

PENGEMBANGAN BANK SOAL MATA KULIAH PSIKOLOGI PERKEMBANGAN

Anna Armeini Rangkuti

Abstract

Valid and reliable test is an essential instrument to measure the student's learning achievement. Standardized test available at the time it is required will make the instructional process more effective and efficient. This study aims at building a test bank with standardized tests for Development Psychology Course as one of Basic Courses for Education that has to be taken by all students of education. The test items were developed by employing the method which could expose psychometric property or psychomotoric characteristics of each item. For this purpose 95 test items were tried out to 224 students of State University of Jakarta. From the whole number of test items, 46 were analyzed by using Rasch Model (Item Response Theory) and the remaining 49 were analyzed with Classical Test Theory. The study resulted in 29 valid test items from Rasch Model and 42 valid test items from Classical Test Theory. The valid test items can be put into the test bank and ready for use.

Key words : test, test bank, Rasch Model, Classical Test Theory.

Abstrak

Tes yang andal dan sah diperlukan sebagai suatu alat untuk mengukur hasil belajar peserta didik. Tersedianya tes baku yang dapat dipergunakan ketika diperlukan sangat membantu proses pembelajaran yang lebih efektif dan efisien. Penelitian ini bertujuan untuk mengembangkan suatu bank tes yang berisikan tes baku untuk matakuliah Psikologi Perkembangan sebagai suatu mata kuliah dasar kependidikan yang harus ditempuh oleh semua mahasiswa pendidikan. Butir tes dikembangkan dengan menggunakan metode yang mengandung ciri-ciri psikometrik atau psikomotorik untuk setiap butir tes. Untuk mencapai tujuan penelitian ini 95 butir tes diujicobakan kepada 224 mahasiswa Universitas Negeri Jakarta. Dari keseluruhan jumlah butir tes, 46 dianalisis dengan menggunakan *Rasch Model (Item Response Theory)*. Oleh karena itu, sangat diperlukan adanya tes baku yang menggunakan *Rasch Model (Item Response Theory)* dan 49 sisanya dianalisis dengan menggunakan teori tes klasik. Penelitian ini menghasilkan 29 butir tes yang valid dari *Rasch Model* dan 42 butir tes yang valid dari teori tes klasik. Butir-butir tes dari yang valid tersebut dapat dimasukkan ke dalam bank tes dan siap dipergunakan.

Kata kunci: tes, bank tes, Rasch Model, teori tes klasik

PENDAHULUAN

Mata kuliah Psikologi Perkembangan merupakan salah satu mata kuliah wajib bagi mahasiswa kependidikan. Mata kuliah Psikologi Perkembangan berada pada rumpun Mata Kuliah Dasar Kependidikan (MKDK) di samping tiga mata kuliah lainnya: Pengantar Ilmu Pendidikan, Teori Belajar Pembelajaran, dan Profesi Kependidikan. Pada garis besarnya, mata kuliah Psikologi Perkembangan memiliki muatan tentang pemahaman perkembangan peserta didik. Mata kuliah Psikologi Perkembangan penting untuk dikuasai oleh mahasiswa kependidikan yang akan berkecimpung di dunia pendidikan nantinya agar

dapat mengenali dan memahami situasi perkembangan psikologis peserta didik.

Di Universitas Negeri Jakarta sebagai salah satu LPTK (Lembaga Pendidikan Tenaga Kependidikan), pengelolaan administratif Mata Kuliah Dasar Kependidikan berada di bawah Fakultas Ilmu Pendidikan. Semua mahasiswa kependidikan di UNJ ketika mengambil mata kuliah MKDK akan berkoordinasi dengan bagian MKDK FIP UNJ. Pemilihan dosen-dosen yang akan mengajar MKDK dilakukan oleh bagian MKDK FIP UNJ. Khusus untuk mata kuliah Psikologi Perkembangan, dosen-dosen pengajarnya diutamakan yang berasal dari bidang keilmuan Psikologi.

