

## **Evaluasi Formatif Berbantuan Komputer Untuk Meningkatkan Kemampuan Analisis Butir Soal Mahasiswa FKIP Universitas Bengkulu**

**Zamzaili<sup>1</sup>, Baki Swita<sup>2</sup>**  
<sup>1,2</sup>Universitas Bengkulu

Email Koresponden: [zamzaili58@gmail.com](mailto:zamzaili58@gmail.com)

### **Abstrak**

Tujuan penelitian ini untuk mengetahui besar sumbangan efektif evaluasi formatif berbantuan komputer terhadap kemampuan analisis butir soal mahasiswa dan membandingkan rata-rata kemampuan analisis butir soal mahasiswa yang diajar dengan evaluasi formatif berbantuan komputer dengan rata-rata kelas diajar tanpa bantuan komputer. Teknik sampling yang digunakan adalah simple random sampling dengan jumlah sampel 32 mahasiswa untuk kelas eksperimen dan kelas kontrol. Instrumen penelitian adalah tes kemampuan analisis butir soal dengan indikator indek kesukaran, daya beda, validitas dan reliabilitas. Teknik analisis data menggunakan anava dua jalur. Hasil penelitian: (1) terdapat sumbangan yang efektif evaluasi formatif berbantuan komputer terhadap kemampuan analisis butir sebesar 52,47%; (2) terdapat pengaruh yang signifikan kemampuan statistik terhadap kemampuan analisis butir soal; dan (3) rata rata kemampuan analisis butir mahasiswa yang mengikuti evaluasi formatif berbantuan komputer (65,63) lebih tinggi dari rata-rata yang diajar tanpa bantuan komputer (59,13).

**Kata kunci :** Evaluasi formatif, Komputer, Analisis butir soal

### **Abstract**

The purpose of this study was to determine the effective contribution of computer-assisted formative evaluation to the student's item analysis skills and to compare the average student-taught item analysis skills with computer-assisted formative evaluation with the average class being taught without computer assistance. The sampling technique used was simple random sampling with a sample size of 32 students for the experimental class and the control class. The research instrument was a test item analysis ability with indicators of difficulty index, difference power, validity and reliability. The data analysis technique uses a two-way ANOVA. The results of the study: (1) there is an effective contribution of computer-assisted formative evaluation to the item analysis ability of 52.47%; (2) there is a significant effect of statistical ability on the item analysis ability; and (3) the average item analysis ability of students who participate in the evaluation computer-assisted formative (65.63) is higher than the average taught without computer assistance (59.13).

**Keywords :** Formative evaluation, Computer, Item analysis

## 1. Pendahuluan

Evaluasi adalah proses untuk mengetahui pencapaian hasil dan efektivitas pembelajaran. Karena itu evaluasi merupakan salah satu komponen pokok yang selalu ada dalam pembelajaran. Dengan kata lain, sebuah pembelajaran tidak dapat dipisahkan dengan kegiatan evaluasi. Sebenarnya evaluasi memiliki dua fungsi utama yaitu untuk mengetahui pencapaian hasil belajar peserta didik dan mengetahui keberhasilan mengajar guru. Pengetahuan tentang hasil belajar peserta didik terkait dengan sejauh mana peserta didik telah mencapai tujuan pembelajaran atau kompetensi-kompetensi yang telah ditetapkan. Keberhasilan mengajar guru terkait dengan sejauh mana guru sebagai manajer belajar peserta didik dalam hal merencanakan, mengelola, memimpin, dan mengevaluasi.

Pengembangan evaluasi pembelajaran serta peningkatan kualitas guru seharusnya mendapat perhatian untuk perbaikan dan peningkatan mutu pendidikan. Evaluasi adalah kegiatan untuk mengumpulkan informasi tentang bekerjanya sesuatu, yang selanjutnya informasi tersebut digunakan untuk menemukan alternatif yang tepat dalam mengambil sebuah keputusan. Evaluasi proses pembelajaran harus dilakukan secara berkesinambungan mulai sejak peserta didik berinteraksi dalam pembelajaran sampai akhir pembelajaran. Untuk melaksanakan evaluasi proses pembelajaran diperlukan alat atau instrumen yang biasa disebut tes. Evaluasi hasil pembelajaran dapat dilakukan bila pendidik sudah dapat menyusun tes sebagai alat untuk mengukur hasil belajar. Bentuk tes yang digunakan tergantung dari kegiatan apa yang hendak dinilai atau kompetensi apa yang ingin dicapai.

