

**PENGEMBANGAN INSTRUMEN PENILAIAN
KOMPETENSI MATEMATIKA MATERI STATISTIKA SISWA SMP
BERDASARKAN KURIKULUM 2013**



Oleh:

KANA HIDAYATI

NIM. 11701261010

**Disertasi ini ditulis untuk memenuhi sebagian persyaratan
untuk mendapatkan gelar Doktor Pendidikan**

**PROGRAM STUDI PENELITIAN DAN EVALUASI PENDIDIKAN
PROGRAM PASCASARJANA
UNIVERSITAS NEGERI YOGYAKARTA
2018**

ABSTRAK

KANA HIDAYATI: *Pengembangan Instrumen Penilaian Kompetensi Matematika Materi Statistika Siswa SMP Berdasarkan Kurikulum 2013. Disertasi. Yogyakarta: Program Pascasarjana, Universitas Negeri Yogyakarta, 2018.*

Tujuan utama penelitian ini adalah menghasilkan instrumen penilaian yang dapat mengukur dengan tepat pencapaian kompetensi matematika siswa SMP untuk aspek sikap, pengetahuan, dan keterampilan pada materi statistika berdasarkan Kurikulum 2013.

Penelitian pengembangan ini meliputi tiga tahapan yaitu menyusun instrumen awal, uji coba, dan interpretasi hasil uji coba. Penyusunan instrumen awal memuat deskripsi implementasi kegiatan penyusunan instrumen. Validitas isi instrumen diperoleh melalui *expert judgment*. Uji coba dilakukan untuk memperoleh bukti kualitas instrumen secara empiris. Interpretasi hasil uji coba merupakan tahapan untuk menentukan perlu tidaknya revisi dan uji coba lanjut. Subjek uji coba adalah siswa SMP N 5 Yogyakarta, SMP Muhammadiyah 2 Yogyakarta, dan SMPIT Abu Bakar Yogyakarta. Responden uji coba pertama sebanyak 688 siswa dan uji coba kedua sebanyak 696 siswa. Validitas konstruk untuk instrumen penilaian aspek sikap diperoleh melalui *Lisrel* menggunakan *second order* CFA, sedangkan untuk instrumen penilaian aspek pengetahuan dan keterampilan diperoleh melalui *Lisrel* menggunakan *first order* CFA. Karakteristik butir pada instrumen penilaian aspek pengetahuan dan keterampilan diperoleh melalui *Parscale* menggunakan IRT Poltomus GPCM. Reliabilitas instrumen meliputi koefisien Omega (ω) yang dihitung berdasarkan nilai λ dari *Lisrel* dan koefisien Alpha dengan SPSS.

Berdasarkan penelitian dihasilkan: (a) instrumen penilaian aspek sikap berupa skala untuk mengukur karakter jujur siswa berbentuk penskalaan subjek, skala untuk mengukur karakter tanggung jawab siswa berbentuk penskalaan subjek, dan skala sikap matematis berbentuk penskalaan respons, dan (b) instrumen penilaian aspek pengetahuan dan keterampilan berupa tes uraian dan tugas proyek. Keseluruhan instrumen secara teoretis dan empiris terbukti berkualitas baik. Kesesuaian butir dengan indikator pada keseluruhan instrumen adalah baik dengan indeks V Aiken sebesar 0,88–1,00 dan indeks keselarasan tes sebesar 0,93. Skala untuk mengukur karakter jujur, karakter tanggung jawab, dan sikap matematis memiliki validitas konstruk dan reliabilitas berkategori baik dengan koefisien Omega masing-masing sebesar 0,86, 0,89, dan 0,81. Tes dan tugas memiliki validitas konstruk dan reliabilitas berkategori baik dengan koefisien Alpha sebesar 0,78 untuk tes dan 0,70 untuk tugas. Seluruh butir tes dan tugas berkategori baik dengan daya beda sebesar 0,31-2,01 dan indeks kesukaran sebesar -2,33 sampai -0,47 untuk tes, daya beda sebesar 0,46–1,11 dan indeks kesukaran sebesar -2,01 sampai -0,91 untuk tugas. Tes memberikan informasi akurat pada theta sebesar -3,0 sampai 0,8, sedangkan tugas pada theta sebesar -3,0 sampai 0,5.

Kata kunci: *instrumen, penilaian, statistika, SMP, Kurikulum 2013*

ABSTRACT

KANA HIDAYATI: *Developing Mathematics Competency Assessment Instrument of Statistical Material of Junior High School Students Based on Curriculum 2013. Dissertation. Yogyakarta: Postgraduate, Yogyakarta State University, 2018.*

The main purpose of this research is to produce assessment instrument that can accurately measure the achievement of SMP student's mathematics competency for attitude, knowledge, and skills aspects of statistical materials based on Curriculum 2013.

The development research covers three steps, such as compose the preliminary instrument, do trials, and do interpretation. The composing of preliminary instruments contains the description of implementation of composing instrument activities. The content validity of instrument is obtained through expert judgment. Trials are conducted to obtain evidence of instrument quality empirically. Interpretation of trial result is a step to determine whether or not a revision and further trial are necessary. Trial subjects were the students of SMP N 5 Yogyakarta, SMP Muhammadiyah 2 Yogyakarta, and SMPIT Abu Bakar Yogyakarta. The respondents of first trial were 688 students and the second trial were 696 students. The construct validity of assessment instrument for attitude aspect is obtained through Lisrel using a second order CFA, while for knowledge and skills aspects assessment instrument is obtained by Lisrel using a first order CFA. The item characteristic of assessment instrument for knowledge and skills aspects is obtained through Parscale using the IRT Polytomus GPCM. The instruments reliability includes the Omega coefficients (ω) calculated based on λ value of Lisrel and Alpha coefficients with SPSS.

Based on the research, it is produced: (a) the assessment instrument of attitude aspects in the form of a scale to measure the student's honest character in subject scaling, a scale to measure the student's responsibility character in subject scaling, and mathematical attitude scale in response scaling form, (b) the assessment instrument of knowledge and skills aspects in the form of essay test and project task. The entire instruments both theoretically and empirically are proven to be good quality. The item suitability to the indicator on the entire instruments is good by V Aiken index of 0.88 - 1.00 and the test alignment index of 0.93. The scale for measuring honest character, responsibility character, and mathematical attitude have good categorical in construct validity and reliability by Omega coefficient 0.86, 0.89, and 0.81. The test and task have good categorical in both construct validity and reliability by Alpha coefficients of 0.78 for test and 0.70 for task. All items of test and task have good categorical by item discrimination of 0.31 - 2.01 and item difficulty -2.33 to -0.47 for test and item discrimination of 0.46 - 1.11 and item difficulty -2.01 to -0.91 for task. The test give accurate information on theta -3.0 to 0.8 whereas the task on theta -3.0 to 0.5.

Key words: *instrument, assessment, statistics, junior high school, curriculum 2013*

PERNYATAAN KEASLIAN KARYA

Yang bertanda tangan di bawah ini:

Nama mahasiswa : Kana Hidayati

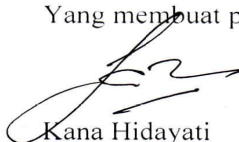
Nomor mahasiswa : 11701261010

Program studi : Penelitian dan Evaluasi Pendidikan

Dengan ini menyatakan bahwa disertasi ini merupakan hasil karya saya sendiri dan belum pernah diajukan untuk memperoleh gelar doktor di suatu perguruan tinggi, dan sepanjang pengetahuan saya dalam disertasi ini tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 23 Juli 2018

Yang membuat pernyataan.



Kana Hidayati







NIM 11701261010

LEMBAR PENGESAHAN
PENGEMBANGAN INSTRUMEN PENILAIAN
KOMPETENSI MATEMATIKA MATERI STATISTIKA SISWA SMP
BERDASARKAN KURIKULUM 2013

KANA HIDAYATI
NIM 11701261010

Dipertahankan di depan Dewan Penguji Disertasi
Program Pascasarjana Universitas Negeri Yogyakarta
Tanggal: 23 Mei 2018

DEWAN PENGUJI

Prof. Dr. Marsigit, MA. (Ketua/Penguji)		20/8 2018
Dr. Edi Istiyono (Sekretaris/Penguji)		20/8 ' 2018
Prof. Dr. Budiyo (Pembimbing Utama/Penguji)		20/8 2018
Dr. Sugiman (Pembimbing/Penguji)		16/8 2018
Dr. Heri Retnawati (Penguji)		16/8 / 2018
Prof. Kumaidi, Ph.D. (Penguji)		15/8 / 2018

Yogyakarta, **23 MAY 2018**

Program Pascasarjana
Universitas Negeri Yogyakarta
Direktur.


Prof. Dr. Marsigit, MA.
NID 19570719 198303 1 004

KATA PENGANTAR

Puji dan syukur ke hadirat Allah SWT, yang telah melimpahkan rahmat dan karunia-Nya, sehingga disertasi berjudul “Pengembangan Instrumen Penilaian Kompetensi Matematika Materi Statistika Siswa SMP Berdasarkan Kurikulum 2013” ini dapat terselesaikan. Penulis menyampaikan ucapan terima kasih yang tak terhingga kepada yang terhormat:

1. Direktur Jenderal Pendidikan Tinggi Republik Indonesia yang telah memberikan biaya pendidikan berupa BPPDN dan hibah Penelitian Disertasi Doktor (PDD) yang sangat berarti bagi penulis dalam penyelesaian disertasi ini.
2. Rektor Universitas Negeri Yogyakarta yang telah memberikan izin dan kesempatan kepada penulis untuk studi lanjut serta pemberian fasilitas bantuan studi untuk staf UNY yang sangat membantu kelancaran penyelesaian disertasi.
3. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta yang telah memberikan izin, dorongan, dan dukungan untuk menyelesaikan studi S3.
4. Direktur Program Pascasarjana Universitas Negeri Yogyakarta beserta staf, atas segala kebijakan, fasilitas, perhatian, dan dorongan sehingga penulis dapat menyelesaikan penulisan disertasi ini.
5. Bapak Prof. Dr. Badrun Kartowagiran dan Dr. Edi Istiyono selaku ketua dan sekretaris Program Studi Penelitian dan Evaluasi Pendidikan yang senantiasa memberikan arahan, kemudahan, dan dorongan semangat kepada penulis.
6. Bapak Prof. Dr. Budiyo dan Bapak Dr. Sugiman selaku pembimbing yang sangat membantu dan penuh perhatian dalam mengarahkan, membimbing, dan memberi dorongan semangat hingga disertasi ini dapat terselesaikan.
7. Bapak Prof. Dr. Marsigit, MA. dan Bapak Prof. Kumaidi, Ph.D. selaku penguji atas segala pertanyaan dan saran serta arahan yang sangat penting bagi disertasi ini agar menjadi lebih baik.

8. Ibu Dr. Heri Retnawati selaku reviewer dan penguji atas telaah kritis, pertanyaan, dan saran yang sangat membantu guna kelengkapan disertasi ini.
9. Kepala Dinas Pendidikan Kota Yogyakarta beserta jajarannya, yang telah memberikan izin dan kemudahan dalam melakukan penelitian.
10. Bapak Dr. Ali Mahmudi, Bapak Dr. Jailani, Bapak Dr. Ariyadi Wijaya, Ibu Endang Listyani, MS., Ibu Mathilda Susanti, M.Si., dan Ibu Dr. Heri Retnawati atas segala saran dan masukannya serta pemberian penilaian dalam kegiatan validasi instrumen yang dikembangkan dalam penelitian ini.
11. Bapak Dr. Suharno (Kepala SMP N 5 Yogyakarta), Ibu Nilawati Isdwiantari, S.Pd. (Kepala SMP Muhammadiyah 2 Yogyakarta), dan Bapak Herry Purwanto, S.Pd. (Kepala SMPIT Abu Bakar Yogyakarta) yang telah memberikan izin melaksanakan penelitian di sekolah.
12. Para guru dan siswa SMP N 5 Yogyakarta, SMP Muhammadiyah 2 Yogyakarta, dan SMPIT Abu Bakar Yogyakarta yang berperan aktif dan sangat membantu dalam pengambilan data penelitian ini.
13. Bapak Rifai, M.Pd., Bapak Slamet Riyadi, M.Pd., Ibu Laksmi Indrawati, S.Pd., Ibu Titin Sumarni, S.Pd., Ibu Elly Arliani, M.Si., Ibu Rosita Kusumawati, M.Si., dan Ibu Retno Subekti, M.Si. yang telah memberikan saran dan masukan pada kegiatan FGD.
14. Seluruh dosen Program Studi PEP dan segenap rekan sejawat di Jurusan Pendidikan Matematika FMIPA UNY, yang telah memberikan dukungan yang sangat berarti dalam penyelesaian disertasi ini.
15. Segenap teman-teman mahasiswa S3 Program Studi PEP yang saling memberikan motivasi dalam penyelesaian studi dan semua pihak yang tidak dapat penulis sebutkan satu per satu yang telah membantu dalam penyelesaian disertasi ini.

Secara khusus, penulis sampaikan terima kasih yang sedalam-dalamnya untuk ayah (alm) dan ibu, ayah dan ibu mertua, dan suami tercinta AR Sugeng Riyadi, M.Ud. beserta anak-anak tersayang: Ahmad Abdus Salam AR, Ahmad Badrus Salam AR, Arsyada Alfiana Kana Putri, dan Ahmad Darus Salam AR, yang

senantiasa mendoakan, mendampingi, dan memberikan dukungan penuh kepada penulis selama menempuh studi.

Semoga Allah SWT melimpahkan rahmat, nikmat, dan keberkahan kepada semua pihak tersebut di atas, atas segala bantuan yang telah diberikan kepada penulis. Semoga disertasi ini bermanfaat bagi pengembangan ilmu kependidikan khususnya di bidang pengukuran dan penilaian pencapaian kompetensi matematika materi statistika siswa SMP serta praktek penilaian pencapaian kompetensi dalam kegiatan pendidikan pada umumnya.

Yogyakarta, 23 Juli 2018

Kana Hidayati

NIM 11701261010

DAFTAR ISI

	Halaman
ABSTRAK	ii
ABSTRACT	iii
PERNYATAAN KEASLIAN.....	iv
LEMBAR PERSETUJUAN	v
KATA PENGANTAR.....	vi
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN.....	xiii
BAB I. PENDAHULUAN	
A. Latar Belakang Masalah	1
B. Identifikasi Masalah	11
C. Pembatasan Masalah	15
D. Rumusan Masalah	16
E. Tujuan Pengembangan	16
F. Spesifikasi Produk yang Dikembangkan	16
G. Manfaat Pengembangan	17
H. Asumsi Pengembangan	19
BAB II. KAJIAN PUSTAKA	
A. Kajian Teori	20
1. Hakikat Penilaian Pencapaian Kompetensi dalam Matematika ..	20
2. Penilaian Matematika SMP Berdasarkan Kurikulum 2013 dan Implementasinya pada Materi Statistika	27
3. Karakter Jujur	43
4. Karakter Tanggung Jawab	47
5. Sikap Matematis	49
6. Teknik Penilaian Diri, Tes Uraian, dan Tugas Proyek	51
7. Penyusunan Instrumen Penilaian	57
8. Rubrik	67
9. Kualitas Perangkat Penilaian.....	69
10. Teori Respons Butir untuk Data Politomus.....	86

B.	Kajian Penelitian yang Relevan	98
C.	Kerangka Pikir	102
D.	Pertanyaan Penelitian	107
BAB III.	METODE PENELITIAN	
A.	Model Pengembangan	108
B.	Prosedur Pengembangan	112
C.	Desain Uji Coba Produk	118
1.	Desain Uji Coba	118
2.	Subjek Coba	119
3.	Teknik dan Instrumen Pengumpulan Data	120
4.	Teknik Analisis Data	120
BAB IV.	HASIL PENELITIAN DAN PENGEMBANGAN	
A.	Hasil Pengembangan Produk Awal	134
B.	Hasil Uji Coba Produk	154
1.	Uji Coba I	154
2.	Uji Coba II	167
C.	Revisi Produk.....	183
D.	Kajian Produk Akhir.....	184
E.	Keterbatasan Penelitian	197
BAB V.	SIMPULAN DAN SARAN	
A.	Simpulan tentang Produk	200
B.	Saran Pemanfaatan Produk	203
C.	Diseminasi dan Pengembangan Produk Lebih Lanjut	205
DAFTAR PUSTAKA.....		207

DAFTAR TABEL

	Halaman
Tabel 1. Taksonomi dari Revisi Taksonomi Bloom.....	38
Tabel 2. Validator Instrumen Penilaian	116
Tabel 3. Kategorisasi Karakter Jujur.....	131
Tabel 4. Kategorisasi Karakter Tanggung Jawab	132
Tabel 5. Kategorisasi Sikap Matematis	132
Tabel 6. Daftar Guru dalam Kegiatan Studi Awal	136
Tabel 7. Kisi-Kisi Skala untuk Mengukur Karakter Jujur.....	142
Tabel 8. Kisi-Kisi Skala untuk Mengukur Karakter Tanggung Jawab.....	143
Tabel 9. Kisi-Kisi Skala untuk Mengukur Sikap Matematis	144
Tabel 10. Kisi-Kisi Instrumen Penilaian Aspek Pengetahuan dan Keterampilan.....	147
Tabel 11. Hasil Perhitungan Indeks V Aiken pada Skala untuk Mengukur Karakter Jujur	150
Tabel 12. Hasil Perhitungan Indeks V Aiken pada Skala untuk Mengukur Karakter Tanggung Jawab	151
Tabel 13. Hasil Perhitungan Indeks V Aiken pada Skala Sikap Matematis	151
Tabel 14. Hasil Perhitungan Indeks V Aiken pada Tes	152
Tabel 15. Hasil Perhitungan Indeks V Aiken pada Tugas	153
Tabel 16. Hasil <i>Second Order</i> CFA Skala untuk Mengukur Karakter Jujur pada Uji Coba I	156
Tabel 17. Hasil <i>Second Order</i> CFA Skala untuk Mengukur Karakter Tanggung Jawab pada Uji Coba I	157
Tabel 18. Hasil <i>Second Order</i> CFA Skala Sikap Matematis pada Uji Coba I ..	158
Tabel 19. Hasil <i>First Order</i> CFA Tes pada Uji Coba I.....	160
Tabel 20. Hasil Estimasi Parameter Butir Tes pada Uji Coba I	162
Tabel 21. Hasil <i>First Order</i> CFA Tugas pada Uji Coba I	164
Tabel 22. Hasil Estimasi Parameter Butir Tugas pada Uji Coba I	166
Tabel 23. Hasil <i>Second Order</i> CFA Skala untuk Mengukur Karakter Jujur pada Uji Coba II	170
Tabel 24. Hasil <i>Second Order</i> CFA Skala untuk Mengukur Karakter Tanggung Jawab pada Uji Coba II	173
Tabel 25. Hasil <i>Second Order</i> CFA Skala Sikap Matematis pada Uji Coba II	176
Tabel 26. Pencapaian Kompetensi Siswa pada Karakter Jujur.....	178
Tabel 27. Pencapaian Kompetensi Siswa pada Karakter Tanggung Jawab	179
Tabel 28. Pencapaian Kompetensi Siswa pada Sikap Matematis.....	180

DAFTAR GAMBAR

	Halaman
Gambar 1. Prediksi Hasil Penilaian Afektif di Sekolah	26
Gambar 2. Tiga Langkah Membangun Suatu Tes di Kelas	60
Gambar 3. Contoh Perhitungan Indeks Keselarasan Porter untuk Matriks 2x2	79
Gambar 4. Kerangka Pikir	106
Gambar 5. Model Pengembangan Instrumen	111
Gambar 6. Grafik <i>Eigen Value</i> untuk Tes	161
Gambar 7. Grafik <i>Eigen Value</i> untuk Tugas.....	165
Gambar 8. Model Pengukuran Skala untuk Mengukur Karakter Jujur pada Uji Coba II	169
Gambar 9. Model Pengukuran Skala untuk Mengukur Karakter Tanggung Jawab pada Uji Coba II	171
Gambar 10. Model Pengukuran Skala Sikap Matematis pada Uji Coba II	174
Gambar 11. Pencapaian Kompetensi Siswa pada Karakter Jujur	179
Gambar 12. Pencapaian Kompetensi Sikap pada Karakter Tanggung Jawab ...	180
Gambar 13. Pencapaian Kompetensi Siswa pada Sikap Matematis	181
Gambar 14. Pencapaian Kompetensi Pengetahuan	182
Gambar 15. Pencapaian Kompetensi Keterampilan	182

DAFTAR LAMPIRAN

		Halaman
Lampiran	1. <i>Focus Group Discussion</i> (FGD)	
	a. Hasil Pelaksanaan <i>Focus Group Discussion</i> (FGD)	220
	b. Daftar Hadir FGD	224
Lampiran	2. Validasi Instrumen	
	a. Lembar Telaah Instrumen	225
	b. Rekap Hasil Validasi Instrumen secara Kualitatif	248
	c. Data Hasil Penilaian Para Pakar terhadap Kesesuaian Butir Instrumen dengan Indikator	260
	d. Pernyataan Validator	262
Lampiran	3. Indeks Validitas Isi	
	a. Hasil Perhitungan Indeks V Aiken	268
	b. Tabel <i>Validity Coefficient</i> (V) dari Aiken	270
	c. Hasil Perhitungan Indeks Keselarasan Porter	271
Lampiran	4. Kisi-Kisi Instrumen	
	a. SKL, KI, dan KD Materi Statistika	274
	b. Kisi-Kisi Instrumen Penilaian Aspek Sikap	275
	c. Kompetensi Dasar, Indikator, <i>Abstract Continuum</i> , dan Kisi-Kisi Instrumen Penilaian Aspek Pengetahuan dan Keterampilan	278
Lampiran	5. Data Hasil Uji Coba Instrumen Penilaian Aspek Sikap	
	a. Data Hasil Uji Coba I pada Skala untuk Mengukur Karakter Jujur	280
	b. Data Hasil Uji Coba II pada Skala untuk Mengukur Karakter Jujur	284
	c. Data Hasil Uji Coba I pada Skala untuk Mengukur Karakter Tanggung Jawab	288
	d. Data Hasil Uji Coba II pada Skala untuk Mengukur Karakter Tanggung Jawab Siswa	292
	e. Data Hasil Uji Coba I pada Skala Sikap Matematis	296
	f. Data Hasil Uji Coba II pada Skala Sikap Matematis	300

Lampiran	6.	Data Hasil Uji Coba Instrumen Penilaian Aspek Pengetahuan dan Keterampilan	
	a.	Data Hasil Uji Coba Tes	304
	b.	Data Hasil Uji Coba Tugas	307
Lampiran	7.	Hasil Analisis Data Instrumen Penilaian Aspek Sikap	
	a.	Hasil Lisrel untuk Uji <i>Multivariate Normality</i> pada Data Uji Coba I dan Uji Coba II Skala untuk Mengukur Karakter Jujur	310
	b.	Hasil Lisrel untuk <i>Second Order</i> CFA pada Data Uji Coba I Skala untuk Mengukur Karakter Jujur	312
	c.	Hasil Lisrel untuk <i>Second Order</i> CFA pada Data Uji Coba II Skala untuk Mengukur Karakter Jujur	319
	d.	Hasil Lisrel untuk Uji <i>Multivariate Normality</i> pada Data Uji Coba I dan Uji Coba II Skala untuk Mengukur Karakter Tanggung Jawab	342
	e.	Hasil Lisrel untuk <i>Second Order</i> CFA pada Data Uji Coba I Skala untuk Mengukur Karakter Tanggung Jawab	344
	f.	Hasil Lisrel untuk <i>Second Order</i> CFA pada Data Uji Coba II Skala untuk Mengukur Karakter Tanggung Jawab	350
	g.	Hasil Lisrel untuk Uji <i>Multivariate Normality</i> pada Data Uji Coba I dan Uji Coba II Skala Sikap Matematis	364
	h.	Hasil Lisrel untuk <i>Second Order</i> CFA pada Data Uji Coba I Skala Sikap Matematis	366
	i.	Hasil Lisrel untuk <i>Second Order</i> CFA pada Data Uji Coba II Skala Sikap Matematis	373
Lampiran	8.	Koefisien Reliabilitas Instrumen Penilaian Aspek Sikap	
	a.	Hasil Perhitungan Koefisien Reliabilitas Omega pada Skala untuk Mengukur Karakter Jujur	395
	b.	Hasil Perhitungan Koefisien Reliabilitas Omega pada Skala untuk Mengukur Karakter Tanggung Jawab	396
	c.	Hasil Perhitungan Koefisien Reliabilitas Omega pada Skala Sikap Matematis	397
Lampiran	9.	Hasil Analisis Data Instrumen Penilaian Aspek Pengetahuan dan Keterampilan	
	a.	Hasil Lisrel untuk Uji <i>Multivariate Normality</i> pada Data Uji Coba Tes	398
	b.	Hasil Lisrel untuk <i>First Order</i> CFA pada Data Uji Coba Tes..	399
	c.	Hasil Lisrel untuk Uji <i>Multivariate Normality</i> pada Data Uji Coba Tugas	410
	d.	Hasil Lisrel untuk <i>First Order</i> CFA pada Data Uji Coba Tugas	411

Lampiran	10. Koefisien Reliabilitas Instrumen Penilaian Aspek Pengetahuan dan Keterampilan	
	a. Hasil Analisis SPSS untuk Koefisien Reliabilitas Alpha dan Perhitungan Koefisien CR pada Data Uji Coba Tes	416
	b. Hasil Analisis SPSS untuk koefisien Reliabilitas Alpha dan Perhitungan Koefisien CR pada Data Uji Coba Tugas	417
Lampiran	11. Hasil Analisis <i>Parscale</i> pada Instrumen Penilaian Aspek Pengetahuan dan Keterampilan	
	a. Hasil Analisis <i>Parscale</i> pada Data Uji Coba Tes	418
	b. Hasil Analisis <i>Parscale</i> pada Data Uji Coba Tugas	436
Lampiran	12. Hasil Pencapaian Kompetensi Siswa	
	a. Pencapaian Kompetensi Aspek Sikap	445
	b. Pencapaian Kompetensi Aspek Pengetahuan dan Keterampilan	448
Lampiran	13. Surat Ijin Penelitian	450
Lampiran	14. Surat Keterangan Kepala Sekolah Tempat Pelaksanaan Uji Coba Instrumen	451

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Sistem pendidikan di Indonesia menetapkan matematika sebagai salah satu mata pelajaran yang diberikan kepada semua siswa pada pendidikan formal mulai dari jenjang sekolah dasar hingga perguruan tinggi. Pembelajaran matematika di Indonesia, hingga saat ini terus berkembang mengikuti perkembangan berbagai teori belajar, teknologi informasi dan komunikasi, serta tuntutan dalam kehidupan sosial. Pembelajaran matematika di sekolah dimaksudkan untuk membekali siswa dengan kemampuan berpikir kritis, logis, sistematis, analitis, dan kreatif agar siswa memiliki kemampuan bertahan hidup pada kondisi jaman yang penuh ketidakpastian, kompetitif, dan terus mengalami perubahan.

Salah satu komponen penting yang mempengaruhi keberhasilan pencapaian tujuan pembelajaran matematika di sekolah adalah penilaian yang dilakukan. Penilaian dalam pembelajaran matematika dilakukan bukan sekedar untuk menentukan skor melainkan untuk meningkatkan kualitas belajar siswa. Selain itu, penilaian dalam pembelajaran matematika semestinya juga dilakukan sesuai dengan karakteristik matematika. Menurut NCTM (*National Council of Teachers of Mathematics*), penilaian dalam pembelajaran matematika merupakan proses

menghimpun keterangan tentang pengetahuan matematika siswa, kecakapan siswa dalam menggunakan matematika, sikap siswa terhadap matematika, serta membuat kesimpulan dari berbagai bukti untuk berbagai kepentingan (NCTM, 1995: 3). Sejalan dengan pengertian tersebut, Van de Walle (2007: 82-83) mengemukakan bahwa secara keseluruhan hal-hal yang dinilai dalam pembelajaran matematika meliputi konsep dan prosedur (pemahaman konsep dan keterampilan matematika), proses matematika (kemampuan *problem solving*), serta sikap dan keyakinan siswa mengenai matematika. Secara garis besar, hal ini menunjukkan bahwa penilaian pencapaian kompetensi siswa dalam pembelajaran matematika mencakup aspek sikap, pengetahuan, dan keterampilan yang saat ini berusaha diwujudkan pemerintah melalui pelaksanaan Kurikulum 2013.

Kurikulum 2013 merupakan kurikulum terbaru yang secara bertahap mulai dilaksanakan pada tahun pelajaran 2013/2014 untuk kelas I dan IV Sekolah Dasar/Madrasah Ibtida'iyah (SD/MI), Kelas VII Sekolah Menengah Pertama/Madrasah Tsanawiyah (SMP/MTs), dan kelas X Sekolah Menengah Atas/Sekolah Menengah Kejuruan/Madrasah Aliyah (SMA/SMK/MA). Berdasarkan surat Nomor: 179342/MPK/KR/2014 tanggal 5 Desember 2014, pemerintah menetapkan berbagai kebijakan di antaranya adalah tetap menerapkan Kurikulum 2013 di sekolah-sekolah yang telah tiga semester menerapkan, yaitu sejak tahun pelajaran 2013/2014 dan menjadikan sekolah-sekolah tersebut sebagai sekolah pengembangan dan percontohan penerapan Kurikulum 2013. Kebijakan ini dipertegas

dengan dikeluarkannya Peraturan Menteri Pendidikan dan Kebudayaan (Permendikbud) Nomor 160 Tahun 2014 tanggal 11 Desember 2014 tentang pemberlakuan Kurikulum tahun 2006 dan Kurikulum 2013. Kebijakan pemerintah tersebut menunjukkan bahwa Kurikulum 2013 tetap diterapkan pada sekolah-sekolah sasaran yang telah menggunakan Kurikulum 2013 mulai tahun pelajaran 2013/2014. Melalui pelaksanaan Kurikulum 2013, pemerintah berkomitmen untuk terus berupaya melakukan berbagai perbaikan termasuk di antaranya dalam penilaian pencapaian kompetensi siswa.

Berdasarkan hasil refleksi pelaksanaan Kurikulum 2013 yang dilakukan Direktorat Pembinaan Sekolah Menengah Pertama (SMP) pada tahun 2014 menunjukkan bahwa berbagai masalah telah terjadi terkait dengan penilaian. Berdasarkan laporan kuantitatif, sekitar 34% guru yang pernah dilatih kurang memahami penilaian berdasarkan Kurikulum 2013. Secara lebih rinci, sekitar 22% guru kesulitan dalam penilaian aspek sikap dan 14% guru kesulitan dalam penilaian aspek keterampilan (Kemendikbud, 2014c). Selain itu, lebih dari 50% guru menyatakan belum dapat merancang, melaksanakan, dan mengolah hasil penilaian dengan baik khususnya pada penilaian aspek sikap. Para guru juga banyak yang masih kurang percaya diri dalam melakukan penilaian aspek keterampilan karena belum sepenuhnya memahami cara menyusun instrumen dan rubrik penilaian keterampilan (Kemendikbud, 2015). Hal ini menunjukkan bahwa terkait dengan

penilaian, para guru masih sangat memerlukan panduan dan berbagai referensi baik terkait cara melakukan penilaian maupun contoh instrumen yang digunakan.