Berdasarkan pengamatan yang dilakukan dan

pengalaman yang diperoleh, dosen-dosen diberikan otoritas dalam membawakan mata kuliah ini mulai dari pemilihan referensi sampai pada penulisan soal sebagai alat mengevaluasi hasil pembelajaran. Silabus pembelajaran disediakan oleh Bagian MKDK FIP dan selanjutnya para dosenlah yang mengelola pembelajaran sampai mahasiswa memperoleh nilai akhir. Kondisi ini di satu sisi menguntungkan pihak dosen karena memiliki otoritas dan kebebasan mengelola proses pembelajaran sampai menentukan standar kelulusan mahasiswa. Namun di sisi lain, kualitas atau pencapaian hasil pembelajaran oleh mahasiswa menjadi tidak terstandar. Misalnya: dua orang mahasiswa yang berasal dari prodi kependidikan yang berbeda dan memperoleh nilai A pada mata kuliah Psikologi Perkembangan tidak dapat dijamin kesetaraan kualitas dan kompetensinya pada mata kuliah tersebut. Hal ini karena proses pembelajaran yang berbeda dan terutama adalah soal-soal evaluasi yang disajikan juga berbeda.

Harapan yang seharusnya terpenuhi adalah terjaminnya kesetaraan standar kelulusan pada mata kuliah MKDK, khususnya Psikologi Perkembangan, sehingga antara satu mahasiswa dengan mahasiswa lain yang berasal dari prodi berbeda dan dosen pengajar berbeda dapat dibandingkan satu sama lain. Pada akhirnya, kualitas dan kompetensi mahasiswa kependidikan, khususnya terkait mata kuliah bidang kependidikan dapat menaikkan peluang dalam kompetisi di lapangan pekerjaan. Selain itu, mutu institusi LPTK dapat lebih terjaga di tengah-tengah *stakeholder*.

Salah satu cara yang dapat dilakukan adalah dengan pengembangan bank soal yang sudah teruji properti psikometriknya. Properti psikometrik yang dimaksud seperti tingkat kesukaran soal. Selanjutnya, soal-soal yang ada di dalam bank soal akan dapat digunakan oleh dosen pengajar mata kuliah MKDK khususnya Psikologi Perkembangan. Dengan demikian, diharapkan mahasiswa dari beragam program studi kependidikan yang telah mengikuti proses pembelajaran dapat dijamin kesetaraan kompetensinya terutama dalam mata kuliah Psikologi Perkembangan.

KAJIAN PUSTAKA

Bank soal

Bank soal merupakan kumpulan soal-soal yang telah teruji properti psikometriknya. Properti psikometrik yang dimaksud adalah tingkat kesukaran, daya pembeda, analisis distraktor, dan sebagainya. Proses yang dilakukan untuk mendapatkan properti

psikometrik yang baik dinamakan kegiatan menganalisis butir soal. Secara lebih luas, kegiatan menganalisis butir soal adalah suatu proses pengumpulan, peringkasan, dan penggunaan data jawaban peserta didik untuk membuat keputusan tentang setiap penilaian (Nitko, 1996, dalam Safari, 2004). Tujuan selanjutnya adalah untuk merevisi atau tidak lagi menggunakan soal yang tidak berfungsi secara efektif serta untuk mengetahui informasi diagnostik pada siswa apakah mereka sudah/belum memahami materi yang telah diajarkan (Aiken, 1994). Soal yang baik kualitasnya adalah soal yang dapat memberikan informasi secara tepat tentang penguasaan materi oleh peserta didik dan dapat membedakan antara peserta didik yang memiliki kemampuan tentang materi dengan yang tidak.

Kegiatan pengembangan bank soal tidak dapat dipisahkan dari kegiatan standarisasi tes prestasi. Menurut Anastasi dan Urbina (1997, dalam Safari, 2005), langkah-langkah untuk pengembangan suatu bank soal dapat dilakukan mulai dari analisis soal secara kualitatif (terkait isi dan bentuknya), dan kuantitatif (terkait ciri-ciri statistiknya).

Rasch Model

Model Rasch adalah bagian dari aplikasi konsep *Item Response Theory* (selanjutnya disebut IRT). Kelebihan IRT adalah bahwa (a) IRT tidak berdasarkan grup *dependent*, (b) skor peserta didik dideskripsikan bukan test *dependent*, (c) model ini menekankan pada tingkat butir soal bukan tes, (d) IRT tidak memerlukan paralel tes untuk menentukan reliabilitas tes, dan (e) IRT suatu model yang memerlukan suatu pengukuran ketepatan untuk setiap skor tingkat kemampuan.

Selanjutnya, Hambleton dan Swaminathan (1985, dalam Safari, 2005) menyatakan bahwa tujuan utama IRT adalah memberikan kesamaan antara statistik soal dan estimasi kemampuan. Ada tiga keuntungan IRT sebagai berikut.

1. Asumsi banyak soal yang diukur pada *trait* yang sama, perkiraan tingkat kemampuan peserta didik adalah *independent*.
2. Asumsi pada populasi tingkat kesukaran, daya pembeda merupakan independen sampel yang menggambarkan untuk tujuan kalibrasi soal.
- c. Statistik yang digunakan untuk menghitung tingkat kemampuan siswa diperkirakan dapat terlaksana.

