

EVALUASI KURIKULUM PEMBELAJARAN SEKOLAH DASAR MENGUNAKAN MODEL CIPP *STUFFLEBEAM*

Ahmad Nasirudin
Universitas Negeri Jakarta
E-mail: anasirudin00@gmail.com

Abstrak: Penelitian ini berusaha mengevaluasi kurikulum pembelajaran mandiri Sekolah Dasar di Indonesia dengan menerapkan model konteks, masukan, proses, produk (CIPP). Untuk tujuan ini, penelitian menggunakan desain metode campuran, di mana guru (N = 400) telah dipilih dan disurvei secara acak, dan pengawas sekolah (N = 3) telah diwawancarai. Selanjutnya dilakukan analisis isi untuk menganalisis isi kurikulum berdasarkan konstruksi model CIPP. Temuan mengungkapkan bahwa tujuan kurikulum cukup berkorelasi dengan konteks. Masukan, proses, dan produk juga cukup berkontribusi pada kebutuhan pendidikan dan masyarakat Indonesia. Namun kurikulum tersebut gagal memenuhi kebutuhan siswa Sekolah Dasar dan masyarakat dalam menyediakan berbagai sumber pengetahuan, kompetensi mengajar, dan kesempatan pelatihan. Studi tersebut merekomendasikan perlunya mengembangkan kurikulum Sekolah Dasar berdasarkan kebutuhan pendidikan anak dan kebutuhan pengembangan masyarakat. Hal ini menyarankan agar penelitian di masa depan dilakukan dalam berbagai aspek pendidikan anak untuk prasekolah, evaluasi berkelanjutan dan komprehensif,

Kata Kunci: Evaluasi, Kurikulum, Pembelajaran Mandiri, *Stufflebeam*, Model CIPP

Abstract: This study seeks to evaluate the self-learning curriculum of elementary schools in Indonesia by applying a context, input, process, product (CIPP) model. To this end, the study used a mixed methods design, in which teachers (N = 400) were randomly selected and surveyed, and school supervisors (N = 3) were interviewed. Furthermore, content analysis is carried out to analyze curriculum content based on the CIPP model construction. The findings reveal that curriculum objectives are moderately correlated with context. Inputs, processes and products also contribute significantly to the needs of Indonesian education and society. However, the curriculum fails to meet the needs of primary school students and the community in providing multiple sources of knowledge, teaching competencies, and training opportunities. The study recommends the need to develop an elementary school curriculum based on children's educational needs and community development needs. This suggests that future research should be carried out in various aspects of child education for preschool, continuous and comprehensive evaluation,

Keywords: Evaluation, Curriculum, Independent Learning, *Stufflebeam*, CIPP Model

PENDAHULUAN

Pengasuhan anak usia dini menjadi jalan penting untuk mengukur perkembangan negara karena persiapan anak-anak untuk masa depan dianggap sebagai keniscayaan yang beradab. Sekolah Dasar merupakan tahapan penting di mana anak-anak belajar, mengembangkan keterampilan yang bervariasi, dan mendapatkan perkembangan anak yang terintegrasi. Sekolah Dasar bukan hanya lingkungan pengajaran tetapi juga lingkungan untuk pengembangan indera, kecenderungan, dan persiapan anak yang inklusif.

Al-Ghamdi (2005) menyatakan bahwa kurikulum pembelajaran mandiri di Sekolah Dasar mendapat kritik dan saran dari spesialis, peneliti, dan staf sejak penerapan pertama. Kritik tersebut termasuk menilai Kembali kurikulum, menilai kembali keefektifannya, dan menentukan kebutuhan pengembangan konten, materi, dan peralatan agar sesuai dengan minat dan perkembangan masyarakat. Untuk mendapatkan evaluasi yang lebih akurat dan efektif, proses kurikulum belajar mandiri di Sekolah Dasar harus didasarkan pada model yang andal dan terstandarisasi yang sebelumnya telah digunakan oleh studi peer-review. Dalam hal ini, Dar Assalam (2010) menunjukkan bahwa ada sejumlah besar model penilaian yang berbeda (misalnya, model berpusat pada tujuan Tyler; model evaluasi bebas tujuan; model Hammond; model evaluasi berorientasi manajemen; konteks, masukan, proses, produk (CIPP), yang memiliki tujuan penggunaan yang berbeda menurut evaluator yang berbeda dan latar belakang ilmiah serta pengalaman praktis mereka. Dari model ini, Model CIPP dipilih untuk studi ini karena model ini dikenal luas di seluruh dunia karena keandalan dan kepraktisannya.

Secara kontekstual, kurikulum pembelajaran mandiri Sekolah Dasar di Indonesia tidak mengalami perkembangan yang memadai, juga tidak pernah dievaluasi sejak pengukuhan tahun 1976 (Ameel et al., 2009). Kurikulum kurang fokus pada keterampilan penting di Sekolah Dasar seperti membaca dan menulis, yang membutuhkan kebutuhan untuk menulis ulang tujuan dan rencana kegiatan anak agar sesuai dengan tren terkini dalam pendidikan Sekolah Dasar modern. Oleh karena itu, penelitian ini mencoba untuk mengevaluasi kurikulum pembelajaran mandiri Sekolah Dasar dengan menggunakan model Stufflebeam, dan untuk mengidentifikasi sejauh mana tujuan kurikulum pembelajaran mandiri Sekolah Dasar berkorelasi dengan konteksnya dan kontribusi masukan, proses, dan keluaran dalam mencapai kebutuhan pendidikan anak-anak dan kebutuhan perkembangan masyarakat Indonesia.

Anak dianggap sebagai inti dari pendidikan di Sekolah Dasar yang tujuan, isi, metode, dan caranya berkembang. Pengalaman langsung anak mewakili awal dari kurikulum pendidikannya. Oleh karena itu, spesialis yang peduli dengan perencanaan program masa kanak-kanak bergantung pada penyesuaian kurikulum dan metode pendidikan anak agar sesuai dengan persiapan, kebutuhan, dan kemampuan anak.

Sassila (2010) mencatat bahwa karakteristik anak di Sekolah Dasar baik yang berkaitan dengan kebutuhan fisik, sosial, maupun emosional memerlukan iklim yang sesuai sehingga memungkinkan terjadinya interaksi antar anak di Sekolah Dasar. Hal ini dapat dilakukan dengan menyediakan semua alat dan fasilitas fisik dan teknis. Kebutuhan

dapat dianggap sebagai topik yang paling relevan dari keseluruhan pertumbuhan, secara umum, dan pertumbuhan di masa kanak-kanak prasekolah, khususnya, di mana anak mengalami perkembangan fisik, sosial, emosional, dan mental yang cepat. Kebutuhan pendidikan anak SD bervariasi sebagai berikut:

- a. Kebutuhan psikologis dan sosial seperti kebutuhan akan cinta dan penerimaan dari orang lain, keamanan emosional, kebutuhan akan penghargaan sosial, kesuksesan dan kesempurnaan, penegasan diri, kebebasan dan kemandirian, serta penerimaan kekuasaan dan permainan;
- b. Kebutuhan mental dan kognitif seperti penelitian dan eksplorasi, dan kebutuhan untuk memperoleh keterampilan linguistik (Yakhlif, 2014).

