

KUALITAS SOAL PILIHAN GANDA PENILAIAN AKHIR SEMESTER MATA PELAJARAN MATEMATIKA

THE QUALITY OF THE MULTIPLE CHOICE ITEMS OF FINAL SEMESTER ASSESSMENT ON MATHEMATIC SUBJECT

Oleh: Nisa Fauziyyah, Pendidikan Guru Sekolah Dasar, Fakultas Ilmu Pendidikan, Universitas Negeri Yogyakarta, nisafah3@gmail.com

Abstrak

Penelitian ini bertujuan untuk mengetahui kualitas butir soal pilihan ganda Penilaian Akhir Semester I Mata Pelajaran Matematika Kelas V Tahun Pelajaran 2018/2019 di Sekolah Dasar Se-Kecamatan Wates. Penelitian ini merupakan penelitian kuantitatif deskriptif. Populasi penelitian adalah Siswa Kelas V Sekolah Dasar sebanyak 747 siswa dan sampel 264 siswa yang ditentukan dengan rumus Slovin. Metode pengumpulan data menggunakan metode dokumentasi. Teknik analisis data secara kuantitatif untuk mengetahui reliabilitas, tingkat kesukaran, indeks daya pembeda, dan efektivitas pengecoh menggunakan program MicroCat ITEMAN versi 3.00. Hasil penelitian adalah sebagai berikut. (1) 12 soal (80%) termasuk kategori sedang dan 3 soal (20%) termasuk kategori sukar. (2) 8 soal (53,33%) memiliki indeks daya pembeda sangat baik, 4 soal memiliki indeks daya pembeda (26,67%) cukup baik, 1 soal memiliki indeks daya pembeda (6,67%) sedang, dan 2 soal (13,33%) memiliki indeks daya pembeda buruk. (3) Pengecoh berfungsi pada 13 soal (87%) dan tidak berfungsi pada 2 soal (13%). (4) Soal memiliki validitas yang cukup yaitu sebesar 0,411. (5) Soal memiliki reliabilitas yang cukup yaitu sebesar 0,661.

Kata kunci: kualitas butir soal, mata pelajaran matematika.

Abstract

This research attempts to determine the quality of the Multiple Choice Items of First Semester Final Assessment's Academic Year 2018/2019 on 5th Grade Elementary School Mathematics subject. This research type was a descriptive quantitative. The populations were students of grade 5 consist of 747 students and the samples were 264 students which determined by Slovin Formula. The data were collected through documentations. Data analysis technique was carried out quantitatively by using MicroCat ITEMAN program version 3.00. The results show that. (1) 12 items are (80%) medium and 3 items are (20%) difficult. (2) Based on the discrimination index, 8 items are (53,33%) very good, 4 items are (26,67%) good, 1 item is (6,67%) fair, and 2 items are (13,33%) poor. (3) Distraction on 13 items (87%) are good and distraction on 2 items (13%) are poor. (4) The items have a sufficient validity that's 0,412. (5) The items have a sufficient reliability that's 0,661.

Keywords: item quality, mathematics subject.

PENDAHULUAN

Evaluasi hasil belajar memegang peranan yang penting dalam proses pendidikan. Menurut Arifin (2013: 5),

evaluasi adalah suatu proses yang sistematis dan berkelanjutan untuk menentukan kualitas (nilai dan arti) dari sesuatu, berdasarkan pertimbangan dan

kriteria tertentu dalam rangka pengambilan keputusan. Melalui evaluasi hasil belajar, dapat diketahui tingkat penguasaan peserta didik terhadap kompetensi yang ditetapkan, kesulitan-kesulitan yang dialami peserta didik dalam proses belajar, serta efisiensi dan efektivitas strategi pembelajaran yang digunakan guru.

Evaluasi erat kaitannya pula dengan penilaian dan pengukuran. Menurut Griffin dan Nix (Widoyoko, 2009: 4) ketiganya memiliki hubungan yang hierarkis. Evaluasi didahului dengan penilaian, sedangkan penilaian didahului dengan pengukuran.

