

# PENGEMBANGAN PERANGKAT LUNAK KOMPUTER UNTUK MENGEVALUASI SOAL TES

Hindayati Mustafidah\*

Program Pendidikan Matematika Universitas Muhammadiyah Purwokerto

**Abstract:** *This research aims to develop computer software used to evaluate some test items such as calculation and determination for validity, reliability, difference power, and difficulty indexes. This software is developed using C++ programming language. This computer software is called "evaluasi". An algorithm is designed before making a thinking runs easily at the program making. This software consists of two menus that can be chosen for the kind of test items. Two menus are for data entry for student's test scores, and four menus are for evaluating the kind of test items. The kinds of items consist of a multiple choice test and an essay test. There are two ways for inputting data, from the keyboard or from the data which were saved in file. This system prepares four types of evaluation that can be selected by user. They are validity, reliability, difference power, and difficulty indexes; each type of them is for multiple choice item and essay item.*

**Kata kunci:** program komputer, soal tes, perangkat lunak, indeks kesukaran, program evaluasi

## PENDAHULUAN

Keberhasilan pendidikan sebagian besar bergantung pada dua faktor, yaitu pengajar (dalam hal ini guru) dan siswa. Untuk mengetahui tingkat keberhasilan pendidikan, dilakukan evaluasi pendidikan. Evaluasi pendidikan di tingkat sekolah bisa dilaksanakan dengan mengevaluasi tingkat penguasaan materi siswa terhadap suatu mata pelajaran. Tingkat penguasaan materi pelajaran ini dapat diketahui dengan suatu metode dan alat. Metode yang bisa digunakan misalnya ujian tulis, ujian lisan, tugas-tugas, ujian pendek/singkat (kuis), dan sebagainya; sedangkan alat yang digunakan misalnya soal-soal yang berbentuk uraian (*essay*) ataupun pilihan ganda (*objective*) beserta hasil pekerjaannya. Pene-tapan metode atau alat tersebut bergantung antara lain pada jumlah siswa, karakteristik

mata pelajaran, waktu yang tersedia, dan kemampuan tenaga pengajar.

Sebelum alat tes hasil belajar suatu mata pelajaran digunakan dalam pengukuran, alat tes hasil belajar tersebut harus telah dipertimbangkan secara masak. Tes hasil belajar tersebut harus sudah dikenal ciri-cirinya menurut tipenya, telah direncanakan, dan telah disusun berdasarkan petunjuk-petunjuk penyusunannya. Selanjutnya setelah alat tes hasil belajar tersebut digunakan dalam pengukuran, maka seorang guru memperoleh umpan-balik tentang mutu alat tes hasil belajar tersebut, yaitu apakah alat tes yang dianggap bermutu sebelum suatu pengukuran, akan tetap bermutu setelah pengukuran dilaksanakan. Umpan-balik mengenai mutu alat tes hasil belajar akan diperoleh, apabila seorang guru memeriksa skor-skor yang diperoleh

---

\*Alamat korespondensi: Dampit RT 02/07 Dukuhwaluh, Kembaran, Banyumas 53182

dari suatu pengukuran tersebut, Pemeriksaan tersebut meliputi pemeriksaan analisis item-item tes, pemeriksaan taraf reliabilitas tes, pemeriksaan validitas tes, dan pemeriksaan praktibilitas tes (Masidjo, 2004) atau yang sering disebut sebagai evaluasi soal tes.

Kenyataan yang ada, soal tes yang telah disusun jarang sekali dilakukan evaluasi. Salah satu penyebabnya adalah karena banyaknya rumus yang digunakan untuk menghitung, sementara waktu yang tersedia sangat terbatas. Di samping hasil penghitungan dengan tangan (secara manual) sangat beresiko terhadap kesalahan pemasukan data, juga hasil perhitungan yang didapatkan kurang teliti. Berdasarkan hasil survei di Program Studi Pendidikan Matematika, mahasiswa yang telah mengambil tugas akhir untuk materi evaluasi soal ulangan umum di beberapa sekolah baik SLTP, SMU, maupun SMK, perhitungan untuk mengevaluasi soal tes 99% dilakukan secara manual. Hal ini merupakan salah satu penghambat dalam menyelesaikan tugas akhir, karena waktu banyak tersita untuk melakukan perhitungan-perhitungan analisis soal. Demikian juga yang terjadi di sekolah-sekolah juga perguruan tinggi, pengajar (guru atau dosen) enggan melakukan evaluasi terhadap soal-soal tes yang telah disunnya; dengan alasan yang sama, yaitu terlalu menyita waktu, tenaga, dan pikiran dalam melakukan perhitungan padahal waktu yang tersedia sangat terbatas; sedangkan tugas dan kegiatan yang lain masih banyak yang belum dikerjakan.

