
PERBANDINGAN TINGKAT KESUKARAN, DAYA PEMBEDA BUTIR SOAL DAN RELIABILITAS TES BENTUK PILIHAN GANDA BIASA DAN PILIHAN GANDA ASOSIASI MATA PELAJARAN EKONOMI

Nani Hanifah

Program Studi Bimbingan Konseling
Fakultas Ilmu Pendidikan dan Pengetahuan Sosial
Universitas Indraprasta PGRI Jakarta
hanifah.nani@ymail.com

Abstract; The purpose of this research was to determine the difference of the item difficulty levels, the item discrimination powers, and the test reliabilities between the ordinary multiple choice (OMC) test and the association multiple choice (AMC) test. The population of the research was the second grade students of SLTPN 15 Bekasi, consisting 112 respondents. This quasi experimental research had utilized two group, randominized subjects post-test only design, using two treatments for each respondent, by giving them the OMC test and the AMC test after three week. The result of this research showed that the items difficulty levels of the AMC test was higher or the AMC test was more difficult than the OMC test. The research also indicated that the items discrimination powers of the AMC test was higher or the AMC test was more effective to differentiate between the higher ability testee and the lower ability testee than the items the OMC test. This research also indicated that not different test reliability between the AMC test and the OMC test.

Keywords: the difficulty level, the discrimination power, the reliability test, the ordinary multiple choice (OMC), the association multiple choice (AMC).

Abstrak; Penelitian ini bertujuan untuk mengetahui perbedaan tingkat kesukaran, daya pembeda butir soal dan reliabilitas tes bentuk pilihan ganda biasa (PGB) dan pilihan ganda asosiasi (PGA). Penelitian ini dilakukan kepada siswa kelas II SLTP 15 Bekasi di semester 2, yang terdiri dari 112 siswa. Ini adalah penelitian kuasi eksperimen dengan rancangan dua kelompok subyek diacak dengan hanya pemberian tes akhir. Penelitian dilakukan dengan cara setiap siswa dikenai dua kali perlakuan yaitu diberikan tes dengan bentuk pilihan ganda biasa, tiga minggu kemudian diberikan lagi tes dengan bentuk pilihan ganda asosiasi. Hasil penelitian menunjukkan bahwa butir soal tes bentuk pilihan ganda asosiasi memiliki tingkat kesukaran lebih tinggi atau bentuk pilihan ganda asosiasi lebih sukar daripada tes bentuk pilihan ganda biasa. Penelitian ini juga menyimpulkan bahwa butir soal tes bentuk pilihan ganda asosiasi memiliki daya pembeda lebih tinggi atau tes bentuk pilihan ganda asosiasi lebih mampu membedakan peserta tes yang berkemampuan tinggi dan yang berkemampuan rendah daripada tes bentuk pilihan ganda biasa. Disimpulkan juga bahwa tidak ada perbedaan reliabilitas tes antara tes bentuk pilihan ganda asosiasi dan tes bentuk pilihan ganda biasa.

Kata-kata kunci: tingkat kesukaran, daya pembeda, reliabilitas tes, tes bentuk Pilihan ganda biasa (PGB), tes bentuk pilihan ganda asosiasi (PGA).

PENDAHULUAN

Setiap sekolah mengadakan evaluasi (penilaian), mulai dari evaluasi formatif sampai evaluasi sumatif termasuk Ujian Sekolah (US) dan Ujian Nasional (UN). Data hasil penilaian dapat memberikan informasi yang akurat untuk pengambilan keputusan tentang daya serap siswa terhadap materi pelajaran, perbaikan proses belajar pembelajaran, bimbingan dan konseling terhadap siswa yang bermasalah, kenaikan kelas dan kelulusan, apabila diperoleh dari hasil pengukuran prestasi belajar yang dilakukan dengan cermat dengan alat ukur yang disebut tes prestasi belajar yang di dalamnya terdapat tes subyektif dan tes obyektif.

Tes obyektif khususnya bentuk tes pilihan ganda di Indonesia digunakan pada tes formatif dan tes sumatif yang juga di dalamnya terdapat ujian sekolah, ujian nasional dan digunakan juga pada ujian masuk perguruan tinggi. Keunggulan penggunaan bentuk tes pilihan ganda adalah dapat mencakup banyak pokok bahasan dan mudah penyekorannya. Jenis tes bentuk pilihan ganda antara lain pilihan ganda biasa dan pilihan ganda asosiasi. Sekarang ini tes bentuk pilihan ganda asosiasi pada ujian sekolah dan ujian nasional di SD, SMP, SMA/SMK tidak digunakan lagi. Sedangkan pada ujian masuk perguruan tinggi bentuk tes pilihan ganda asosiasi masih digunakan sehingga peserta tes yang pada umumnya lulusan SMA/SMK mengalami kesulitan dalam menjawab soal-soal bentuk pilihan ganda asosiasi karena mereka tidak terbiasa menghadapi bentuk soal tersebut.

Tidak digunakannya bentuk soal pilihan ganda asosiasi di sekolah-sekolah karena soal-soal bentuk pilihan ganda asosiasi dianggap lebih sulit dibandingkan dengan bentuk soal pilihan ganda biasa dalam pengerjaan soal maupun dalam pembuatannya tapi lebih dapat membedakan siswa yang

mempunyai kemampuan tinggi dengan siswa yang mempunyai kemampuan rendah. Apakah benar pandangan ini? Untuk menjawab pertanyaan tersebut maka perlu diadakan penelitian tentang perbandingan tingkat kesukaran, daya pembeda butir soal dan reliabilitas tes bentuk pilihan ganda biasa dan bentuk pilihan ganda asosiasi.

Masalah yang akan dibahas dalam penelitian ini dapat dirumuskan sebagai berikut: (1). Apakah terdapat perbedaan tingkat kesukaran butir soal tes bentuk pilihan ganda biasa dan pilihan ganda asosiasi? (2). Apakah terdapat perbedaan daya pembeda butir soal tes bentuk pilihan ganda biasa dan pilihan ganda asosiasi? (3). Apakah terdapat perbedaan reliabilitas tes bentuk pilihan ganda biasa dan pilihan ganda asosiasi?

TINJAUAN PUSTAKA

Hakikat Tes

Anas Sudijono (2006:93) mengatakan bahwa tes adalah alat atau prosedur yang digunakan dalam rangka pengukuran dan penilaian.

Sedangkan menurut Ratna sayekti (2008:5), tes adalah seperangkat butir atau pernyataan yang dibuat untuk diberikan kepada siswa dengan syarat-syarat tertentu atau tes adalah prosedur yang sistematis untuk mengobservasi tingkah laku.

