

# LAPORAN PENELITIAN

RESEARCH GROUP  
NAMA RISET GRUP: EVALUASI PEMBELAJARAN IPA

JUDUL:

PENINGKATAN KOMPETENSI PROFESIONAL CALON GURU IPA  
DALAM PENGEMBANGAN ASSESSMENT BERSTANDAR SURVEI  
BENCHMARKING INTERNASIONAL (PISA) BERORIENTASI DAYA  
SAING GLOBAL



Oleh:

<b>Dr. Dadan Rosana, M.Si</b>	<b>/NIP. 19591212 198702 1 001</b>	<b>Ketua</b>
<b>Eko Widodo, M.Pd</b>	<b>/NIP. 19591212 198702 1 001</b>	<b>Anggota</b>
<b>Wita Setianingsih, M.Pd</b>	<b>/NIP. 19800422 200501 2 001</b>	<b>Anggota</b>
<b>Didik Setyawarno, M.Pd</b>	<b>/NIP. 19881013 201504 1 004</b>	<b>Anggota</b>
<b>Haedar Ahmad Hanafi</b>	<b>/NIM. 15312241025</b>	<b>Mahasiswa</b>
<b>Yustar Afif Priambodo</b>	<b>/NIM. 15312241028</b>	<b>Mahasiswa</b>

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS NEGERI YOGYAKARTA  
Agustus 2018

**LEMBAR PENGESAHAN  
LAPORAN PENELITIAN RESEARCH GROUP**

1. Judul Penelitian : Peningkatan Kompetensi Profesional Calon Guru IPA dalam Pengembangan Assessment Berstandar Survei Benchmarking Internasional (PISA) Berorientasi Daya Saing Global
2. Ketua Peneliti :
- a. Nama lengkap : Dr. Dadan Rosana, M.Si.
- b. Jabatan : Lektor Kepala
- c. Program Studi : Pendidikan IPA
- d. Alamat : Citra Ringin Mas C-13 Purwomartani, Kalasan, Sleman, Yogyakarta
- e. Telepon : +6281392859303
- f. e-mail : danrosana@uny.ac.id
3. Nama Research Group : Evaluasi Pembelajaran IPA
4. Tim Pengabdian :

No	Nama, Gelar	NIP	Bidang Keahlian
1.	Didik Setyawarno, S.Pd.Si., M.Pd.	19881013 201504 1 004	Pendidikan IPA
2.	Drs. Eko Widodo, M.Pd.	19591212 198702 1 001	Pendidikan IPA
3.	Wita Setianingsih, S.Pd., M.Pd.	19800422 200501 2 001	Pendidikan IPA

5. Mahasiswa yang terlibat :

No	Nama	NIM	Prodi
1.	Haedar Ahmad Hanafi	15312241025	Pend. Ilmu Pengetahuan Alam
2.	Yustar Afif Priambodo	15312241028	Pend. Ilmu Pengetahuan Alam

6. Lokasi Penelitian : FMIPA Universitas Negeri Yogyakarta
7. Waktu Penelitian : 3 Februari 2018 s/d 10 Agustus 2018
8. Dana yang diusulkan : Rp. 20.000.000,00

Mengesahkan,  
Dekan,



Dr. Hartono, M.Si.  
NIP 19620329 198702 1 002

Yogyakarta, 9 Agustus 2018  
Ketua Pelaksana



Dr. Dadan Rosana, M.Si.  
NIP 19690202 199303 1 002

## **KATA PENGANTAR**

Puji syukur ke hadirat Allah Swt, Tuhan YME, yang telah memberikan rahmat yang tidak terhingga kepada kita semua sehingga Laporan Penelitian Kelompok Bidang Keahlian (Research Group) dengan judul “Peningkatan kompetensi profesional calon guru IPA dalam pengembangan assessment berstandar *survei benchmarking* internasional (PISA) berorientasi daya saing global” telah selesai dengan baik. Laporan Penelitian ini dirancang dalam bentuk program kerjasama pada Kelompok Bidang Keahlian (*Research Group*) Evaluasi Pembelajaran IPA sebagai bagian dari kelompok bidang keahlian di Jurusan Pendidikan IPA FMIPA Universitas Negeri Yogyakarta. Kegiatan ini disusun relevan dengan Tugas Pokok dan Fungsi Program Studi Pendidikan IPA yang berkomitmen untuk meningkatkan kualitas pendidikan IPA khususnya di tingkat Sekolah Menengah Pertama.

Semoga laporan kegiatan ini nantinya dapat dirasakan manfaatnya oleh berbagai pihak, khususnya para Guru IPA SMP di DIY. Aamiin.

Yogyakarta, 10 Agustus 2018

Penyusun,

Tim Peneliti RG Evaluasi  
Prodi Pendidikan IPA

## ABSTRAK

Standar guru Ilmu Pengetahuan Alam (IPA) di Indonesia masih perlu ditingkatkan kualitasnya, khususnya bila dikaitkan dengan *survei benchmarking* internasional. Salah satu survei tersebut adalah Programme for International Student Assessment (PISA), mengukur apa yang diketahui siswa dan apa yang dapat dia lakukan (aplikasi) dengan pengetahuannya. Tujuan dari penelitian ini adalah, meningkatkan kompetensi profesional calon guru IPA dalam pengembangan assessment berstandar *survei benchmarking* internasional (PISA) agar berdaya saing global, dengan sasaran strategis mahasiswa calon guru di Lembaga Pendidikan Tenaga Kependidikan (LPTK). Tahapan penelitian yang dilakukan meliputi; (1) mendapatkan hasil pengujian validitas dan reliabilitas butir soal untuk pengukuran assessment berstandar *survei benchmarking* internasional (PISA), (2) menghasilkan pemetaan kompetensi akademik calon guru khususnya kemampuan dalam mengembangkan assessment berstandar *survei benchmarking* internasional (PISA), (3) meningkatkan kompetensi calon guru dalam pengembangan assessment berstandar *survei benchmarking* internasional (PISA) melalui workshop dan pendampingan.

Metodelogi yang dikembangkan dalam penelitian ini adalah model analisis butir test klasik dan *Item Respons Theory* serta Research & Development model spiral sebagaimana yang direferensikan oleh Cennamo dan Kalk (2005:6). Dalam model spiral ini dikenal 5 (lima) fase pengembangan yakni: (1) definisi (*define*), (2) desain (*design*), (3) peragaan (*demonstrate*), (4) pengembangan (*develop*), dan (5) penyajian (*deliver*). Penelitian dilakukan di Jurusan Pendidikan IPA FMIPA Universitas Negeri Yogyakarta, dengan melibatkan mahasiswa angkatan tahun 2014 yang telah mengambil matakuliah Evaluasi Pembelajaran IPA dan sedang mempersiapkan penulisan Tugas Akhir Skripsi.

Hasil Penelitian ini adalah; (1) dihasilkannya perangkat test HOTS berbasis PISA yang valid berdasarkan validasi isi dan konstruk oleh dosen ahli evaluasi, (2) hasil pemetaan kompetensi akademik calon guru khususnya kemampuan dalam mengembangkan assessment berstandar *survei benchmarking* internasional (PISA) adalah; 6,06 % berada pada kategori sangat baik, 71,21 % baik, dan 22,73 pada kategori kurang, (3) terdapat pengaruh yang signifikan dari workshop dan pendampingan terhadap peningkatan kompetensi calon guru dalam pengembangan assessment berstandar *survei benchmarking* internasional (PISA) berdasarkan uji beda nilai pretest dan posttest dengan *Independent Sample T-Test* diperoleh nilai Sig (*2-tailed*) sebesar  $0,018 < 0,025$  dan nilai *effect size* sebesar 0,86.

**Kata kunci;** *Assessment terstandar, survei benchmarking internasional (PISA), kompetensi profesional, calon guru IPA*

## DAFTAR ISI

HALAMAN DEPAN .....	
PRAKATA .....	
HALAMAN PENGESAHAN .....	
ABSTRAK/RINGKASAN.....	
DAFTAR ISI .....	
BAB I      PENDAHULUAN .....	
BAB II     TINJAUAN PUSTAKA .....	
BAB III    METODE PENELITIAN .....	
DAFTAR PUSTAKA .....	
LAMPIRAN .....	

## **BAB I**

### **PENDAHULUAN**

#### **A. Latar Belakang Masalah**

Latar belakang permasalahan penelitian ini adalah terkait dengan belum dikembangkannya assessment literasi sains dan high order thinking skills, khususnya bila dikaitkan dengan *survei benchmarking* internasional seperti Programme for International Student Assessment (PISA). PISA merupakan sistem ujian yang diinisiasi oleh Organisation for Economic Cooperation and Development (OECD), untuk mengevaluasi sistem pendidikan dari 72 negara di seluruh dunia. Setiap tiga tahun, siswa berusia 15 tahun dipilih secara acak, untuk mengikuti tes dari tiga kompetensi dasar yaitu membaca, matematika dan sains. PISA mengukur apa yang diketahui siswa dan apa yang dapat dia lakukan (aplikasi) dengan pengetahuannya. Tema survei digilir setiap 3 tahun, tahun 2015 fokus temanya adalah kompetensi sains.

Pemerintah Indonesia mulai memberikan perhatian serius terhadap survei atau pemetaan Internasional karena terkait dengan daya saing bangsa di era global. Kepala Badan Penelitian dan Pengembangan (Balitbang) Kemdikbud, Totok Suprayitno, menyampaikan bahwa peningkatan capaian Indonesia tahun 2015 cukup memberikan optimisme, meskipun masih rendah dibanding rerata OECD. Berdasar nilai rerata, terjadi peningkatan nilai PISA Indonesia di tiga kompetensi yang diujikan. Peningkatan terbesar terlihat pada kompetensi sains, dari 382 poin pada tahun 2012 menjadi 403 poin di tahun 2015. Dalam kompetensi matematika meningkat dari 375 poin di tahun 2012 menjadi 386 poin di tahun 2015. Kompetensi membaca belum menunjukkan peningkatan yang signifikan, dari 396 di tahun 2012 menjadi 397 poin di tahun 2015. Peningkatan tersebut mengangkat posisi Indonesia 6 peringkat ke atas bila dibandingkan posisi peringkat kedua dari bawah pada tahun 2012.

Sedangkan, berdasar nilai median, capaian membaca siswa Indonesia meningkat dari 337 poin di tahun 2012 menjadi 350 poin di tahun 2015. Nilai matematika melonjak 17 poin dari 318 poin di tahun 2012, menjadi 335 poin di tahun 2015. Lonjakan tertinggi terlihat pada capaian sains yang mengalami kenaikan dari 327 poin di tahun 2012 menjadi 359 poin di tahun 2015. Peningkatan capaian median yang lebih tinggi dari mean ini merupakan indikator yang baik dari sisi peningkatan akses dan pemerataan kualitas secara inklusif.

Kepala Pusat Penilaian Pendidikan Balitbang (Kapuspendik Balitbang) Kemendikbud mengatakan secara konsisten terjadi peningkatan cakupan sampling mahasiswa calon guru IPA Indonesia yaitu sebanyak 46 persen di tahun 2003 menjadi 53 persen di tahun 2006. Selanjutnya, angka tersebut naik ke 63,4 persen di tahun 2012, dan menjadi 68,2 persen di tahun 2015. “Peningkatan cakupan sampling ini merupakan bukti capaian wajib belajar 9 tahun dan ekspansi menuju wajar 12 Tahun dan inklusi kepesertaan murid Indonesia dalam pendidikan membuahkan hasil” jelasnya, di Jakarta, Selasa (6/12/2016).

Hal yang terpenting dari survei benchmarking internasional seperti PISA ini adalah bagaimana kita melakukan tindak lanjut berdasar diagnosa yang dihasilkan dari survei tersebut. Peningkatan capaian yang terjadi harus terus ditingkatkan dengan meningkatkan mutu pendidikan di Indonesia. Bila laju peningkatan tahun 2012-2015 dapat dipertahankan, maka pada tahun 2030 capaian kita akan sama dengan capaian rerata negara-negara OECD. Karena itu, menjadi sangat penting untuk memasukkan assessment PISA ini dalam pembelajaran bagi seorang calon guru. Standar guru Ilmu Pengetahuan Alam (IPA) di Indonesia masih perlu ditingkatkan kualitasnya, khususnya bila dikaitkan dengan *survei benchmarking* internasional.

#### **A. Rumusan Masalah**

Berdasarkan latar belakang masalah, maka rumusan masalah dari penelitian ini adalah, bagaimana meningkatkan kompetensi profesional calon guru IPA dalam pengembangan assessment berstandar *survei benchmarking* internasional (PISA) agar berdaya saing global, dengan sasaran strategis mahasiswa calon guru di Lembaga Pendidikan Tenaga Kependidikan (LPTK)?

Rumusan di atas dijabarkan menjadi rumusan yang lebih rinci sebagai berikut;

1. bagaimana menghasilkan pemetaan kompetensi akademik calon guru khususnya kemampuan dalam mengembangkan assessment berstandar *survei benchmarking* internasional (PISA)?
2. bagaimana mendapatkan hasil pengujian validitas dan reliabilitas butir soal untuk pengukuran assessment berstandar *survei benchmarking* internasional (PISA)?
3. bagaimana mengembangkan model peningkatan kompetensi calon guru dalam pengembangan assessment berstandar *survei benchmarking* internasional (PISA)?

## **B. Tujuan**

Berdasarkan rumusan masalah dan latar belakang masalah, maka tujuan dari penelitian ini adalah, meningkatkan kompetensi profesional calon guru IPA dalam pengembangan assessment berstandar *survei benchmarking* internasional (PISA) agar berdaya saing global, dengan sasaran strategis mahasiswa calon guru di Lembaga Pendidikan Tenaga Kependidikan (LPTK).

Tujuan di atas dijabarkan menjadi tujuan sebagai berikut;

4. menghasilkan pemetaan kompetensi akademik calon guru khususnya kemampuan dalam mengembangkan assessment berstandar *survei benchmarking* internasional (PISA),
5. mendapatkan hasil pengujian validitas dan reliabilitas butir soal untuk pengukuran assessment berstandar *survei benchmarking* internasional (PISA),
6. mengembangkan model peningkatan kompetensi calon guru dalam pengembangan assessment berstandar *survei benchmarking* internasional (PISA).

## **C. Manfaat dan Urgensi Penelitian**

Sesuai dengan tujuan dan latar belakang permasalahan, maka manfaat dari penelitian ini, diantaranya adalah:

### **1. Manfaat Secara Praktis**

- a. Model peningkatan kompetensi calon guru dalam pengembangan assessment penilaian tes literasi sains dan high order thinking skills terstandar PISA sangat penting untuk mendukung suksesnya tujuan perubahan kurikulum 2013.
- b. Pola pengembangan model peningkatan kompetensi Dosen dalam penyusunan assessment tes literasi sains dan high order thinking skills terstandar PISA dapat dijadikan referensi yang sangat tepat sebagai best practice penentuan standar penilaian yang dapat diterapkan di dalam pembelajaran microteaching untuk membekali mahasiswa sebelum PPL.

### **2. Urgensi Penelitian**

- a. Untuk melakukan mendapatkan hasil penelitian yang dapat menyelesaikan masalah bangsa dan masyarakat dengan fokus bidang pendidikan dalam mengembangkan inovasi sistem penilaian khususnya tes literasi sains dan high order thinking skills terstandar PISA .