Para pendidik dan peneliti telah mengidentifikasi keseriusan kurikulum dan peran penting yang dimainkannya dalam mendidik generasi pelajar, dalam membuat mereka memperoleh keterampilan dan pengetahuan yang membantu mereka dalam pengembangan terintegrasi dari kepribadian mereka dan kemajuan masyarakat mereka. Kurikulum mencerminkan budaya masyarakat dengan segenap unsur keyakinan agama, pola berpikir, perilaku, dan metode pendidikan (Sharif, 2007). Jad (2011) mengidentifikasi evaluasi kurikulum Sekolah Dasar sebagai proses pengumpulan informasi tentang semua komponen kurikulum, untuk mengungkapkan keterampilan dan bakat anak yang mungkin mempengaruhi pencapaian tujuan pertumbuhan seimbang anak Sekolah Dasar secara keseluruhan, untuk memutuskan tentang kurikulum itu sendiri. Evaluasi dilakukan dengan menggunakan metodologi penelitian yang akurat berdasarkan kriteria yang dapat dipercaya untuk menguji keefektifan yang sesuai dengan konteks yang ingin dikembangkan kurikulum. Dari model evaluasi tersebut, model Stufflebeam merupakan evaluasi yang paling banyak diadopsi untuk program, institusi, dan kurikulum yang digunakan secara luas dan terkenal dengan efikasi, komprehensibilitas, dan reliabilitasnya, yang diadaptasi dalam penelitian ini.

Model CIPP

Model ini dirancang pertama kali oleh Danial-Stufflebeam pada akhir 1960-an untuk membantu mengevaluasi program, institusi, dan kurikulum di bidang pendidikan, dan kemudian model tersebut telah digunakan lebih lanjut di berbagai bidang selain pendidikan seperti program sosial, bisnis, dan militer. (Stufflebeam, 2003). Tujuan model ini adalah untuk membantu penilai yang mengevaluasi program sekolah, kursus pengajaran, dan desain kurikulum, yang mencakup empat dimensi yang berbeda: konteks, masukan, proses, dan produk, dan tujuan akhirnya bukanlah "membuktikan tetapi untuk

meningkatkan" (Stufflebeam, 2003. hal.31).

Fase pertama adalah konteks, yang terutama berkaitan dengan mengidentifikasi kebutuhan populasi sasaran, mendefinisikan masalah, menetapkan tujuan, mengidentifikasi sumber daya, dan menetapkan kriteria untuk menilai hasil (Stufflebeam & Shinkfield,2007). Ini terutama didasarkan pada menemukan jawaban atas pertanyaan, "Apa yang perlu dilakukan?" Zhang et al. (2011) menyatakan bahwa evaluasi konteks adalah dasar evaluasi. Inimewakili kerangka umum dari setiap model evaluasi yang mencakupfilosofi budaya, politik, ekonomi, dan pendidikan lingkungansekitarnya. Perannya terletak dalam memverifikasi definisi tujuanpendidikan dan kesesuaiannya dengan kebutuhan siswa. Informasitentang kelompok sasaran dikumpulkan dan batasan evaluasi ditentukan.

Komponen kedua adalah input, yang mengacu pada semua rencana,strategi, dan anggaran pendekatan yang dipilih untuk implementasi. Stufflebeam (2003) menyatakan bahwa masukan "pengguna evaluasi bantuan untuk merancang upaya perbaikan, mengembangkan proposal pendanaan yang dapat dipertahankan, merinci rencana aksi, mencatat rencana alternatif yang dipertimbangkan, dan mencatat dasar untuk memilih satu pendekatan di atas yang lain" (hlm.32) Menurut Karatas dan Fer (2009), masukan berkaitan dengan penyediaan informasi tentang potensi dan sumber daya tubuh, yang akan melaksanakan kurikulum, meninjau literatur dan kurikulum serupa, berkonsultasi dengan spesialis. Selanjutnya, strategi alternatif untuk implementasi dan evaluasi dari segi biaya, utilitas, jadwal, dan kemungkinan hambatan dan mengidentifikasi rencana yang mungkin lebih sesuai dengan kebutuhan desain metode implementasi kurikulum. Ini melayani pengambilan keputusan dan mengadakan lokakarya dan seminar.

Mengenai proses, Stufflebeam (2003) berpendapat bahwa evaluator perlu memprediksi, menilai, mengamati cacat dalam rencana atau pelaksanaannya, dan kemudian memberikan umpan balik tentang bagaimana melaksanakan perbaikan dalam rencana tindakan. Berkenaan dengan kurikulum pendidikan, itu berarti menggunakan metode yang berbeda untuk mengungkapkan bagaimana menerapkan dan mengimplementasikan kurikulum untuk mendeteksi kelayakan rencana implementasi dan pelaksanaan. Ini sesuai dengan peran evaluasi struktural kurikulum dalam mengumpulkan informasi dan data tentang kemajuan proses pendidikan untuk memberikan umpan balik, yang akan membantu mendeteksi kekurangan saat menerapkan kurikulum. Ini akan meningkatkan rencana operasional yang dikembangkan oleh perencana kurikulum, dan dalam interpretasi hasil, dan mengidentifikasi alasan di balik hasil tertentu (Al-Qararaa,

Al-Qaisi, & Al-Rofai)

Komponen terakhir adalah produk, yang artinya mengukur hasil belajar yang diinginkan dan tidak diinginkan. Faktor ini membantu untuk mengidentifikasi apakah kebutuhan siswa dan penerima bantuan telah terpenuhi dan sejauh mana. Ini juga membantu dalam menemukan efek samping yang disengaja dan tidak diinginkan, dan untuk membuat keputusan apakah akan melanjutkan, menghentikan, atau membuat rencana perbaikan (Stufflebeam, 2003). Menurut Attieh (2008), evaluasi keluaran kurikulum berarti pengumpulan dan analisis ketentuan yang berkaitan dengan keberhasilan kurikulum dan dampak yang disengaja dan tidak diinginkan serta hasil yang positif dan negatif. Kalender keluaran sesuai dengan evaluasi akhir kurikulum, dan bertujuan untuk menentukan sejauh mana maksud dan tujuan metodologi tercapai. Ini terkait dengan konteks, masukan, dan proses.