Cronbach (2004:26) mengemukakan bahwa tes adalah merupakan suatu prosedur yang sistematis untuk mengamati atau mendeskripsikan satu atau lebih karakteristik seseorang dengan menggunakan standar numerik atau sistem kategori.

Menurut Anne Anastasi (2007:25), tes adalah alat pengukur yang mempunyai standar obyektif sehingga dapat digunakan secara meluas serta betul-betul dapat digunakan untuk mengukur dan membandingkan keadaan psikis atau tingkah laku individu.

Crocker dan Algina dalam Suke Silverius (2001:4) mengatakan bahwa tes adalah suatu proses baku untuk memperoleh sampel tingkah laku dari suatu ranah tertentu.

Dari beberapa pengertian tes yang dikemukakan di atas dapat disimpulkan bahwa tes adalah cara atau alat yang digunakan untuk mengukur dan menilai pencapaian suatu tujuan yang telah ditetapkan. Tes hasil belajar adalah tes yang diberikan setelah materi pelajaran diberikan kepada siswa berupa tes lisan, tes tertulis atau tes perbuatan untuk mengetahui keberhasilan pencapaian tujuan pengajaran.

Norman E. Gronlund (2002:106) mengemukakan bahwa ada enam prinsip yang harus diperhatikan dalam menyusun tes hasil belajar yaitu (1) tes hasil belajar hendaknya mengukur hasil belajar yang telah ditentukan dengan jelas dan sesuai dengan tujuan pengajaran, (2) tes hasil belajar hendaknya mengukur sampel yang representatif dari hasil belajar dan bahan-bahan yang tercakup dalam pengajaran, (3) tes hasil belajar hendaknya mencakup jenis-jenis pertanyaan yang paling sesuai untuk mengukur hasil belajar yang diinginkan, (4) tes hasil belajar hendaknya direncanakan agar hasilnya sesuai dengan hasil atau tujuan yang diinginkan untuk keperluan tertentu, (5) tes hasil belajar hendaknya dibuat dengan reliabilitas yang tinggi dan kemudian harus ditafsirkan dengan hati-hati, (6) tes hasil belajar hendaknya dipakai untuk memperbaiki hasil belajar.

Menurut Slamento (2008:19) syarat-syarat tes yang baik adalah (1) sah (valid) artinya tes harus dapat mengukur yang seharusnya diukur, (2) andal (reliabel) artinya jika digunakan kepada sekelompok siswa yang sama/equivalen beberapa kali dalam waktu berbeda, hasilnya akan sama, (3) obyektif artinya tidak dicampuri pengaruh subyektif dari pihak penilai, (4)

seimbang dalam bahan pelajaran, tingkat kesukaran dan tujuan pengajaran, (5) membedakan artinya hasil tes hendaknya dapat membedakan prestasi belajar tiap siswa yang berhasil, kurang berhasil dan tidak berhasil, (6) norma artinya hasil tes mudah diukur menurut norma atau aturan tertentu, (7) jujur artinya tes mengemukakan masalah yang wajar bukan jebakan, (8) praktis artinya ditinjau dari pembiayaan efisien dan mudah dilaksanakan.

Bloom mengemukakan dalam Suharsimi Arikunto (2009:7), ada tiga ranah/domain yang disebut taksonomi tujuan pengajaran yang harus diperhatikan dalam pembuatan tes yaitu ranah kognitif (pengetahuan:*knowledge*, pemahaman:*comprehension*, penerapan:*application*, penugasan:*analysis*, penyimpulan:*synthesis*, penilaian:*evaluation*), ranah afektif (pendapat:*opinion*, sikap:*attitude*, nilai:*value*) dan ranah psikomotor (berhubungan dengan keterampilan dan gerakan tubuh).

Kualitas Tes

Menurut Wayan Nurkencana (2002:141), tes berkualitas baik dapat dianalisis dari empat kriteria adalah (1) validitas artinya tes tersebut dapat mengukur apa yang hendak diukur, (2) reliabilitas artinya tes tersebut dapat menunjukkan hasil yang konsisten (tetap) walaupun sudah berkali-kali tes tersebut dilakukan, (3) tingkat daya pembeda artinya tes tersebut dapat membedakan siswa menurut tingkat kemampuannya, (4) tingkat kesukaran artinya memiliki keseimbangan antara butir soal sukar, sedang dan mudah.

Sedangkan menurut Suharsimi arikunto (2009:56), persyaratan sebuah tes dikatakan baik jika memiliki (1) validitas, menunjukkan ketepatan, (2) reliabilitas, menunjukkan ketetapan, (3) obyektivitas, menunjukkan tidak

dipengaruhi oleh subyektifitas, (4) kepraktisan, artinya mudah dilaksanakan, tidak menuntut peralatan yang sulit dan mudah mengoreksinya, (5) ekonomis menunjukkan efisiensi artinya tidak membutuhkan waktu yang lama, biaya yang mahal dan tenaga yang banyak.

Sumadi Suryabrata (2007:32) mengidentifikasi enam persyaratan tes yang baik yaitu (1) tes harus reliabel, (2) tes harus valid, (3) tes harus obyektif, (4) tes harus diskriminatif, (5) tes harus komprehensif, (6) tes harus mudah digunakan.

Dalam persyaratan tes yang baik, tes harus reliabel artinya tes tersebut tetap hasilnya walaupun sudah berkali-kali dilakukan. Tes harus valid artinya tes tersebut tepat mengukur kemampuan yang hendak diukur. Tes harus obyektif artinya menilai sesuai dengan kemampuan siswa, tes harus diskriminatif artinya hasil tes tersebut dapat membedakan siswa yang mempunyai kemampuan rendah, sedang dan tinggi. Tes harus komprehensif artinya soal tes mencakup keseluruhan materi yang telah diajarkan. Tes harus mudah digunakan artinya tes tersebut dapat dilaksanakan dengan mudah sesuai dengan waktu, tenaga, biaya dan sarana/prasarana yang ada di sekolah tersebut.

Bentuk Tes

Menurut Ngalim Purwanto (2001:35), tes yang biasa dilakukan guru terdiri dari tes lisan (*oral test*), tes tertulis (*written test*) dan tes perbuatan (*practice test*). Tes tertulis dibagi menjadi tes obyektif dan tes uraian (*essay*). Tes obyektif terdiri dari tes pilihan ganda, tes benar salah, tes menjodohkan, dan tes isian, sedangkan tes uraian terdiri dari uraian bebas dan uraian terbatas.

Bentuk atau teknik tes menurut Depdiknas (2007:4) yaitu unjuk kerja (*performance*), penugasan (*project*), hasil kerja (*product*), tertulis (*paper & pen*),

portofolio (*portfolio*). Sikap (*attitude*), penilaian diri (*self evaluation*). Bentuk tes pilihan ganda terdapat pada teknik tes tertulis.