- b. Memberikan peluang yang lebih tinggi untuk meningkatkan kompetensi calon guru sebelum turun kelapangan menerapkan kurikulum 2013.
- c. Meningkatkan, menguatkan, dan menjaga kesinambungan periset dan institusi untuk melaksanakan Riset di Universitas Negeri Yogyakarta.
- d. Meningkatkan kompetensi periset Universitas Negeri Yogyakarta pada bidang prioritas model penerapan kurikulum dan model model pembelajaran.
- e. Mengembangkan keilmuan terkini dan pemanfaatannya untuk menyelesaikan permasalahan yang berkembang di masyarakat.

#### **D. Luaran Penelitian dan Kontribusi Terhadap Ilmu Pengetahuan**

Penelitian ini memiliki kontribusi untuk peningkatan kualitas penilaian, dan peningkatan kualitas mahasiswa. Oleh karena itu, luaran dari penelitian ini antara lain adalah:

2. **Model peningkatan kompetensi calon guru dalam mengembangkan assessment tes literasi sains dan high order thinking skills terstandar PISA untuk pembelajaran IPA terpadu.**
3. **Publikasi artikel ilmiah pada jurnal nasional/internasional yang terakreditasi.**  
Pengembangan Model Model peningkatan kompetensi calon guru dalam mengembangkan assessment tes literasi sains dan high order thinking skills terstandar PISA untuk pembelajaran IPA terpadu adalah bersifat aktual dan orisinal karena baru dikembangkan dan belum diteliti secara lebih mendalam, oleh karena itu sangat berpeluang untuk dipublikasikan baik di jurnal nasional maupun internasional.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **A. Penilaian dalam Kurikulum 2013 (K-13)**

Secara konseptual, kurikulum adalah suatu respon pendidikan terhadap kebutuhan masyarakat dan bangsa dalam membangun generasi muda bangsanya. Secara pedagogis, kurikulum adalah rancangan pendidikan yang memberi kesempatan untuk mahasiswa calon guru IPA mengembangkan potensi dirinya dalam suatu suasana belajar yang menyenangkan dan sesuai dengan kemampuan dirinya untuk memiliki kualitas yang diinginkan masyarakat dan bangsanya (2). Berkaitan dengan pentingnya kurikulum, lebih jauh Iftikhar Uddin Khwaja menyatakan bahwa *“one of the most important activities of the university or school is the development of curriculum or course outlines in consonance with the national and international demands and realities”*.

Kurikulum pendidikan tingkat sekolah di Indonesia telah mengalami perubahan secara berkelanjutan. Kurikulum 2013 (K-13) telah menggantikan kurikulum berbasis kompetensi (KTSP). K-13 dan KTSP pada dasarnya sama-sama menekankan penguasaan kompetensi. Kurikulum 2013 (K-13) dicirikan dengan adanya kompetensi inti dan kompetensi dasar. Dru Riddle, et al (2016:239) menyatakan bahwa sebagai berikut.

*“Competency: “An observable ability of a health professional, integrating multiple components such as knowledge, skills, values, and attitudes. Since competencies are observable, they can be measured and assessed to ensure their acquisition”*.

Berdasarkan pengertian tersebut dapat dipahami bahwa kompetensi merupakan kemampuan yang dapat diamati yang mengintegrasikan berbagai komponen seperti pengetahuan, keterampilan, nilai, dan bakat yang dapat diukur dan dinilai. Kompetensi Inti dalam kurikulum 2013 merupakan terjemahan atau operasionalisasi SKL dalam bentuk kualitas yang harus dimiliki mereka yang telah menyelesaikan pendidikan pada satuan pendidikan tertentu atau jenjang pendidikan tertentu, gambaran mengenai kompetensi utama yang dikelompokkan ke dalam aspek sikap, pengetahuan, dan keterampilan (afektif, kognitif, dan psikomotor) yang harus dipelajari mahasiswa calon guru IPA untuk suatu jenjang sekolah, kelas dan mata pelajaran. Kompetensi Inti harus menggambarkan kualitas yang seimbang antara pencapaian hard skills dan soft skills (Kelitbang, 2013:5).

Kompetensi Inti berfungsi sebagai unsur pengorganisasi (*organising element*) Kompetensi Dasar. Sebagai unsur pengorganisasi, kompetensi Inti merupakan pengikat untuk organisasi vertikal dan organisasi horizontal Kompetensi Dasar. Organisasi vertikal Kompetensi Dasar adalah keterkaitan antara konten Kompetensi Dasar satu kelas atau jenjang pendidikan ke kelas/jenjang di atasnya sehingga memenuhi prinsip belajar yaitu terjadi suatu akumulasi yang berkesinambungan antara konten yang dipelajari siswa. Organisasi horizontal adalah keterkaitan antara konten Kompetensi Dasar satu mata pelajaran dengan konten Kompetensi Dasar dari mata pelajaran yang berbeda dalam satu pertemuan mingguan dan kelas yang sama sehingga terjadi proses saling memperkuat.

Kompetensi Dasar merupakan kompetensi setiap mata pelajaran untuk setiap kelas yang diturunkan dari Kompetensi Inti. Kompetensi Dasar adalah konten atau kompetensi yang terdiri atas sikap, pengetahuan, dan ketrampilan yang bersumber pada kompetensi inti yang harus dikuasai mahasiswa calon guru IPA (Kelitbang, 2013:7). Kompetensi tersebut dikembangkan dengan memperhatikan karakteristik mahasiswa calon guru IPA, kemampuan awal, serta ciri dari suatu mata pelajaran. Mata pelajaran sebagai sumber dari konten untuk menguasai kompetensi bersifat terbuka dan tidak selalu diorganisasikan berdasarkan disiplin ilmu yang sangat berorientasi hanya pada filosofi esensialisme dan perenialisme. Mata pelajaran dapat dijadikan organisasi konten yang dikembangkan dari berbagai disiplin ilmu atau non disiplin ilmu yang diperbolehkan menurut filosofi rekonstruksi sosial, progresifisme, atau pun humanisme. Karena filosofi yang dianut dalam kurikulum adalah eklektik seperti dikemukakan di bagian landasan filosofi, maka nama mata pelajaran dan isi mata pelajaran untuk kurikulum yang akan dikembangkan tidak perlu terikat pada kaedah filosofi esensialisme dan perenialisme.

Penilaian adalah proses memberikan atau menentukan nilai kepada objek tertentu berdasarkan suatu kriteria tertentu. Penilaian merupakan kegiatan menafsirkan atau mendeskripsikan hasil pengukuran. Penilaian adalah proses untuk mengambil keputusan dengan menggunakan informasi yang diperoleh melalui pengukuran hasil belajar, baik yang menggunakan instrumen tes maupun non tes. Esensi dari pengukuran (*measurement*) adalah kuantifikasi atau penetapan angka tentang karakteristik atau keadaan individu menurut aturan-aturan tertentu (Dadan Rosana, 2013:35). Lebih jauh Pengukuran (*measurement*) adalah proses pemberian angka atau usaha memperoleh deskripsi numerik dari suatu tingkatan dimana seseorang mahasiswa calon guru IPA telah mencapai karakteristik tertentu. Penilaian merupakan

rangkaian kegiatan untuk memperoleh, menganalisis, dan menafsirkan data tentang proses dan hasil belajar mahasiswa calon guru IPA yang dilakukan secara sistematis dan berkesinambungan, sehingga menjadi informasi yang bermakna dalam pengambilan keputusan. Dengan demikian penilaian merupakan serangkaian kegiatan untuk memperoleh informasi atau data mengenai proses dan hasil belajar siswa.

Penilaian Pencapaian Kompetensi mahasiswa calon guru IPA dalam kurikulum 2013 mencakup kompetensi sikap, pengetahuan, dan keterampilan yang dilakukan secara berimbang sehingga dapat digunakan untuk menentukan posisi relatif setiap mahasiswa calon guru IPA terhadap standar yang telah ditetapkan. Cakupan penilaian merujuk pada ruang lingkup materi, kompetensi mata pelajaran/kompetensi muatan/kompetensi program, dan proses. Penilaian dilakukan dengan cara menganalisis dan menafsirkan data hasil pengukuran capaian kompetensi siswa yang dilakukan secara sistematis dan berkesinambungan sehingga menjadi informasi yang bermakna dalam pengambilan keputusan (Kemdikbud, 2016:5).

Kurikulum 2013 merupakan kurikulum berbasis kompetensi yang menekankan pembelajaran berbasis aktivitas yang bertujuan memfasilitasi siswa memperoleh sikap, pengetahuan, dan keterampilan. Hal ini berimplikasi pada penilaian yang harus meliputi sikap, pengetahuan, dan keterampilan baik selama proses (formatif) maupun pada akhir periode pembelajaran (sumatif). Berikut adalah beberapa hal yang perlu diperhatikan dalam melaksanakan penilaian:

1. Penilaian diarahkan untuk mengukur pencapaian Kompetensi Dasar (KD) pada Kompetensi Inti (KI-1, KI-2, KI-3, dan KI-4).
2. Penilaian menggunakan acuan kriteria, yaitu penilaian yang dilakukan dengan membandingkan capaian siswa dengan kriteria kompetensi yang ditetapkan. Hasil penilaian baik yang formatif maupun sumatif seorang siswa tidak dibandingkan dengan skor siswa lainnya namun dibandingkan dengan penguasaan kompetensi yang dipersyaratkan.
3. Penilaian dilakukan secara terencana dan berkelanjutan. Artinya semua indikator diukur, kemudian hasilnya dianalisis untuk menentukan kompetensi dasar (KD) yang telah dikuasai dan yang belum, serta untuk mengetahui kesulitan belajar siswa .
4. Hasil penilaian dianalisis untuk menentukan tindak lanjut, berupa program peningkatan kualitas pembelajaran, program remedial bagi siswa yang pencapaian kompetensinya di

bawah KBM/KKM, dan program pengayaan bagi siswa yang telah memenuhi KBM/KKM. Hasil penilaian juga digunakan sebagai umpan balik bagi orang tua/wali siswa dalam rangka meningkatkan kompetensi siswa.

## **B. Butir Soal Berstandar PISA**

PISA merupakan singkatan dari *Programme Internationale for Student Assesment* yang merupakan suatu bentuk evaluasi kemampuan dan pengetahuan yang dirancang untuk siswa usia 15 tahun . PISA sendiri merupakan proyek dari Organization for Economic Co-operation and Development (OECD) yang pertama kali diselenggarakan pada tahun 2000 untuk bidang membaca, matematika dan sains. Ide utama dari PISA adalah hasil dari sistem pendidikan harus diukur dengan kompetensi yang dimiliki oleh siswa dan konsep utamanya adalah literasi.

Dalam melakukan studi ini, setiap negara harus mengikuti prosedur operasi standar yang telah ditetapkan, seperti pelaksanaan uji coba dan survei, penggunaan tes dan angket, penentuan populasi dan sampel, pengelolaan dan analisis data, dan pengendalian mutu. Desain dan implementasi studi berada dalam tanggung jawab konsorsium internasional yang beranggotakan the Australian Council for Educational Research (ACER), the Netherlands National Institute for Educational Measurement (Citogroep), the National Institute for Educational Policy Research in Japan (NIER), dan WESTAT United States.

Tujuan PISA adalah untuk mengukur prestasi literasi membaca, matematika, dan sains bagi siswa usia 15 tahun. Bagi Indonesia, manfaat yang dapat diperoleh antara lain untuk mengetahui posisi prestasi literasi siswa di Indonesia bila dibandingkan dengan prestasi literasi siswa di negara lain dan faktor – faktor yang mempengaruhinya. Dasar penilaian prestasi literasi membaca, matematika, dan sains dalam PISA memuat pengetahuan yang terdapat dalam kurikulum dan pengetahuan yang bersifat lintas kurikulum. Masing-masing aspek literasi yang diukur adalah sebagai berikut:

1. Membaca: memahami, menggunakan, dan merefleksikan dalam bentuk tulisan.
2. Matematika: mengidentifikasi dan memahami serta menggunakan dasar-dasar matematika yang diperlukan seseorang dalam menghadapi kehidupan sehari-hari.

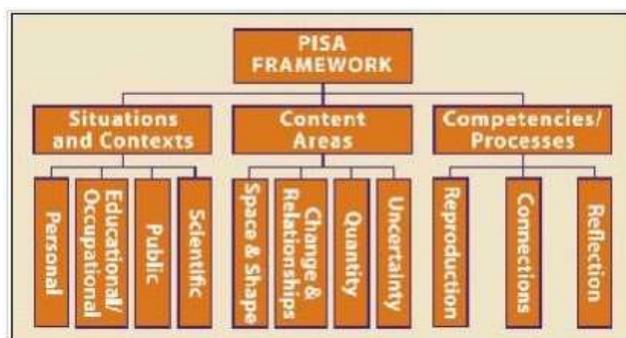
3. Sains: menggunakan pengetahuan dan mengidentifikasi masalah untuk memahami fakta-fakta dan membuat keputusan tentang alam serta perubahan yang terjadi padalingkungan.

Soal-soal PISA sangat menuntut kemampuan penalaran dan pemecahan masalah. Seorang siswa dikatakan mampu menyelesaikan masalah apabila ia dapat menerapkan pengetahuan yang telah diperoleh sebelumnya ke dalam situasi baru yang belum dikenal. Di dalam soal-soal PISA terdapat delapan ciri kemampuan kognitif yaitu :

1. *Thinking and reasoning*
2. *Argumentation*
3. *Communication*
4. *Modelling*
5. *Problem posing and solving*
6. *Representation, using symbolic*
7. *Formal and technical language and operations*
8. *Use of aids and tools*

Kedelapan kemampuan kognitif itu sangat sesuai dengan tujuan pembelajaran IPA yang terdapat pada kurikulum . Dari penjelasan di atas dapat disimpulkan bahwa soal-soal PISA bukan hanya menuntut kemampuan dalam penerapan konsep saja, tetapi lebih kepada bagaimana konsep itu dapat diterapkan dalam berbagai macam situasi, dan kemampuan siswa dalam bernalar dan berargumentasi tentang bagaimana soal itu dapat diselesaikan.

Framework PISA IPA berdasarkan tiga dimensi: (i) isi atau konten; (ii) proses yang perlu dilakukan siswa ketika mengamati suatu gejala, menghubungkan gejala itu dengan IPA, kemudian memecahkan masalah yang diamatinya itu; dan (iii) situasi dan konteks. Seperti terlihat pada gambar berikut ini:



Gambar 2.1 *PISA IPA Framework*

### **C. Kemampuan Berpikir Tingkat Tinggi**

Taksonomi Bloom dianggap merupakan dasar bagi berpikir tingkat tinggi (*Higher Order Thinking*). Pemikiran ini didasarkan bahwa beberapa jenis pembelajaran memerlukan proses kognisi yang lebih daripada yang lain, tetapi memiliki manfaat- manfaat lebih umum.

Berdasarkan Taksonomi Bloom yang telah direvisi (Anderson&Karthwoll, 2001), silahkan anda review kembali materi Unit 1, di sana terdapat tiga aspek dalam ranah kognitif yang menjadi bagian dari kemampuan berpikir tingkat tinggi atau higher-level thinking atau high order thinking (HOT). Ketiga aspek itu adalah aspek analisis-sintesis, aspek evaluasi dan aspek mencipta. Sedang tiga aspek lain dalam ranah yang sama, yaitu aspek mengingat, aspek memahami, dan aspek aplikasi, masuk dalam bagian intelektual berpikir tingkat rendah atau lower-order thinking.