Masalah ini dapat diatasi jika kita bisa memanfaatkan salah satu hasil perkembangan ilmu pengetahuan dan teknologi (IPTEK) yang berupa perangkat komputer. Di samping dari segi kecepatan yang tinggi, komputer juga lebih teliti dalam penghitungan angka atau bilangan dibandingkan dengan hasil penghitungan secara manual. Tidak atau kurang dimanfaatkannya komputer dalam membantu pekerjaan manusia disebabkan tidak atau kurang mengerti dalam mengoperasikan komputer itu sendiri. Sebagai contoh dalam bidang pen-

didikan, untuk melakukan evaluasi terhadap soal tes, dalam penentuan tingkat validitas, reliabilitas, indeks kesukaran, dan daya pembeda soal tidak dilakukan dengan bantuan perangkat komputer. Hal ini disebabkan tidak tahu bagaimana menggunakan komputer untuk keperluan evaluasi soal tersebut. Hal ini mengakibatkan pekerjaan mengevaluasi soal tes terasa sangat berat dan sangat menjemukan, di samping sangat menyita waktu. Kalau keadaan seperti ini dibiarkan berlarut-larut, maka soal-soal tes yang digunakan sebagai alat ukur tingkat penguasaan materi siswa dan tingkat keberhasilan guru dalam mengajar tidak bisa menjadi pengukur sebagaimana yang diharapkan. Hal ini disebabkan karena soal tes itu sendiri belum diketahui kualitasnya sebagai alat ukur.

Berdasarkan uraian di atas, permasalahan dalam penelitian ini adalah bagaimana mengembangkan sebuah perangkat lunak (program komputer) yang mempunyai kemampuan untuk menghitung dan menentukan tingkat validitas, reliabilitas, indeks kesukaran, dan daya pembeda soal ujian atau tes. Dengan menggunakan bantuan program komputer, di samping lebih teliti hasil perhitungannya, tidak perlu menghitung secara langsung dengan rumus-rumus matematika yang memerlukan waktu dan tenaga banyak, sehingga pekerjaan mengevaluasi soal tidak lagi menjadi beban berat yang harus dipikul.

## METODE PENELITIAN

Penelitian ini dilakukan di Laboratorium Pengembangan Pendidikan Matematika - FKIP - UMP. Alat yang dipergunakan adalah sebuah komputer (PC) bersistem operasi Windows' 98 dengan perangkat lunak Turbo C++ 3.0 untuk mengimplementasikan ke dalam program komputer.

Adapun langkah-langkah operasional penelitian yang telah dilakukan adalah:

1. Mengkonversikan formula matematika penghitungan validitas, reliabilitas, indeks kesukaran, dan daya pembeda soal baik yang berbentuk soal pilihan ganda

maupun soal uraian ke dalam bahasa komputer. Soal yang berbentuk pilihan ganda penghitungan validitasnya menggunakan rumus sebagai berikut (Arikunto, 2002):

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\{N\sum X^2 - (\sum X)^2\}\{N\sum Y^2 - (\sum Y)^2\}}}$$

Keterangan :

$r_{xy}$  : Koefisien korelasi bentuk soal  
 $N$  : Banyaknya siswa  
 $X$  : Skor butir soal nomor tertentu  
 $Y$  : Skor total

Untuk menghitung validitas soal uraian digunakan rumus Korelasi *Product Moment* (Waridjan, 1991) sebagai berikut:

$$r_{xy} = \frac{\sum xy}{\sqrt{\{(\sum x^2) (\sum y^2)\}}}$$

Keterangan :

$r_{xy}$  = Koefisien korelasi antara x dan y  
 $x = X - \bar{X}$   
 $X$  = Skor butir soal nomor tertentu  
 $\bar{X}$  = Rata-rata skor butir soal nomor tertentu  
 $y = Y - \bar{Y}$   
 $Y$  = Skor total  
 $\bar{Y}$  = Rata-rata skor total

