara sistematis dan praktis

yang dapat digunakan oleh pimpinan pondok pesantren untuk mengevaluasi program *tahfiz* al-Qur'an. Pengembangan diawali dari kajian teori, preliminari penelitian, dan temuan secara empirik di lapangan maka dapat disusun draf awal konsep pengembangan model evaluasi. Draft awal yang telah disusun kemudian dikritisi oleh para pakar melalui teknik Delphi terdiri dari pakar evaluasi, pengukuran, metodologi, pakar *tahfiz* al-Qur'an dan pondok pesantren. Hasil dari Delphi menjadikan komponen dari draf awal banyak mengalami perubahan terutama isi dari komponen konteks evaluasi, susunan, jumlah instrumen, susunan komponen model evaluasi program *tahfiz* al-Qur'an, ketepatan penggunaan istilah dalam evaluasi, format penulisan petunjuk pada kolom instrumen, dan penggunaan bahasa Indonesia secara baku dan kata serapan.

Pada saat dilaksanakan proses validasi isi draf awal model evaluasi program pembelajaran *tahfiz* al-Qur'an oleh para pakar secara beriringan dilaksanakan *Focus Group Discussion* (FGD). Hasil dari FGD adalah 1 menentukan standar pokok evaluasi program *tahfiz* al-Qur'an terdiri atas konteks program, input program, proses program, dan produk program. Selain standar pokok ditambah dengan standar pendukung berupa kenaikan jumlah santri tiap tahun, dan jumlah wisudawan *tahfiz* al-Qur'an tiap tahun, perkembangan nilai raport *tahfiz* al-Qur'an, jumlah santri berprestasi bidang *tahfiz* al-Qur'an.

FGD selain menghasilkan standar minimal evaluasi tersebut juga menghasilkan kriteria standar minimal proses menghafal al-Qur'an, metode pemeliharaan hafalan al-Qur'an. Hasil kesepakatan dari proses FGD dijadikan standar untuk menentukan kriteria evaluasi kualitas pelaksanaan program pembelajaran *tahfiz* al-Qur'an di masing-masing pondok pesantren pesantren.

Instrumen setelah menjadi draf awal kemudian divalidasi oleh pakar pengukuran, pakar evaluasi, pakar metodologi penelitian, dan pakar *tahfiz* al-Qur'an, dilanjutkan uji keterbacaan instrumen untuk menguji kelayakan isi instrumen oleh calon pengguna

instrumen (praktisi), dilakukan uji coba pertama menggunakan subjek uji coba 33 orang untuk menerima saran dan masukan dari kekurangan pada instrumen, uji coba kedua menggunakan subjek uji coba 49 orang untuk melihat kelayakan instrumen dan melihat validitas dan reliabilitas instrumen, dan uji coba ketiga menggunakan subjek uji coba 224 orang untuk melihat kelayakan instrumen, melihat validitas, melihat reliabilitas, dan menguji kecocokan model dengan data empiris.

Menjawab pertanyaan kedua yang berbunyi bagaimana cara pelaksanaan evaluasi program pembelajaran *tahfiz* al-Qur'an secara tepat menggunakan model evaluasi program pembelajaran *tahfiz* al-Qur'an Coni P2 di pondok pesantren?. Komponen yang dievaluasi mengikuti standar komponen yang telah ditentukan oleh standar evaluasi Coni P2. Komponen model evaluasi ini telah mengalami beberapa tahapan uji coba instrumen dan dilakukan perbaikan-perbaikan, sehingga menghasilkan model evaluasi yang cocok digunakan di pondok pesantren. Teknik yang digunakan untuk analisis data adalah teknik analisis deskriptif dengan menggunakan rerata skor dan rerata skor total kemudian dikonfirmasi dengan standar konfensi dengan menggunakan kriteria penilaian yang telah disepakati. Tahapan-tahapan uji coba instrumen dijelaskan berikut ini.

Hasil evaluasi di Pondok Pesantren Al-Ittifaqiah (PP.I). Evaluasi konteks program pembelajaran *tahfiz* al-Qur'an terdiri atas gambaran lingkungan sudah sesuai dengan konteks program, kebutuhan yang diperlukan oleh program, dan peluang bagi pondok dan santri dalam pelaksanaan program ini sudah "sangat baik". Evaluasi input terdiri atas kualitas santri, kualifikasi guru, dan sarana, dinilai oleh pimpinan, guru, dan santri. Memiliki kriteria "sangat baik" memang dalam pelaksanaan input sangat selektif baik dari rekrut santri, rekrut guru, dan persiapan sarana belajar guna menunjang keberhasilan program *tahfiz* al-Qur'an. Rekrut santri melalui tes kemampuan dasar, wawancara dan bahkan harus mengisi surat

perjanjian kesanggupan menghafal sebanyak 30 juz. Sarana belajar disiapkan berupa *muşalla* putra dan putri, asrama santri *tahfiz* putra dan putri, belum tersedia secara khusus untuk pelaksanaan *tahfiz* al-Qur'an, disinilah letaknya kekurangan pondok pesantren Al-Ittifaqiah sehingga pelaksanaan proses belajar *tahfiz* dilaksanakan di *muşalla*, kantor madrasah Aliyah, *waiting room*, asrama putra dan putri, dan di rumah guru. Disinilah letak kesenjangan di Pondok Pesantren Al-Ittifaqiah.