Melalui evaluasi terhadap proses pembelajaran, guru memperoleh informasi tentang pencapaian keberhasilan peserta didik kesulitan-kesulitan yang ditemui, kesalahan-kesalahan yang dilakukan dan memberikan perbaikan serta sekaligus untuk merevisi program pembelajaran. Evaluasi yang dapat memberikan masukan pada guru terhadap kelemahan dan kekuatan proses pembelajaran yang telah dilaksanakan disebut dengan *evaluasi formatif*. Disamping itu fungsi lain dari evaluasi formatif juga dapat memberikan pertimbangan pada program pengajaran yang sedang berlangsung, hal ini dikatakan oleh Bhola (2002) bahwa : *formative evaluation is a method of judging the worth of a program while the program activities are forming or happening, formative evaluation focuses on the process*. Biasanya evaluasi formatif ini dilakukan oleh guru disetiap akhir satuan pelajaran. Bagi guru yang jarang melakukan evaluasi formatif akan mengalami kesulitan memilih pendekatan mengajar mana yang tepat digunakan dalam proses pembelajaran berikutnya.

Supaya kegiatan evaluasi dapat memberikan informasi yang tepat tentang pencapaian keberhasilan dan kesulitan yang dialami peserta didik dan kelemahan proses pembelajaran oleh guru untuk dilakukan perbaikan maka diperlukan instrumen evaluasi atau tes yang berkualitas. Menurut Azwar (1996) untuk memperoleh tes yang baik dan berkualitas dilakukan melalui analisis butir soal guna mengetahui validitas, reliabilitas, taraf kesukaran dan daya beda tes. Banyak mahasiswa calon guru yang mengalami kesulitan melakukan analisis butir soal



karena memerlukan konsep statistik dan analisis data yang cukup banyak. Analisis kualitas tes merupakan suatu tahap yang harus ditempuh untuk mengetahui derajat kualitas suatu tes, baik tes secara keseluruhan maupun butir soal yang menjadi bagian dari tes tersebut. Tes sebagai alat evaluasi diharapkan menghasilkan nilai yang objektif dan akurat. Jika tes yang digunakan guru kurang baik, maka hasil yang diperolehpun tentunya kurang baik. Oleh sebab itu, tes yang digunakan guru harus memiliki kualitas yang lebih baik dilihat dari berbagai segi. Tes hendaknya disusun sesuai dengan prinsip dan prosedur penyusunan tes. Setelah digunakan perlu diketahui apakah tes tersebut berkualitas baik atau kurang baik. Untuk mengetahui apakah suatu tes yang digunakan termasuk baik atau kurang baik, maka perlu dilakukan analisis butir menggunakan data dari beberapa peserta didik atau sampel kecil (Millman dan Greene, 1993).

Analisis tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks. Indeks tingkat kesukaran ini pada umumnya dinyatakan dalam bentuk proporsi yang besarnya berkisar 0 - 1 (Aiken, 1994). Semakin besar indeks tingkat kesukaran yang diperoleh dari hasil hitungan, berarti semakin mudah soal dan sebaliknya makin kecil daya beda makin sulit soal itu. Suatu butir tes dikatakan valid apabila mampu mengukur apa yang mau diungkap dari variabel tersebut secara tepat. Menurut Arikunto (2012) pengujian validitas biasanya dilakukan dengan mengkorelasikan skor tiap butir dengan skor keseluruhan butir menggunakan rumus teknik korelasi *product moment* angka kasar. Daya pembeda soal adalah kemampuan suatu butir soal dapat membedakan peserta didik yang telah menguasai materi yang ditanyakan dan peserta didik yang belum menguasai materi yang ditanyakan. Indeks daya pembeda setiap butir soal biasanya juga dinyatakan dalam bentuk proporsi. Semakin tinggi indeks daya pembeda soal berarti semakin mampu soal yang bersangkutan membedakan peserta didik yang telah memahami materi dengan peserta didik yang belum memahami materi. Indeks daya pembeda berkisar antara -1,00 sampai dengan +1,00. Semakin tinggi daya pembeda suatu soal, maka semakin baik soal itu. Reliabilitas tes adalah untuk mengetahui tingkat ketepatan (*precision*) dan keajegan (*consistency*) skor tes. Indeks reliabilitas berkisar antara 0 - 1. Semakin tinggi koefisien reliabilitas suatu tes, makin tinggi pula ketepatannya atau dikatakan tes dapat dipercaya. Untuk melihat kemampuan analisis butir soal perlu diadakan evaluasi, baik evaluasi formatif maupun evaluasi sumatif.