Suharsimi Arikunto (2009:23), secara garis besar alat evaluasi digolongkan menjadi dua macam yaitu tes dan non tes.

Tes dapat berupa tes formatif dan sumatif. Tes formatif adalah tes yang dilakukan setiap akhir satu atau beberapa pokok bahasan, seperti ulangan harian, UTS (Ulangan Tengah Semester) dan tugas, tujuannya lebih kepada perbaikan proses pembelajaran. Tes formatif termasuk dalam penilaian proses pembelajaran. Tes sumatif adalah tes yang dilakukan pada akhir semester atau akhir tahun, berupa UAS (Ulangan Akhir Semester), Ujian Nasional (UN) dan Ujian Sekolah (US), tujuannya untuk menentukan kenaikan tingkat/kelas dan kelulusan. Tes formatif termasuk dalam penilaian hasil pembelajaran.

Tes juga dapat berupa *pre test* dan *post tes*. Pre tes adalah tes yang diberikan sebelum diberikan materi pelajaran, bertujuan untuk mengetahui seberapa jauh pengetahuan yang dikuasai siswa tentang materi pelajaran yang akan diberikan dan menarik perhatian dan konsentrasi siswa terhadap materi pelajaran yang akan diberikan. Sedangkan *post test* adalah tes yang diberikan kepada siswa setelah siswa diberikan materi pelajaran, bertujuan untuk mengetahui pencapaian keberhasilan terhadap materi yang telah diberikan.

Bentuk Soal Pilihan Ganda

Menurut Depdiknas (2007:31) bentuk soal pilihan ganda terdiri dari dua bagian yaitu pokok soal (*stem*) yang berisi permasalahan yang akan ditanyakan dan sejumlah pilihan atau kemungkinan jawaban (*option*). Dari sejumlah pilihan jawaban yang disediakan hanya ada satu jawaban yang

paling tepat, sedangkan kemungkinan jawaban lain disebut pengecoh (*distractor*). Ditinjau dari ragamnya, bentuk soal pilihan ganda terdiri dari: pilihan ganda biasa, pilihan ganda asosiasi, pilihan ganda hubungan sebab akibat.

Saifuddin Azwar mengatakan (2006:75) bahwa bentuk soal pilihan ganda mempunyai keunggulan dan kelemahan. Keunggulannya: (1). Dapat memuat soal yang banyak sehingga mencakup materi dan tujuan yang luas. (2). Mudah dan cepat dalam pengoreksian jawaban dan pemberian skor. (3). Dengan menggunakan lembar jawaban lebih efisien. (4). Obyektivitas tinggi. Sedangkan kelemahannya adalah (1). Dalam membuat soal sulit, makan banyak waktu, pikiran dan tenaga. (2). Hanya dapat mengungkapkan kompetensi untuk tujuan kognitif saja. (3). Kemungkinan jawaban benar hanya tebakan semata.

Bentuk Soal Pilihan Ganda Biasa

Ign Masidjo (2005:48) mengemukakan, pada bentuk soal pilihan ganda biasa, siswa dihadapkan pada pernyataan atau pertanyaan yang berisi permasalahan dan sejumlah alternatif jawaban, siswa harus memilih satu jawaban yang paling benar atau yang paling tepat.

Juga dikatakan oleh Don F. Blood dan William C. Budd (2002:81) bahwa soal pilihan ganda biasa terdiri dari pernyataan yang menyatakan sebuah masalah dan biasanya ada tiga sampai lima alternatif jawaban, salah satunya jawaban benar, yang lainnya disebut pengecoh. Dengan kata lain bentuk pilihan ganda biasa jawaban benarnya tunggal.

Contoh soal bentuk pilihan ganda biasa:

1. Salah satu Negara anggota ASEAN adalah
 - a. Amerika
 - b. Filipina
 - c. Cina
 - d. Australia

Jawabannya adalah b. Filipina

Bentuk Soal Pilihan Ganda Asosiasi (Kompleks)

Depdikbud (2007:46) mengemukakan bahwa bentuk pilihan ganda asosiasi ini hampir sama dengan bentuk pilihan ganda biasa tetapi bentuk pilihan ganda asosiasi mempunyai kombinasi pernyataan, dari kombinasi pernyataan itu hanya ada satu jawaban yang betul.

Sedangkan Suryabrata (2007:48) mengatakan bahwa bentuk pilihan ganda asosiasi disebut juga bentuk pilihan ganda kombinasi karena alternatif jawaban betulnya kemungkinan tunggal dan kemungkinan merupakan gabungan dari beberapa alternatif jawaban betul. Contoh bentuk soal pilihan ganda asosiasi:

Petunjuk: untuk butir-butir soal berikut pilihlah:

- a. Jika hanya (1), (2), dan (3) betul
- b. Jika hanya (1) dan (3) betul
- c. Jika hanya (2) dan (4) betul
- d. Jika hanya (4) betul
- e. Jika semuanya betul

Contoh soal:

Peraturan dalam koperasi berupa...

1. Al
2. AD
3. AU
4. ART

Jawaban: c

Perbedaan antara bentuk soal pilihan ganda biasa dan bentuk soal pilihan ganda asosiasi hanya pada alternatif jawabannya. Pada bentuk soal pilihan ganda biasa, tiap alternatif jawaban berbentuk jawaban tunggal. Sedangkan pada bentuk soal pilihan ganda asosiasi, alternatif jawabannya dapat berupa gabungan jawaban betul.

Menurut Ign Masidjo (2005:49), untuk pilihan ganda kompleks (asosiasi) cara menjawabnya lebih kompleks. Terdapat perbedaan antara bentuk pilihan

ganda biasa dan pilihan ganda asosiasi dalam cara menjawab.

Dalam menjawab soal bentuk pilihan ganda asosiasi lebih sulit dari pada menjawab soal bentuk pilihan ganda biasa karena dalam menjawab soal bentuk pilihan ganda biasa, peserta tes hanya menempuh satu tahap yaitu memilih satu jawaban dari beberapa alternatif jawaban yang disediakan. Sedangkan untuk menjawab soal bentuk pilihan ganda asosiasi peserta tes menempuh tiga tahap. Tahap pertama, mengidentifikasi satu atau lebih jawaban betul di antara jawaban (1), (2), (3), dan (4). Tahap kedua, menghubungkan jawaban yang telah diidentifikasi dengan petunjuk pengerjaannya. Tahap ketiga, memilih alternatif jawaban dalam bentuk huruf.