Hasil evaluasi proses pembelajaran *tahfiz* al-Qur'an adalah "sangat bagus". Guru sudah bekerja sangat maksimal, guru menguasai hafalan al-Qur'an cukup, saat mengajar guru tidak membawa *muşhaf*. Guru sudah menerapkan hukum tajwid pada santri dengan tepat dengan bukti pada saat pelaksanaan observasi di kelas dijumpai saat santri kurang tepat dalam pengucapan huruf al-Qur'an pada ayat, maka guru akan menegor dengan memberi aba-aba atau ketukan meja kemudian dijelaskan, dan dipraktikkan oleh santri secara langsung.

Hasil penilaian dari observasi di kelas santri putra memiliki nilai 0,71 "tingkat reliabilitas "baik" dan hasil observasi pada santri putri menghasilkan nilai 0,70 "tingkat reliabilitas "baik". Proses observasi diawali dengan melihat saat santri telah dihadapan guru kemudian santri menyerahkan buku kendali *tahfiz* al-Qur'an. Jika santri akan memulai mengaji maka terlebih dahulu membaca *ta'awudz*, membaca materi hafalan, dan diakhiri dengan bacaan *tashdiq* (*şadaqallabul 'adhim*), santri mundur dan menjabat tangan guru sambil dicium dan santri menerima buku kendali dan kembali duduk di tempat semula.

Media pembelajaran di Pondok Pesantren Al-Ittifaqiah sudah dimanfaatkan secara sangat baik dan maksimal, santri diwajibkan menggunakan *muşhaf* murni pojok, harus menggunakan tulisan *rasam Usmani*. Motivasi belajar santri terlihat memiliki antusias yang tinggi, seperti terlihat dari hasil penilaian evaluasi ini adalah "sangat baik". Motivasi belajar *tahfiz* al-Qur'an rata-rata terbangun dari diri santri sendiri untuk

masuk di program *tahfiz* ini, karena di pondok ini sering diadakan penampilan para *hafiz-hafizah*, juara kelas, juara umum, juara favorit, dihadiri oleh seluruh warga pondok pesantren, untuk menerima hadiah berupa piagam, pembebasan uang SPP.

Teknik pemeliharaan hafalan al-Qur'an sudah "baik". Guna menjaga kualitas hafalan maka diperlukan frekwensi pengulangan lebih banyak, karena hafalan jika tidak dilakukan pengulangan akan lupa dan hilang dari ingatan, oleh karena itu pondok pesantren Al-Ittifaqiah memiliki program *simaan* al-Qur'an bagi santri dan juga guru.

Evaluasi produk kompetensi *tahfiz* al-Qur'an dinilai oleh guru dan santri, hasil evaluasi "sangat baik" dibuktikan dengan penguasaan hafalan secara baik disertai ketepatan aplikasi hukum tajwid, dan penguasaan ayat *mutasyabihat* (ayat serupa tetapi tak sama). Pemahaman ayat-ayat yang serupa tetapi tidak sama juga merupakan ciri kualitas hafalan baik. Indikasi hafalan kualitas baik (lancar) memiliki 3-5 kali kesalahan dalam menyetor hafalan, tidak membuka atau sesekali membuka *mus'haf* saat mengulangi hafalan sendiri, berani disimak oleh orang lain, bersedia diuji seperti MHQ.

Hasil evaluasi di Pondok Pesantren Raudhatul Ulum. Penilaian konteks sudah "sangat baik" program *tahfiz* al-Qur'an sudah tergambar secara jelas sehingga dapat dipahami bersama untuk menjalankan programnya. Kebutuhan program sudah tergambar secara jelas, demikian juga peluang dari adanya program ini dapat memberi peluang bagi para santri untuk mengikuti program *tahfiz* al-Qur'an sehingga menjadi *hafiz-hafizah* yang berkualitas tinggi untuk mengajarkan kepada masyarakat di masa mendatang.