Dalam Taksonomi Bloom, kemampuan melibatkan analisis, evaluasi dan mengkreasi dianggap berpikir tingkat tinggi (Pohl, 2000). Menurut Krathwohl (2002) dalam *A revision of Bloom's Taxonomy: an overview - Theory Into Practice* menyatakan bahwa indikator untuk mengukur kemampuan berpikir tingkat tinggi meliputi:

1. Menganalisis
  - a. Menganalisis informasi yang masuk dan membagi-bagi atau menstrukturkan informasi ke dalam bagian yang lebih kecil untuk mengenali pola atau hubungannya
  - b. Mampu mengenali serta membedakan faktor penyebab dan akibat dari sebuah skenario yang rumit.
  - c. Mengidentifikasi/merumuskan pertanyaan
2. Mengevaluasi
  - a. Memberikan penilaian terhadap solusi, gagasan, dan metodologi dengan menggunakan kriteria yang cocok atau standar yang ada untuk memastikan nilai efektivitas atau manfaatnya.
  - b. Membuat hipotesis, mengkritik dan melakukan pengujian
  - c. Menerima atau menolak suatu pernyataan berdasar kan kriteria yang telah ditetapkan
3. Mengkreasi
  - a. Membuat generalisasi suatu ide atau cara pandang terhadap sesuatu
  - b. Merancang suatu cara untuk menyelesaikan masalah

- c. Mengorganisasikan unsur-unsur atau bagian-bagian menjadi struktur baru yang belum pernah ada sebelumnya.

Stein dan Lane(1996) dikutip oleh Tony Thomson dalam Jurnal International Electronic Journal of Mathematics Education (2008) mendefinisikan berpikir tingkat tinggi adalah *the use of complex, nonalgorithmic thinking to solve a task in which there is not a predictable, well-rehearsed approach or pathway explicitly suggested by the task, task instruction, or a worked out example*. Menurut Stein berpikir tingkat tinggi menggunakan pemikiran yang kompleks, non algorithmic untuk menyelesaikan suatu tugas, ada yang tidak dapat diprediksi, menggunakan pendekatan yang berbeda dengan tugas yang telah ada dan berbeda dengan contoh.

Untuk mengajarkan keterampilan berpikir tingkat tinggi memang tidak mudah, contohnya kemampuan menarik kesimpulan, pertama-tama proses kognitif inferring harus dipecah ke dalam langkah-langkah sebagai berikut: (a) mengidentifikasi pertanyaan atau fokus kesimpulan yang akan dibuat, (b) mengidentifikasi fakta yang diketahui, (c) mengidentifikasi pengetahuan yang relevan yang telah diketahui sebelumnya, dan (d) membuat perumusan prediksi hasil akhir. Karena itulah, kita perlu memperhatikan prinsip-prinsip dalam pembelajaran keterampilan berpikir di kelas pembelajaran, diantaranya adalah sebagai berikut:

1. keterampilan berpikir tidak otomatis dimiliki siswa
2. keterampilan berpikir bukan merupakan hasil langsung dari pembelajaran suatu bidang studi
3. Pada kenyataannya siswa jarang melakukan transfer sendiri keterampilan berpikir ini, sehingga perlu adanya latihan terbimbing
4. Pembelajaran keterampilan berpikir memerlukan model pembelajaran yang berpusat kepada siswa (*student-centered*).

Selain beberapa prinsip di atas, satu hal yang tidak kalah pentingnya dalam melatih keterampilan berpikir adalah perlunya latihan-latihan yang intensif. Seperti halnya keterampilan yang lain, dalam keterampilan berpikir siswa perlu mengulang untuk melatihnya walaupun sebenarnya keterampilan ini sudah menjadi bagian dari cara berpikirnya. Latihan rutin yang dilakukan siswa akan berdampak pada efisiensi dan otomatisasi keterampilan berpikir yang telah dimiliki siswa.

Selain itu Levie dan Levie dalam Azhar Arzad (2009: 9) yang membaca kembali hasil-hasil penelitian tentang belajar melalui stimulus gambar dan stimulus kata atau visual dan verbal menyimpulkan bahwa stimulus visual membuahkan hasil belajar yang lebih baik untuk tugas-tugas seperti mengingat, mengenali, mengingat kembali, dan menghubungkan fakta dan konsep. Sedangkan stimulus verbal memberikan hasil belajar yang lebih baik apabila pembelajaran itu melibatkan ingatan yang berurut-urutan (sekuensial). Karena itulah maka dalam dunia pendidikan ada 3 model seorang siswa dalam menerima suatu pelajaran, *I hear and I forget* (saya mendengar dan saya akan lupa), *I see and I remember* (saya melihat dan saya akan ingat), *I do and I understand* (saya melakukan dan saya akan mengerti).

Untuk mengembangkan *Higher Level Questions* maka dalam pembuatan soal-soal ulangan, guru perlu memperhatikan beberapa hal berikut ini:

1. Soal hendaknya menggunakan stimulus, stimulus yang baik hendaknya menyajikan informasi yang jelas, padat, mengandung konsep/gagasan inti permasalahan, dan benar secara fakta.
2. Soal yang dikembangkan harus sesuai dengan kondisi pembelajaran yang dilaksanakan di dalam kelas maupun di luar kelas yang berhubungan dengan kehidupan sehari-hari
3. Soal mengukur keterampilan berpikir kritis
4. Soal mengukur keterampilan pemecahan masalah

## **BAB III**

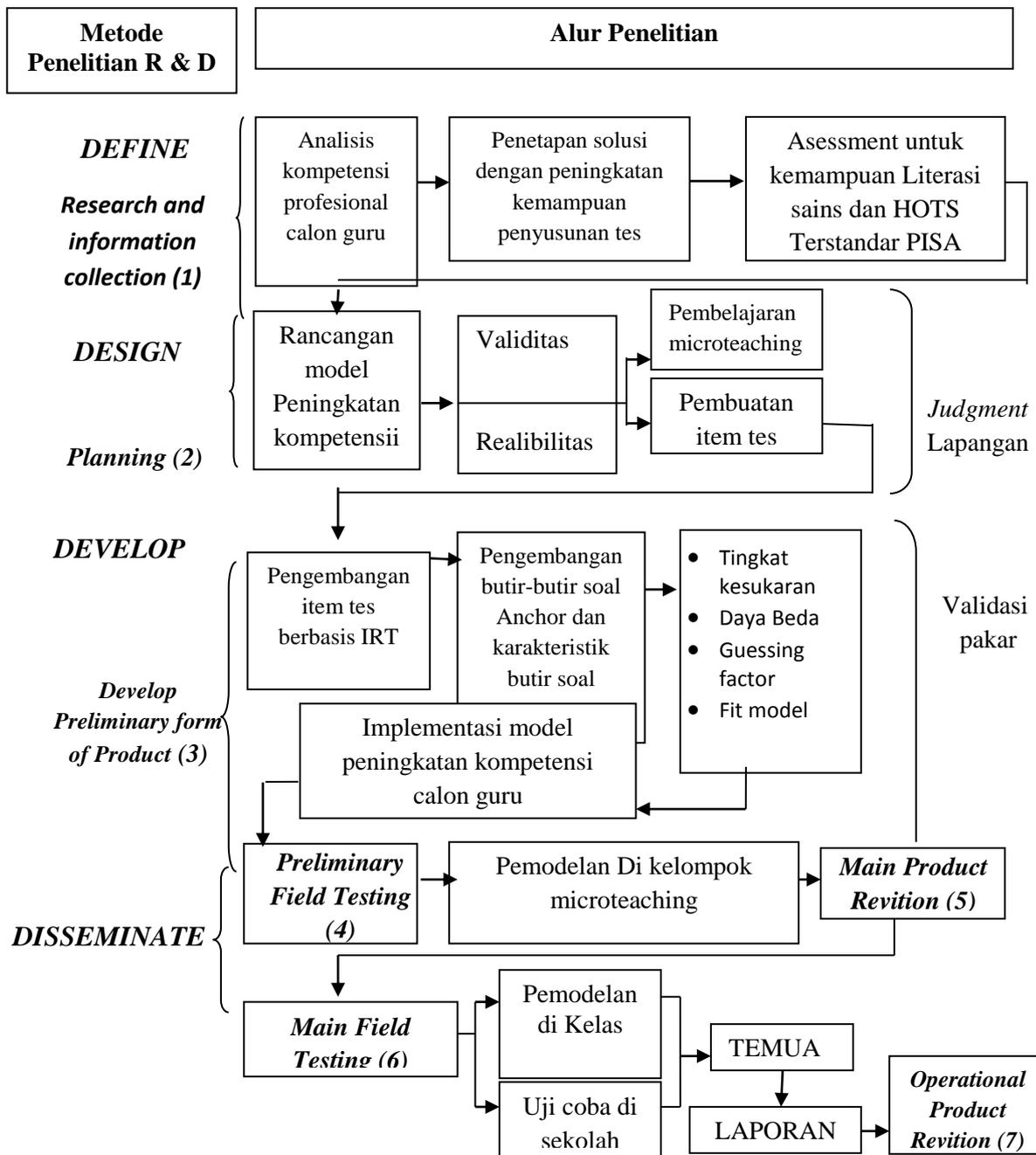
### **METODE PENELITIAN**

#### **A. Paradigma dan Disain Penelitian**

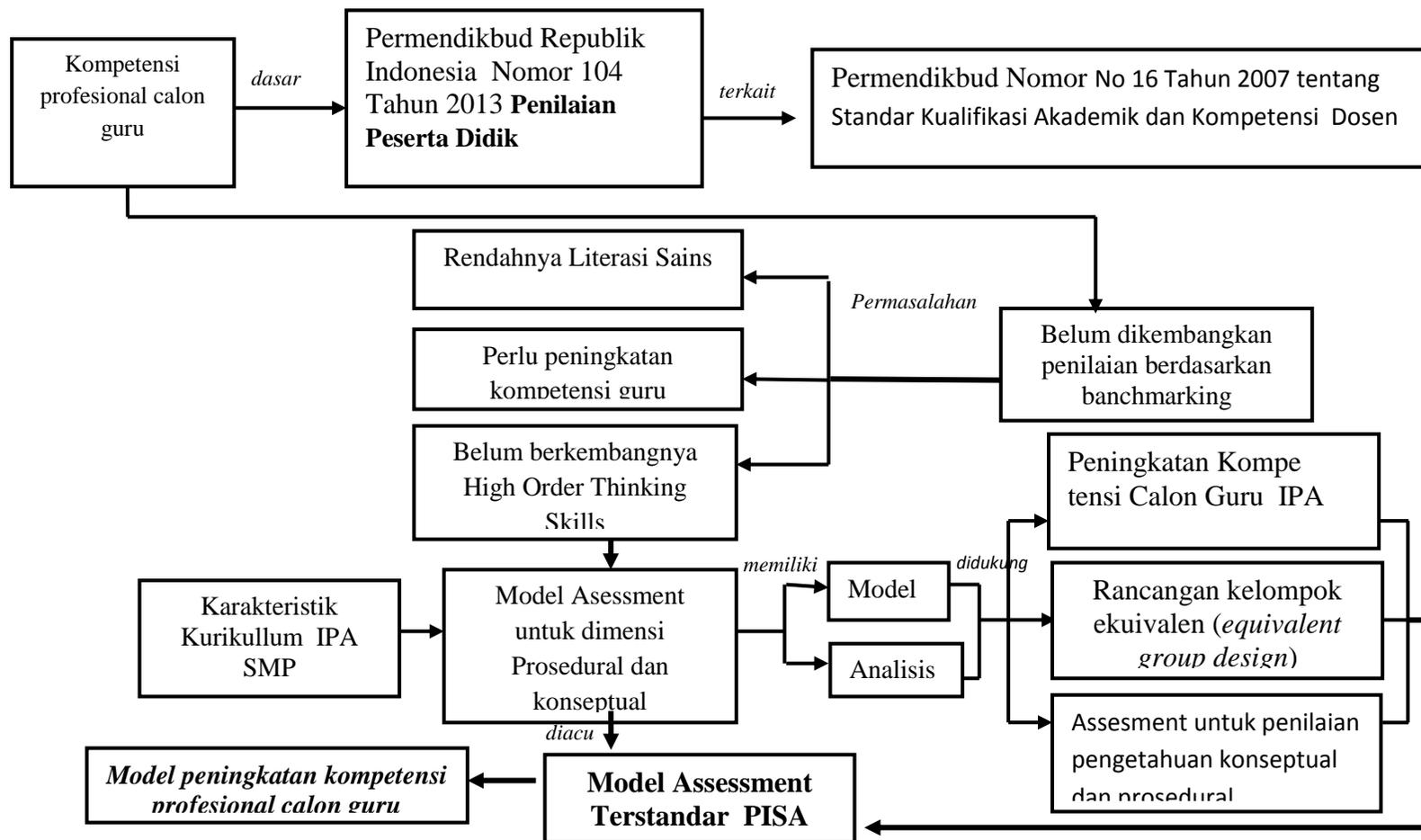
Guru adalah tokoh sentral dalam dunia pendidikan. Seorang guru bertanggung jawab memberikan pertolongan pada mahasiswa calon guru IPA untuk mencapai tingkat kedewasaan agar mampu berdiri sendiri. Mandiri dalam memenuhi tugasnya sebagai makhluk sosial dan sebagai individu. guru harus memiliki kualifikasi akademik dan kompetensi sebagai agen pembelajaran, sehat jasmani dan rohani, serta memiliki kemampuan untuk mewujudkan tujuan pendidikan nasional. Kompetensi sebagai agen pembelajaran pada jenjang pendidikan dasar dan menengah serta pendidikan anak usia dini meliputi; (1) Kompetensi pedagogik, (2) Kompetensi kepribadian, (3) Kompetensi professional, dan (4) Kompetensi sosial, sebagaimana tertuang dalam Peraturan Menteri Pendidikan Nasional Republik Indonesia No 16 Tahun 2007 tentang Standar Kualifikasi Akademik dan Kompetensi Guru.

Penilaian hasil belajar oleh pendidik dilakukan secara berkesinambungan untuk memantau proses, kemajuan, dan perbaikan hasil dalam bentuk ulangan harian, ujian tengah semester, ujian akhir semester, dan ujian kenaikan kelas. Penilaian hasil belajar oleh pendidik digunakan untuk menilai pencapaian kompetensi mahasiswa ; bahan penyusunan laporan hasil belajar; dan memperbaiki proses pembelajaran. Hal ini di atur dalam Permendikbud nomor 104 tahun 2014. Perbedaan yang mendasar dengan permendikbud-permendikbud sebelumnya yang berisi pedoman penilaian, yakni Permendikbud nomor 66 tahun 2013 tentang standar penilaian, Permendikbud nomor 81A tahun 2013 tentang implementasi kurikulum 2013 dan paling akhir Permendikbud no 59 tahun 2014 tentang kurikulum 2013 SMA/MA yang disahkan pada tanggal 2 Juli 2014, adalah pada rentangan nilai dan penulisan angka pada rapor.

Permasalahan yang kemudian muncul berkaitan dengan hal ini adalah masih lemahnya kemampuan calon guru dalam mengembangkan sistem penilaian yang sesuai dengan tuntutan kurikulum. Karena itu diperlukan upaya peningkatan kompetensi akademik calon guru dalam pengembangan assessment penilaian yang sesuai dengan benchmarking survei internasional seperti PISA agar dapat meningkatkan daya saing global lulusan. Untuk itu, desain penelitian ditunjukkan oleh gambar 4., dan paradigma penelitian seperti ditunjukkan di gambar 5.