Dalam bidang pendidikan, psikologi, maupun variabel-variabel sosial lainnya, kegiatan pengukuran biasanya menggunakan tes sebagai alat ukur. Menurut Mardapi (2008: 67), tes merupakan salah satu cara yang digunakan untuk menaksir besarnya kemampuan seseorang secara tidak langsung, yaitu melalui respon terhadap stimulus atau pertanyaan. Pendapat senada juga disampaikan Widoyoko (2009: 45) yang mengemukakan bahwa tes merupakan salah satu alat untuk melakukan pengukuran, yaitu alat untuk mengumpulkan informasi karakteristik suatu objek yang berupa kecakapan peserta didik, minat, motivasi, dan sebagainya.

Agar hasil pengukuran menunjukkan skor yang sebenarnya dari

objek yang diukur, maka diperlukan tes yang baik. Menurut Lubis (2008: 44), tes hasil belajar dikatakan baik jika memenuhi syarat validitas dan reliabilitas. Sementara itu, menurut Sudijono (2015: 370), tes hasil belajar yang baik adalah tes yang butir soalnya telah dianalisis sebagai upaya untuk mengidentifikasi apakah butir soal telah dapat menjalankan fungsinya sebagai alat ukur yang memadai atau belum. Penganalisan dapat dilakukan dari tingkat kesukaran, daya pembeda, dan fungsi pengecoh. Menurut Daryanto (2008: 177) cara menilai tes yaitu: (1) dengan meneliti secara jujur soal – soal yang sudah disusun; (2) mengadakan analisis soal meliputi Taraf Kesukaran, Daya Pembeda, dan Efektivitas Pengecoh; (3) mengadakan *checking* Validitas, dan (4) mengadakan *checking* Reliabilitas.

Analisis perangkat soal ditentukan oleh validitas dan reliabilitas soal. Widoyoko (2009: 128) mengemukakan bahwa instrumen dikatakan valid apabila instrumen tersebut dapat dengan tepat mengukur apa yang hendak diukur. Sementara itu, reliabilitas merujuk pada ketetapan atau keajegan suatu alat ukur untuk menilai apa yang dinilainya (Sudjana, 2011: 16). Selanjutnya, analisis butir soal dilakukan dengan menghitung aspek daya pembeda, tingkat kesukaran, dan efektivitas pengecoh/*distractor*. Ketiga aspek ini diperlukan agar butir tes sebagai

alat ukur dapat menjalankan fungsinya dengan baik.

Analisis butir soal memegang peranan yang penting dalam penyusunan soal. Menurut Arifin (2009:246) analisis butir soal dirancang untuk mengetahui cacat dalam butir tes sehingga dapat diperbaiki sebelum digunakan pada tes berikutnya, serta digunakan untuk mengetahui tes yang diberikan terlalu sulit atau terlalu mudah untuk dikerjakan oleh peserta didik. Sementara itu, menurut Menurut Asmawi Zainul dan Noehi Nasution (Widoyoko, 2014: 130-131) terdapat beberapa alasan mengapa diperlukan analisis butir soal yaitu: (1) untuk mengetahui kekautan dan kelemahan butir tes, sehingga dapat dilakukan seleksi dan revisi butir soal; (2) untuk menyediakan informasi tentang spesifikasi butir soal secara lengkap, sehingga akan lebih memudahkan bagi pembuat soal yang akan memenuhi kebutuhan ujian dalam bidang dan tingkat tertentu; (3) untuk segera dapat mengetahui masalah yang terkandung dalam butir soal, seperti kemenduaan butir soal, kesalahan meletakkan kunci jawaban, soal yang terlalu sukar dan terlalu mudah, atau soal yang mempunyai daya beda rendah; dan (4) untuk dijadikan alat guna menilai butir soal yang akan disimpan dalam kumpulan soal.

Penilaian Akhir Semester (PAS) adalah salah satu penilaian yang dilakukan guru terhadap pencapaian kompetensi peserta didik pada akhir Semester I. Pada penilaian ini, biasanya digunakan jenis tes tertulis yang meliputi tes uraian dan tes objektif. Berdasarkan wawancara dengan salah satu tim guru pembuat soal Penilaian Akhir Semester I pada 13 November 2018, soal Penilaian Akhir Semester I disusun oleh tim pembuat soal dan seorang verifikator. Tim tersebut ditunjuk oleh Dinas Pendidikan, Pemuda dan Olahraga (Dikpora) Kabupaten Kulon Progo. Alur pembuatan soal yaitu pemetaan kompetensi dasar, penyusunan kisi-kisi soal, pembuatan soal, dan verifikasi oleh verifikator. Tim pembuat soal memiliki waktu dua minggu untuk menyusun butir soal Penilaian Akhir Semester. Soal yang telah selesai ditulis kemudian diserahkan kepada verifikator. Apabila terdapat soal yang catat atau kurang sesuai, biasanya langsung dikoreksi oleh verifikator. Namun demikian, belum dilakukan analisis empirik terhadap butir-butir soal yang digunakan dalam Penilaian Akhir Semester I, termasuk dalam Mata Pelajaran Matematika Kelas V. Dengan demikian, belum diketahui pula bagaimanakah kualitas butir soal pada Penilaian Akhir Semester I Mata Pelajaran Matematika kelas V.