Hasil perhitungan validitas, kemudian digolongkan ke dalam kriteria sebagai berikut:

0,80 <  $r_{xy}$  ≤ 1,00, adalah validitas sangat tinggi  
 0,60 <  $r_{xy}$  ≤ 0,80, adalah validitas tinggi  
 0,40 <  $r_{xy}$  ≤ 0,60, adalah validitas cukup  
 0,20 <  $r_{xy}$  ≤ 0,40, adalah validitas rendah  
 0,00 <  $r_{xy}$  ≤ 0,20, adalah validitas sangat rendah

$r_{xy} \leq 0,00$ , adalah tidak valid  
 (Waridjan, 1991)

Menurut Arikunto (2002), reliabilitas instrumen untuk soal objektif dapat dicari dengan menggunakan rumus KR-20 sebagai berikut:

$$r_{11} = \left[ \frac{k}{k-1} \right] \left[ \frac{V_t - \sum pq}{V_t} \right]$$

$r_{11}$  = Reliabilitas instrumen  
 $k$  = Banyak butir soal  
 $V_t$  = Variansi total  
 $p$  = Proporsi yang menjawab butir soal dengan betul (proporsi objek yang mempunyai skor)  
 $q$  = Proporsi subjek yang mempunyai skor 0 ( $q = 1 - p$ )

Menurut Suherman (2001: 167) untuk soal uraian digunakan rumus sebagai berikut:

$$r_{11} = \left( \frac{n}{n-1} \right) \left( 1 - \frac{\sum S_t^2}{S_t^2} \right)$$

Keterangan:

$N$  = Banyak butir soal  
 $\sum S_t^2$  = Jumlah variansi skor tiap item  
 $S_t^2$  = Variansi skor total

Tolok ukur untuk mengukur derajat reliabilitas alat evaluasi dapat menggunakan tolok ukur sebagai berikut:

0,00 <  $r_{11}$  ≤ 0,20, derajat reliabilitas sangat rendah  
 0,20 <  $r_{11}$  ≤ 0,40, derajat reliabilitas rendah  
 0,40 <  $r_{11}$  ≤ 0,60, derajat reliabilitas sedang  
 0,60 <  $r_{11}$  ≤ 0,80, derajat reliabilitas tinggi  
 0,80 <  $r_{11}$  ≤ 1,00, derajat reliabilitas sangat tinggi

Penghitungan indeks kesukaran soal objektif digunakan rumus menurut Arikunto (2002):

$$IK = \frac{B}{N}$$

Keterangan :

IK = Indeks kesukaran untuk tiap butir soal  
 $B$  = Banyaknya siswa yang menjawab benar setiap butir soal

N = Banyaknya siswa yang memberikan jawaban pada soal yang dimaksudkan

Untuk mencari indeks kesukaran untuk soal uraian digunakan rumus (Suherman, 2001):

$$IK = \frac{A + B - (2NS_{\min})}{2N(S_{\max} - S_{\min})}$$

IK = Indeks kesukaran

A = Jumlah skor kelompok atas

B = Jumlah skor kelompok bawah

N = Siswa kelompok atas atau bawah

$S_{\max}$  = Skor tertinggi tiap soal uraian

$S_{\min}$  = Skor terendah tiap soal uraian

Klasifikasi taraf kesukaran :

IK < 0,00, soal terlalu sukar

0,00 < IK < 0,30, soal sukar

0,30 ≤ IK < 0,70, soal sedang

0,70 ≤ IK < 1,00, soal mudah

IK = 1,00, soal terlalu mudah

Untuk mencari daya pembeda soal objektif digunakan rumus menurut Arikunto (2002):

$$Dp = \frac{B_a}{J_a} - \frac{B_b}{J_b}$$

Keterangan :

Dp = Daya pembeda

$J_a$  = Jumlah siswa kelompok atas

$J_b$  = Jumlah siswa kelompok bawah

$B_a$  = Banyak siswa kelompok atas yang menjawab benar

$B_b$  = Banyak siswa kelompok bawah yang menjawab benar.