Hasil evaluasi input Pondok Pesantren Raudhatul Ulum (PP.RU) sudah "sangat baik" ditandainya setiap santri yang masuk program *tahfiz* al-Qur'an dengan direkrut secara hati-hati dan cermat agar tidak mengalami kesulitan saat menghafal al-Qur'an. Guru juga direkrut secara baik seperti dari Pondok Pesantren Assalam, Sungai Lilin,

Banyu Asin, Sumatera Selatan, Pondok Pesantren Gontor, Ponorogo, Jawa Timur, LIPIA Jakarta, Pondok Pesantren (*ma'had tahfiz* al-Qur'an) Al-Iman, Yaman, *ma'had tahfiz* Al-Qur'an di Siria, *Ma'had tahfiz* al-Qur'an di Madinah, *ma'had tahfiz* Al-Qur'an. Al-Irsyad, Pekalongan, Jawa Tengah, Darul *Hufaz*, Berenung, Lampung Selatan, Sumatera Selatan, dan dari Pondok Pesantren Raudhatul Ulum sendiri.

Sarana belajar sudah baik dengan disedikannya masjid 2 lantai, muşalla Putri, Kantor Pusat, Gedung Serba Guna dan Olah Raga, Toko Pelajar, Kantin Putra dan Putri, dan Klinik. Ruang setoran hafalan al-Qur'an menggunakan muşalla, asrama, kelas, rumah guru. Semua yang dibutuhkan untuk menunjang keterlaksanaan program pembelajaran *tahfiz* al-Qur'an sudah sangat baik.

Media belajar menghafal al-Qur'an sudah dimanfaatkan dengan "sangat baik". Menggunakan al-Qur'an standar yaitu al-Qur'an pojok/sudut yang memiliki kriteria tertentu. Evaluasi motivasi belajar *tahfiz* al-Qur'an bagi santri memiliki kriteria "sangat baik". Kriteria baik tersebut ditunjukkan pada kehadiran santri mencapai 80% setiap bulan kehadiran santri didokumentasikan, mencapai target satu hari satu halaman bahkan ada yang melebihi target karena ada yang sekali setor minimal 3 halaman, sehingga jangka waktu setahun 6 bulan dapat menyelesaikan 30 juz.

Penilaian teknik pemeliharaan hafalan al-Qur'an memiliki kriteria "baik". Kriteria tersebut dijelaskan bahwa hafalan yang telah dikuasai dilakukan *muraja'ah* menyetor ulang (menyetor ulang) kepada guru, *muraja'ah* individu (*nderes*), *simaan* dengan teman sejawat.

Hasil evaluasi proses pembelajaran *tahfiz* al-Qur'an di Pondok Pesantren Raudhatul Ulum (PP.RU) menghasilkan hasil evaluasi dengan kriteria "sangat baik". Hasil penilaian dari observasi di kelas santri putra memiliki nilai 0,46 "tingkat reliabilitas baik" dan hasil observasi pada santri putri menghasilkan nilai 0,48 "tingkat reliabilitas baik".

Evaluasi kompetensi *tahfiz* al-Qur'an sudah "sangat baik" dibuktikan dengan penguasaan hafalan secara baik disertai kete-

patan aplikasi hukum tajwid, dan penguasaan ayat *mutasyabihat*.

Penilaian akhlak mulia memiliki kriteria “baik”. Kriteria tersebut ditunjukkan pada individu santri seperti memiliki jiwa tenang yang terpancar dalam diri santri seperti santri yang diwawancarai, berpenampilan menarik menggunakan pakaian sederhana namun tetap terlihat anggun dan menarik, tidak mudah berkeluh kesah seperti saat menjumpai ayat yang sulit dihafal dengan berulang kali mengulangi, mampu menyelesaikan permasalahannya dengan bijak, jujur dalam berkata, disiplin dalam bertindak.

Penilaian terhadap amal salih memiliki kriteria “baik”. Kriteria amal salih ditunjukkan pada individu santri berupa mampu berinteraksi dengan sesama dan masyarakat seperti memberi ucapan salam sambil tersenyum saat bertemu dengan orang lain, partisipasi pada kegiatan pondok, menghadiri panggilan. Santri mampu beradaptasi terhadap lingkungan seperti menghadiri acara yang diadakan oleh pondok pesantren, kerja bakti lingkungan pondok pesantren, dan menjaga kelestarian alam. Santri mampu memberikan bantuan kepada orang lain seperti mengantar teman sakit di bawa berobat ke dokter.

Hasil evaluasi di Pondok Pesantren Raudhatul Qur’an (PP.RQ). Hasil evaluasi konteks pada gambaran lingkungan program *tahfiz* al-Qur’an “sangat baik” karena dirumuskan oleh pimpinan pondok pesantren, para pendiri dan pejuang pondok pesantren seperti visi adalah sebagai *central excellent* pendidikan Islam, khususnya *tahfiz* al-Qur’an dan penguasaan bahasa Arab, sehingga terbentuknya generasi Qur’ani yang berilmu dan berakhlak mulia. Penilaian kebutuhan untuk menunjang jalannya program *tahfiz* al-Qur’an dinilai “baik”. Penilaian peluang adalah pada klasifikasi “baik”. Peluang yang diharapkan adanya program ini adalah menyediakan tempat pelaksanaan program, peluang bagi para santri memiliki prestasi *tahfiz* al-Qur’an 30 juz, memiliki prestasi baik tingkat kabupaten, propinsi, nasional, dan hingga internasional.