METODE PENELITIAN

Jenis Penelitian

Penelitian ini menggunakan jenis penelitian kuantitatif deskriptif. Tujuan penelitian ini untuk mengetahui kualitas butir soal pilihan ganda Penilaian Akhir Semester I Mata Pelajaran Matematika Kelas V Sekolah Dasar Se-Kecamatan Wates Tahun Pelajaran 2018/2019.

Tempat dan Waktu Penelitian

Penelitian ini bertempat di Sekolah Dasar se-Kecamatan Wates, Kabupaten Kulon Progo. Waktu pengambilan data penelitian ini yaitu pada bulan Februari 2019.

Populasi dan Sampel

Populasi yang digunakan dalam penelitian ini yaitu seluruh siswa SD kelas V se- Kecamatan Wates yang menerapkan kurikulum 2013, sejumlah 747 siswa. Sampel yang digunakan sejumlah 264 siswa dengan menggunakan rumus Slovin.

Teknik Pengumpulan Data

Teknik pengumpulan data dalam penelitian ini menggunakan teknik dokumentasi. Metode dokumentasi adalah

Teknik Analisis Data

1. Reliabilitas

Reliabilitas dapat dihitung dengan formula Alpha sebagai berikut:

$$\alpha = \frac{1}{k} \left(\frac{\sum x^2}{N} - \frac{(\sum x)^2}{N^2} \right)$$

Keterangan:

= koefisien reliabilitas alpha

k = banyaknya belahan

= varians skor belahan 1

= varians skor belahan 2

= Varians skor tes (x)

(Azwar, 2002: 185)

Reliabilitas dikategorikan ke dalam lima korelasi, yaitu korelasi sangat rendah, rendah, cukup, tinggi, dan sangat tinggi. Hal ini berpedoman pada pendapat Basuki dan Hariyanto (2014: 119) mengenai tingkat reliabilitas.

Tabel 1. Tingkat Reliabilitas

Koefisien Reliabilitas	Tingkat Reliabilitas
0,00 – 0,19	Sangat Rendah
0,20 – 0,39	Rendah
0,40 – 0,69	Cukup
0,70 – 0,89	Tinggi
0,90 – 1,00	Sangat Tinggi

(Basuki dan Hariyanto, 2014: 144)

2. Tingkat Kesukaran

Untuk menghitung tingkat kesukaran butir soal, digunakan rumus berikut ini:

$$K = \frac{n_i}{N}$$

Keterangan:

n_i = banyaknya siswa yang menjawab aitem dengan benar

N = banyaknya siswa yang menjawab aitem

(Azwar, 2002: 134)

Untuk menafsirkan indeks kesukaran tersebut, dapat digunakan kriteria sebagai berikut:

Tabel 2. Kriteria Tingkat Kesukaran

Tingkat Kesulitan	Kualitas Butir Soal
0,91 – 1,00	Sangat mudah, butir soal tidak baik, tidak digunakan
0,71 – 0,90	Mudah, butir soal kurang baik, direvisi
0,31 – 0,70	Sedang, butir soal cukup baik, digunakan
0,21 – 0,30	Sulit, butir soal kurang baik, direvisi
0,00 – 0,20	Sangat sulit, butir soal tidak baik, tidak digunakan

(Widoyoko, 2014: 133)

3. Indeks Daya Pembeda

Untuk menghitung daya pembeda, digunakan rumus berikut ini:

$$D_p = \frac{n_{iT} - n_{iR}}{N_T - N_R}$$

n_{iT} = banyaknya penjawab aitem dengan benar dari Kelompok Tinggi

N_T = banyak penjawab aitem dari Kelompok Tinggi

n_{iR} = banyaknya penjawab aitem dengan benar dari Kelompok Rendah

N_R = banyak penjawab aitem dari Kelompok Tinggi Rendah

(Azwar, 2002: 138)