Banyak studi telah dilakukan pada konteks kurikulum pendidikan, penerapan model CIPP dan evaluasi kurikulum pendidikan tinggi. Namun, sedikit yang diketahui tentang evaluasi kurikulum Sekolah Dasar. Dari yang terkait dengan evaluasi kurikulum pendidikan tinggi, Karatas dan Fer (2009) mengidentifikasi bagaimana kurikulum bahasa Inggris di Universitas Yeldiz di Turki berkorelasi dengan model CIPP. Para peneliti melakukan survei dan wawancara kepada 35 guru dan 415 siswa. Tanggapan mereka telah dianalisis, dan temuan terungkap meskipun pendapat mereka bervariasi pada item model yang berbeda, mereka cenderung mendukung gagasan komponen model CIPP. Demikian pula, Tunk (2010) mengevaluasi kurikulum tahun persiapan Universitas Ankara menggunakan model CIPP. Empat ratus enam siswa dan 12 instruktur telah disurvei dan diwawancarai. Selain, analisis konten telah dijalankan pada kurikulum bahasa Inggris. Hasil penelitian mengungkapkan bahwa kurikulum Universitas Ankara melayani sebagian model CIPP dan masalah tertentu perlu dikembangkan.

Adapun tujuan dari penelitian ini antara lain :

1. Sejauh mana kurikulum pembelajaran mandiri Sekolah Dasar di Indonesia cocok dengan kebutuhan Pendidikan anak-anak dan kebutuhan perkembangan masyarakat Indonesia dengan menggunakan model Stufflebeam?
2. Seberapa dekat korelasi tujuan kurikulum pembelajaran mandiri Sekolah Dasar dengan konteks menggunakan model Stufflebeam?
3. Seberapa jauh masukan dari kurikulum belajar mandiri Sekolah Dasar berkontribusi untuk mencapai kebutuhan pendidikan anak dan kebutuhan perkembangan masyarakat Indonesia menggunakan model Stufflebeam?

4. Sejauh mana proses kurikulum belajar mandiri di Sekolah Dasar berkontribusi untuk mencapai kebutuhan pendidikan anak-anak dan kebutuhan perkembangan masyarakat Indonesia dengan menggunakan model Stufflebeam?

METODE PENELITIAN

Bagian ini menjelaskan metodologi yang digunakan, pemilihan sampel, dan alat pengumpulan data dengan langkah-langkah yang diikuti untuk memastikan keandalan dan validitasnya.

Pendekatan Studi

Metodologi kualitatif digunakan untuk mengumpulkan data pada kartu yang sesuai dengan akses mendalam ke umpan balik dari pengawas untuk mengungkap realitas kurikulum belajar mandiri untuk Sekolah Dasar dan kartu analisis isi. Metodologi kuantitatif digunakan dengan menggunakan model evaluasi (yang dibangun) Stufflebeam untuk mensurvei pandangan sampel tentang realitas kurikulum belajar mandiri yang saat ini diterapkan di Sekolah Dasar.

Populasi dan Sampling

Populasi penelitian terdiri dari (a) 756 guru Sekolah Dasar (pengawas, kepala sekolah, dan guru) menurut Data Manajemen Sekolah Dasar 2020 yang ada di Kantor Koryandik Kec. Tenjo Kab. Bogor, (b) 3 pengawas penanggung jawab Sekolah Dasar di Dinas Pendidikan Kecamatan Tenjo Kabupaten Bogor, dan (c) buku kurikulum pembelajaran mandiri yang diadopsi di Sekolah Dasar untuk dimasukkan dalam analisis isi. Sampel acak dipilih dari populasi asli yang terdiri dari Guru di 34 Sekolah Dasar di Kecamatan Tenjo Kab. Bogor, dari total 36 Sekolah Dasar (90%) dengan cara undian; kemudian, pengawas untuk menjawab kuesioner secara acak sederhana. Empat ratus peserta menanggapi (yaitu, 98% dari populasi). Contoh dari penelitian ini — yang telah diwawancarai — adalah sampel yang bertujuan untuk mempelajari suatu fenomena secara mendalam. Lima belas supervisor dari 27 supervisor secara sukarela diterima untuk berpartisipasi dalam wawancara.

Alat Penelitian

Untuk mencapai tujuan penelitian, tiga alat digunakan. Berikut ini adalah penjelasan rinci:

Alat pertama: Kuesioner. Kuesioner digunakan untuk memperoleh pandangan anggota populasi penelitian tentang isi aktual dari kurikulum pembelajaran mandiri Sekolah Dasar dengan menggunakan model CIPP, yang dikonstruksi dalam bentuk dan isi

menurut sumber berikut. Kerangka umum evaluasi kurikulum pendidikan seperti yang didefinisikan oleh Stufflebeam terdiri dari empat dimensi:

Validasi kuesioner. Validitas wajah dinilai oleh penilai, di mana draf pertama kuesioner diberikan kepada sekelompok 13 anggota fakultas yang mengkhususkan diri dalam kurikulum dan Sekolah Dasar. Draft tersebut dinilai dalam kaitannya dengan kesesuaian konten dan bidang untuk berbagai aspek mata pelajaran. Mereka juga diminta untuk mengajukan modifikasi. Penilai memberikan pengamatan mereka, dan meminta amandemen dan proposal. Akibatnya, peneliti mengubah sesuai, sehingga item menjadi 40, bukan 64.

Alat kedua: Kartu wawancara. Untuk mencapai tujuan dan prosedur penelitian diperlukan pembuatan kartu wawancara sampel pengawas Sekolah Dasar di lapangan. Data yang diperoleh dari wawancara berkontribusi untuk menjawab berbagai pertanyaan studi. Untuk menyusun kartu wawancara, wawancara bertujuan untuk mendapatkan lebih banyak pendapat dan informasi tentang realitas kurikulum belajar mandiri di Sekolah Dasar.

Validitas kartu wawancara. Untuk mencapai validitas wawancara yang digunakan dalam penelitian ini, telah disiapkan draft formulir kartu terdiri dari sejumlah pertanyaan di masing-masing bidang model CIPP, yang diberikan kepada tiga reviewer yang mengkhususkan diri di bidang Sekolah Dasar dan kurikulum. Mereka menilai kuesioner dalam hal kesesuaian pertanyaan dan relevansi item dengan masalah yang diteliti dari penelitian saat ini. Peninjau memberikan komentar mereka dan peneliti telah mengubah pertanyaan berdasarkan komentar mereka.

Analisis data

Dengan menggunakan instrument penelitian untuk memperoleh hasil yang valid, yaitu angket, wawancara Untuk menjawab tiga pertanyaan pertama dari penelitian ini, data dari angket dan wawancara telah dianalisis, sedangkan data kualitatif yang diperoleh dari kartu analisis isi digunakan untuk menjawab pertanyaan.

1. Sejauh mana kurikulum pembelajaran mandiri Sekolah Dasar di Indonesia cocok dengan kebutuhan Pendidikan anak-anak dan kebutuhan perkembangan masyarakat Indonesia dengan menggunakan model Stufflebeam?
2. Seberapa dekat korelasi tujuan kurikulum pembelajaran mandiri Sekolah Dasar dengan konteks menggunakan model Stufflebeam?

3. Seberapa jauh masukan dari kurikulum belajar mandiri Sekolah Dasar berkontribusi untuk mencapai kebutuhan pendidikan anak dan kebutuhan perkembangan masyarakat Indonesia menggunakan model Stufflebeam?
4. Sejauh mana proses kurikulum belajar mandiri di Sekolah Dasar berkontribusi untuk mencapai kebutuhan pendidikan anak-anak dan kebutuhan perkembangan masyarakat Indonesia dengan menggunakan model Stufflebeam?