Gambar 4. Desain Penelitian



Gambar 5. Paradigma Penelitian

Disain penelitian menggunakan metode penelitian *Research and Development* dan alur penelitian yang dijelaskan pada Gambar 2. Fase *define* atau *research and information collection* (Borg dan Gall, 1983: 776) merupakan fase penelitian dan pengumpulan data awal berupa studi literatur, analisis kebutuhan dan studi lapangan. Fase *design* atau *planning* (Borg dan Gall, 1983: 777) merupakan rancangan produk yang akan dihasilkan, meliputi tujuan penggunaan produk, pengguna produk dan deskripsi komponen-komponen produk. Fase *develop* atau *develop preliminary form of product* (Borg dan Gall, 1983: 781) merupakan pengembangan produk awal. Fase *Disseminate* ada empat langkah pengembangan, yaitu *preliminary field testing* (Borg dan Gall, 1983: 782) yang merupakan ujicoba lapangan awal, *main product revision* (Borg dan Gall, 1983: 782) atau revisi hasil ujicoba, *main field testing* (Borg dan Gall, 1983: 783) atau ujicoba lapangan utama serta *operational product revision* (Borg dan Gall, 1983: 784) atau penyempurnaan produk hasil ujicoba lapangan.

## **1. Prosedur Penelitian**

Mengacu pada desain penelitian tersebut, prosedur penelitian dilaksanakan melalui tahapan-tahapan dalam *research and development* (R & D). Tujuan utama R & D untuk mengembangkan dan memvalidasi suatu program atau model yang akan digunakan agar tujuan menjadi efektif dan siap untuk diimplementasikan. Tahapan-tahapan R & D diformulasikan menjadi model 4-D (*Four-D Models*) (Thiagarajan, 1975: 5) dan disesuaikan Borg dan Gall (1983: 775) yaitu:

### **a. Define (D-1)/Research and Information Collection**

#### **1) Analisis Teori/ Studi Literatur**

Tahap ini menganalisis secara teori model model untuk penyetaraan kualitas tes, meliputi:

- Ada tiga jenis rancangan penyetaraan tes yang dapat digunakan, yaitu rancangan kelompok tunggal (*single group design*), rancangan kelompok ekuivalen (*equivalent group design*), dan rancangan dengan butir tes jangkar. Dalam rancangan kelompok tunggal digunakan satu kelompok peserta yang merespons dua perangkat tes (X dan Y). Parameter butir dari kedua perangkat tes diestimasi secara terpisah dengan mengkalibrasi parameter kemampuan peserta tes atau parameter butir. Berdasarkan rancangan ini, dengan mengkalibrasi parameter kemampuan peserta, maka parameter butir dari perangkat tes X dan Y sudah berada pada skala yang sama.
- Idealnya untuk menyetarakan skor dari beberapa perangkat tes, maka perangkat tes- perangkat tes tersebut diberikan pada responden yang sama. Dengan membandingkan kemampuan peserta tes dari dua/lebih perangkat tes maka penyetaraan dua perangkat tes dapat dilakukan. Kenyataan di lapangan,

rancangan ini sulit dilakukan karena adanya faktor kelelahan, belajar, dan adanya faktor latihan untuk tes kedua atau berikutnya. Selain itu, akan terdapat kesulitan dalam hal merencanakan waktu yang cukup bagi responden untuk mengikuti tes lebih dari satu kali (Miyatun dan Mardapi, 2000).

- Pada rancangan kelompok ekuivalen digunakan dua kelompok peserta ekuivalen ( $K_1$  dan  $K_2$ ) dan dua perangkat tes (X dan Y). Kelompok peserta  $K_1$  mengerjakan perangkat tes X dan kelompok peserta  $K_2$  mengerjakan perangkat tes Y. Mengingat kelompok  $K_1$  dan  $K_2$  adalah ekuivalen, maka kedua kelompok dianggap tunggal. Penentuan konstanta konversi berikutnya seperti rancangan kelompok tunggal. Keuntungan rancangan ini dapat menghindari efek negatif yang disebabkan karena latihan dan kelelahan peserta tes, sedangkan kekurangannya ada kemungkinan bias yang disebabkan karena tidak mudah untuk membuat distribusi kemampuan dua kelompok peserta tes yang benar-benar ekuivalen (Sukirno, 2007: 310).
- Pada rancangan tes jangkar (*anchor test design*) biasanya digunakan jika masalah keamanan tes menjadi salah satu pertimbangan penting dan memungkinkan untuk menyelenggarakan beberapa tes dalam satu waktu. Pada desain ini masing-masing perangkat tes mempunyai beberapa butir yang sama (*common item*) dan masing-masing kelompok mengerjakan perangkat tes yang berbeda. Pada desain ini terdapat dua variasi, yaitu (Chong dan Osborn, 2005): 1) jika *common item* diperhitungkan dalam pemberian skor disebut *internal common item*; dan 2) jika *common item* tidak diperhitungkan dalam pemberian skor disebut *external common item*.

## 2) Analisis Tugas/Needs Assessment

Tahap ini mengidentifikasi keterampilan-keterampilan proses utama dan menganalisisnya dalam set-set sub-sub keterampilan yang diperlukan. Analisis ini menjamin kekomprehensifan tugas-tugas dalam mencapai kompetensi literasi sains dan High Order Thinking Skills sesuai dengan tuntutan tes PISA. Aktivitas penelitian meliputi pengembangan rancangan tes, memuat tujuan penilaian yang akan dilakukan, tempo (waktu yang ditempuh) untuk pelaksanaan pengujian, pesan utama pembelajaran (sasaran pembelajaran dan garis besar topik materi uji), indikator butir soal (ciri-ciri penguasaan materi uji dan pencapaian sasaran pembelajaran), serta jumlah dan bentuk butir soal (per-indikator, per topik, dan keseluruhan tes). Sebaran butir soal dalam tes seharusnya memperhatikan keseimbangan tuntutan penguasaan sesuai dengan pesan kurikulum, sehingga memberi nuansa keterwakilan topik bahasan..

## 3) Analisis Konsep/Needs Assessment

Tahap ini mengidentifikasi konsep-konsep utama yang diajarkan, menyusunnya dalam hirarki dan menguraikan dalam tema-tema utama. Adapun langkah-langkah konstruksi tes yang ditempuh adalah sebagai berikut: 1) menetapkan tujuan tes; 2) analisis berdasarkan benchmarking survei internasional; 3) analisis buku pelajaran dan sumber materi belajar lainnya; 4) membuat kisi-kisi; 5) penulisan tujuan instruksional khusus; 6) penulisan soal; 7) telaah soal (*face validity*); 8) reproduksi tes terbatas; 9) uji coba tes; 10) analisis hasil uji coba; 11) revisi soal, dan 12) merakit soal menjadi tes. Metode penyetaraan menurut teori responsi butir berfungsi untuk menentukan konstanta konversi. Hal ini mengingat bahwa penyetaraan antara dua perangkat tes atau lebih dapat dilakukan jika konstanta konversi telah diketahui. Nilai konversi yang dihasilkan kemudian di substitusi dalam persamaan skala pada rancangan penyetaraan yang digunakan. Ada beberapa metode penyetaraan tes yang dapat digunakan dan faktor-faktor yang mempengaruhi keakuratan metode penyetaraan tes. Dalam teori responsi butir terdapat empat metode penyetaraan tes, yaitu: regresi, rerata sigma, rerata dan sigma tegar, dan kurva karakteristik

#### **b. Develop (D-3)/Develop Preliminary form of Product**

Tahap ini memperoleh persetujuan untuk meningkatkan kualitas assessment yang digunakan dalam peningkatan kompetensi calon guru dalam menyusun assessment terstandar PISA. Sejumlah ahli diminta untuk mengevaluasi program penyetaraan berbasis IRT yang sudah dirancang, meliputi empat metode penyetaraan tes, yaitu: regresi, rerata sigma, rerata dan sigma tegar, dan kurva karakteristik kemudian berbasis *feedback* para ahli model yang telah dihasilkan dimodifikasi/direvisi untuk menjadi lebih tepat, efektif, dan bermanfaat serta teknik kualitasnya tinggi.

Jenis penelitian yang digunakan pada saat uji coba adalah kuasi eksperimen. Kuasi eksperimen digunakan karena peneliti tidak dapat mengendalikan secara utuh variabel (variabel internal dan eksternal) peserta didik sehingga penelitian ini menggunakan jenis kuasi eksperimen. Menurut Sugiyono (2013:77), kuasi eksperimen digunakan karena terdapat variabel-variabel luar yang dapat mempengaruhi pelaksanaan eksperimen, sehingga peneliti tidak dapat sepenuhnya mengontrol kelas kontrol dalam eksperimen.

Pada penelitian ini, peneliti ingin mengetahui perbedaan hasil workshop dan pembelajaran tentang pembuatan soal test HOTS berbasis PISA dan pembuatan soal HOTS menggunakan Buku Evaluasi Pembelajaran Sains yang telah ditulis oleh ketua tim peneliti dan diterbitkan oleh UNY Press. Sehingga dalam penelitian ini, peneliti membuat dua kelompok kelas, yaitu kelas eksperimen dan kelas kontrol. Penelitian ini menggunakan desain penelitian *nonequivalent pretest-posttest control group design* untuk mengetahui pengaruh perlakuan terhadap kedua subjek peneliti.

Kelas eksperimen diberikan perlakuan pembelajaran dengan perangkat pembuatan soal test HOTS berbasis PISA, sedangkan kelas kontrol diberikan perlakuan pembelajaran sesuai dengan pembuatan soal HOTS menggunakan Buku Evaluasi Pembelajaran Sains. Berikut adalah gambar desain *pretest* dan *posttest*.

E	O <sub>1</sub>	X <sub>1</sub>	O <sub>2</sub>
K	O <sub>3</sub>	X <sub>2</sub>	O <sub>4</sub>

Gambar 6. Desain *pretest* dan *posttest*

**Keterangan:**

- E = kelas eksperimen dengan pembuatan soal test HOTS berbasis PISA
- K = kelas kontrol dengan pembuatan soal HOTS menggunakan Buku Evaluasi Pembelajaran Sains
- X1 = perlakuan kelas eksperimen dengan pembuatan soal test HOTS berbasis PISA
- X2 = perlakuan kelas kontrol dengan pembuatan soal HOTS menggunakan Buku Evaluasi Pembelajaran Sains
- O<sub>1</sub> = hasil *pretest* pada kelas eksperimen
- O<sub>2</sub> = hasil *posttest* pada kelas eksperimen
- O<sub>3</sub> = hasil *pretest* pada kelas kontrol
- O<sub>4</sub> = hasil *posttest* pada kelas kontrol

Berdasarkan desain tersebut, kedua kelompok kelas (kelas eksperimen dan kelas kontrol) diberikan diberikan *pretest* dan *posttest*. Pada tes awal (*pretest*), kedua kelas ini diberikan tes yang sama. Namun yang menjadi pembeda adalah bahwa kelas eksperimen diberi *treatment* atau perlakuan tertentu yaitu diberikan pembelajaran dengan Pembuatan soal test HOTS berbasis PISA , sedangkan kelas kontrol diberikan *treatment* dengan pembuatan soal HOTS menggunakan Buku Evaluasi Pembelajaran Sains. Setelah diberikan perlakuan yang berbeda, peneliti melakukan tes akhir (*posttest*) untuk mengetahui pengaruh Pembuatan soal test HOTS berbasis PISA terhadap kemampuan mahasiswa calon guru IPA.

### **c. Dessiminate (D-4)**

Pengujian perangkat hasil pengembangan (*Preliminary Field Testing*) dilakukan di bekerjasama dengan Dinas Pendidikan dan Kebudayaan dan Perguruan Tinggi Negeri melalui tahap ujicoba dan sudah direvisi (*Main Product Revision*), maka perangkat-perangkat diterapkan dalam pembuatan tes literasi sains dan high order thinking skillsterstandar PISA (*Main Field Testing*), kemudian diobservasi segala variabel yang menjadi fokus/tujuan pengembangan.

## **2. Teknik Pengumpulan Data**

Teknik dokumentasi digunakan untuk mengumpulkan data-data yang berkaitan dengan penelitian yang akan dilakukan, seperti kisi-kisi soal, perangkat soal, kunci jawaban, serta respon mahasiswa yang diambil dari kelas yang menjadi sampel penelitian. Respon dalam penelitian ini didapatkan setelah mahasiswa mengerjakan satu set instrumen Tes literasi sains dan high order thinking skillsterstandar PISA yang memuat butir-butir soal pilihan ganda dengan empat alternatif jawaban. Instrumen tersebut dibuat oleh tim peneliti.

## **3. Validitas Instrumen**

Instrumen Tes literasi sains dan high order thinking skills terstandar PISA yang telah disusun selanjutnya divalidasi agar instrumen tersebut benar-benar menggambarkan apa yang hendak diukur (Aiken 1980). Butir-butir yang telah dikembangkan dibuat berdasarkan kisi-kisi yang terdistribusi proporsional sesuai uraian materi yang tercantum dalam kurikulum, agar secara validitas isi (konten) atau validitas teoritis memenuhi syarat. Koefisien validitas isi dalam penelitian ini diolah dari skor yang diberikan oleh *expert judgement*. Hasil *judgment* selanjutnya diolah menggunakan formula Aiken (Aiken, 1985):

$$V = \frac{\sum s}{[n(c - 1)]}$$

dimana :  $s = r - lo$

lo = angka penilaian validitas yang terendah

c = angka penilaian validitas tertinggi

r = angka yang diberikan oleh penilai

Digunakan empat *ratings category* yaitu “tidak relevan, kurang relevan, relevan, dan sangat relevan” maka sebaiknya indeks aiken sebesar 0,87 ( $\alpha=0,05$ ) atau 0,93 ( $\alpha=0,01$ ) (Aiken, 1985). Namun menurut Sireci & Geisinger (1995: 246-247), koefisien validitas sekitar 0,7 masih dapat diterima dan dianggap memuaskan. Berdasarkan analisis menggunakan formula aiken didapatkan indeks rata-rata validitas isi sebesar 0,945. Validitas isi bagi butir-butir instrumen Tes literasi sains dan high order thinking skills terstandar PISA bergerak dari angka 0,867 sampai 1. Dengan demikian, dapat disimpulkan bahwa butir-butir dalam instrumen tes yang digunakan telah valid.

#### 4. Teknik Analisis Data

##### a. Analisis Karakteristik Instrumen Tes IPA

Analisis data dimulai dengan mendeskripsikan kelayakan (karakteristik) instrumen Tes literasi sains dan high order thinking skills terstandar PISA menggunakan pendekatan klasik dan pendekatan modern. Untuk pendekatan klasik digunakan program ITEMAN Versi 3.0. Berdasarkan pendekatan tersebut, butir dikatakan memenuhi syarat sebagai butir yang baik jika: (a) Tingkat kesukaran butir berada dalam rentang 0,3-0,7 (Allen & Yen, 1979; Zulaiha, 2008); (b) Daya beda butir lebih besar dari 0,3 (Reynolds, Livingston, & Wilson, 2010; Kartowagiran, 2012); dan (c) setidaknya pengecoh dipilih oleh 2% peserta tes (Brown dalam Fernandes, 1984). Selain melihat karakteristik butir, juga akan dilihat koefisien reliabilitasnya. Instrumen dianggap reliabel jika koefisien reliabilitasnya minimal 0,7 (Linn, 1989; Mardapi, 2014).

Analisis berdasarkan pendekatan modern dilakukan menggunakan *model Rasch* dengan bantuan program WINSTEPS. Program WINSTEPS dipilih karena memiliki kelebihan, yaitu: (a) dapat menganalisis data dikotomus dan politomus beserta kombinasinya; (b) tersedianya hasil analisis baik berdasarkan teori tes klasik maupun teori respon modern; (c) hasil analisis teori respon modern didasarkan pada model *likelihood maximum* menggunakan model logistik satu parameter; (d) mampu melakukan prediksi terhadap data yang hilang (*missing data*); dan (e) kalibrasi yang dilakukan dalam permodelan Rasch secara sekaligus dalam tiga hal yakni skala pengukuran, orang (*person*), dan butir soal (*item*) (Linacre, 2012; Sumintono & Widhiarso, 2015).