Untuk menafsirkan indeks daya pembeda tersebut, dapat digunakan kriteria sebagai berikut:

Tabel 3. Kriteria Evaluasi Indeks Daya Pembeda

Indeks Diskriminasi	Evaluasi
0,40 atau lebih	Bagus sekali
0,30 – 0,39	Lumayan bagus tapi mungkin masih perlu peningkatan
0,20 – 0,29	Belum memuaskan, perlu diperbaiki
Kurang dari 0,20	Jelek dan harus dibuang

(Azwar, 2002: 140)

4. Efektivitas Pengecoh

Efektivitas pengecoh dilihat dari banyaknya siswa yang memilih pengecoh tersebut. Daryanto (2007: 193) mengungkapkan bahwa pengecoh dapat dikatakan berfungsi baik jika paling sedikit dipilih oleh 5% peserta tes (Daryanto, 2007: 193). Pengecoh yang tidak baik direkomendasikan untuk diganti atau direvisi.

HASIL PENELITIAN DAN PEMBAHASAN

Hasil Penelitian

1. Reliabilitas

Reliabilitas dihitung menggunakan formula Alpha dengan bantuan program Iteman MicroCat versi 3.00. Hasil perhitungan koefisien reliabilitas soal pilihan ganda Penilaian Akhir Semester Tahun Pelajaran 2018/2019 Kelas V pada 26 Sekolah Dasar di Kecamatan Wates yang mengimplementasikan Kurikulum 2013 adalah 0,661. Hasil tersebut

menunjukkan bahwa soal pilihan ganda Penilaian Akhir Semester I Tahun Pelajaran 2018/2019 Mata Pelajaran Kelas V memiliki tingkat reliabilitas yang cukup.

2. Tingkat Kesukaran

Indeks kesukaran butir soal dihitung dengan bantuan Program Iteman MicroCat versi 3.00. Untuk mengetahui hasil perhitungan indeks kesukaran butir soal pada *output* program MicroCat Iteman versi 3.00, maka dapat dilihat pada bagian “*Prop Correct*” dari statistik butir soal. Berdasarkan hasil analisis butir soal pilihan ganda Penilaian Akhir Semester I Mata Pelajaran Matematika Kelas V Sekolah Dasar Tahun Pelajaran 2018/2019 didapatkan hasil sebagai berikut.

Tabel 4. Tingkat Kesukaran Butir Soal

Butir Soal	Prop. Correct	Kriteria
1	0,617	Sedang
2	0,490	Sedang
3	0,167	Sulit
4	0,652	Sedang
5	0,318	Sedang
6	0,485	Sedang
7	0,496	Sedang
8	0,405	Sedang
9	0,235	Sulit
10	0,383	Sedang
11	0,246	Sulit
12	0,322	Sedang
13	0,303	Sedang
14	0,420	Sedang
15	0,345	Sedang

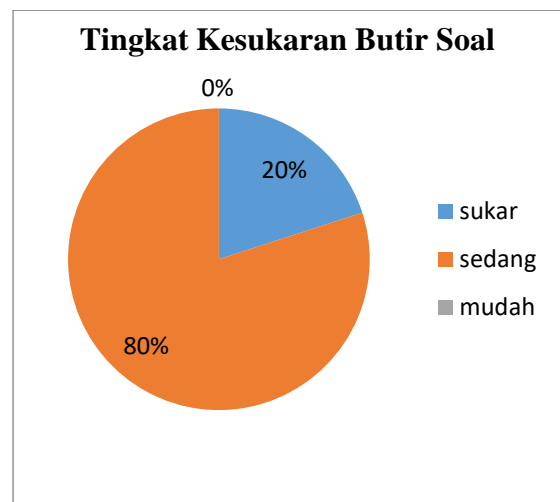
Berdasarkan nilai *prop. correct* di atas, indeks kesukaran butir soal kemudian diinterpretasikan ke dalam tiga kriteria berdasarkan nilai yang sudah ditetapkan

yaitu mudah, sedang, dan sulit. Nilai indeks kesukaran 0,00 – 0,30 termasuk butir soal sukar, indeks kesukaran 0,31 – 0,70 termasuk butir soal sedang, dan indeks kesukaran 0,71 – 1,00 termasuk butir soal mudah.