Hasil penelitian Maulana (2018) tentang Analisis Butir Soal Tes Matematika Siswa SMPN1 Indrapuri menemukan soal uraian terdiri dari 20% soal sukar, 0% atau tidak ada butir soal mudah dan 50% butir soal kategori sedang. Daya beda soal uraian terdiri dari 0% atau tidak ada soal masuk kategori cukup, 46,6% butir soal masuk dalam kategori baik, 20% kategori daya beda baik sekali, 6,6% kategori daya beda jelek, Validitas soal uraian 66% valid dan soal tidak valid 33,3%, sedangkan reliabilitas cukup tinggi yaitu 0,76.



Melakukan evaluasi terhadap proses pembelajaran merupakan kemampuan yang harus dikuasai mahasiswa calon guru. Hal ini dijelaskan dalam permendiknas No.16 tahun 2017 bahwa kompetensi pedagogik yang harus dikuasai guru adalah (a) mengembangkan instrumen penilaian dan evaluasi proses dan hasil belajar, (b) mengadministrasikan penilaian proses dan hasil belajar secara berkesinambungan dengan menggunakan berbagai instrumen (c) menganalisis hasil penilaian proses dan hasil belajar untuk berbagai tujuan. Dalam mengembangkan instrumen penilaian perlu dilakukan dengan analisis butir soal baik secara kualitatif maupun secara kuantitatif.

Dalam mengerjakan analisis butir banyak peserta didik yang mempunyai hasil perhitungan yang keliru walaupun sudah menggunakan rumus yang tepat dan benar. Hal ini disebabkan karena ketelitian dari setiap proses perhitungan ketika melakukan pembulatan desimal sehingga hasil akhirnya sangat jauh berbeda dengan hasil sebenarnya. Kesulitan lain yang ditemukan peserta didik adalah ketika menemukan nilai tinggi ordinat pada kurva normal untuk nilai proporsi taraf kesukaran tertentu, ini disebabkan peserta didik belum memahami konsep interpolasi. Guna mengatasi kesulitan dan kesalahan yang dilakukan peserta didik dalam mengerjakan analisis butir soal, maka setelah evaluasi formatif dilaksanakan perbaikan pembelajaran dilakukan dengan bantuan komputer. Penggunaan komputer untuk perhitungan yang sederhana dilakukan dengan bantuan excel dan untuk mengetahui hasil akhir digunakan program iteman 3.0.

Kenyataan yang ditemukan di sekolah Zamzaili (2018) yaitu: (a) hasil kegiatan penyusunan tes, guru hampir tidak menemui kendala dalam memahami materi penyusunan tes yang sesuai dengan kurikulum 2013; (b) guru belum terbiasa menyusun tes hots dari masalah matematika yang sama, tetapi setelah diberi penjelasan tentang kata kerja yang digunakan maka guru sudah dapat membuat berbagai level kognitif tes dari satu masalah matematika yang sama; dan (c) pada materi analisis kualitas tes menggunakan software iteman 3.0 dan iteman 4.3 memang ada kendala yang ditemukan seperti pengetikan hasil jawaban guru pada notepad banyak yang salah sehingga program tidak jalan (tidak direspon iteman). Masih banyak guru yang belum bisa membaca prin out iteman, karena perlu pengetahuan rumus korelasi bagian total untuk interpretasinya.