Evaluasi input terhadap kualitas santri “sangat baik”. Santri direkrut untuk masuk di pada program *tahfiz* al-Qur’an melalui seleksi administrasi saja dengan mengisi formulir dan menyerahkan foto, belum dilakukan seleksi santri yang memiliki kemampuan untuk menghafal al-Qur’an. Penilaian kualifikasi guru *tahfiz* al-Qur’an memiliki kriteria “sangat baik”. Kriteria baik tersebut memang guru betul-betul direkrut dari beberapa pondok pesantren yang betul-betul berkualitas tinggi seperti Al-Amin, Tredian, Madura, Jawa Timur, Al-Mukhlis, Bandung, Raudhatul Qur’an, Ogan Ilir, Sumatera Selatan, Raudhatul Ulum, Ogan Ilir Sumatera Selatan, Al-Ittifaqiah. Evaluasi sarana belajar memiliki klasifikasi “sangat baik”. Klasifikasi baik tersebut dapat dijelaskan bahwa pondok pesantren memiliki gedung 4 unit, lokal belajar formal 12 unit, kantor 4 unit, ruang asrama/inap santri 8 unit, rumah guru 3 unit, masjid 1 unit, toilet 11 unit, penerang listrik PLN 1 unit.

Evaluasi proses dimulai kinerja guru memiliki nilai pada posisi agak rendah sekalipun pada klasifikasi “baik” karena masih di bawah angka 4. Guru menyiasati cara mengajar kepada santri yang masih banyak pemula menghafal al-Qur’an sebarang kemampuan individu yang terpenting hafalan lancar tidak terbata-bata, Evaluasi pemanfaatan media yang digunakan untuk menghafal al-Qur’an memiliki nilai “sangat baik”. Sebagai penunjang penguatan hafalan santri selalu diperdengarkan dengan lagu-lagu *murattal* al-Qur’an melalui Radio Santri. Qori’ dari Timur Tengah seperti Syekh Imam besar Makkah Abdurrahman Assudaisy, Syekh Abdurrahman Al-Hudzaifi, dan Syekh Al-Minsyari dan Syekh Mushthofa Ismail. Evaluasi motivasi belajar *tahfiz* al-Qur’an bagi santri memiliki kriteria “baik”. Kriteria baik tersebut ditunjukkan pada kehadiran santri setiap bulan, mencapai target setengah halaman sampai satu halaman tiap hari, patuh terhadap tata tertib. Pelaksanaan *muroja’ah* pengulangan hafalan rata-rata 2-5 halaman tiap hari disetor kepada guru.

Hasil penilaian dari observasi di kelas santri putra memiliki nilai 0,80 “tingkat reliabilitas sangat baik/*excellent*” karena $>0,75$ dan hasil observasi pada santri putri menghasilkan nilai 0,70 “tingkat reliabilitas baik”. Observasi diawali dari melihat posisi santri masing-masing duduk berbaris menjadi dua baris panjang dan berdekatan, mereka menyiapkan atau melancarkan materi hafalan masing-masing sementara santri yang paling depan dua orang maju terlebih dahulu. Penilaian teknik pemeliharaan hafalan al-Qur’an memiliki kriteria “sangat baik”. Kriteria tersebut dijelaskan bahwa hafalan yang telah dikuasai dalam jiwa maka harus dijaga agar tidak hilang dari ingatan dengan *muraja’ah* menyeter ulang (menyeter ulang) kepada guru, *muraja’ah* individu (*nderes*), *simaan* dengan teman sejawat dan mendengarkan *murattal* dari Radio santri.

Kompetensi *tahfiz* al-Qur’an “sangat baik” dibuktikan dengan penguasaan hafalan secara baik disertai ketepatan aplikasi hukum tajwid, dan penguasaan ayat *mutasyabihat*. Penilaian akhlak mulia kriteria “sangat baik”. Kriteria tersebut ditunjukkan pada individu santri seperti sopan saat bicara dengan orang lain, berpenampilan menarik, tidak mudah berkeluh kesah dalam menghadapi kesulitan, mampu menyelesaikan permasalahannya dengan bijak, jujur dalam berkata, disiplin dalam bertindak, dan memiliki rasa empati terhadap orang lain. Penilaian terhadap amal salih memiliki kriteria “sangat baik”. Kriteria amal salih ditunjukkan memberi ucapan salam saat bertemu dengan orang lain, partisipasi pada kegiatan yang diselenggarakan oleh pondok pesantren, menghadiri panggilan jika dipanggil oleh guru atau orang lain.