Untuk mencari daya pembeda untuk soal uraian digunakan rumus (Suherman, 2001):

$$Dp = \frac{A - B}{N(S_{\max} - S_{\min})}$$

Keterangan :

Dp = Daya pembeda

A = Jumlah skor kelompok atas

B = Jumlah skor kelompok bawah

N = Jumlah siswa kelompok atas atau bawah

$S_{\max}$  = Skor tertinggi setiap soal uraian

$S_{\min}$  = Skor terendah setiap soal uraian

Klasifikasi daya pembeda :

$Dp \leq 0,00$ , jelek sekali

$0,00 < Dp \leq 0,20$ , jelek

$0,20 < Dp \leq 0,40$ , cukup

$0,40 < Dp < 0,70$ , baik

$0,70 < Dp < 1,00$ , sangat baik

2. Menyusun algoritma pemrograman. Algoritma disusun untuk mempermudah alur pikir pemrograman, dan menunjukkan urutan logika pemrograman.

3. Membuat kode program komputer untuk penghitungan validitas, reliabilitas, indeks kesukaran, dan daya pembeda soal pilihan ganda dan uraian. Kode program disusun menggunakan bahasa pemrograman Turbo C++ versi 3.0.

4. Membuat kasus tes untuk persiapan testing terhadap program komputer yang telah dibuat. Kasus tes ini berupa data nilai siswa.

5. Evaluasi dan testing program komputer.

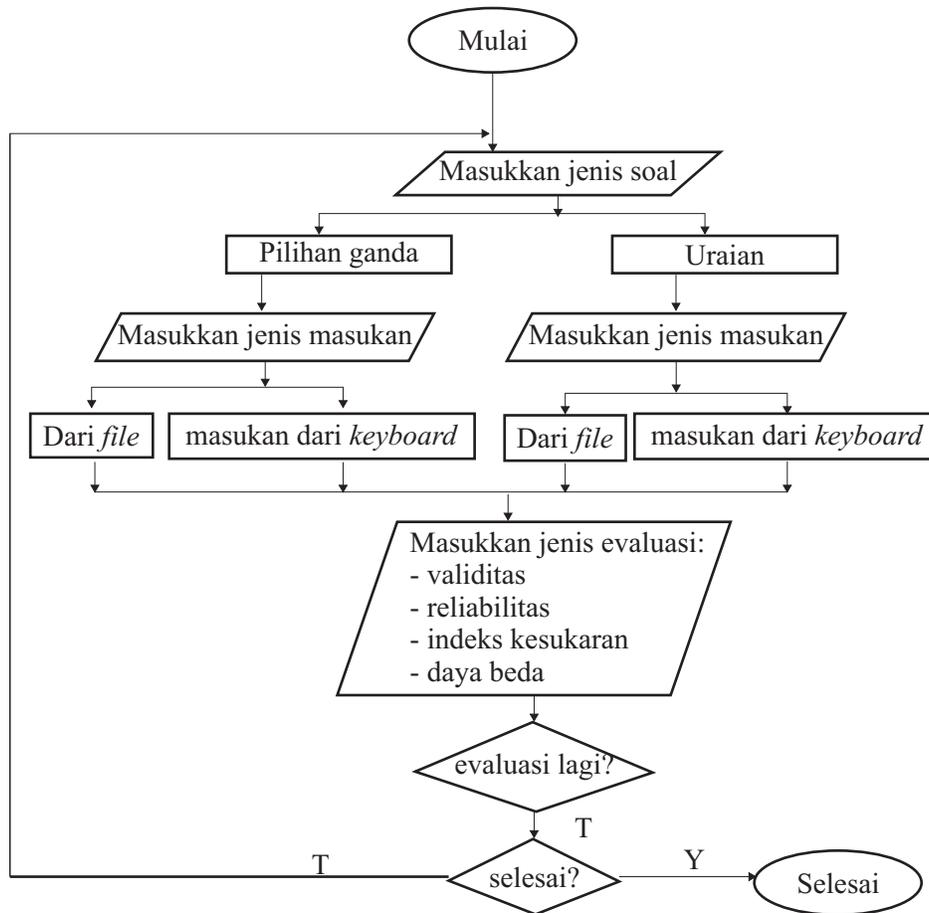
Desain program komputer untuk evaluasi soal tes ini dirancang sebagai berikut :

(1) Program komputer terdiri dari dua pilihan jenis soal (pilihan ganda atau uraian), dua jenis cara pemasukan data (data dari *file* atau data langsung dimasukkan dari papan ketik komputer), dan empat macam pilihan evaluasi, yaitu: validitas, reliabilitas, indeks kesukaran, dan daya pembeda yang bisa dipilih salah satu oleh pemakai; (2) Data nilai siswa berperan sebagai masukan (*input*) program; dan (3) Data nilai yang dimasukkan akan ditampilkan keluaran (*output*) program berupa: (a) tingkat validitas tiap butir soal; (b) tingkat reliabilitas soal; (c) indeks kesukaran tiap butir soal; dan (d) daya pembeda tiap butir soal.