Hasil refleksi terkait penilaian yang dilakukan pemerintah di atas, ternyata juga sesuai dengan hasil penelitian Retnawati (2015) tentang hambatan guru matematika SMP di DIY dalam menerapkan Kurikulum 2013. Hasil penelitian menunjukkan bahwa hambatan yang dialami para guru terkait penilaian di antaranya adalah guru kesulitan merencanakan penilaian aspek sikap, memilah aspek pengetahuan dan keterampilan, sistem penilaian yang digunakan dirasakan terlalu rumit, dan guru juga kurang yakin dengan kesahihan dan objektivitas penilaian yang dilakukan. Hasil penelitian ini pun senada dengan hasil wawancara yang dilakukan pada para guru matematika SMP di DIY yang telah menerapkan Kurikulum 2013 mulai tahun pelajaran 2013/2014.

Berdasarkan hasil wawancara diperoleh informasi bahwa para guru masih kesulitan melakukan penilaian khususnya untuk aspek sikap dan keterampilan walaupun telah ada contoh dari pemerintah. Penilaian yang mesti dilakukan dirasakan para guru menjadi lebih banyak dan referensi instrumen penilaian dari pemerintah ada yang kurang tepat atau bahasanya ada yang sulit dipahami siswa. Terjadinya perubahan kebijakan menjadikan batasan materi terasa tidak jelas sehingga penilaiannya juga menjadi kurang terarah. Para guru juga menyatakan masih sangat memerlukan referensi terkait contoh instrumen penilaian berdasarkan Kurikulum 2013. Adapun terkait buku panduan dari pemerintah, menurut para guru banyak

memuat soal kemampuan berpikir tingkat tinggi sementara para guru masih memerlukan contoh soal untuk mengukur kemampuan berpikir tingkat menengah. Kesulitan ini dialami para guru hampir pada semua materi termasuk statistika.

Materi statistika merupakan salah satu materi yang mengalami perubahan dalam pelaksanaan Kurikulum 2013. Materi ini dalam pembelajaran matematika SMP pada Kurikulum Tingkat Satuan Pendidikan (KTSP) hanya diajarkan pada kelas XI, namun berdasarkan Kurikulum 2013 mulai diajarkan pada kelas VII dan VIII. Perubahan ini tentunya juga berdampak pada kesiapan penilaian yang dilakukan guru. Para guru mengungkapkan bahwa masih sangat memerlukan referensi terkait instrumen penilaian untuk mengukur pencapaian kompetensi aspek sikap, pengetahuan, dan keterampilan yang sesuai dengan Kurikulum 2013, mudah digunakan para guru, dan telah teruji kualitasnya.

Mengacu pada hasil refleksi yang telah dilakukan, berbagai perubahan telah dilakukan pemerintah dengan begitu cepat dalam pelaksanaan Kurikulum 2013. Berdasarkan revisi Kurikulum 2013 yang diterapkan mulai tahun pelajaran 2016/2017, pemerintah melakukan penyederhanaan kegiatan penilaian yang harus dilakukan para guru. Kegiatan penilaian dibuat menjadi lebih sederhana, mudah dilakukan, tidak saling tumpang tindih, tidak membebani para guru dan siswa, namun tetap berdasarkan prinsip dan kaidah penilaian.

Penyederhanaan tersebut di antaranya adalah bahwa penilaian untuk aspek sikap (sosial dan spiritual) tidak wajib dilakukan semua guru mata pelajaran melainkan cukup dilakukan oleh guru PPKn dan guru pendidikan agama. Penilaian

akhir sikap siswa dilakukan oleh wali kelas, sedangkan guru mata pelajaran selain PPKn dan pendidikan agama hanya menilai aspek akademik sesuai mata pelajaran yang diajarkan saja (Kemendikbud, 2015). Berpijak pada kebijakan terbaru ini, guru mata pelajaran matematika tidak wajib melakukan penilaian aspek sikap. Namun demikian, mengingat hakikat penilaian dalam pembelajaran matematika sebaiknya para guru matematika tetap menilai aspek sikap walaupun sewajarnya saja baik dalam rangka memperkuat karakter siswa melalui pembelajaran matematika maupun untuk menambah referensi atau masukan kepada sekolah terkait penilaian sikap siswa.

Penilaian sikap siswa dalam pembelajaran matematika tetap sangat penting diperhatikan utamanya berkaitan dengan karakter atau akhlak mulia pada siswa sesuai tujuan pendidikan nasional bangsa Indonesia. Menurut Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional pasal 1 ayat 1 disebutkan bahwa pendidikan adalah usaha sadar dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar siswa secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa dan negara. Berdasarkan undang-undang tersebut, menunjukkan bahwa pendidikan di Indonesia harus memuat pendidikan karakter.

Pendidikan karakter semestinya juga terintegrasi dalam penilaian matematika. Artinya, melalui penilaian dalam pembelajaran matematika juga diungkap karakter siswa sehingga sangat membantu para guru mengambil langkah dalam upaya memperkuat karakter baik pada siswa. Kumaidi (2014: 8-9) mengemukakan bahwa

Kurikulum 2013 menuntut guru tidak sekedar menilai penguasaan materi ajar melainkan lebih dari itu mencakup perkembangan afeksi seperti keberimanan, kejujuran, kedisiplinan, dan karakter baik lainnya. Hal ini menunjukkan bahwa penilaian aspek sikap dalam Kurikulum 2013 lebih dimaksudkan sebagai penguatan karakter siswa.

Penilaian yang berkualitas sangat diperlukan guna tercapainya tujuan pembelajaran matematika di sekolah. Melalui penilaian yang berkualitas, pencapaian kompetensi siswa dapat dinilai dengan baik dan memenuhi akurasi serta akuntabilitas hasil penilaian. Penilaian berkualitas memerlukan instrumen penilaian yang berkualitas pula. Anderson & Morgan (2008: xi) menyatakan bahwa kualitas penilaian pendidikan tergantung pada kualitas instrumen yang digunakan. Hal ini menunjukkan bahwa kualitas instrumen harus benar-benar diperhatikan agar diperoleh hasil penilaian yang berkualitas dan dapat dipertanggungjawabkan.

Terkait dengan kualitas instrumen penilaian, hingga saat ini secara umum digunakan kriteria validitas dan reliabilitas. Selain itu, khusus untuk instrumen penilaian aspek pengetahuan semestinya juga didukung terpenuhinya kriteria sebagai instrumen penilaian yang baik terkait parameter butir, fungsi informasi, dan kesalahan baku pengukurannya. Penggunaan instrumen penilaian yang valid, reliabel, dan memenuhi kriteria sebagai instrumen yang baik akan memberikan hasil penilaian yang akuntabel dan akurat. Hal ini dikarenakan adanya kesesuaian antara apa yang dipelajari siswa berdasarkan standar dalam kurikulum dengan apa yang muncul dari siswa pada penilaian serta adanya konsistensi hasil penilaian.

Pembuktian validitas suatu instrumen pada umumnya meliputi validitas isi dan validitas konstruk. Bukti validitas isi dilakukan dengan analisis rasional melalui *expert judgment* dan bukti validitas konstruk dengan analisis faktor. Mengacu pada perkembangan teori pengukuran hingga saat ini, menunjukkan bahwa validitas isi suatu instrumen penilaian juga dapat diperoleh melalui studi keselarasan (*alignment*) antara penilaian dengan standar dalam kurikulum. Ananda (2003: 2) dan Bholá, Impara, & Buckendahl (2003: 21) menyatakan bahwa hasil uji keselarasan dapat digunakan sebagai bukti validitas isi. Hal ini berarti bahwa hasil studi keselarasan dapat dijadikan sebagai bukti validitas isi yang mendukung interpretasi skor hasil penilaian.

Berdasarkan penelitian di beberapa negara telah dihasilkan sejumlah metode *alignment* yang terus berkembang. Beberapa metode tersebut di antaranya adalah metode analisis Webb (Webb, 1997), metode La Marca (La Marca et al., 2000), revisi taksonomi Bloom (Anderson & Krathwohl, 2001), *Survey of Enacted Curriculum/SEC* (Porter & Smithson, 2001), dan Proyek 2061 dari *American Association for the Advancement of Science/AAAS* (Stern & Algrehn, 2002). Penggunaan berbagai metode uji keselarasan tersebut pada umumnya dipilih berdasarkan relevansinya dengan karakteristik mata pelajaran dan kurikulum yang digunakan negara yang bersangkutan. Modifikasi berbagai metode juga berkembang dalam sejumlah penelitian di berbagai negara. Secara umum metode-metode uji keselarasan tersebut masih terbatas penggunaannya pada instrumen penilaian pengetahuan.

Selain memperhatikan berbagai kriteria di atas, instrumen penilaian yang berkualitas tentunya akan diperoleh apabila disusun melalui langkah-langkah yang tepat dan dapat dipertanggungjawabkan. Berbagai teknik penyusunan instrumen telah banyak dikemukakan para ahli dan berbagai penelitian pengembangan instrumen dalam pembelajaran matematika juga cukup banyak dilakukan. Namun memperhatikan karakteristik materi matematika yang khas maka khusus untuk pembelajaran matematika masih perlu dikembangkan dan dikaji lebih lanjut.

Instrumen penilaian yang berkualitas juga ditentukan oleh ketepatan memilih teknik penilaian dan bentuk instrumen yang digunakan selain memperhatikan beberapa kriteria kualitas instrumen dan langkah-langkah penyusunannya. Teknik penilaian untuk aspek sikap yang umumnya digunakan dalam pembelajaran matematika adalah observasi, wawancara, penilaian diri, dan penilaian antar teman dengan bentuk instrumen berupa lembar observasi, pedoman wawancara, dan angket dalam bentuk penskalaan respons. Pengembangan instrumen penilaian aspek sikap menggunakan penskalaan subjek yang langsung terkait dengan materi dalam matematika masih sangat jarang. Padahal, instrumen berbentuk penskalaan subjek memiliki nilai praktis yang tinggi dan banyak digunakan para perancang skala psikologi (Azwar, 2005: 32). Mengingat bahwa penilaian aspek sikap sangat terkait dengan karakter maka penggunaan skala psikologi berbentuk penskalaan subjek sangat relevan digunakan dalam rangka mengukur karakter siswa.

Penilaian aspek pengetahuan dan keterampilan dalam pembelajaran matematika perlu memperhatikan bahwa keduanya saling terkait. Oleh sebab itu,

pemilihan teknik penilaian dan bentuk instrumen dapat dirancang secara bersamaan. Teknik penilaian untuk aspek pengetahuan dan keterampilan yang umumnya digunakan dalam pembelajaran matematika adalah tes, penugasan kinerja atau proyek, dan portofolio. Bentuk instrumen penilaian pengetahuan dengan teknik tes dapat berupa tes objektif dan tes uraian sedangkan untuk penilaian keterampilan dapat dilakukan dalam bentuk penugasan kinerja atau proyek baik bersifat individu atau kelompok. Penggunaan bentuk tes uraian dan tugas kinerja atau proyek dalam pembelajaran matematika sangat sesuai dipilih para guru dalam melakukan penilaian aspek pengetahuan dan keterampilan khususnya yang telah menerapkan Kurikulum 2013. Namun, pada umumnya kualitas tes dan tugas baik yang dibuat para guru maupun contoh dalam buku panduan belum sepenuhnya didukung bukti kualitas instrumen secara empiris. Walaupun berbagai penelitian pengembangan instrumen berupa tes uraian atau tugas telah disertai bukti kualitas secara empiris namun belum banyak yang menerapkan penggunaan teori pengukuran yang relevan seperti *Item Response Theory* (IRT) untuk data politomus.

Berdasarkan berbagai uraian di atas, penelitian ini dimaksudkan untuk mengembangkan instrumen penilaian yang dapat mengukur dengan tepat pencapaian kompetensi siswa pada aspek sikap, pengetahuan, dan keterampilan pada mata pelajaran matematika. Instrumen penilaian yang dikembangkan difokuskan pada materi statistika SMP berdasarkan Kurikulum 2013. Instrumen penilaian hasil pengembangan dalam penelitian ini secara teoretis dan empiris teruji kualitasnya dan terbukti berkualitas baik. Kriteria kualitas instrumen penilaian meliputi validitas isi

melalui *expert judgment* untuk semua instrumen dan khusus instrumen penilaian aspek pengetahuan dilengkapi dengan bukti validitas isi melalui uji keselarasan. Bukti empiris kualitas instrumen penilaian pengetahuan dan keterampilan juga ditinjau dari parameter butir, fungsi informasi, dan kesalahan baku pengukurannya yang mengacu pada IRT politomus. Pengujian validitas konstruk dalam penelitian ini dilakukan untuk semua instrumen dengan menggunakan teknik analisis faktor konfirmatori (*Confirmatory Factor Analysis/CFA*). Selain itu, bukti kualitas instrumen yang dikembangkan juga dilengkapi dengan menghitung koefisien reliabilitasnya baik reliabilitas konstruk maupun reliabilitas instrumen ditinjau dari konsistensi internalnya.

B. Identifikasi Masalah

Berdasarkan latar belakang masalah di atas, dapat diidentifikasi berbagai permasalahan sebagai berikut.

1. Mengacu pada tujuan pendidikan di Indonesia dan hakikat penilaian matematika maka penilaian pencapaian kompetensi siswa pada mata pelajaran matematika semestinya dilakukan secara menyeluruh pada aspek sikap, pengetahuan, dan keterampilan. Walaupun penilaian aspek sikap pada penerapan Kurikulum 2013 akhirnya tidak diwajibkan bagi guru mata pelajaran selain PPKn dan pendidikan agama, sebaiknya para guru matematika tetap melakukan penilaian aspek sikap terutama dalam rangka penguatan karakter siswa. Kenyataannya hingga saat ini,

para guru matematika masih mengalami kesulitan dalam melakukan penilaian khususnya untuk aspek sikap. Para guru juga masih ragu dengan kualitas instrumen yang disusunnya baik untuk aspek sikap, pengetahuan, dan keterampilan. Permasalahan ini juga dialami oleh para guru matematika SMP di Kota Yogyakarta yang menerapkan Kurikulum 2013. Oleh sebab itu, adanya pengembangan instrumen penilaian pencapaian kompetensi matematika siswa untuk aspek sikap, pengetahuan, dan keterampilan tentunya sangat membantu para guru matematika dalam tugasnya melakukan penilaian.

2. Instrumen penilaian yang digunakan para guru matematika yang menggunakan Kurikulum 2013 pada umumnya mengacu pada panduan dan contoh dari pemerintah. Hal ini tentunya masih terbatas ketersediaannya termasuk untuk materi statistika. Ketersediaan berbagai referensi terkait instrumen penilaian pencapaian kompetensi siswa pada materi statistika untuk aspek sikap, pengetahuan, dan keterampilan yang relevan dengan Kurikulum 2013 kenyataannya masih diperlukan para guru matematika yang menerapkan Kurikulum 2013. Oleh sebab itu, tambahan referensi terkait contoh instrumen penilaian untuk aspek sikap, pengetahuan, dan keterampilan pada materi statistika yang terbukti berkualitas baik dan bersifat fleksibel untuk digunakan sangat bermanfaat bagi para guru matematika.
3. Materi statistika sangat terkait dengan data dalam kehidupan sehari-hari sehingga melalui penilaian pada materi ini juga dapat diintegrasikan penguatan karakter

siswa khususnya karakter jujur dan tanggung jawab. Oleh sebab itu, penilaian pencapaian kompetensi pada materi ini, selain difokuskan pada aspek pengetahuan dan keterampilan juga dapat dilakukan penilaian aspek sikap baik terkait dengan sikap matematis dalam pembelajaran matematika maupun karakter jujur dan tanggung jawab siswa. Kenyataannya, saat ini belum banyak referensi instrumen penilaian yang secara khusus mengukur karakter jujur, tanggung jawab, dan sikap matematis siswa.

4. Teknik penilaian aspek sikap yang umumnya digunakan para guru dalam pembelajaran matematika adalah observasi, wawancara, penilaian diri, dan penilaian antar teman. Bentuk instrumen yang biasa digunakan adalah lembar observasi, pedoman wawancara, dan angket. Penilaian aspek sikap dengan skala psikologi pada umumnya masih menggunakan model penskalaan respons. Skala psikologi yang disusun menggunakan model penskalaan subjek masih jarang digunakan. Padahal, skala psikologi dengan model penskalaan subjek memiliki nilai praktis yang tinggi karena dapat mengungkap kondisi psikologi siswa secara lebih mendalam. Oleh sebab itu perlu dikembangkan instrumen penilaian aspek sikap menggunakan model penskalaan subjek yang disesuaikan dengan muatan materi dalam pembelajaran matematika.
5. Referensi yang ada untuk instrumen penilaian aspek pengetahuan dan keterampilan seperti tes uraian dan tugas, belum banyak yang disertai bukti kualitas instrumen secara empiris menggunakan teori pengukuran yang sesuai

untuk data politomus yakni teori respons butir politomus. Oleh sebab itu, pengembangan tes uraian dan tugas menggunakan pendekatan teori respons butir politomus perlu dilakukan. Hal ini sangat berguna dalam menambah referensi terkait instrumen penilaian pencapaian kompetensi siswa untuk aspek pengetahuan dan keterampilan yang secara empiris terbukti berkualitas baik dengan analisis menggunakan teori pengukuran yang tepat.

6. Pembuktian kualitas suatu instrumen penilaian secara umum ditinjau dari validitas isi, validitas konstruk, dan reliabilitasnya serta karakteristik butirnya terutama untuk instrumen penilaian aspek pengetahuan dan keterampilan. Terkait dengan validitas isi biasanya dilakukan dengan *expert judgment* dan didukung dengan perhitungan indeks validitas isi V Aiken. Perkembangan teori pengukuran menunjukkan bahwa bukti validitas isi suatu instrumen penilaian aspek pengetahuan dapat diperoleh melalui uji keselarasan. Hasil uji keselarasan memperkuat bukti validitas isi yang menunjukkan kesesuaian butir-butir instrumen dengan standar dalam kurikulum yang digunakan. Namun, hingga saat ini masih sangat sedikit informasi mengenai studi keselarasan antara penilaian dengan standar yang ditetapkan pemerintah dalam kurikulum di Indonesia khususnya dalam pembelajaran matematika. Oleh sebab itu, pembuktian kualitas instrumen dengan uji keselarasan perlu dilakukan karena sangat bermanfaat terutama dalam memperkuat bukti validitas isi suatu instrumen penilaian aspek pengetahuan yang dikembangkan.

C. Pembatasan Masalah

Mengingat berbagai pertimbangan, masalah dalam penelitian ini dibatasi pada pengembangan instrumen penilaian pencapaian kompetensi siswa untuk aspek sikap, pengetahuan, dan keterampilan pada mata pelajaran matematika SMP. Pengembangan instrumen penilaian untuk aspek sikap, pengetahuan, dan keterampilan dilakukan pada materi statistika kelas VII berdasarkan Kurikulum 2013.

Penilaian aspek sikap dalam penelitian ini dibatasi pada karakter jujur, karakter tanggung jawab, dan sikap matematis siswa. Hal ini mengingat bahwa materi statistika sangat terkait dengan data baik hasil pengamatan maupun pengukuran dalam kehidupan sehari-hari sehingga karakter jujur dan tanggung jawab sangat relevan untuk dinilai melalui materi ini. Penilaian terhadap sikap matematis perlu dilakukan kaitannya dengan materi matematika pada umumnya karena sikap matematis siswa juga sangat berpengaruh pada hasil belajar siswa.

Penilaian sikap dilakukan dengan teknik penilaian diri (*self assessment*) dengan instrumen berbentuk skala psikologis karena bersifat lebih praktis dan fleksibel untuk digunakan para guru. Selain itu, skala psikologis sangat memperhatikan perbedaan antar individu sehingga dapat mengungkap karakter siswa secara lebih mendalam. Adapun untuk instrumen penilaian pengetahuan dibatasi pada tes uraian dan tugas proyek individu. Tes bentuk uraian dan tugas proyek individu dipilih dengan pertimbangan bahwa tes dan tugas ini mampu mengungkap pengetahuan dan keterampilan siswa dalam materi statistika secara lebih detail dan mendalam pada tiap individu.

D. Rumusan Masalah

Berdasarkan uraian di atas, dirumuskan masalah sebagai berikut.

1. Bagaimanakah instrumen penilaian yang dapat mengukur dengan tepat pencapaian kompetensi matematika siswa untuk aspek sikap, pengetahuan, dan keterampilan pada materi statistika SMP berdasarkan Kurikulum 2013?
2. Bagaimanakah kualitas instrumen penilaian yang dihasilkan?
3. Bagaimanakah pencapaian kompetensi siswa berdasarkan instrumen penilaian yang dihasilkan tersebut?

E. Tujuan Pengembangan

Tujuan pengembangan dalam penelitian ini adalah sebagai berikut.

1. Menghasilkan instrumen penilaian yang dapat mengukur dengan tepat pencapaian kompetensi matematika siswa untuk aspek sikap, pengetahuan, dan keterampilan pada materi statistika SMP berdasarkan Kurikulum 2013.
2. Mendeskripsikan kualitas instrumen penilaian yang dihasilkan.
3. Mendeskripsikan pencapaian kompetensi siswa berdasarkan instrumen penilaian yang dihasilkan.

F. Spesifikasi Produk yang Dikembangkan

Produk utama yang dihasilkan dari penelitian ini adalah instrumen penilaian pencapaian kompetensi siswa pada mata pelajaran matematika materi statistika SMP

berdasarkan Kurikulum 2013 untuk aspek sikap, pengetahuan dan keterampilan. Secara keseluruhan, instrumen penilaian pencapaian kompetensi tersebut secara teoretis dan empiris terbukti berkualitas baik.

Instrumen penilaian aspek sikap yang dihasilkan berupa skala untuk mengukur karakter jujur, skala untuk mengukur karakter tanggung jawab, dan skala sikap matematis. Skala untuk mengukur karakter jujur dan tanggung jawab siswa berbentuk penskalaan subjek yang terkait langsung dengan materi statistika. Skala sikap matematis berbentuk penskalaan respons yang terkait dengan materi matematika secara umum. Adapun instrumen penilaian aspek pengetahuan dan keterampilan berupa tes uraian dan tugas proyek individu. Tes uraian dan tugas proyek individu dikembangkan dengan pendekatan IRT politomus menggunakan *Generalized Partial Credit Model* (GPCM) dan khusus untuk tes terdiri atas butir-butir yang mengacu pada revisi taksonomi Bloom.

G. Manfaat Pengembangan

Secara teoretis, penelitian ini memberikan manfaat dan sumbangan di bidang pengukuran dan penilaian pendidikan. Penyusunan instrumen penilaian yang dideskripsikan dalam penelitian ini, secara teoritis dapat menjadi referensi terkait teknik penyusunan instrumen penilaian khususnya dalam pembelajaran matematika. Adapun terkait dengan kegiatan analisisnya, secara teoritis juga dapat menjadi referensi bagi para guru matematika dan praktisi pendidikan pada umumnya terutama terkait analisis guna menentukan bukti kualitas suatu instrumen penilaian yang

meliputi validitas isi, validitas konstruk, reliabilitas, dan penentuan karakteristik butir. Hal ini khususnya terkait tinjauan analisis parameter butir, fungsi informasi, dan kesalahan baku pengukuran dengan pendekatan IRT Politomus menggunakan GPCM serta adanya bukti validitas isi berupa uji keselarasan antara penilaian dengan standar dalam Kurikulum 2013.

Adapun secara praktis, manfaat penelitian ini di antaranya adalah:

1. Bagi guru matematika, produk berupa instrumen penilaian yang dihasilkan dalam penelitian ini dapat menjadi tambahan referensi terkait contoh instrumen penilaian aspek sikap, pengetahuan, dan keterampilan khususnya pada materi statistika SMP berdasarkan Kurikulum 2013.
2. Bagi siswa, instrumen penilaian yang dihasilkan dari penelitian ini dapat mengungkap pencapaian kompetensi siswa pada mata pelajaran matematika khususnya materi statistika secara tepat pada aspek sikap, pengetahuan, dan keterampilan.
3. Bagi pengambil kebijakan bidang pendidikan, hasil penelitian dapat digunakan sebagai bahan pertimbangan terkait kegiatan pengembangan instrumen penilaian aspek sikap, pengetahuan, dan keterampilan yang sesuai dengan Kurikulum 2013 khususnya dalam pendidikan matematika.
4. Bagi peneliti lain, hasil penelitian ini dapat dijadikan sebagai referensi pendukung dalam rangka penelitian lanjutan (pengembangan) ataupun penelitian sejenis (perluasan) terutama terkait pengembangan instrumen penilaian pencapaian kompetensi matematika siswa untuk aspek sikap, pengetahuan, dan

keterampilan. Selain itu, secara lebih khusus juga terkait bentuk instrumen penilaian berupa skala psikologi menggunakan model penskalaan subjek, pembuktian validitas isi melalui uji keselarasan antara penilaian dengan standar dalam kurikulum yang digunakan, serta pengembangan tes uraian dengan pendekatan teori respons butir politomus menggunakan *Generalized Partial Credit Model* (GPCM).

H. Asumsi Pengembangan

Pengembangan instrumen penilaian dalam penelitian ini didasarkan pada asumsi sebagai berikut.

1. Kemampuan siswa SMP di DIY berdistribusi normal dan terletak pada rentang negatif tak terhingga sampai positif tak terhingga.
2. Indikator-indikator kompetensi yang ditetapkan untuk instrumen penilaian aspek pengetahuan dan keterampilan dapat diukur.
3. Konstruk karakter jujur, karakter tanggung jawab, dan sikap matematis didukung oleh kajian teori yang relevan.
4. Pada butir yang mengikuti pendekatan IRT politomus menggunakan GPCM, siswa yang berkemampuan lebih tinggi memperoleh skor lebih tinggi daripada siswa yang berkemampuan lebih rendah.

BAB II

KAJIAN PUSTAKA

A. Kajian Teori

1. Hakikat Penilaian Pencapaian Kompetensi dalam Matematika

Penilaian merupakan komponen penting yang menunjang keberhasilan pencapaian tujuan pembelajaran matematika di sekolah. Berbagai definisi tentang penilaian telah dikemukakan para ahli. Popham (2003: xi) menyatakan bahwa penilaian adalah proses pengumpulan informasi untuk mengambil keputusan. Senada dengan pendapat tersebut, Overton (2008: 62) juga menyatakan bahwa penilaian adalah proses pengumpulan informasi untuk memonitor kemajuan siswa dan pengambilan keputusan dalam bidang pendidikan. Hal ini juga dipertegas Kizlik (2009: 7) yang menyatakan bahwa penilaian adalah proses pengumpulan informasi berkaitan dengan tujuan pembelajaran. Nitko & Brookhart (2011: 6) mengemukakan bahwa penilaian merupakan proses mendapatkan informasi yang digunakan untuk pengambilan keputusan tentang siswa, kurikulum, program dan sekolah serta kebijakan pendidikan. Berdasarkan beberapa pendapat di atas, dapat disimpulkan bahwa penilaian merupakan kegiatan mengumpulkan dan mengolah informasi mengenai kemajuan belajar siswa yang digunakan sebagai tolok ukur ketercapaian tujuan pembelajaran dan pengambilan keputusan tentang siswa dalam kegiatan belajarnya.

Terkait dengan penilaian dalam matematika, menurut NCTM (2000: 22) disebutkan bahwa penilaian dalam matematika seharusnya mendukung proses pembelajaran matematika dan memberi informasi yang berguna bagi guru dan siswa. Secara lebih detail, de Lange (2000: 102) mengemukakan lima prinsip utama dalam penilaian matematika yaitu: (1) penilaian ditujukan untuk meningkatkan kualitas belajar dan proses pembelajaran, (2) penilaian harus dirancang sedemikian rupa sehingga siswa benar-benar mendemonstrasikan apa yang mereka kuasai, (3) penilaian bersifat operasional guna mencapai tujuan pembelajaran matematika, (4) kualitas alat penilaian tidak dilihat pada mudahnya pemberian skor secara objektif, dan (5) alat penilaian bersifat praktis. Selain itu, perkembangan teknologi juga dapat dimanfaatkan para guru matematika untuk kegiatan penilaian karena teknologi dapat didesain dan dimanfaatkan guna memberikan respons cepat terhadap hasil kerja siswa (Leachy, 2012). Hal ini menunjukkan bahwa penilaian dalam matematika sebaiknya dilakukan dengan memanfaatkan perkembangan teknologi.

Selain memperhatikan prinsip-prinsip tersebut, penilaian dalam matematika semestinya juga dilakukan sesuai dengan karakteristik matematika. Van de Walle (2007: 17) menyatakan bahwa hal-hal yang dinilai dalam pembelajaran matematika meliputi pemahaman konsep, keterampilan matematika, kemampuan *problem solving*, serta sikap dan keyakinan mengenai matematika. Pemahaman konsep sangat penting dalam pembelajaran matematika.

Siswa harus memiliki pemahaman mengenai konsep-konsep matematika secara mendalam termasuk keterkaitan antar konsep agar mampu menguasai matematika dengan baik. Hal ini mengingat bahwa konsep dalam matematika terbangun dengan adanya saling keterkaitan antar konsep. Sebagai contoh: Berapakah luas suatu persegi yang memiliki keliling 20 satuan? Untuk dapat menyelesaikan soal tersebut, siswa harus memiliki pemahaman konsep yang baik tentang keliling dan luas persegi serta keterkaitannya.

Keterampilan matematika juga merupakan bagian penting dalam matematika. Siswa harus mahir melakukan keterampilan matematika agar mampu menyelesaikan masalah atau soal matematika dengan benar. Keterampilan matematika merupakan prosedur yang dilakukan siswa dalam melaksanakan tugas-tugas matematika seperti membuat grafik, menghitung atau mengkalkulasi, dan mengukur objek berdasarkan alat ukur yang benar. Misalnya, ketika guru menugaskan siswa untuk menggambar diagram lingkaran dari suatu data tertentu, maka selain memerlukan pemahaman konsep yang baik, siswa juga harus memiliki keterampilan matematika yang baik untuk dapat membuat diagram lingkaran dengan benar.

Selain pemahaman konsep dan keterampilan matematika, pemecahan masalah juga merupakan esensi dari kekuatan matematika. Siswa tidak saja harus memahami konsep matematika secara mendalam, memiliki keterampilan matematika yang baik, namun yang lebih penting lagi adalah siswa dapat

memanfaatkan keduanya untuk memecahkan masalah dalam matematika. Secara garis besar ada dua kategori permasalahan dalam matematika yaitu permasalahan tertutup dan permasalahan terbuka. Permasalahan tertutup (*closed problem*) dimaksudkan sebagai permasalahan yang memiliki satu jawaban yang benar namun bisa diperoleh dengan satu atau banyak cara untuk menyelesaikannya. Contoh: (a) Tentukan hasil dari $3+4$; (b) Jika harga sebuah pensil \$2 maka berapakah harga dua buah pensil? Adapun permasalahan terbuka memiliki banyak jawaban benar dan banyak cara dalam penyelesaiannya. Contoh: (a) Buatlah persamaan yang dapat dibuat dari angka 2,3, dan 5; (b) Buatlah persamaan yang hasilnya 9 (Bahar & Maker, 2015: 1533-1534). Hal ini menunjukkan bahwa guru matematika dapat menggunakan jenis permasalahan tertutup maupun terbuka dalam kegiatan penilaian yang dilakukan.