Untuk menjawab pertanyaan ketiga yang berbunyi bagaimanakah komponen-komponen dan indikator-indikator model evaluasi program pembelajaran agar dapat digunakan sebagai acuan penyusunan instrumen evaluasi program pembelajaran *tahfiz* al-Qur’an di pondok pesantren. Dijawab menggunakan analisis deskriptif untuk mendeskripsikan komponen model evaluasi dan

secara kuantitatif untuk konstruk model evaluasi program *tahfiz* al-Qur’an.

Model yang digunakan untuk mengevaluasi program pembelajaran *tahfiz* al-Qur’an adalah menggunakan model evaluasi hasil R & D dari model evaluasi Stufflebean (CIPP) dan model evaluasi kesenjangan Malcolm Provus (*discrepancy*). Berdasarkan temuan-temuan di atas setelah dikaitkan dengan komponen dan indikator program pembelajaran *tahfiz* al-Qur’an yaitu gambaran lingkungan, kebutuhan, peluang, kualitas santri, kualifikasi guru, sarana belajar, kinerja guru, media belajar, motivasi belajar santri, pemeliharaan hafalan al-Qur’an, kompetensi *tahfiz*, kompetensi akhlak mulia, dan kompetensi amal salih kesemuanya merupakan indikator yang berhasil disusun hasil dari pengembangan model evaluasi.

Model evaluasi hasil pengembangan telah diuji secara empirik dan telah memenuhi kriteria *Goodness of Fit Index* menggunakan *Confirmatory Factor Analysis* (CFA). Instrumen hasil pengembangan telah memiliki validitas $> 0,3$ dan memiliki reliabilitas $> 0,70$. Dengan demikian, instrumen model evaluasi yang dikembangkan melalui R&D dapat digunakan untuk mengevaluasi program pembelajaran *tahfiz* al-Qur’an di pondok pesantren. Komponen evaluasi dan indikator pembelajaran *tahfiz* al-Qur’an berhasil disusun secara sistematis. Model yang digunakan untuk mengevaluasi program *tahfiz* al-Qur’an adalah model CIPP dan Provus. Model evaluasi ini merupakan hasil dari pengembangan yang diberi nama Coni P2. Model ini mengevaluasi program *tahfiz* al-Qur’an menggunakan komponen dari CIPP yang terdiri dari konteks, input, proses, dan produk.

Hasil analisis komponen dan indikator model evaluasi program pembelajaran *tahfiz* al-Qur’an diuji secara empirik guna mendapatkan kriteria valid dan reliabel pada konstruk instrumen dan menguji model yang dihipotesiskan dengan menggunakan data hasil uji coba di lapangan.

Uji signifikansi sebagai syarat Analisis Konfirmatori. Instrumen sebelum dilakukan analisis menggunakan konfirmatori terlebih

dahulu diuji menggunakan teknik *Exploratory* untuk mengetahui apakah instrumen layak atau tidak diuji menggunakan analisis konfirmatori. Berikut hasil uji signifikansi instrumen. Instrumen konteks nilai *Kaiser Meiyer Olkin Measure of Sampling Adequacy* (KMO) sebesar 0,929. Uji *Barletts Tes of Sphericity* nilai *Chi-Square* 3584,060, df 190, dan signifikansi 0,00. Instrumen input memiliki KMO sebesar 0,875, uji *Barletts Tes of Sphericity* nilai *Chi-Square* 2234,886, df 3,78 dan signifikansi 0,00. Instrumen proses memiliki nilai KMO sebesar 0,816, uji *Barletts Tes of Sphericity* nilai *Chi-Square* 3403,122, df 780, dan signifikansi 0,00. Instrumen produk memiliki nilai KMO sebesar 0,967, uji *Barletts Tes of Sphericity* nilai *Chi-Square* 6428,326, df 561, dan signifikansi 0,00. Seluruh KMO lebih besar dari 0,5 artinya instrumen sudah bagus dan nilai signifikansi seluruh instrumen sudah di bawah 0,05. Dengan demikian, seluruh instrumen dinyatakan berdistribusi normal sehingga sudah memenuhi syarat untuk dianalisis menggunakan *Confirmatory Factor Analysis*. Rangkuman nilai hasil uji signifikansi instrumen Coni P2 ditampilkan pada tabel berikut ini.

Tabel 6. Hasil Uji KMO Instrumen Uji Coba Ketiga

| Komponen | KMO | Barlet's Test of Sphericity | | |
|----------|-------|-----------------------------|-----|------|
| | | X ² | Df | Sig |
| Konteks | 0,929 | 3584,060 | 190 | 0,00 |
| Input | 0,875 | 2234,886 | 378 | 0,00 |
| Proses | 0,816 | 3403,122 | 780 | 0,00 |
| Produk | 0,967 | 6428,326 | 561 | 0,00 |

Model pengukuran. Instrumen konteks dianalisis menggunakan CFA menghasilkan seperti berikut ini.

