

KUALITAS SOAL UJIAN DEPARTEMEN PENDIDIKAN BAHASA DAERAH FPBS UPI TAHUN AKADEMIK 2016/2017

Haris Santosa Nugraha

Departemen Pendidikan Bahasa Daerah FPBS UPI

Pos-el: harissantosa89@upi.edu

Abstrak

Penelitian ini dilatarbelakangi dari belum disusunnya tolak ukur penyusunan alat evaluasi mata kuliah berdasarkan uji kelayakan soal-soal ujian di Departemen Pendidikan Bahasa Daerah FPBS UPI. Tujuan penelitian ini adalah untuk mendeskripsikan kualitas soal ujian mata kuliah Departemen Pendidikan Bahasa Daerah FPBS UPI tahun akademik 2016/2017 di yang dikaji berdasarkan uji validitas, reliabilitas, tingkat kesukaran, dan daya beda sebagai dasar penentuan kelayakan soal-soal ujian. Penelitian ini didesain menggunakan pendekatan kuantitatif dengan metode deskriptif-analisis. Sumber data diambil secara *purposive* pada soal-soal mata kuliah kategori pembelajaran, kebahasaan, kesastraan, dan kebudayaan yang menggunakan jenis soal pilihan ganda dengan jumlah 160 item soal. Teknik yang digunakan dalam mengumpulkan data yaitu teknik studi dokumentasi. Teknik pengolahan data dilakukan dengan cara kompilasi, tabulasi, skoring, dan interpretasi. Hasil dari penelitian ini adalah, 1) tingkat validitas soal yang diuji sebanyak 59% valid serta 41% sisanya tidak valid; 2) tingkat reliabilitas soal yang diuji berada dalam kategori sedang; 3) tingkat kesukaran soal 58% sedang, 20% mudah, 14% sukar, 7% sangat mudah, dan 2% sisanya sangat sukar; 4) daya pembeda soal 43% cukup baik, 39% buruk, 10% baik, 6% sangat buruk, dan 2% sisanya sangat baik; serta 5) tingkat kelayakan soal ujian yang dianalisis yakni 67% layak, 12% harus direvisi, dan 21% harus diganti.

Kata Kunci: *validitas, reliabilitas, tingkat kesukaran, daya beda, dan kelayakan*

THE QUALITY OF TEST ITEMS OF SUNDANESE LANGUAGE EDUCATION DEPARTMENT OF FPBS UPI IN 2016/2017 ACADEMIC YEAR

Abstract

This research was motivated by the fact that the benchmarking for subjects evaluation device has not been compiled based on test item feasibility in the Sundanese Education Department of FPBS UPI. The purpose of this study was to describe the quality of test items in the Department of Sundanese Language Education FPBS UPI in 2016/2017 academic year, which were examined based on tests of validity, reliability, level of difficulty, and power of the different on the basis for determining the feasibility of exam question items. This study employed a quantitative approach with descriptive-analysis methods. Data sources were chosen purposively on test items of learning, language, literature, and culture subjects' categories and used multiple-choice question type consisted of 160 exam question items. The technique used in collecting the data was document study technique. Data processing techniques were done by compiling, tabulating, scoring, and interpreting. The results of this study were: 1) questions validity is 59% and the remaining 41% is not valid; 2) the questions reliability is mediocre; 3) question level of difficulty are distributed as it follows: 58% is moderate, 20% is easy, 14% is difficult, 7% is very easy, and the remaining 2% is very difficult; 4) the differentiating question 43% is good enough, 39% is poor, 10% is good, 6% is very poor, and the

remaining 2% is very good; and 5) the question items feasibility split in 67% is feasible, 12% must be revised, and 21% must be replaced.

Keywords: validity, reliability, level of difficulty, power of the different, and feasibility

PENDAHULUAN

Evaluasi tes hasil belajar merupakan salah satu cara untuk mengukur ketercapaian suatu proses pembelajaran yang telah dilaksanakan. Dengan adanya tes hasil belajar, pendidik dapat mengetahui kelebihan ataupun kekurangan proses pembelajaran yang telah dilakukannya. Sehingga, hasil evaluasi itu dapat dijadikan sebagai tolak ukur perbaikan terhadap kekurangan-kekurangan yang telah dilakukan. Seperti halnya memperbaiki cara mengajar, memilih metode pembelajaran yang tepat, penggunaan media pembelajaran, dan lain sebagainya.

Tujuan dari tes hasil belajar adalah untuk memperoleh informasi tentang kekurangan sebuah soal sebagai “petunjuk” dalam mengadakan perbaikan (Arikunto, 2013 hal. 222). Informasi sekaitan dengan hasil belajar tersebut dapat dilakukan dengan pengolahan hasil tes dengan baik. Menurut Purwanto (dalam Sridadi, 2002 hal. 27) pengolahan tes hasil belajar yang baik dapat dilakukan dengan cara analisis soal serta menghitung validitas dan reabilitas soal. Hal tersebut dilakukan sebagai upaya untuk mengidentifikasi soal-soal yang digunakan merupakan soal yang baik, kurang baik, dan soal yang buruk.