Berdasarkan uraian di atas, penilaian aspek pengetahuan dan keterampilan dalam matematika tentunya bersifat saling terkait. Hasil belajar matematika tidak hanya dilihat ketika siswa memahami konsep dan mampu memecahkan masalah melainkan juga harus dilihat kemahirannya dalam melakukan keterampilan matematika. Proses siswa dalam memanfaatkan pemahaman konsep dan keterampilan matematika untuk memecahkan suatu masalah dalam matematika melalui penalaran yang dimilikinya harus dinilai secara menyeluruh. Oleh sebab itu, penilaian aspek pengetahuan dan keterampilan dalam pembelajaran matematika semestinya saling terkait dan tidak dipisahkan.

Adapun terkait dengan sikap, Baron & Byrne (1987: 46) secara spesifik memberikan definisi tentang sikap yaitu ketetapan diri seseorang dalam mengevaluasi baik positif atau negatif terhadap orang lain, diri sendiri, benda, atau masalah. Sikap bersifat tetap sepanjang waktu sehingga perasaan yang bersifat sesaat tidak dapat dikatakan sebagai sikap. Hal ini senada dengan Ajzen (1989: 241) yang mengartikan sikap sebagai disposisi individu untuk merespons positif atau negatif terhadap objek, orang, lembaga, atau peristiwa. Vaughan & Hogg (1995: 5) mendukung definisi tersebut dengan menyatakan bahwa sikap bersifat relatif permanen yakni bertahan dalam waktu dan situasi. Perasaan sesaat di satu tempat bukan merupakan sikap karena sikap merupakan sebuah kondisi perasaan dan keyakinan yang relatif tetap serta berupa kecenderungan perilaku yang signifikan terhadap objek atau peristiwa. Berdasarkan beberapa pendapat di atas dapat disimpulkan bahwa sikap adalah keadaan dalam diri seseorang yang menggerakkan untuk melakukan tindakan atau perilaku dengan kondisi perasaan yang tetap atau tidak sesaat baik bersifat positif atau negatif terhadap situasi atau kondisi di lingkungan sekitarnya.

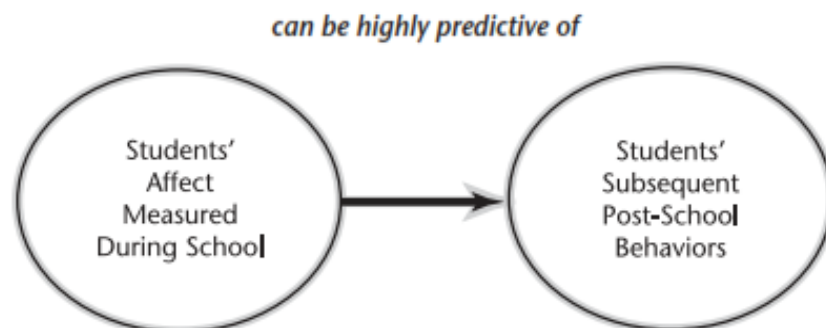
Memperhatikan tujuan pendidikan matematika terkait pembentukan sikap siswa, maka dalam penilaiannya pun semestinya juga memperhatikan sikap siswa terhadap matematika. Sikap siswa dalam menghadapi matematika dan keyakinannya mengenai matematika penting untuk diperhatikan karena sering mempengaruhi prestasi siswa dalam matematika (NCTM, 2000). Sikap siswa

terhadap matematika baik positif atau negatif mempengaruhi cara siswa melakukan sesuatu dalam matematika. Pengertian tentang sikap dalam matematika di antaranya dikemukakan Arcavi & Schoenfeld (2006: 2) yang menyatakan bahwa “*mathematical thinking related attitudes is intellectual predispositions towards doing mathematics and solving problems including perspectives on what are mathematics and mathematical activity*”. Pernyataan ini menjelaskan bahwa berfikir matematis yang dihubungkan sebagai sikap adalah kecenderungan intelektual terhadap matematika dan pemecahan masalah yang termasuk di dalamnya adalah perspektif tentang apa matematika dan aktivitas matematika. Hal ini dapat diungkapkan dengan kata lain yakni sikap yang dimaksud dapat disebut sebagai sikap matematis.

Selain sikap matematis, melalui penilaian dalam pembelajaran matematika juga semestinya berkontribusi mewujudkan tujuan pendidikan nasional bangsa Indonesia terutama terkait pengembangan potensi siswa untuk memiliki akhlak mulia. Hal ini menunjukkan bahwa melalui penilaian dalam pembelajaran matematika semestinya juga ikut berperan dalam pendidikan karakter yang harus dimiliki para siswa Indonesia seperti menghargai ajaran agama yang dianut, jujur, tanggung jawab, gotong royong, toleransi, sopan, santun, dan lain-lain.

Menurut Lickona (1992: 12), penekanan pendidikan karakter mulia (*good character*) meliputi tiga kelompok karakter yang baik yaitu *moral knowing*,

moral feeling, dan *moral action*. Artinya, karakter mulia (*good character*) memuat pengetahuan tentang kebaikan yang selanjutnya menimbulkan komitmen atau niat terhadap kebaikan dalam diri seseorang dan akhirnya benar-benar diwujudkan dalam perilaku yakni melakukan kebaikan. Terkait pembentukan karakter yang baik dalam pembelajaran matematika, Sujono (1988: 19-20) mengemukakan bahwa dengan belajar matematika dapat dikembangkan watak atau karakter seseorang. Hal ini senada dengan Soedjadi (2000: 66-67) yang menyebutkan bahwa pembelajaran matematika tidak hanya memuat nilai edukasi yang mencerdaskan siswa melainkan juga nilai edukasi yang membentuk pribadi siswa. Penilaian aspek sikap yang dilakukan selama di sekolah, menurut Popham (2003: 107) berkontribusi terhadap perilaku siswa setelah lulus. Prediksi tersebut dapat digambarkan sebagai berikut.



Gambar 1. Prediksi Hasil Penilaian Afektif di Sekolah (Popham, 2003: 107)

Gambar 1 menunjukkan bahwa afeksi siswa yang diukur selama di sekolah sangat berpengaruh terhadap perilaku siswa setelah lulus. Oleh sebab itu, pada

penilaian aspek sikap dalam pembelajaran matematika, selain menilai sikap matematis sebaiknya juga menilai karakter siswa.

Berdasarkan uraian di atas, dapat diambil kesimpulan bahwa penilaian dalam pembelajaran matematika di sekolah secara garis besar mencakup aspek sikap, pengetahuan, dan keterampilan. Namun, instrumen penilaian yang digunakan pun seharusnya disesuaikan dengan karakteristik dan kekhasan matematika. Penilaian pada aspek sikap hendaknya terkait dengan sikap matematis siswa dan pendidikan karakter. Jenis karakter yang dinilai dapat disesuaikan dengan materi yang sedang dipelajari siswa dan dilakukan baik dalam level *moral knowing*, *moral feeling*, maupun *moral action*. Adapun untuk penilaian aspek pengetahuan dan keterampilan dalam pembelajaran matematika saling terkait di antaranya karena sama-sama memperhatikan aspek pemahaman konsep siswa sehingga dalam penyusunan instrumen penilaiannya sebaiknya dirancang secara beriringan atau bersamaan.

2. Penilaian Matematika SMP Berdasarkan Kurikulum 2013 dan Implementasinya pada Materi Statistika

Penilaian pada mata pelajaran matematika SMP berdasarkan Kurikulum 2013 dilaksanakan dengan memperhatikan berbagai peraturan pemerintah yang terus mengalami perubahan atau revisi berdasarkan perkembangan yang terjadi di lapangan. Berdasarkan Permendikbud Nomor 23 Tahun 2016 tentang Penilaian Hasil Belajar oleh Pendidik dan Satuan Pendidikan pada Pendidikan Dasar dan Pendidikan Menengah disebutkan bahwa penilaian adalah proses pengumpulan

dan pengolahan informasi untuk mengukur pencapaian hasil belajar siswa. Penilaian hasil belajar siswa pada pendidikan dasar dan pendidikan menengah meliputi aspek sikap, pengetahuan, dan keterampilan. Secara umum karakteristik kompetensi aspek sikap berdasarkan Kurikulum 2013 pada awal mulai diberlakukan mengacu pada taksonomi domain afektif dari Krathwohl, karakteristik kompetensi aspek pengetahuan mengacu pada revisi taksonomi Bloom oleh Lorin W. Anderson dan David R. Krathwohl, dan karakteristik kompetensi dimensi keterampilan mengacu pada karakteristik kompetensi keterampilan dari Dyer (Kemendikbud, 2013). Berikut uraian masing-masing taksonomi tersebut.

a. Taksonomi Domain Afektif dari Krathwohl

Taksonomi domain afektif dari Krathwohl merupakan salah satu taksonomi yang mungkin paling dikenal dari berbagai taksonomi domain afektif yang berkembang. Seels & Glasgow (1990: 28) mengemukakan bahwa taksonomi tersebut sesuai dengan prinsip internalisasi yang mengacu pada proses yang mempengaruhi keadaan seseorang terhadap suatu objek dari tingkat kesadaran umum hingga diinternalisasi dan konsisten terdapat pada perilaku seseorang. Taksonomi domain afektif dari Krathwohl meliputi lima tahap yaitu *receiving*, *responding*, *valuing*, *organization*, dan

characterization by value (Krathwohl, Bloom, & Masia, 1964). Masing-masing tahap diuraikan sebagai berikut.

- 1) *Receiving* (menerima), menggambarkan tahap menyadari atau sensitif terhadap keberadaan ide-ide tertentu, materi, atau fenomena dan bersedia untuk mentolerirnya yakni kesediaan untuk menyadari adanya suatu fenomena di lingkungannya. Tahap ini merupakan pembuka indera seseorang terhadap dunia luar. Ada tiga proses untuk tahap ini yaitu: pertama, adanya kesadaran tentang apa yang sedang terjadi; kedua, kesediaan menerima apa yang terjadi sebagai stimulus; dan ketiga, kemauan untuk mengontrol atau memilih stimulus mana yang akan diperhatikan lebih lanjut.
- 2) *Responding* (menanggapi), merupakan tahap kedua taksonomi yang berupa pemberian respons atau reaksi terhadap gejala atau fenomena yang ada di lingkungannya. Pada tahap ini seseorang tidak saja memperhatikan fenomena tetapi juga bereaksi. Tingkat yang tinggi pada tahap ini adalah minat. Proses menanggapi ini lebih menekankan pada keinginan yang bersangkutan.
- 3) *Valuing* (menilai), merupakan tahap yang berkaitan dengan penilaian atau penghargaan yang diberikan pada suatu objek, fenomena, atau tingkah laku. Penilaian berdasar pada internalisasi dari serangkaian nilai tertentu yang diekspresikan ke dalam tingkah laku. *Valuing* melibatkan

penentuan nilai, keyakinan atau sikap yang menunjukkan derajat internalisasi, dan komitmen. Derajat rentangannya mulai dari menerima suatu nilai hingga pada tingkat komitmen. Kalau pada tahap menanggapi, seseorang yang melakukannya baru menunjukkan rasa senang dan gembira dapat memberikan penanggapan, dalam jenjang penghargaan ini sudah sampai pada rasa keterikatan atau memiliki terhadap suatu stimulus. Tahap penghargaan terbagi atas tiga kategori yaitu penerimaan suatu nilai, kecenderungan (*preferensi*) akan suatu nilai, dan keterikatan (*commitment*) akan suatu nilai tertentu. Tahap menilai atau menghargai tidak hanya menerima nilai tapi juga mampu menilai konsep atau fenomena baik atau buruk.

- 4) *Organization* (pengorganisasian), merupakan tahap keempat yang memadukan nilai-nilai yang berbeda, menyelesaikan konflik, dan membentuk suatu sistem nilai yang konsisten. Pada tahap *organization*, nilai satu dengan nilai lain dikaitkan, konflik antar nilai diselesaikan, dan mulai membangun sistem nilai internal yang konsisten. Pengorganisasian terjadi apabila seseorang berada dalam situasi dimana terdapat lebih dari satu nilai.
- 5) *Characterization by a Value or Value Complex* (karakterisasi nilai atau nilai yang kompleks), merupakan tahap tertinggi yang berarti melakukan tindakan secara konsisten atau tetap sesuai dengan nilai-nilai yang telah terinternalisasi dalam diri tiap individu. Tahap ini adalah tahap tertinggi

pada domain afektif karena pada tahap ini suatu nilai dan sikap bukan saja disenangi, dihargai, diterima, digunakan dalam kehidupan, dan diorganisasikan dengan nilai lainnya melainkan sudah menyatu pada dirinya.

b. Revisi Taksonomi Bloom

Taksonomi untuk aspek pengetahuan pada Kurikulum 2013 mengacu pada revisi taksonomi Bloom oleh Lorin W. Anderson dan David R. Krathwohl. Revisi taksonomi Bloom meliputi dua dimensi yaitu dimensi proses kognitif dan dimensi pengetahuan. Dimensi proses kognitif memuat enam kategori yaitu: *remember*, *understand*, *apply*, *analyze*, *evaluate*, dan *create*. Adapun dimensi pengetahuan memuat empat kategori yaitu *factual knowledge*, *conceptual knowledge*, *procedural knowledge*, dan *metacognitive knowledge*. Pengertian masing-masing kategori dalam dimensi proses kognitif dan dimensi pengetahuan pada revisi taksonomi Bloom beserta contoh-contohnya dalam pembelajaran matematika diuraikan sebagai berikut (Anderson & Krathwohl, 2001: 66-88).

1) Dimensi Proses Kognitif

a) *Remember*

Remember (mengingat), merupakan upaya memperoleh kembali pengetahuan dari ingatan atau memori yang telah lampau. Mengingat terdiri

dari *recognition* dan *recalling*. *Recognition* terkait dengan identifikasi guna menempatkan pengetahuan masa lampau terkait pengetahuan yang dipelajari. Contoh: mengenal simbol π , \subset , \notin , Σ , μ , dan seterusnya. *Recalling* merupakan proses kognitif mengambil pengetahuan masa lampau yang sesuai, contoh: mengingat bahwa sudut siku-siku besarnya 90° , modulus dari sekumpulan data adalah nilai data yang paling sering muncul atau mempunyai frekuensi tertinggi.

b) *Understand*

Understand (mengerti/memahami), artinya membangun suatu pengertian dari berbagai sumber, misalnya: bacaan, pesan, dan komunikasi. Mengerti/memahami meliputi aktivitas *classification* dan *comparing*. *Classification* berarti mengelompokkan atau mengkategorikan sesuatu dalam satu kategori yang diawali dari suatu informasi atau contoh yang spesifik. Contoh: mengelompokkan sekumpulan bilangan dalam bilangan genap dan bilangan ganjil. *Comparing* berarti mengidentifikasi perbedaan dan persamaan dari dua atau lebih situasi, kejadian, objek, ide, atau permasalahan. Membandingkan dapat dilakukan dengan cara memetakan, mengontraskan, dan mencocokkan. Contoh: membandingkan metode substitusi dan eliminasi untuk menyelesaikan persamaan kuadrat, membandingkan jenis data yang cocok disajikan dalam diagram lingkaran dan grafik garis.

c) *Apply*

Apply (menerapkan), merupakan proses kognitif menggunakan suatu cara atau prosedur untuk menyelesaikan masalah atau melakukan percobaan. Menerapkan terdiri dari *executing* dan *implementing*. *Executing* berarti proses kognitif menyelesaikan masalah atau melakukan percobaan menggunakan cara atau prosedur yang seharusnya. Contoh: menggunakan rumus yang tepat dalam menghitung rata-rata dari sekumpulan data. *Implementing* berkaitan dengan pemilihan dan penggunaan cara untuk hal-hal baru atau asing. Contoh: menentukan luas daerah tertentu dengan menggunakan integral, menggunakan grafik garis untuk menggambarkan data tentang suhu udara suatu daerah dalam waktu tertentu.

d) *Analyze*

Analyze (menganalisis) adalah proses kognitif dalam memecahkan suatu masalah dengan memisahkan tiap bagian masalah dan mencari keterkaitannya serta mencari tahu bagaimana keterkaitan itu dapat menimbulkan masalah. Menganalisis meliputi proses kognitif *attributeing* dan *organizing*. *Attributeing* muncul ketika siswa menemukan masalah dan membangun ulang unsur masalahnya. Contoh: menandai prosedur mengalikan dua bilangan dengan tiga digit. *Organizing* diawali siswa dengan mengidentifikasi unsur yang paling sesuai dan penting dari suatu masalah, dilanjutkan dengan membangun hubungan yang cocok berdasarkan informasi yang diperoleh. Contoh: menyusun keterangan tentang berbagai

istilah dalam statistika dan formulanya, menyusun keterangan tentang kondisi yang mendasari penggunaan tabel atau diagram tertentu.

e) *Evaluate*

Evaluate (mengevaluasi) adalah proses kognitif memberi penilaian berdasarkan standar dan kriteria yang ada. Mengevaluasi meliputi *checking* dan *critiquing*. *Checking* terkait dengan pengujian hal-hal yang tidak konsisten atau kegagalan dari suatu operasi dan mengarah pada penetapan sejauh mana suatu rencana telah dapat berjalan dengan baik. Contoh: memeriksa kebenaran suatu argumen yang diberikan tentang dimensi tiga. *Critiquing* berkaitan dengan berpikir kritis yakni siswa mampu melihat sisi negatif dan positif dari suatu hal dan menilainya dengan standar atau kriteria yang telah ditetapkan. Contoh: memberikan penilaian mengapa penggunaan metode tertentu lebih baik daripada metode yang lain dalam memecahkan suatu masalah matematika dalam kehidupan sehari-hari.

f) *Create*

Create (mencipta), merupakan proses kognitif meletakkan unsur-unsur secara bersama guna membentuk kesatuan yang bersifat koheren dan mengarahkan siswa menghasilkan produk baru dengan mengorganisasikan unsur-unsur menjadi bentuk yang berbeda dengan sebelumnya. Mencipta meliputi *generating* dan *producing*. *Generating* berarti merepresentasikan atau merumuskan masalah dan menemukan alternatif hipotesis yang diperlukan. Contoh: menghipotesiskan kecenderungan suatu data.

Menggeneralisasikan merupakan proses berpikir divergen yang menjadi inti dari berpikir kreatif. *Producing* terkait dengan perencanaan untuk menyelesaikan masalah sampai membuat suatu produk. Contoh: menghasilkan jaring-jaring balok yang berbeda dari jaring-jaring balok yang dicontohkan, menyajikan data melalui animasi komputer.

2) Dimensi Pengetahuan

a) *Factual Knowledge*

Factual knowledge (pengetahuan faktual) memuat elemen-elemen dasar guna mempelajari suatu disiplin ilmu atau menyelesaikan masalah dalam disiplin ilmu tersebut. Pengetahuan faktual meliputi: (1) pengetahuan tentang terminologi; dan (2) pengetahuan tentang detail-detail dan elemen-elemen spesifik. Pengetahuan tentang terminologi memuat pengetahuan tentang label, simbol verbal, dan simbol non verbal (angka, kata, gambar, tanda). Contoh: pengetahuan simbol 2017, $>$, Σ , π . Pengetahuan tentang detail-detail dan elemen-elemen spesifik memuat pengetahuan tentang peristiwa, tanggal, lokasi, sumber informasi, dan sejenisnya. Contoh: pengetahuan tentang $4+3 = 7$, data tentang jumlah penduduk suatu daerah tertentu.

b) *Conceptual Knowledge*

Conceptual knowledge (pengetahuan konseptual) merupakan pengetahuan tentang klasifikasi, kategori, dan hubungan antara dua atau

lebih kategori pengetahuan yang lebih tertata dan kompleks. Pengetahuan konseptual meliputi: (1) pengetahuan tentang kategori dan klasifikasi; (2) pengetahuan tentang prinsip dan generalisasi; dan (3) pengetahuan tentang teori, struktur, dan model. Pengetahuan tentang kategori dan klasifikasi meliputi kategori, kelas, divisi, dan susunan spesifik dalam suatu disiplin ilmu. Contoh: pengetahuan tentang pengertian bilangan asli, pengertian persegi, pengertian statistik dan statistika, pengertian parameter. Prinsip dan generalisasi memuat fakta dan peristiwa yang spesifik serta pendeskripsian proses dan interelasi di antara kategori dan detail-detail suatu disiplin ilmu. Contoh: pengetahuan tentang prinsip dari modus ponens, silogisme, atau modus tollens. Pengetahuan tentang struktur, teori, dan model meliputi pengetahuan tentang paradigma, teori, epistemologi, model dalam disiplin-disiplin ilmu guna mendeskripsikan, memahami, menjelaskan, dan memprediksi fenomena. Contoh: pengetahuan tentang teorema Pythagoras.

c) *Procedural Knowledge*

Procedural knowledge (pengetahuan prosedural) merupakan pengetahuan mengenai cara melakukan sesuatu. Pengetahuan prosedural terkait dengan pertanyaan “bagaimana”. Pengetahuan prosedural meliputi: (1) pengetahuan tentang keterampilan dalam bidang tertentu dan algoritma; (2) pengetahuan tentang metode dan teknik dalam bidang tertentu; dan (3) pengetahuan tentang kriteria guna menentukan kapan harus menggunakan suatu cara atau prosedur tertentu yang tepat. Pengetahuan tentang

keterampilan dalam bidang tertentu dan algoritma, misalnya cara menjumlahkan 4 dan 2 (algoritma) merupakan *pengetahuan prosedural* sedangkan jawabannya 6 merupakan *pengetahuan faktual*. Pengetahuan tentang metode dan teknik dalam bidang tertentu memuat tentang bagaimana proses berpikir dalam menyelesaikan masalah, bukan hasil pemikiran atau penyelesaian masalahnya. Contoh: pengetahuan tentang bagaimana cara melukis segitiga samakaki dan samasisi, bagaimana cara membuat diagram lingkaran dari suatu data. Adapun contoh pengetahuan tentang kriteria guna menentukan kapan menggunakan suatu prosedur yang tepat, misalnya: pengetahuan tentang aturan yang digunakan dalam melakukan operasi campuran dari beberapa bilangan, pengetahuan tentang pembuatan sudut dalam penyajian data menggunakan diagram lingkaran.

d) *Metacognitive Knowledge*

Metacognitive knowledge (pengetahuan metakognitif) bersifat relatif baru dan termasuk dalam dimensi pengetahuan. Pengetahuan metakognitif meliputi: (1) pengetahuan strategis; (2) pengetahuan tentang tugas-tugas kognitif yang meliputi pengetahuan kontekstual dan kondisional; dan (3) pengetahuan diri. Pengetahuan strategis merupakan pengetahuan tentang strategi dalam pemecahan masalah. Contoh: mengetahui bahwa menggunakan strategi mengerjakan soal-soal matematika berbeda dengan membuat puisi. Pengetahuan tentang tugas-tugas kognitif yang meliputi pengetahuan kontekstual dan kondisional merupakan pengetahuan tentang

tuntutan beragam tugas kognitif. Pengetahuan kontekstual terkait dengan strategi belajar dan berpikir, sedangkan pengetahuan kondisional terkait tentang situasi dan kondisi kapan dan mengapa suatu strategi tertentu digunakan. Contoh: mengetahui bahwa metode pemfaktoran bentuk kuadrat tertentu mengandung kekurangan dan kelebihan dibanding dengan metode lain. Pengetahuan diri memuat pengetahuan tentang kelemahan, kekuatan, bakat, minat, dan motivasi yang terkait dengan kognisi dan belajar. Artinya, pengetahuan diri ini memuat kesadaran tentang tingkat pengetahuan yang dimiliki oleh diri sendiri. Contoh: menyadari bahwa materi tertentu dalam statistika sudah dipahami dan ada materi lain yang belum dipahami.

Secara umum, tabel taksonomi dari revisi taksonomi Bloom disajikan pada Tabel 1 sebagai berikut (Anderson & Krathwohl, 2001: 28).

Tabel 1. Taksonomi dari Revisi Taksonomi Bloom

<i>The Knowledge Dimension</i>	<i>The Cognitive Process Dimension</i>					
	<i>1 Remember</i>	<i>2 Understand</i>	<i>3 Apply</i>	<i>4 Analyzes</i>	<i>5 Evaluate</i>	<i>6 Create</i>
<i>A. Factual Knowledge</i>						
<i>B. Conceptual Knowledge</i>						
<i>C. Procedural Knowledge</i>						
<i>D. Metacognitive Knowledge</i>						

Berdasarkan Tabel 1, tiap sel mewakili dimensi pengetahuan dan dimensi proses kognitif. Misalnya, sel *A1* berarti pengetahuan faktual dan proses kognitif mengingat. Contoh: mengingat simbol x , π , \subset , \notin . Sel *B3* berarti pengetahuan konseptual dan proses kognitif mengaplikasikan. Contoh: menghitung $25 \times 4 - 25$, menentukan nilai rata-rata, median, dan modus dari sekumpulan data tertentu. Adapun sel *C6* berarti pengetahuan prosedural dan proses kognitif menciptakan. Contoh, membuat bangun sebarang yang memiliki luas 96 cm^2 , membuat diagram lingkaran dan menyajikannya dalam tampilan semenarik mungkin menggunakan media grafis atau komputer.

c. Taksonomi Dyer untuk Aspek Keterampilan

Kompetensi keterampilan pada Kurikulum 2013 mengacu pada konsep kreativitas dari inovator yang dikemukakan Dyer, Gregersen, & Christensen (2009: 2) berdasarkan hasil riset yang dilakukan selama tujuh tahun terhadap 25 enterpreneur inovatif dan survei pada lebih dari 3000 orang. Hasil riset yang memberikan pencerahan terkait kemampuan menemukan sesuatu yang inovatif dan menjadi bagian dari perilaku para inovator tersebut ditulis dalam *The Innovator's DNA* (2009: 3). Hasil riset menunjukkan bahwa inovator memiliki empat aktivitas utama yang berupa pola tindakan hingga membuat seorang inovator menemukan hal-hal baru

yaitu *questioning*, *observing*, *experimenting*, dan *networking*. Keempat pola tindakan tersebut secara bersama membantu para inovator dalam kegiatan *associating* (mengasosiasi atau menalar) sehingga dapat menumbuhkan wawasannya. Akhirnya dapat disimpulkan bahwa terdapat lima keterampilan penggerak inovator yaitu *questioning*, *observing*, *experimenting*, *networking* dan mengintegrasikan keempatnya dengan dukungan kecakapan bernalar (*associating*) untuk membentuk wawasan terbaru. Secara singkat pengertian masing-masing keterampilan adalah sebagai berikut (Dyer, Gregersen, & Christensen, 2009: 5).

- 1) *Questioning* (menanya), dimaksudkan bahwa dengan menanya, seorang inovator dapat menggali dan mempertimbangkan kemungkinan-kemungkinan baru. Kecerdasan kreatif tidak berasal dari kemampuan menjawab dengan benar, tetapi digali dengan keterampilan menanya.
- 2) *Observing* (mengamati), merupakan kegiatan inovator yang menemukan banyak ide karena kegiatannya yang sering mengamati secara langsung dan detail mengenai apa yang terjadi dan dialami.
- 3) *Experimenting* (mencoba) merupakan kegiatan yang membuat inovator lebih cepat melihat reaksi, berinteraksi dengan objek, dan membuktikan bisa atau tidaknya sebuah ide dijalankan. Hal ini karena tanpa percobaan dan kegagalan yang berulang tidak mungkin dilahirkan sebuah inovasi

yang sukses. Seorang inovator tanpa henti mencoba dan mencoba pengalaman baru serta menjelajahi pengalaman yang berbeda dari sebelumnya.

- 4) *Networking* (membentuk jejaring) merupakan kegiatan berinteraksi dengan orang-orang yang berasal dari berbagai latar belakang sehingga didapatkan pandangan dari berbagai perspektif yang berbeda. Hal ini menunjukkan bahwa kecerdasan berinteraksi dan berkolaborasi merupakan faktor penting yang menunjang daya inovasi. Para inovator pada umumnya menghabiskan waktu lebih banyak pada kegiatan ini dibandingkan orang lain dan mampu menggali sukses dari hal tersebut.
- 5) *Associating* (menalar), merupakan kemampuan menghubungkan hal-hal yang tampak tidak berhubungan sama sekali menjadi sesuatu yang baru.

Berdasarkan berbagai uraian di atas, menunjukkan bahwa penilaian dalam pembelajaran matematika SMP berdasarkan Kurikulum 2013 meliputi aspek sikap, pengetahuan, dan keterampilan yang semestinya dilakukan dengan memperhatikan karakteristik materi yang dipelajari para siswa. Adapun terkait instrumen penilaian yang digunakan, berbagai teknik penilaian dan bentuk instrumen, baik formal maupun informal dapat digunakan. Hal ini sebagaimana telah tertuang dalam Permendikbud Nomor 23 Tahun 2016 tentang Standar Penilaian Pendidikan yang menyebutkan

bahwa penilaian hasil belajar oleh pendidik dapat dilakukan dalam bentuk ulangan, penugasan, dan/atau bentuk lain yang diperlukan dan instrumen penilaiannya dapat berupa tes, pengamatan, penugasan perseorangan atau kelompok, dan bentuk lain yang sesuai dengan karakteristik kompetensi dan tingkat perkembangan siswa.

Salah satu materi dalam mata pelajaran matematika SMP berdasarkan Kurikulum 2013 yang dipelajari dari kelas VII hingga IX dan dipilih dalam penelitian ini untuk dikembangkan instrumen penilaiannya adalah statistika. Materi ini memiliki sifat khas yakni sangat berkaitan dengan data dalam kehidupan sehari-hari. Oleh sebab itu, melalui pembelajaran materi ini dapat dikembangkan karakter mulia siswa yakni sikap jujur dan tanggung jawab selain sikap matematis kaitannya dengan materi matematika pada umumnya.

Penilaian aspek sikap pada materi statistika dapat ditekankan pada level *moral knowing* karena agak menyulitkan bagi para guru matematika kalau penilaian dilakukan dengan pengamatan pada level tindakan atau *moral action*. Penilaian pada level ini juga bersesuaian dengan taksonomi afektif dari Krathwohl mulai dari *receiving* (menerima), *responding* (menanggapi), *valuing* (menilai), dan *organization* (pengorganisasian). Penilaian sikap pada level *moral knowing* dapat dilakukan secara lebih mudah melalui penilaian diri (*self assessment*). Adapun untuk penilaian

aspek pengetahuan dan keterampilan pada materi statistika dapat dilakukan melalui tes uraian dan tugas proyek individu. Melalui tes uraian, penguasaan materi siswa dapat dilihat dan diukur secara lebih mendalam dan melalui pengerjaan tugas proyek individu dapat diketahui dan diungkap aspek keterampilan siswa dengan lebih detail secara perseorangan.

3. Karakter Jujur

Jujur merupakan akhlak mulia atau karakter baik yang harus dimiliki para siswa Indonesia. Bisri (2009: 8) mengemukakan bahwa jujur berarti mengatakan atau melakukan tindakan sesuai dengan yang sebenarnya. Orang yang jujur akan berkata atau berbuat sesuai dengan kenyataan apa adanya tanpa dikurangi atau ditambahi. Hal ini berarti bahwa orang yang jujur berperilaku apa adanya sesuai dengan yang sesungguhnya, tidak dibuat-buat, tidak berbohong, tidak dikurangi-kurangi, tidak ditambah-tambahi, dan tidak menyembunyikan informasi.