Evaluasi formatif yang diterapkan sebagai perlakuan dalam penelitian ini adalah evaluasi formatif berbantuan komputer. Evaluasi formatif berbantuan komputer yang digunakan untuk mengerjakan analisis butir soal dapat mengurangi kesalahan-kesalahan yang dilakukan oleh peserta didik dan akan membantu peningkatan kemampuan analisis butir soal. Dengan mengadakan evaluasi formatif dapat dikumpulkan data atau informasi tentang kekuatan dan kelemahan yang terjadi dalam proses pembelajaran. Data ini merupakan bahan untuk melakukan perbaikan proses pembelajaran. Evaluasi formatif berbantuan komputer dilakukan pada kelompok kecil siswa yang mengalami kesalahan pada setiap butir soal diatas 60%. Perbaikan terhadap kesalahan yang dilakukan mahasiswa dimulai dari penjelasan memahami pembentukan konsep dan keterampilan menghitung analisis butir. Penelitian ini dimaksudkan untuk



melaksanakan evaluasi formatif berbantuan komputer untuk mengetahui: (1) besar sumbangan efektif evaluasi formatif berbantuan komputer terhadap kemampuan analisis butir; (2) pengaruh kemampuan statistik terhadap kemampuan analisis butir soal; dan (3) perbedaan rata rata kemampuan analisis butir mahasiswa yang diajar dengan evaluasi formatif berbantuan komputer dan rata-rata yang diajar tanpa bantuan komputer.

## 2. Metode

Penelitian ini dilaksanakan pada FKIP Universitas Bengkulu. Penelitian dilakukan hampir mendekati satu semester, yaitu mulai Juli 2020 sampai dengan bulan Nopember 2020. Disain penelitian ini adalah desain faktorial  $2 \times 2$ , ini disesuaikan dengan data yang diharapkan adanya perbedaan kemampuan analisis butir soal sebagai akibat dari perlakuan yang diberikan. Dalam penelitian ini dilibatkan tiga variabel, yakni variabel bebas, variabel atribut, dan variabel terikat. Variabel bebas adalah evaluasi formatif berbantuan komputer, yang dibedakan menjadi dua kelompok, yaitu evaluasi formatif dengan komputer sebagai kelompok eksperimen dan evaluasi formatif tanpa komputer sebagai kelompok kontrol. Sebagai variabel atribut adalah kemampuan statistik. Gambaran disain penelitian ini seperti pada tabel berikut.

**Tabel 1. Rancangan Penelitian**

Evaluasi Formatif Kemampuan Statistik	Dengan Komputer (A1)	Tanpa Komputer (A2)
Kemampuan statistik tinggi (B1)	A1B1	A2B1
Kemampuan statistik rendah (B2)	A1B2	A2B2

Teknik pengambilan sampel dilakukan dengan teknik *simple random sampling*. Dari 4 kelas pada prodi pendidikan matematika semester 5 dan semester 7 dipilih dua prodi yang digunakan dalam penelitian ini. Instrumen mengukur kemampuan analisis butir soal adalah berupa tes yang dikembangkan sebanyak 8 butir soal essay. Bentuk tes yang dikembangkan adalah tes uraian. Untuk menilai validitas isi tes kemampuan analisis butir digunakan 4 orang panelis yang menguasai materi analisis butir. Hasil penilaian panelis menunjukkan bahwa kualitas butir tes kemampuan analisis butir yang dinilai berkisar antara *cukup sesuai* dan *sesuai*. Oleh karena itu, pada tahap ini hanya ada revisi terhadap aspek kognitif, dan beberapa catatan dari panelis yang dipertahankan untuk revisi butir tes pada tahap selanjutnya. Koefisien reliabilitas hasil penilaian panelis dihitung dengan menggunakan rumus anava Hoyt. Dari hasil perhitungan diperoleh koefisien reliabilitas penilaian panelis sebesar 0,849 dengan galat baku pengukuran 1,499. Hal ini berarti bahwa tes yang dikembangkan memiliki tingkat konsistensi yang cukup tinggi, karena bila dalam jangka waktu tertentu dan dalam kondisi yang sama para panelis menilai kembali tes tersebut maka keragaman skor total yang dicapai hanya berkisar  $\pm 1,5$ . Jadi bila pada penilaian awal seseorang



panelis memiliki skor total 65, maka pada penilaian berikutnya kemungkinan penyimpangan skornya adalah  $1,5 \pm 65$  atau skor total yang akan dicapai berkisar antara 63,5 sampai dengan 66,5.