Menurut Nurgiantoro (2013 hal. 190) soal yang baik adalah soal yang berkualitas, yakni soal yang didukung oleh butir-butir yang baik, efektif, dan dapat dipertanggungjawabkan. Kegiatan uji kualitas soal dilakukan oleh pendidik untuk sebagai sarana menentukan kelayakan soal dalam rangka meningkatkan mutu soal yang telah dibuatnya. Kegiatan ini dilakukan sebagai proses pengumpulan, peringkasan, dan penggunaan informasi dari jawaban siswa sebagai dasar membuat keputusan tentang

setiap penilaian (Nitko, 1996 hal. 308). Di samping itu, tujuan analisis butir soal juga membantu dalam meningkatkan hasil tes melalui revisi atau membuang soal yang tidak efektif, serta mengetahui informasi diagnostik siswa tentang pemahaman terhadap materi yang telah diajarkan (Aiken, 1994: 63). Oleh karena itu, untuk menentukan estimasi kualitas sebuah butir soal perlu dilakukan analisis terhadap kesahihan alat tes, konsistensi dengan keseluruhan tes, taraf kesukaran yang berimbang, dan memiliki daya pembeda yang baik.

Untuk mengukur kesahihan alat tes dapat diuji menggunakan uji validitas soal. Menurut Nurgiantoro (2013 hal. 152) proses pengujian validitas alat tes dilakukan dengan beberapa pertimbangan, *pertama* validitas merujuk pada kelayakan interpretasi yang dibuat berdasarkan skor hasil tes yang berkaitan dengan penggunaan tertentu bukan terhadap instrumennya itu sendiri. *Kedua* validitas adalah masalah kadar (*matter of degree*) maka dihindari pemikiran bahwa hasil tes itu valid dan tidak valid. *Ketiga*, penilaian validitas tes mesti terkait dengan tujuan penggunaan hasil tes. Berdasarkan hal tersebut, maka proses validasi dilakukan dengan cara pengumpulan bukti-bukti yang menunjukkan dasar saintifik penafsiran skor yang direncanakan dari hasil skor tes. Penafsiran hal uji tersebut digunakan sebagai upaya menginterpretasi apakah alat tes tersebut dapat digunakan untuk mengukur apa yang seharusnya diukur atau belum (Sugiyono, 2010 hal. 137).

Untuk mengukur konsistensi soal dengan keseluruhan tes dapat dilakukan dengan menguji reliabilitas butir soal. Uji reliabilitas secara luas berkenaan uji kepercayaan, kejelasan, konsistensi, keandalan, atau kestabilan berdasarkan

kriteria yang telah ditetapkan. Suatu tes dapat dikatakan reliabel jika selalu memberikan hasil yang sama bila diteskan pada kelompok yang sama pada waktu atau kesempatan yang berbeda (Anastasi, 1997 hal. 103). Berdasarkan hal tersebut, maka reliabilitas suatu tes merupakan tingkat atau derajat konsistensi tes yang bersangkutan. Suatu alat tes memiliki reliabilitas yang baik bila alat ukur itu memiliki konsistensi yang handal walaupun dikerjakan oleh siapapun (dalam level yang sama), di manapun dan kapanpun berada.

Taraf kesukaran yang berimbang adalah soal yang sebarannya tidak terlalu mudah juga tidak terlalu sukar. Soal yang terlalu mudah akan mengakibatkan siswa tidak tertantang untuk mempertinggi usaha dalam memecahkannya, sehingga menyebabkan kejenuhan dan tidak berkembangnya nalar siswa. Sebaliknya soal yang terlalu sukar akan menyebabkan siswa putus asa atau hilang semangat untuk mencoba mengerjakan soal karena dianggap diluar jangkauan kemampuannya. Tapi, bukan berarti soal yang terlalu mudah atau sukar tidak boleh digunakan. Hal ini tergantung dari tujuan penggunaannya. Oleh karena itu, proporsi taraf kesukaran soal harus diperhatikan dengan cara menganalisis indeks kesukaran soal 0,0 – 1,0 (Arikunto, 2013 hal. 225).

Berdasarkan daya pembedanya, butir soal yang baik adalah butir soal yang dapat membedakan kemampuan peserta didik yang berkemampuan tinggi dan yang berkemampuan rendah atau yang sudah dan belum menguasai materi (Arikunto, 2013 hal, 226). 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 siswa atau mahasiswa yang telah memahami materi dengan siswa atau mahasiswa yang belum memahami materi.