Tingkat Kesukaran Soal (*Difficulty Level*)

Saifudin Azwar (2006:129) mengatakan bahwa tingkat kesukaran butir soal adalah proporsi antara banyaknya peserta tes yang menjawab butir soal dengan benar dengan banyaknya peserta tes. Hal ini berarti makin banyak peserta tes yang menjawab butir soal dengan benar maka makin besar indeks tingkat kesukaran, yang berarti makin mudah butir soal itu. Sebaliknya makin sedikit peserta tes yang menjawab butir soal dengan benar maka soal tersebut makin sukar.

Sedangkan menurut HJX. Fernades (2004:60), tingkat kesukaran butir soal diukur oleh persentase siswa yang menjawab soal dengan benar. Jika soal mudah maka indeks kesukaran lebih tinggi. Soal dengan nilai p yang mendekati 0 adalah soal yang sangat sulit, sedangkan soal dengan nilai p mendekati 1 adalah soal yang sangat mudah. Indeks tingkat kesukaran yang sangat baik adalah 0,3 sampai 0,7.

Bistok Sirait (2009:301) berpendapat bahwa tinggi rendahnya

tingkat kesukaran suatu butir soal dapat disebabkan oleh kerumitan (kompleksitas) pokok soal dan kondisi pilihan jawaban yang disediakan. Dalam arti keluasan pertanyaan soal sering membingungkan peserta tes dan alternatif jawaban yang homogen atau kalimat soal sulit dipahami atau mempunyai pengertian ganda bagi peserta tes.

Menurut Anastasi dan Susan Urbina (2007:128), tingkat kesukaran soal berkaitan dengan persentase peserta yang menjawab soal dengan benar. Semakin mudah butir soal, makin besarlah persentasenya. Jika tingkat kesukaran 70 % ($p=0,70$), soal tersebut dianggap lebih mudah dibandingkan jika tingkat kesukaran soalnya 15 % ($p=0,15$). yang paling baik adalah soal yang mempunyai tingkat kesukaran 0,50.

Wayan Nurkencana (2002:161) mengemukakan tentang klasifikasi tingkat kesukaran soal yaitu jika nilai p (prosentase tingkat kesukaran soal) = 0,81 – 1,00 butir soal mudah sekali, jika nilai $p= 0,61- 0,80$ butir soal mudah, jika nilai $p= 0,41 – 0,60$ butir soal sedang, jika nilai $p= 0,21 – 0,40$ butir soal sukar, jika nilai $p= 0,00 – 0,20$ butir soal sukar sekali.

Nana Sujana (2002:135) mengatakan bahwa terdapat tingkat kesukaran mudah, sedang dan sukar. Tingkat kesukaran yang baik adalah 0,25 sampai 0,75. Soal yang mempunyai tingkat kesukaran di bawah 0,25 berarti soal itu sukar, sebaliknya soal yang mempunyai tingkat kesukaran di atas 0,75 adalah soal itu terlalu mudah.

Berdasarkan teori-teori yang telah dikemukakan, tingkat kesulitan soal adalah seberapa mudah dan seberapa sulitnya suatu soal bagi siswa. Tingkat kesukaran dinyatakan dengan persentase siswa yang menjawab soal dengan benar. Makin besar persentase siswa yang menjawab soal dengan benar, makin mudah soal itu. Sebaliknya makin kecil persentase siswa yang menjawab soal

dengan benar, makin sukar soal itu. Dengan demikian soal yang dijawab benar oleh 85 % siswa, dinyatakan mempunyai tingkat kesukaran 0,85, tentunya soal ini lebih mudah dari soal yang mempunyai tingkat kesukaran 20 % (0,20).

Daya Pembeda Soal (*Discrimination Power*)

Dali S. Naga (2002:67) mengatakan bahwa daya pembeda soal adalah kemampuan soal dengan skornya dapat membedakan peserta tes dari kelompok tinggi dan kelompok rendah. Dengan kata lain makin tinggi daya pembeda soal makin banyak peserta dari kelompok tinggi yang dapat menjawab soal dengan benar dan makin sedikit peserta tes dari kelompok rendah yang dapat menjawab soal dengan benar. Agar dapat diterima maka nilai D (discrimination: daya pembeda soal) adalah 0,30 atau lebih. Sedangkan untuk dapat dinyatakan cukup memuaskan adalah 0,40 ke atas.

Menurut Charles D. Hopkins dan Richard L. Antes (1999:159), indeks daya pembeda soal mengukur bagaimana baiknya sebuah soal membedakan tingkat kemampuan siswa. Indeks daya pembeda soal bernilai $-1,00$ sampai $+1,00$. Makin tinggi nilai D sebuah soal, makin baik soal tersebut membedakan siswa yang berkemampuan tinggi dari siswa yang berkemampuan rendah. Soal yang mempunyai nilai $D = 0,40$ dianggap sangat efektif dalam membedakan siswa yang berkemampuan tinggi dan berkemampuan rendah. Soal yang mempunyai nilai D antara 0,20 sampai 0,39 dianggap memuaskan. Soal yang mempunyai nilai D lebih rendah dari 0,39 harus diperbaiki.

Sedangkan fernandes (2004:27) mengatakan bahwa daya pembeda soal adalah kemampuan memisahkan siswa pandai dan siswa kurang. Sedangkan indeks daya pembeda soal adalah

perbedaan persentase dari 27 % siswa yang mendapat nilai tinggi (kelompok atas) dan 27 % siswa yang mendapat nilai rendah (kelompok bawah). Soal yang mempunyai indeks daya pembeda antara 0,15 sampai 0,20 atau lebih tinggi menunjukkan daya pembeda yang baik.

Berdasarkan pendapat-pendapat yang telah dikemukakan di atas, daya pembeda soal adalah kemampuan soal untuk membedakan kelompok peserta tes berkemampuan tinggi dan kelompok peserta tes yang berkemampuan rendah. Nilai daya pembeda dinyatakan melalui indeks daya pembeda. Makin tinggi atau makin besar indeks daya pembeda soal, makin besar soal tersebut dapat membedakan antara kelompok tinggi dan kelompok rendah. Untuk menghitung daya pembeda soal mengambil dari 27 % kelompok tinggi dan 27 % dari kelompok rendah. Nilai indeks daya pembeda soal berkisar antara -1 sampai $+1$. Jika siswa kelompok tinggi dapat menjawab soal dengan benar lebih besar dari siswa kelompok rendah maka indeks daya pembeda positif. Jika siswa kelompok tinggi dan rendah sama-sama dapat menjawab soal dengan benar maka soal tersebut mempunyai daya pembeda nol. Jika siswa kelompok rendah dapat menjawab soal dengan benar lebih besar dari kelompok tinggi maka soal tersebut mempunyai daya pembeda negatif. Soal yang mempunyai indeks daya pembeda negatif dan nol, dibuang karena soal tersebut tidak dapat membedakan siswa yang mempunyai kemampuan tinggi dan rendah.