Selanjutnya, terdapat empat macam model IRT (Hambleton, 1993, dalam Safari, 2005) sebagai berikut.

1. Model satu parameter (Model Rasch), yaitu untuk menganalisis data yang hanya menitikberatkan pada parameter tingkat kesukaran soal

2. Model dua parameter, yaitu untuk menganalisis data yang hanya menitikberatkan pada parameter tingkat kesukaran dan daya pembeda soal.
3. Model tiga parameter, yaitu untuk menganalisis data yang menitikberatkan pada parameter tingkat kesukaran soal, daya pembeda soal, dan menebak (*guessing*).
4. Model empat parameter, yaitu untuk menganalisis data yang menitikberatkan pada parameter tingkat kesukaran soal, daya beda soal, menebak, dan penyebab lain.

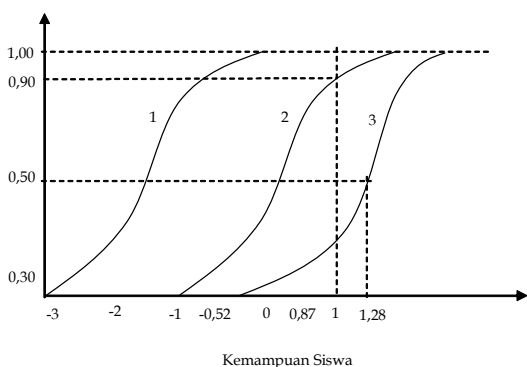
Pada penelitian ini yang akan diaplikasikan adalah Model Rasch. Model Rasch adalah suatu formulasi matematik yang menghubungkan kemungkinan/ peluang (*probability*) hasil (menjawab benar atau salah) jika peserta didik menjawab suatu butir soal sesuai dengan karakteristik orang dan butir soal tersebut (Hayat, 1995).

Peluang menjawab benar setiap butir pada model Rasch atau model satu parameter dapat diketahui dengan menggunakan rumus berikut ini.

$$P_i(0) = \frac{e^{IX^\circ - bi}}{1 + e^{D(O - bi)}} \quad \text{atau} \quad P_i(0) = \frac{1}{1 + e^{D(E - bi)}}$$

Estimasi data yang lebih teliti dan akurat hasilnya adalah menggunakan komputer seperti menggunakan program *Bigsteps*. Dalam program *Bigsteps*, estimasi data digunakan metode *Approximation Maximum Likelihood* (PROX) dan *Unconditional Maximum Likelihood* (UCON). Untuk mendapatkan hasil yang akurat, estimasi data dengan komputer dapat melakukan iterasi maksimum untuk metode PROX, misal bisa sampai 20 kali kemudian dilanjutkan dengan metode UCON sampai dengan 50 kali tergantung banyaknya data. Perbedaan hasil kalibrasi pada setiap iterasi semakin lama semakin kecil dan akan berhenti bila prosesnya sudah terpenuhi (*converge*) atau lebih kecil dari 0,01.

Peluang menjawab benar



Gambar 1. Contoh Kurva Ciri Soal Model Satu Parameter atau Rasch

Pengembangan Bank Soal Berdasarkan Rasch Model

Pengembangan bank soal dengan menggunakan *Rasch Model* akan menghasilkan soal-soal yang teruji tingkat kesukarannya. Tingkat kesukaran soal adalah peluang untuk menjawab benar suatu soal pada tingkat kemampuan tertentu yang biasanya dinyatakan dalam bentuk indeks.

Fungsi tingkat kesukaran butir soal biasanya dikaitkan dengan tujuan tes. Misalnya untuk keperluan ujian semester digunakan butir soal yang memiliki tingkat kesukaran sedang, untuk keperluan seleksi digunakan butir soal yang memiliki tingkat kesukaran tinggi/sukar, dan untuk keperluan diagnostik biasanya digunakan butir soal yang memiliki tingkat kesukaran rendah/mudah.

Tingkat kesukaran butir soal memiliki dua kegunaan, yaitu kegunaan bagi pengajar, pengujian, dan pengajaran (Nitko, 1996, dalam Safari, 2005). Kegunaannya bagi pengajar adalah (1) sebagai pengenalan konsep terhadap pembelajaran ulang dan memberi masukan kepada peserta didik tentang hasil belajar pengajar, dan (2) memperoleh informasi tentang penekanan kurikulum atau mencurigai terhadap butir soal yang bias.