Hasil pengujian konstruk instrumen konteks menunjukkan hasil sesuai dengan ketentuan *goodness of fit* ditandai dengan nilai Lamda (λ) pada masing-masing butir semua butir telah menunjukkan valid semua, nilai *Lamda* terendah 0,71 ($\lambda > 0,30$), *Chi-Square*= 195,48, df=167, q -value= 0,06506 ($> 0,05$), RMSEA= 0,028 ($\leq 0,08$), dan GFI= 0,97.

Dengan demikian, instrumen konteks seluruh butirnya dinyatakan telah valid.

Hasil pengujian konstruk instrumen input telah memenuhi persyaratan *goodness of fit statistik* ditunjukkan pada nilai Lamda (λ) terendah 0,52 ($\lambda > 0,30$), *Chi-Square* = 390,00, df= 347, q -value= 0,0554 ($> 0,05$), RMSEA= 0,024 ($\leq 0,08$), GFI= 0,90 (GFI = 0,90). Dengan demikian, semua butir instrumen input dinyatakan valid.

Instrumen proses memiliki empat dimensi yang diukur (1) kinerja guru memiliki 10 item butir, (2) pemanfaatan media memiliki 10 item butir, (3) motivasi belajar santri memiliki 11 item butir, dan (4) teknik pemeliharaan hafalan memiliki sembilan item butir.

Hasil analisis konfirmatori instrumen proses memiliki nilai Lamda (λ) terendah 0,69 ($\lambda > 0,30$), *Chi-Square* (δ)= 759,48,122, df= 736, q -value= 0,26670 ($> 0,05$), RMSEA= 0,012 ($\leq 0,08$), GFI= 0,90 (GFI $\geq 0,90$), nilai GFI sudah baik memiliki nilai pada posisi standar terendah.

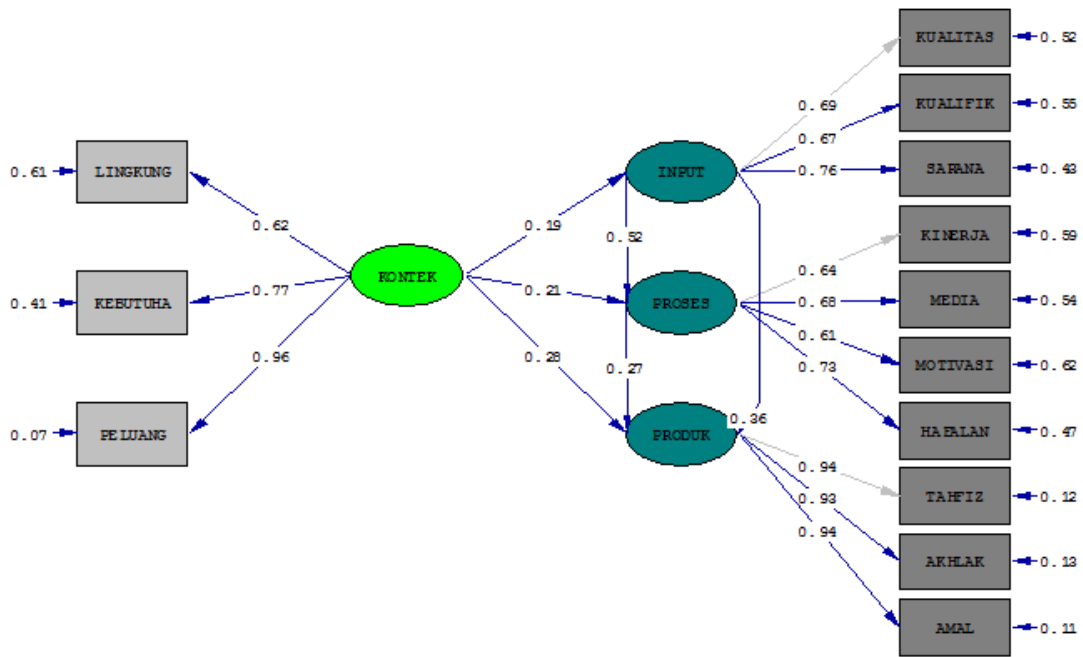
Instrumen produk dianalisis menggunakan Konfirmatori terdiri atas tiga dimensi pertama, kompetensi *taahfiz* al-Qur'an memiliki 13 item butir, kedua, kompetensi akhlak mulia memiliki 13 item butir, dan ketiga, kompetensi amal salih memiliki delapan item butir. Hasil analisis konfirmatori instrumen produk nilai Lamda (λ) terendah 0,74 ($\lambda > 0,30$), *Chi-Square* (δ)= 562,70, df=524 q -value= 0,11745 ($> 0,05$), RMSEA= 0,018 ($\leq 0,08$), GFI= 0,90 (GFI $\geq 0,90$), GFI memenuhi persyaratan sekalipun posisi terendah. Berikut ini rangkuman hasil analisis instrumen evaluasi program *taahfiz* al-Qur'an.