## HASIL DAN PEMBAHASAN

### Algoritma Program

Algoritma umum dari rancangan program yang dihasilkan dapat dilihat pada Gambar 1 berikut ini.



Gambar 1. Rancangan Program

### Hasil Keluaran Program

Tampilan awal program adalah sebagai berikut.

Masukkan jenis soal: 1. Pilihan Ganda  
2. Uraian  
3. Selesai  
Pilihan Anda (1/2/3): 1

PROGRAM EVALUASI SOAL TES  
PROGRAM EVALUASI SOAL TES

Menu ini merupakan menu pilihan jenis soal yang terdiri dari dua macam, yaitu soal pilihan ganda atau soal uraian. Jika dipilih 1, maka yang akan dijalankan adalah evaluasi untuk jenis soal pilihan ganda. Pada menu berikutnya pemakai harus menentukan pilihan apakah data nilai berasal dari sebuah *file* yang sudah disiapkan atau data nilai langsung diketikkan melalui papan ketik komputer saat program ini dijalankan. Jika yang dipilih adalah 1 (*file*), maka sebelumnya pemakai harus sudah mempunyai *file* data nilai yang tersimpan dalam media penyimpanan komputer. Sebagai contoh *file* data yang sudah disiapkan ada-

lah data.txt, maka isian pilihan yang harus dimasukkan adalah sebagai berikut.

Data nilai berupa:  
 1. *File*  
 2. Masukan dari *keyword*

Pilihan Anda: 1

---

Nama *file* datanya?: Data.txt  
 Berapa jumlah siswa/  
 respondennya?: 5

jumlah datanya: 30  
 jumlah butir soal: 6

siswa ke 1: 0 1 1 1 0 1  
 siswa ke 2: 1 0 0 1 1 1  
 siswa ke 3: 1 1 1 1 1 0  
 siswa ke 4: 0 0 1 0 1 0  
 siswa ke 5: 1 1 0 0 1 0

Setelah data terbaca oleh program, selanjutnya pemakai diminta untuk memilih salah satu jenis evaluasi soal sesuai dengan menu berikut ini.

| Jenis Evaluasi      |   |
|---------------------|---|
| * (V)aliditas       | * |
| * (R)eliabilitas    | * |
| * (D)aya Beda       | * |
| * (I)ndek Kesukaran | * |
| * (K)eluar          | * |

Pilih salah satu (V/R/D/I/K):

Jika pemakai memilih menu validitas dengan menekan huruf **V** pada papan ketik, maka keluaran program yang dihasilkan adalah sebagai berikut.

Hasil evaluasi validitas soal pilihan ganda :

Harga Rxy untuk butir soal no. 1 : 0.48038 --> validitas cukup  
 Harga Rxy untuk butir soal no. 2 : 0.48038 --> validitas cukup  
 Harga Rxy untuk butir soal no. 3 : 0.08006 --> validitas sangat rendah  
 Harga Rxy untuk butir soal no. 4 : 0.88070 --> validitas sangat tinggi  
 Harga Rxy untuk butir soal no. 5 : -0.19612 --> butir soal tidak valid  
 Harga Rxy untuk butir soal no. 6 : 0.32026 --> validitas rendah

Jika pemakai memilih menu reliabilitas dengan menekan huruf **R** pada pa-

pan ketik, maka keluaran program yang dihasilkan adalah sebagai berikut.

Hasil evaluasi reliabilitas soal pilihan ganda :  
 r11: -0.055 --> soal tidak reliabel

Jika pemakai memilih menu daya beda soal dengan menekan huruf **D** pada

papan ketik, maka keluaran program yang dihasilkan adalah sebagai berikut.