Reliabilitas Tes

Nama lain dari reliabilitas adalah ketetapan, kepercayaan, kemantapan, keterandalan, keajegan atau kestabilan. Sedangkan yang dimaksud dengan reliabilitas tes adalah sebagai berikut: menurut Anne Anastasi dan Susan Urbina (2007:63), reliabilitas tes adalah kestabilan skor yang diperoleh dari orang

yang sama ketika diuji ulang dengan tes yang sama pada situasi yang berbeda. Suatu tes dikatakan reliabel jika hasil tes sama pada seseorang walaupun telah diadakan pengtesan ulang pada waktu yang berbeda.

T. Raka Joni (2006:36) mengatakan bahwa suatu tes dikatakan reliabel sebuah tes adalah ketetapan hasil tes. Sebuah reliabilitas tes memberi nilai yang sama kepada individu yang sama pada setiap hasil tes.

Sedangkan Syaifuddin Azwar (2006:180) mengatakan bahwa istilah reliabilitas sering disamakan dengan *consistency*, *stability*, yang pada prinsipnya menunjukkan sejauh mana pengukuran itu dapat memberikan hasil yang relatif tidak berbeda jika dilakukan pengukuran kembali terhadap subyek yang sama.

Crocker dan Algina (1996:6) menyebutkan bahwa tinggi rendahnya koefisien reliabilitas dipengaruhi oleh beberapa faktor antara lain: (1) panjang tes, (2) kecepatan (batas waktu) mengerjakan tes, (3) homogenitas kelompok, (4) kesukaran butir soal.

Menurut Dali S. Naga (2008:50), cara menghitung reliabilitas yaitu: (1) reliabilitas stabilitas (uji-ulang uji), (2) reliabilitas ekuivalensi (uji-uji setara), (3) reliabilitas konsistensi internal (pilah paro: Spearman Brown, politomi: 1,2,3,4,5: Alpha Cronbach, dan dikotomi: 0-1: Kuder-Richardson).

Berdasarkan teori-teori yang telah dikemukakan di atas, dapat disimpulkan bahwa reliabilitas suatu tes adalah keajegan suatu tes. Tes dikatakan reliabel apabila menunjukkan hasil yang sama walaupun dilakukan tes berulang ulang. Reliabilitas dinyatakan dengan angka dan dikenal sebagai koefisiensi reliabilitas. Makin tinggi koefisien reliabilitas sebuah tes makin tinggi pula reliabilitas tes tersebut.

Kajian Penelitian Yang Relevan

Hasil penelitian yang dilakukan oleh Yahya Umar (2010:80), menunjukkan bahwa soal bentuk pilihan ganda (biasa dan asosiasi) cukup efektif untuk mengungkapkan pengetahuan tentang proses pada bidang sosiologi dan efektif mengungkapkan kemampuan analisis dan penilaian hubungan kausal pada mata pelajaran Fisika.

Hasil penelitian yang dilakukan oleh Sudjarwadi (2009:101), menunjukkan bahwa soal bentuk pilihan ganda kompleks (asosiasi) memberikan tingkat kesukaran, daya pembeda dan reliabilitas yang tinggi dibandingkan dengan soal pilihan ganda biasa pada mata pelajaran Kimia.

Hasil penelitian yang dilakukan oleh Maria Ulfa (2008:130), menunjukkan bahwa soal bentuk pilihan ganda asosiasi mempunyai tingkat kesulitan yang lebih tinggi dari pada soal pilihan ganda biasa dan soal benar salah pada mata pelajaran Agama.

Hasil penelitian yang dilakukan oleh A.D. Hehahia (2007:90), menunjukkan bahwa terdapat pengaruh banyaknya alternatif soal bentuk pilihan ganda (biasa dan asosiasi) terhadap reliabilitas, tingkat kesukaran, daya pembeda butir soal.

METODOLOGI PENELITIAN

Variabel bebas dalam penelitian ini adalah bentuk soal tes yang terdiri dari: 1) bentuk soal pilihan ganda biasa, 2) bentuk soal pilihan ganda asosiasi. Sedangkan variabel terikatnya adalah: 1) tingkat kesukaran butir soal, 2) daya pembeda butir soal, 3) reliabilitas tes.

Penelitian dilaksanakan di SMPN 15 Bekasi pada mata pelajaran ekonomi. Populasi penelitian adalah siswa kelas II sebanyak 320 orang. Untuk mengambil sampel penelitian menggunakan teknik *Stratified Proporsional Random Sampilng* sehingga didapat sampel penelitian sebanyak 112 orang.

Pengumpulan data dilakukan dengan menggunakan instrumen tertulis yang berjumlah 36 butir berupa tes bentuk pilihan ganda biasa dan pilihan ganda asosiasi. Rancangan penelitian yang digunakan adalah rancangan dua kelompok subyek diacak dengan pemberian tes akhir. Penelitian ini termasuk jenis penelitian kuasi eksperimen. Prosedur pemberian tes dilakukan dengan cara: setiap responden dikenai dua kali perlakuan yaitu diberikan tes dengan butir soal bentuk pilihan ganda biasa dan pilihan ganda asosiasi. Antara tes pertama dan kedua dilaksanakan dalam selang waktu tiga minggu. Setiap kali pelaksanaan tes, kedua tes tersebut diujikan bersamaan kepada masing-masing separuh responden dengan butir soal bentuk pilihan ganda biasa yang diberi nama tes tipe B dan tes dengan butir soal bentukpilihan ganda asosiasi yang diberi nama tes tipe A. Responden yang berdampingan mendapat tipe tes yang berbeda. Responden yang pada tes pertama memperoleh tes tipe A, pada tes kedua memperoleh tes tipe B. Responden yang pada tes pertama memperoleh tes tipe B, pada tes kedua memperoleh tes tipe A.

Langkah-langkah dalam menganalisis data adalah:

Langkah pertama adalah menganalisis butir soal kedua tes

berdasarkan data lembaran jawaban responden untuk memperoleh data tingkat kesukaran, daya pembeda dan reliabilitas tes masing-masing kedua tes. Terlebih dahulu dihitung rata-rata masing-masing variabel tersebut sehingga diperoleh rata-rataTK_B, rata-rataTK_A, rata-rata DP_B, rata-rata DP_A, rata-rata r_B, rata-rata r_A. Untuk menghitung tingkat kesukaran, daya pembeda butir soal dan reliabilitas tes menggunakan rumus tingkat kesukaran, rumus daya pembeda butir soal dan rumus reliabilitas tes, kemudian hasil perhitungan tingkat kesukaran butir soal dihubungkan dengan kriteria tingkat kesukaran butir soal yaitu: (0,000 - 0,244) = sukar; (0,245 - 0,744) = sedang; dan rentang (0,745 - 1,000) = mudah.