Dari hasil analisis di atas, diketahui bahwa soal yang termasuk kategori sukar ada 3 soal (20%) dan soal yang termasuk kategori sedang ada 12 soal (80%). Persebaran 15 butir soal berdasarkan indeks kesukarannya adalah sebagai berikut.

Tabel 5. Distribusi Butir Soal Berdasarkan Tingkat Kesukaran

Kategori	Nomor Butir Soal	Jumlah Butir Soal	Persentase
Mudah (0,71 – 1,00)	-	0	-
Sedang (0,31 – 0,70)	1, 2, 4, 5, 6, 7, 8, 10, 12, 13, 14, 15	12	80%
Sukar (0,00 – 0,30)	3, 9, 11	3	20%



Gambar 1. Distribusi Butir Soal menurut Tingkat Kesukaran

3. Indeks Daya Pembeda

Berdasarkan hasil analisis butir soal pilihan ganda Penilaian Akhir Semester I Mata Pelajaran Matematika Kelas V Sekolah Dasar Tahun Pelajaran 2018/2019 didapatkan hasil sebagai berikut.

Tabel 6. Indeks Daya Pembeda Butir Soal

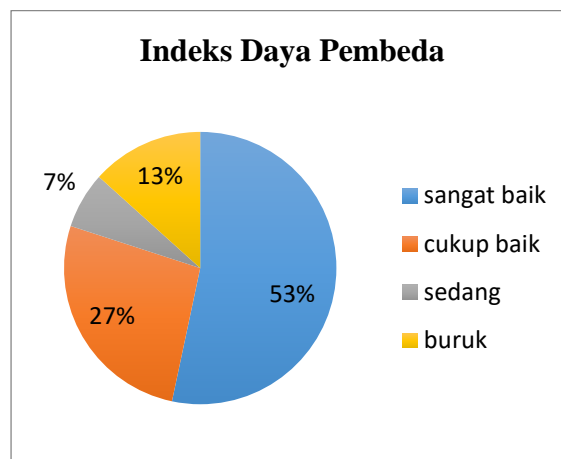
Butir Soal	Point Biser	Kriteria	Evaluasi
1	0,419	Bagus	Diterima
2	0,581	Bagus	Diterima
3	0,145	Buruk	Ditolak
4	0,499	Bagus	Diterima
5	0,582	Bagus	Diterima
6	0,442	Bagus	Diterima
7	0,482	Bagus	Diterima
8	0,450	Bagus	Diterima
9	0,345	Cukup bagus	Diterima
10	0,376	Cukup bagus	Diterima
11	0,269	Belum memuaskan	Direvisi
12	0,192	Buruk	Ditolak
13	0,594	Bagus	Diterima
14	0,366	Cukup bagus	Diterima
15	0,373	Cukup bagus	Diterima

Berdasarkan nilai *point biser* di atas, indeks daya pembeda butir soal kemudian diinterpretasikan ke dalam empat kriteria yaitu sangat baik, cukup baik, sedang, dan buruk. Indeks daya pembeda 0,40 atau lebih termasuk sangat baik, indeks daya pembeda 0,30 – 0,39 termasuk cukup baik, indeks daya pembeda 0,20 – 0,29 termasuk sedang, dan indeks daya pembeda kurang dari 0,20 termasuk buruk. Dari hasil analisis di atas, diketahui bahwa soal yang memiliki indeks daya pembeda sangat baik ada 8 soal (53,33%), soal yang

memiliki indeks daya pembeda cukup baik ada 4 soal (26,67%), soal yang memiliki indeks daya pembeda sedang ada 1 soal (6,67%), dan soal yang memiliki indeks daya pembeda buruk ada 2 soal (13,33%). Persebaran 15 butir soal berdasarkan indeks daya pembedanya adalah sebagai berikut.

Tabel 7. Distribusi Butir Soal Berdasarkan Indeks Daya Pembeda

Kriteria	Nomor Butir Soal	Jumlah Butir Soal	Persentase
Sangat Baik	1, 2, 4, 5, 6, 7, 8, 13	8	53,33%
Cukup Baik	9,10, 14, 15	4	26,67%
Sedang	11	1	6,67%
Buruk	3, 12	2	13,33%



Gambar 2. Distribusi Soal Berdasarkan Indeks Daya Pembeda

4. Efektivitas Pengecoh

Berdasarkan hasil analisis butir soal pilihan ganda Penilaian Akhir Semester I Mata Pelajaran Matematika Kelas V Sekolah Dasar Tahun Pelajaran 2018/2019 didapatkan hasil sebagai berikut.