Adapun kegunaannya bagi pengujian dan pengajaran adalah (1) pengenalan konsep yang diperlukan untuk diajarkan ulang, (2) tanda-tanda terhadap kelebihan dan kelemahan pada kurikulum, (3) memberi masukan kepada peserta didik, (4) tanda-tanda kemungkinan adanya butir soal yang bias, dan (5) merakit tes yang memiliki ketepatan data soal.

Tingkat kesukaran butir soal juga dapat digunakan untuk memprediksi alat ukur itu sendiri (soal) dan kemampuan peserta didik dalam memahami materi yang diajarkan guru. Misalnya satu butir soal termasuk kategori mudah, maka prediksi terhadap informasi ini adalah (1) pengecoh butir soal itu tidak berfungsi dan (2) sebagian besar siswa menjawab benar butir soal itu; artinya bahwa sebagian besar siswa telah memahami materi yang ditanyakan.

Bila suatu butir soal termasuk kategori sukar, maka prediksi terhadap informasi ini adalah (1) butir soal itu "mungkin" salah kunci jawaban; (2) butir soal itu mempunyai dua atau lebih jawaban yang benar; (3) materi yang ditanyakan belum diajarkan atau belum tuntas pembelajarannya; sehingga kompetensi minimum yang harus dikuasai siswa belum tercapai; (4) materi yang diukur tidak sesuai, ditanyakan dengan menggunakan bentuk soal yang diberikan; serta (5) pernyataan atau kalimat soal terlalu kompleks dan panjang.

METODOLOGI PENELITIAN

Metode yang digunakan dalam penelitian ini adalah metode *research and development*. Metode ini bertujuan untuk menghasilkan produk tertentu dan menguji keefektifan produk tersebut. Pada penelitian ini, produk yang dimaksud adalah bank soal yang telah teruji properti psikometriknya. Berikut adalah pemaparan hal-hal yang terkait dengan metodologi penelitian.

Prosedur Penelitian

Berikut ini beberapa prosedur penelitian sebagai berikut.

1. Menyusun soal-soal berdasarkan silabus resmi oleh MKDK dan beberapa referensi Psikologi Perkembangan yang secara umum digunakan oleh para dosen pengajar.
2. Menguji coba soal-soal secara kualitatif (keterbacaan) dan kuantitatif. Subjek yang digunakan untuk proses uji coba adalah kelas-kelas mata kuliah Psikologi Perkembangan (MKDK).
3. Menganalisis data yang diperoleh menggunakan program komputer *IteMan* dan *Bigsteps*.
4. Melakukan pemilihan soal-soal yang layak dimasukkan sebagai bagian dari bank soal mata kuliah Psikologi Perkembangan.

Subjek Penelitian

Populasi penelitian adalah mahasiswa kependidikan di Universitas Negeri Jakarta. Sampel yang digunakan berjumlah 224 mahasiswa yang berasal dari fakultas penyelenggara program kependidikan di UNJ. Sampel berjenis kelamin laki-laki dan perempuan dan berasal dari berbagai angkatan.

Teknik Pengambilan Sampel

Teknik pengambilan sampel adalah *nonprobability sampling* dengan cara kuota *sampling*. Teknik ini dipilih karena sumber data yang sangat luas sehingga untuk menentukan subjek mana yang akan dijadikan sumber data maka pengambilan sampelnya berdasarkan jumlah sampel yang ditetapkan (Sugiyono, 2007). Tempat pengambilan sampel adalah kelas-kelas yang sedang menyelenggarakan mata kuliah Psikologi Perkembangan.

Instrumen Penelitian

Instrumen yang akan digunakan dalam penelitian ini terdiri dari kumpulan soal-soal mata kuliah Psikologi Perkembangan. Soal-soal disusun berdasarkan materi pada silabus yang diterbitkan oleh bagian MKDK Fakultas Ilmu Pendidikan. Instrumen yang digunakan dibagi ke dalam dua paket soal: paket soal A (berisi 46 soal) dan paket soal B (berisi 49 soal). Bobot

mata kuliah ini 2 SKS. Muatan materi pada mata kuliah ini ada enam, yaitu konsep dasar psikologi perkembangan, prinsip dan karakteristik perkembangan, aspek-aspek perkembangan yang menjadi perbedaan individual, teori-teori perkembangan, dan tahap-tahap perkembangan.

Teknik Analisis Data

Data penelitian yang diperoleh akan dianalisis melalui dua tahap sebagai berikut.