Berdasarkan penjelasan di atas, pengujian kualitas soal sangat penting untuk dilakukan. Salah satunya dapat dilakukan untuk mengukur kualitas soal-soal ujian yang ada di Departemen Pendidikan Bahasa Daerah (DPBD) FPBS UPI. Berdasarkan hasil observasi pra-penelitian, soal-soal yang diujikan dalam ujian setiap semester di DPBD FPBS UPI secara berkala diperbaharui oleh dosen-dosen pengampu mata kuliah. Tapi, proses pembaharuan soal-soal tersebut dilakukan tanpa melakukan proses pengujian butir soal terlebih dahulu. Pembaharuan soal hanya didasari dengan alasan agar tidak sama persis dengan soal sebelumnya. Padahal mungkin saja soal yang diganti itu memiliki kualitas soal yang baik dan yang tidak diganti mempunyai kualitas sebaliknya. Hal ini terjadi dimungkinkan karena kesibukan ataupun pemahaman sebagian pengampu terhadap pengujian butir soal masih kurang.

Jika masalah ini dibiarkan dan terus berlanjut, dikhawatirkan proses evaluasi yang dilakukan pun tidak akan berjalan dengan optimal. Dampaknya penilaian terhadap mahasiswa pun menjadi kurang objektif, karena alat ukur yang digunakan mungkin saja tidak memiliki kualitas yang baik. Untuk itu, pengujian butir soal perlu dilakukan untuk mengkaji dan menelaah setiap butir soal sebelum soal digunakan. Pengujian soal ini diharapkan dapat mengukur keakuratan soal, tingkat konsistensi soal, tingkat kesukaran soal (sukar, sedang, atau mudah), tingkat daya pembeda soal (sangat buruk, buruk, cukup, dan baik), dan tafsiran kelayakan soal (berkualitas, harus direvisi, atau ditolak untuk digunakan).

Berdasarkan latar belakang di atas, permasalahan yang teridentifikasi dalam penelitian ini adalah belum adanya pengujian secara empiris dan terkontrol terhadap soal-soal ujian di DPBD FPBS UPI. Hal ini tentu harus menjadi perhatian dalam rangka menjaga kualitas

dan keobjektifan proses dan hasil perkuliahan. Agar penelitian lebih terfokus dan terukur, maka soal-soal ujian yang akan dianalisis dalam penelitian ini dibatasi hanya soal UAS yang dikelompokkan ke dalam soal-soal pembelajaran, kebahasaan, kesastraan, dan kebudayaan yang dilaksanakan pada Semester Genap Tahun Akademik 2016/2017 saja.

Adapun manfaat dari penelitian ini adalah untuk memberikan informasi sebagai bahan refleksi terhadap pembuat soal tentang keakuratan, konsistensi, kesukaran, daya pembeda yang pada akhirnya dapat memberikan tafsiran terhadap kualitas soal, apakah layak digunakan atau sebaliknya serta sebagai sumber perencanaan penyusunan soal pembelajaran, bahasa, sastra dan budaya Sunda kedepannya di DPBD FPBSD UPI.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif-analisis. Adapun langkah-langkah metode deskriptif dilakukan dengan cara mengadopsi teori Stephen (2004:18) yakni (a) mengumpulkan informasi faktual secara rinci dan menggambarkan gejala-gejala di lapangan; (b) mengidentifikasi masalah-masalah yang dihadapi; (c) membuat perbandingan-perbandingan; dan (d) menentukan implikasi dari pengalaman dalam merencanakan perbaikan di masa yang akan datang.

Populasi dalam penelitian ini merupakan soal-soal ujian DPBD FPBS UPI Tahun Akademik 2016/2017. Sampelnya adalah soal-soal ujian DPBD FPBS UPI semester genap tahun akademik 2016/2017 yang menggunakan jenis soal pilihan Ganda. Adapun teknik pengambilan sampel dilakukan dengan menggunakan teknik *purposive sampling*. Data yang dijadikan sampel penelitian terdiri dari empat kelompok Mata Kuliah (MK), yakni kelompok MK

Pembelajaran, MK Kebahasaan, MK Kesastraan, dan MK Kebudayaan. Objek dalam penelitian ini berkembang apa adanya, tidak dimanipulasi oleh peneliti dan kehadiran peneliti tidak mempengaruhi dinamika pada objek tersebut. Penelitian ini bersifat *ex post facto*, artinya data dikumpulkan setelah semua peristiwa yang diperhatikan terjadi (Azwar, 2004:9). Hal ini dikarenakan jawaban mahasiswa terhadap soal ujian yang dianalisis merupakan data yang telah dilaksanakan sebelumnya.

Instrumen yang dipakai untuk menjaring data yang digunakan adalah program analisis butir soal yang dibuat dalam program *Microsoft Excel 2013*. Instrumen dikembangkan oleh peneliti untuk mendapatkan gambaran yang dapat menjawab tujuan penelitian ini. Jawaban mahasiswa terhadap soal-soal ujian diharapkan dapat memberikan gambaran tentang kualitas soal yang diujikan. Untuk mendapatkan hasil yang baik, sebelum instrumen penelitian digunakan, terlebih dahulu divalidasi oleh *profesional judgement* dan ahli evaluasi pembelajaran.