Hasil perhitungan daya pembeda butir soal dihubungkan dengan kriteria daya pembeda butir soal yaitu: (0,395 - 1,000) = amat baik; (0,295 - 0,394) = baik, 0,195 - 0,294 = cukup, -0,00 - 0,194 = jelek.

Langkah kedua adalah pengujian persyaratan analisis, meliputi uji normalitas menggunakan uji Liliefors dan uji homogenitas variansi menggunakan uji F.

Langkah ketiga adalah pengujian hipotesis menggunakan uji t.

Tabel 1. Rancangan Penelitian Kuasi Eksperimen

Perlakuan Kateristik yg dibandingkan	Tes bentuk soal pilihan ganda biasa (PGB)	Tes bentuk soal pilihan ganda asosiasi (PGA)
Tingkat kesulitan butir soal (TK)	TKB	TKA
Daya pembeda butir soal (DP)	DPB	DPA
Reliabilitas tes (r)	r _B	r _A

HASIL PENELITIAN DAN PEMBAHASAN**a. Hasil penelitian****1. Deskripsi data tingkat kesukaran butir soal**

Deskripsi data tingkat kesukaran soal disajikan dalam rangkuman tabel 2:

Tabel 2. Rangkuman Data Tingkat Kesukaran Butir Soal

	Tes bentuk PGB	Tes bentuk PGA
Banyaknya butir soal	36	36
Rata-rata	0,6281	0,5209
Simpangan baku	0,1604	0,1591
Median	0,6445	0,5600
Tingkat kesukaran ter-Kecil	0,327	0,244
Tingkat kesukaran ter-Besar	0,848	0,789
Butir soal mudah	4 = 11,11 %	2 = 5,55 %
Butir soal sedang	31 = 86,11 %	29 = 80,56 %
Butir soal sukar	1 = 2,78 %	5 = 13,89 %

2. Deskripsi data daya pembeda butir soal.

Deskripsi data daya pembeda soal disajikan dalam rangkuman berbentuk tabel sebagai berikut:

Tabel 3. Rangkuman Data Daya Pembeda Butir Soal

	Tes bentuk PGB	Tes bentuk PGA
Banyaknya butir soal	36	36
Rata-rata	0,2438	0,3233
Simpangan baku	0,0245	0,0318
Median	0,245	0,3215
Daya pembeda ter-Kecil	0,160	0,265
Daya pembeda ter-Besar	0,276	0,411
Banyaknya butir soal dengan daya pembeda:		
Amat baik	0 = 0 %	3 = 8,33 %
Baik	0 = 0 %	25 = 69,44 %
Cukup	35 = 97,22 %	8 = 22,23 %
Jelek	1 = 2,78 %	0 = 0 %

3. Deskripsi reliabilitas tes soal

Deskripsi data reliabilitas tes disajikan dalam rangkuman berbentuk tabel sebagai berikut:

Tabel 4: Rangkuman Data Reliabilitas Tes

No.	Reliabilitas tes PGB (r_B)	No.	Reliabilitas tes PGA (r_A)
1	0,775	1	0,702
2	0,690	2	0,520
3	0,608	3	0,656
4	0,454	4	0,777
5	0,700	5	0,590
6	0,818	6	0,760
7	0,596	7	0,786
	Rata-rata $r_B=0,6630$		Rata-rata $r_A=0,6844$

4. Pengujian Persyaratan Analisis

a. Pengujian Normalitas

Pengujian normalitas data tingkat kesukaran, daya pembeda butir soal dan reliabilitas tes dilakukan dengan menggunakan uji-L. Ringkasan hasil uji normalitas pada tabel sebagai berikut:

Tabel 5. Hasil Uji Normalitas Data Tingkat Kesukaran Butir Soal

Tes dengan bentuk soal	N	L _{hitung}	L _{tabel $\alpha=0,05$}	Kesimpulan	
				H ₀	Distribusi
PGB	36	0,1353	0,1477	Diterima	Normal
PGA	36	0,1118	0,1477	Diterima	Normal

Keterangan: PGB=Pilihan Ganda Biasa. PGA=Pilihan Ganda Asosiasi

Tabel 6. Hasil Uji Normalitas Data Daya Pembeda Butir Soal

Tes dengan bentuk soal	N	L _{hitung}	L _{tabel $\alpha=0,05$}	Kesimpulan	
				H ₀	Distribusi
PGB	36	0,1366	0,1477	Diterima	Normal
PGA	36	0,0779	0,1477	Diterima	Normal

Keterangan: PGB=Pilihan Ganda Biasa. PGA=Pilihan Ganda Asosiasi

Tabel 7. Hasil Uji Normalitas Data Reliabilitas Tes

Tes dengan bentuk soal	N	L _{hitung}	L _{tabel $\alpha=0,05$}	Kesimpulan	
				H ₀	Distribusi
PGB	7	0,1165	0,300	Diterima	Normal
PGA	7	0,2358	0,300	Diterima	Normal

Keterangan: PGB=Pilihan Ganda Biasa. PGA=Pilihan Ganda Asosiasi

5. Pengujian Homogenitas

Pengujian homogenitas data tingkat kesukaran, daya pembeda butir soal dan reliabilitas tes dilakukan dengan menggunakan uji-F. Ringkasan hasil uji homogenitas pada tabel sebagai berikut:

Tabel 8. Hasil Uji Homogenitas

Variabel	N	F _{hitung}	F _{tabel $\alpha=0,05$}	Kesimpulan	
				H ₀	Distribusi
Tingkat Kesukaran	36	1,0158	1,76	Diterima	Homogen
Daya pembeda	36	1,6667	1,76	Diterima	Homogen
Reliabilitas	7	1,5469	4,21	Diterima	Homogen

6. Pengujian Hipotesis

Pengujian hipotesis data tingkat kesukaran, daya pembeda butir soal dan reliabilitas tes dilakukan dengan menggunakan uji-t. Ringkasan hasil uji hipotesis pada tabel sebagai berikut:

Tabel 9. Hasil Uji Hipotesis

Variabel	N	t _{hitung}	t _{tabel} $\alpha=0,05$	Kesimpulan	
				H ₀	H ₁
Tingkat Kesukaran	36	2,8435	1,667	Ditolak	Diterima
Daya pembeda	36	11,866	1,667	Ditolak	Diterima
Reliabilitas	7	0,3133	1,782	Diterima	Ditolak

b. Pembahasan

Dari deskripsi data hasil penelitian menunjukkan bahwa tingkat kesukaran terkecil (soal paling sukar) pada tes bentuk pilihan ganda asosiasi (0,244) dibandingkan dengan tes bentuk pilihan ganda biasa (0,327). Sedangkan tingkat kesukaran terbesar (soal paling mudah) pada tes bentuk pilihan ganda asosiasi (0,789) dibandingkan dengan tes bentuk pilihan ganda biasa (0,848). Pada kriteria tingkat kesukaran soal, makin kecil angka yang diperoleh maka makin sulit soal itu. Hal ini menunjukkan bahwa tes bentuk pilihan ganda asosiasi lebih sukar dibandingkan dengan tes bentuk pilihan ganda biasa.