Analisis berdasarkan pendekatan modern mula-mula dilakukan dengan menguji asumsi unidimensi, asumsi independensi lokal, asumsi invariansi parameter (invariansi parameter butir dan invariansi parameter kemampuan), dan kecocokan model (*item fit* dan *person fit*). Pembuktian asumsi unidimensi akan dilakukan melalui analisis faktor, yaitu dengan melihat kontribusi faktor pertama terhadap varians tes serta banyaknya titik (plot) sebelum siku pada pada *scree plot*. Namun, sebelum melakukan analisis faktor, terlebih dahulu dilakukan pengujian kelayakan sampel analisis dengan menggunakan uji KMO-MSA dan uji Bartlet's dengan kriteria *Kaiser-Meyer Olkin Measure Sampling Adequacy* (KMO-MSA) lebih dari 0,5 dan signifikansi uji Bartlet's kurang dari 0,05.

Asumsi independensi akan dibuktikan menggunakan matriks varians-kovarians antarkemampuan responden. Jika nilai kovarians antarinterval kemampuan, kecil atau mendekati nol, maka asumsi independensi lokal telah terpenuhi (Hambleton, Swaminathan & Rogers, 1991: 56). Sementara itu, asumsi invariansi parameter akan dibuktikan menggunakan *scatter plot* dan koefisien korelasinya. Jika korelasinya positif dan tinggi, maka asumsi invariansi parameter telah terpenuhi (Retnawati, 2014: 8).

Analisis karakteristik instrumen berdasarkan pendekatan ini ditentukan sebagai berikut: (a) untuk setiap butir soal, sebaran respon peserta tes cocok dengan modelnya; (b) untuk setiap butir soal estimasi tingkat kesukaran butir berkisar antara  $-2 \logit \leq b_i \leq 2 \logit$  (Hambleton & Swaminathan, 1985); dan (c) tes akan memberikan informasi yang baik jika  $TIF \geq 10$  (Hambleton dalam Wiberg, 2004). Untuk mengetahui kecocokan butir (*item fit*) maupun kecocokan pola respon peserta tes (*person fit*) berdasarkan model Rasch, maka dapat dilihat dari nilai *Outfit Mean Square* (MNSQ). Butir tes dikatakan fit model jika memiliki nilai MNSQ berkisar dari 0,5 sampai 1,5 (Linacre, 2002).

Tabel 1. Rentang Nilai *Outfit Mean Square* (MNSQ) (Linacre, 2002)

Nilai	Implikasi bagi Pengukuran
> 2,0	Merusak sistem pengukuran
1,5 – 2,0	Tidak mempunyai makna bagi pengukuran
0,5 – 1,5	Bermanfaat bagi pengukuran
< 0,5	Tidak bermanfaat bagi pengukuran, tetapi tidak merusak

#### **b. Kecocokan Tingkat Kesukaran yang Telah Ditetapkan oleh Dosen dengan Tingkat Kesukaran Hasil Empirik**

Kecocokan tingkat kesukaran yang telah ditetapkan oleh dosen terhadap kecocokan tingkat kesukaran hasil empirik akan dilihat menggunakan uji *chi square*. Uji ini digunakan untuk menentukan perbedaan frekuensi observasi ( $O_i$ ) dengan frekuensi ekspektasi atau frekuensi harapan ( $E_i$ ) suatu kategori tertentu yang dihasilkan (West, 2008). Adapun perhitungannya, digunakan tabel kontingensi 3 x 3 (tiga baris dan tiga kolom). Secara terperinci, uji *chi square* akan digunakan untuk melihat:

a) Kecocokan yang telah ditetapkan oleh dosen (tim peneliti) dengan tingkat kesukaran hasil empirik yang dianalisis menggunakan pendekatan klasik (program ITEMAN).

b) Kecocokan yang telah ditetapkan oleh dosen (tim peneliti) dengan tingkat kesukaran hasil empirik yang dianalisis menggunakan pendekatan modern (program WINSTEPS, berbasis Rasch Model).

Hasil analisis disimpulkan cocok apabila *Asymp sig* lebih kecil dari taraf signifikansi yang ditetapkan ( $\alpha = 0,05$ ).

**c. Estimasi Skor Tes literasi sains dan high order thinking skills terstandar PISA**

Skor IPA mahasiswa dapat diestimasi setelah parameter kemampuan dan parameter butir (tingkat kesukaran) diketahui. Adapun persamaan untuk mengestimasi Skor Tes literasi sains dan high order thinking skillsterstandar PISA berdasarkan kedua model penskoran tersebut:

**a) Model Penskoran Pembobotan (Lord, 1980: 73)**

$$y = \sum_{i=1}^n w_i P_i (\theta)$$

dengan :  $y$  = skor berdasarkan model penskoran pembobotan

$w_i$  = bobot item ke- $i$

$P_i(\theta)$  = peluang mahasiswa dengan kemampuan  $\theta$  dalam menjawab item  $i$   
dengan benar

**b) Model Penskoran *Number-Right* (Lord, 1980: 230)**

$$\xi_a = \Sigma^{(a)} P_g(\theta)$$

dengan :  $\xi_a$  = number right score mahasiswa  $a$

$\Sigma^{(a)}$  = banyaknya item yang dijawab oleh mahasiswa  $a$

$P_g(\theta)$  = peluang mahasiswa  $a$  dengan kemampuan  $\theta$  dalam menjawab item  $g$   
dengan benar

Skor IPA yang telah diestimasi selanjutnya dilihat karakteristik distribusinya berdasarkan rerata, simpangan baku, skewness, dan kurtosis.

#### **d. Hasil Penerapan Model Penskoran terhadap Skor Tes literasi sains dan high order thinking skills terstandar PISA**

Skor yang diestimasi berdasarkan model penskoran *WD* maupun *NR* selanjutnya akan dilihat ada tidaknya perubahan peringkat skor mahasiswa akibat penerapan model penskoran yang berbeda. Hal tersebut dapat diketahui dengan cara melihat koefisien korelasi intraklasnya (*Intraclass Correlation Coefficient*).

#### **5. Subyek Penelitian**

Dalam langkah ujicoba lapangan awal dan ujicoba terbatas model peningkatan kompetensi akademik calon guru dalam pengembangan assessment penilaian tes literasi sains dan high order thinking skills terstandar PISA pada pembelajaran IPA secara terpadu ini, digunakan subyek penelitian adalah mahasiswa semester 6 Program Studi Pendidikan IPA tahun 2018 yang sedang mengikuti mata kuliah Evaluasi Pembelajaran IPA dengan mengambil sampel beberapa kelompok terdiri dari 10 mahasiswa yang setiap kelompok diampu oleh dua dosen.

## BAB IV

### HASIL PENELITIAN DAN PEMBAHASAN

#### A. Hasil Penelitian

Penelitian ini merupakan penelitian kuasi eksperimen dengan desain penelitian *pretest-posttest nonequivalent control group design* yang dilaksanakan di dua kelas, yaitu kelas eksperimen dan kelas kontrol. Pembelajaran IPA di kelas eksperimen menggunakan workshop dan pendampingan pembuatan soal HOTS berbasis benchmarking survey internasional (PISA) sedangkan pembelajaran di kelas kontrol menggunakan materi HOTS berdasarkan buku Evaluasi Pembelajaran Sains di Jurusan Pendidikan IPA FMIPA UNY. Data yang diperoleh peneliti dalam penelitian ini adalah profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) dan hasil tes HOTS berstandar PISA mahasiswa calon guru IPA. Data tersebut merupakan data dari sampel sebanyak 37 mahasiswa kelas A dan 28 dari kelas I yang mengambil mata kuliah Evaluasi Pembelajaran Sains tahun 2018 di Jurusan Pendidikan IPA FMIPA UNY.

#### 1. Validitas dan Reliabilitas Soal *Pretest-Posttes*

Soal terlebih dahulu divalidasi isi dan konstruk oleh 2 orang dosen dikelompok bidang keahlian Evaluasi Pembelajaran di Prodi Pendidikan IPA FMIPA UNY sebelum diujicobakan di kelas C yang mengambil mata kuliah Evaluasi Pembelajaran tahun 2018 di Jurusan Pendidikan IPA FMIPA UNY sebagai uji validitas empiris. Hasil validasi isi dan konstruk menghasilkan 20 butir soal yang kemudian diujicobakan di kelas C sebagai uji empiris. Hasil dari pengujian di kelas C dianalisis content dan construct untuk mengetahui validitas dan realibilitas instrumen soal yang kemudian menjadi soal *pretest* dan *posttest* pada kelas eksperimen (kelas A) dan kelas control (kelas D).

Soal yang valid dan reliabel setelah dilakukan uji di kelas C sebanyak 20 soal juga sehingga masing-masing instrumen soal *pretest* dan soal *posttest* terdiri dari 20 butir soal.

Butir soal *pretest* tidak persis sama dengan butir soal *posttest*, perubahan dilakukan dengan mengacu pada kisi-kisi yang memiliki tingkat kesulitan butir soal yang sama. .

## 2. Data Hasil Observasi Keterlaksanaan Pembelajaran

Keterlaksanaan pembelajaran dengan penggunaan Pelatihan pengembangan soal HOTS berstandar PISA dinilai dengan lembar observasi keterlaksanaan pembelajaran yang telah divalidasi. Lembar observasi keterlaksanaan pembelajaran terdiri dari kegiatan pendahuluan, persiapan, penjelajahan, pertanyaan, penyelidikan, refleksi, dan penutup. Data hasil observasi keterlaksanaan pembelajaran dengan penggunaan Pelatihan pengembangan soal HOTS berstandar PISA dapat dilihat pada Tabel 2.

Tabel 2. Data Keterlaksanaan Pembelajaran IPA dengan Penggunaan Pelatihan pengembangan soal HOTS berstandar PISA

Pertemuan Ke-	Presentase Keterlaksanaan (%)		Kategori
	Kegiatan		
	Dosen	Mahasiswa calon guru IPA	
1	92,75	88,25	Sangat baik
2	94,25	89,75	Sangat baik
3	96,00	96,50	Sangat baik

Berdasarkan Tabel 2, keterlaksanaan pembelajaran dengan menggunakan pelatihan pengembangan soal HOTS berstandar PISA belum sepenuhnya dilakukan dengan sempurna. Namun demikian, presentase kegiatan dosen pada setiap pertemuan semakin meningkat, sehingga menunjukkan kemajuan yang signifikan. Demikian pula dengan perkembangan keterlaksanaan kegiatan mahasiswa calon guru IPA, sudah terlihat meningkat dari pertemuan satu ke pertemuan setelahnya. Walaupun belum mencapai sempurna, keterlaksanaan pembelajaran dengan pelatihan pengembangan soal HOTS berstandar PISA baik pada kegiatan dosen maupun mahasiswa calon guru IPA memperoleh kategori sangat baik.

## 3. Data Hasil Uji Kompetensi Dalam Pengembangan Soal HOTS Berbasis

### **Benchmarking Survey Internasional (PISA) Mahasiswa Calon Guru IPA**

Data profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) diperoleh dari data kemampuan awal dan kemampuan akhir mahasiswa calon guru IPA. Data kemampuan awal mahasiswa calon guru IPA diperoleh dari pemberian pretest kepada mahasiswa calon guru IPA baik di kelas kontrol maupun kelas eksperimen, sebelum dilaksanakan pembelajaran pada kelas eksperimen dengan pelatihan penyusunan soal HOTS berbasis PISA dan kelas control dengan menggunakan buku Evaluasi Pembelajaran Sains. Data kemampuan akhir mahasiswa calon guru IPA diperoleh dari posttest kepada mahasiswa calon guru IPA setelah dilaksanakan pembelajaran pada kelas eksperimen dan kelas kontrol. Profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) melalui angket pada kelas eksperimen dan kelas kontrol dapat dilihat melalui Tabel 3.

Tabel 3. Data profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) Mahasiswa calon guru IPA Berdasar Hasil Pretest dan Postest

Kelas	Kemampuan Awal (Pretest)				Kemampuan Akhir (Postest)			
	Rerata	Simp Baku	Skor		Rerata	Simp Baku	Skor	
			Min	Max			Min	Max
Eksperimen	53,25	4,15	45,00	60,00	65,32	4,15	62,00	92,00
Kontrol	51,50	4,25	43,00	59,00	55,28	3,75	49,00	78,00

Berdasarkan tabel 3 di atas, terdapat peningkatan kemampuan profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) baik kelas eksperimen maupun kelas kontrol. Data kemampuan profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) secara lengkap dapat dilihat pada lampiran

#### 4. Data *N*-Gain Profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) Mahasiswa calon guru IPA

Data kemampuan awal dan kemampuan akhir profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA diuji dengan Gain ternormalisasi (*N*- Gain). Data *N*-Gain diperoleh dari hasil pretest sebelum dilaksanakan pembelajaran dan hasil postest sesudah dilaksanakan pembelajaran baik pada kelas eksperimen maupun kelas kontrol. Hasil data *N*-Gain dapat dilihat pada Tabel 4.

Tabel 4. *N*-Gain Profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) Mahasiswa calon guru IPA

Kelas	Rerata		<i>N</i> -Gain	Simpangan Baku	Kategori
	Sebelum Pembelajaran	Sesudah Pembelajaran			
Eksperimen	53,25	65,32	0,71	0,21	Tinggi

Kontrol	51,50	55,28	0,28	0,15	Rendah
---------	-------	-------	------	------	--------

Tabel 5. Kriteria Gain Ternormalisasi

<b>Skor Gain Ternormalisasi</b>	<b>Interpretasi</b>
$g \geq 0,7$	Tinggi
$0,3 \leq g < 0,7$	Sedang
$g < 0,3$	Rendah

(Sumber: Hake, 2000:1)

Berdasarkan Tabel 4, hasil rerata sesudah dilaksanakan pembelajaran menunjukkan hasil yang lebih tinggi dari pada hasil rerata sebelum dilaksanakan pembelajaran baik pada kelas eksperimen maupun pada kelas kontrol. Hasil *N-Gain* profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA pada kelas eksperimen menunjukkan hasil yang lebih tinggi (kategori tinggi) dari pada hasil *N-Gain* pada kelas control (pada kategori rendah).

## 5. Profil Kemampuan Hasil tes HOTS berstandar PISA Mahasiswa calon guru IPA

Profil kemampuan hasil tes HOTS berstandar PISA mahasiswa calon guru IPA diperoleh dari dua data, yaitu data hasil posttest kelas eksperimen 37 mahasiswa dan kelas kontrol 29 mahasiswa. Data *posttest* diperoleh dari tes hasil tes HOTS berstandar PISA setelah mahasiswa calon guru IPA diberikan pembelajaran. *Pretest* dan *posttest* diberikan pada kedua kelompok, yaitu kelompok kelas eksperimen dan kelompok kelas kontrol. Profil kemampuan hasil tes HOTS berstandar PISA mahasiswa calon guru IPA dapat dilihat pada Tabel 6.

Tabel 6. Profil Kemampuan Hasil tes HOTS berstandar PISA Mahasiswa calon guru IPA

Kriteria	Profil Kemampuan berdasarkan hasil Posttest		
	Skor	Jumlah Mahasiswa	Prosentase (%)
Tinggi	100 - 80	4	6,06
Sedang	79 -50	47	71,21
Rendah	< 50	15	22,73

## 6. Uji Prasyarat Hipotesis

Uji prasyarat hipotesis dilakukan peneliti untuk menentukan statistika yang akan digunakan dalam penelitian ini, parametrik atau nonparametrik. Uji prasyarat meliputi uji normalitas dan homogenitas. Uji stastistik yang digunakan untuk uji normalitas dan uji homogenitas yaitu menggunakan program SPSS *for windows* versi 22.