Menurut Asmani (2011: 22), kejujuran merupakan perilaku yang menjadikan diri sebagai orang yang selalu dapat dipercaya. Hal ini diwujudkan pada perkataan dan tindakan, baik terhadap diri sendiri maupun pada orang lain. Senada dengan pengertian tersebut, Zuhdi (2015: 26) mengemukakan bahwa kejujuran adalah perilaku seseorang yang didasarkan pada upaya menjadikan dirinya selalu dapat dipercaya dalam perkataan dan perbuatannya. Kemendikbud (2013) menyebutkan bahwa jujur adalah perilaku yang didasarkan pada upaya

menjadikan dirinya sebagai orang yang selalu dapat dipercaya dalam perkataan, tindakan, dan pekerjaan. Hal ini berarti bahwa kejujuran seseorang tercermin dari tindakannya untuk tidak bohong, lurus hati, dapat dipercaya kata-katanya, dan tidak berkhianat. Apabila seseorang berkata tidak sesuai dengan kebenaran dan kenyataan sebenarnya atau tidak mengakui suatu hal sesuai dengan apa adanya, maka orang tersebut dapat dikatakan tidak jujur.

Karakteristik perilaku jujur menurut Emosda (2011: 152) adalah perilakunya diikuti dengan hati yang lurus atau ikhlas, berkata sesuai kenyataan, dan bertindak sesuai bukti dan kebenaran yang ada. Rusyan (2006) mengemukakan beberapa hal yang mendorong terbentuknya sifat jujur, yaitu: (a) membiasakan berkata sesuai dengan yang dilakukan, (b) mengakui kebenaran yang dimiliki orang lain dan kesalahan yang ada pada diri sendiri jika memang memiliki kesalahan, (c) selalu ingat bahwa semua perbuatan dilihat Allah SWT, (d) meyakini bahwa sifat jujur mengantarkan manusia pada derajat yang terhormat, (e) bersikap bijaksana sesuai aturan hukum, (f) meyakini bahwa dengan jujur, berarti menjaga diri di akhirat kelak. Sa'aduddin (2006: 31) juga mengemukakan beberapa indikator dalam kejujuran yaitu:

- a) Jujur pada diri sendiri. Kejujuran yang paling berharga pada diri manusia adalah ketika berani tampil apa adanya atau jujur pada diri sendiri, apa adanya tidak mengharapkan orang lain melihat dirinya lebih dari kenyataan.
- b) Jujur dalam perkataan. Artinya, ketika berkomunikasi hendaknya dilakukan dengan berkata yang benar, jujur, lurus dan tidak berbelit-belit dalam

menyampaikan sebuah informasi. Hal ini disebabkan apabila memberikan informasi dilakukan dengan berbelit-belit maka akan memungkinkan terjadinya kesalahan informasi yang diterima oleh penerima informasi dari pengirim informasi.

- c) Jujur dalam usaha. Kejujuran yang berupa kemauan yakni aktifitas psikis yang mengandung usaha aktif dan berhubungan dengan pelaksanaan suatu tujuan yang merupakan titik akhir dari gerakan yang menuju suatu arah. Kemauan sebagai pendorong timbulnya perbuatan berdasarkan berbagai pertimbangan akal/pikir yang menentukan benar salahnya perbuatan dan pertimbangan perasaan yang menentukan baik buruknya atau halus tidaknya perbuatan.
- d) Benar dalam berjanji. Artinya bahwa dalam menjalani kehidupan, seseorang pada umumnya pernah mengucapkan janji maka yang harus dilakukan adalah menepati janjinya dan tidak boleh mengingkarinya.

Adapun indikator jujur dalam pendidikan menurut Kemendikbud (2015) meliputi: (a) tidak menyontek dalam mengerjakan ulangan/ujian, (b) tidak melakukan plagiarisme (menyalin atau mengambil karya orang lain tanpa menyebutkan sumber), (c) mengungkapkan perasaan apa adanya, (d) menyerahkan barang temuan kepada yang berhak atau berwenang, (e) membuat laporan berdasarkan informasi atau data apa adanya, (f) mengakui kekurangan atau kesalahan yang dimiliki, dan (g) menyampaikan informasi sesuai fakta yang ada. Koellhoffer (2009: 4) juga mengemukakan bahwa kejujuran dalam

akademik berarti jujur dalam peraturan pendidikan. Seseorang secara akademik dikatakan jujur apabila tidak melakukan tindakan plagiarisme. Artinya, tidak menggunakan pekerjaan orang lain atau tidak menyalin pekerjaan orang lain tanpa izinnnya.

Menurut Koss (2011: 39), ketidakjujuran akademik merupakan suatu perilaku atau tindakan seseorang yang meminjam dan menyalin tugas dari orang lain, termasuk menyalin jawaban pada saat ujian. Perilaku ketidakjujuran akademik lainnya adalah menulis jawaban ujian pada kaki atau tangan, pakaian, dinding, kertas, meja atau kursi, serta saling bertukar jawaban dengan teman menggunakan kode-kode tertentu. Adapun menurut Kibler yang dikutip Ercegovac & Richardson (2004: 301), ketidakjujuran akademik adalah suatu bentuk plagiarisme dan kecurangan yang dilakukan siswa dalam bentuk menerima atau memberikan bantuan secara tidak sah dalam ujian serta menerima penghargaan untuk pekerjaan yang tidak dilakukannya.

Perilaku tidak jujur lain dalam kegiatan akademik adalah siswa memanfaatkan kesempatan yang ada untuk saling bertukar jawaban dengan temannya atau menyontek saat pengawas ujian keluar ruangan. Selain itu, karena melihat teman-teman yang lain sedang menyontek, siswa pun mengikuti dan ikut memanfaatkan situasi untuk saling menyontek juga termasuk ketidakjujuran akademik. Hal ini sebagaimana dikemukakan Schmelkin (2008: 4) yang berpendapat bahwa bentuk ketidakjujuran akademik yaitu plagiarisme, saling meniru dengan teman lain saat ujian, mencuri tes atau soal, dan membeli kertas

jawaban. Ketidakjujuran akademik jelas merupakan perbuatan yang tidak dibenarkan dalam sistem pendidikan di Indonesia. Terkait dengan kejujuran dalam akademik ini, pengakuan terhadap suatu ketidakjujuran merupakan perbuatan positif dan hal ini merupakan karakter baik yang harus diapresiasi dan dihargai namun tentunya tidak menjadikan penilaian yang maksimal pada kinerja seorang siswa. Hal ini perlu diperhatikan mengingat hasil penilaian terhadap kinerja siswa yang sesungguhnya semestinya benar-benar terbebas dari ketidakjujuran akademik.

Berdasarkan pendapat beberapa ahli di atas, dapat diambil kesimpulan tentang pengertian karakter jujur secara konseptual. Karakter jujur yang dimaksud dalam penelitian ini secara konseptual adalah perilaku seseorang yang menunjukkan adanya kesesuaian antara ucapan dan perbuatan dengan kenyataan yang sebenarnya. Karakter jujur dimaksudkan sebagai karakter mulia yang hendaknya dimiliki setiap siswa. Adapun secara operasional jujur diartikan dengan berperilaku apa adanya sesuai kenyataan, mengakui setiap perbuatan baik positif maupun negatif, mengerjakan tugas atau ujian sesuai kemampuan sendiri, dan dapat dipercaya atas kepemilikan orang lain.

4. Karakter Tanggung Jawab

Sebagaimana karakter jujur, karakter baik yang harus dimiliki para siswa Indonesia adalah dimilikinya karakter tanggung jawab pada diri tiap siswa.

Pengertian tentang tanggung jawab di antaranya dikemukakan Lickona (1992) yang mengartikan tanggung jawab sebagai melaksanakan kewajiban atau pekerjaan baik di dalam keluarga, sekolah, maupun tempat bekerja dengan memberikan yang terbaik dan dilakukan sepenuh hati. Hal ini senada dengan pengertian tanggung jawab menurut Miller (2005: 43) yang menyatakan bahwa tanggung jawab adalah siap menanggung atas perilaku atau tindakan yang dilakukan dan melakukan hal benar pada waktu yang tepat sehingga orang lain mempercayai. Hal ini menunjukkan bahwa seseorang yang bertanggungjawab akan teguh dalam mengambil keputusan dan siap menanggung konsekuensi atau risiko yang ada. Seseorang yang bertanggung jawab akan selalu siap dan berani menanggung segala sesuatunya atau dengan kata lain berani memberikan jawab dan menanggung akibatnya.

Tanggung jawab menurut Zuchdi (2011) dijabarkan sebagai berikut: (a) memenuhi kewajiban diri, (b) dapat dipercaya, (c) dapat mengontrol diri sendiri, (d) gigih, (e) mempersiapkan diri menjadi yang terbaik, (f) tepat waktu, (g) disiplin diri, dan (h) dapat bekerja sama dengan teman dalam satu tim. Adapun menurut Kemendikbud (2015), tanggung jawab diartikan sebagai perilaku seseorang untuk melaksanakan tugas dan kewajiban yang seharusnya dia lakukan terhadap diri sendiri, lingkungan (alam, sosial dan budaya), masyarakat, negara dan Tuhan Yang Maha Esa. Perilaku yang mencerminkan adanya tanggung jawab dapat berupa: (a) melaksanakan tugas individu dengan baik, (b) menerima risiko atas segala tindakan yang dilakukan, (c) tidak menuduh/menyalahkan

orang lain tanpa bukti yang akurat, (d) mengembalikan barang yang dipinjam, (e) mengakui dan meminta maaf atas kesalahan yang dilakukan, (f) menepati janji, (g) tidak menyalahkan orang lain untuk kesalahan dirinya sendiri, dan (h) melaksanakan apa yang pernah dikatakan tanpa diminta/disuruh.

Berdasarkan berbagai pendapat di atas, dapat disimpulkan bahwa secara konseptual, tanggung jawab adalah sikap seseorang yang teguh dalam mengambil keputusan serta siap dan berani menanggung risiko atau konsekuensi atas setiap keputusannya. Berdasarkan definisi konseptual ini, didefinisikan secara operasional bahwa tanggung jawab adalah perilaku untuk melaksanakan tugas dan kewajiban sebagaimana yang seharusnya sesuai kemampuannya dengan baik dan sepuh hati (komitmen pada tugas), siap menanggung risiko atau konsekuensi yang ada (sportif), dan mengakui jika melakukan kesalahan serta mampu menyelesaikan tugas hingga terselesaikan sampai tuntas (disiplin).

5. Sikap Matematis

Sikap matematis yang baik sangat diperlukan guna menunjang keberhasilan proses pembelajaran matematika. Definisi tentang sikap matematis di antaranya dikemukakan oleh Katagiri (2006: 12) yang menyatakan bahwa:

Mathematical thinking is like an attitude, as in it can be expressed as a state of "attempting to do" or "working to do" something. It is not limited to results represented by actions, as in "the ability to do," or "could do" or "couldn't do" something.

Katagiri menegaskan bahwa *mathematical thinking* seperti sebuah sikap yang di dalamnya dapat dinyatakan sebagai keadaan "mencoba untuk melakukan" atau

"bekerja untuk melakukan" sesuatu. Hal ini tidak terbatas pada hasil yang diwakili oleh tindakan, seperti "kemampuan untuk melakukan" atau "bisa melakukan" atau "tidak bisa melakukan" sesuatu. Berdasarkan definisi tersebut, sikap matematis merujuk pada berpikir matematis yang berkaitan dengan sikap dan tindakan yang dilakukan siswa ketika berhadapan dengan masalah atau soal matematika. Sikap matematis dalam diri siswa tidak cukup hanya dilihat ketika siswa mengerjakan. Sikap matematis siswa juga dapat dilihat ketika siswa menunjukkan apa yang ditanya, informasi apa saja yang digunakan dalam perhitungan, memutuskan rumus mana yang akan digunakan, dan ketika siswa menghasilkan suatu penyelesaian dari masalah atau soal matematika yang dihadapi.

Mengacu pada pendapat Katagiri di atas, sikap matematis secara konseptual dapat diartikan sebagai suatu kecenderungan bertindak secara suka atau tidak suka terhadap suatu aktifitas pemecahan masalah matematika serta sikap seorang siswa dalam proses pembelajaran dalam konteks pemecahan masalah matematika dan aktivitas matematika. Lebih lanjut, secara lebih operasional Katagiri (2006: 13) mengemukakan bahwa sikap matematis yang dimaksud sebagai *mathematics attitude* meliputi empat faktor yaitu: (1) mencoba memahami masalah dan tujuan serta substansi masalah dengan jelas secara mandiri, (2) mencoba mengambil tindakan logis, (3) mencoba untuk mengekspresikan hal-hal dengan jelas dan ringkas, dan (4) mencoba mencari penyelesaian yang lebih baik.

6. Teknik Penilaian Diri, Tes Uraian, dan Tugas Proyek

Penilaian pencapaian kompetensi matematika semestinya dilakukan secara menyeluruh pada aspek sikap, pengetahuan, dan keterampilan. Terkait dengan teknik penilaian yang dipilih dalam melakukan penilaian tentunya sangat tergantung dengan kondisi siswa dan muatan materi yang diujikan. Secara umum untuk mengukur dan mengungkap hasil belajar siswa pada aspek pengetahuan pada pembelajaran matematika dapat digunakan teknik tes bentuk uraian karena melalui tes uraian dapat lebih diketahui pemahaman konsep dan kemampuan *problem solving* siswa. Namun mengingat kemampuan *problem solving* juga memerlukan keterampilan matematika, maka penilaian aspek pengetahuan dan keterampilan dapat dirancang bersamaan. Penilaian sikap sifatnya untuk melengkapi sehingga akan lebih fleksibel dilaksanakan guru melalui teknik penilaian diri. Adapun sikap yang dinilai bisa disesuaikan dengan karakteristik materi matematika yang dipelajari.

a. Penilaian Diri (*Self Assessment*)

Penilaian diri merupakan teknik penilaian terhadap diri sendiri dengan mengidentifikasi kelebihan dan kekurangan sikapnya dalam berperilaku. Hasil penilaian diri siswa dapat digunakan sebagai data konfirmasi tentang perkembangan sikap siswa. Instrumen penilaian diri pada umumnya berupa lembar penilaian diri yang berisi butir-butir pernyataan

sikap positif yang diharapkan dengan format kolom Ya dan Tidak, *Likert scale*, atau skala lainnya. Satu lembar penilaian diri dapat digunakan untuk melakukan penilaian sikap spiritual dan sikap sosial sekaligus (Kemendikbud, 2015). Adapun Weeden, Winter, & Broadfoot (2002: 73) mendefinisikan penilaian diri sebagai suatu proses review yang melibatkan diri sendiri dalam: (a) merefleksikan pengalaman masa lalu, (b) mengingat apa yang terkait dengan pengalaman yang dipelajari, dan (c) mencoba menambah ide yang lebih jelas tentang apa yang dipelajari. Hal ini senada dengan Noonan & Duncan (2005: 2) yang mengemukakan bahwa penilaian diri merupakan penilaian yang dilakukan oleh diri sendiri terhadap pekerjaan atau proses belajar diri sendiri.

Ross (2006: 5) menegaskan bahwa penilaian diri berkontribusi pada peningkatan prestasi dan perilaku siswa. Melalui penilaian diri dapat diketahui kecenderungan siswa ketika bereaksi terhadap situasi-situasi tertentu. Tujuan penilaian diri menurut Shepard (Brennan, 2006: 632) adalah bukan untuk memberi nilai angka, namun untuk memperoleh pengetahuan yang bisa digunakan untuk pembelajaran selanjutnya. Inti penilaian diri adalah untuk memperbaiki kualitas belajar individu secara berkelanjutan. Penilaian diri membantu proses regulasi diri siswa untuk mengamati dan menginterpretasikan kualitas diri sendiri. Proses regulasi menurut Ross (2006: 5) mencakup pengamatan diri, justifikasi diri, dan reaksi diri.

Berdasarkan beberapa pendapat di atas, dapat disimpulkan bahwa penilaian diri (*self assessment*) merupakan bentuk penilaian terhadap diri sendiri guna mengetahui kelebihan dan kekurangan sikapnya dalam berperilaku dan digunakan dalam rangka meningkatkan kualitas pembelajaran yang dilakukan. Penilaian diri dapat dilakukan dalam bentuk pengamatan diri atau pemberian respons terhadap butir-butir instrumen penilaian yang dirancang baik dalam skala *Likert* atau skala lainnya. Penggunaan teknik penilaian diri dalam penilaian aspek sikap pada mata pelajaran matematika sangat relevan dan mudah untuk dilakukan guru.

b. Tes Uraian dan Tugas Proyek

Penggunaan instrumen dalam penilaian berupa tes di antaranya dikemukakan Gronlund & Linn (1990: 8) dan Crocker & Algina (1986: 11) yang menyatakan bahwa tes merupakan salah satu alat penilaian. Definisi tentang tes di antaranya dikemukakan Thissen & Wainer (2001: 5) yang mendefinisikan tes sebagai sampel kecil dari performan individu, yang terukur dengan baik dalam kondisi terkontrol dan memberikan gambaran secara akurat mengenai kemampuan individu untuk waktu yang cukup lama. Berdasarkan pendapat ini menunjukkan bahwa tes merupakan salah satu alat penilaian yang sistematis dan memuat sampel perilaku serta dimaksudkan untuk memperoleh gambaran mengenai kemampuan peserta tes.

Terkait dengan tes uraian, Gronlund & Linn (1990: 212-213) dan Mehrens & Lehmann (1973: 207) menyatakan bahwa tes uraian dapat dikelompokkan menjadi dua yaitu tes uraian dengan respons terbatas dan tes uraian dengan respons diperluas. Tes uraian dengan respons terbatas dimaksudkan bahwa jawaban siswa dibatasi dalam format, isi, dan respons. Adapun tes uraian dengan respons diperluas, memberikan siswa kebebasan untuk mendiskusikan, memberikan informasi berdasar fakta, mengorganisasi jawaban, mengintegrasikan gagasan, dan mengevaluasi gagasan yang sesuai dengan anggapan siswa.

Mardapi (2016: 95) juga menyatakan bahwa tes uraian ada dua bentuk yaitu tes uraian objektif dan tes uraian non objektif. Bentuk tes uraian objektif memiliki jawaban yang sudah pasti dan hanya satu jawaban yang benar. Adapun bentuk tes uraian non objektif, jawabannya bersifat luas dan tidak hanya satu jawaban yang benar, tergantung argumentasi siswa. Perkembangan dalam pembelajaran matematika menunjukkan bahwa kedua jenis tes uraian ini dapat digunakan dalam pembelajaran matematika. Hal ini mengingat bahwa jenis permasalahan dalam matematika sangat menekankan pada pemecahan masalah matematika dan untuk menilai kemampuan matematika tingkat tinggi pada umumnya digunakan tes uraian non objektif yang berisi *open ended problem* (permasalahan terbuka).

Kebaikan tes uraian menurut Ornstein (1992: 1) adalah: (a) efektif digunakan untuk menentukan seberapa baik para siswa dapat berpikir secara

logis, meneliti, mengevaluasi, menyatukan, memecahkan masalah, dan membuat hipotesa, (b) dapat menunjukkan kebaikan siswa dalam mendukung pandangan, mengorganisir pemikiran, dan menciptakan gagasan, metode, dan solusi, (c) kompleksitas pertanyaan tes uraian dan kompleksitas berpikir yang diharapkan dari siswa dapat disesuaikan dengan kemampuan, umur, dan pengalaman siswa, dan (d) dapat disusun dalam waktu yang relatif singkat. Adapun kelebihan tes uraian menurut Walstad (2006: 6) adalah: (1) berpotensi besar menilai tingkat pemahaman siswa pada level yang lebih tinggi, (2) siswa memiliki kebebasan untuk menyiapkan, memilih, dan menyajikan gagasan dalam kata-kata mereka sendiri sebagai jawaban, (3) guru memiliki kesempatan melihat siswanya membuat jawaban, (4) cocok untuk tes prestasi yang berhubungan dengan analisa permasalahan, aplikasi konsep, atau evaluasi keputusan.

Tes uraian juga memiliki beberapa kelemahan. Ornstein (1992: 1) mengungkapkan kelemahan tes uraian yaitu: (1) perlu mempertimbangkan waktu untuk membaca dan mengevaluasi jawaban serta adanya subjektifitas dalam penskorannya, dan (2) panjang jawaban dan kompleksitas jawaban dapat mendorong permasalahan reliabilitas dalam penskoran. Adapun kelemahan tes uraian menurut Walstad (2006: 4) adalah: (a) skor dapat berbeda ketika dinilai guru yang sama pada waktu yang berbeda atau guru yang berbeda pada waktu yang sama, (b) memerlukan waktu yang lama untuk mengoreksi, (c) memperhatikan jumlah siswa yang besar pada setiap

kelas, kecil kemungkinan guru mempunyai banyak waktu untuk menggunakan tes uraian, dan (d) memerlukan biaya yang relatif besar.

Penilaian dalam pembelajaran matematika untuk aspek pengetahuan dan keterampilan saling terkait dan melengkapi. Penilaian keterampilan juga melibatkan pengetahuan siswa dalam menyelesaikannya. Oleh sebab itu, salah satu teknik penilaian yang dapat diterapkan dalam pembelajaran matematika untuk mengukur aspek keterampilan yang melibatkan pengetahuan adalah penilaian proyek. Hal ini sebagaimana dikemukakan Fan (2004: 11) bahwa penilaian proyek dapat menilai domain kognitif, kemampuan pemecahan masalah, dan keterampilan berpikir kreatif. Lebih lanjut dikemukakan bahwa teknik penilaian proyek cocok digunakan sebagai salah satu alternatif penilaian dalam pembelajaran matematika.

Penggunaan penilaian proyek pada penerapan Kurikulum 2013 dimaksudkan sebagai tugas yang berupa rangkaian kegiatan yang dimulai dari perencanaan, pengumpulan data, pengorganisasian data, pengolahan dan penyajian data, hingga pelaporan. Penilaian proyek dapat dilakukan untuk mengukur satu atau beberapa Kompetensi Dasar (KD). Menurut Kemendikbud (2015), hal-hal yang perlu dipertimbangkan dalam penilaian tugas proyek adalah: (1) pengelolaan, yakni kemampuan siswa memilih topik, mencari informasi, dan mengelola waktu pengumpulan data, serta penulisan laporan; (2) relevansi, artinya topik, data, dan produk sesuai dengan Kompetensi Dasar (KD); (3) keaslian, artinya produk yang

dihasilkan siswa merupakan hasil karyanya, dengan mempertimbangkan kontribusi guru berupa petunjuk dan dukungan terhadap proyek siswa; dan (4) inovasi dan kreativitas, yakni hasil proyek siswa terdapat unsur-unsur kebaruan dan menemukan sesuatu yang berbeda dari biasanya.

7. Penyusunan Instrumen Penilaian

Instrumen penilaian merupakan kunci utama dari keberhasilan pelaksanaan kegiatan penilaian pendidikan. Kualitas hasil penilaian sangat tergantung dari kualitas instrumen yang digunakan. Oleh sebab itu, instrumen penilaian hendaknya disusun dan dikembangkan dengan baik melalui langkah-langkah yang tepat dan dapat dipertanggungjawabkan sehingga hasil penilaian yang diperoleh pun juga akurat dan dapat dipertanggungjawabkan. Hal ini sebagaimana dikemukakan Anderson & Morgan (2008: 4) yang menyatakan bahwa kualitas penilaian pendidikan tergantung pada kualitas instrumen yang digunakan. Jika instrumen penilaian dirancang dengan cara yang buruk, maka penilaian yang dilakukan hanya akan membuang-buang waktu dan biaya. Sebaliknya, jika instrumen penilaian dirancang dengan baik maka hasilnya dapat mendukung kualitas informasi yang diperoleh dari hasil penilaian bahkan menjadi salah satu faktor yang dapat meningkatkan hasil belajar siswa.

Brookhart & Nitko (2011: 5) mengemukakan lima prinsip dasar dalam melakukan penilaian yaitu: (1) menentukan dengan jelas apa yang hendak dinilai, (2) memastikan bahwa teknik penilaian yang dipilih cocok untuk menilai apa yang hendak dinilai, (3) memastikan bahwa teknik penilaian yang dipilih sesuai

dengan kebutuhan siswa, (4) jika memungkinkan, memastikan penggunaan berbagai indikator hasil belajar untuk tiap target penilaian, dan (5) memastikan bahwa ketika menginterpretasikan hasil penilaian, telah diperoleh informasi yang cukup banyak tentang siswa. Adapun langkah-langkah pengembangan instrumen menurut Wilson (2015: 18-19) meliputi: (1) peta konstruk, (2) butir-butir, (3) skor butir, dan (4) pengukuran. Berdasarkan *Queensland Studies Authority* (2010: 4) prosedur pengembangan instrumen penilaian dapat dilakukan guru melalui langkah-langkah sebagai berikut: (1) mulailah dengan tujuan umum dalam silabus yakni menentukan tujuan umum yang akan dinilai, (2) mengacu pada matriks standar untuk deskripsi yang relevan dengan tujuan umum yang dipilih, (3) mengembangkan instrumen yang memungkinkan siswa menunjukkan atribut yang dinilai, dan (4) mengembangkan kriteria pada lembar instrumen berdasarkan atribut yang dipilih dari matriks standar.

Kegiatan penyusunan instrumen penilaian khususnya yang berbentuk tes sebaiknya memperhatikan tentang kesinambungan pembelajaran (*learning continuum*) terkait keluasan dan kedalaman materi sehingga penilaian menjadi lebih efektif. Hal sebagaimana dikemukakan Subali (2016: 12) bahwa kisi-kisi tes untuk pengukuran hasil belajar harus mengacu pada *learning continuum* yang berupa *abstract continuum*. Hal ini terlebih mengingat kualifikasi penguasaan materi yang terjadi di Indonesia saat ini cenderung tidak terstruktur akibat penerapan kurikulum terus mengalami perubahan sehingga belum

memperhatikan esensi pembelajaran berkesinambungan (*learning continuum*). Padahal, kesesuaian antara proses pengukuran dengan penyajian materi dalam proses pembelajaran menjadi faktor yang perlu diperhatikan mengingat kondisi ini berpengaruh terhadap prestasi yang diperoleh oleh siswa. Oleh sebab itu, instrumen penilaian berupa tes dalam pembelajaran matematika sebaiknya juga dibuat berurutan dengan menggunakan *learning continuum* yang disajikan dalam *abstract continuum* agar proses penilaian lebih terarah.

Rumusan *learning continuum* sebagai *abstract continuum* dalam penyusunan instrumen penilaian matematika dapat dilakukan dengan mengkaji indikator-indikator Kompetensi Dasar (KD) yang telah ditetapkan. Hal ini sebagaimana dilakukan Kusaeri (2012: 21) yang melakukan perumusan *learning continuum* pada penyusunan tes diagnostik aljabar. Apabila instrumen penilaian disusun dengan memperhatikan aspek pembelajaran yang berkesinambungan (*learning continuum*) maka penilaian yang dilakukan benar-benar akan dapat mengukur pencapaian kompetensi siswa secara lebih jelas dan terarah. *Learning continuum* yang disusun sebagai *abstract continuum* membantu guru dalam meningkatkan kualitas penilaian karena guru dapat dengan urut dan lengkap mengukur dan menilai pencapaian kompetensi siswa pada aspek pengetahuan dan keterampilan.

Secara garis besar, bentuk instrumen penilaian yang digunakan dalam kegiatan penilaian pendidikan dapat dibedakan dalam dua bentuk yakni tes dan non tes. Penyusunan instrumen penilaian baik berbentuk tes maupun non tes

harus dilakukan melalui langkah-langkah yang tepat dan dapat dipertanggungjawabkan agar benar-benar diperoleh instrumen penilaian yang berkualitas baik. Berbagai referensi terkait langkah-langkah dalam menyusun instrumen tes dan non tes dapat diuraikan sebagai berikut.

a) Instrumen Tes

Secara umum, Oriondo & Antonio (1998: 34) mengemukakan langkah-langkah penyusunan instrumen tes dalam lima langkah yaitu: (1) perencanaan tes, (2) ujicoba tes, (3) menetapkan validitas tes, (4) menetapkan reliabilitas, dan (5) menafsirkan tes. Adapun Popham (2009: 62) mengemukakan tiga langkah membangun suatu instrumen tes di kelas yang digambarkan sebagai berikut.



Gambar 2. Tiga Langkah Membangun Suatu Tes di Kelas (Popham, 2009: 62)

Berdasarkan Gambar 2, langkah pertama adalah menentukan tujuan pembelajaran yang akan dibuat penilaiannya, langkah kedua adalah mengidentifikasi tes berdasarkan kesimpulan yang mendukung keputusan, dan langkah ketiga adalah membangun tes yang menghasilkan tes yang valid berdasarkan kesimpulan yang diperoleh.

Teknik perencanaan tes juga dikemukakan Oriondo & Antonio (1989: 34) yakni: (a) menentukan tujuan, (b) menyiapkan tabel spesifikasi, (c) memilih format butir yang sesuai, (d) menulis butir tes, dan (e) mengedit butir-butir tes. Adapun secara lebih spesifik, langkah-langkah menyusun tes, menurut Mardapi (2016: 95) adalah: (1) menyusun spesifikasi tes, (2) menulis tes, (3) menelaah tes, (4) melakukan uji coba tes, (5) menganalisis butir tes, (6) memperbaiki tes, (7) merakit tes, (8) melaksanakan tes, dan (9) menafsirkan hasil tes.

Menurut Gronlund & Linn (1990: 217-221), untuk menyusun tes berbentuk uraian, ada beberapa hal yang perlu diperhatikan yaitu: pertanyaan sebaiknya menunjukkan kenyataan yang ada, menampilkan estimasi batas waktu rata-rata untuk setiap pertanyaan, dan menghindari penggunaan pertanyaan opsional. Selain itu, Mardapi (2016: 105-106) juga memberikan gambaran hal-hal yang harus diperhatikan dalam membuat tes uraian yang bersifat non objektif sebagai berikut: (a) gunakan kata-kata seperti: mengapa, jelaskan, uraikan, hitunglah, bandingkan, tafsirkan, dan buktikan, (b) hindari menggunakan pertanyaan: apa, siapa, dan bila, (c) gunakan bahasa Indonesia yang baku, (d) hindari kata-kata yang dapat ditafsirkan ganda, (e) buat petunjuk mengerjakan soal, (e) buat kunci jawaban, dan (f) buat pedoman penskoran. Adapun untuk membuat tes uraian objektif, hal-hal yang harus diperhatikan meliputi: (a) menulis soal berdasarkan kisi-kisi pada indikator dan (b) mengedit pertanyaan dalam hal: apakah pertanyaan mudah dimengerti?, apakah data yang digunakan benar?, apakah tata letak keseluruhan baik?, apakah pemberian bobot skor sudah

tepat?, apakah kunci jawaban sudah benar?, dan apakah waktu mengerjakan tes cukup?.