Disini penulis terlebihdahulu akan menganalisis butir soal yang di buat dengan menggunakan aplikasi Bilog. Instrumen pengumpul data berupa 40 butir soal dengan 400 responden. Data penelitian diperoleh dari hasil dari intrumen yang telah di bagikan yang kemudian telah di uji validitas dan uji reliabilitas. Uji validitas menggunakan rumus korelasi *product moment* dan uji reliabilitas menggunakan rumus *cronbach alpha*. Sedangkan analisis daya pembeda, tingkat kesukaran, dan nilai *guessing* berdasarkan pendekatan IRT

Tabel 1. Kriteria Butir Soal yang Baik Berdasarkan Teori Respon Butir

Parameter	Nilai	Keterangan
Daya Pembeda	0,4 s/d 2	Baik
Tingkat Kesukaran	-2 s/d 2	Baik
<i>Pseudo Guessing</i>	$0 \text{ s/d } \frac{1}{k}$ (k = jumlah alternatif jawaban)	Baik

Tiga parameter logistik (3PL) dilakukan dengan menggunakan bantuan program BILOG-MG versi 3.0. Kriteria untuk melihat butir soal yang baik dalam teori respon butir digunakan pendapat yang dikemukakan oleh Hadi (2011; Huda & Mardapi, 2015) yaitu sebagai berikut:

Dalam IRT 3PL, daya pembeda disimbolkan dengan parameter “ \diamond ” atau biasa disebut *discrimination parameters*. Pendapat lain menurut Hambleton, Swaminathan & Rogers dalam Song Gao (2011) bahwa rentang normal untuk parameter “ \diamond ” atau daya pembeda adalah $0 - 2$. Menurut Kurniawan (2015) juga menyebutkan bahwa daya pembeda yang baik berkisar antara $0 \leq \diamond \leq 2$. Dengan demikian, dalam analisis data ini digunakan rentang untuk daya pembeda yang baik berkisar antara $0 \leq \diamond \leq 2$. Tingkat kesukaran disimbolkan dengan parameter “ \diamond ” atau biasa disebut *difficulty parameters*. Tingkat kesukaran yang baik yaitu memiliki rentang $-2 \leq \diamond \leq 2$ (Kurniawan, 2015). Nilai *Guessing* disimbolkan dengan parameter “ \diamond ” atau biasa disebut *pseudo-guessing*. *Guessing* mempunyai rentang $0 \leq \diamond \leq 1$ (Baker, 2001). Berdasarkan Tabel 1 di atas, bahwa untuk *guessing* berkisar antara 0 s/d $1/k$, k merupakan jumlah alternatif jawaban.

HASIL DAN PEMBAHASAN

Analisis butir soal untuk mengetahui daya pembeda, tingkat kesukaran, dan nilai *guessing* dilakukan menggunakan bantuan software BILOG-MG versi 3.0. Analisis ini menggunakan model tiga parameter logistik (3PL). Bisa dilihat secara keseluruhan sebagai hasil keluaran (*output*) BILOG-MG 3 pada Fase I, Fase II untuk Mengestimasi Parameter Soal, Fase III untuk Menaksir Fungsi Informasi, Kesalahan Baku Pengukuran, dan Kemampuan Peserta Tes. Hasil analisis menggunakan software BILOG-MG versi 3.0 dengan 3PL diperoleh data sebagai berikut:

Hasil Analisis PHASE 1

Perhitungan Statistik Instrumen Penilaian Menggunakan BILOG-MG

Gambar 1. Hasil Analisis PHASE 1

ITEM STATISTICS FOR SUBTEST TEST0001					ITEM*TEST CORRELATION		
ITEM	NAME	#TRIED	#RIGHT	PCT	LOGIT/1.7	PEARSON	BISERIAL
1	ITEM0001	397.0	297.0	74.8	-0.64	0.286	0.390
2	ITEM0002	397.0	263.0	66.2	-0.40	0.376	0.486
3	ITEM0003	397.0	217.0	54.7	-0.11	0.416	0.523
4	ITEM0004	397.0	274.0	69.0	-0.47	0.307	0.403
5	ITEM0005	397.0	302.0	76.1	-0.68	0.278	0.382
6	ITEM0006	397.0	366.0	92.2	-1.45	0.139	0.256
7	ITEM0007	397.0	301.0	75.8	-0.67	0.350	0.480
8	ITEM0008	397.0	239.0	60.2	-0.24	0.354	0.449
9	ITEM0009	397.0	282.0	71.0	-0.53	0.320	0.424
10	ITEM0010	397.0	308.0	77.6	-0.73	0.361	0.503
11	ITEM0011	397.0	289.0	72.8	-0.58	0.301	0.404
12	ITEM0012	397.0	154.0	38.8	0.27	0.327	0.416
13	ITEM0013	397.0	373.0	94.0	-1.61	0.202	0.401
14	ITEM0014	397.0	314.0	79.1	-0.78	0.270	0.382
15	ITEM0015	397.0	267.0	67.3	-0.42	0.366	0.476
16	ITEM0016	397.0	284.0	71.5	-0.54	0.271	0.360
17	ITEM0017	397.0	286.0	72.0	-0.56	0.319	0.426
18	ITEM0018	397.0	230.0	57.9	-0.19	0.251	0.316
19	ITEM0019	397.0	313.0	78.8	-0.77	0.271	0.382
20	ITEM0020	397.0	325.0	81.9	-0.89	0.229	0.335
21	ITEM0021	397.0	376.0	94.7	-1.70	0.182	0.377
22	ITEM0022	397.0	254.0	64.0	-0.34	0.417	0.535
23	ITEM0023	397.0	213.0	53.7	-0.09	0.348	0.437
24	ITEM0024	397.0	276.0	69.5	-0.49	0.354	0.465
25	ITEM0025	397.0	291.0	73.3	-0.59	0.350	0.471
26	ITEM0026	397.0	353.0	88.9	-1.22	0.229	0.380
27	ITEM0027	397.0	268.0	67.5	-0.43	0.348	0.453
28	ITEM0028	397.0	321.0	80.9	-0.85	0.288	0.416
29	ITEM0029	397.0	338.0	85.1	-1.03	0.270	0.415
30	ITEM0030	397.0	311.0	78.3	-0.76	0.337	0.473
31	ITEM0031	397.0	195.0	49.1	0.02	0.380	0.476
32	ITEM0032	397.0	226.0	56.9	-0.16	0.317	0.400
33	ITEM0033	397.0	244.0	61.5	-0.27	0.318	0.404
34	ITEM0034	397.0	154.0	38.8	0.27	0.350	0.445
35	ITEM0035	397.0	268.0	67.5	-0.43	0.264	0.343
36	ITEM0036	397.0	315.0	79.3	-0.79	0.315	0.446
37	ITEM0037	397.0	380.0	95.7	-1.83	0.164	0.363
38	ITEM0038	397.0	117.0	29.5	0.51	0.279	0.369
39	ITEM0039	397.0	161.0	40.6	0.22	0.337	0.427
40	ITEM0040	397.0	331.0	83.4	-0.95	0.264	0.394

Tabel di atas merupakan hasil perhitungan statistik oleh BILOG-MG versi 3.0 terhadap 40 butir soal yang dikembangkan. Berdasarkan hasil tersebut, kolom #RIGHT merupakan jumlah peserta tes yang menjawab soal dengan benar. Dengan demikian, diperoleh hasil perhitungan *correlation biserial* (korelasi biserial) yang berbeda-beda tiap butir soal. Semua butir soal akan dikalibrasi karena nilai korelasi biserialnya besar dari $-0,15$. Sehingga semua butir soal digunakan dapat dikalibrasi

pada analisis berikutnya untuk mengetahui karakteristik butir soal dengan daya pembeda, tingkat kesukaran, dan nilai *guessing* pada 40 butir.