Uji normalitas bertujuan untuk mengetahui data yang dikumpulkan terdistribusi normal atau tidak. Kriteria pada uji normalitas yaitu jika nilai Sig. (Signifikansi) atau nilai probabilitas < 0,05 maka data tidak terdistribusi normal, jika nilai Sig. (Signifikansi) atau nilai probabilitas 0,05 maka data terdistribusi normal. Hasil analisis ditampilkan

dalam tabel *Shapiro-Wilk*. Sedangkan uji homogenitas bertujuan untuk mengetahui seragam tidaknya varian sampel dari suatu populasi yang sama. Perhitungan uji homogenitas dengan melihat tabel *Levene Statistic* dengan kriteria nilai *Levene Statistic* > 0,05 maka data homogen.

Peneliti melakukan uji prasyarat terhadap hasil pretest mahasiswa calon guru IPA baik pada kelas eksperimen maupun pada kelas kontrol. Adapun hasil uji prasyarat dapat dilihat pada Tabel 7 dan Tabel 8.

Tabel 7. Hasil Uji Normalitas Kemampuan Awal Mahasiswa calon guru IPA

Variabel	Kelas	Nilai Sig.
Hasil pretest HOTS berstandar PISA	Eksperimen	0,269
	Kontrol	0,224

Tabel 8. Hasil Uji Homogenitas Kemampuan Awal Mahasiswa calon guru IPA

Variabel	Levene Statistic	df1	df2	Sig.
Hasil pretest HOTS berstandar PISA	0,024	1	66	0,900

Berdasarkan uji normalitas dan uji homogenitas, diketahui data peneliti normal dan homogen. Sehingga, penelitian ini menggunakan statistika parametrik dengan metode *Independent T Test*.

## 7. Hasil Uji Hipotesis

Penelitian ini menggunakan statistika parametrik dengan metode *Independent Sample T-Test* dan metode *Multivariate Analysis of Variance* (MANOVA). Pengujian hipotesis dengan metode *Independent Sample T-Test* dilakukan dengan menggunakan program *SPSS for windows* versi 22 kemudian peneliti melanjutkan dengan uji pengaruh (*effect size*). Perhitungan *effect size* dilakukan untuk mengetahui seberapa besar pengaruh penggunaan Workshop dan pendampingan berbasis terhadap profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA. *Effect size* juga digunakan untuk mengetahui seberapa besar pengaruh Workshop dan pendampingan terhadap hasil tes HOTS berstandar PISA mahasiswa calon guru IPA.

Sedangkan pengujian hipotesis dengan metode *Multivariate Analysis of Variance* (MANOVA) dilakukan dengan menggunakan program *SPSS for windows* versi 22. Metode ini digunakan untuk mengetahui pengaruh Workshop dan pendampingan terhadap profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking

survey internasional (PISA) dan hasil tes HOTS berstandar PISA mahasiswa calon guru IPA di Jurusan Pendidikan IPA FMIPA UNY.

**a. Uji Pengaruh Workshop dan pendampingan terhadap Profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) Mahasiswa calon guru IPA**

Uji hipotesis dengan metode *Independent Sample T-Test* digunakan untuk mengetahui apakah ada atau tidaknya pengaruh Workshop dan pendampingan pembuatan soal HOTS berbasis benchmarking survey internasional (PISA) terhadap profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA. Uji hipotesis ini menggunakan nilai gain ternormalisasi angket profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) dan data hasil tes simulasi profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA. Adapun hipotesis dari penelitian ini adalah sebagai berikut.

$H_0$  = Tidak ada pengaruh penggunaan *Workshop dan pendampingan pengembangan soal HOTS berbasis PISA* terhadap profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA di Jurusan Pendidikan IPA FMIPA UNY .

$H_a$  = Ada pengaruh Workshop dan pendampingan terhadap profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA di Jurusan Pendidikan IPA FMIPA UNY.

Dasar pengambilan keputusan pada uji hipotesis dengan metode *Independent Sample T-Test* adalah sebagai berikut.

- 1) Jika  $\text{Sig. (2-tailed)} > 0,025$ , maka  $H_0$  diterima
- 2) Jika  $\text{Sig. (2-tailed)} < 0,025$ , maka  $H_0$  ditolak

Hasil uji hipotesis dengan metode *Independent Sample T-test* berdasar nilai gain ternormalisasi angket profil kompetensi dalam pengembangan soal HOTS

berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA menggunakan *SPSS for windows* versi 16 dapat diketahui pada Tabel 9.

Tabel 9. Workshop dan pendampingan terhadap Profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA)

F	T	Sig. (2-tailed)
4,616	3,646	0,001

Berdasarkan output test statistika uji hipotesis dengan metode *Independent Sample T-Test* nilai Sig. (2-tailed) pada angket profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) sebesar 0,001 lebih kecil dari 0,025. Oleh karena itu, sebagaimana dasar pengambilan keputusan pada uji hipotesis dengan metode *Independent Sample T-Testi* dapat diketahui bahwa  $H_a$  diterima dan  $H_o$  ditolak. Hipotesis alternatif yang diterima yaitu ada pengaruh Workshop dan pendampingan terhadap profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA di Jurusan Pendidikan IPA FMIPA UNY.

Selain melakukan uji hipotesis dengan metode *Independent Sample T-Test* berdasar angket profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA, peneliti juga melakukan uji hipotesis dengan metode *Independent Sample T-Test* berdasar tes simulasi profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) menggunakan *SPSS for windows* versi 22 dapat diketahui pada Tabel 10.

Tabel 10. Penggunaan Workshop dan pendampingan terhadap Profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA)

F	T	Sig. (2-tailed)
6,428	12,982	0,000

Berdasarkan output test statistika uji hipotesis dengan metode *Independent Sample T-Test* nilai Sig. (2-tailed) pada tes simulasi bencana gempa bumi sebesar 0,000 lebih kecil dari 0,025. Oleh karena itu, sebagaimana dasar pengambilan

keputusan pada uji hipotesis dengan metode *Independent Sample T-Test* dapat diketahui bahwa  $H_a$  diterima dan  $H_o$  ditolak. Hipotesis alternatif yang diterima yaitu ada pengaruh penggunaan *Subject Specific Pedagogy* (SSP) IPA berbasis kegiatan workshop dan pendampingan terhadap profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA di Jurusan Pendidikan IPA FMIPA UNY.

Peneliti kemudian melakukan perhitungan *effect size* pada sikap kesiapsiagaan bencana gempa bumi berdasar angket yang diberikan kepada mahasiswa calon guru IPA . Besar *effect size* yang diperoleh yaitu 1,004 dengan kategori tinggi. Peneliti juga melakukan perhitungan *effect size* pada tes simulasi profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA . Hasil *effect size* yang diperoleh pada sikap profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) berdasar tes simulasi sebesar 3,790 dengan kategori tinggi.

#### **b. Uji Pengaruh Workshop dan pendampingan terhadap Hasil tes HOTS berstandar PISA Mahasiswa calon guru IPA di**

Uji hipotesis dengan metode *Independent Sample T-Test* digunakan untuk mengetahui apakah ada atau tidaknya pengaruh Workshop dan pendampingan terhadap hasil tes HOTS berstandar PISA mahasiswa calon guru IPA di Jurusan Pendidikan IPA FMIPA UNY. Uji hipotesis ini menggunakan nilai gain ternormalisasi hasil tes HOTS berstandar PISA mahasiswa calon guru IPA. Nilai gain ternormalisasi diperoleh dari nilai *pretest* dan nilai *possttest* mahasiswa calon guru IPA . Adapun hipotesis dari penelitian ini adalah sebagai berikut.

$H_o$  = Tidak ada pengaruh Workshop dan pendampingan terhadap hasil tes HOTS berstandar PISA mahasiswa calon guru IPA di Jurusan Pendidikan IPA FMIPA UNY.

Ha = Ada pengaruh Workshop dan pendampingan terhadap hasil tes HOTS berstandar PISA peserta di Jurusan Pendidikan IPA FMIPA UNY.

Dasar pengambilan keputusan pada uji hipotesis dengan metode *Independent Sample T-Test* adalah sebagai berikut.

- 1) Jika Sig. (2-tailed) > 0,025, maka Ho diterima
- 2) Jika Sig. (2-tailed) < 0,025, maka Ho ditolak

Hasil uji hipotesis dengan metode *Independent Sample T-test* berdasar nilai gain ternormalisasi hasil tes HOTS berstandar PISA mahasiswa calon guru IPA menggunakan *SPSS for windows* versi 22 dapat diketahui pada Tabel 11.

Tabel 11. Pengaruh workshop dan pendampingan terhadap Hasil tes HOTS berstandar PISA Mahasiswa calon guru IPA

F	T	Sig. (2-tailed)
1,302	2,592	0,023

Berdasarkan output test statistika uji hipotesis dengan metode *Independent T-Test* nilai Sig. (2-tailed) pada hasil tes HOTS berstandar PISA mahasiswa calon guru IPA sebesar 0,013 lebih kecil dari 0,025. Oleh karena itu, sebagaimana dasar pengambilan keputusan pada uji hipotesis dengan metode *Independent T-Testi* dapat diketahui bahwa Ha diterima dan Ho ditolak. Hipotesis alternatif yang diterima yaitu ada pengaruh workshop dan pendampingan terhadap hasil tes HOTS berstandar PISA mahasiswa calon guru IPA di Jurusan Pendidikan IPA FMIPA UNY.

Peneliti kemudian melakukan perhitungan *effect size* pada hasil tes HOTS berstandar PISA mahasiswa calon guru IPA. Besar *effect size* yang diperoleh yaitu 0,932 dengan kategori tinggi.

## B. Pembahasan

Penelitian ini bertujuan mengetahui pengaruh Workshop dan pendampingan terhadap profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA dan mengetahui pengaruh

Workshop dan pendampingan terhadap hasil tes HOTS berstandar PISA mahasiswa calon guru IPA. Penelitian ini menggunakan *pretest-posttest control group design*. Populasi pada penelitian ini yaitu mahasiswa calon guru IPA yang mengambil mata kuliah Evaluasi Pembelajaran IPA tahun 2018. Pengambilan sampel dilakukan dengan *cluster sampling* atau teknik sampling daerah. Teknik *cluster sampling* digunakan karena pemilihan sampel berdasarkan kelompok populasi yang telah ditentukan. Kelas yang ditentukan sebagai sampel adalah kelas A dan kelas I. Setelah menentukan kelas sebagai sampel, peneliti menentukan secara acak untuk menentukan mana kelas eksperimen dan mana kelas kontrol. Berdasarkan proses acak tersebut, kelas A sebagai kelas eksperimen dan kelas I sebagai kelas kontrol. Walaupun kelompok sampel telah ditentukan, peneliti tetap melakukan uji normalitas dan uji homogenitas populasi yang diolah menggunakan program SPSS *for windows* versi 22 sebagai data pendukung. Hasil yang diperoleh yaitu sampel yang telah ditentukan sebagai sampel memiliki data yang terdistribusi normal namun dan homogen.

Perbedaan kedua kelas tersebut terletak pada pembelajaran yang berlangsung. Kelas eksperimen menggunakan Workshop dan pendampingan pengembangan soal HOTS berbasis PISA, sedangkan kelas kontrol menggunakan pengembangan soal HOTS berdasarkan Buku Evaluasi Pembelajaran IPA yang disusun oleh Tim Peneliti dan telah diterbitkan UNY Press. Kedua sampel tersebut diberikan tema pembelajaran yang sama, yaitu materi IPA yang diajarkan di Sekolah Menengah Pertama. Kegiatan inti pada pembelajaran kelas eksperimen difasilitasi dengan workshop dan pendampingan pengembangan soal HOTS berbasis benchmarking survey internasional (PISA). Pada tahap persiapan, mahasiswa calon guru IPA mempersiapkan materi yang akan dipelajari. Melalui materi yang telah dipersiapkan, mahasiswa calon guru IPA melakukan tahap penjelajahan yang meliputi sejumlah kegiatan yang dilakukan mahasiswa selama pembelajaran berlangsung. Selanjutnya, mahasiswa calon guru IPA mengumpulkan banyak pertanyaan berdasar apa yang dilakukan pada tahap penjelajahan. Pada tahap pengembangan soal, pertanyaan-pertanyaan inilah yang kemudian akan diselidiki oleh mahasiswa calon guru IPA untuk menemukan jawaban dari pertanyaan yang telah diajukan. Tahap terakhir, yaitu refleksi, kegiatan pembahasan kembali mengenai hasil pengembangan soal yang dilakukan mahasiswa calon guru IPA.

Sebelum kelas eksperimen dan kelas kontrol diberikan perlakuan, peneliti memberikan *pretest* pada kedua kelas yang bertujuan untuk mengetahui awal mahasiswa calon guru IPA . *Pretest* untuk mengetahui kemampuan awal profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA). Setelah *pretest* dilaksanakan pada kedua kelas, peneliti memberikan perlakuan pada kelas eksperimen dan kelas kontrol. Di akhir pembelajaran, peneliti memberikan *posttest* dan tes pada kelas eksperimen dan kelas kontrol sebagai pengukur kemampuan akhir mahasiswa calon guru IPA . *Posttest* yang dilaksanakan pada kedua kelas adalah sama, namun demikian pada soal *posttest* yang untuk mengukur kemampuan akhir hasil tes HOTS berstandar PISA dilakukan pengacakan susunan nomor soal.

Peneliti melaksanakan proses kegiatan belajar mengajar selama 4 pertemuan dengan total durasi 8 sks. Selama pertemuan tersebut, peneliti dapat berhasil melaksanakan seluruh tahapan dalam kegiatan. Terbukti dengan hasil observasi keterlaksanaan pembelajaran yang mencapai diatas 90 % dan memiliki kategori sangat baik. Pada akhir di pertemuan keempat, peneliti mengambil data profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA pada kelas eksperimen dan kelas kontrol dengan melaksanakan tes.. Hasil tes menunjukkan profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA sesuai dengan indicator. Skor yang diperoleh dari penilaian tes simulasi kemudian dikonversi dalam skala 100 dan hasil yang diperoleh kemudian dibandingkan tabel kategori. Selanjutnya, saat rangkaian pembelajaran selesai, peneliti memberikan *posttest* untuk mengukur profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) dan hasil tes HOTS berstandar PISA mahasiswa calon guru IPA pada kelas eksperimen dan kelas kontrol.