Terkait penyusunan instrumen tes berdasarkan Kurikulum 2013, sebaiknya butir-butir tes disusun dalam enam tingkatan menurut revisi taksonomi Bloom, yaitu: pengetahuan (C1), pemahaman (C2), penerapan (C3), analisis (C4), evaluasi (C5), dan mencipta (C6). Proporsi di antara tingkatan pengetahuan tersebut pada tingkat pendidikan SMP adalah C1–C2 berkisar 20%, C3-C4 berkisar 55%, C5 berkisar 15%, dan C6 berkisar 10% (Kemendikbud, 2013). Instrumen bentuk tes merupakan salah satu instrumen yang relevan untuk menilai kompetensi aspek pengetahuan. Oleh sebab itu, mengingat bahwa penilaian pengetahuan dan keterampilan dalam pembelajaran matematika semestinya dilakukan secara beriringan maka secara umum dapat dirumuskan langkah-langkah penyusunan instrumen penilaian pengetahuan dan keterampilan dalam pembelajaran matematika yaitu: (1) menentukan tujuan penilaian dan teknik penilaian yang digunakan, (2) menentukan indikator dan merumuskan *abstract continuum*, (3) menyusun kisi-kisi, dan (4) menulis butir-butir instrumen dan pedoman penskorannya.

b) Instrumen Non Tes

Instrumen non tes pada umumnya digunakan untuk menilai pencapaian kompetensi siswa pada aspek sikap. Bentuk instrumen non tes dapat berbentuk lembar observasi, angket (kuisisioner), skala psikologi (skala sikap), maupun

inventori (daftar cek). Secara umum, langkah-langkah penyusunan instrumen penilaian sikap dapat dilakukan dengan mengacu pada langkah-langkah pengembangan instrumen penilaian afektif menurut Mardapi (2016: 132) sebagai berikut: (1) menentukan spesifikasi instrumen, (2) menulis instrumen, (3) menentukan skala instrumen, (4) menentukan sistem penskoran, (5) menelaah instrumen, (6) melakukan uji coba, (7) menganalisis instrumen, (8) merakit instrumen, (9) melaksanakan pengukuran, dan (10) menafsirkan hasil pengukuran. Adapun terkait telaah instrumen penilaian sikap dapat dilakukan dengan menelaah tentang: (a) apakah butir pertanyaan atau pernyataan sesuai dengan indikator, (b) bahasa yang digunakan apa sudah komunikatif dan menggunakan tata bahasa yang benar, (c) apakah butir pertanyaan atau pernyataan tidak bias, (d) apakah format instrumen menarik untuk dibaca, dan (e) apakah jumlah butir sudah tepat, sehingga tidak menjemukan menjawabnya (Mardapi, 2016: 140).

Apabila instrumen non tes yang digunakan berupa skala psikologi, maka menurut Azwar (2005: 11) penyusunannya dapat dilakukan dengan menggunakan langkah-langkah sebagai berikut: (a) menetapkan konstruk psikologis yang akan diukur, (b) menetapkan indikator perilaku, (c) menetapkan penskalaan dan pemilihan format, (d) menulis butir dan melakukan review butir, (e) uji coba, (f) analisis butir, (g) seleksi butir, (h) reliabilitas dan validitas, dan (i) instrumen final. Lebih lanjut, menurut Azwar (2005: 41-51), pemberian skor

dalam penyusunan skala psikologi berkaitan dengan penskalaan yakni proses penentuan letak stimulus atau respons tertentu pada suatu kontinum psikologis.

Azwar (2005: 29-32) mengemukakan tiga pendekatan utama terkait metode penskalaan yaitu metode yang berorientasi pada stimulus, metode yang berorientasi pada subjek, dan metode yang berorientasi pada respons. Penskalaan stimulus berorientasi pada stimulus yakni kuantifikasi titik letak stimulus pada kontinum akan menjadi nilai skala bagi stimulus tersebut. Penskalaan respons adalah prosedur penempatan pilihan jawaban pada suatu kontinum kuantitatif sehingga titik angka pilihan jawaban tersebut menjadi nilai atau skor yang diberikan bagi masing-masing jawaban. Penskalaan subjek bertujuan meletakkan individu-individu pada suatu kontinum penilaian sehingga kedudukan relatif individu menurut suatu atribut yang diukur dapat diperoleh. Prosedur ini memiliki nilai praktis yang tinggi dan banyak digunakan para perancang skala psikologi.

Azwar (2005: 35-40) juga mengemukakan beberapa kaidah penulisan butir yang perlu diperhatikan dalam penyusunan skala psikologi di antaranya adalah: (a) gunakan kalimat dan kata-kata yang jelas, sederhana, dan mudah dimengerti oleh responden tapi tetap mengikuti tata tulis yang benar dan tata bahasa Indonesia yang baku; (b) tulis butir dengan hati-hati sehingga tidak menimbulkan penafsiran ganda pada istilah yang digunakan; (c) selalu mengingat bahwa penulisan butir mengacu pada indikator perilaku atau komponen atribut sehingga jangan menulis butir yang langsung menanyakan atribut yang hendak

diungkap; (d) selalu perhatikan indikator perilaku yang hendak diungkap sehingga stimulus dan pilihan jawaban tetap relevan dengan tujuan pengukuran; (e) cobalah menguji pilihan-pilihan jawaban yang telah ditulis, adakah perbedaan arti atau makna antara dua pilihan yang berbeda sesuai dengan ciri atribut yang sedang diukur, apabila tidak maka butir yang bersangkutan tidak akan memiliki daya beda (*discriminating power*); (f) perhatikan bahwa isi butir tidak boleh mengandung *social desirability* yakni butir yang isinya sesuai dengan keinginan sosial umumnya atau dianggap baik oleh norma sosial, dan (g) untuk menghindari *stereotype* jawaban, sebagian dari butir perlu dibuat dalam arah *favorable* dan sebagian lain dibuat dalam arah tidak *favorable*.

Anderson & Morgan, (2008: 100) juga mengemukakan langkah-langkah penyusunan instrumen yang berbentuk kuisisioner sebagai berikut: (a) tentukan apa tujuan dari kuisisioner dan bagaimana data akan digunakan, (b) mengembangkan *blueprint* yang meliputi: penentuan responden, fokus area, jenis butir, dan *coding* atau *scoring* serta cara mengadministrasikannya, (c) menulis butir, menggunakan kelompok (atau panel) yang terdiri dari orang-orang berpengalaman untuk mereview dan memperbaiki butir, dan desain tata letak sehingga mudah bagi responden untuk menggunakan dan entri data dapat terproses secara efisien, (d) menentukan rencana analisis data untuk memproses informasi yang terkumpul serta menetapkan indikator dan variabel pengukuran untuk analisis statistik berikutnya, (e) uji coba atau uji lapangan kuisisioner untuk menetapkan kesesuaian butir dan kategori respons, dan (f) menganalisis data uji

coba kuesioner, memperbaiki kuesioner, dan menghasilkan kuesioner akhir untuk diadministrasikan.

Pendapat lain tentang penilaian afektif di kelas juga dikemukakan Popham (2003: 109) yang menyatakan bahwa penilaian afektif di kelas dapat dilakukan di antaranya dengan *self report assessment* dan *Likert inventories*. Apabila menggunakan *Likert inventory* menurut Popham (1995: 186) dapat dilakukan dengan langkah-langkah sebagai berikut: (1) memilih variabel sikap yang akan dinilai, (2) membuat serangkaian pernyataan yang bersifat mendukung (*favorable*) dan tidak mendukung (*unfavorable*) mengenai variabel afektif yang akan dinilai tersebut, (3) meminta beberapa orang untuk mengklasifikasikan pernyataan sebagai pernyataan positif atau negatif, (4) menentukan banyaknya pilihan dan pernyataan untuk pilihan respons tersebut pada setiap pernyataan, (5) menyusun inventori yang memuat petunjuk dan arah kepada para siswa tentang cara menanggapi dan mengisi inventori yang harus diselesaikan secara sendiri-sendiri, (6) mengelola penggunaan inventori baik untuk siswa yang dididik sendiri atau jika mungkin (sebagai uji coba) kepada siswa lainnya, (7) melakukan penskoran, dan (8) mengidentifikasi dan mengeliminasi pernyataan yang gagal berfungsi dengan baik dalam pengukuran.

Terkait dengan kegiatan pengukuran atribut psikologis, menurut Azwar (2005: 3) termasuk sangat sukar bahkan mungkin tidak akan pernah dapat dilakukan dengan validitas, reliabilitas, dan objektivitas yang tinggi. Hal ini antara lain disebabkan karena: (a) atribut psikologi bersifat laten atau tidak

tampak, (b) butir-butir skala psikologi didasari indikator-indikator perilaku yang jumlahnya terbatas, (c) respons subjek dipengaruhi faktor lain seperti kondisi, prosedur, administrasi, dan lain-lain, (d) atribut psikologi manusia stabilitasnya tidak tinggi, dan (e) interpretasi bersifat normatif, terdapat lebih banyak sumber kesalahan atau *error*.

Berdasarkan uraian di atas, dapat diperoleh langkah-langkah penyusunan instrumen bentuk non tes guna penilaian sikap dalam pembelajaran matematika yaitu: (1) menentukan tujuan penilaian sikap dan jenis sikap yang hendak dinilai, (2) melakukan kajian teori, (3) memilih teknik penilaian, bentuk instrumen, dan model skala pengukuran, (4) menyusun kisi-kisi, dan (5) menulis butir-butir instrumen, sistem penskoran, dan cara penafsirannya.

8. Rubrik

Rubrik untuk melakukan penskoran pada penilaian hasil belajar siswa dapat dibedakan menjadi rubrik holistik dan rubrik analitik (Nitko, 2001: 102). Rubrik holistik digunakan apabila penskor hanya memberikan satu buah skor (*single rating*) berdasarkan penilaian secara keseluruhan. Artinya, penskoran dilakukan terhadap proses keseluruhan tanpa melakukan penilaian pada bagian komponen secara terpisah. Rubrik holistik biasanya digunakan apabila kesalahan pada bagian dari proses masih dapat ditolerir dan kualitas keseluruhannya masih cukup tinggi. Adapun rubrik analitik, penskoran mula-mula dilakukan atas bagian-bagian secara terpisah, kemudian skor yang diperoleh dijumlahkan untuk

memperoleh skor total. Hal ini berarti bahwa menggunakan rubrik analitik penskoran pada tiap butir dilakukan secara parsial untuk tiap langkah.

Keputusan tentang pemilihan pendekatan holistik atau analitik pada penskoran perlu terlebih dahulu mempertimbangkan penggunaan hasil akhirnya. Apabila diinginkan skor sumatif secara keseluruhan, lebih baik memilih pendekatan holistik. Sebaliknya, jika tujuannya ialah umpan balik formatif, maka sebaiknya menggunakan pendekatan analitik. Komponen dari suatu rubrik dapat berupa deskriptor yang menyatakan apa yang diharapkan dari para siswa pada tiap-tiap tingkatan kompetensi. Sebuah deskriptor menguraikan dengan jelas kepada siswa seperti apa kompetensi itu sebenarnya pada tiap-tiap tingkatan dan bagaimana performans mereka dibedakan dari performans lainnya untuk tiap-tiap kriteria. Namun, ada juga rubrik yang tidak memiliki deskriptor, yakni hanya berupa kriteria-kriteria dan label-label untuk membedakan berbagai tingkatan kompetensi.

Pemilihan jenis rubrik dalam penyusunan suatu rubrik penskoran juga perlu memperhatikan bahwa jenis rubrik yang satu tidaklah lebih baik dari yang lain, yang penting ialah mana yang sesuai untuk tujuan yang diinginkan. Hal ini mengingatkan bahwa menurut Popham (1995: 152), penilaian performans memiliki tiga sumber kesalahan (*sources of error*) yaitu: (a) *scoring-instrument flaws*, pedoman penskoran instrumen yang tidak jelas sehingga sukar untuk digunakan oleh penilai, umumnya karena komponen-komponen tersebut sukar untuk diamati (*unobservable*); (b) *procedural flaws*, prosedur yang digunakan tidak

baik sehingga mempengaruhi hasil penskoran; (c) *teachers' personal-bias errors*, penilai (*rater*) cenderung sukar menghilangkan masalah *personal bias*, yakni ada kemungkinan penilai mempunyai masalah *generosity error* artinya cenderung memberi nilai yang tinggi, walaupun kenyataan sebenarnya hasil pekerjaan siswa tidak baik atau sebaliknya. Masalah lain adalah adanya kemungkinan terjadinya subjektifitas penilai sehingga sukar baginya untuk memberi nilai yang objektif.

Berdasarkan uraian di atas, dapat disimpulkan bahwa penyusunan rubrik penskoran dapat dilakukan dengan pendekatan holistik dan analitik. Selain itu, suatu rubrik penskoran ada yang memiliki deskriptor dan ada pula yang tidak memiliki deskriptor. Rubrik penskoran pada tes uraian dan tugas pada materi matematika akan lebih tepat apabila dilakukan dengan pendekatan analitik. Artinya penskoran dilakukan secara parsial untuk tiap langkah pada tiap butir menggunakan deskriptor yang dirancang sesuai dengan pencapaian kompetensi yang diharapkan dicapai siswa melalui butir tes atau butir tugas tersebut.

9. Kualitas Perangkat Penilaian

Gable (1986: 17) menyarankan tiga hal yang harus dipertimbangkan dalam membuat perangkat penilaian, yaitu validitas, reliabilitas, dan kepraktisan. Adapun menurut Popham (2003: 42), tiga hal yang harus diperhatikan dalam kegiatan pengukuran adalah validitas, reliabilitas, dan *assessment bias*. Senada dengan pendapat Popham, menurut Darling-Hammond et al. (2013: 1), penilaian yang berkualitas tinggi hendaknya memenuhi kriteria valid, reliabel, dan adil. Berdasarkan pendapat-pendapat tersebut, menunjukkan bahwa ada dua hal utama

yang harus diperhatikan terkait kualitas perangkat penilaian yang digunakan dalam kegiatan pembelajaran yakni validitas dan reliabilitas.

a. **Validitas**

Berkenaan dengan validitas, Crocker & Algina (1986: 56) membagi validitas menjadi tiga, yaitu validitas isi (*content validation*), validitas konstruk (*construct validation*), dan validitas yang dihubungkan dengan kriteria (*criterion-related validation*). Menurut Kerlinger (1986: 78), validitas dibagi menjadi tiga, yaitu validitas isi (*content validity*), validitas yang dihubungkan kriteria (*criterion validity*), dan validitas konstruk (*construct validity*).

Secara umum, penilaian terhadap pencapaian hasil belajar menekankan pada validitas isi, penilaian bidang psikologi menekankan pada validitas konstruk, sedangkan penilaian pada sistem seleksi lebih menekankan pada validitas kriteria. Oleh sebab itu, mengingat bahwa penelitian ini lebih berorientasi pada pencapaian hasil belajar siswa pada aspek sikap, pengetahuan, dan keterampilan maka penekanannya lebih kepada validitas isi dan validitas konstruk.

1. **Validitas Isi (*Content Validity*)**

Validitas isi dimaksudkan untuk mengetahui isi dari suatu alat ukur sudah representatif atau belum untuk mengukur domain performans yang hendak diukur (Crocker & Algina, 1986: 218). Validitas isi secara mendasar

dapat diperoleh melalui pendapat ahli (*expert judgment*). Menurut Sireci & Bond (2014: 100) menyatakan bahwa bukti validitas isi suatu instrumen khususnya tes dapat dilakukan dengan pendekatan tradisional dan modern.

a) Bukti Validitas Isi dengan Pendekatan Tradisional

Pembuktian validitas isi suatu instrumen penilaian diperoleh dengan memeriksa sejauh mana butir-butir instrumen mencerminkan dan tidak mencerminkan domain konten. Metode yang paling umum digunakan untuk membuktikan validitas berdasarkan konten adalah melalui *expert judgment* yang mencermati butir dalam hal: (a) menguji sasaran yang diuji, (b) menilai sejauh mana butir mewakili konten yang dimaksudkan, dan (c) menilai sejauh mana butir relevan dengan domain yang diuji (Sireci, 2014: 102).

Terpenuhinya validitas isi, selain dibuktikan secara kualitatif terkait kelayakan instrumen menurut pakar juga perlu diperkuat dengan indeks validitas isi yang menunjukkan tingkat kevalidan instrumen berdasarkan penilaian para pakar. Lawshe (1975: 567) mengusulkan rasio validitas isi (*Content Validity Ratio/CVR*) untuk mengukur derajat kesepakatan para pakar pada satu butir. Tingkat validitas isi dinyatakan dalam indikator tunggal yang besarnya berkisar dari -1 sampai 1. Lawshe mengusulkan bahwa setiap penilai yang terdiri dari panel ahli atau pakar menjawab pertanyaan untuk setiap butir dengan tiga pilihan

jawaban yaitu (1) esensial, (2) berguna tapi tidak esensial, dan (3) tidak diperlukan. Rumus untuk menghitung CVR adalah sebagai berikut.

$$CVR = \frac{n e^{-\left(\frac{N}{2}\right)}}{\frac{N}{2}} \quad (1)$$

Keterangan :

n_e : banyaknya pakar yang memberi respons esensial pada suatu butir,

N : banyaknya pakar.

Perhitungan CVR dilakukan pada tiap butir. Menurut Lawshe, secara umum jika lebih dari setengah panelis menunjukkan bahwa butir adalah penting/esensial, maka butir tersebut setidaknya telah dapat dikatakan memiliki validitas isi yang baik. Selain itu, skor CVR tiap butir juga bisa dibandingkan dengan skor minimal CVR pada taraf penerimaan 0,05 (Lawshe, 1975: 568).

Pendekatan lain terkait koefisien validitas isi diusulkan oleh Aiken (1985: 132). Formula V Aiken untuk menghitung koefisien validitas isi didasarkan pada hasil penilaian dari panel ahli atau pakar sebanyak n orang terhadap suatu butir dari segi sejauh mana butir tersebut mewakili domain atau konstruk yang diukur. Nilai koefisien V Aiken berkisar antara 0–1. Rumus V Aiken adalah sebagai berikut.

$$V = \frac{s}{[n(c-1)]} \quad (2)$$

dimana $s = \sum n_i (r_i - l_o)$

Keterangan :

V : indeks validitas dari Aiken,

n_i : banyaknya pakar yang memilih kriteria i ,

r_i : kriteria ke- i ,

l_o : rating terendah,

n : banyaknya pakar,

c : banyaknya rating/kriteria.

b) Bukti Validitas Isi dengan Pendekatan Modern

Menurut Sireci & Bond (2014: 101), pendekatan baru yang berkembang terkait dengan pembuktian validitas isi adalah melalui uji keselarasan (*alignment*). Keselarasan antara penilaian dan standar dapat didefinisikan sebagai tingkat kesepakatan yang mengukur konsistensi antara standar atau isi kurikulum untuk mata pelajaran tertentu dengan penilaian yang dilakukan untuk mengukur hasil belajar siswa (Bhola, Impara, & Buckendahl, 2003: 21). Hal ini menunjukkan bahwa tujuan uji keselarasan adalah untuk membentuk tingkat kecocokan antara penilaian dan konten mata pelajaran sebagaimana tercantum dalam kompetensi dasar tiap mata pelajaran.

Keselarasan antara penilaian dan standar penting untuk diperhatikan karena ketidakselarasan antara penilaian dan standar dalam kurikulum menjadikan interpretasi hasil penilaian tidak valid dan tidak akurat. Hal ini sebagaimana diungkapkan Biggs (2003: 14) yang

menyatakan bahwa sulit untuk secara akurat memperoleh informasi prestasi siswa yang sesuai dengan tujuan yang diinginkan saat penilaian tidak selaras dengan standar atau kurikulum. Lebih lanjut menurut Wiggins & McTighe (2001: 51), mengungkapkan bahwa tanpa keselarasan menjadikan hasil yang diharapkan tercapai akan terbatas karena siswa tidak akan belajar terkait apa yang sedang dinilai. Penilaian seharusnya memberikan informasi tentang seberapa baik siswa telah mencapai kompetensi yang diharapkan.

Keselarasan merupakan kondisi dasar bagi berfungsinya sistem pendidikan berbasis standar (Smith & O'Day, 1990: 7). Senada dengan pendapat tersebut, Fuhrman (2001: 9) menyatakan bahwa keselarasan merupakan landasan penting yang harus dapat dipenuhi terutama dalam pendidikan berbasis standar. Selain itu, Gamoran et al. (1997: 2) juga menyatakan bahwa keselarasan antara penilaian dan standar berperan dalam menutup kesenjangan prestasi. Hal ini berarti bahwa perbedaan antara siswa berkinerja rendah dan siswa berkinerja tinggi mengalami penurunan. Keselarasan yang tinggi antara penilaian dan standar dalam kurikulum diharapkan dapat meningkatkan hasil belajar siswa, meningkatkan efisiensi suatu reformasi pendidikan, dan menjadi masukan berharga bagi keputusan yang akuntabel.

Studi keselarasan memberikan informasi tentang sejauh mana dan seberapa baik antara penilaian dan kurikulum memfasilitasi dan

mampu meningkatkan hasil belajar siswa. Logika yang mendasari adalah bahwa jika standar menentukan apa dan seberapa baik siswa harus belajar dan tes mengukur apa yang mereka ketahui dan dapat melakukan, maka keduanya harus disinkronkan (Herman & Webb, 2007: 1). Hal ini diperkuat pendapat Airasian & Miranda (2002: 253) yang menyatakan bahwa ketidakselarasan antara penilaian dan standar dalam kurikulum dapat menyebabkan berbagai kesulitan dalam kelanjutan proses pendidikan. Pentingnya keselarasan juga diungkapkan Tindal (2005: 87) yang menyatakan bahwa hasil dari studi keselarasan dapat digunakan untuk menentukan apakah restrukturisasi penilaian diperlukan atau tidak. Jika restrukturisasi diperlukan, hasil keselarasan akan membantu untuk mengidentifikasi perubahan yang diperlukan dalam penilaian.

Hampir semua model atau metode uji keselarasan, butir penilaian dikategorikan berdasarkan pada setidaknya dua kriteria terkait dengan standar dalam kurikulum (Bhola, Impara, & Buchendahl, 2003). Dua kriteria yang paling umum digunakan adalah konten dan kompleksitas kognitif. Definisi konten dan kompleksitas kognitif dalam setiap kriteria pun dapat berbeda di antara metode keselarasan yang berkembang. Selain itu, jumlah kategori dalam setiap kriteria juga bervariasi antar metode. Beberapa metode atau model *alignment* yang berkembang saat ini di antaranya adalah: (a) Model Webb, (b) Metode La Marca, (c)

Metode *Survey of Enacted Curriculum (SEC)*, (d) Metode revisi taksonomi Bloom, dan (e) Metode keselarasan Proyek 2061 dari *American Association for the Advancement of Science (AAAS)*.

Terkait dengan penggunaan revisi taksonomi Bloom untuk digunakan sebagai salah satu metode uji keselarasan, di antaranya karena revisi taksonomi Bloom telah dikenal dan diakui dalam penelitian pendidikan. Selain itu, studi empiris menunjukkan bahwa menggunakan revisi taksonomi Bloom sebagai salah satu alat untuk menguji keselarasan menghasilkan tingkat keandalan yang lebih tinggi antar-penilai daripada beberapa taksonomi lainnya (Nasstrom & Henriksson, 2008). Revisi taksonomi Bloom telah diterapkan pada semua tingkat pendidikan, dari pendidikan pra-sekolah hingga pendidikan tinggi dan hampir di semua jenis pelajaran akademik. Penggunaan taksonomi ini tidak terbatas di Amerika Serikat, tapi telah digunakan hampir di seluruh dunia (Squire, 2001). Hal ini menunjukkan bahwa penggunaan taksonomi ini telah diakui dan digunakan oleh banyak negara termasuk Indonesia.

Penggunaan metode revisi taksonomi Bloom untuk uji keselarasan mengacu pada Anderson & Krathwohl (2001: 101-104) yang mengemukakan bahwa revisi taksonomi Bloom juga bermanfaat untuk menilai kesesuaian antara tujuan pembelajaran dan penilaian. Adapun langkah-langkahnya adalah: pertama, mengidentifikasi tujuan-

tujuan pokok pembelajaran dan menentukan kotak-kotak pada tabel taksonomi yang relevan; kedua, mengidentifikasi penilaian-penilaian pokoknya dan menentukan kotak-kotak pada tabel taksonomi yang relevan. Berdasarkan kedua langkah tersebut, perhatikan apakah penekanan pada setiap tujuan tercermin dalam penilaiannya. Jika kotak-kotak dan penekanan yang dihasilkan oleh dua langkah tersebut bersesuaian, berarti memang terdapat kesesuaian antara tujuan dan penilaiannya.

Perkembangan dalam uji keselarasan menunjukkan bahwa metode revisi taksonomi Bloom untuk menguji keselarasan antara penilaian dan standar dalam kurikulum dapat dimodifikasi dengan model Porter dalam hal perhitungan indeks keselarasannya. Rentang indeks keselarasan dari 0 (tidak ada keselarasan) hingga 1 (keselarasan sempurna). Hal ini sebagaimana telah dilakukan Nasstrom & Henriksson (2008) yang melakukan uji keselarasan antara penilaian dan standar dalam kurikulum dengan menggunakan revisi taksonomi Bloom terkait kompleksitas kognitifnya. Adapun rumus indeks keselarasan dari Porter (P) adalah sebagai berikut.

$$P = 1 - \frac{\sum_{k=1}^k \sum_{j=1}^j |a_{jk} - b_{jk}|}{2} \quad (3)$$

Keterangan:

j : banyaknya baris,

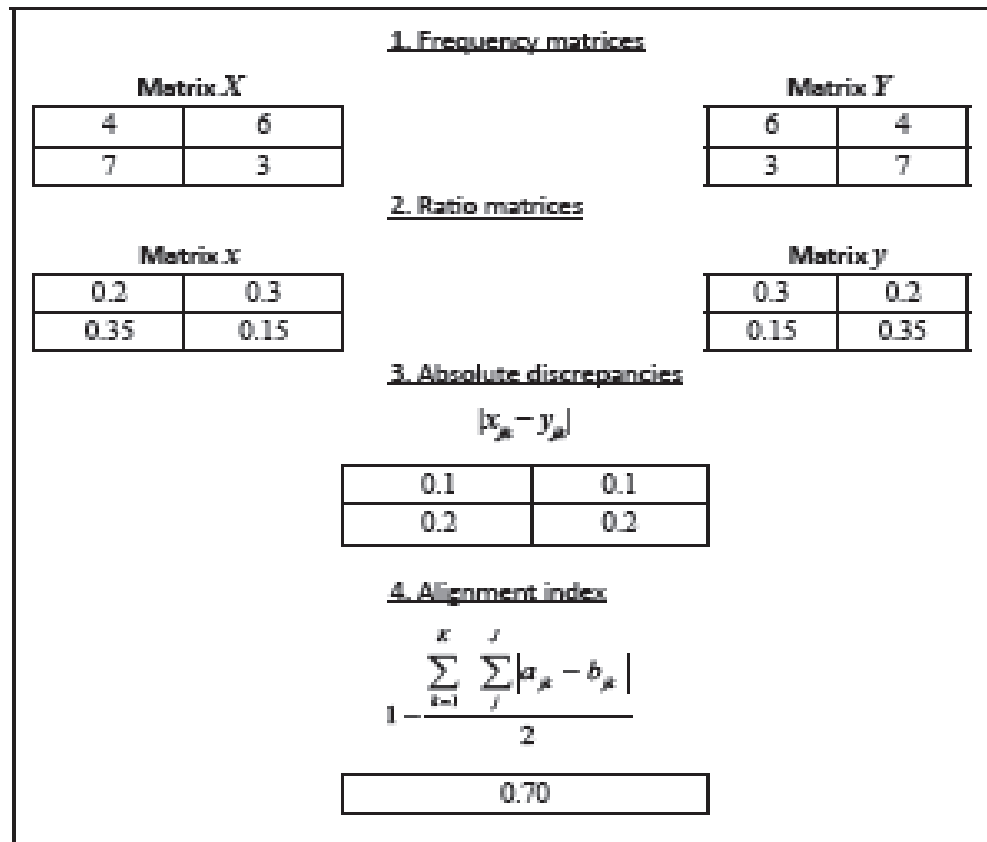
k : banyaknya kolom dalam setiap matriks X dan Y ,

a_{jk} dan b_{jk} : rasio dalam sel-sel pada baris j dan kolom k untuk masing-masing matriks rasio x dan y .

Secara umum, indeks keselarasan Porter (P) dapat ditentukan dalam empat langkah sebagai berikut (Ndiovu & Mji, 2012: 4):

- 1) Membuat matriks frekuensi untuk dua dokumen yang dibandingkan, misalnya diberi label sebagai matriks X dan matriks Y .
- 2) Setiap sel dalam matriks X dan Y , dihitung rasio dalam sel dengan cara membandingkan angka dalam sel dengan jumlah angka dalam masing-masing matriks dan beri label matriks rasio sebagai matriks x dan matriks y .
- 3) Setiap baris j dan kolom k dalam matriks rasio, dihitung nilai absolut dari perbedaan antara rasio dalam sel x_{jk} dan y_{jk} atau $|x_{jk} - y_{jk}|$,
- 4) Menghitung indeks keselarasan dengan rumus Porter. Mengacu rumus Porter pada persamaan 3, dalam hal ini $|x_{jk} - y_{jk}|$ dimaksudkan sebagai $|a_{jk} - b_{jk}|$.

Sebagai contoh sederhana perhitungan indeks keselarasan dari Porter untuk matriks 2x2 disajikan pada Gambar 3 sebagai berikut.



Gambar 3. Contoh Perhitungan Indeks Keselarasan Porter untuk Matriks 2x2 (Ndiovu & Mji, 2012: 4).

2) Validitas Konstruk (*Construct Validity*)

Validitas konstruk merujuk pada sejauhmana suatu instrumen mengukur suatu konstruk teoretik atau trait yang hendak diukur (Allen & Yen, 1979: 108). Hal ini berarti validitas konstruk mempersoalkan sejauh mana skor-skor hasil pengukuran dengan instrumen yang digunakan merefleksikan konstruk teoretik yang mendasari penyusunan instrumen tersebut. Mardapi (2008: 21) juga menegaskan bahwa validitas konstruk

menunjukkan sejauhmana instrumen mengungkap suatu konstruk teori yang diukurnya atau suatu trait yang dikembangkan dalam penyusunan instrumen. Menurut Nunnaly (1981: 141), beberapa aspek penting terkait validitas konstruk adalah: (a) menspesifikasikan domain yang diukur dengan jelas dan dapat dilihat, (b) menentukan seberapa kuat keterkaitan antar aspek pada variabel yang diukur, dan (c) menentukan keberadaan dan seberapa pentingnya variabel yang dijelaskan dari konstruk. Salah satu metode yang diakui para ahli dan sering digunakan untuk menentukan validitas konstruk adalah analisis faktor.

Kerlinger (1986) menyatakan bahwa maksud dan kegunaan dasar analisis faktor ada dua yaitu: (a) mengeksplorasi wilayah-wilayah variabel guna mengetahui dan menunjukkan faktor-faktor yang diduga melandasi variabel-variabel itu, dan (b) menguji hipotesis tentang relasi-relasi antar variabel. Analisis faktor dibedakan menjadi dua yaitu *Exploratory Factor Analysis* (EFA) dan *Confirmatory Factor Analysis* (CFA).

Model analisis yang cukup komprehensif dan banyak digunakan untuk uji validitas konstruk adalah analisis faktor konfirmatori (*Confirmatory Factor Analysis*/CFA). Hair et al. (2010: 663) mengemukakan bahwa CFA merupakan salah satu cara untuk menguji seberapa baik variabel-variabel yang dihadirkan atau ditetapkan dapat mewakili suatu kostruk. Hal ini menunjukkan bahwa penggunaan CFA dilakukan jika faktor penyusun konstruk telah ditentukan terlebih dahulu.