Pengumpulan data dilakukan dengan teknik studi dokumentasi terhadap soal-soal UAS Semester Genap Tahun Akademik 2016/2017 menggunakan instrumen penelitian yang telah disusun. Jika data penelitian telah terkumpul, kemudian data tersebut dianalisis. Teknik analisis data yang digunakan adalah teknik analisis statistik deskriptif. Analisis statistik deskriptif digunakan untuk menganalisis data dengan cara mendeskripsikan atau menggambarkan data-data yang telah terkumpul sebagaimana adanya. Data yang terkumpul selanjutnya diolah menggunakan statistik deskriptif dengan cara (1) dikompilasi, (2) ditabulasi, (3) diberi skor dan dipersentase, dan (4) diinterpretasi.

Untuk menganalisis butir soal pada penelitian ini digunakan program analisis soal pada *Microsoft Excel 2013*. Excel

merupakan sebuah program pengolahan data yang biasa dinamakan "spreadsheet". Karena program ini dapat digunakan untuk mengolah data yang berupa angka ataupun lainnya. Ada dua cara mengolah data dengan Ms. Excel, yaitu (1) melalui program bantu khusus perhitungan statistik dan (2) melalui fungsi statistik yang terdapat di dalam Excel. Tetapi, dikarenakan di dalam program ini tidak tersedia fasilitas uji validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan penentuan kelayakan soal maka program ini sengaja dibuat untuk mempermudah cara pengolahan data dengan memanfaatkan fungsi statistik.

HASIL DAN PEMBAHASAN

Pada bagian ini dideskripsikan hasil yang telah dicapai pada penelitian kualitas soal ujian DPBD FPBS UPI Tahun Akademik 2016/2017. Ada empat kelompok soal yang diuji kualitasnya yakni kelompok soal bidang pembelajaran, kebahasaan, kesastraan, dan kebudayaan. Soal diuji secara kuantitatif dengan cara menganalisis validitas, reliabilitas, tingkat kesukaran, daya beda, dan kelayakan soal mata kuliah tahun akademik 2016/2017.

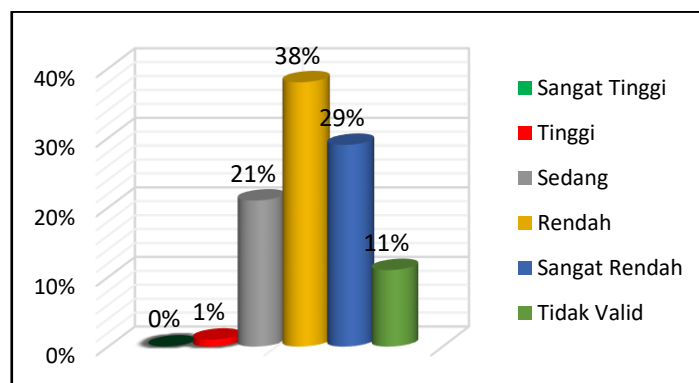
Validitas Soal Ujian

Uji validitas soal dilakukan untuk mengukur tingkat keandalan atau keshahihan alat ukur. Item soal yang

dinyatakan valid merupakan soal yang digunakan untuk mengukur apa yang seharusnya diukur (Sugiyono, 2010:137). Kegunaan validitas pada penelitian ini yaitu untuk mengetahui sejauh mana ketepatan dan kecermatan instrumen pengukuran dalam melakukan fungsi ukurnya agar soal yang diperoleh merupakan soal sesuai dengan tujuan pengukuran. Dukungan setiap butir soal dinyatakan dalam bentuk korelasi sehingga untuk mendapatkan validitas suatu butir soal dilakukan perhitungannya menggunakan program analisis dalam *Ms. Excel 2013*.

Adapun interpretasi validitas soal mengacu kepada pendapat Djiwandono (2011 hal. 168) yang mengurutkan tingkat kadar korelasi berhubungan dengan koefisiensi validitas. Indeks koefisien validitas berkisar antara 0,0 – 1,0. Hasil uji validitas 0,0 - 0,19 termasuk kategori sangat rendah, 0,20 – 0,39 kategori rendah, 0,40 – 0,59 kategori sedang, 0,60 – 0,79 kategori tinggi, dan 0,80 – 1,00 kategori sangat tinggi.

Berdasarkan hasil analisis rekapitulasi uji validitas soal ujian DPBD FPBS UPI yang mencakup kelompok bidang pembelajaran, kebahasaan, kesastraan, dan kebudayaan. Tingkat validitas soal ujian DPBD FPBS UPI 38% rendah, 29% sangat rendah, 21% sedang, 11% tidak valid, 1% tinggi, dan 0% sangat tinggi. Jika disajikan dalam grafik seperti di bawah ini.