Berdasarkan data pada Tabel di atas, diperoleh estimasi perhitungan daya pembeda, tingkat kesukaran, dan nilai *guessing* terhadap soal pada konten *quantity* sebagai instrumen program pelaksanaan kurikulum di sekolah dasar dari sebagai berikut:

Hasil Analisis PHASE 2

Estimasi Perhitungan Karakteristik Butir Soal Menggunakan BILOG MG

Gambar 2. Hasil Analisis PHASE 2

SUBTEST TEST0001; ITEM PARAMETERS AFTER CYCLE 6

ITEM	INTERCEPT S.E.	SLOPE S.E.	THRESHOLD S.E.	LOADING S.E.	ASYMPTOTE S.E.	CHISQ (PROB)	DF
ITEM0001	0.475 0.144*	0.537 0.100*	-0.884 0.350*	0.473 0.088*	0.243 0.090*	11.2 (0.2654)	9.0
ITEM0002	0.248 0.135*	0.696 0.129*	-0.356 0.227*	0.571 0.106*	0.189 0.072*	4.9 (0.8463)	9.0
ITEM0003	-0.088 0.135*	0.768 0.139*	0.114 0.164*	0.609 0.110*	0.139 0.055*	11.6 (0.1698)	8.0
ITEM0004	0.255 0.154*	0.597 0.121*	-0.427 0.303*	0.512 0.104*	0.240 0.086*	0.7 (0.9995)	8.0
ITEM0005	0.513 0.144*	0.525 0.100*	-0.978 0.367*	0.465 0.089*	0.250 0.092*	3.0 (0.9626)	9.0
ITEM0006	1.387 0.156*	0.446 0.111*	-3.111 0.751*	0.407 0.101*	0.261 0.099*	5.3 (0.3765)	5.0
ITEM0007	0.556 0.143*	0.689 0.132*	-0.807 0.284*	0.568 0.108*	0.240 0.088*	2.5 (0.9809)	9.0
ITEM0008	0.043 0.140*	0.620 0.118*	-0.069 0.232*	0.527 0.100*	0.178 0.068*	7.0 (0.6322)	9.0
ITEM0009	0.300 0.161*	0.657 0.133*	-0.458 0.296*	0.549 0.111*	0.263 0.088*	2.1 (0.9894)	9.0
ITEM0010	0.693 0.124*	0.644 0.113*	-1.076 0.281*	0.541 0.095*	0.197 0.079*	15.6 (0.0478)	8.0
ITEM0011	0.339 0.166*	0.639 0.135*	-0.531 0.321*	0.538 0.114*	0.282 0.092*	8.5 (0.4811)	9.0
ITEM0012	-0.644 0.189*	0.686 0.151*	0.939 0.187*	0.565 0.124*	0.134 0.048*	16.8 (0.0324)	8.0
ITEM0013	1.668 0.185*	0.590 0.129*	-2.827 0.550*	0.508 0.111*	0.260 0.099*	1.3 (0.9871)	7.0
ITEM0014	0.614 0.153*	0.569 0.115*	-1.078 0.373*	0.495 0.100*	0.279 0.098*	4.2 (0.8990)	9.0

ITEM0015	0.156 0.171*	0.818 0.192*	-0.191 0.233*	0.633 0.149*	0.260 0.080*	5.0 (0.7577)	8.0
ITEM0016	0.321 0.154*	0.507 0.102*	-0.634 0.368*	0.452 0.091*	0.253 0.091*	12.2 (0.2017)	9.0
ITEM0017	0.199 0.195*	0.848 0.207*	-0.234 0.263*	0.647 0.158*	0.340 0.085*	6.6 (0.6740)	9.0
ITEM0018	-0.108 0.170*	0.460 0.105*	0.234 0.345*	0.418 0.096*	0.220 0.080*	23.8 (0.0046)	9.0
ITEM0019	0.595 0.153*	0.553 0.115*	-1.076 0.385*	0.484 0.101*	0.280 0.099*	5.8 (0.7629)	9.0
ITEM0020	0.750 0.140*	0.478 0.098*	-1.570 0.449*	0.431 0.089*	0.258 0.097*	11.9 (0.2208)	9.0
ITEM0021	1.738 0.179*	0.541 0.117*	-3.212 0.651*	0.476 0.103*	0.251 0.097*	1.8 (0.9402)	6.0
ITEM0022	0.146 0.151*	0.864 0.171*	-0.169 0.193*	0.654 0.130*	0.199 0.071*	12.8 (0.1179)	8.0
ITEM0023	-0.220 0.170*	0.710 0.146*	0.310 0.203*	0.579 0.119*	0.186 0.064*	7.1 (0.6239)	9.0
ITEM0024	0.321 0.142*	0.660 0.133*	-0.486 0.265*	0.551 0.111*	0.218 0.081*	12.6 (0.1827)	9.0
ITEM0025	0.443 0.146*	0.690 0.136*	-0.643 0.279*	0.568 0.112*	0.241 0.085*	9.1 (0.4271)	9.0
ITEM0026	1.185 0.143*	0.509 0.097*	-2.330 0.489*	0.453 0.086*	0.246 0.095*	4.2 (0.8417)	8.0
ITEM0027	0.178 0.168*	0.758 0.157*	-0.235 0.250*	0.604 0.125*	0.254 0.081*	8.8 (0.4588)	9.0
ITEM0028	0.763 0.132*	0.534 0.095*	-1.430 0.367*	0.471 0.084*	0.228 0.089*	12.5 (0.1857)	9.0
ITEM0029	0.968 0.139*	0.548 0.102*	-1.768 0.404*	0.480 0.090*	0.245 0.094*	4.0 (0.9146)	9.0
ITEM0030	0.607 0.157*	0.753 0.166*	-0.807 0.300*	0.601 0.133*	0.290 0.094*	11.1 (0.1947)	8.0
ITEM0031	-0.444 0.208*	0.911 0.217*	0.488 0.160*	0.673 0.160*	0.186 0.057*	4.1 (0.8513)	8.0
ITEM0032	-0.152 0.176*	0.664 0.143*	0.229 0.236*	0.553 0.119*	0.214 0.071*	4.0 (0.9121)	9.0
ITEM0033	0.041 0.151*	0.560 0.112*	-0.073 0.277*	0.489 0.098*	0.204 0.076*	3.4 (0.9477)	9.0
ITEM0034	-0.623 0.180*	0.721 0.154*	0.865 0.173*	0.585 0.125*	0.123 0.046*	8.4 (0.3005)	7.0
ITEM0035	0.023 0.205*	0.655 0.153*	-0.036 0.316*	0.548 0.128*	0.324 0.087*	3.3 (0.9517)	9.0
ITEM0036	0.679 0.147*	0.646 0.124*	-1.051 0.326*	0.543 0.105*	0.260 0.094*	5.8 (0.7599)	9.0
ITEM0037	1.886 0.199*	0.554 0.132*	-3.406 0.725*	0.485 0.116*	0.254 0.098*	0.2 (0.9966)	4.0
ITEM0038	-1.020 0.248*	0.711 0.184*	1.434 0.228*	0.579 0.150*	0.122 0.042*	5.5 (0.5948)	7.0
ITEM0039	-0.566 0.179*	0.668 0.147*	0.847 0.194*	0.556 0.122*	0.133 0.051*	10.6 (0.1562)	7.0
ITEM0040	0.842 0.144*	0.534 0.097*	-1.577 0.402*	0.471 0.086*	0.261 0.097*	10.4 (0.3201)	9.0
					* STANDARD ERROR		
LARGEST CHANGE =		0.006158		299.7 331.0		(0.8907)	