Pengujian menggunakan CFA ini merupakan kebalikan dari *Exploratory Factor Analysis (EFA)*. Artinya, *Exploratory Factor Analysis (EFA)* digunakan jika jumlah faktor yang akan terbentuk tidak ditentukan terlebih dahulu.

Analisis faktor konfirmatori dapat dilakukan dengan model persamaan struktural (*Structural Equation Modeling/SEM*). Model persamaan struktural memiliki dua bagian penting yaitu model pengukuran yang dilaksanakan dengan analisis konfirmatori dan model struktural yang dilakukan dengan model regresi. Model pengukuran berfungsi untuk mengetahui sejauhmana ketepatan pengukuran yang telah dilakukan sedangkan model struktural untuk menguji suatu teori atau hubungan antar konsep. Menurut Jöreskog & Sörbom (1996: 115), model pengukuran digunakan untuk menetapkan variabel terukur yang tepat sebagai indikator pada variabel laten.

Dua alasan yang mendasari penggunaan SEM menurut Yamin & Kurniawan (2009: 3-4) adalah: (1) SEM mempunyai kemampuan untuk mengestimasi hubungan antar variabel yang bersifat *multiple relationship*. Hubungan ini dibentuk dalam model struktural yang menggambarkan hubungan antara konstruk *dependent* dan *independent*; (b) SEM mempunyai kemampuan untuk menggambarkan pola hubungan antara konstruk laten (*unobserved*) dan variabel manifes (*manifest variable* atau variabel indikator). Analisis faktor menggunakan SEM dapat dilakukan dengan

bantuan *software Lisrel*. Adapun terkait dengan ukuran sampel dalam SEM, Comrey & Lee (1992) yang dikutip MacCallum et al. (1999: 840) mengemukakan kriteria ukuran sampel sebagai berikut: 100 (*poor*), 200 (*fair*), 300 (*good*), 500 (*very good*), dan ≥ 1000 (*excellent*).

Metode estimasi parameter pada SEM meliputi: *Instrument Variable* (IV), *Two Stage Least Square* (2SLS), *Unweighted Least Square* (ULS), *Generalized Least Square* (GLS), *Maximum Likelihood* (ML), *Weighted Least Square* (WLS), dan *Diagonally Weighted Least Square* (DWLS) (Wijanto, 2008: 44). Adapun pengujian hipotesis pada SEM mengasumsikan bahwa: (a) model persamaan struktural yang dipakai sebagai landasan adalah benar, dan (b) data yang digunakan untuk menguji model mengikuti distribusi normal multivariat bersama (*joint multivariate normal distribution*).

Jika data sampel yang dianalisis tidak memenuhi asumsi distribusi normal multivariat bersama maka nilai statistik *chi square* yang dipakai dalam menunjukkan ketepatan model fit secara keseluruhan akan meningkat dan kesalahan standar yang dipakai untuk menguji signifikansi parameter akan menurun. Apabila data yang dianalisis ternyata tidak memenuhi asumsi distribusi normal multivariat bersama maka ada empat pendekatan umum yang dapat digunakan untuk menangani data non normal tersebut yakni: (a) menggunakan estimator yang tahan terhadap ketidaknormalan data misalnya GLS dan WLS, (b) menggunakan penskalaan kuat (*robust*) dan penyesuaian

chi square, (c) menggunakan teknik *bootstrapping* untuk menghitung nilai kritis kai kuadrat, nilai parameter dan kesalahan standar, dan (4) memadukan indikator dalam sebuah paketan butir (Widhiarso, 2012: 1-2).

Penerapan model SEM pada tes hasil belajar juga perlu dilakukan terutama pada tes yang mengandung beberapa Kompetensi Dasar (KD). Melalui penerapan SEM, dapat diketahui signifikansi indikator yang mewakili KD terhadap kompetensi pokok yang diharapkan. Selain itu juga dapat diketahui eksistensi dan urutan suatu KD dalam sumbangannya terhadap kompetensi pokok. Walaupun tes hasil belajar pada dasarnya telah mempunyai domain yang jelas, namun tidak berarti tidak diperlukan lagi pembuktian terhadap validitas konstruknya. Bukti validitas konstruk pada tes hasil belajar dapat digunakan sebagai konfirmasi antara validitas isi dari tes yang telah dibuat dengan data empirik setelah dilakukan uji coba. Selain itu, hasil uji validitas konstruk juga dapat digunakan untuk menunjukkan unidimensionalitas alat ukur berdasarkan data hasil pengukuran yang diperoleh dalam uji coba.

3) Reliabilitas

Instrumen penilaian yang baik, selain valid hendaknya juga reliabel. Instrumen penilaian yang reliabel akan memberikan hasil yang relatif sama pada setiap pengukuran, meskipun waktu pengukuran berbeda. Menurut Nunnally (1981: 12), reliabilitas adalah ketetapan atau keajegan suatu instrumen penilaian yang memberikan hasil yang relatif sama dalam waktu

yang berlainan. Hal ini berarti reliabilitas mengandung muatan stabilitas (tidak berubah-ubah) dan konsisten (ajeg). Bila instrumen dipakai berulang-ulang untuk mengukur gejala yang sama dan hasil yang diperoleh relatif stabil atau konsisten, maka instrumen tersebut dapat dikatakan reliabel.

Secara empiris, tinggi rendahnya reliabilitas ditunjukkan oleh suatu angka yang disebut koefisien reliabilitas. Makin tinggi koefisien reliabilitas, makin dekat nilai skor amatan dengan skor yang sesungguhnya, sehingga nilai skor amatan dapat digunakan sebagai pengganti komponen skor yang sesungguhnya. Makin tinggi koefisien reliabilitas sesuatu instrumen, maka kemungkinan kesalahan yang terjadi akan makin kecil ketika orang membuat keputusan berdasarkan skor yang diperoleh.

Koefisien reliabilitas suatu instrumen dapat diperoleh dengan memperhatikan konsistensi internalnya. Salah satu metode yang populer dalam psikometri dan sering digunakan untuk menentukan reliabilitas berdasarkan konsistensi internal adalah koefisien α dari Cronbach (*Cronbach's Alpha*). Formula untuk menghitung koefisien α adalah sebagai berikut (Cronbach, 1951: 299).

$$\alpha = \frac{p}{p-1} \left(1 - \frac{\sum_{i=1}^p \sigma_{x_i}^2}{\sigma_u^2} \right) \quad (4)$$

Keterangan:

p : banyaknya butir x_i ,

u : $x_1 + x_2 + \dots + x_p$.

Selain *Cronbach's Alpha*, pada penggunaan SEM juga dapat diperoleh reliabilitas konstruk (*Construct Reliability/CR*). Formula untuk menghitung CR adalah sebagai berikut (Wijanto, 2008: 175).

$$\text{Construct Reliability} = \frac{(\sum \text{std.loading})^2}{(\sum \text{std.loading})^2 + \sum e_j} \quad (5)$$

Keterangan:

std. loading: *Standarized Loading Factor* (SLF),

e: *error variances*.

Hair, et al (2010: 688) mengemukakan bahwa estimasi $CR > 0,7$ tergolong baik, sementara CR antara 0,6 dan 0,7 tergolong dapat diterima, asalkan indikator validitas konstruk model tergolong baik.

Menurut Umar (2014: 394), pada dasarnya tidak ada perbedaan yang mendasar antara koefisien *Cronbach's Alpha* dengan *Construct Reliability* (CR). Hal ini juga ditekankan Widhiarso (2009: 41) yang mengemukakan bahwa penggunaan koefisien *Cronbach's Alpha* dapat dilakukan apabila dari hasil analisis faktor ditemukan struktur data yang bersifat unidimensi. Pendapat-pendapat tersebut menunjukkan bahwa penggunaan koefisien *Cronbach's Alpha* dengan *Construct Reliability* (CR) lebih sesuai untuk pengukuran yang bersifat unidimensi.

Widhiarso (2009: 42-47) menyatakan bahwa jika koefisien α diaplikasikan pada pengukuran yang bersifat multidimensi maka akan didapatkan hasil yang *underestimate*. Oleh sebab itu, apabila data memiliki

struktur multidimensi maka estimasi reliabilitas sebaiknya menggunakan koefisien reliabilitas yang dapat diaplikasikan pada model pengukuran multidimensi. Beberapa koefisien reliabilitas yang dapat digunakan untuk model pengukuran multidimensi tersebut secara umum dapat dibedakan dalam dua jenis yaitu: (1) koefisien reliabilitas yang diawali dengan analisis faktor konfirmatori dengan pendekatan SEM, meliputi: koefisien reliabilitas konstruk Omega dari McDonald dan koefisien reliabilitas komposit Raykov; dan (2) koefisien reliabilitas yang tidak memerlukan prosedur analisis faktor, meliputi: koefisien *Cronbach's Alpha* berstrata atau koefisien reliabilitas komposit Mosier.

10. Teori Respons Butir untuk Data Poltomus

Instrumen penilaian dalam pembelajaran matematika seperti tes uraian dan tugas untuk mengukur kinerja atau hasil belajar siswa pada aspek pengetahuan dan keterampilan pada umumnya disusun dengan menggunakan format respons lebih dari dua kategori (politomus). Hal ini dilakukan dengan pertimbangan agar instrumen tersebut memberikan informasi lebih banyak dibandingkan dengan tes yang diskor secara dikotomus.

Berdasarkan data respons butir yang diskor lebih dari dua kategori (politomus) tersebut, model teori respons butir politomus diperlukan untuk menggambarkan hubungan yang tidak linear antara penempuh tes dengan kemampuan θ dan probabilitas penempuh tes menjawab respons butir pada

kategori tertentu. Asumsi teori respons butir, model teori respons butir politomus, dan pengertian fungsi informasi dalam teori respons butir diuraikan sebagai berikut.

a. Asumsi Teori Respons Butir

Model respons dalam teori respons butir mempunyai makna bahwa probabilitas subjek untuk menjawab butir dengan benar tergantung pada kemampuan subjek dan karakteristik butir. Hal ini berarti siswa dengan kemampuan tinggi akan mempunyai probabilitas menjawab benar lebih besar jika dibandingkan dengan siswa yang mempunyai kemampuan rendah. Hambleton & Swaminathan (1985: 16) dan Hambleton, Swaminathan, & Rogers (1991: 9) menyatakan bahwa ada dua asumsi utama yang mendasari teori respons butir yaitu unidimensi dan independensi lokal.

Unidimensi berarti bahwa tes hanya berisi satu kemampuan yang diukur oleh butir-butir yang disusun pada tes tersebut. Asumsi unidimensi dapat dibuktikan jika tes mengandung hanya satu komponen dominan yang mengukur performansi suatu subjek. Adapun independensi lokal, terjadi jika kemampuan-kemampuan yang mempengaruhi performansi dijadikan konstan, maka respons subjek terhadap pasangan butir yang manapun akan independen secara statistik satu sama lain. Asumsi ini akan terpenuhi apabila jawaban peserta terhadap sebuah butir soal tidak mempengaruhi jawaban peserta terhadap butir soal yang lain.

b. Model Teori Respons Butir Politomus

Selain asumsi-asumsi yang harus dipenuhi, hal penting yang perlu diperhatikan dalam teori respons butir adalah pemilihan model respons butir yang tepat. Pemilihan model yang tepat akan mengungkap keadaan yang sesungguhnya dari data tes sebagai hasil pengukuran. Ada beberapa model respons butir yang dapat digunakan untuk menganalisis butir tes yang penskorannya bersifat politomus. Beberapa model teori respons butir politomus yang telah diusulkan para ahli antara lain: *Nominal Response Model (NRM)*, *response model for multiple-choice items*, *Rating Scale Model (RSM)*, *Partial Credit Model (PCM)*, *Graded Response Model (GRM)*, *sequential model for ordered response*, dan *Generalized Partial Credit Model (GPCM)* (Van der Linden & Hambleton, 1997: 30). Semua model tersebut berdasarkan asumsi bahwa respons penempuh tes pada suatu butir tes bergantung pada kemampuan penempuh tes yang diukur menggunakan skala unidimensi.

Pada dasarnya model respons butir politomus dapat dikategorikan dalam dua macam model, yaitu model respons nominal dan ordinal, tergantung pada asumsi tentang karakteristik data. Model respons nominal dapat diterapkan pada butir yang mempunyai alternatif jawaban yang tidak terurut dengan adanya berbagai tingkat kemampuan yang diukur, sedangkan model respons ordinal digunakan bila respons peserta pada sebuah butir dapat diskor dalam banyaknya kategori tertentu yang tersusun dalam

kecakapan. Sebagai contoh, butir-butir skala sikap tipe Likert yang diskor menggunakan pedoman penskoran kategori respons terurut, merupakan penskoran ordinal. Adapun butir-butir tes matematika dapat diskor dengan menggunakan sistem kredit parsial yakni poin-poin untuk melengkapi langkah menuju jawaban benar dihargai dan ini juga merupakan penskoran ordinal.

Beberapa model teori respons butir politomus yang sering digunakan oleh para ahli di antaranya adalah *Graded Response Model (GRM)*, *Partial Credit Model (PCM)*, dan *Generalized Partial Credit Model (GPCM)*. Penggunaan masing-masing model tersebut secara ringkas dapat diuraikan sebagai berikut.

1) *Graded Response Model (GRM)*

Model GRM tepat digunakan ketika respons penempuh tes terhadap butir dapat digolongkan sebagai respons kategori yang berurutan dan tingkatan penyelesaiannya cenderung meningkat. GRM adalah generalisasi dari model logistik dua parameter (2-PL) pada model teori respons butir dikotomus. Respons peserta suatu butir j dalam GRM dapat dikategorikan menjadi $m + 1$ skor kategori terurut, $k = 0, 1, 2, \dots, m$, dimana m banyaknya langkah dalam menyelesaikan dengan benar butir j , dan indeks kesukaran dalam setiap langkah juga terurut. Tingkat kesukaran tiap kategori butir tes pada model GRM disusun secara berurutan sehingga jawaban peserta tes terurut dari kategori yang rendah

hingga kategori yang tinggi. Model GRM tepat digunakan ketika respons penempuh tes terhadap butir dapat digolongkan sebagai respons kategori yang berurutan dan tingkatan penyelesaiannya cenderung meningkat, seperti yang ada pada skala penilaian *Likert*.

Samejima mengasumsikan bahwa respons penempuh tes pada butir ke i diklasifikasikan ke dalam $m_i + 1 = K_i$ urutan kategori. Oleh sebab itu, penempuh tes yang memperoleh skor kategori tinggi menunjukkan kemampuan yang lebih tinggi dari penempuh tes yang memperoleh skor kategori rendah. Skor kategori untuk butir i berupa bilangan bulat x , dimana $x=0, 1, 2, \dots, m$. Sebagai contoh, butir dengan jumlah kategori respons butir $K = 5$, penempuh tes memperoleh skor $x = 0, 1, 2, 3, 4$. Penskoran dengan lima pilihan jawaban, terdapat $m_i = 4$ parameter tingkat kesukaran (*threshold*) ($j = 1, 2, 3, 4$) antara pilihan respons. Hubungan parameter butir dan kemampuan peserta untuk kasus homogen (a_j sama dalam setiap langkah) dalam GRM dinyatakan sebagai berikut (Muraki, & Bock, 1997: 7).

$$P_{jk}(\theta) = P_{jk}^*(\theta) - P_{j,k+1}^*(\theta) \quad (6)$$

dengan,

$$P_{jk}^*(\theta) = \frac{\exp[Da_j(\theta - b_{jk})]}{1 + \exp[Da_j(\theta - b_{jk})]}$$

$$P_{j0}^*(\theta) = 1 \text{ dan } P_{j,m+1}^*(\theta) = 0$$

Keterangan:

a_j : indeks daya beda butir j,

b_{jk} : indeks kesukaran kategori k butir j,

θ : kemampuan peserta,

$P_{jk}(\theta)$: probabilitas peserta berkemampuan θ yang memperoleh skor kategori k pada butir j,

$P_{jk}^*(\theta)$: probabilitas peserta berkemampuan θ yang memperoleh skor kategori k atau lebih pada butir j,

D : faktor skala ($D = 1,7$).

2) *Partial Credit Model* (PCM)

PCM merupakan perluasan model Rasch, diasumsikan bahwa setiap butir mempunyai daya beda yang sama. Seperti halnya pada GRM, PCM juga cocok untuk butir yang diskor dalam kategori berjenjang, tetapi indeks kesukaran dalam setiap langkah tidak perlu terurut, suatu langkah dapat lebih sukar dari pada langkah berikutnya. Menurut Muraki & Bock (1997: 16), PCM mempunyai bentuk umum sebagai berikut.

$$P_{jk}(\theta) = \frac{\exp \sum_{v=0}^k (\theta - b_{jv})}{\sum_{h=0}^m \exp \sum_{v=0}^h (\theta - b_{jv})}, k = 0, 1, 2, \dots, m \quad (7)$$

Keterangan:

$P_{jk}(\theta)$: probabilitas peserta berkemampuan θ memperoleh skor kategori k pada butir j,

θ : kemampuan peserta,

$m + 1$: banyaknya kategori butir j,

b_{jk} : indeks kesukaran kategori k pada butir j,

dan didefinisikan bahwa: $\sum_{h=0}^k (\theta - b_{jh}) \equiv 0$, $\sum_{h=0}^h (\theta - b_{jh}) \equiv \sum_{h=1}^h (\theta - b_{jh})$

Skor kategori suatu butir dalam PCM menunjukkan banyaknya langkah untuk menyelesaikan dengan benar butir tersebut. Jadi, skor kategori yang lebih tinggi menunjukkan kemampuan yang lebih tinggi dari pada skor kategori yang lebih rendah. Khusus pada PCM, jika suatu butir memiliki dua kategori, maka persamaan 7 menjadi persamaan model Rasch. Oleh sebab itu, PCM dapat diterapkan untuk butir politomus dan dikotomus.

3) *Generalized Partial Credit Model (GPCM)*

GPCM merupakan bentuk umum dari PCM, GPCM dinyatakan dengan bentuk umum sebagai berikut (Muraki & Bock, 1997: 154).

$$P_{jh}(\theta) = \frac{\exp\left[\sum_{v=0}^h Z_{jv}(\theta)\right]}{\sum_{e=1}^{m_j} \exp\left[\sum_{v=0}^e Z_{jv}(\theta)\right]} , k = 0, 1, 2, \dots, m_j \quad (8)$$

dan, $Z_{jh}(\theta) = Da_j(\theta - b_{jh}) = Da_j(\theta - b_j + d_h)$, $b_{j0} = 0$,

Keterangan:

- $P_{jk}(\theta)$: probabilitas peserta dengan kemampuan θ memperoleh skor kategori k pada butir j,
- θ : kemampuan peserta,
- a_j : indeks daya beda butir j,
- b_{jk} : indeks kesukaran kategori k pada butir j,
- b_j : indeks kesukaran lokasi butir j disingkat parameter butir lokasi,
- d_k : parameter kategori k,
- m_j+1 : banyaknya kategori butir j,
- D : faktor skala ($D=1,7$).

Parameter b_{jk} , oleh Masters (Muraki & Bock, 1997: 17) disebut sebagai parameter tahap butir. Parameter ini merupakan titik potong antara kurva $P_{jk}(\theta)$ dan $P_{jk-1}(\theta)$. Kedua kurva tersebut hanya berpotongan di satu titik pada skala θ . Adapun untuk $a_j > 0$ berlaku:

Jika $\theta = b_{jk}$, maka $P_{jk}(\theta) = P_{jk-1}(\theta)$

Jika $\theta > b_{jk}$, maka $P_{jk}(\theta) > P_{jk-1}(\theta)$

Jika $\theta < b_{jk}$, maka $P_{jk}(\theta) < P_{jk-1}(\theta)$, $k = 1, 2, 3, \dots, m_j$.

Nilai b_{jk} bersifat tidak terurut dan $P_{jk}(\theta)$ merupakan fungsi respons kategori butir dari GPCM.

Berbagai penelitian telah dilakukan untuk mengetahui ketepatan penggunaan model-model tersebut. Dodd melakukan analisis dengan GRM dan GPCM terhadap pengukuran sikap (De Ayala, 1993:179).

Model-model ini digunakan untuk menganalisis sikap guru terhadap keterampilan berkomunikasi dan analisis kemampuan penalaran non verbal. Selain itu, penerapan GPCM cocok dengan kondisi penskoran respons butir tes berbentuk uraian dalam praktik di lapangan, yaitu skor butir diberikan berdasarkan banyaknya langkah yang dijawab benar tanpa memperhatikan urutan langkah.

Estimasi parameter butir dan kemampuan pada IRT politomus di antaranya dapat dilakukan dengan bantuan software *Parscale* dari SSI (Muraki & Bock, 1997). Selain memperhatikan terpenuhinya asumsi dalam teori respons butir politomus, faktor lain yang perlu diperhatikan dalam mengestimasi parameter butir dan kemampuan adalah ukuran sampel. Ukuran sampel yang diperlukan untuk memperoleh nilai parameter butir dan kemampuan yang memenuhi syarat terkait dengan model yang digunakan menurut Reise & Yu (De Ayala, 1993: 178) adalah minimal 500. Hal ini didukung Jurich and Goodman (2009) yang juga menyatakan bahwa dalam analisis menggunakan program *Parscale* sebaiknya dilakukan minimal dengan 500 peserta tes.

Sebelum diperoleh estimasi kemampuan peserta tes, terlebih dahulu dilakukan kegiatan kalibrasi butir yakni memperoleh informasi mengenai estimasi parameter butir dan mengecek kecocokan model. Kalibrasi ini dapat dilakukan jika data respons peserta terhadap tes telah diperoleh (Du Toit, 2003). Ada dua pendekatan yang dapat digunakan

untuk memperoleh estimasi parameter butir yakni *Marginal Maximum Likelihood* (MML) dan *Marginal Maximum A Posteriori* (MMAP). MML efisien untuk semua model respons butir dan bentuk tes yang panjang maupun yang pendek.

c) Fungsi Informasi

Berdasarkan teori respons butir, istilah fungsi informasi digunakan untuk butir dan tes. Fungsi informasi butir merupakan suatu metode untuk menjelaskan kekuatan suatu butir pada perangkat soal dan menyatakan kekuatan atau sumbangan butir soal dalam mengungkap kemampuan laten (*latent trait*) yang diukur dengan tes tersebut. Melalui fungsi informasi butir diketahui butir yang cocok dengan model sehingga membantu dalam seleksi butir soal. Adapun fungsi informasi tes merupakan jumlah dari fungsi informasi butir-butir tes tersebut (Hambleton & Swaminathan, 1985: 94). Berkaitan dengan hal ini, nilai fungsi informasi tes akan tinggi jika butir-butir penyusun tes mempunyai fungsi informasi yang tinggi pula. Hambleton & Swaminathan (1985: 101) menyatakan bahwa fungsi informasi sangat berguna untuk: (1) konstruksi tes, (2) pemilihan butir, (3) penilaian presisi pengukuran, (4) komparasi sejumlah tes, (5) penentuan bobot dan penskoran, dan (6) komparasi berbagai metode penskoran. Secara umum, fungsi informasi tes dinyatakan sebagai berikut (Hambleton, Swaminathan, & Rogers, 1991: 94):

$$I(\theta) = \sum_{i=1}^n I_i(\theta) \quad (9)$$

Keterangan:

$I(\theta)$: fungsi informasi tes yang terdiri atas n butir,

$I_i(\theta)$: fungsi informasi butir ke-i.

Fungsi informasi butir pada model logistik dengan probabilitas jawaban benar ($P_i(\theta)$) dan probabilitas jawaban salah ($Q_i(\theta)=1-P_i(\theta)$) dirumuskan sebagai berikut:

$$I_i(\theta) = \frac{[P_i'(\theta)]^2}{P_i(\theta)Q_i(\theta)} \quad (10)$$

$I_i(\theta)$ adalah informasi yang diberikan butir i pada θ sedangkan $P_i'(\theta)$ adalah derivasi $P_i(\theta)$ terhadap θ ((Hambleton, Swaminathan, & Rogers, 1991: 91). Derivasi $P_i(\theta)$ terhadap θ di atas menunjukkan koefisien arah lengkungan respons butir. Koefisien arah ini berubah-ubah menurut nilai θ .

Fungsi informasi butir $I_j(\theta)$, pada model politomus adalah informasi yang disumbangkan oleh butir tertentu j pada kisaran θ . Fungsi informasi butir pada respons butir politomus diberikan oleh Samejima (Muraki, 1993: 353) sebagai berikut.

$$I_j(\theta) = D^2 a_j^2 \sum_{c=1}^{m_j} |T_c - \bar{T}_j(\theta)|^2 P_{jc}(\theta) \quad (11)$$

dimana

$$\bar{T}_j(\theta) = \sum_{c=1}^{m_j} T_c P_{jc}(\theta)$$

$I_j(\theta)$ adalah fungsi informasi butir j, D adalah tetapan yang dapat mempunyai nilai 1 atau 1,7, a_j merupakan parameter daya beda butir j. $\bar{T}_j(\theta)$ adalah fungsi respons butir untuk butir yang diskor politomus. Adapun fungsi informasi tes dapat dihitung dengan rumus sebagai berikut.

$$I(\theta) = \sum_{j=1}^n I_j(\theta) \quad (12)$$

Nilai-nilai parameter butir dan kemampuan peserta merupakan hasil estimasi. Oleh sebab itu, kebenarannya bersifat probabilistik dan tidak terlepas dari adanya kesalahan pengukuran. Kesalahan pengukuran standar (*Standard Error of Measurement, SEM*) dalam teori respons butir berkaitan erat dengan fungsi informasi. Fungsi informasi dengan SEM mempunyai hubungan yang berbanding terbalik kuadratik, semakin besar fungsi informasi maka SEM semakin kecil atau sebaliknya. Jika nilai fungsi informasi dinyatakan dengan $I(\theta)$ dan nilai estimasi SEM dinyatakan dengan $\hat{SEM}(\theta)$ maka hubungan keduanya, menurut Hambleton, Swaminathan, & Rogers (1991: 94) dinyatakan sebagai berikut.

$$\hat{SEM}(\theta) = \frac{1}{\sqrt{I(\theta)}} \quad (13)$$

Kriteria untuk menentukan karakteristik instrumen tes yang baik, harus dipilih sejumlah butir yang memiliki fungsi informasi tinggi. Namun, berbagai literatur tidak ada yang menyebutkan besarnya fungsi informasi minimal yang harus dimiliki oleh suatu butir yang baik.

B. Kajian Penelitian yang Relevan

Sejumlah penelitian telah dilakukan terkait dengan pengembangan instrumen penilaian dalam pembelajaran matematika beserta studi validitas dan reliabilitasnya. Berbagai penelitian tentang studi keselarasan antara penilaian dengan standar dalam kurikulum tertentu juga telah dilakukan. Berikut beberapa penelitian yang relevan.

Terkait dengan kegiatan penilaian dalam kelas matematika, Suurtamm, Koch, & Arden (2010) dari Kanada meneliti tentang praktek penilaian yang dilakukan para guru di kelas matematika. Hasil penelitian menunjukkan bahwa para guru menggunakan berbagai bentuk dan teknik penilaian untuk meningkatkan hasil belajar siswa. Praktek penilaian yang dilakukan di antaranya menggunakan tes, jurnal, observasi, *self assessment*, dan kuis. Praktek ini ternyata memungkinkan guru untuk mengetahui pemikiran matematika siswa dan menjadi bahan umpan balik yang diberikan kepada siswa. Selain itu, para guru juga menggunakannya untuk meningkatkan pembelajaran. Hasil penelitian juga menunjukkan bahwa kolaborasi, koherensi, dan keyakinan guru sangat mendukung dalam penerapan praktek penilaian yang dilakukan.

Secara lebih spesifik, penelitian tentang pengembangan instrumen juga dilakukan oleh Suksawang, Sujiva, & Soydhurum (2010) yang melakukan penelitian tentang pengembangan tes hasil belajar matematika untuk siswa berbakat di kelas 10 dengan menggunakan proses berpikir tingkat tinggi di Thailand. Sampel penelitian adalah 241 siswa kelas 10 dari sekolah menengah negeri yang mengambil program matematika dan sains. Tes yang dikembangkan terdiri atas 28 butir. Berdasarkan hasil

penelitian diperoleh bahwa dalam hal kemampuan matematika, 2,90% dari sampel menunjukkan kemampuan ekstrim; 31,12% dari sampel menunjukkan kemampuan yang luar biasa; 39,83% dari sampel menunjukkan kemampuan yang tinggi; 14,11% dari sampel menunjukkan kemampuan sedang; dan 12,03% dari sampel menunjukkan kemampuan rendah.

Penelitian tentang pengembangan instrumen juga dilakukan Wheeler (2010) di Amerika Serikat. Pengembangan instrumen dalam penelitian ini dilakukan untuk materi persamaan. Berdasarkan hasil penelitian diperoleh delapan indikator yang harus dicapai siswa terkait pemahaman tentang persamaan dan dihasilkan instrumen yang terdiri dari 18 butir yang bersifat dikotomus dalam bentuk pilihan, benar/salah, dan isian kosong. Instrumen yang dihasilkan telah melalui uji lapangan, *review* ahli, dan wawancara dengan siswa serta uji reliabilitas yang memberikan bukti bahwa instrumen terdiri dari butir-butir yang dianggap efektif.

Adapun terkait penelitian pengembangan instrumen berbentuk skala sikap di antaranya dilakukan Yesil (2013) di Turki. Penelitian ini bertujuan untuk mengembangkan alat ukur sikap tanggung jawab siswa dalam belajar berdasarkan persepsi masing-masing siswa. Penelitian ini melibatkan 579 siswa Sekolah Dasar. Analisis terhadap validitas konstruk instrumen yang dikembangkan meliputi EFA, CFA, perhitungan korelasi butir, dan daya beda butir. Adapun reliabilitas alat ukur di antaranya dilihat dari koefisien konsistensi internal dengan *Cronbach's Alpha* dan koefisien determinasinya. Melalui penelitian ini dihasilkan instrumen pengukur sikap

tanggung jawab dalam pembelajaran yang terbukti memiliki validitas isi, validitas konstruk, dan reliabilitas yang baik.

Berbagai penelitian terkait studi keselarasan antara penilaian dengan standar dalam kurikulum tertentu telah dilakukan di berbagai negara. Kulm, Wilson, & Kitchen (2005) meneliti tentang keselarasan antara standar dan butir penilaian matematika di Amerika. Prosedur proyek 2061 dari *AAAS* diterapkan untuk menganalisis isi dan kualitas butir penilaian dengan standar yang digunakan. Masih di Amerika, Webb (2006) menerapkan model Webb untuk mengevaluasi keselarasan antara standar dan penilaian matematika untuk kelas 3 - 8 dan kelas 10 di Wisconsin. Delapan pakar berpartisipasi dalam menganalisis keselarasan. Para pakar tersebut terdiri dari ahli materi matematika, pengawas bidang matematika, guru matematika, dan mahasiswa program doktor pendidikan matematika. Hasil penelitian menunjukkan bahwa keselarasan antara standar dan penilaian terdapat pada empat dari tujuh butir penilaian.