Grafik 1 Persentase Kriteria Reliabilitas Soal

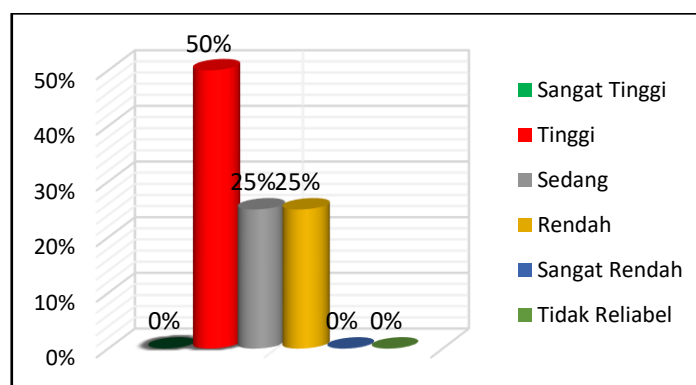
Dari data tersebut, tingkat validitas soal ujian kriteria rendah merupakan tingkat yang paling banyak dibandingkan kriteria lainnya. Kriteria tersebut didapatkan berdasarkan hasil uji korelasi *product moment* pada item-item soal yang di ditafsirkan berdasarkan hasil komparasi nilai t_{itung} dengan t_{tabel} . Semakin besar t_{itung} dibanding t_{tabel} akan berpengaruh pada lebih tingginya tingkat validitas suatu item soal, begitupun sebaliknya. Sedangkan jika data dikelompokkan berdasarkan keterangan data valid dan tidak valid maka 59% soal dinyatakan valid dan 41% sisanya dinyatakan tidak valid (*invalid*). Gambaran persentase rekapitulasi uji validasi soal ujian ini didasarkan pada interpretasi besaran koefisien korelasi setiap item soal menggambarkan item-item soal masih rendah dan belum dapat mengukur apa yang seharusnya diukur. Jadi, berdasarkan hal tersebut menunjukkan bahwa validitas soal ujian DPBD FPBS UPI Semester Genap Tahun Akademik 2016/2017 secara umum masih berada dalam kategori rendah.

Reliabilitas Soal Ujian

Uji reliabilitas dilakukan untuk mengetahui tingkat konsistensi jawaban pada instrumen tes. Analisis ini pada

hakikatnya menguji keajegan soal tes yang di dalamnya berupa seperangkat butir soal apabila diberikan berulang kali pada objek yang sama. Instrumen yang baik secara akurat memiliki jawaban yang konsisten dengan hasil yang relatif sama (Arikunto, 2013: 86). Indeks reliabilitas berkisar antara 0,0 – 1,0. Semakin tinggi koefisien reliabilitas suatu tes atau semakin mendekati 1,0, makin tinggi pula keajegan/ketepatannya. Adapun kriteria tingkat reliabilitas adalah 0,0 - 0,19 kategori sangat rendah, 0,20 – 0,39 kategori rendah, 0,40 – 0,59 kategori sedang, 0,60 – 0,79 kategori tinggi, dan 0,80 – 1,00 kategori sangat tinggi.

Berdasarkan hasil analisis, kelompok soal yang memiliki kategori koefisien tinggi yakni kelompok kebahasaan dengan nilai 0,75 dan kelompok kesastraan dengan nilai 0,67 atau sebanyak 50%. Kelompok soal dengan kategori sedang yakni kelompok kebudayaan dengan nilai 0,52 atau sebanyak 25%. Kelompok soal dengan kategori rendah yakni kelompok pembelajaran dengan nilai koefisien 0,28 atau sebanyak 25%. Sedangkan tidak ada kategori sangat baik, sangat rendah dan tidak reliabel atau sebanyak 0%. Jika disajikan dalam grafik seperti di bawah ini.



Grafik 2 Persentase Kriteria Reliabilitas Soal

Kategori reliabilitas yang ditemukan dalam grafik di atas menunjukkan bahwa keterandalan soal ada yang sudah “baik” tetapi adapula yang

masih “perlu peningkatkan”. Keterandalan yang dimaksud meliputi ketepatan atau kecermatan hasil pengukuran dan keajegan atau kestabilan

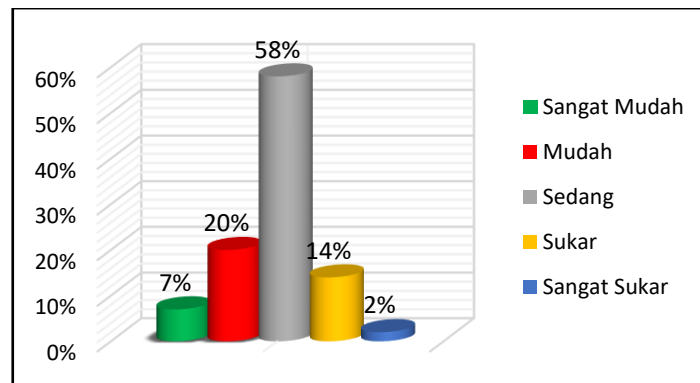
dari hasil pengukuran. Sehingga apabila dilakukan pengujian beberapa kali pada soal tes ini, akan memberikan hasil yang tetap atau relatif sama. Berdasarkan hal tersebut, tentu sangat diperlukan adanya revisi terhadap soal-soal yang digunakan agar keterandalan soal yang digunakan dapat diyakini keajegannya.