Tabel 8. Efektivitas Pengecoh Butir Soal

Butir Soal	A	B	C	D
	<i>Prop. Endorsing</i>	<i>Prop. Endorsing</i>	<i>Prop. Endorsing</i>	<i>Prop. Endorsing</i>
1	0,189	0,617	0,110	0,080
2	0,470	0,326	0,091	0,114
3	0,170	0,508	0,167	0,155
4	0,045	0,189	0,110	0,652
5	0,318	0,462	0,144	0,057
6	0,485	0,311	0,140	0,064
7	0,163	0,212	0,117	0,496
8	0,125	0,405	0,326	0,144
9	0,148	0,254	0,235	0,303
10	0,383	0,212	0,231	0,170
11	0,189	0,348	0,212	0,246
12	0,261	0,261	0,322	0,144
13	0,303	0,144	0,311	0,235
14	0,197	0,201	0,420	0,167
15	0,235	0,345	0,205	0,011

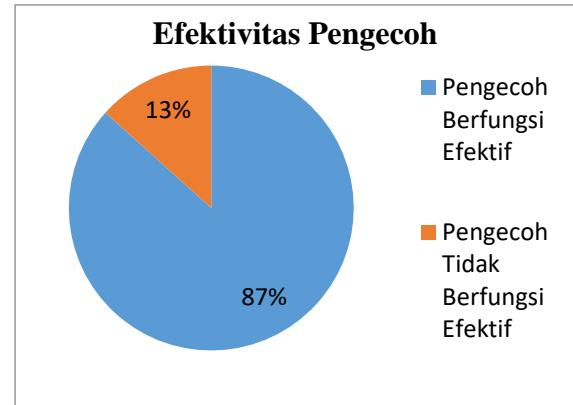
Dari tabel mengenai hasil analisis efektivitas pengecoh butir soal pilihan ganda Penilaian Akhir Semester I Mata Pelajaran Matematika Kelas V Tahun Pelajaran 2018/2019 di Kecamatan Wates, maka didapatkan persentase hasil analisis sebagai berikut.

Tabel 9. Distribusi Butir Soal Berdasarkan Efektivitas Pengecoh

Evaluasi	Nomor Soal	Jumlah Butir Soal	Persen-tase
Berfungsi	1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 15	13	87%
Tidak Berfungsi	3, 12	2	13%

Dari tabel di atas, didapatkan data bahwa terdapat 13 butir soal yang pengecohnya berfungsi efektif dengan persentase sebesar 87%. Ketiga belas butir

soal ini dapat diterima dan digunakan sebagai bagian keseluruhan tes. Sementara itu, pengecoh pada dua butir soal yaitu nomor 3 dan 12 perlu dianalisis dan direvisi agar dapat menjalankan fungsi dengan baik.



Gambar 3. Distribusi Soal Berdasarkan Efektivitas Pengecoh

Pembahasan

1. Reliabilitas

Menurut Azwar (2017: 7), hasil suatu pengukuran dapat dipercaya hanya apabila dalam beberapa kali pelaksanaan pengukuran terhadap kelompok subjek yang sama diperoleh hasil yang relatif sama, selama aspek yang diukur dalam diri subek memang belum berubah. Reliabilitas menunjuk pada ketetapan atau keajegan suatu alat ukur untuk menilai apa yang dinilainya (Sudjana, 2011: 16).

Analisis reliabilitas didapat dengan bantuan program MicroCat ITEMAN versi 3.00. hasil analisis reliabilitas dapat dilihat pada *output* program MicroCat ITEMAN versi 3.00 di bagian skala statistik tes yaitu koefisien *alpha*. Dari hasil pengolahan data reliabilitas yang telah dilaksanakan

oleh peneliti untuk mengetahui tingkat keajegan atau ketetapan soal, didapatkan data bahwa soal tersebut memiliki tingkat reliabilitas sebesar 0,661. Sesuai dengan pendapat Basuki dan Hariyanto (2014: 119), bahwa apabila hasil perhitungan reliabilitas berada di antara koefisien 0,40 - 0,69 maka termasuk ke dalam tingkat reliabilitas korelasi cukup.