Tabel 7. Rangkuman Hasil Model Pengukuran

| Index | Komponen | | | |
|----------|----------|--------|--------|--------|
| | Konteks | Input | Proses | Produk |
| χ^2 | 195,48 | 390,00 | 759,48 | 562,70 |
| df | 167 | 347 | 736 | 524 |
| p-value | 0,06506 | 0,0554 | 0,2667 | 0,1174 |
| RMSEA | 0,028 | 0,024 | 0,012 | 0,018 |
| GFI | 0,973 | 0,90 | 0,90 | 0,90 |

Model evaluasi secara empiris dinyatakan fit dan telah memenuhi persyaratan yang ditentukan oleh struktural CFA, yaitu nilai $Chi-Square = 71,46$, $df = 59$, $p\text{-value} =$

$0,12805 (> 0,05)$, $RMSEA = 0,031 (\leq 0,08)$, $GFI = 0,953 (GFI > 0,90)$. Model empiris disajikan pada Gambar 1.



Chi-Square=71.46, df=59, P-value=0.12805, RMSEA=0.031

Gambar 1. Model Empiris Coni P2

Rata-rata nilai koefisien pada instrumen konteks, input, proses, dan produk dari tiga pondok pesantren yaitu pondok pesantren Al-Ittifaqiah, Raudhatul Qur'an, dan Raudhatul Ulum, reliabilitas konteks 0,955, input 0,945, proses 0,931 dan produk 0,982. Nilai koefisien tersebut menunjukkan pada posisi lebih dari 0,7. Dengan demikian, seluruh instrumen dinyatakan reliabel artinya bahwa semua instrumen evaluasi program pembelajaran *tabfiz* al-Qur'an dapat diterapkan di semua pondok pesantren di berbagai daerah.

Simpulan

Berdasarkan uraian dan analisis data dapat disimpulkan sebagai berikut.

Instrumen model evaluasi program pembelajaran *tabfiz* al-Qur'an Coni P2 dikembangkan dengan cara melakukan kajian teori, temuan di lapangan, Delphi, FGD, uji keterbacaan, uji coba, uji coba kedua, dan uji coba ketiga.

Cara implementasi model Coni P2 pada pembelajaran *tabfiz* al-Qur'an adalah melakukan evaluasi menggunakan perangkat evaluasi Coni P2 dimulai dari konteks program, input program, proses pembelajaran, dan produk pembelajaran *tabfiz* al-Qur'an, kemudian dibandingkan dengan standar, dan dilanjutkan dengan rekomendasi perbaikan.

Hasil pelaksanaan evaluasi program pembelajaran *tabfiz* al-Qur'an menggunakan model Coni P2 berhasil mendeteksi kesenjangan dan merekomendasikan pada PP.I perbaikan sarana belajar, PP.RU perbaikan sarana belajar, dan PP.RQ perbaikan sarana belajar, kinerja guru, dan motivasi belajar santri.

Manfaat yang diperoleh oleh pimpinan pondok pesantren dari pelaksanaan evaluasi program *tabfiz* al-Qur'an menggunakan model Coni P2 adalah informasi kesenjangan proses pelaksanaan program dan informasi kesenjangan tersebut dijadikan sebagai dasar

untuk melakukan perbaikan perencanaan selanjutnya.

Model Coni P2 terdiri konteks, input, proses, dan produk, dan terbagi menjadi 13 indikator. Komponen dan indikator instrumen telah teruji secara empirik menggunakan analisis CFA menghasilkan semua butir dinyatakan valid karena semua butir memiliki *Chi-Square* kecil, *p-value* > 0,05, RMSEA < 0,08 dan GFI > 0,90.

Berdasarkan hasil observasi dan wawancara model evaluasi Coni P2 sangat efektif digunakan untuk mengevaluasi program pembelajaran *tahfiz* al-Qur'an, karena model Coni P2 memiliki sifat yang sederhana sehingga mudah dilaksanakan, praktis karena dapat dilakukan kapan saja, dan ekonomis karena pelaksana evaluasi adalah warga pondok pesantren sendiri.

Saran

Pondok pesantren baik tingkat menengah dan tingkat tinggi (mahasiswa), rumah *tahfiz*, dan lembaga yang menyelenggarakan program *tahfiz* al-Qur'an apapun namanya hendaknya selalu melaksanakan evaluasi program untuk mengukur ketercapaian program dan mengetahui kelemahan untuk dilakukan perbaikan dan meningkatkan kualitas program periode berikutnya.

Model evaluasi Coni P2 hendaknya digunakan oleh pimpinan pondok pesantren untuk mengevaluasi program *tahfiz* al-Qur'an agar pondok pesantren lebih tertib dan teratur dalam mendokumentasikan data-data yang digunakan untuk mengevaluasi program yang dicanangkan.