Tingkat Kesukaran Soal Ujian

Tingkat kesukaran (TK) merupakan indeks yang menyatakan tingkat kemudahan atau kesulitan suatu butir soal bagi peserta tes. Menurut Arikunto (2013 hal. 225), indeks

kesukaran sering diklasifikasikan menjadi sukar, sedang, dan mudah. Soal yang dinyatakan sukar adalah soal yang tingkat kesukaran (P) 0,00 - 0,30, soal yang sedang tingkat kesukaran (P) 0,31 - 0,70, dan soal yang mudah tingkat kesukaran (P) 0,71 - 1,00.

Berdasarkan hasil analisis, tingkat kesukaran soal ujian DPBD FPBS UPI adalah 58% sedang, 20% mudah, dan 14% sukar, sangat mudah 7%, dan sangat sukar 2%.



Grafik 3 Persentase Tingkat Kesukaran Soal

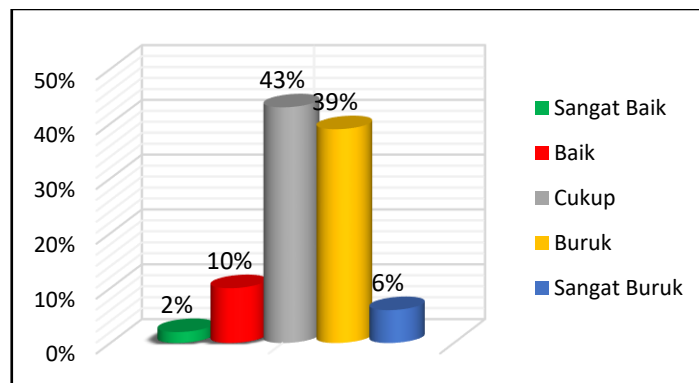
Proporsi sebaran tingkat kesukaran soal ujian dari grafik di atas termasuk baik, mengingat bahwa butir soal yang baik adalah butir yang tidak terlalu sukar dan tidak terlalu mudah. Jika butir soal terlalu sukar sehingga semua peserta tes tidak bisa menjawab, atau butir soal terlalu mudah sehingga semua pembelajar dapat menjawab benar, berarti butir tersebut tidak dapat mencerminkan capaian hasil belajar pembelajar. Artinya, butir-butir soal yang demikian tidak dapat membedakan antara peserta tes kelompok tinggi (peserta tes yang banyak menjawab benar) dan peserta tes kelompok rendah (peserta tes yang banyak menjawab salah).

Daya Beda Soal Ujian

Uji daya pembeda dilakukan untuk menentukan apakah soal tersebut dapat membedakan antara

siswa yang pintar dengan siswa yang kurang pintar (Arikunto, 2013 hal. 226). Angka yang menunjukkan besarnya daya pembeda disebut indeks diskriminatif (D) yang berkisar antara 0,00 - 1,00. Soal yang dinyatakan memiliki daya pembeda sangat buruk adalah soal dengan kategori 0,0 - 0,19, kategori rendah 0,20 - 0,39, kategori sedang 0,40 - 0,59, kategori tinggi 0,60 - 0,79, dan kategori sangat tinggi 0,80 - 1,00.

Berdasarkan hasil analisis daya penelitian, ditemukan lima kategori daya pembeda yang diurutkan berdasarkan yang paling dominan, yakni 43% soal termasuk kategori cukup baik, 39% kategori buruk, 10% kategori baik, 6% kategori sangat buruk, dan 2% kategori sisanya sangat baik. Kategori daya beda ini digambarkan dalam grafik sebagai berikut.



Grafik 4 Persentase Daya Beda Soal

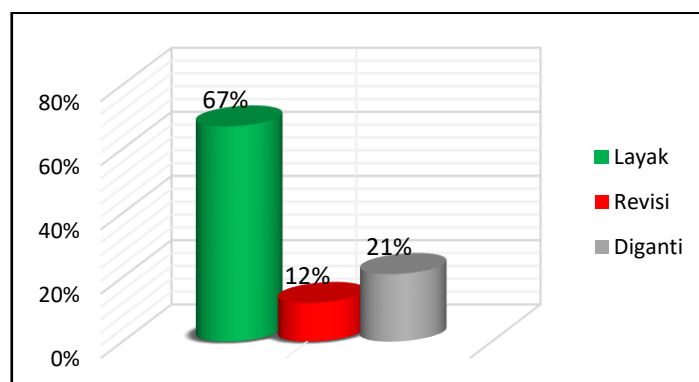
Data di atas menunjukkan bahwa daya pembeda soal ujian sebagian besar didominasi kategori cukup dan baik. Hal ini mengindikasikan bahwa alat tes sudah cukup baik untuk membedakan kemampuan siswa yang berkemampuan tinggi dan berkemampuan rendah.