Berdasarkan beberapa kriteria penilaian butir tes yang telah ditetapkan, maka dari hasil analisis butir yang diuji terhadap 8 butir, diperoleh 7 butir tes dapat digunakan, karena memenuhi semua kriteria validitas, daya beda, taraf kesukaran dan reliabilitas yang telah ditetapkan, sedangkan 1 butir lainnya dinyatakan gugur karena tidak memenuhi taraf kesukaran dan daya beda. Butir soal no.1 pada pokok bahasan taraf kesukaran butir karena memiliki daya beda 0,2812. Jumlah butir yang dapat digunakan pada setiap pokok bahasan seperti pada Tabel 2.

**Tabel 2. Hasil Uji Coba Tes Kemampuan Analisis Butir**

No	p	r	D	$r_{11}$	Ket
1	0.5625	0.6159	<b>0.2812</b>	0.9452	<b>gugur</b>
2	0.6063	0.8641	0.3320		dipakai
3	0.5688	0.9820	0.3273		dipakai
4	0.5813	0.9068	0.4940		dipakai
5	0.6313	0.9661	0.5780		dipakai
6	0.5500	0.9028	0.4268		dipakai
7	0.6000	0.6713	0.3133		dipakai
8	0.6313	0.9661	0.5780		dipakai

Validasi tes kemampuan butir analisis soal dilakukan secara teoritis (isi) maupun empiris. Validasi isi dilakukan dengan penilaian panelis yang terdiri dari dua orang guru dan dua orang dosen ahli bidang studi. Dari hasil penilaian panelis dilakukan penyempurnaan tes tahap pertama. Validasi empiris atau uji coba pada siswa bukan kelas sampel dilakukan dengan analisis butir untuk dimaksudkan untuk menentukan tingkat kesukaran butir tes, daya beda butir tes, validitas butir dengan koefisien biserial titik dan koefisien reliabilitas tes.

Data hasil belajar geometri dianalisa dengan teknik statistik deskriptif dan statistik inferensial. Statistik deskriptif digunakan untuk deskripsi data, sedangkan statistik inferensial digunakan untuk menguji persyaratan analisis data dan menguji hipotesis penelitian. Teknik analisis yang digunakan untuk menguji hipotesis penelitian adalah analisis varian  $2 \times 2$ . Pada dasarnya teknik analisis tertentu baru dapat diterapkan, apabila persyaratan yang mendasarinya telah terpenuhi. Uji persyaratan analisis ini dimaksudkan untuk mengetahui apakah data yang sudah terkumpul memenuhi persyaratan untuk di analisis. Pengujian hipotesis dalam penelitian ini menggunakan analisis varians dua jalur (*disain faktorial*  $2 \times 2$ ). Asumsi yang harus dipenuhi dengan analisis varians dua jalur adalah data penelitian harus berdistribusi normal dan varians antar data ubahan penelitian harus homogen.

Analisis data menggunakan anava dua jalur, apabila hasil analisis menunjukkan adanya interaksi, maka dilanjutkan dengan uji Tuckey untuk melihat efek interaksi atau efek sederhana (*simple effect*) mana yang unggul. Apabila terjadi efek sederhana yang berlawanan, maka disimpulkan adanya

interaksi antara evaluasi formatif dengan kemampuan statistik dalam pengaruhnya terhadap kemampuan analisis butir soal. Artinya bagi mahasiswa yang memiliki level berpikir rendah yang mengikuti evaluasi formatif berbantuan komputer memiliki hasil kemampuan analisis butir soal yang lebih tinggi dari pada siswa yang mengikuti evaluasi formatif tanpa bantuan komputer. Sebaliknya bagi siswa yang memiliki level berpikir tinggi yang mengikuti evaluasi formatif dengan komputer memiliki kemampuan analisis butir soal yang lebih tinggi dari pada siswa yang mengikuti evaluasi formatif dengan bantuan komputer.