Tahap pertama dengan program *IteMan* untuk menguji tingkat kesukaran soal, daya pembeda, kualitas distraktor, dan koefisien reliabilitas soal. Konsep dasar teknik ini adalah *Classical Test Theory*, di mana kualitas soal akan sangat ditentukan oleh kondisi sampel yang digunakan (*sample dependent*). Teknik analisis ini digunakan pada kedua jenis paket soal A dan B.

Tahap kedua dengan teknik analisis kuantitatif berdasarkan *Item Response Theory* parameter logistik. Konsep dasar dan hasil perhitungan dengan teknik *Item Response Theory* tidak akan tergantung pada kondisi sampel. Program komputer yang digunakan untuk kepentingan analisis data adalah program *Bigsteps*. Menurut Setiadi (1999), *Bigsteps* adalah *software* yang diciptakan secara khusus untuk analisis dan kalibrasi butir soal. *Bigsteps* diterbitkan dan dikembangkan pertama sekali oleh Benjamin D. Wright dan John Linacre dari University of Chicago, AS, pada tahun 1992. Penyusunan *software* ini berdasarkan pada pendekatan *Rasch Model*. *Output* analisis dengan *Bigsteps* akan menghasilkan informasi mengenai skor yang diperoleh, estimasi tingkat kemampuan peserta tes, estimasi tingkat kesukaran butir soal dalam bentuk skala baku yang disebut *Logit (log odds unit)*, indeks kesalahan pengukuran, kecocokan antara data dan model, daya pembeda, dsb. Namun dalam penelitian ini yang akan dipertimbangkan adalah kecocokan antara data dengan model, estimasi tingkat kemampuan peserta tes, dan estimasi tingkat kesukaran butir soal.

HASIL DAN PEMBAHASAN

Pelaksanaan tes dalam rangka pengumpulan data menggunakan dua jenis paket soal: paket soal A dan B. Data yang diperoleh dari pelaksanaan tes dengan paket soal A dianalisis dengan program *IteMan* dan *Bigsteps*. Sementara data yang diperoleh dari pelaksanaan tes dengan paket soal B dianalisis dengan program *IteMan* saja. Berikut akan dipaparkan tentang pelaksanaan tes, proses analisis data, dan hasil

analisis data pada setiap paket soal.

Paket Soal A

1. Pelaksanaan Tes Menggunakan Paket Soal A

Tes dilaksanakan sebanyak dua kali dan masing-masing terdiri dari dua kelas mahasiswa. Total seluruh peserta ujian adalah 122 mahasiswa. Pelaksanaan ujian pada waktu yang berbeda karena peserta ujian berasal dari program studi kependidikan yang berbeda.

2. Proses dan Hasil Analisis Data

a. Analisis Data dengan Iteman

Data yang diperoleh dari 122 mahasiswa yang mengikuti tes dianalisis dengan *Iteman* untuk mengedrop soal yang tidak berkorelasi positif dengan tes secara keseluruhan. Jumlah soal yang dianalisis sebanyak 46 soal. Soal yang didrop karena nilai poin biserialnya negatif adalah soal nomor 1 dan 19. Masing-masing nilai koefisien korelasi poin biserial untuk kedua soal itu adalah -0.024 dan -0.012. Sementara soal yang ditandai dengan “*check the key*” oleh *Iteman* tidak didrop dengan asumsi selama korelasi poin biserial positif maka soal tersebut termasuk valid secara konstruk. Adanya tanda ‘*check the key*’ dikarenakan cukup banyak subjek yang memilih distraktor soal dibandingkan kunci jawaban. Namun, selama ada peserta tes yang menjawab sesuai dengan kunci jawaban, maka soal tersebut diasumsikan layak untuk dianalisis pada tahap selanjutnya. Kemudian, hasil analisis yang diperoleh dapat digunakan untuk menentukan tingkat kesukaran, daya diskriminasi, dan analisis distraktor. Berikut ini penjabaran dari ketiga hal tersebut.

- 1) Tingkat kesukaran diperoleh berdasarkan angka yang tertera sebagai *proportional correct* (p). Nilai p yang baik adalah jika berada di sekitar nilai 0.5 atau ada juga yang berpendapat nilai p yang baik jika berada di antara 0.25-0.75. Semakin tinggi nilai p dan mendekati 1 maka menunjukkan soal yang semakin mudah. Sebaliknya, jika nilai p semakin kecil dan mendekati 0 maka soal semakin sukar.
- 2) Daya diskriminasi (D) adalah sejauh mana suatu soal melakukan diferensiasi dengan benar di antara para peserta tes pada objek yang menjadi sasaran pengukuran oleh tes yang disajikan. Nilai D dapat diketahui dari hubungan antara skor soal dengan skor total. Nilai D yang baik ditunjukkan dari nilai korelasi yang positif-signifikan dan tinggi (mendekati 1). Rumus korelasi yang dapat digunakan antara lain dengan rumus korelasi biserial dan poin biserial.