Dari deskripsi data hasil penelitian daya pembeda butir soal menunjukkan pada tes bentuk pilihan ganda asosiasi, butir soal amat baik: 3 = 8,33 %, baik: 25 = 69,44 %, cukup: 8 = 22,23 % dan jelek: 0 %, dibandingkan pada tes bentuk pilihan ganda biasa, butir soal amat baik: 0 %, baik: 0 %, cukup: 35 = 97,22 %, jelek: 1 = 2,78 %. Hal ini menunjukkan bahwa tes bentuk pilihan ganda asosiasi mempunyai daya pembeda lebih baik daripada tes bentuk pilihan ganda biasa, dengan kata lain tes bentuk pilihan ganda asosiasi lebih mampu membedakan siswa yang berkemampuan tinggi dan siswa berkemampuan rendah.

Dari deskripsi data hasil penelitian reliabilitas tes menunjukkan

bahwa koefisien reliabilitas tes bentuk pilihan ganda biasa sebesar 0,6630 sedangkan koefisien reliabilitas tes bentuk pilihan ganda asosiasi sebesar 0,6844. Hal ini menunjukkan bahwa reliabilitas tes bentuk pilihan ganda asosiasi tidak lebih tinggi dari tes bentuk pilihan ganda biasa atau tidak menunjukkan perbedaan yang berarti, hasilnya hampir sama.

Hasil pengujian normalitas data tingkat kesukaran butir soal menunjukkan bahwa $L_{hitung} = 0,1353$ lebih kecil dari $L_{tabel} = 0,1477$ ($L_{hitung} < L_{tabel} \alpha = 0,05$) pada butir soal tes bentuk pilihan ganda biasa, berarti H_0 diterima menunjukkan bahwa sampel penelitian berasal dari kelompok populasi yang berdistribusi normal.

Hasil pengujian normalitas data tingkat kesukaran butir soal, menunjukkan bahwa $L_{hitung} = 0,1118$ lebih kecil dari $L_{tabel} = 0,1477$ ($L_{hitung} < L_{tabel} \alpha = 0,05$) pada butir soal tes bentuk pilihan ganda asosiasi, berarti H_0 diterima menunjukkan bahwa sampel penelitian berasal dari kelompok populasi yang berdistribusi normal.

Hasil pengujian normalitas data daya pembeda butir soal, menunjukkan bahwa $L_{hitung} = 0,1366$ lebih kecil dari $L_{tabel} = 0,1477$ ($L_{hitung} < L_{tabel} \alpha = 0,05$) pada butir soal tes bentuk pilihan ganda biasa, berarti H_0 diterima menunjukkan

bahwa sampel penelitian berasal dari kelompok populasi yang berdistribusi normal.

Hasil pengujian normalitas data tingkat kesukaran butir soal, menunjukkan bahwa $L_{hitung} = 0,0779$ lebih kecil dari $L_{tabel} = 0,1477$ ($L_{hitung} < L_{tabel} \alpha = 0,05$) pada butir soal tes bentuk pilihan ganda asosiasi, berarti H_0 diterima menunjukkan bahwa sampel penelitian berasal dari kelompok populasi yang berdistribusi normal.

Hasil pengujian normalitas data reliabilitas tes, menunjukkan bahwa $L_{hitung} = 0,1165$ lebih kecil dari $L_{tabel} = 0,3000$ ($L_{hitung} < L_{tabel} \alpha = 0,05$) butir soal tes bentuk pilihan ganda biasa berarti H_0 diterima menunjukkan bahwa sampel penelitian berasal dari kelompok populasi yang berdistribusi normal.

Hasil pengujian normalitas data reliabilitas tes, menunjukkan bahwa $L_{hitung} = 0,2358$ lebih kecil dari $L_{tabel} = 0,3000$ ($L_{hitung} < L_{tabel} \alpha = 0,05$) butir soal tes bentuk pilihan ganda asosiasi, berarti H_0 diterima menunjukkan bahwa sampel penelitian berasal dari kelompok populasi yang berdistribusi normal.

Hasil pengujian homogenitas data tingkat kesukaran butir soal menunjukkan $L_{hitung} = 1,0158$ lebih kecil dari $L_{tabel} = 1,7600$ ($L_{hitung} < L_{tabel} \alpha = 0,05$), berarti H_0 diterima, menunjukkan bahwa sampel penelitian berasal dari kelompok populasi yang homogen.

Hasil pengujian homogenitas data daya pembeda butir soal menunjukkan $L_{hitung} = 1,6667$ lebih kecil dari $L_{tabel} = 1,7600$ ($L_{hitung} < L_{tabel} \alpha = 0,05$), berarti H_0 diterima, menunjukkan bahwa sampel penelitian berasal dari kelompok populasi yang homogen.

Hasil pengujian homogenitas data reliabilitas tes, menunjukkan $L_{hitung} = 1,5469$ lebih kecil dari $L_{tabel} = 4,210$ ($L_{hitung} < L_{tabel} \alpha = 0,05$), berarti H_0

diterima, menunjukkan bahwa sampel penelitian berasal dari kelompok populasi yang homogen.

Hasil pengujian hipotesis pertama menunjukkan bahwa $t_{hitung} = 2,8435$ lebih besar dari $t_{tabel} \alpha 0,05 = 1,667$ ($t_{hitung} > t_{tabel}$) berarti H_0 ditolak. H_1 diterima. Dengan demikian hipotesis pertama diterima yaitu butir soal bentuk pilihan ganda asosiasi mempunyai tingkat kesukaran yang lebih tinggi daripada butir soal pilihan ganda biasa, keduanya dari kisi-kisi soal yang sama.

Hasil pengujian hipotesis kedua menunjukkan bahwa $t_{hitung} = 11,866$ lebih besar dari $t_{tabel} \alpha 0,05 = 1,667$ ($t_{hitung} > t_{tabel}$) berarti H_0 ditolak. H_1 diterima. Dengan demikian hipotesis kedua diterima yaitu butir soal bentuk pilihan ganda asosiasi mempunyai daya pembeda yang lebih tinggi daripada butir soal pilihan ganda biasa. Berarti butir soal pilihan ganda asosiasi lebih mampu membedakan antara peserta tes yang berkemampuan tinggi dan peserta tes yang berkemampuan rendah. keduanya dari kisi-kisi soal yang sama.