Kelayakan Soal Ujian

Uji kelayakan soal merupakan tahapan menentukan layak tidaknya soal yang telah dianalisis diterima sebagai soal berkualitas, harus direvisi, atau ditolak dan jangan digunakan. Proses penentuan kelayakan soal digunakan acuan dari uji validitas, uji reliabilitas, tingkat kesukaran, dan daya pembeda. Soal yang dinyatakan layak diterima adalah soal yang kriteria validitasnya valid, kriteria realibilitasnya minimal sedang, kriteria tingkat kesukaran

sedang/sukar, dan kriteria daya pembedanya cukup/baik/sangat baik. Soal yang dinyatakan harus direvisi adalah yang soal kriteria validitasnya valid, kriteria realibilitasnya minimal sedang, kriteria tingkat kesukaran mudah/sedang, dan kriteria daya pembedanya cukup/buruk. Sedangkan soal yang dinyatakan ditolak atau tidak layak digunakan adalah yang soal dengan kriteria validitasnya tidak valid, kriteria realibilitasnya rendah, kriteria tingkat kesukaran mudah/sukar, dan kriteria daya pembedanya buruk/sangat buruk.

Berdasarkan hasil analisis uji kelayakan soal ujian DPBD FPBS UPI Semester Genap 2016/2017 sebanyak 67% soal ujian dinyatakan layak, 12% harus direvisi, dan 21% harus diganti. Jika disajikan dalam grafik seperti di bawah ini.



Grafik 5 Persentase Kelayakan Soal

Berdasarkan komposisi kelompok mata kuliah dalam grafik di atas,

kelompok kebahasaan termasuk kelompok yang soalnya paling banyak

masuk kategori layak digunakan, diikuti soal kebudayaan, kesastraan, dan terahir pembelajaran. Dengan adanya analisis terhadap soal pembelajaran bahasa di atas, setidaknya telah memberikan suatu prosedur sistematis yang memberikan informasi sangat khusus terhadap butir tes yang disusun. Analisis tes ini dilaksanakan sebabagai salah satu kegiatan yang perlu dilakukan dalam rangka meningkatkan mutu instrumen tes, baik mutu keseluruhan tes maupun mutu tiap butir soal yang menjadi bagian dari tes tersebut. Tes sebagai alat evaluasi diharapkan dapat menghasilkan nilai atau skor yang objektif dan akurat. Oleh karena itu perlu diusahakan agar tes yang diberikan kepada mahasiswa sedapat mungkin cukup baik dan bermutu. Tes yang baik tentunya dapat digunakan berulang-ulang dengan sedikit perubahan. Sebaiknya tes yang buruk supaya dibuang atau tidak digunakan untuk memberi nilai kepada mahasiswa. Sebuah tes dapat dikatakan layak sebagai alat pengukur bila memenuhi persyaratan tes yang mencakup valid, reliabel, mempunyai daya pembeda dan tingkat kesukaran soal yang baik.

Adapun soal-soal yang harus direvisi dan diganti karena tidak memenuhi kriteria di atas ternyata mengalami kesalahan dalam faktor materi, konstruksi, dan faktor bahasa. Dalam faktor materi sebagian besar soal ditolak karena soal yang diujikan kurang sesuai dengan indikator dalam tercantum dalam RPS. Selain itu soal ditemukan memiliki pilihan jawaban tidak homogen dan logis serta jawaban benar yang lebih dari satu.

Faktor konstruksi soal-soal ujian yang direvisi atau ditolak karena menyantumkan gambar, grafik, atau tabel yang kurang berfungsi dengan baik sehingga memberikan multitafsir terhadap jawaban yang dipilih. Pokok soalpun ada yang memberi arahan terhadap jawaban benar sehingga hampir seluruh mahasiswa dapat menjawab soal

dengan benar. Selain itu, pokok soal masih ada menggunakan pernyataan yang negatif. Sedangkan dalam faktor bahasa masih terdapat penggunaan ejaan dan penulisan kata/kalimat yang digunakan dalam soal kurang tepat atau salah sehingga memberikan kerancuan terhadap jawaban yang dianggap benar oleh siswa.