Nilai D berada di antara -1 s/d +1. Semakin positif dan mendekati 1 maka daya diskriminasi soal semakin baik. Dalam analisis daya diskriminasi tes kognitif dengan paket soal A ini digunakan koefisien korelasi poin biserial.

- 3) Analisis distraktor, yaitu analisis terhadap alternatif jawaban bukan kunci yang merupakan pengecoh sehingga diharapkan orang yang dikenai tes dan berada pada kelompok bawah (kelompok yang tidak punya kemampuan) akan memilih jawaban bukan kunci yang menjadi distraktor dan relatif merata pada pilihan jawaban salah tersebut. Dengan demikian, akan dapat dibedakan penjawab soal yang tahu jawaban kunci dengan penjawab soal yang sebenarnya tidak tahu jawaban kunci. Analisis distraktor diketahui dari proporsional *endorsing* dan korelasi biserial atau point biserial. Jika proporsional *endorsing* pada setiap alternatif jawaban selain kunci/distraktor < 0.10 maka sebaiknya alternatif jawaban tersebut direvisi total atau dihilangkan saja. Jika dilihat dari nilai koefisien korelasi biserial/poin biserial, maka distraktor yang baik adalah koefisien korelasi biserial/poin biserial yang bernilai negatif dan signifikan.

Berikut adalah hasil analisis data menggunakan program *Iteman* secara umum terhadap paket soal A. Statistik yang diperoleh adalah mean, median, varians, standar deviasi, minimum, maksimum, koefisien *Alpha*, dan *standart error of measurement*.

Tabel 1. Hasil Analisis Data Paket Soal A Menggunakan Program *Iteman*

Statistik	Nilai
Mean	24.697
Varians	71.162
Standar Devasi	8.436
Minimum	9
Maximum	41
Median	23
Koefisien Alpha	0.88
SEM	2.916

Mean empiris yang diperoleh adalah 24.697, hal ini berarti rata-rata skor subjek penelitian sebesar nilai tersebut. Varians yang diperoleh sebesar 71.162, artinya keragaman data cukup besar dan data heterogen. Standar deviasi menunjukkan standar besarnya penyimpangan dari mean, yaitu sebesar 8.436. Skor terendah/minimum yang dicapai subjek adalah 9 dan skor tertinggi/maksimum adalah 41. Hal

ini menggambarkan adanya subjek yang hanya menjawab 9 soal dengan benar dari sejumlah 46 soal yang disajikan. Namun demikian, ada juga subjek yang menjawab 41 soal dengan benar. Nilai tengah/median yang diperoleh adalah 23. Kemudian, reliabilitas instrumen 'tes A' ini sebesar 0.88 dengan *standar error of measurement* sebesar 2.916. *Standar error of measurement* memberikan interpretasi secara relatif terhadap makna koefisien reliabilitas. Tes yang lebih reliabel adalah tes yang paling kecil varians *error*-nya dalam melakukan pengukuran. Dalam hal ini, perlu membandingkan hasil analisis data paket soal A dengan paket soal B nantinya.

Sementara itu, hubungan antara varians *error* dengan *standart error of measurement* digambarkan dalam rumus berikut ini:

$$s_e = s_x \sqrt{1 - r_{xx'}}$$

Se = standar error of measurement

Sx = akar varians error

R_{xx'} = koefisien reliabilitas tes

b. Analisis Data dengan *Bigsteps*

1). *Person* dan item yang *didrop* pada *run I* dan *II*

Setelah memperoleh soal valid dari analisis dengan *Iteman* (sebanyak 44 soal), selanjutnya dilakukan analisis dengan *Bigsteps*. Analisis dengan *Bigsteps* memiliki konsep *Item Response Theory* dengan satu parameter (*Rasch Model*). Proses iterasi dilakukan selama beberapa kali untuk mengedrop *person* dan *item*/soal. Proses mengedrop *person* biasanya disebut dengan *Run I*, sementara proses mengedrop *item*/soal biasanya disebut dengan *Run II*.

Untuk menyeleksi data statistik orang yang tidak *fit* dengan model, maka digunakan kriteria sebagai berikut nilai '*outfit*' statistik lebih besar dari +2.0 atau nilai '*poin biserial*' negatif.