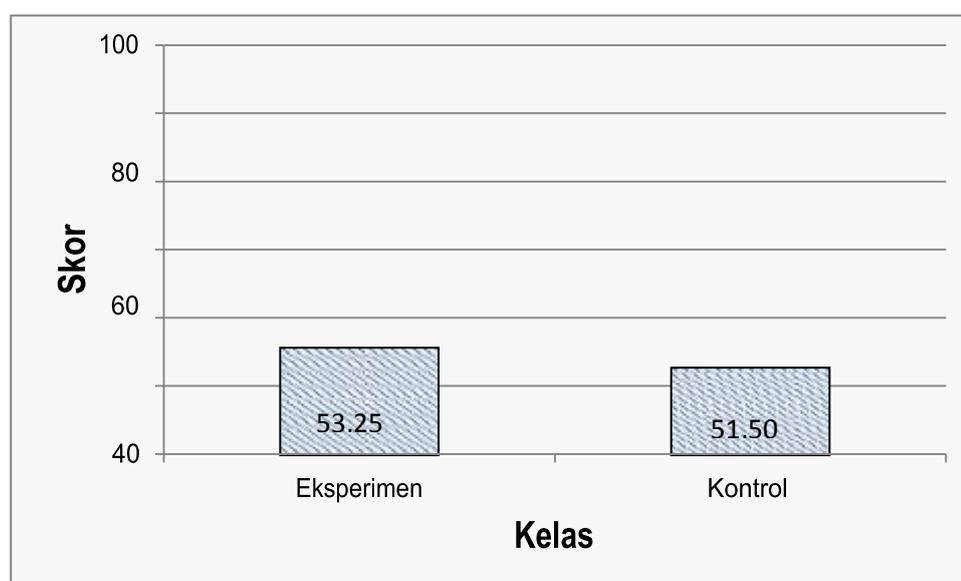
#### **1. Pengaruh Workshop dan pendampingan terhadap Profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) Mahasiswa calon guru IPA**

Mahasiswa calon guru IPA yang memiliki kompetensi tinggi dalam pengembangan soal HOTS berbasis PISA diharapkan dapat menjadi agen perubahan dalam sistem penilaian di tingkat sekolah menengah pertama. Profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) pada

kelas eksperimen difasilitasi dengan pembelajaran dengan pelatihan pengembangan soal HOTS berstandar PISA .

Data profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA diperoleh dari sebelum diberikan perlakuan (*pretest*), saat pembelajaran berlangsung sebagai pemantapan pada tahapan kegiatan workshop dan pendampingan , dan setelah pemberian perlakuan (*posttest*). Hasil data profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) diperoleh dari hasil test dengan menggunakan perangkat soal yang telah divalidasi oleh dosen ahli.

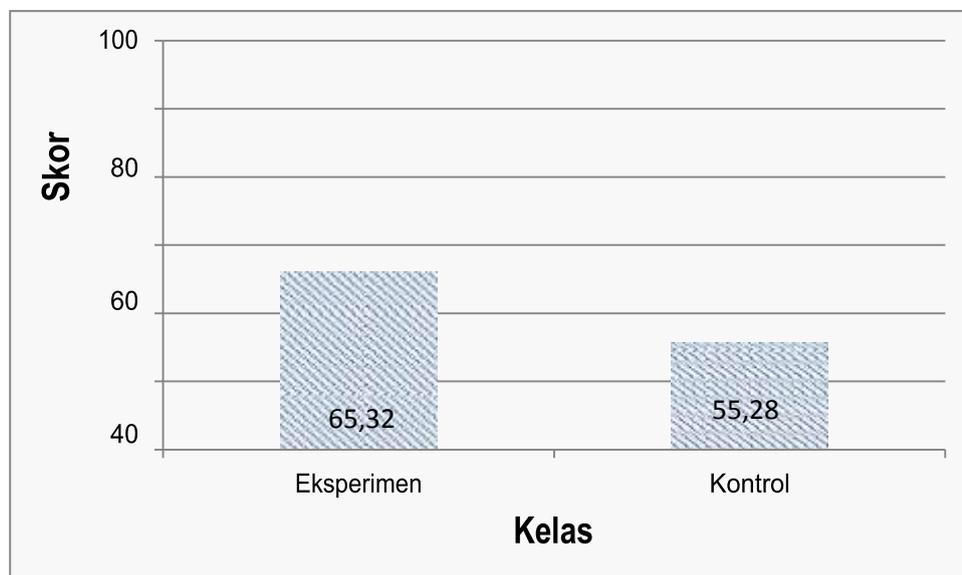
Data profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA dilihat dari kemampuan awal berdasar soal pretest kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA tersaji pada Gambar 6.



Gambar 6. Histogram Data Kemampuan Awal Profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) Mahasiswa calon guru IPA

Data kemampuan awal profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA kelas eksperimen lebih tinggi dari kelas kontrol. Selisih skor antara kelas eksperimen dan kelas kontrol sebanyak 1,75. Setelah diberi perlakuan, data kemampuan akhir profil

kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA dapat dilihat pada Gambar 7.



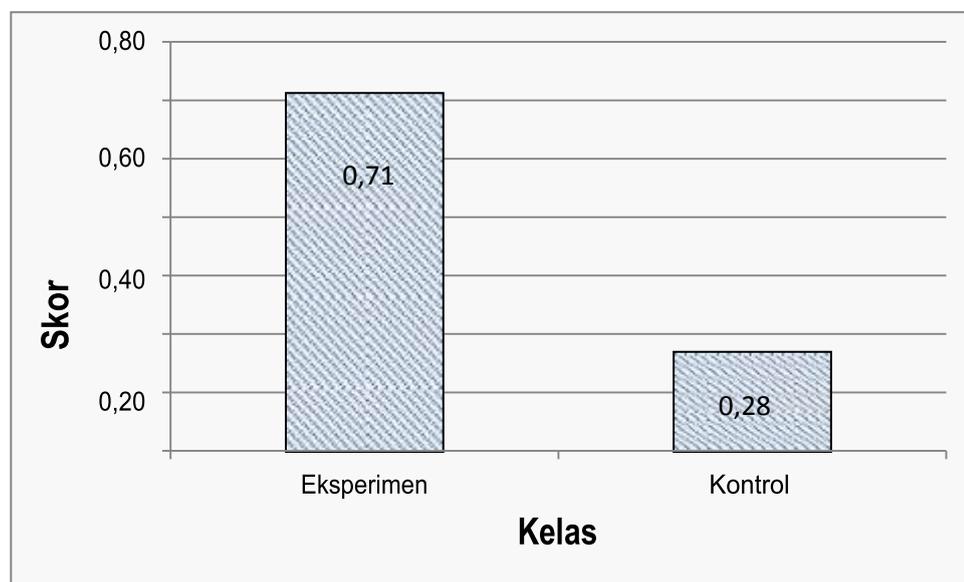
Gambar 7. Histogram Data Kemampuan Akhir Profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) Mahasiswa calon guru IPA

Data kemampuan akhir profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) menunjukkan bahwa kelas eksperimen memiliki skor yang lebih tinggi dibanding dengan kelas kontrol dengan selisih skor sebanyak 10,04.

Berdasarkan hasil analisis kemampuan awal dan kemampuan akhir profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA), terdapat peningkatan skor rata-rata *pretest* dan *posttest* pada angket profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) .

Data di atas menunjukkan bahwa terdapat peningkatan pada kedua kelompok kelas. Hasil pada kelas eksperimen mengalami peningkatan skor sebesar 10,04, sedangkan

pada kelas kontrol mengalami peningkatan sebesar 1,75. Dari hasil tersebut dapat dilihat bahwa peningkatan hasil tes kelompok kontrol tidak sebesar pada kelas eksperimen. Hal tersebut terjadi sebagai implikasi penggunaan kegiatan workshop dan pendampingan pada kelas eksperimen. Peneliti juga melihat adanya peningkatan melalui skor gain ternormalisasi profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA yang ditunjukkan pada Gambar 8.



Gambar 8. Histogram Skor Gain Ternormalisasi Profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) Mahasiswa calon guru IPA

Berdasarkan histogram di atas diketahui bahwa skor gain ternormalisasi profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA kelas eksperimen lebih tinggi dibanding profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) kelas kontrol. Kelas eksperimen memiliki nilai gain ternormalisasi profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) sebesar 0,71 dengan kategori tinggi, sedangkan pada kelas kontrol memiliki nilai gain ternormalisasi profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) sebesar 0,28 dengan kategori rendah.

peningkatan sikap profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) dengan berdasarkan data kemampuan awal dan kemampuan akhir mahasiswa calon guru IPA, sehingga diketahui bahwa kedua kelompok kelas mengalami peningkatan profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) dari sebelum diberikan perlakuan hingga setelah diberi perlakuan.

Hasil uji hipotesis yang dilakukan dengan metode *Independent Sample T-Test* menunjukkan bahwa nilai Sig (*2-tailed*) pada test kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) lebih rendah dari 0,025, yaitu nilai masing-masing sebesar 0,001 dan 0,000. Berdasarkan nilai Sig (*2-tailed*), penelitian ini menolak  $H_0$  dan menerima  $H_a$ . Hipotesis alternatif yaitu terdapat pengaruh kegiatan workshop dan pendampingan terhadap peningkatan profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA.

Selain menggunakan metode *Independent Sample T-Test*, penelitian ini juga melihat uji pengaruh melalui nilai *effect size*. *Effect size* digunakan untuk mengukur perbedaan di antara kedua kelompok kelas dan untuk menunjukkan besar pengaruh variabel bebas yang diintervensikan pada kelompok sampel percobaan. Nilai *effect size* yang diperoleh pada angket profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) sebesar 1,002 dan nilai *effect size* yang diperoleh pada tes simulasi profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) sebesar 1,932. Menurut Cohen (2011:617), jika besar nilai *effect size* yang diperoleh lebih dari 1,00 maka uji pengaruh tersebut memiliki efek yang besar. Sehingga, penerapan Kegiatan workshop dan pendampingan pengembangan soal HOTS berbasis PISA memiliki pengaruh yang tinggi atau signifikan terhadap peningkatan profil kompetensi dalam pengembangan soal HOTS berbasis benchmarking survey internasional (PISA) mahasiswa calon guru IPA.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **A. Kesimpulan**

Sesuai dengan tujuan penelitian, hasil yang telah di dapatkan dari kegiatan penelitian ini, adalah; (1) dihasilkannya perangkat test HOTS berbasis PISA yang valid berdasarkan validasi isi dan konstruk oleh dosen ahli evaluasi, (2) hasil pemetaan kompetensi akademik calon guru khususnya kemampuan dalam mengembangkan assessment berstandar *survei benchmarking* internasional (PISA) adalah; 6,06 % berada pada kategori sangat baik, 71,21 % baik, dan 22,73 pada kategori kurang, (3) terdapat pengaruh yang signifikan dari workshop dan pendampingan terhadap peningkatan kompetensi calon guru dalam pengembangan assessment berstandar *survei benchmarking* internasional (PISA) berdasarkan uji beda nilai pretest dan posttest dengan *Independent Sample T-Test* diperoleh nilai Sig (2-tailed) sebesar  $0,018 < 0,025$  dan nilai *effect size* sebesar 0,71.

#### **B. Saran**

Soal berpikir tingkat tinggi (High Order Thinking Skills) yang berorientasi pada survey benchmarking internasional sangat penting dikembangkan dalam pembelajaran di sekolah di Indonesia. Karena itu orientasinya bukan sekedar untuk menghadapi pemetaan internasional, tetapi yang lebih penting adalah bagaimana meningkatkan kualitas sumber daya manusia di Indonesia melalui jalur pendidikan formal. Dengan mengembangkan HOTS maka akan dihasilkan lulusan yang lebih kreatif, inovatif dan siap menghadapi tantangan di era revolusi Industri 4.0.

#### **C. Ucapan Terimakasih**

Terimakasih diucapkan kepada FMIPA Universitas Negeri Yogyakarta yang telah memberikan pendampingan manajemen pengelolaan penelitian serta pendanaannya melalui DIPA FMIPA UNY tahun 2018.

## DAFTAR PUSTAKA

- Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing*. New York: Longman.
- Angoff, W. H. 1982. Uses of Difficulty and Discrimination Indices for Detecting Item Bias In RA Berk. *Handbook of Methods for Detecting Item Bias*. Baltimore: John Hopkins University Press.
- Asmin. 2004. Implementasi Teori Responsi Butir dan Fungsi Informasi Butir Tes dalam Pengujian Hasil Belajar Akhir di Sekolah. *Jurnal Pendidikan dan Kebudayaan*, X (48): 234-245.
- Barnard. John. J. 1996. *In Search for Equity in Educational Measurement: Traditional Versus Modern Equating Methods*. Makalah: Disampaikan pada ASEESA National Conference di HSRC Conference Centre. Pretoria: Afrika Selatan.
- Center for Excellence in Learning and Teaching (CELT). (2011). A model of learning objectives. Iowa State University. Retrieved March 2011, from <http://www.celt.iastate.edu/teaching/RevisedBlooms1.html>.
- Clark, D. (2010). Bloom's taxonomy of learning domains: The three types of learning. **Big Dog & Little Dog's Performance Juxtaposition**. Edmonds, WA: Author. Retrieved from <http://www.nwlink.com/~donclark/hrd/bloom.html>.
- Cracolice, M.S., Deming, J.C. & Ehlert, B. (2008). Concept learning versus problem solving: a cognitive difference. *Journal of Chemical Education*. 85 (6), 873-878.
- Rustad, S. A. Munandar, dan Dwiyanto. (2004). *Analisis Prasarana dan Sarana Pendidikan SD/MI, SMP/MTs dan /SMK*. Jakarta: Balitbang, Depdiknas.
- Wiyanto. (2004). *Kegiatan Laboratorium IPA untuyk Mengembangkan Kemampuan Berpikir*. Prosiding Makalah Konvensi Nasional Pendidikan Indonesia (Konaspi)V di Surabaya, 9 Oktober 2004. ISBN: 979-445-001-4.
- Camilli, Gregory, dan Lorrie A. Shepard. 1994. *Methods for Identifying Biased Test Items*. California: Sage Publication.
- Chong Ho Yu dan Sharon E. Osborn. 2005. Test Equating by Common Items and Common Subject: Concepts and Applications. *Practical Assessment, Research & Evaluation*. X (4): 187-198.
- Crocker, Linda, & Algina, James. 1986. *Introduction to classical and modern test theory*. New York: Holt, Rinehart and Winston, Inc.
- Djaali. 2004. *Pengukuran Dalam Bidang Pendidikan*. Jakarta: Program Pascasarjana Universitas Negeri Jakarta.

- Dorans, N. J. (2004). Equating, concordance, and expectation. *Applied Psychological Measurement*, 28 (4),227-246.
- Gronlund, Norman. E. 1985. *Measurement and Evaluation in Teaching*. New York: Macmillan Publishing Company.
- Hambleton, Ronald K, Swaminathan, H., dan Jane Rogers, H. 1991. *Fundamentals of Item Response Theory*. London: SagePublications.
- Hambleton, Ronald K., dan Swaminathan, H. 1985. *Item Response Theory: Principle and Applications*. Boston: Kluwer Nijhoff Publishing.
- Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.),*Journal of Educational measurement* (4th ed., pp. 187{220). Westport, CT: Greenwood.
- Jihad, Asep, Abdul Haris. 2011. *Evaluasi Pembelajaran*. Multi Pressindo: Yogyakarta.
- Kim, S., von Davier, A. A., & Haberman, S. (2008). Il-sample equating using a synthetic linking function. *Journal of Educational Measurement*, 45, 325{342}
- Kolen, Michael J., dan Robert L. Brennan. 2004. *Test Equating, Scaling, and Linking: Methods and Practices*. New York: Springer.
- Kolen, Michael J., dan Robert L. Brennan. 1995. *Test Equating*. New York: Springer Verlag.
- Kumaidi. 2000. Standardisasi Butir Soal. *Jurnal Pendidikan dan Kebudayaan*. V(5): 132-143.
- Livingstone, S. A., Doran, N. J. dan Wright, N. K. 1990. What Combination of Sampling and Equating Methods Work Best?. *Applied Measurement in Education*. III (2): 73-95.
- Livingston, S. A., & Kim, S. (2009). The circle-arc method for equating in ll samples. *Journal of Educational Measurement*, 46, 330{343}
- Lord, F. M. (2009). The standard error of equipercentile equating. *Journal of Educational Statistics*,7, 165{174}
- Lord, Frederick, M.1990. *Aplications of Item Response Theory to Practical Testing Problems*. New Jersey: LawrenceErlbaum Associates, Publishers.
- Mary J.Allen and Wendy M Yen, 1989, *Introduction to Measurement Theory*, California: Broke.
- McDonald, Roderick P. 1991. *Test Theory: A Unified Treatment*. New Jersey: Lawrence Erlbaum Associatiates Publisher.
- Naga, Dali, S. 1992. *Pengantar Teori Sekor Pada Pengukuran Pendidikan*. Jakarta: Besbats.