Penelitian tentang studi keselarasan juga dilakukan Gunilla & Henriksson (2008) di Swedia dengan meneliti kelebihan dan kekurangan berbagai model atau metode uji keselarasan yang berbeda dengan mengacu pada kriteria teoretis yang ditetapkan. Hasil penelitian menunjukkan bahwa revisi taksonomi Bloom dan taksonomi Porter adalah model yang paling tepat. Menggunakan data empiris, selanjutnya diperoleh bahwa revisi taksonomi Bloom lebih inklusif dan eksklusif dibandingkan dengan taksonomi Porter. Indeks keandalan antar rater secara signifikan juga lebih baik pada revisi taksonomi Bloom daripada taksonomi Porter.

Kesimpulan dari penelitian ini adalah bahwa revisi taksonomi Bloom merupakan model terbaik untuk membuktikan adanya keselarasan. Selain itu, hasil penelitian juga menunjukkan bahwa revisi taksonomi Bloom adalah model atau metode yang paling tepat digunakan dalam mata pelajaran Kimia dan Matematika.

Secara lebih spesifik, Nasstrom (2009) juga melakukan penelitian terkait kegunaan revisi taksonomi Bloom untuk mengevaluasi keselarasan antara standar dan penilaian di Swedia dengan fokus penelitian untuk mengetahui perbedaan konsistensi panelis yang berasal dari guru dan ahli penilaian dalam uji keselarasan. Hasil penelitian menunjukkan bahwa dalam penggunaan revisi taksonomi Bloom untuk studi keselarasan, terdapat perbedaan antara panelis yang terdiri dari guru dan panelis yang terdiri dari para ahli penilaian. Para panelis yang terdiri dari ahli penilaian ternyata lebih konsisten dalam memberikan interpretasinya atas standar.

Ndlovu & Mji (2012) juga melakukan studi keselarasan antara penilaian matematika dengan kerangka penilaian dari TIMSS pada kinerja para siswa di Afrika Selatan dalam tes *benchmark* internasional. Tujuan dari penelitian ini adalah untuk menyelidiki tingkat keselarasan antara kerangka penilaian matematika dari TIMSS tahun 2003 pada kelas 8 dengan hasil penilaian matematika berdasarkan standar pada kurikulum yang diperbaiki untuk mata pelajaran matematika kelas 8 di Afrika Selatan. Metode penelitian ini adalah analisis dokumen. Berdasarkan hasil penelitian diperoleh indeks keselarasan Porter sebesar 0,751 untuk keselarasan antara penilaian berdasarkan standar dan kerangka penilaian dari TIMSS.

Anderson et al. (2015: 27) melakukan studi keselarasan dengan mengukur keselarasan butir-butir tes matematika melalui sistem online dengan mengendalikan efek pakar (*rater*). Penilaian para pakar dilakukan dengan format yang lebih sederhana memuat telaah kuantitatif dan kualitatif yang disajikan secara online. Para pakar memberikan penilaian terkait keselarasan butir dengan standar dalam empat tingkatan (0: *no alignment*, 1: *vague alignment*, 2: *somewhat aligned*, 3: *directly aligned*). Pakar memberikan penilaian tidak ada keselarasan apabila menilai 0-1 dan ada keselarasan apabila menilai 2-3. Para pakar juga memberikan masukan secara kualitatif apabila penilaian yang diberikan dalam tingkatan 0-1. Hasil studi ini juga menunjukkan bahwa penggunaan metode online untuk mengukur keselarasan butir tidak memerlukan pertemuan langsung para pakar, mengurangi biaya, meningkatkan fleksibilitas waktu, dan menawarkan berbagai efisiensi terutama dalam menilai keselarasan suatu bank soal yang besar.

C. Kerangka Pikir

Penerapan Kurikulum 2013 merupakan salah satu upaya pemerintah guna meningkatkan kualitas pendidikan agar sesuai dengan perkembangan zaman. Kurikulum 2013 diterapkan untuk mencetak generasi yang siap menghadapi masa depan sehingga menitikberatkan pada proses pembelajaran yang mendorong siswa memiliki kompetensi sikap, pengetahuan, dan keterampilan yang jauh lebih baik. Oleh sebab itu, mengacu pada penerapan Kurikulum 2013, hakikat penilaian dalam pembelajaran matematika dan tujuan pendidikan di Indonesia, maka penilaian

pencapaian kompetensi matematika siswa semestinya dilakukan pada aspek sikap, pengetahuan, dan keterampilan. Penilaian aspek sikap berdasarkan Kurikulum 2013 mencakup perkembangan afeksi yang tidak sekedar dibatasi pada sikap untuk berbuat namun lebih dari itu terkait dengan karakter siswa.

Salah satu materi yang mengalami perubahan dalam pelaksanaan Kurikulum 2013 adalah statistika. Materi ini berdasarkan Kurikulum 2006 atau Kurikulum Tingkat Satuan Pendidikan (KTSP) hanya diajarkan pada kelas IX, namun menggunakan Kurikulum 2013 mulai diajarkan di kelas VII dan VIII. Statistika merupakan materi yang memiliki kekhasan karena terkait langsung dengan data dalam kehidupan sehari-hari. Pengintegrasian pendidikan karakter yang harus dimiliki siswa seperti jujur dan tanggung jawab melalui penilaian pada materi ini sangat relevan dilakukan selain juga memperhatikan penilaian aspek pengetahuan dan keterampilan serta sikap matematis siswa.

Hingga saat ini, pemerintah berkomitmen untuk tetap melaksanakan Kurikulum 2013 secara bertahap dengan terus melakukan perbaikan seiring perkembangan yang terjadi. Berdasarkan hasil refleksi yang dilakukan pemerintah, wawancara peneliti, serta hasil penelitian menunjukkan bahwa para guru matematika termasuk di DIY pada umumnya masih kesulitan dalam melakukan penilaian berdasarkan Kurikulum 2013. Kesulitan yang terjadi tidak hanya terkait cara penilaiannya melainkan juga dalam penyusunan instrumen hingga penentuan kualitasnya khususnya untuk aspek sikap dan keterampilan. Para guru matematika

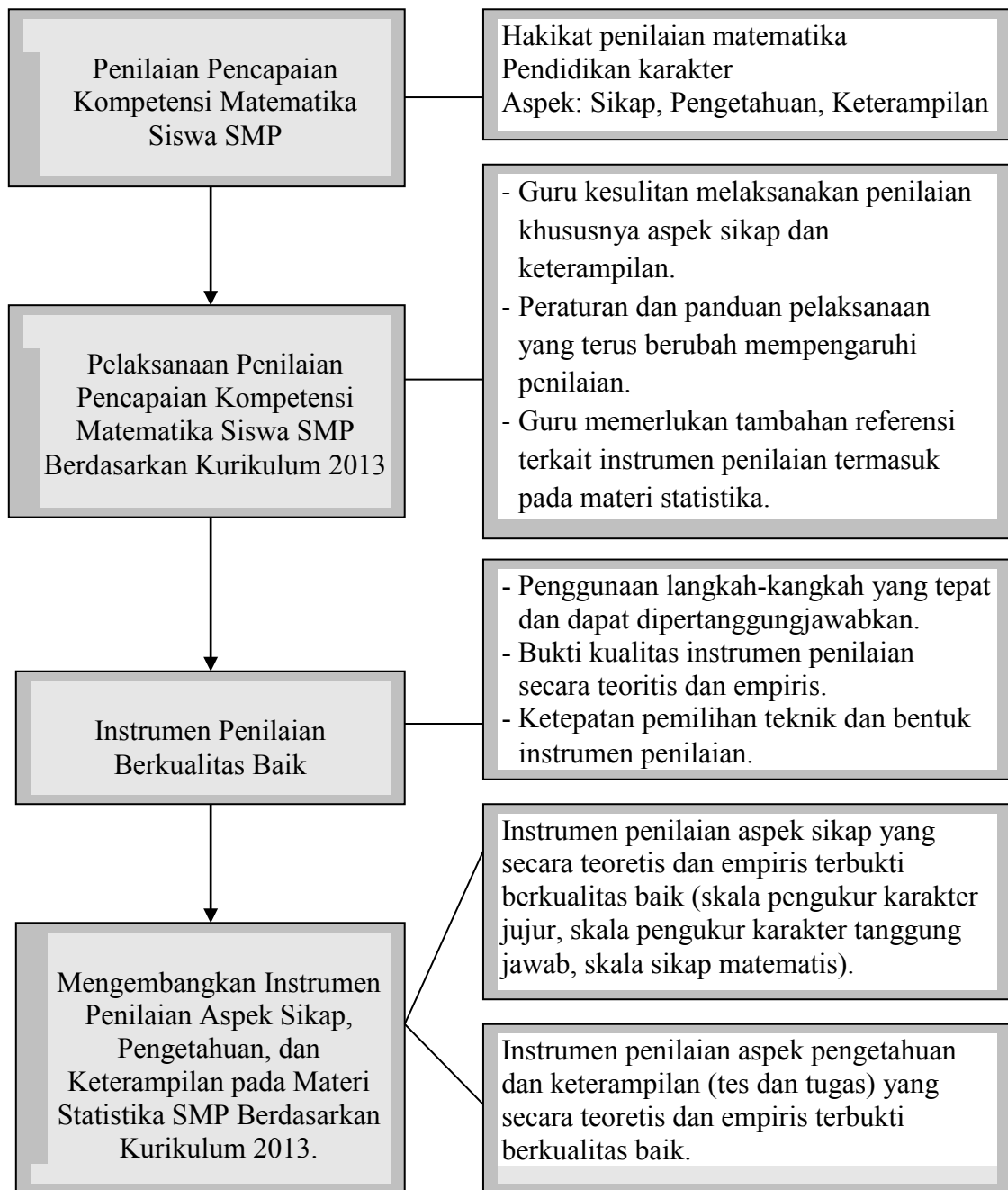
juga masih memerlukan referensi terkait contoh instrumen penilaian untuk aspek sikap, pengetahuan, dan keterampilan yang berkualitas baik sehingga dapat digunakan para guru termasuk di antaranya pada materi statistika.

Penilaian yang berkualitas sangat dipengaruhi oleh penggunaan instrumen yang berkualitas pula. Instrumen yang berkualitas diperoleh melalui langkah-langkah penyusunan yang tepat dan dapat dipertanggungjawabkan. Adapun untuk menetapkan bahwa suatu instrumen penilaian dikatakan berkualitas baik adalah dengan memperhatikan validitas, reliabilitas, dan parameter butir yang memenuhi kriteria sebagai instrumen penilaian yang baik. Instrumen penilaian yang berkualitas semestinya juga memiliki keselarasan dengan standar dalam kurikulum. Apabila penilaian yang dilakukan tidak selaras dengan standar dalam kurikulum yang digunakan, maka tujuan pendidikan sebagaimana tertuang dalam kurikulum tidak dapat tercapai secara optimal. Selain itu, hasil uji keselarasan bisa menjadi sumber bukti validitas isi karena mengungkap sejauh mana instrumen penilaian mencerminkan standar yang telah ditetapkan dalam kurikulum. Berbagai penelitian pengembangan instrumen penilaian pencapaian kompetensi siswa pada mata pelajaran matematika yang didukung studi keselarasan antara penilaian dan standar yang ditetapkan dalam kurikulum belum banyak dilakukan di Indonesia. Padahal, selama lebih dari satu dekade, berbagai metode atau model untuk membuktikan adanya keselarasan antara penilaian dan standar dalam kurikulum terus berkembang.

Selain memperhatikan langkah-langkah penyusunan instrumen yang tepat dan terpenuhinya berbagai kriteria kualitas instrumen, penilaian yang berkualitas tentunya juga ditentukan oleh ketepatan dalam memilih teknik dan bentuk instrumen yang digunakan. Berbagai teknik penilaian dan bentuk instrumen dipilih berdasarkan karakteristik materi dan tujuan pelaksanaan penilaian. Penilaian aspek sikap yang terkait dengan karakter siswa pada materi statistika dapat dilakukan dengan teknik penilaian diri dengan bentuk instrumen berupa suatu skala psikologi. Adapun instrumen penilaian pengetahuan dan keterampilan dalam materi statistika dapat dilakukan melalui tes bentuk uraian dan tugas proyek individu.

Berdasarkan uraian di atas, melalui penelitian ini dikembangkan instrumen penilaian untuk aspek sikap, pengetahuan, dan keterampilan pada mata pelajaran matematika materi statistika SMP berdasarkan Kurikulum 2013. Penilaian aspek sikap dilakukan melalui penilaian diri menggunakan skala untuk mengukur karakter jujur siswa, skala untuk mengukur karakter tanggung jawab siswa, dan skala sikap matematis. Penilaian aspek pengetahuan dan keterampilan dilakukan dengan tes uraian dan tugas proyek individu. Bukti kualitas instrumen meliputi bukti teoretis dan empiris. Hasil penelitian diharapkan sangat bermanfaat bagi para guru, siswa, praktisi pendidikan, dan pengambil kebijakan pendidikan terkait kegiatan pengembangan instrumen penilaian. Produk hasil penelitian juga dapat menjadi tambahan referensi terkait contoh instrumen penilaian aspek sikap, pengetahuan, dan keterampilan pada mata pelajaran matematika materi statistika SMP berdasarkan Kurikulum 2013.

Secara garis besar, kerangka pikir penelitian yang menggambarkan alur proses berpikir penelitian ini disajikan pada Gambar 4 sebagai berikut.



Gambar 4. Kerangka Pikir

D. Pertanyaan Penelitian

Berdasarkan rumusan masalah dan kajian teori di atas, dalam penelitian ini dirinci pertanyaan-pertanyaan penelitian sebagai berikut.

1. Bagaimanakah instrumen penilaian yang dapat mengukur dengan tepat pencapaian kompetensi matematika siswa untuk aspek sikap pada materi statistika SMP berdasarkan Kurikulum 2013?
2. Bagaimanakah instrumen penilaian yang dapat mengukur dengan tepat pencapaian kompetensi matematika siswa untuk aspek pengetahuan dan keterampilan pada materi statistika SMP berdasarkan Kurikulum 2013?
3. Bagaimana kualitas instrumen penilaian aspek sikap yang dihasilkan, ditinjau dari validitas isi, validitas konstruk, dan reliabilitasnya?
4. Bagaimana kualitas instrumen penilaian aspek pengetahuan dan keterampilan yang dihasilkan ditinjau dari validitas isi, validitas konstruk, reliabilitas, parameter butir, fungsi informasi, dan kesalahan baku pengukurannya?
5. Bagaimanakah pencapaian kompetensi aspek sikap siswa berdasarkan penggunaan instrumen penilaian aspek sikap yang dihasilkan?
6. Bagaimanakah pencapaian kompetensi aspek pengetahuan dan keterampilan siswa berdasarkan penggunaan instrumen penilaian aspek pengetahuan dan keterampilan yang dihasilkan?

BAB III

METODE PENELITIAN

A. Model Pengembangan

Penelitian ini merupakan penelitian pengembangan yang bertujuan untuk menghasilkan instrumen penilaian pencapaian kompetensi matematika siswa pada aspek sikap, pengetahuan, dan keterampilan yang terbukti secara teoretis dan empiris berkualitas baik. Instrumen penilaian yang dikembangkan difokuskan pada materi statistika SMP berdasarkan Kurikulum 2013. Model pengembangan yang digunakan untuk menghasilkan instrumen penilaian pencapaian kompetensi siswa dalam penelitian ini merupakan modifikasi model pengembangan instrumen dari Wilson (2005: 18-19) dan pengembangan instrumen tes dari Oriondo & Antonio (1998: 34).

Menurut Wilson (2005: 18-19), langkah-langkah pengembangan instrumen penilaian meliputi: (a) *construct map*, (b) *items*, (c) *item scores*, dan (d) *measures*. Adapun langkah-langkah pengembangan instrumen tes dari Oriondo & Antonio (1998: 34) adalah: (a) *planning the test*, (b) *trying out the test*, (c) *establishing test validity*, (d) *establishing test reliability*, dan (e) *interpreting the test score*. Berdasarkan kedua model tersebut, secara garis besar langkah-langkah pengembangan instrumen penilaian dalam penelitian ini dibagi dalam dua tahap. Tahap I merupakan tahap pendahuluan dan tahap II merupakan kegiatan pengembangan instrumen penilaian pencapaian kompetensi meliputi: (1) penyusunan instrumen awal, (2) uji coba instrumen, dan (3) interpretasi hasil uji coba. Hal ini

berdasarkan pertimbangan bahwa pengembangan instrumen penilaian dalam penelitian ini benar-benar didasari oleh adanya kebutuhan para guru matematika SMP yang telah menerapkan Kurikulum 2013 terkait instrumen penilaian yang digunakan. Selain itu, juga mengingat bahwa instrumen penilaian untuk mengukur pencapaian kompetensi matematika siswa pada aspek sikap, pengetahuan, dan keterampilan dapat berupa tes dan non tes.

Tahap I merupakan tahap pendahuluan yang meliputi telaah Kurikulum 2013, identifikasi masalah riil dan kajian kebutuhan, dan kajian berbagai referensi terkait instrumen penilaian dalam pembelajaran matematika. Kegiatan telaah Kurikulum 2013 dilakukan dengan mengkaji berbagai kebijakan pemerintah terkait penerapan Kurikulum 2013 untuk jenjang SMP, mencermati Standar Kompetensi Lulusan (SKL) dan Kompetensi Inti (KI) untuk jenjang SMP serta Kompetensi Dasar (KD) untuk mata pelajaran matematika SMP materi statistika berdasarkan Kurikulum 2013.

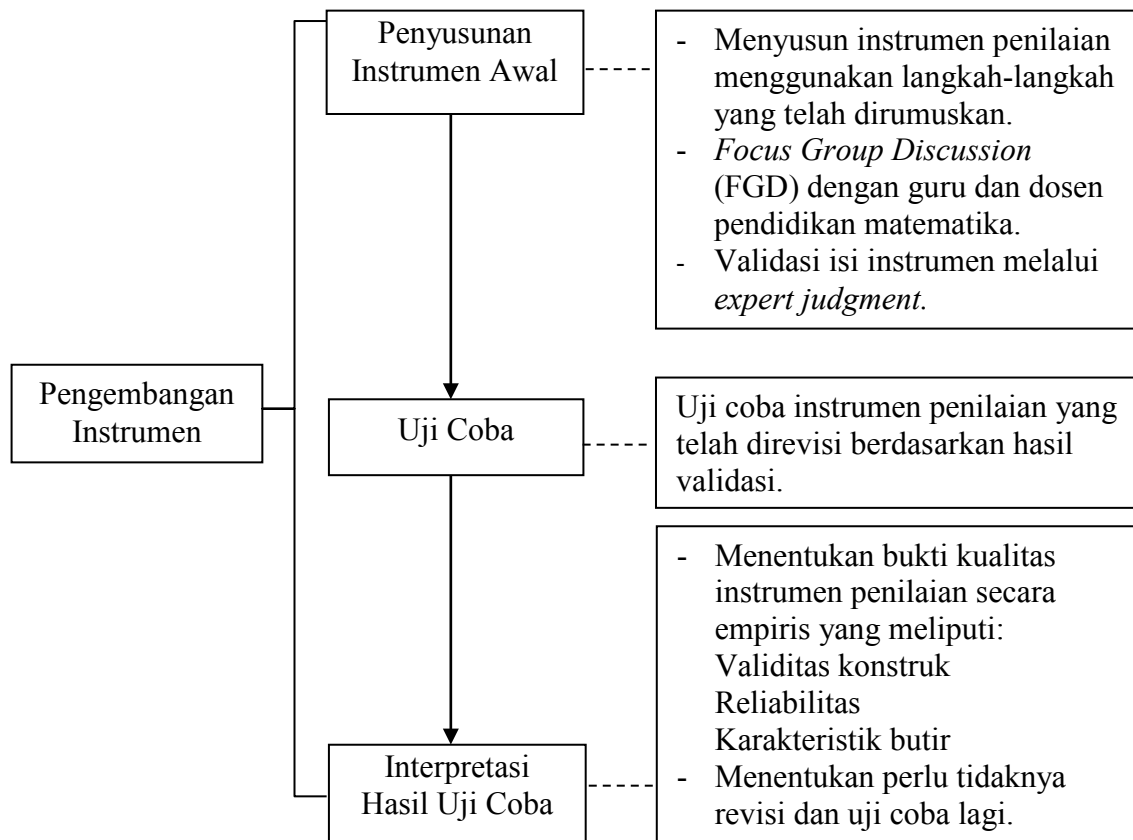
Kegiatan identifikasi kebutuhan riil dan kajian kebutuhan dilakukan untuk mendapatkan informasi realita permasalahan di lapangan khususnya terkait dengan penilaian dan kebutuhan ketersediaan referensi instrumen penilaian pencapaian kompetensi siswa bagi para guru. Kegiatan ini dilakukan melalui penelusuran informasi dan wawancara terhadap para guru matematika SMP yang telah menerapkan Kurikulum 2013 di DIY. Adapun pengkajian pustaka dilakukan dengan melakukan penelusuran berbagai referensi yang terkait dengan penilaian pencapaian kompetensi siswa dalam pembelajaran matematika, langkah-langkah penyusunan instrumen penilaian pencapaian kompetensi sikap, pengetahuan, dan keterampilan

dalam pembelajaran matematika, kualitas instrumen penilaian termasuk uji keselarasan antara penilaian dengan standar dalam kurikulum. Penelusuran kajian pustaka juga dilakukan dengan mengkaji berbagai hasil penelitian yang relevan.

Berdasarkan hasil ketiga kegiatan pada tahap pendahuluan selanjutnya dilakukan elaborasi dengan mengolah ketiga unsur tersebut. Hasil elaborasi menjadi gambaran untuk mengembangkan instrumen penilaian pencapaian kompetensi siswa pada aspek sikap, pengetahuan, dan keterampilan. Berdasarkan hasil kegiatan tahap I dirumuskan langkah-langkah penyusunan instrumen penilaian untuk aspek sikap, pengetahuan, dan keterampilan yang secara teoritis dianggap mampu menghasilkan instrumen penilaian pencapaian kompetensi siswa yang berkualitas baik pada mata pelajaran matematika.

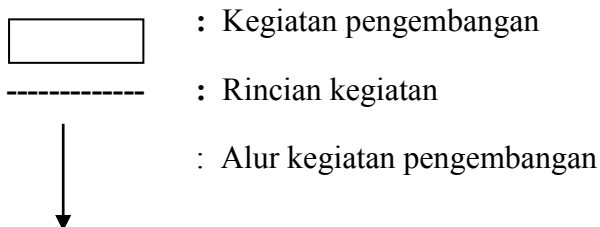
Tahap II merupakan tahap pengembangan instrumen penilaian pencapaian kompetensi siswa pada mata pelajaran matematika SMP materi statistika berdasarkan Kurikulum 2013 menggunakan langkah-langkah penyusunan instrumen yang dirumuskan pada tahap I. tahap II merupakan kegiatan pengembangan instrumen penilaian pencapaian kompetensi meliputi: (1) penyusunan instrumen awal, (2) uji coba instrumen, dan (3) interpretasi hasil uji coba. Berdasarkan kegiatan tahap II ini dihasilkan instrumen penilaian pencapaian kompetensi matematika siswa SMP materi statistika untuk aspek sikap, pengetahuan, dan keterampilan yang secara teoretis dan empiris terbukti berkualitas baik.

Secara visual, langkah-langkah pengembangan instrumen pada penelitian tahap II ini disajikan pada Gambar 5 sebagai berikut.



Gambar 5. Model Pengembangan Instrumen

Keterangan:



Berdasarkan Gambar 5, penyusunan instrumen awal dimaksudkan sebagai kegiatan menyusun instrumen penilaian berdasarkan langkah-langkah penyusunan instrumen yang telah dirumuskan pada tahap I. Instrumen awal yang dihasilkan

selanjutnya didiskusikan dalam kegiatan *Focus Group Discussion* (FGD) guna memperoleh masukan agar diperoleh instrumen awal yang lebih baik. Langkah berikutnya adalah instrumen yang dihasilkan pada tahap penyusunan instrumen awal ini divalidasi melalui *expert judgment*. Berdasarkan hasil akhir proses validasi yakni diperolehnya penilaian layak terhadap instrumen penilaian yang dikembangkan, selanjutnya dilakukan uji coba dan ditindaklanjuti dengan interpretasi hasil uji coba yakni menentukan bukti kualitas instrumen berdasarkan data empiris dan menentukan perlu tidaknya revisi dan uji coba lagi.

B. Prosedur Pengembangan

Pengembangan instrumen penilaian dalam penelitian ini merupakan kegiatan penelitian tahap II dari keseluruhan kegiatan penelitian. Mengacu pada langkah pengembangan instrumen penilaian sebagaimana disajikan pada Gambar 5, kegiatan pengembangan diawali dengan penyusunan instrumen awal yang dalam penelitian ini dilakukan dalam tiga kegiatan yaitu:

- a) Menyusun instrumen penilaian pencapaian kompetensi matematika siswa pada materi statistika SMP berdasarkan Kurikulum 2013 untuk aspek sikap, pengetahuan, dan keterampilan menggunakan langkah-langkah penyusunan instrumen yang telah dirumuskan berdasarkan hasil kegiatan penelitian tahap I.

- b) Melakukan *Focus Group Discussion* (FGD) dengan guru dan dosen pendidikan matematika untuk mendiskusikan instrumen penilaian yang telah disusun.
- c) Melakukan validasi instrumen penilaian yang dikembangkan kepada para pakar yang terdiri dari dua pakar pendidikan matematika, dua pakar statistika, dan dua pakar penilaian hasil belajar matematika.

Instrumen penilaian yang dikembangkan dalam penelitian ini difokuskan pada salah satu materi mata pelajaran matematika SMP berdasarkan Kurikulum 2013 yakni statistika. Sebagaimana telah dikemukakan sebelumnya, materi ini dipilih dengan pertimbangan bahwa materi ini pada kurikulum sebelumnya yakni Kurikulum 2006 atau Kurikulum Tingkat Satuan Pendidikan (KTSP) hanya dipelajari di kelas IX, namun berdasarkan Kurikulum 2013 mulai dipelajari pada kelas VII dan VIII. Pertimbangan lainnya adalah materi ini merupakan salah satu materi yang diujikan dalam tes berstandar internasional seperti TIMSS dan PISA.

Materi statistika sangat berkaitan erat dengan data yang diperoleh baik dari hasil pengamatan maupun pengukuran. Materi ini dapat diaplikasikan secara luas dalam berbagai bidang kehidupan bermasyarakat (Van de Walle, 2008: 194). Berkaitan dengan data yang diperoleh, ketepatan data dan kesesuaian antara data sesungguhnya dengan data yang disampaikan sangatlah penting. Hal ini menunjukkan bahwa dalam materi statistika selain bersifat matematis juga memuat karakter jujur dan tanggung jawab baik dalam berpikir, bersikap, maupun bertindak.

Memperhatikan karakteristik materi statistika tersebut, instrumen penilaian aspek sikap yang dikembangkan difokuskan pada sikap matematis, karakter jujur, dan

karakter tanggung jawab. Teknik penilaian aspek sikap berupa penilaian diri dipilih untuk dikembangkan dalam penelitian ini. Hal ini berdasarkan pertimbangan bahwa mengacu pada revisi Kurikulum 2013, penilaian untuk aspek sikap yang dilakukan oleh guru mata pelajaran seperti matematika menilai aspek sikap sewajarnya dan hanya menilai aspek akademik sesuai bidang yang diajarkan saja. Guru mata pelajaran matematika dalam melakukan penilaian aspek sikap lebih bersifat untuk menambah referensi atau masukan terkait penilaian aspek sikap siswa. Oleh sebab itu, penilaian diri dipilih karena merupakan teknik penilaian yang bersifat fleksibel untuk digunakan guru matematika dalam melakukan penilaian aspek sikap.

Instrumen penilaian aspek pengetahuan dan keterampilan dalam penelitian ini dirancang secara bersamaan karena keduanya saling terkait. Instrumen penilaian aspek pengetahuan yang dikembangkan berupa tes uraian dan instrumen penilaian aspek keterampilan berupa tugas proyek individu. Tes uraian dipilih dengan pertimbangan mampu mengungkap kemampuan siswa secara lebih detail. Tes uraian yang disusun mengacu pada materi statistika kelas VII. Adapun tugas proyek individu dipilih karena lebih mudah diterapkan dan mampu mengungkap kemampuan siswa secara individu dalam menunjukkan keterampilan melalui penyelesaian tugas yang dikerjakannya.

Kegiatan yang dilakukan setelah seluruh instrumen penilaian selesai disusun adalah melakukan *Focus Group Discussion* (FGD) dengan guru dan dosen pendidikan matematika. Forum FGD ini dipilih untuk mendiskusikan instrumen yang telah disusun tersebut, mengingat bahwa menurut Krueger & Casey (2000: 5), FGD

adalah suatu metode diskusi secara mendalam yang melibatkan kelompok kecil yang homogen (6-12 orang) untuk mendiskusikan topik tertentu. Kegiatan FGD yang dilakukan dalam penelitian ini melibatkan 4 orang guru matematika SMP yang menerapkan Kurikulum 2013 di DIY dan 3 orang dosen pendidikan Matematika. Daftar guru dan dosen pendidikan matematika yang terlibat dalam kegiatan FGD secara lengkap dapat dilihat pada Lampiran 1b. Hasil pelaksanaan FGD selanjutnya ditindaklanjuti diskusi dengan pembimbing.

Berdasarkan analisis hasil FGD dan diskusi dengan pembimbing, diperoleh rancangan produk instrumen penilaian yang dikembangkan. Selain dirancang produk berupa instrumen penilaian, pada tahap ini juga dirancang perangkat lain dalam penelitian yaitu lembar telaah untuk uji validitas. Rancangan produk yang telah dikembangkan berdasarkan hasil FGD dan diskusi pembimbing selanjutnya dilakukan penyempurnaan produk. Produk inilah yang selanjutnya divalidasi dengan menggunakan *expert judgment*.

Validasi instrumen melalui *expert judgment* atau penilaian pakar dilakukan dengan menelaah instrumen penilaian yang telah disusun ditinjau dari kontennya. Telaah untuk instrumen penilaian aspek sikap meliputi: (a) kesesuaian butir dengan indikator, (b) bahasa yang digunakan, (c) pernyataan butir yang bersifat tidak bias, (d) kejelasan pilihan jawaban, dan (e) ketepatan pedoman penskoran. Adapun untuk instrumen penilaian aspek pengetahuan dan keterampilan meliputi: (a) kesesuaian butir dengan indikator, (b) materi/substansi, (c) konstruksi, dan (d) bahasa. Para pakar secara khusus melakukan penilaian terhadap kesesuaian antara butir dengan indikator

dalam bentuk skala *Likert* dengan lima pilihan jawaban. Penilaian ini dikenakan pada seluruh instrumen penilaian yang dikembangkan baik untuk aspek sikap, pengetahuan, dan keterampilan. Selain itu, untuk tujuan uji keselarasan guna memperkuat bukti validitas isi, lembar telaah juga dilengkapi format untuk penilaian pakar terkait muatan revisi taksonomi Bloom pada tiap butir tes. Pengujian keselarasan hanya dilakukan pada instrumen penilaian aspek pengetahuan dengan mengacu pada karakteristik kompetensi yang digunakan dalam Kurikulum 2013 yakni revisi taksonomi Bloom.