Hipotesis penelitian ketiga mengatakan bahwa diduga tes bentuk pilihan ganda asosiasi mempunyai reliabilitas tes yang lebih tinggi daripada tes bentuk pilihan ganda biasa. Dari hasil pengujian hipotesis ketiga menunjukkan bahwa $t_{hitung} = 0,3133$ lebih besar dari $t_{tabel} = \alpha 0,05 = 1,782$ ($t_{hitung} < t_{tabel}$) berarti H_0 diterima. H_1 ditolak. Dengan demikian hipotesis penelitian ketiga ditolak, yaitu reliabilitas tes bentuk pilihan ganda asosiasi tidak lebih tinggi dari tes bentuk pilihan ganda biasa karena reliabilitas tes bentuk pilihan ganda asosiasi (0,6844) tidak mempunyai perbedaan yang berarti, tidak jauh berbeda atau hampir sama dengan tes bentuk pilihan ganda biasa (0,6630). keduanya dari kisi-kisi soal yang sama.

SIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian dan pembahasan maka disimpulkan sebagai berikut:

1. Butir soal tes bentuk pilihan ganda asosiasi mempunyai tingkat kesukaran yang lebih tinggi daripada butir soal bentuk pilihan ganda biasa.
2. Butir soal tes bentuk pilihan ganda asosiasi mempunyai daya pembeda yang lebih tinggi daripada butir soal bentuk pilihan ganda biasa.
3. Reliabilitas tes bentuk pilihan ganda asosiasi tidak mempunyai perbedaan yang berarti, tidak jauh berbeda atau hampir sama dengan reliabilitas tes bentuk pilihan ganda biasa.

Saran

Berdasarkan kesimpulan yang telah dikemukakan maka disarankan:

1. Siswa SMP dan SMA hendaknya sering dilatih untuk mengerjakan soal bentuk pilihan ganda asosiasi agar terbiasa dan tidak asing lagi sehingga dapat dengan mudah mengerjakan soal bentuk pilihan ganda asosiasi pada seleksi masuk perguruan tinggi negeri negeri.
2. Dalam penyelenggaraan tes prestasi belajar mulai dari tes formatif sampai tes sumatif termasuk Ujian Nasional (UN)/Ujian Sekolah (US) hendaknya diikutsertakan soal bentuk pilihan ganda asosiasi bersama soal bentuk lain agar soal tes atau ujian lebih bervariasi.
3. Perlu menambah jumlah soal bentuk pilihan ganda asosiasi yang lebih banyak di bank soal. Untuk memperoleh butir soal yang bermutu hendaknya melalui uji coba yang terus menerus sampai menemukan butir soal yang memenuhi syarat.

DAFTAR PUSTAKA

- Anastasi, Anne dan Susan Urbina. *Tes Psikologi*. Jakarta: Prenhallindo, 2007.
- Arikunto, Suharsimi. *Dasar-dasar Evaluasi Pendidikan*. Jakarta: Bumi Aksara, 2009.
- Azwar, Saifuddin. *Tes Prestasi*. Yogyakarta: Pustaka Pelajar, 2006..
- Azwar, Saifuddin. *Reliabilitas, Validitas, Interpretasi dan Komputasi*. Yogyakarta: Liberty, 2006.
- Depdikbud. *Pedoman Penulisan Soal Hasil Belajar IPS*. Jakarta: Balitbang, Sisjian, 2007.
- Depdiknas. *Pedoman Penulisan Soal Bentuk Pilihan Ganda IPS SMP*. Jakarta: Dinas Pendas, 2004.
- Blood, Don F. dan Budd William. *Educational Measurement and Evaluation*. New York: Harper and Row, 2002.
- Crocker, L dan J. Algina. *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston Inc, 1996.
- Crombach, Lee J. *Essential Of Psychological Testing*. New York: Harper & Row, 2004.
- Fernandes, H.J.X. *Testing and Measurement*. Jakarta: Depdikbud, 2004.
- Gronlund, Norman E. *Contruction Achievement Test*. Englewood Chiffs: Prentise-Hall, 2002.
- Hehahia, A.D. *Pengaruh Jumlah Alternatif Tes Obyektif Bentuk Pilihan Ganda Terhadap Reliabilitas, Tingkat Kesukaran dan Daya Pembeda*. Tesis. Jakarta: PPsUNJ, 2007.
- Hopkins, Charles D dan Antes, Richard L. *Classroom Measurement and Evaluation*. Illionis, F.E. Peacock, 1999.

- Joni , T. Raka. *Pengukuran dan Penilaian Pendidikan*. Surabaya: Karya Anda, 2006.
- Masidjo, Ign. *Penilaian Hasil Belajar Siswa di Sekolah*. Yogyakarta: Kanisius, 2005.
- Naga, Dali. S. *Pengantar Teori Skor pada Pengukuran Pendidikan*. Jakarta: Gunadarma, 2002.
- Nurkencana, Wayan dan Supartana. *Evaluasi Hasil Belajar*. Surabaya: Usaha Nasional, 2002
- Purwanto, Ngalm. *Prinsip-Prinsip dan Teknik Evaluasi Pengajaran*. Bandung: Remaja Rosdakarya. 2001.
- Sirait, Bistok. *Bahan Pengajaran Untuk Mata Kuliah Evaluasi Hasil Belajar Siswa*. Jakarta: Depdikbud, 2009.
- Slameto. *Evaluasi pendidikan*. Jakarta: Bina Aksara, 2008.
- Silverius, Suke. *Evaluasi Hasil belajar dan Umpan Balik*. Jakarta: Gramedia, 2001.
- Sujarwadi, H.S. *Perbedaan Karakteristik Soal dan Reliabilitas Tes pilihan Ganda Biasa dengan pilihan Ganda Kompleks Mata Pelajaran Kimia*. Tesis. Yogyakarta: PPs UGM, 2009.
- Sujana, Nana. *Penilaian Hasil Proses Belajar*. Bandung: Remaja Rosdakarya, 2002.
- Sudijono, Anas. *Pengantar Evaluasi Pendidikan*. Jakarta: Grasindo Persada, 2006.
- Suryabrata, Sumadi. *Pengembangan Tes Hasil Belajar*. Jakarta: Rajawali, 2007.
- Ulfa, Maria. *Perbandingan Validitas dan Tingkat Kesukaran Butir Soal Antara Bentuk Tes Pilihan Ganda Biasa, Pilihan Ganda Asosiasi dan Benar Salah pada Mata Pelajaran Pendidikan Agama Islam*. Tesis. Jakarta: PPs UNJ, 2008.
- Umar, Yahya. *Berbagai Permasalahan Bentuk Soal Uraian dan Pilihan Ganda dalam Ujian. Buletin Pengujian dan Penilaian*. Jakarta: Depdikbud, 2010.