PARAMETER	MEAN	STN	DEV				
ASYMPTOTE	0.230	0.052					
SLOPE	0.638	0.113					
LOG(SLOPE)	-0.465	0.176					
THRESHOLD	-0.702	1.132					

Berdasarkan data setimasi perhitungan karakteristik butir soal di atas diperoleh rekapitulasi hasil estimasi perhitungan karakteristik yang meliputi daya pembeda, tingkat kesukaran, dan nilai *guessing* beserta keterangannya sebagai berikut:

Tabel 1. Rekapitulasi Hasil Estimasi Perhitungan Karakteristik Instrumen Penilaian

Butir Soal Nomor	Daya Pembeda		Tingkat Kesukaran		Nilai Guessing	
	a	Keterangan	B	Keterangan	c	Keterangan
1	0,537	Baik	-0,884	Baik	0,243	Baik
2	0,696	Baik	-0,356	Baik	0,189	Baik
3	0,768	Baik	0,114	Baik	0,139	Baik
4	0,597	Baik	-0,427	Baik	0,240	Baik
5	0,525	Baik	-0,974	Baik	0,250	Baik
6	0,446	Baik	-3,111	Tidak baik	0,261	Sedang
7	0,689	Baik	-0,807	Baik	0,240	Baik
8	0,620	Baik	-0,069	Baik	0,178	Baik
9	0,657	Baik	-0,458	Baik	0,263	Baik
10	0,644	Baik	-1,076	Baik	0,197	Baik
11	0,639	Baik	-0,531	Baik	0,282	Baik
12	0,686	Baik	0,939	Baik	0,134	Baik
13	0,590	Baik	-2,827	Tidak baik	0,260	Sedang
14	0,569	Baik	-1,078	Baik	0,279	Baik
15	0,818	Baik	-0,191	Baik	0,260	Baik
16	0,507	Baik	-0,634	Baik	0,253	Baik
17	0,848	Baik	-0,234	Baik	0,340	Baik
18	0,460	Baik	0,234	Baik	0,220	Baik
19	0,553	Baik	-1,076	Baik	0,280	Baik
20	0,487	Baik	-1,570	Baik	0,258	Sedang
21	0,541	Baik	-3,212	Tidak baik	0,251	Sedang
22	0,864	Baik	-0,169	Baik	0,199	Baik
23	0,710	Baik	0,310	Baik	0,186	Baik
24	0,660	Baik	-0,468	Baik	0,218	Baik
25	0,690	Baik	-0,643	Baik	0,241	Baik
26	0,509	Baik	-2,330	Tidak baik	0,246	Baik
27	0,758	Baik	-0,235	Baik	0,254	Baik
28	0,534	Baik	-1,430	Baik	0,228	Baik
29	0,548	Baik	-1,768	Baik	0,245	Baik
30	0,753	Baik	-0,807	Baik	0,290	Sedang
31	0,911	Baik	0,488	Baik	0,186	Baik
32	0,664	Baik	0,229	Baik	0,214	Baik
33	0,560	Baik	-0,073	Baik	0,204	Baik
34	0,721	Baik	0,865	Baik	0,123	Baik
35	0,655	Baik	-0,036	Baik	0,324	Sedang
36	0,646	Baik	-1,051	Baik	0,260	Sedang
37	0,554	Baik	-3,406	Tidak baik	0,254	Baik
38	0,711	Baik	1,434	Baik	0,122	Baik
39	0,664	Baik	0,847	Baik	0,133	Baik

40	0,534	Baik	-1,577	Baik	0,261	Baik
----	-------	------	--------	------	-------	------

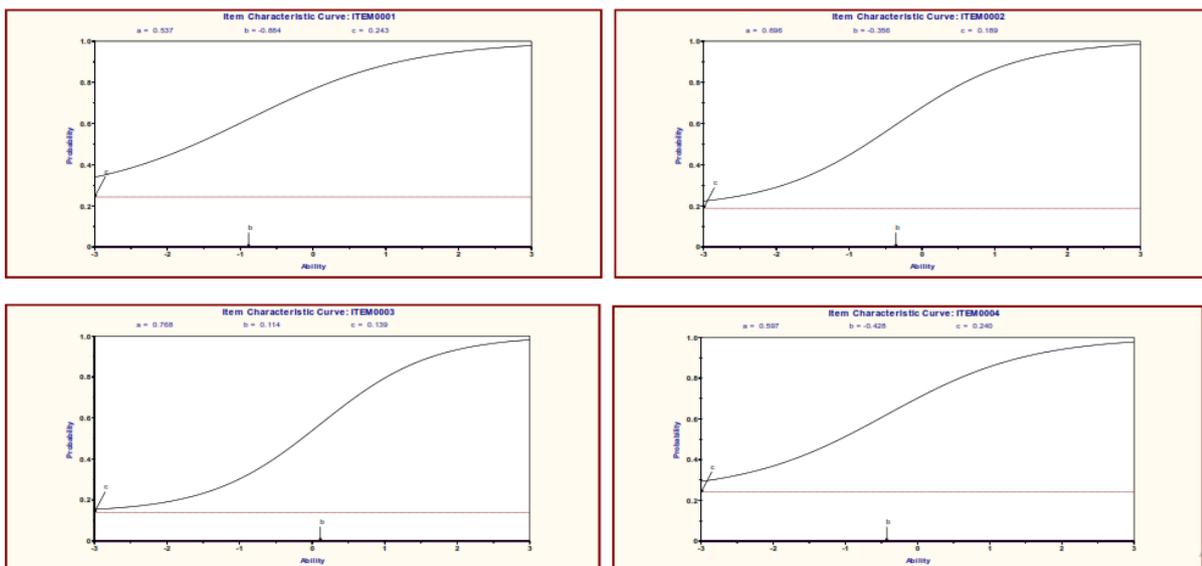
Dari di atas, jumlah butir soal yang layak dianalisis menggunakan teori respon butir 3PL adalah 40 butir soal. Estimasi parameter a dari ke-40 butir soal tersebut dinyatakan mempunyai nilai daya pembeda yang baik karena nilai daya pembeda yang dihasilkan berada pada rentang $0 - 2$. Daya pembeda merupakan kemampuan suatu instrumen penilaian dalam membedakan antara peserta didik yang sudah menguasai materi dengan peserta didik yang belum menguasai materi. Semakin tinggi daya beda suatu item, semakin baik suatu item dalam membedakan kemampuan penempuh tes (Purwo, 2010). Sesuai hasil estimasi perhitungan tersebut, ke-40 butir soal dinyatakan baik dalam membedakan peserta didik yang sudah menguasai materi dengan peserta didik yang belum menguasai materi.