SIMPULAN

Berdasarkan hasil dan pembahasan di atas dapat disimpulkan bahwa tingkat validitas soal yang diuji sebanyak atau 59% dinyatakan valid sedangkan 41% sisanya dinyatakan tidak valid. Jumlah soal yang valid lebih besar dibandingkan soal yang tidak valid, hal ini menunjukkan bahwa soal ujian yang diujikan sebagian besar dapat mengukur kompetensi mahasiswa yang sejajar dengan materi atau isi mata kuliah yang diberikan, tapi tidak sedikit juga yang sebaliknya. Tingkat reliabilitas soal yang diuji berada dalam kategori sedang. Hal ini menunjukkan bahwa soal ujian memiliki keterandalan cukup baik. Keterandalan yang dimaksud meliputi ketepatan atau kecermatan hasil pengukuran dan keajegan atau kestabilan dari hasil pengukuran. Sehingga apabila dilakukan pengujian beberapa kali pada soal tes ini, akan memberikan hasil yang tetap atau relatif sama.

Jika dilihat dari tingkat kesukaran soal 58% termasuk kategori sedang, 20% mudah, 14% sukar, 7% sangat mudah, dan 2% sisanya sangat sukar. Hasil pengujian ini menyatakan bahwa soal yang diujikan sudah memiliki proporsi yang cukup baik. Daya pembeda soal ujian 43% dianggap cukup baik, 39% buruk, 10% baik, 6% sangat buruk, dan 2 % sisanya sangat baik. Rata-rata daya pembeda pada soal ini berada pada posisi cukup baik sehingga mampu membedakan jawaban antara siswa yang pandai (berkemampuan tinggi) dengan siswa yang kurang pandai (berkemampuan rendah). Adapun tingkat

kelayakan soal ujian yang dianalisis yakni 67% layak, 12% harus direvisi, dan 21% harus diganti. Kriteria yang ditemukan dari analisis ini sebagian besar soal tergolong memenuhi persyaratan soal yang memiliki kualitas baik dari uji validitas, reliabilitas, tingkat kesukaran, dan daya pembedanya.

Berdasarkan simpulan di atas, hasil analisis soal ujian DPBD FPBS UPI menunjukkan hasil yang cukup baik. Namun masih diperlukan perbaikan soal-soal ujian pada kelompok pembelajaran, kebahasaan, kesastraan, maupun kebudayaan dalam pengujian berikutnya. Oleh karena itu, masih diperlukan revisi atau perbaikan soal-soal ujian baik kelompok pembelajaran, kebahasaan, kesastraan, maupun kebudayaan dalam pengujian berikutnya. Diperlukan kemampuan, kejelian, serta tolok ukur hasil analisis soal yang baik untuk meningkatkan kualitas soal ujian yang akurat dan dapat mengukur kompetensi pembelajaran yang telah dicapai mahasiswa. Soal-soal yang perlu diperbaiki dipilih dari materi yang telah disampaikan serta disajikan dalam soal yang memiliki kualitas tinggi. Untuk mewujudkan hal tersebut diperlukan kemampuan, kejelian, serta pengalaman yang baik dari dosen-dosen pengampu untuk meningkatkan kualitas kelayakan soal-soal ujian agar hasilnya akan lebih akurat dalam mengukur kompetensi pembelajaran yang telah dicapai mahasiswa.

PUSTAKA RUJUKAN

- Aiken, Lewis R. (1994). *Psychological Testing and Assessment, (Eight Edition)*, Boston: Allyn and Bacon.
- Anastasi, Anne and Urbina, Susana. (1997). *Psicoholological Testing (Seventh Edition)*. New Jersey: Prentice-Hall, Inc.
- Arikunto, Suharsimi. (2013.) *Dasar-dasar Evaluasi Pendidikan*. Jakarta: PT. Bumi Aksara.
- Azwar, Saifuddin. 2004. *Metode Penelitian*. Yogyakarta: Pustaka Pelajar.
- Djiwandono, Soenardi. (2011). *Tes Bahasa: Pegangan bagi Pengajar Bahasa*. Jakarta: PT Indeks.
- Nitko, Anthony J. (1996). *Educational Assessment of Students, Second Edition*. Ohio: Merrill an imprint of Prentice Hall Englewood Cliffs.
- Nurgiantoro, Burhan. (2013). *Penilaian Pembelajaran Bahasa berbasis Kompetensi*. Yogyakarta: BPFY-Yogyakarta.
- Sridadi, M. P. (2002). *Analisis Butir Soal Pilihan Ganda*. Majalah Olahraga, 8(26), 26-37.
- Stephen P Robbins, 2004. *Perilaku Organisasi, Konsep, Kontroversi dan Aplikasi. Alih Bahasa: Hadyana Pujaatmaka*. Jakarta: Penerbit PT. Bhuana Ilmu Populer.
- Sugiyono. 2010. *Statistika untuk Penelitian*. Bandung: Alfabeta.

UCAPAN TERIMA KASIH

Terima kasih penulis sampaikan kepada semua pihak yang telah membantu penelitian ini, terutama kepada Penyunting Jurnal Lokabasa atas dimuatnya tulisan ini.