- Nitko, Anthony. J. 1992. *Criterion Reference Testing Workshop: Handouts and Reading Material Tidak dipublikasikan*. Cipayung, Bogor: Examination Development Unit (Puslitbang Sisjian).
- Miyatun, Erna., dan Djemari Mardapi. 2000. Komparasi Metode Penyetaraan Tes Menurut Teori Responsi Butir. *Jurnal Penelitian dan Evaluasi*. II (3): 124-132.
- Peraturan Pemerintah No. 19 Th 2005 Tentang: Standar Nasional Pendidikan (SNP). Bandung: Citra Umbara.
- Peterson, N.S., Kolen, M.J., dan Hoover, H.D. 1989. Scaling, Norming, and Equating. In R.L. Linn (Ed), *Educational Measurement*. New York: Macmillan.
- Rahayu, Wardani. 2008. Pengaruh Metode Linking Terhadap Banyak Butir False Positive pada Pendeteksian DIF Berdasarkan Teori Responsi Butir. *Disertasi*. Jakarta: Universitas Negeri Jakarta.
- Ridho, Ali. 2007. Karakteristik Psikometrik Tes Berdasarkan Pendekatan Teori Tes Klasik dan Teori Respon Aitem. *Jurnal Insan Media*. II (2): 1-28.
- Setiadi, Hari. 1998. Bank Soal yang Dikalibrasi dengan Konsep IRT Memecahkan Permasalahan Ujian-ujian Sistematis yang Diadakan pada Periode-periode Tertentu, *Jurnal Kajian Dikbud IV* (13).
- Setiadi, Hari. 2009. Permasalahan dan Solusinya dalam Pelaksanaan Ujian Nasional di Masa Mendatang, *Matahari: Jurnal Penelitian dan Pendidikan*. X (1): 66-74.
- Skaggs, G. (2005). Accuracy of random groups equating with very ll amples. *Journal of Educational Measurement*, 42, 309{330}
- Susongko, Purwo. 2005. Penyetaraan Parameter Butir Secara Konkuren untuk Menguji Secara Statistik Keberadaan Item Function (DIF). Makalah: Disampaikan pada Seminar Nasional *Hasil Penelitian tentang Evaluasi Hasil Belajar serta Pengelolaannya*. Pascasarjana UNY Didukung oleh Direktorat P2TK & KPT dan HEPI, Yogyakarta, 14-15 Mei 2005.
- Sukirno, D. S. 2007. Penyetaraan Tes UAN: Mengapa dan Bagaimana. *Jurnal Cakrawala Pendidikan*. XXVI (3): 305-321.
- Syarifah. 2007. Persyaratan Analisis Instrumen Sebagai Prasyarat Ketepatan Hasil Analisis Dalam Penelitian Pendidikan. *Cakrawala Pendidikan*. XXVI (2): 15-27.
- Swediati, Nonny. 1997. *Metode untuk Penyetaraan (Equating) Sekor Tes Secara Klasik*. Pusat Pengujian Balitbang Dikbud: Jakarta.
- Wibowo, Mungin Eddy. 2011. Kondisi Psikologis Mahasiswa dalam Menghadapi Ujian Nasional, *Buletin BNSP: Media Komunikasi dan Dialog Standar Pendidikan*. VI (1): 7-11.

- Widhiarso, Wahyu. 2011. Aplikasi Teori Respon Butir untuk Menguji Invariansi Pengukuran Psikologis Guna Keperluan Survei dan Seleksi Pekerjaan. *Jurnal Psikobuana*. III (2): 104-117.
- Tumilisar, A.V.J. 2006. Akurasi Relatif Penyetaraan Sekor Tes untuk Sampel Berukuran 300 Ditinjau dari Metode Penyetaraan dan Teknik Penghalusan. *Jurnal Pendidikan Penabur*. V (6): 1-19.
- Zhu, W. 1998. Test Equating: What, Why, How?. *Research Quarterly for Exercises and Sport*. Wayne State University.

## **LAMPIRAN**

- 1. Contoh soal yang dikembangkan dalam kegiatan penelitian**
- 2. Berita Acara kegiatan seminar awal dan akhir penelitian**

## CONTOH SOAL SOAL HOTS BERBASIS PISA YANG DIKEMBANGKAN

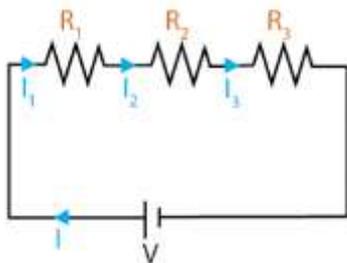
Nama :  
.....

Kelas :  
.....

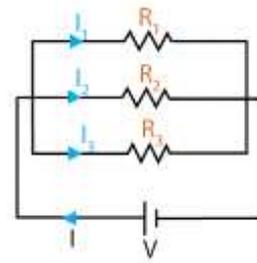
Bacalah artikel berikut dan kerjakan soal di bawah ini!

### HAMBATAN SERI DAN PARALEL

Rangkaian listrik terdiri atas dua jenis yaitu seri dan paralel. Selain itu ada juga gabungan dari dua jenis rangkaian listrik, yang disebut rangkaian campuran sehingga di dalam rangkaian tersebut ada seri dan paralel. Perbedaan jenis rangkaian listrik ini terletak pada cara merangkainya. Rangkaian listrik seri disusun secara berurutan, sedangkan rangkaian listrik paralel disusun secara bersusun sejajar atau paralel. Penerapan rangkaian seri dalam kehidupan sehari-hari dapat ditemukan pada lampu senter.



Gambar 1. Rangkaian Seri



Gambar 1. Rangkaian Paralel

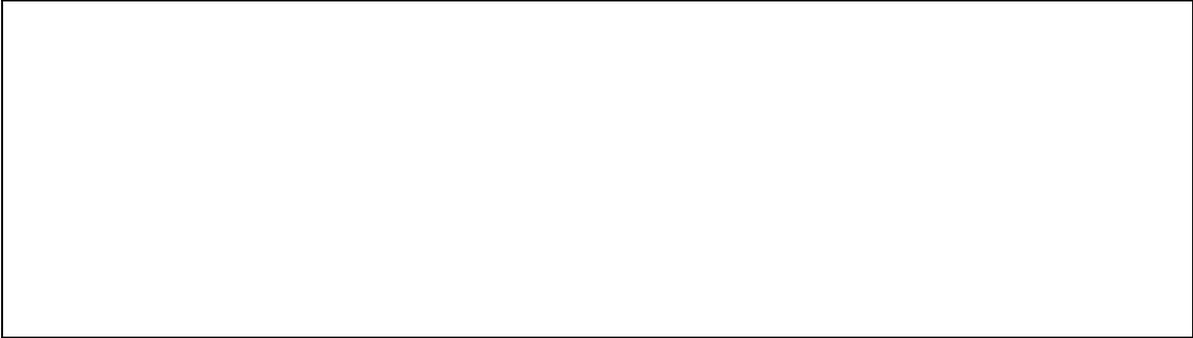
Rangkaian listrik seri, paralel, dan campuran memiliki karakteristik dan rumus yang berbeda. Rangkaian tersebut mempunyai kelebihan dan kekurangan masing-masing.

#### A. Soal Uraian

Kerjakan soal berikut berdasarkan percobaan yang telah Anda lakukan dan bacaan di atas

1. Buatlah contoh disain rangkaian listrik sederhana yang di susun secara campuran!

2. Dengan menggunakan lampu sebagai hambatan pada rangkaian listrik yang di susun seri dengan paralel dan temukan kelebihan dan kekurangannya masing-masing berdasarkan Gambar 1 dan 2 di atas!



**B. Soal Benar dan Salah**

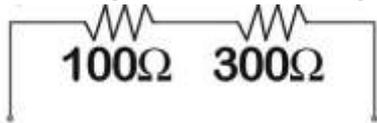
Pilih B jika pernyataan benar dan S jika pernyataan salah!

No.	Pernyataan	Jawaban
1.	Radio, kulkas, dan barang elektronik lainnya <b>tidak dapat disamakan</b> dengan hambatan listrik.	<b>B S</b>
2.	Menghitung hambatan pengganti pada rangkaian seri berbeda dengan cara menghitung hambatan pengganti pada dan paralel	<b>B S</b>

**C. Soal Pilihan Ganda**

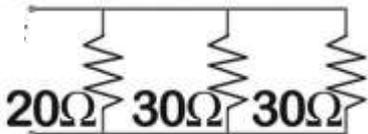
Pilihlah salah satu jawaban yang paling tepat!

1. Suatu rangkaian di susun sebagai berikut.



Nilai hambatan total dari rangkaian di atas adalah ... .

- a. 100 Ω
  - b. 200 Ω
  - c. 300 Ω
  - d. 400 Ω
2. Rangkaian listrik di susun sebagai berikut.



Nilai hambatan total dari rangkaian di atas adalah ... .

- a. 8,57 Ω
- b. 10,65 Ω
- c. 14,47 Ω
- d. 16,65 Ω

**D. Soal Menjodohkan**

Jawablah pertanyaan berikut dengan menjodohkan jawaban yang ada di sisi kanan pertanyaan!

No.	Soal	Pilihan jawaban
1.	Jika suatu hambatan atau resistor masing-masing bernilai $R$ dalam jumlah yang banyak disusun secara seri, maka nilai hambatan total menjadi semakin .....	a. Kecil b. Besar
2.	Jika suatu hambatan atau resistor masing-masing bernilai $R$ dalam jumlah yang banyak disusun secara paralel, maka nilai hambatan total menjadi semakin .....	c. Dua kali d. Ohmmeter e. Stabil

### E. Soal Jawaban Singkat

Jawablah pertanyaan berikut dengan benar!

1. Alat ukur nilai hambatan listrik secara langsung adalah .....
2. Jika pada suatu rangkaian listrik ada satu lampu yang mati, maka yang lain juga ikut mati adalah salah satu ciri rangkaian listrik yang disusun secara .....

## IX

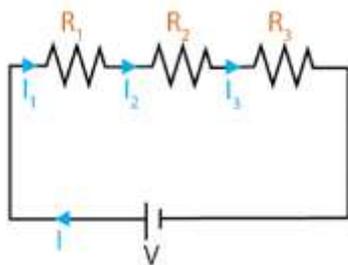
Nama :  
.....

Kelas :  
.....

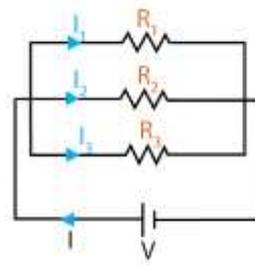
Bacalah artikel berikut dan kerjakan soal di bawah ini!

### HAMBATAN SERI DAN PARALEL

Rangkaian listrik terdiri atas dua jenis yaitu seri dan paralel. Selain itu ada juga gabungan dari dua jenis rangkaian listrik, yang disebut rangkaian campuran sehingga di dalam rangkaian tersebut ada seri dan paralel. Perbedaan jenis rangkaian listrik ini terletak pada cara merangkainya. Rangkaian listrik seri disusun secara berurutan, sedangkan rangkaian listrik paralel disusun secara bersusun sejajar atau paralel. Penerapan rangkaian seri dalam kehidupan sehari-hari dapat ditemukan pada lampu senter.



Gambar 1. Rangkaian Seri



Gambar 1. Rangkaian Paralel

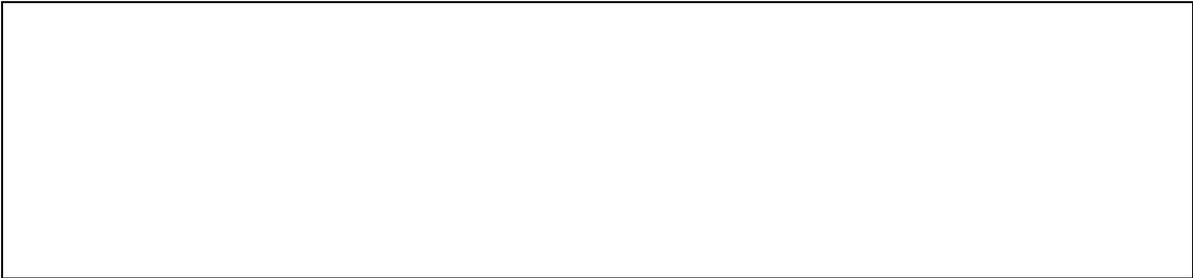
Rangkaian listrik seri, paralel, dan campuran memiliki karakteristik dan rumus yang berbeda. Rangkaian tersebut mempunyai kelebihan dan kekurangan masing-masing.

### F. Soal Uraian

Kerjakan soal berikut berdasarkan percobaan yang telah Anda lakukan dan bacaan di atas

3. Buatlah contoh disain rangkaian listrik sederhana yang di susun secara campuran!

4. Dengan menggunakan lampu sebagai hambatan pada rangkaian listrik yang di susun seri dengan paralel dan temukan kelebihan dan kekurangannya masing-masing berdasarkan Gambar 1 dan 2 di atas!



**G. Soal Benar dan Salah**

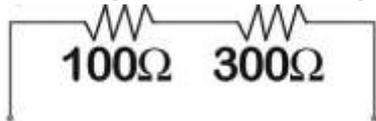
Pilih B jika pernyataan benar dan S jika pernyataan salah!

No.	Pernyataan	Jawaban
1.	Radio, kulkas, dan barang elektronik lainnya <b>tidak dapat disamakan</b> dengan hambatan listrik.	<b>B S</b>
2.	Menghitung hambatan pengganti pada rangkaian seri berbeda dengan cara menghitung hambatan pengganti pada dan paralel	<b>B S</b>

**H. Soal Pilihan Ganda**

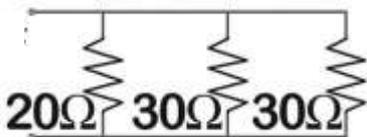
Pilihlah salah satu jawaban yang paling tepat!

3. Suatu rangkaian di susun sebagai berikut.



Nilai hambatan total dari rangkaian di atas adalah ... .

- e. 100 Ω
  - f. 200 Ω
  - g. 300 Ω
  - h. 400 Ω
4. Rangkaian listrik di susun sebagai berikut.



Nilai hambatan total dari rangkaian di atas adalah ... .

- e. 8,57 Ω
- f. 10,65 Ω
- g. 14,47 Ω
- h. 16,65 Ω

**I. Soal Menjodohkan**

Jawablah pertanyaan berikut dengan menjodohkan jawaban yang ada di sisi kanan pertanyaan!

No.	Soal	Pilihan jawaban
1.	Jika suatu hambatan atau resistor masing-masing bernilai $R$ dalam jumlah yang banyak disusun secara seri, maka nilai hambatan total menjadi semakin .....	f. Kecil g. Besar
2.	Jika suatu hambatan atau resistor masing-masing bernilai $R$ dalam jumlah yang banyak disusun secara paralel, maka nilai hambatan total menjadi semakin .....	h. Dua kali i. Ohmmeter j. Stabil

**J. Soal Jawaban Singkat**

**Jawablah pertanyaan berikut dengan benar!**

3. Alat ukur nilai hambatan listrik secara langsung adalah .....
4. Jika pada suatu rangkaian listrik ada satu lampu yang mati, maka yang lain juga ikut mati adalah salah satu ciri rangkaian listrik yang disusun secara .....