2. Tingkat Kesukaran

Beberapa ahli mengemukakan pendapatnya terkait besar indeks kesukaran (p) yang ideal. Menurut Reynolds, dkk (2012: 149), untuk memaksimalkan variabilitas di antara peserta tes, level kesukaran item yang optimal adalah 0,5 yang mengindikasikan bahwa 50% peserta tes menjawab item dengan benar dan 50% peserta tes menjawab salah. Namun demikian, lebih lanjut Reynolds, dkk (2012: 149) mengungkapkan bahwa terkadang perlu untuk memilih soal yang mempunyai indeks kesukaran di bawah 0,50 dan beberapa soal yang mempunyai indeks kesukaran di atas 0,50. Pendapat senada juga disampaikan Azwar (2002: 135) bahwa walaupun tidak selalu benar, namun umumnya indeks kesukaran yang berada di sekitar 0,50 dianggap yang terbaik. Pendapat di atas juga didukung oleh pendapat Widoyoko (2014: 133) yang menyatakan bahwa untuk tes hasil belajar, tingkat kesulitan yang dianggap baik adalah berkisar 0,50. Suatu butir soal yang

memiliki nilai p yang semakin dekat dengan 0,50, maka butir soal tersebut semakin baik bagi kelompok tertentu. Begitu pula apabila semakin jauh tingkat kesulitan dari 0,50, maka semakin kurang informasi yang diperoleh tentang butir soal dan kelompok peserta tes.

Sementara itu, menurut Widoyoko (2014: 133), hubungan antara tingkat kesulitan dengan kualitas butir soal yaitu (1) butir soal dengan indeks kesukaran sebesar 0,31 – 0,70 dapat digunakan sebagai bagian dari tes keseluruhan; (2) butir soal dengan indeks kesukaran 0,21 – 0,30 atau 0,71 – 0,90 perlu diperbaiki; dan (3) butir soal dengan indeks kesukaran 0,00 – 0,20 atau 0,91 – 1,00 sebaiknya tidak digunakan.

Berdasarkan pendapat ahli di atas, maka evaluasi pada soal Penilaian Akhir Semester I Mata Pelajaran Matematika Kelas V Tahun Pelajaran 2018/2019 ini adalah sebagai berikut. Butir soal nomor 1, 2, 4, 5, 6, 7,8, 10, 12, 13, 14, dan 15 termasuk soal sedang. Dengan demikian, butir soal nomor 1, 2, 4, 5, 6, 7,8, 10, 12, 13, 14, dan 15 diterima sebagai bagian dari tes keseluruhan. Butir soal nomor 3, 9, dan 11 termasuk soal yang sulit sehingga sebaiknya perlu dilakukan analisis yang mendalam dan direvisi.

3. Indeks Daya Pembeda

Soal nomor 1, 2, 4, 5, 6, 7, 8, dan 13 memiliki indeks daya pembeda di atas

0,40. Menurut kriteria evaluasi dari Azwar (2002: 140), soal dengan indeks 0,40 atau lebih termasuk bagus sekali dan dapat membedakan antara siswa yang berkemampuan tinggi dan siswa yang berkemampuan rendah. Dengan demikian, delapan soal tersebut dapat diterima sebagai bagian dari keseluruhan tes.

Soal nomor 9, 10, 14, dan 15 memiliki indeks daya pembeda antara 0,30 – 0,39. Menurut kriteria evaluasi dari Azwar (2002: 140), soal dengan indeks daya pembeda antara 0,30 – 0,39 termasuk soal yang cukup baik, namun masih perlu peningkatan. Dengan demikian, lima butir soal dalam Penilaian Akhir Semester I dapat diperbaiki agar indeks daya pembedanya menjadi lebih baik.

Soal nomor 11 memiliki indeks daya pembeda 0,20 – 0,29. Menurut kriteria evaluasi dari Azwar (2002: 140), soal dengan indeks daya pembeda antara 0,20 – 0,29 belum memuaskan dan perlu diperbaiki. Dengan demikian, Dengan demikian, butir soal nomor 11 dapat diperbaiki agar indeks daya pembedanya menjadi lebih baik.

Soal nomor 3 dan 12 memiliki indeks daya pembeda kurang dari 0,20. Menurut kriteria dari Azwar (2002: 140), soal dengan indeks daya pembeda kurang dari 0,20 jelek dan harus dibuang. Dengan demikian, butir soal nomor 3 dan 12 sebaiknya tidak digunakan.