Terdapat 122 subjek dan 44 soal yang dianalisis. *Mean score* dari *summary of person* sebesar 9.9 yang menunjukkan rata-rata skor mentah dari seluruh subjek yang mengerjakan 44 soal. Rata-rata kemampuan subjek sebesar -1.73 dalam Logit. *Mean error* (rata-rata kesalahan pengukuran) = 0.44 yang menunjukkan rata-rata dari standar eror seluruh subjek. *Mean error* item lebih kecil dibandingkan dengan *mean error person* (0.31 < 0.44). Dikarenakan lebih banyaknya data yang dianalisis pada soal (sebanyak 144 subjek) maka reliabilitasnya lebih besar (0.94 dibandingkan dengan 0.32). Pada *Summary of Items*, *mean measure* sebesar 0.00 karena memang program *Bigsteps* ini membuat rata-rata kalibrasi seluruh soal sebesar 0, kecuali bila dikalibrasi dengan *anchor item* dari bank soal yang sudah ada sebelumnya.

Hasil standarisasi baik pada *infit* maupun *outfit* (untuk mengetahui apakah suatu data fit dengan model) menghasilkan distribusi data tersebut di mana mean sama dengan atau mendekati 0 dan nilai standar deviasinya sama dengan atau mendekati 1.

c. *Person Statistics: Outfit Order*

Data ini memperlihatkan subjek yang tidak fit dengan model Rasch. Terdapat 3 subjek yang harus di drop dari analisis karena memiliki nilai *outfit* yang lebih besar dari +2.0 atau nilai poin biserial yang negatif.

d. *Items Statistics: Outfit order*

Pada data ini diperoleh gambaran soal yang tidak fit dengan model Rasch, karena memiliki nilai *outfit* 3.4 (lebih besar dari +2.0) dan 14 item lain karena nilai koefisien poin biserial negatif.

Berdasarkan hasil analisis tersebut diperoleh hasil total soal yang gugur dengan menggunakan program *Iteman* dan *Bigsteps* sebanyak 17 soal dan soal yang valid sebanyak 29 soal. Selain itu, berdasarkan hasil analisis diperoleh gambaran *map of persons and items* yang kurang fit. Hal ini dikarenakan kemampuan subjek yang mengerjakan paket soal A kurang fit dengan tingkat kesukaran soal sehingga terlihat soal-soal yang disajikan kurang dapat dikerjakan dengan baik oleh subjek, atau soal terlalu sulit untuk subjek dengan kemampuan demikian.

Terjadinya ketidakcocokan atau ketidak-fit-an tingkat kesukaran soal dengan kemampuan subjek dikarenakan berbedanya dosen pengajar di berbagai kelas MKDK yang mahasiswanya dipakai sebagai subjek penelitian ini. Improvisasi terhadap aplikasi isi silabus mata kuliah berbeda-beda antardosen meskipun silabus yang digunakan sama. Selain itu, belum ada pemakaian buku ajar yang standar untuk pembelajaran mata kuliah Psikologi Perkembangan sebagai mata kuliah MKDK.

Paket Soal B

1. Pelaksanaan tes menggunakan paket soal B

Tes dilaksanakan sebanyak dua kali dan masing-masing terdiri dari dua kelas mahasiswa. Total seluruh peserta ujian adalah 102 mahasiswa. Pelaksanaan ujian pada waktu yang berbeda karena peserta ujian berasal dari program studi kependidikan yang berbeda.

2. Proses dan Hasil Analisis Data

Analisis Data dengan *Iteman*

Data yang diperoleh dari 104 mahasiswa yang mengikuti tes dianalisis dengan *Iteman* untuk mendrop soal yang tidak berkorelasi positif dengan tes secara keseluruhan. Jumlah soal yang dianalisis sebanyak 49 soal. Soal yang didrop karena nilai poin biserialnya

negatif berjumlah 7 soal. Nilai koefisien korelasi poin biserial untuk ketujuh soal itu berkisar antara -0.012 sampai dengan -0.096. Sementara soal yang ditandai dengan 'check the key' oleh *Iteman* tidak didrop dengan asumsi selama korelasi poin biserial positif maka soal tersebut termasuk valid secara konstruk.