Estimasi parameter b atau tingkat kesukaran dari ke-40 butir soal tersebut termasuk kategori tingkat kesukaran yang baik karena berada pada rentang -2 hingga 2 . Tingkat kesukaran yang tidak baik terlihat dari hasil estimasi yang tidak berada pada rentang $-2 \leq b \leq 2$. Butir soal tersebut memiliki nilai tingkat kesukaran < -2 , sehingga termasuk dalam kategori butir soal yang mudah. Hal tersebut sejalan dengan pernyataan (Kurniawan, 2015) yang menyatakan bahwa nilai tingkat kesukaran > 2 termasuk dalam kategori butir soal yang sulit dan nilai tingkat kesukaran < -2 termasuk dalam kategori butir soal yang mudah, sedangkan nilai tingkat kesukaran yang berada pada rentang $-2 \leq b \leq 2$ termasuk dalam kategori butir soal yang sedang. Dengan demikian, butir 40 butir termasuk dalam kategori soal yang sedang.

Selanjutnya untuk nilai *guessing* dalam Tabel 1 berkisar antara 0 s/d $1/k$, k merupakan jumlah alternatif jawaban. Jumlah alternatif jawaban dalam penelitian ini adalah empat, sehingga *guessing* yang baik berkisar antara 0 s/d $0,25$. Berdasarkan perhitungan teori respon butir 3PL menggunakan BILOG-MG versi 3.0, ke-40 butir soal tersebut menunjukkan bahwa nilai *guessing* hanya sampai pada $0,4$ dan tidak lebih dari $0,5$. Hal tersebut sejalan dengan pernyataan Song Gao (2011) menyatakan bahwa “*The highest probability of guessing cannot be greater than 0,5; otherwise, it would not be called knowledge-based answer instead of guessing*”. Kemungkinan terbesar dari *guessing* (menebak) tidak lebih dari $0,5$; dengan kata lain, tidak dapat dikatakan jawaban dari sebuah pengetahuan hanya dari *guessing* (menebak). Artinya, nilai *guessing* dalam sebuah jawaban tidak akan lebih dari $0,5$ sehingga hal tersebut tidak dapat dijadikan patokan atau pedoman kebenaran jawaban

dari sebuah pengetahuan tersebut hanya dengan menebak. Dengan demikian, kemampuan menebak tidak dapat benar-benar mengetahui kemampuan seseorang. Menebak dapat terjadi terutama pada soal bentuk pilihan ganda dan hal tersebut dapat menyebabkan kesalahan skor variansi terutama kesalahan pada validitas dan reliabilitas soal (Roger, 1999).

Dalam menyusun soal bentuk pilihan ganda, kemampuan dalam menyusun alternatif jawaban merupakan aspek yang sangat penting. Penggunaan distractor yang tidak baik akan mengurangi kualitas soal. Tidak berfungsinya distractor dengan baik secara empirik ditunjukkan oleh hasil penelitian yang dilakukan Attali & Bar-Hillel (2003) yang menyimpulkan bahwa baik penempuh tes maupun pembuat soal memiliki kecenderungan yang sama dalam memilih jawaban atau menempatkan kunci jawaban yaitu pada jawaban yang berada di tengah. Hal tersebut, tentunya meningkatkan kesempatan penempuh tes dalam menebak jawaban. Semakin tinggi kemampuan menebak penempuh tes, tentunya akan memperlemah daya pembeda soal. Semakin rendah daya pembeda soal akan menyebabkan semakin homogen skor yang diperoleh, dan semakin homogen skor yang diperoleh akan memperlemah reliabilitas soal tersebut (Allen & Yen, 1979). Sehingga adanya soal tidak valid dan mempunyai reliabilitas rendah juga dapat disebabkan kemampuan menebak peserta didik yang tinggi dalam menjawab tes yang diberikan. Berikut adalah contoh grafik ICC butir soal nomor 1, 2, 3, dan 4 pada hasil analisis menggunakan BILOG-MG versi 3.0:



Gambar 2. Item Characteristic Curve (ICC) Butir Soal Nomor 1,2,3, dan 4

Model 3PL melibatkan tiga parameter yaitu daya pembeda (a), tingkat kesukaran (b), dan nilai *guessing* (c). Berdasarkan gambar di atas, diketahui nilai a butir soal nomor 1 yaitu 0,537 dengan nilai b sebesar $-0,884$ dan nilai c sebesar 0,243. Sedangkan nilai a butir soal nomor 2 yaitu 0,696 dengan nilai b sebesar $-0,536$ dan nilai c sebesar 0,189, nilai a butir soal nomor 3 yaitu 0,768 dengan nilai b sebesar 0,114 dan nilai c sebesar 0,139, dan nilai a butir soal nomor 4 yaitu 0,597 dengan nilai b sebesar $-0,428$ dan nilai c sebesar 0,240. Berdasarkan hal tersebut diketahui bahwa nilai a untuk butir soal nomor 1 lebih rendah dari butir soal nomor 2, sehingga ICC pada butir soal nomor 1 tampak lebih landai dibandingkan butir soal nomor 2, serta nilai a untuk butir soal nomor 3 lebih tinggi dari butir soal nomor 4, sehingga ICC pada butir soal nomor 4 tampak lebih landai dibandingkan butir soal nomor 3.

Dengan demikian semakin rendah nilai daya pembeda (a) suatu butir soal maka ICC akan tampak lebih landai, sebaliknya semakin tinggi nilai daya pembeda (b) suatu butir soal maka ICC akan tampak lebih menanjak. Nilai $b_1 = -0,884$ dapat diartikan bahwa diperlukan kemampuan minimal $-0,884$ untuk menjawab butir soal dengan benar. Sama halnya dengan butir soal nomor 1, butir soal nomor 2 dengan nilai $b_2 = 0,356$ dapat diartikan bahwa diperlukan kemampuan minimal 0,356 untuk menjawab butir soal dengan benar, butir soal nomor 3 dan 4 dengan nilai $b_3 = 0,114$ dan $b_4 = -0,428$ dapat diartikan bahwa diperlukan kemampuan minimal 0,114 dan $-0,428$ untuk menjawab butir soal dengan benar.