4. Efektivitas Pengecoh

Efektivitas distraktor dilihat dari dua kriteria, yaitu distraktor dipilih oleh siswa dari Kelompok Rendah dan memilih distraktor tersebut tersebar relatif proposional pada masing-masing distraktor yang ada (Azwar, 2002: 142). Hasil analisis butir soal yang dilakukan dengan bantuan ITEMAN MicroCat versi 3.00 menunjukkan bahwa: (1) butir soal nomor 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, dan 15 memiliki pengecoh yang berfungsi dengan efektif untuk mengecoh siswa kurang memahami materi pembelajaran; dan (2) butir soal nomor 3 dan 12 memiliki pengecoh yang berfungsi dengan kurang efektif.

Apabila ditelaah lebih lanjut, kegagalan efektivitas pengecoh pada butir soal nomor 3 ialah karena kunci jawaban yang tersedia lebih dari satu. Soal nomor 3 memiliki kunci jawaban yaitu $-m$. Pada butir soal nomor 3, jawaban $-m$ ini terdapat pada dua pulihan jawaban yaitu pilihan jawaban a dan pilihan jawaban c. Selain terdapat dua pilihan jawaban yang sama dan keduanya merupakan jawaban benar, terdapat satu pilihan jawaban lagi yang benar yaitu pilihan jawaban b. Pilihan jawaban b berbunyi: $-m$. Jawaban tersebut dapat pula dinilai benar karena $-m$ sama nilainya dengan $-m$. Hal ini

dikarenakan kedua pecahan tersebut ($—$ dan $—m$) merupakan pecahan yang senilai.

SIMPULAN DAN SARAN

Simpulan

Berdasarkan pemaparan hasil penelitian dan pembahasan dapat ditarik kesimpulan bahwa kualitas butir soal pilihan ganda pada Penilaian Akhir Semester I Mata Pelajaran Matematika Kelas V Sekolah Dasar Se-Kecamatan Wates Tahun Pelajaran 2018/2019 beragam. Ditinjau dari tingkat kesukaran, 12 soal (80%) termasuk kategori mudah dan 3 soal (20%) termasuk sulit. Ditinjau dari indeks daya pembeda, 8 soal (53,33%) sangat baik, 4 soal (26,67%) cukup baik, 1 soal (6,67%) sedang, dan 2 soal (13,33%) buruk. Ditinjau dari efektivitas pengecoh, pengecoh berfungsi pada 13 soal (87%) dan tidak berfungsi pada 2 soal (13%). Ditinjau dari koefisien *alpha* sebesar 0,661 menunjukkan bahwa tingkat reliabilitas cukup.

Saran

Berdasarkan hasil penelitian, pembahasan, serta kesimpulan, maka peneliti mengajukan beberapa saran diantaranya bagi guru dan sekolah agar dilakukan uji coba dan analisis butir soal pada pembuatan soal Penilaian Akhir Semester sehingga diketahui kekuatan dan

kelemahan butir soal sebagai bahan evaluasi untuk menentukan butir soal yang dapat digunakan dan direvisi.

DAFTAR PUSTAKA

- Arifin, Z. (2013). *Evaluasi pembelajaran: prinsip, teknik, prosedur*. Bandung: Remaja Rosdakarya.
- Azwar, S. (2002). *Tes prestasi: fungsi pengembangan pengukuran prestasi belajar*. Yogyakarta: Pustaka Pelajar.
- Daryanto. (2008). *Evaluasi Pendidikan*. Jakarta: Rineka Cipta.
- Lubis, M. (2008). *Evaluasi pendidikan nilai*. Yogyakarta: Pustaka Pelajar.
- Basuki, I. & Haryanto. (2014). *Asesmen pembelajaran*. Bandung: Remaja Rosdakarya Offset.
- Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan nontes*. Yogyakarta: Mitra Cendekia.
- Sudjana, N. (2009). *Penilaian hasil proses belajar mengajar*. Bandung: Remaja Rosdakarya.
- Reynolds, C.R., Livingston, R.B., & Willson, V. (2010). *Measurement and assessment in education*. New Jersey: Pearson.
- Sudjana, N. (2009). *Penilaian hasil proses belajar mengajar*. Bandung: Remaja Rosdakarya.
- Widoyoko, E. P. (2014). *Penilaian hasil pembelajaran di sekolah*. Yogyakarta: Pustaka Pelajar.