### 3. Hasil dan Pembahasan

Hasil tes kemampuan analisis butir pada kelas eksperimen dan kontrol diperoleh data A1 dan A2, sedangkan dari dokumen diperoleh data kemampuan statistic B1 dan B2. Analisis data penelitian ini menggunakan desain faktorial 2 x 2 dengan Anava dua jalur, oleh karena itu kemampuan statistic mahasiswa dikelompokkan menjadi level tinggi dan level rendah sesuai dengan anava dua jalur. Uji normalitas data penelitian dilakukan dengan menggunakan uji Lilliefors. Hasil perhitungan uji normalitas didapat harga L hitung (Lo) untuk ke empat kelompok data lebih kecil dari harga L tabel (Lt) pada taraf signifikansi  $\alpha = 0,05$ . Hal ini berarti bahwa ke empat kelompok data A1 A2, B1, B2 berasal dari populasi yang berdistribusi normal. Pengujian homogenitas varians dilakukan dengan menggunakan uji Levene. Dari perhitungan diperoleh nilai F hitung Levene Statistic = 1,676 untuk db (3,28), sedangkan nilai F table pada taraf signifikansi  $\alpha = 0,05$  adalah 2,947. Karena nilai F hitung keil dari nilai F table, berarti ke empat kelompok data yang diuji mempunyai varians yang sama atau homogen.

**Tabel 3. Data Kemampuan Analisis butir**

<b>Statistik</b>	<b>A<sub>1</sub></b>	<b>A<sub>2</sub></b>	<b>B<sub>1</sub></b>	<b>B<sub>2</sub></b>
N	16	16	16	16
Mean	65.625	59.125	73.75	51
Median	67.5	60	76.5	51
S D	17.9958	10.0987	10.6864	7.6245
Varians	323.85	101.9833	114.2	58.1333
Skor mak	40	45	52	40
Skor min	86	80	86	62
Jumlah	1050	946	1180	816

#### Pengujian Hipotesis

Hipotesis pertama ‘terdapat sumbangan efektif evaluasi formatif berbantuan komputer terhadap kemampuan analisis butir’ digunakan analisis regresi sederhana diperoleh nilai  $F_h = 12,678$  dan nilai  $F_t = 4,600$ , karena  $F_h > F_t$  berarti terdapat pengaruh yang signifikan evaluasi formatif berbantuan computer



terhadap kemampuan analisis butir. Besarnya sumbangan evaluasi formatif berbantuan computer terhadap kemampuan analisis butir 47,52%

**Tabel 4. Anava Dua Jalur untuk Kemampuan Analisisbutir**

<i>SV</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
Sample A	4232	1	4232	78.9184	0.0000	4.19597
Columns B	364.5	1	364.5	6.7972	0.0145	4.19597
Interaction AxB	800	1	800	14.9184	0.0000	4.19597
Within	1501.5	28	53.625			
Total	6898	31				

Perhitungan hipotesis kedua dilihat pada baris interaksi (AxB) menunjukkan bahwa nilai F hitung = 14,918 lebih besar dari F tabel = 4, 196 untuk taraf signifikansi 5 % dengan derajat kebebasan (dk) = (1, 28) ( $F_h = 14,918 > F_t(0,05) (1,124) = 4,196$  berarti  $H_0$  ditolak dan  $H_1$  diterima, ada interaksi antara kemampuan statistika dengan evaluasi formatif terhadap kemampuan analisis butir, atau kemampuan statistika mempengaruhi kemampuan analisis butir.

Hipotesis ketiga dengan menggunakan Anava dua jalur diperoleh hasil nilai F hitung = 78,198 lebih besar dari nilai F tabel = 4,196 untuk taraf signifikansi 5 %). Ini berarti bahwa  $H_0$  ditolak dan  $H_1$  diterima. Dengan demikian terdapat perbedaan yang sangat signifikan kemampuan analisis butir yang diajar dengan evaluasi formatif berbantuan komputer dan evaluasi formatif tanpa komputer.

### **Pembahasan**

Hasil perhitungan dengan menggunakan anava dua jalur menunjukkan adanya interaksi antara kemampuan statistika dengan evaluasi formatif, hal ini menunjukkan kedua perlakuan dalam penelitian ini mempunyai keunggulan. Mahasiswa yang memiliki kemampuan statistika rendah lebih menyukai perhitungan analisis butir dengan rumus statistika tanpa bantuan komputer hal ini dapat dilihat dari nilai rata rata  $A_2B_2 = 52,625$  dan rata rata  $A_1B_2 = 49,375$ , Sedangkan mahasiswa yang memiliki kemampuan statistika tinggi lebih menyukai perhitungan dengan bantuan computer dapat dilihat dari nilai rata rata  $A_1B_1 = 82,375$  dan rata rata  $A_2B_1 = 65,625$ .