Diseminasi hendaknya dilaksanakan dengan mengikutsertakan pihak pemerintah terutama Kementerian Agama, pengurus forum pondok pesantren, dan kalangan perguruan tinggi *tahfiz* al-Qur'an mulai tingkat kabupaten hingga nasional agar model ini tersosialisasikan melalui pihak-pihak terkait.

Model Evaluasi Coni P2 dapat dikembangkan lebih lanjut tidak terbatas untuk mengevaluasi program *tahfiz* al-Qur'an di pondok pesantren belaka, namun dapat juga digunakan di tingkat madrasah *tahfiz* al-Qur'an, perguruan tinggi *tahfiz* al-Qur'an,

dan lembaga yang menyelenggarakan program *tahfiz* al-Qur'an, dan bahkan dapat untuk mengevaluasi satuan mata pelajaran.

Daftar Pustaka

- Arifin, Z. (2012). *Evaluasi pembelajaran*. Bandung: PT Remaja Rosdakarya.
- Borg, W.R. & Gall, M.D. (1983). *Educational research: An intruduction*. New York & London: Longman. Book Company.
- Cozby, P.C. (2009). *Methods behavioral research (9th ed)*. New York: McGrawHill Companies, Ins, 1221.
- Creswell, J.W. (2009). *Research design (Qualitative, quantitative, and mixed methods approaches) (3rd ed)*. Los Angeles: Sage Publication.
- Cruickshank, D.R. (1990). *Research that inform teacher and teacher educators*. Bloomington: Phi Delta Kappa Educational Foundation.
- Darussalam, G. (2010). Program evaluation in higher education program evaluation in higher education. *The International Journal of Research and Review*, 5, 56-65.
- Davidson, E.J. (2005). *Evaluation methodology basics (The nuts and bolts of sound evaluation)*. London: ECIY 1 SP United Kingdom.
- Eye, A. V. & Mun, E. Y. (2005). *Analyzing rater agreement (Manifest variable methods)*. Marwah, New Jersey: Lawrence Erlbaum Associates, Inc.
- Gall, M.D., Gall, J.P., & Borg, W.R (2007) *Educational research an intruduction (8th ed)*. Library of Congress Cataloging-in Publication Data. New York: Pearson Education, Inc.
- Harto, K. & Abdurrahmansyah. (2011). *Metodologi pembelajaran berbasis active learning (Arab baru pembelajaran PAI di sekolah dan madrasah)*. Yogyakarta: Pustaka Felicha.

- Issac, Stephen & Michel, William B. (1981). *Handbook in research & evaluation*. (2nd ed). San Diego, California: Edits Publisher.
- Joyce, B. Weil, M Colhoun, E. (2009). *Model of teaching (8th ed)*. New Jersey: Pearson Education, Inc Publishing US Allyn & Bocam, One Lake Street Upper Saddle River.
- Keeves, J.P. & Lakomski, G. (1992). *Issues in educational research*. Amsterdam: Pergamon An Imprint of Elsevier Science.
- Nurosis, J.M (1986). *SPSS/PC + for the imbbc/xt/at*. Chicago, SPSS.Inc
- Mardapi, D. (2012). *Pengukuran penilaian & evaluasi pendidikan*. Yogyakarta: Nuha Medika.
- Merten, D.M. (2010). *Research and evaluation in education and psychology (Integrating diversi with quantitative, qualitative, and mixed methods)(3th ed)*. California: Sage Publications, Inc.
- Miles, M.B. & Huberman, A.M. (1985). *Qualitative data analysis*. California: Sage Publication, Inc. 275 South Beverly Drive, Beverly Hills.
- Stufflebeam, D. L, Shinkfield, A. J. (1985). *Systematic evaluation*. U.S.A. Massachusetts: Kluwer-Nijhoff Publishing.
- Sudijono, A. (2003). *Pengantar evaluasi pendidikan*. Jakarta: PT Raja Grafindo Persada.
- Sugiyono. (2008). *Metode penelitian pendidikan (Pendekatan kuantitatif, kualitatif, dan R&D)*. Bandung: Alfabeta.
- Wijanto, S.H. (2008). *Struktural equation modeling dengan lisrel. 8*. Yogyakarta: Graha Ilmu.
- Winkel, W.S. (2009). *Psikologi pengajaran*. Yogyakarta: Media Abadi.
- Worthen, B.R & Sanders, J.R. (1981). *Educational Evaluation: Theory and practice*. Ohio: Charles A. Jones Publishing Company Worthington.
- Zaid, M. A. (2011) (King Khalid University). Language acquisition, linguistic creativity and achievement: Insights from the Qur'an. *KEMANUSIAAN Vol. 18, No. 2*, 75-100.