Hasil evaluasi daya pembeda soal pilihan ganda :

Daya Beda butir soal ke-1 : 0.25 --> daya beda soal cukup  
 Daya Beda butir soal ke-2 : 0.75 --> daya beda soal sangat baik  
 Daya Beda butir soal ke-3 : 0.75 --> daya beda soal sangat baik  
 Daya Beda butir soal ke-4 : 1.00 --> daya beda soal sangat baik  
 Daya Beda butir soal ke-5 : 0.00 --> daya beda soal jelek sekali  
 Daya Beda butir soal ke-6 : 0.50 --> daya beda soal baik

Jika pemakai memilih menu indeks papan ketik, maka keluaran program yang kesukaran dengan menekan huruf **I** pada dihasilkan adalah sebagai berikut.

Hasil evaluasi indeks kesukaran soal pilihan ganda :

Indeks Kesukaran butir soal ke-I : 0.60 --> soal sedang  
 Indeks Kesukaran butir soal ke-2 : 0.60 --> soal sedang  
 Indeks Kesukaran butir soal ke-3 : 0.60 --> soal sedang  
 Indeks Kesukaran butir soal ke-4 : 0.60 -> soal sedang  
 Indeks Kesukaran butir soal ke-5 : 0.80 --> soal mudah  
 Indeks Kesukaran butir soal ke-6 : 0.40 --> soal sedang

Jika pemakai memilih menu keluar dengan menekan huruf **K** pada papan ketik, maka keluaran program yang dihasilkan adalah kembali ke menu utama, yaitu menu untuk menentukan pilihan jenis soal. Jika pemakai memilih jenis soal uraian dengan pilihan masukan data berasal dari *file*, maka tampilan selengkapnya disajikan sebagai berikut sebagai berikut.

Masukkan jenis soal: 1. Pilihan Ganda  
 2. Uraian  
 3. Selesai  
 Pilihan Anda (1/2/3): 2

Data nilai berupa:  
 1. *File*  
 2. Masukan dari *keyword*  
 Pilihan Anda: 1  
 Nama *file* datanya?: Data.txt

Berapa jumlah siswa/  
 respondennya?: 6

jumlah datanya: 30  
 jumlah butir soal: 5  
 siswa ke 1: 2 3 4 2 3  
 siswa ke 2: 1 0 2 5 3  
 siswa ke 3: 2 5 8 2 1  
 siswa ke 4: 3 4 3 4 0  
 siswa ke 5: 8 0 1 4 7  
 siswa ke 6: 9 4 5 8 0

| Jenis Evaluasi      |   |
|---------------------|---|
| * (V)aliditas       | * |
| * (R)eliabilitas    | * |
| * (D)aya Beda       | * |
| * (I)ndek Kesukaran | * |
| * (K)eluar          | * |

Pilih salah satu (V/R/D/I/K):

Hasil evaluasi validitas soal uraian :

Butir soal no. 1: 0.88651 --> validitas sangat tinggi  
 Butir soal no. 2: 0.29817 --> validitas rendah  
 Butir soal no. 3: 0.28682 --> validitas rendah  
 Butir soal no. 4 : 0.56415 --> validitas cukup  
 Butir soal no. 5 : - 0.10251 --> butir soal tidak valid

Hasil evaluasi reliabilitas soal uraian :  
 r11: -0.245 --> soal tidak reliabel

Hasil evaluasi daya pembeda soal uraian :

Daya Beda butir soal ke - 1 : 0.27 --> daya beda soal cukup  
 Daya Beda butir soal ke - 2 : 0.07 --> daya beda soal jelek  
 Daya Beda butir soal ke - 3 : 0.12 --> daya beda soal jelek  
 Daya Beda butir soal ke - 4 : 0.08 --> daya beda soal jelek  
 Daya Beda butir soal ke - 5 : 0.05 --> daya beda soal jelek

---

Hasil evaluasi indeks kesukaran soal uraian :  
Daya Beda butir soal ke - 1 : 0.14 --> soal sukar  
Daya Beda butir soal ke - 2 : 0.27 --> soal sukar  
Daya Beda butir soal ke - 3 : 0.13 --> soal sukar  
Daya Beda butir soal ke - 4 : 0.01 - > soal sukar  
Daya Beda butir soal ke - 5 : 0.17 --> soal sukar

---

Pilihan lain yang bisa dipilih oleh pemakai dalam hal pemasukan data selain data yang berasal dari file yang sudah disediakan adalah masukan data berupa data nilai yang langsung diketikkan dari papan ketik saat program dijalankan. Berikut disajikan contoh tampilan program dengan pilihan pemakai untuk jenis soal uraian dan data dimasukkan dari papan ketik.