## **4. Simpulan dan Saran**

### **Simpulan**

Berdasarkan analisis data penelitian maka dapat disimpulkan: (1) terdapat sumbangan efektif evaluasi formatif berbantuan komputer terhadap kemampuan analisis butir sebesar soal mahasiswa sebesar 47,52%; (2) Terdapat pengaruh kemampuan statistik terhadap kemampuan analisis butir; dan (3) Rata-rata kemampuan analisis butir mahasiswa yang diajar dengan evaluasi formatif berbantuan komputer lebih tinggi dari rata-rata kelas diajar tanpa bantuan komputer.



## Saran

Dari hasil penemuan penelitian disarankan: (1) penelitian ini berusaha mengungkap beberapa usaha yang dapat dilakukan untuk meningkatkan kemampuan analisis butir soal. Temuan bahwa evaluasi formatif berbantuan komputer dapat meningkatkan kemampuan analisis butir soal, maka informasi ini akan merupakan masukan berharga bagi guru pada umumnya dan dosen pada khususnya agar dapat menerapkan evaluasi formatif berbantuan komputer dalam proses pembelajaran; (2) bahwa evaluasi formatif berbantuan komputer dapat membetulkan kesalahan hitung yang dilakukan mahasiswa dengan proses manual, sehingga evaluasi formatif berbantuan komputer perlu dipelajari oleh semua mahasiswa; dan (3) apabila ditemukan bahwa evaluasi formatif berbantuan komputer lebih cocok untuk mahasiswa yang memiliki kemampuan statistika rendah, maka nantinya dalam proses pembelajaran diberlakukan evaluasi formatif tanpa bantuan komputer untuk mahasiswa yang memiliki kemampuan berpikir tinggi.

## Daftar Pustaka

- Aiken, L.R. (1994). *Psychological Testing and Assesment*. Massachusetts: Allyn and Bacon Inc.
- Anastasi, Anne & Urbina, (1997), *Psychological Testing*. New York: Prentice-Hall, Inc.
- Arikunto, S. (2012). *Dasar dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara
- Azwar, S. (1996). *Tes Prestasi Belajar*. Yogyakarta: Pustaka Pelajar.
- Bhola, (2002), *What is Formative Evaluation*, p.1 ([http : // www. sil. org/lingua links/literacy/ formative evaluation. Html](http://www.sil.org/lingua/links/literacy/formative%20evaluation.html)).
- Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test, Theory*. New York: Holt, Rinehart and Winston, Inc.
- Cronbach, L. (1984) *Essential of Psychological Testing*, New York: Harper and Row.
- Djaali, Pudji, M., & Ramly. (2000), *Pengukuran Dalam Bidang Pendidikan*, Jakarta: PPS Universitas Negeri Jakarta.
- Frankel.& Wallen. (1993). *How to Diesign and Evaluate Research in Education*. New York: McGraw Hill. Inc.
- Linn, R. L. & Gronlund, N. E. (1995). *Measurement and Assessment in teaching* (Seventh Edition). Ohio: Merrill, an imprint of Prentice Hall.
- Maulana, F. (2018) *Analisis Butir Soal Tes Matematika Siswa SMPN 1 Indrapuri Tahun Pelajaran 2017/2018*. Skripsi UIN Ar-raniry Aceh
- Millman, J. & Greene, J. (1993). *The Spesification and Development of Tests of Achiiement and Ability in Robert L. Lin (Editor)*. Educational Measurement, Third Edition. Phoenix: American Council on Education, Series on Higher Education Oryx Press.
- Nitko, A.J. (1996). *Educational Assessment of Students, Second Edition*. Ohio: Merrill an imprint of Prentice Hall Englewood Cliffs



Rosana, D. & Setyawarno, D. (2016), *Statistik Terapan untuk Penelitian Bidang Pendidikan*. Yogyakarta: FMIPA UNY

Tessmer, M. (1995), *Planing and Conducting Formative Evaluation*. London: Kogan Page Limited

Zamzaili (2018) *Pelatihan Penyusunan Tes Dan Analisis Kualitas Tes Berbantuan Software Iteaman Bagi Guru MAN Kota Bengkulu*. Laporan PPM FKIP Univ Bengkulu.