Adanya tanda 'check the key' dikarenakan cukup banyak subjek yang memilih distraktor soal dibandingkan kunci jawaban. Namun, selama ada peserta tes yang menjawab sesuai dengan kunci jawaban, maka soal tersebut diasumsikan layak untuk dianalisis pada tahap selanjutnya. Berikut adalah hasil analisis data menggunakan program *Iteman*. Statistik yang diperoleh adalah mean, median, varians, standar deviasi, minimum, maksimum, koefisien *Alpha*, dan *standart error of measurement*.

Tabel 2. Hasil Analisis Data Paket Soal B Menggunakan Program *Iteman*

Statistik	Nilai
Mean	20.98
Varians	12.392
Standar Deviasi	3.52
Minimum	12
Maximum	31
Median	21
Koefisien Alpha	0.225
SEM	3.1

Mean empiris yang diperoleh adalah 20.98, hal ini berarti rata-rata skor subjek penelitian sebesar nilai tersebut. Varians yang diperoleh sebesar 12.392, artinya keragaman data cukup besar, namun data berdasarkan paket soal A lebih heterogen. Standar deviasi menunjukkan standar besarnya penyimpangan dari mean, yaitu sebesar 3.52. Skor terendah/minimum yang dicapai subjek adalah 12 dan skor tertinggi/maksimum adalah 31. Hal ini menggambarkan adanya subjek yang hanya menjawab 12 soal dengan benar dari sejumlah 49 soal yang disajikan. Namun demikian, ada juga subjek yang menjawab 31 soal dengan benar. Nilai tengah/median yang diperoleh adalah 21.

Kemudian, reliabilitas instrumen paket soal B ini sebesar 0.225 dengan *standart error of measurement* sebesar 3.1. Jika dibandingkan dengan paket soal A, maka paket soal B ini lebih rendah reliabilitasnya dan *standart error of measurement*-nya juga lebih besar. Dengan demikian, analisis data pada paket soal B menghasilkan 42 soal valid berdasarkan *output* program *Iteman* yang menggunakan pendekatan *Classical Test Theory*.

PENUTUP

Kesimpulan

Berdasarkan hasil analisis data yang dilakukan dengan menggunakan program *Iteman* dan *Bigstep* terhadap soal-soal, maka diperoleh kesimpulan sebagai berikut.

Pertama, paket soal A yang pada awalnya terdiri dari 46 soal dianalisis menggunakan program *Iteman* dan *Bigsteps*. Analisis tersebut menghasilkan jumlah soal valid sebanyak 29 buah dan soal gugur sebanyak 17 buah.

Kedua, paket soal B yang pada awalnya terdiri dari 49 soal dianalisis menggunakan program *Iteman*. Analisis tersebut menghasilkan jumlah soal valid sebanyak 42 buah dan soal gugur sebanyak 7 buah.

Ketiga, karakteristik soal paket A dengan subjek kurang fit atau cocok. Kemampuan subjek jauh di bawah tingkat kesukaran soal. Kemungkinan hal ini dikarenakan materi yang diajarkan oleh dosen pengajar memiliki muatan yang berbeda dengan konten soal, meskipun menggunakan silabus yang sama.

Saran

Saran yang diajukan berdasarkan hasil penelitian ini adalah bagi peneliti selanjutnya, memperluas sampel yang menjadi subjek penelitian sehingga hasil penelitian dapat lebih menggambarkan populasi mahasiswa kependidikan serta memperbanyak jumlah soal yang diuji sehingga akan memperkaya bank soal mata kuliah terkait.

Bagi pihak pengajar, dalam setiap evaluasi pembelajaran hendaknya menggunakan soal-soal ujian yang sudah teruji properti psikometriknya, sehingga kompetensi mahasiswa dapat dibandingkan antara satu sama lain meskipun berbeda dosen pengajar dan asal program studinya.

DAFTAR PUSTAKA

- Hayat, B. & Setiadi, H. (1998). *Mendesain perangkat tes dengan model Rasch*. Jakarta: Depdikbud
- Safari. (2005). *Teknik analisis butir soal instrumen tes dan nontes*. Jakarta: Assosiasi Pengawas Sekolah Indonesia-Depdiknas
- Setiadi, H. (1999). *Penggunaan program Bigsteps untuk pengembangan bank soal*. Jakarta: Depdikbud
- Sugiyono. (2007). *Metode penelitian pendidikan: Pendekatan kuantitatif, kualitatif, dan R&D*. Bandung: Alfabeta

KETERANGAN PENULIS

Program studi Psikologi Fakultas Ilmu Pendidikan
Universitas Negeri Jakarta.

Anna Armeini Rangkuti, lahir di Padangsidempuan
pada tanggal 24 Mei 1976. Saat ini aktif sebagai Dosen