Masukkan jenis soal: 1. Pilihan Ganda  
2. Uraian  
3. Selesai

Pilihan Anda (1/2/3): 2

Data nilai berupa:

1. *File*
2. Masukan dari *keyword*

---

Pilihan Anda: 2  
Berapa banyaknya responden (siswa)? : 5  
Berapa butir soalnya ? : 2  
Masukkan data nilai satu per satu  
siswa ke : 1, skor butir soal ke- 1 : 43  
siswa ke : 1, skor butir soal ke- 2 : 45  
siswa ke : 2, skor butir soal ke- 1 : 75  
siswa ke : 2, skor butir soal ke- 2 : 23  
siswa ke : 3, skor butir soal ke- 1 : 90  
siswa ke : 3, skor butir soal ke- 2 : 32  
siswa ke : 4, skor butir soal ke- 1 : 54  
siswa ke : 4, skor butir soal ke- 2 : 10  
siswa ke : 5, skor butir soal ke- 1 : 21  
siswa ke : 5, skor butir soal ke- 2 : 34

Datanya :

43 45  
75 23  
90 32  
54 10  
21 34

---

Hasil evaluasi validitas soal uraian :  
Butir soal no. 1: 0.88160 --> validitas sangat tinggi  
Butir soal no. 2: 0.23463 --> validitas rendah

---

Hasil evaluasi reliabilitas soal uraian :  
r11: -0.494 --> soal tidak reliabel

---

Hasil evaluasi daya pembeda soal uraian :  
Daya Beda butir soal ke - 1 : 0.26 --> daya beda soal cukup  
Daya Beda butir soal ke - 2 : 0.06 --> daya beda soal jelek

---

---

Hasil evaluasi indeks kesukaran soal uraian :  
Daya Beda butir soal ke - 1 : 0.04 --> soal sukar  
Daya Beda butir soal ke - 2 : - 0.00 --> soal terlalu sukar

---

Untuk mengakhiri pelaksanaan menu 3 (selesai) pada saat muncul menu evaluasi soal menggunakan program evaluasi ini, dilakukan dengan cara memilih

Masukkan jenis soal :  
1. Pilihan Ganda  
2. Uraian  
3. Selesai  
Pilihan Anda (1/2/3) : 3

### **KESIMPULAN DAN SARAN**

Penelitian ini telah menghasilkan sebuah perangkat lunak komputer berupa program untuk melakukan evaluasi terhadap soal tes. Program ini diberi nama program "evaluasi". Data nilai hasil pekerjaan siswa yang telah diperoleh digunakan sebagai data masukan oleh program dan keluaran program yang dihasilkan berupa hasil evaluasi soal yang meliputi validitas, reliabilitas, daya pembeda, dan indeks kesukaran. Program ini bisa digunakan untuk mengevaluasi soal baik soal pilihan ganda maupun soal uraian. Saran yang bisa disampaikan adalah:

1. Dengan telah dihasilkannya program evaluasi ini, maka evaluasi soal tidak perlu lagi dilakukan secara manual. Bagi para penyusun soal tes, hendaknya selalu melakukan evaluasi terhadap soal tes yang telah disusunnya, sehingga kualitas soal bisa lebih baik sehingga soal benar-benar bisa dijadikan sebagai alat ukur untuk mengukur kemampuan siswa.
2. Karena program evaluasi ini hanya meliputi jenis evaluasi validitas, reliabilitas, daya pembeda, dan indeks kesukaran, maka bisa dikembangkan lagi untuk jenis evaluasi soal yang lain misalnya praktibilitasnya.

### **DAFTAR PUSTAKA**

- Arikunto, S. (2002). *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara.
- Masidjo, I. (2004). *Penilaian Pencapaian Hasil Belajar Siswa di Sekolah*. Yogyakarta: Kanisius.
- Suherman, E. (2001). *Evaluasi Proses dan Hasil Belajar Matematika*. Jakarta: Universitas Terbuka.
- Waridjan. (1991). *Tes Hasil Belajar Gaya Objektif*. Semarang: IKIP Semarang Press.