Nilai *guessing* (c) pada ICC di atas, merupakan peluang menjawab benar dengan menebak tiap item tes yang diberikan. *Guessing* pada ICC menunjukkan jarak asimtot bawah terhadap skala kemampuan dan menyatakan probabilitas peserta ujian yang berkemampuan rendah menjawab item dengan benar (Hambleton, 1991). Nilai *guessing* mempunyai hubungan dengan daya pembeda item. Keduanya sangat menentukan besarnya fungsi informasi item yaitu nilai fungsi yang menyatakan ukuran ketepatan suatu pengukuran (Hambleton & Swaminathan, 1985). Fungsi informasi item dapat dimaknai sebagai reliabilitas dalam teori tes klasik. Nilai fungsi informasi item akan maksimal apabila daya pembeda bernilai setinggi mungkin dan nilai *guessing* sekecil mungkin (Hambleton, 1991). Berdasarkan Tabel di atas, daya pembeda dengan nilai setinggi mungkin dan *guessing* dengan nilai sekecil mungkin terdapat pada butir soal nomor 6,13,21,26,37 dan 38. Ini artinya fungsi informasi atau reliabilitas item bernilai maksimal jika dibandingkan dengan butir soal yang lain.

KESIMPULAN DAN SARAN

Berdasarkan uraian hasila analisis di atas dengan menggunakan bantuan Software Bilog MG versi 3.0 menganalisis 40 butir soal sehingga menghasilkan, daya pembeda ke-40 butir soal termasuk kategori baik karena berada pada rentang 0 sampai 2. Tingkat kesukaran butir soal yang termasuk kategori soal sedang. Butir soal dengan kategori tidak baik dikarenakan nilai b atau tingkat kesukaran kurang dari -2 dan termasuk soal yang mudah. Sedangkan butir soal dengan kategori baik termasuk soal yang sedang. Nilai *guessing* ke-40 butir soal termasuk kategori baik karena nilai c berada pada rentang $0 \leq c \leq 1$.

Berdasarkan temuan penelitian ini, maka dapat dibuat rekomendasi berikut. Ketidakbiasaan narasumber dalam mengerjakan soal-soal non rutin membuat mereka kesulitan.

DAFTAR PUSTAKA

Dari Buku:

Attieh, M. (2008). *Modern curriculum and teaching methods*. Amman: Dar Al-Manaheg for Publishing and Distribution.

Danton, D. (2002). *Focus on quality: Pre kindergarten programs in SREB states*. Board Atlanta: Southern Regional Education.

Jad, M. (2011). *Kindergarten curriculum*. Amman, Jordan: Dar Al-Masirah.

Sharif, S. (2007). *The role of the kindergarten environment in providing children with some preparation skills for reading* (Unpublished master's thesis). Faculty of Education, King Saud University, Riyadh, Saudi Arabia.

Stufflebeam, D. (2003, March). *The CIPP model for evaluation: An update, a review, and a checklist to guide implementation*. Paper presented at the 2003 Annual Conference of Oregon Program Evaluators Network (OPEN), Portland, OR.

Stufflebeam, D., & Shinkfield, A. (2007). *Evaluation theory, models, and applications*. San Francisco, CA: Jossey-Bass.

Dari Jurnal:

Al-Amir, G. (2011). *The impact of the use of interactive programs in the growth of literacy skills for preschool children*. Paper presented at the 2nd International Conference on E Learning Distance Learning is Unique to the Unique Generation, Riyadh.

Al-Ghamdi, H. (2005). *Development of the education system in the Kingdom of Saudi Arabia* Riyadh, Saudi Arabia: Obeikan Library.

- Al-Qararaa, A., Al-Qaisi, T., & Al-Rofai, M. (2011). The effect of the classroom environment on the development of creative thinking among children of the basic stage in Tafilah governorate. *Journal of the College of Education, Jordan*, 35, 83-125.
- Ameel, H., Badawood, A., Khalil, A., Abu Hashim, A., Hanadi, B., Najla, L. & Bashtah, K. (2009). *The reality of kindergarten programs in the Kingdom of Saudi Arabia*. Riyadh, Saudi Arabia: UNICEF.
- Darassalam, G. (2010). Program evaluation in higher education. *The International Journal of Research*, 3(4), 56-65.
- Karatas, H., & Fer, S. (2009). Evaluation of English curriculum at Yıldız Technical University using CIPP model. *Education & Science*, 34(153), 47-60.
- Lin, H., & Ching, M. (2012). Managing the Taiwan Kindergarten Evaluation System. *International Journal of Research Studies in Management*, 11, 77-84.
- Mirzazadeh, A., Gandomkar, R., Hejri, S. M., Hassanzadeh, G., Koochak, E., Golestani, A., Razavi, E. (2016). Undergraduate medical education program renewal: A longitudinal context, input, process, and product evaluation study. *Perspectives on Medical Education*, 5, 15-23.
- Rooholamini, A., Amini, M., Bazrafkan, L., Dehghani, M. R., Esmaeilzadeh, Z., Nabeiei, P., Kojuri, J. (2017). Program evaluation of an integrated basic science medical curriculum in Shiraz Medical School, using CIPP evaluation model. *Journal of Advances in Medical Education & Professionalism*, 5, 148-154.
- Sassila, R. (2010). A proposed proposal to ensure the quality of the educational environment in kindergartens in the Syrian Republic. *University of Damascus Journal*, 26, 235-280
- Smaadi, H., & Marwa, N. (1991). The curriculum developed for kindergarten “self-learning” basic curriculum. *General Presidency for Girls’ Education*, 4(3), 16-18.
- Sobha, K. (2011). Analysis of the content of the curriculum of self-learning of kindergartens and the extent to which it contains the skills of critical and creative thinking. *The Association of Modern Education: Egypt*, 4(11), 35-112.
- Yakhlif, R. (2014). The role of kindergartens in social growth. *Journal of the Academy of Social and Human Studies*, 11, 15-10.
- Zhang, G., Zeller, N., Griffith, R., Metcalf, D., Williams, J., Shea, C., & Misullis, K. (2011). Using the Context, Input, Process, and Product Evaluation Model (CIPP) as a comprehensive framework to guide the planning, implementation, and assessment of service-learning programs. *Journal of Higher Education Outreach and Engagement*, 15, 57-84.

Dari Internet:

- Ministry of Education. (2015). *Education development: The national report of the Kingdom of Saudi Arabia*. Riyadh: Ministry of Education.

Dari Tesis:

Tunk, F. (2010). *Evaluation of an English language teaching program at a public university using CIPP model* (Unpublished master's thesis). Middle East Technical University, Ankara, Turkey.