Para pakar yang terlibat dalam penelitian ini terdiri dari dua pakar pendidikan matematika, dua pakar statistika, dan dua pakar penilaian hasil belajar matematika sebagai berikut.

Tabel 2. Validator Instrumen Penilaian

Keahlian	Jumlah	Nama	Instansi
Pendidikan Matematika	2	Dr. Ariyadi Wijaya	UNY
		Dr. Ali Mahmudi	UNY
Statistika	2	Endang Listyani, MS	UNY
		Mathilda Susanti, M.Si.	UNY
Penilaian Hasil Belajar Matematika	2	Dr. Jailani	UNY
		Dr. Heri Retnawati	UNY

Teknik yang dilakukan untuk kegiatan validasi mengacu pada prinsip teknik *delphi* yang dipakai Dunn (1994: 366) yaitu semua pakar memberikan tanggapan dan penilaian secara terpisah, dan penilaian setiap pakar dihimpun dan dikomunikasikan kembali ke semua pakar. Selain itu, terkait dengan pembuktian validitas isi melalui

uji keselarasan, Porter (2004: 19) menyatakan bahwa dalam studi keselarasan, penilaian dari beberapa pakar dapat dilakukan secara independen.

Adapun langkah-langkah teknik *delphi* yang digunakan pada penelitian ini adalah sebagai berikut:

- 1) Produk berupa instrumen penilaian untuk aspek sikap, pengetahuan, dan keterampilan serta lembar telaah yang telah disusun dan dikembangkan disampaikan kepada para pakar untuk divalidasi. Terkait dengan studi keselarasan, peneliti juga menyampaikan dokumen yang berisi tentang SKL, KI, dan KD mata pelajaran matematika materi statistika serta format untuk menilai keselarasan dengan metode revisi taksonomi Bloom.
- 2) Peneliti merangkum hasil penilaian para pakar dan melakukan revisi berdasarkan masukan dan saran yang diberikan para pakar.
- 3) Peneliti bertemu para pakar untuk mengklarifikasi butir instrumen yang telah direvisi sehingga diperoleh produk akhir yang telah dinyatakan valid secara konten oleh para pakar.

Berdasarkan hasil penilaian para pakar, selain peneliti merangkum dan merevisi, peneliti juga menghitung indeks validitasnya dengan menggunakan rumus V Aiken. Kriteria penentuan valid tidaknya suatu butir adalah dengan membandingkan nilai V_{hitung} dengan nilai V_{tabel} . Nilai V_{tabel} adalah nilai minimum indeks validitas isi berdasarkan banyaknya pakar (validator) pada tabel V dalam Aiken (1985: 134). Karena banyaknya pakar dalam penelitian ini adalah enam dan

banyaknya rating yang dibuat dalam lembar telaah adalah lima maka minimal indeks validitas isi berdasarkan tabel V Aiken adalah 0,79.

Berdasarkan hasil validasi, diperoleh hasil telaah pakar yang bersifat kualitatif dan pernyataan para pakar bahwa instrumen yang dikembangkan telah layak atau siap digunakan. Selain itu juga diperoleh data kuantitatif terkait kesesuaian antara butir instrumen dengan indikator serta khusus untuk instrumen berupa tes juga diperoleh penilaian pakar terkait muatan revisi taksonomi Bloom. Berdasarkan hasil akhir proses validasi yakni diperolehnya penilaian layak terhadap instrumen penilaian yang dikembangkan, selanjutnya dilakukan uji coba. Kegiatan uji coba dilakukan guna memperoleh bukti empiris kualitas instrumen penilaian yang dikembangkan. Hasil uji coba ditindaklanjuti dengan interpretasi hasil uji coba yakni menentukan bukti kualitas instrumen berdasarkan data empiris dan menentukan perlu tidaknya revisi dan uji coba lagi.

C. Desain Uji Coba Produk

1. Desain Uji Coba

Tahap uji coba produk merupakan tahap uji terhadap produk berupa instrumen penilaian yang telah dihasilkan pada tahap penyusunan instrumen awal. Berdasarkan hasil uji coba, diperoleh bukti empiris tentang kualitas instrumen penilaian yang dikembangkan. Uji coba produk dilakukan dengan terlebih dahulu melakukan koordinasi dengan kepala sekolah dan para guru

matematika SMP yang terlibat guna kelancaran pelaksanaan uji coba. Berdasarkan hasil koordinasi selanjutnya dilakukan uji coba produk.

2. Subjek Coba

Kegiatan uji coba dilakukan pada siswa kelas VII pada tiga SMP di kota Yogyakarta yang menerapkan Kurikulum 2013. Pemilihan ketiga SMP ini sebagai tempat uji coba dilakukan dengan teknik *purposive cluster sampling* yakni pemilihan sampel uji coba berdasarkan pada kelompok sekolah berdasarkan pertimbangan atau tujuan tertentu. Mengingat bahwa produk yang dihasilkan dalam penelitian ini ditujukan untuk seluruh siswa baik yang berkemampuan rendah, sedang, maupun tinggi, maka uji coba dilakukan pada SMP yang mewakili kualitas tinggi, sedang, dan berkembang serta mewakili sekolah negeri dan swasta.

Penentuan kualitas sekolah dalam penelitian ini berdasarkan hasil Ujian Nasional (UN) tahun pelajaran 2015/2016. Kondisi enam SMP di kota Yogyakarta yang menerapkan Kurikulum 2013 mulai tahun pelajaran 2013/2014 cukup mewakili sekolah dengan kualitas tinggi, sedang, dan berkembang. Oleh sebab itu, dipilih tiga SMP yang masing-masing mewakili tiap kualitas tersebut yaitu SMP N 5 Yogyakarta mewakili kualitas tinggi, SMP IT Abu Bakar Yogyakarta mewakili kualitas sedang, dan SMP Muhammadiyah 2 Yogyakarta mewakili kualitas berkembang. Kegiatan penelitian dalam rangka uji coba produk dilaksanakan pada bulan Mei 2016 hingga bulan September 2016.

3. Teknik dan Instrumen Pengumpulan Data

Data hasil uji coba dalam penelitian ini berupa data kuantitatif. Data diperoleh berdasarkan isian siswa terhadap instrumen penilaian aspek sikap berupa lembar penilaian diri yang terdiri atas skala untuk mengukur karakter jujur, skala untuk mengukur karakter tanggung jawab, dan skala sikap matematis serta jawaban siswa pada tes uraian dan pengerjaan tugas proyek individu. Teknik pengumpulan data dilakukan dengan menyampaikan instrumen penilaian yang dikembangkan kepada siswa untuk diisi dan dikerjakan. Pelaksanaan pengumpulan data dilakukan setelah terlebih dahulu melakukan koordinasi dengan pihak sekolah yakni kepala sekolah dan guru matematika. Berdasarkan hasil koordinasi, peneliti mengumpulkan data didampingi dan dibantu para guru matematika yang ditugaskan oleh kepala sekolah.

4. Teknik Analisis Data

Berdasarkan data hasil uji coba, ditentukan bukti empiris kualitas instrumen penilaian yang dikembangkan. Bukti empiris kualitas instrumen penilaian aspek sikap meliputi validitas konstruk dan reliabilitasnya. Adapun untuk instrumen penilaian aspek pengetahuan dan keterampilan meliputi validitas konstruk, reliabilitas konstruk, reliabilitas berdasarkan konsistensi internal, dan karakteristik butirnya. Kegiatan analisis data dalam penelitian ini diawali dengan uji persyaratan analisis sebelum melakukan analisis lainnya yakni menentukan kualitas instrumen penilaian secara empiris dan hasil pencapaian kompetensi yang diperoleh siswa.

a. Uji Persyaratan Analisis

Uji persyaratan analisis bertujuan untuk mengetahui apakah data yang dikumpulkan memenuhi persyaratan untuk dianalisis dengan teknik yang direncanakan atau tidak. Menggunakan teknik *Confirmatory Factor Analysis* (CFA) dengan bantuan program *Lisrel*, data yang dianalisis harus memenuhi asumsi bahwa data yang digunakan untuk menguji model mengikuti distribusi normal multivariat bersama (*joint multivariate normal distribution*).

Normalitas data dalam analisis diperlukan agar estimasi parameter yang dihasilkan tidak bias sehingga kesimpulan yang diambil tepat. Jika data yang dianalisis tidak memenuhi asumsi distribusi normal multivariat bersama maka perlu dilakukan pendekatan yang mampu menangani data non normal tersebut. Pendekatan yang dapat digunakan menurut Widhiarso (2012: 1-2) adalah: (a) menggunakan estimator yang tahan terhadap ketidaknormalan data misalnya GLS dan WLS, (b) menggunakan penskalaan yang bersifat tahan terhadap ketidaknormalan data, (c) menggunakan teknik *bootstrapping* untuk menghitung nilai kritis *chi square*, nilai parameter dan kesalahan standar, dan memadukan indikator dalam sebuah paketan butir.

Hasil uji normalitas dapat diperoleh melalui analisis program *Lisrel* dan dilakukan sebelum analisis faktor konfirmatori. Kriteria yang digunakan untuk mengetahui normalitas data adalah jika $p\text{-value} > 0,05$ maka sebaran

normal, dan sebaliknya jika $p\text{-value} \leq 0,05$ maka sebaran tidak normal (Yamin & Kurniawan, 2009: 29).

Adapun untuk analisis butir instrumen tes dan tugas dilakukan berdasarkan IRT politomus. Asumsi yang harus dipenuhi di antaranya adalah ukuran sampel dan unidimensionalitas data (Thorpe & Favia, 2012: 10). Ukuran sampel dalam IRT politomus menurut Reeve & Fayers (2005: 71) adalah sekurang-kurangnya 250, namun ukuran sampel sekitar 500 lebih disarankan guna akurasi estimasi parameter.

b. Analisis Data

1) Teknik Analisis untuk Menentukan Kualitas Instrumen Penilaian

Teknik analisis yang digunakan untuk menentukan kualitas instrumen penilaian berdasarkan data hasil uji coba adalah sebagai berikut.

a) Aspek sikap

Analisis data pada instrumen penilaian aspek sikap yang terdiri atas skala untuk mengukur karakter jujur, skala untuk mengukur karakter tanggung jawab, dan skala sikap matematis dilakukan guna memperoleh bukti empiris kualitas instrumen penilaian ditinjau dari validitas konstruk dan reliabilitasnya. Pembuktian validitas konstruk dilakukan menggunakan *second order* CFA dengan bantuan program *Lisrel* versi 8.51.

Kriteria kevalidan suatu butir indikator dalam mewakili konstruk adalah apabila nilai $t\text{-value} > 1,96$ dan nilai muatan faktor terstandar (*Standardized Loading Factor/SLF*) minimal 0,3 (Igarria et al, 1997: 290; Hair et al, 2010: 119). Mengacu pada kriteria tersebut, pada hasil analisis awal dengan *Lisrel*, apabila ada butir indikator yang tidak valid artinya ada salah satu kriteria yang tidak terpenuhi maka butir tersebut harus dieliminasi dan dilakukan analisis ulang. Namun, apabila hanya butir tersebut yang mewakili suatu indikator maka butir tersebut tidak dieliminasi melainkan diperbaiki atau direvisi untuk selanjutnya dilakukan uji coba lagi. Apabila butir yang dieliminasi tersebut tidak menghilangkan indikator atau dengan kata lain indikator diwakili lebih dari satu butir sehingga masih ada butir lain yang mewakili, maka butir tersebut dapat dieliminasi dan dilanjutkan analisis berikutnya.

Berdasarkan hasil analisis ulang setelah butir indikator yang tidak valid dieliminasi, yang perlu diperhatikan adalah kecocokannya dengan model. Jika ternyata model belum fit, maka dilakukan respesifikasi model dengan melakukan modifikasi indeks sesuai yang disarankan *Lisrel*. Menurut Wijanto (2008: 150), umumnya ada dua saran yang diusulkan dalam modifikasi indeks yaitu: (1) penambahan lintasan di antara variabel laten dengan variabel teramati, dan (2) penambahan kovariansi di antara dua kesalahan atau *error*. Penambahan lintasan bisa dilakukan apabila mempunyai dukungan teori yang cukup

kuat, namun jika tidak sebaiknya tidak dilakukan. Adapun pilihan kedua yakni penambahan kovariansi di antara dua kesalahan sebaiknya dilakukan untuk model pengukuran dari variabel laten yang sama.

Kecocokan model dalam penelitian ini mengacu pada pendapat Garson (2009) yang mengemukakan bahwa dukungan terhadap kecocokan model yang dikembangkan melalui data empiris paling tidak dilihat dari tiga ukuran kecocokan yang mewakili tiga kategori uji kecocokan model yang berbeda. Ketiga kategori uji kecocokan model tersebut adalah uji kecocokan absolut, uji kecocokan inkremental, dan uji kecocokan parsimoni. Jika dua dari tiga kategori tersebut memenuhi kriteria maka model yang dikembangkan cocok dengan data. Hal ini juga sebagaimana dikemukakan Hadi (2007: 117) bahwa jika dua dari tiga kategori tersebut signifikan maka model yang dikembangkan cocok dengan data.

Menurut Suranto, Muhyadi, & Mardapi (2014: 102) mengemukakan bahwa model yang dikembangkan dinyatakan cocok dengan data lapangan apabila sudah terpenuhi dua kriteria dari tiga kriteria yang menjadi ukuran kecocokan absolut yakni *Root Mean Square Error of Approximation* (RMSEA) $\leq 0,08$, *Chi square* yang diperoleh dari pengujian memiliki *p-value* $> 0,05$ dan *Goodness of Fit Index* (GFI) $\geq 0,90$.

Terkait kecocokan model, Hooper, Coughlan, & Mullen (2008: 57) mengemukakan bahwa penggunaan *chi square* bersifat sensitif terhadap ukuran sampel yang digunakan. Artinya, *chi square* cenderung menolak model ketika ukuran sampel yang digunakan adalah besar. Hal ini senada dengan Hair et al. (2009: 645) yang mengemukakan bahwa banyak faktor yang mempengaruhi uji signifikansi *chi square*. Nilai *Chi square* yang tidak signifikan tetap memungkinkan peneliti untuk mengatakan bahwa modelnya termasuk cukup baik, terlebih ketika ukuran sampelnya besar. Namun demikian, berapapun hasil *chi square*, nilai *chi square* dan *p-value* dari model tetap penting dan harus dilaporkan serta melengkapinya dengan indeks kecocokan model yang lainnya (Hair et al., 2009: 645). Wijanto (2008: 155) juga mengemukakan bahwa konsensus para peneliti menyatakan bahwa nilai *chi square* bukan satu-satunya ukuran *Goodness of Fit* (GOF) dan tidak ada satu ukuran GOF yang secara eksklusif mewakili kecocokan keseluruhan model.

Mengacu pada beberapa pendapat di atas, kriteria kecocokan model yang digunakan dalam penelitian adalah dilihat dari nilai RMSEA, *p-value*, dan GFI mewakili kecocokan absolut dan AGFI mewakili kecocokan inkremental. Adapun kriteria kecocokan model yang digunakan adalah RMSEA antara 0,03 sampai 0,08, *p-value*>0,05, $GFI \geq 0,90$, $AGFI \geq 0,90$ (Hair et al, 2009: 641-644). Nilai indeks

kecocokan yang diperoleh dalam penelitian ini dimaksudkan untuk menggambarkan benar atau tidaknya suatu konstruk didukung oleh variabel-variabel laten yang mempunyai indikator-indikator seperti yang telah ditentukan sebelumnya.

Estimasi reliabilitas instrumen penilaian aspek sikap diperoleh dari *output* hasil analisis *second order* CFA dan ditentukan dengan rumus koefisien reliabilitas komposit McDonald atau koefisien Omega (ω). Hal ini sebagaimana dikemukakan Widhiarso (2009: 41-45) bahwa reliabilitas yang diawali dengan analisis faktor konfirmatori dengan pendekatan SEM pada model pengukuran multidimensi di antaranya dapat dilakukan dengan koefisien reliabilitas konstruk McDonald atau koefisien Omega (ω). Model pengukuran dari penilaian aspek sikap dalam penelitian ini mengikuti model pengukuran multidimensi karena tiap konstruk yang diukur terdiri dari beberapa variabel laten. Berikut rumus untuk menghitung koefisien Omega (ω).

$$\omega = \frac{(\sum_{i=1}^I \lambda_i)^2}{(\sum_{i=1}^I \lambda_i)^2 + (\sum_{i=1}^I 1 - \lambda_i^2)} \quad (14)$$

Keterangan: λ_i : *factor loading* terstandar indikator ke-i.

Kriteria koefisien reliabilitas mengacu pada Hair et al. (2010: 688) yakni koefisien reliabilitas $>0,7$ tergolong baik, sementara koefisien reliabilitas $0,6-0,7$ tergolong dapat diterima asalkan indikator validitas konstruk model tergolong baik.

b) Aspek Pengetahuan dan Keterampilan

Analisis data yang dilakukan pada instrumen penilaian aspek pengetahuan berupa tes uraian dan penilaian aspek keterampilan berupa tugas proyek dilakukan guna memperoleh bukti empiris kualitas instrumen penilaian ditinjau dari validitas konstruk, reliabilitas konstruk (*construct reliability*), dan reliabilitas berdasarkan konsistensi internalnya. Selain itu juga dilakukan analisis untuk menentukan karakteristik butir instrumen.

Pembuktian validitas konstruk dilakukan dengan *first order* CFA menggunakan program *Lisrel* versi 8.51. Sebagaimana pada instrumen penilaian sikap, kriteria kevalidan suatu butir indikator dalam mewakili konstruk juga mengacu pada nilai *t-value* > 1,96 dan nilai muatan faktor terstandar (*Standardized Loading Factor/SLF*) minimal 0,3 (Igbaria et al, 1997: 290; Hair et al, 2010: 119). Adapun kriteria kecocokan model juga mengacu pada nilai RMSEA, *p value*, GFI, dan AGFI.

Estimasi reliabilitas konstruk instrumen tes dan tugas yang diperoleh dari *output* analisis *first order* CFA dalam penelitian ini ditentukan dengan rumus *Construct Reliability* (CR). Kriteria *Construct Reliability* (CR) juga mengacu pada Hair et al. (2009: 688) yakni koefisien reliabilitas > 0,7 tergolong baik, sementara koefisien reliabilitas antara 0,6 dan 0,7 tergolong dapat diterima asalkan indikator validitas konstruk model tergolong baik. Adapun reliabilitas instrumen

berdasarkan konsistensi internalnya ditentukan menggunakan koefisien Alpha dari Cronbach (*Cronbach's Alpha*). Perhitungan koefisien Alpha dilakukan dengan program SPSS versi 20. Kriteria koefisien reliabilitas yang baik adalah minimal 0,7 (Nunnally, 1981: 245). Adapun menurut Ebel & Frisbie (1991: 86) menyebutkan bahwa apabila tes digunakan sebagai tes standar maka koefisien reliabilitas sebaiknya antara 0,85-0,95 sedangkan bila digunakan untuk kelas minimal 0,65.

Hasil pembuktian validitas konstruk pada tes dan tugas sekaligus menunjukkan bukti terpenuhi atau tidaknya asumsi unidimensi yang disyaratkan dalam analisis butir berdasarkan IRT politomus. Selain itu, terpenuhinya asumsi unidimensi juga dilihat dari *plot eigenvalue* yang menunjukkan satu komponen dominan (Hambleton, Swaminathan, & Rogers, 1991: 56). *Plot eigenvalue* dalam penelitian ini diperoleh melalui analisis faktor dengan program SPSS versi 20.

Karakteristik butir tes dan tugas diperoleh melalui analisis butir berdasarkan IRT politomus GPCM dengan bantuan program *Parscale* versi 4.1. GPCM dipilih dengan memperhatikan muatan materi statistika sehingga butir-butir tes uraian yang dikembangkan diasumsikan memiliki tingkat kesukaran yang berbeda dan tingkat kesukaran tiap langkah tidak terurut. Terkait dengan ukuran sampel, untuk memperoleh estimasi parameter yang akurat, Reeve & Fayer (2005: 71) merekomendasikan ukuran sampel sekurang-kurangnya 500.

Berdasarkan hasil analisis *Parscale* dengan GPCM, pada sebuah butir diperoleh estimasi parameter daya beda (a), parameter lokasi (b -*global*), dan satu set $m-1$ parameter tingkat kesukaran (b). Parameter daya beda (a) menjelaskan seberapa besar butir dapat membedakan antara individu dengan kemampuan berbeda. Parameter lokasi (b -*global*) menjelaskan mengenai tingkat kesukaran butir yakni kesukaran relatif dari sebuah *step* dibandingkan dengan *step* lain dalam satu butir. Kalibrasi butir data politomus dengan *Parscale* dilakukan dengan membuat sintaks terlebih dahulu untuk menghasilkan statistik butir yang fit, taksiran parameter θ , a , b -*global*, dan d serta berbagai grafik. Parameter b_{jk} diperoleh dengan mengurangkan nilai parameter b -*global_j* dengan nilai parameter d_{jk} . Adapun grafik yang dihasilkan antara lain *Item Characteristic Curve* (ICC), *Item Information Curve* (IIC), dan *Test Information Curve* (TIC).

Hasil penaksiran parameter butir berupa a dan b -*global* dapat dibaca dari output *Parscale* fase 2. Berdasarkan output fase 2 juga dapat dibaca parameter d_{jk} atau parameter kategori. Adapun hasil penaksiran parameter kemampuan (θ) dapat dibaca pada output fase 3. Ketentuan mengenai butir yang memenuhi kriteria sebagai butir yang baik dalam penelitian ini adalah: (1) nilai parameter $a > 0,25$ pada skala logit, dan (2) nilai parameter b -*global* terletak pada selang -3 dan 3 pada skala logit (Wells, Hambleton, & Purwono, 2008).

2) Teknik Analisis Pencapaian Kompetensi Siswa

Pencapaian kompetensi siswa ditentukan berdasarkan skor yang diperoleh dari data penerapan instrumen hasil pengembangan yang terbukti secara empiris berkualitas baik. Penilaian pencapaian kompetensi aspek sikap dilihat dari skor siswa pada skala untuk mengukur karakter jujur siswa, skala untuk mengukur karakter tanggung jawab siswa, dan skala sikap matematis. Adapun pencapaian kompetensi aspek pengetahuan dan keterampilan dilihat dari skor siswa pada tes dan tugas.

Mengacu pada Ebel & Frisbie (1991: 266-280), disebutkan bahwa secara umum ada dua metode pemberian nilai yakni *relative grading* dan *absolut grading*. *Relatif grading* pada umumnya disusun berdasarkan kurva distribusi normal menggunakan nilai tengah dan simpangan baku skor yang diperoleh. Adapun *absolut grading* biasa disebut juga metode berbasis konten sehingga pada umumnya berupa nilai batas (*cutoff score*). *Absolut grading* biasanya digunakan pada penilaian aspek pengetahuan dan keterampilan yang menggunakan instrumen berupa tes dan tugas atau proyek lainnya. Selain metodenya, perlu diperhatikan bahwa kegiatan pemberian nilai pada siswa dipengaruhi oleh berbagai hal seperti keunikan individu, kondisi mental atau psikologisnya, kebebasan dan demokrasi dalam kelas, dan lain sebagainya.

a) Pencapaian Kompetensi Aspek Sikap

Penilaian pencapaian kompetensi siswa pada aspek sikap meliputi penilaian terhadap karakter jujur, karakter tanggung jawab, dan sikap matematis siswa. Penilaian diberikan dalam bentuk kategori dan bersifat *relatif grading*. Kategorisasi penilaian aspek sikap dalam penelitian ini mengacu pada kategori dari Mardapi (2016: 146). Penentuan kategori dilakukan berdasarkan skor siswa (X), rata-rata (\bar{x}), dan simpangan bakunya (s_x).

Kategorisasi penilaian pada skala untuk mengukur karakter jujur, skala untuk mengukur karakter tanggung jawab, dan skala sikap matematis siswa disajikan sebagai berikut.

Tabel 3. Kategorisasi Karakter Jujur

Skor Siswa	Kategori
$X \geq \bar{x} + 1. s_x$	Sangat Kuat
$\bar{x} \leq X < \bar{x} + 1. s_x$	Kuat
$\bar{x} - 1. s_x \leq X < \bar{x}$	Cukup Kuat
$X \leq \bar{x} - 1. s_x$	Kurang Kuat

Tabel 4. Kategorisasi Karakter Tanggung Jawab

Skor Siswa	Kategori
$X \geq \bar{x} + 1. s_x$	Sangat Kuat
$\bar{x} \leq X < \bar{x} + 1. s_x$	Kuat
$\bar{x} - 1. s_x \leq X < \bar{x}$	Cukup Kuat
$X \leq \bar{x} - 1. s_x$	Kurang Kuat

Tabel 5. Kategorisasi Sikap Matematis

Skor Siswa	Kategori Sikap Matematis
$X \geq \bar{x} + 1. s_x$	Sangat Baik
$\bar{x} \leq X < \bar{x} + 1. s_x$	Baik
$\bar{x} - 1. s_x \leq X < \bar{x}$	Cukup Baik
$X \leq \bar{x} - 1. s_x$	Kurang Baik

b) Pencapaian Kompetensi Aspek Pengetahuan dan Keterampilan

Penilaian terhadap pencapaian kompetensi aspek pengetahuan dan keterampilan berupa nilai yang dihitung dari skor siswa. Penilaian yang dilakukan menggunakan metode *absolut grading*. Nilai siswa untuk aspek pengetahuan dihitung dengan persamaan sebagai berikut.

$$\text{Nilai Siswa} = \frac{\text{Skor Siswa}}{\text{Skor Total Tes}} \times 100 \quad (15)$$

Adapun nilai siswa untuk aspek keterampilan dihitung dengan persamaan sebagai berikut.

$$\text{Nilai Siswa} = \frac{\text{Skor Siswa}}{\text{Skor Total Tugas}} \times 100 \quad (16)$$

Berdasarkan nilai siswa tersebut, siswa dinyatakan telah mencapai kompetensi yang diharapkan apabila memenuhi Kriteria Ketuntasan Minimal yang ditetapkan. Penelitian ini menggunakan KKM sebesar 75. Artinya, pencapaian kompetensi pengetahuan dan keterampilan siswa pada materi statistika dianggap tercapai atau tuntas apabila nilai siswa ≥ 75 . Penggunaan KKM ini mengacu pada panduan dari pemerintah dan hasil wawancara dengan para guru matematika SMP di DIY yang telah menerapkan Kurikulum 2013.

BAB IV

HASIL PENELITIAN DAN PENGEMBANGAN

A. Hasil Pengembangan Produk Awal

Pengembangan produk awal dalam penelitian ini dilakukan setelah peneliti melakukan studi pendahuluan yang meliputi telaah Kurikulum 2013, identifikasi masalah riil dan kajian kebutuhan, serta kajian berbagai referensi mengenai instrumen penilaian dalam pembelajaran matematika. Studi pendahuluan dilakukan dalam rangka memperoleh gambaran lebih mendalam tentang pelaksanaan penilaian matematika berdasarkan Kurikulum 2013 di DIY serta mengidentifikasi permasalahan dan kebutuhan para guru matematika khususnya terkait instrumen penilaian pada materi statistika. Adapun berdasarkan telaah Kurikulum 2013 dan kajian berbagai literatur mengenai instrumen penilaian, peneliti merumuskan langkah-langkah penyusunan instrumen penilaian aspek sikap, pengetahuan, dan keterampilan yang digunakan dalam penelitian ini.

Telaah Kurikulum 2013 dimaksudkan untuk mengkaji berbagai peraturan dan kebijakan pemerintah terkait pelaksanaan Kurikulum 2013 mengingat Kurikulum 2013 merupakan kurikulum baru yang diterapkan mulai tahun pelajaran 2013/2014 dan bersifat terbuka dengan berbagai perubahan. Kegiatan telaah Kurikulum 2013 juga dilakukan dengan mencermati Standar Kompetensi Lulusan (SKL) dan

Kompetensi Inti (KI) untuk jenjang SMP serta Kompetensi Dasar (KD) untuk mata pelajaran matematika SMP materi statistika berdasarkan Kurikulum 2013.

Berdasarkan hasil telaah diperoleh bahwa dalam pelaksanaan Kurikulum 2013 terus mengalami perbaikan mengikuti perkembangan yang terjadi di lapangan. Oleh sebab itu, pelaksanaan Kurikulum 2013 dari tahun ke tahun pun mengalami perubahan baik dalam hal penggunaan buku pelajaran dan panduan pelaksanaan maupun acuan peraturan yang digunakan. Rumusan Standar Kompetensi Lulusan (SKL) pada tingkat satuan pendidikan SMP juga terus mengalami perbaikan namun tetap meliputi aspek sikap, pengetahuan, dan keterampilan. Penelitian ini difokuskan pada materi statistika SMP berdasarkan Kurikulum 2013 yang diterapkan pada tahun pelajaran 2015/2016. Standar Kompetensi Lulusan, Kompetensi Inti, dan Kompetensi Dasar materi statistika kelas VII secara lengkap dapat dilihat pada Lampiran 4a.

Kegiatan identifikasi masalah riil dan kajian kebutuhan dilakukan untuk mendapatkan informasi realita permasalahan di lapangan serta mengidentifikasi kebutuhan para guru terkait penilaian pencapaian kompetensi siswa. Kegiatan ini dilakukan melalui penelusuran informasi dengan wawancara terhadap para guru matematika SMP yang telah menerapkan Kurikulum 2013 di DIY dan terlibat dalam penelitian ini. Informasi yang dihimpun meliputi: (a) penilaian aspek sikap, pengetahuan, dan keterampilan yang sudah dilakukan, (b) kesulitan atau hambatan yang dialami, (c) kegiatan penyusunan instrumen, (d) ketersediaan referensi terkait instrumen penilaian berdasarkan Kurikulum 2013 khususnya pada materi statistika,

(e) kegiatan penskoran, (f) kegiatan penentuan kualitas butir, (g) pendekatan teori yang digunakan ketika melakukan analisis butir, (h) solusi yang diharapkan untuk mengatasi permasalahan atau kesulitan yang terjadi, dan (i) Kriteria Ketuntasan Minimal (KKM) yang digunakan. Para guru yang terlibat dalam kegiatan studi awal adalah sebagai berikut.

Tabel 6. Daftar Guru dalam Kegiatan Studi Awal

Nama	Pendidikan	Asal Sekolah
Arw	S1	SMP N 5 Yogyakarta
Sra	S2	SMP N 5 Yogyakarta
Riy	S2	SMP N 5 Yogyakarta
Nip	S1	SMP Muhammadiyah 2 Yogyakarta
Ffa	S1	SMP Muhammadiyah 2 Yogyakarta
Nmt	S1	SMP Muhammadiyah 2 Yogyakarta
Rfi	S2	SMP N 3 Banguntapan
Smr	S2	SMP N 4 Depok
Dnw	S1	SMP IT Abu Bakar Yogyakarta
Npd	S1	SMP IT Abu Bakar Yogyakarta
Liw	S1	SMP N 1 Sleman
Tsm	S1	SMP N 1 Turi

Berdasarkan hasil wawancara diperoleh informasi bahwa penilaian pencapaian kompetensi matematika siswa pada aspek sikap, pengetahuan, dan keterampilan berdasarkan Kurikulum 2013 sudah berusaha dilakukan oleh para guru. Kegiatan penilaian dilakukan mengacu arahan dan panduan yang diperoleh dari berbagai sosialisasi dan pelatihan yang diberikan pemerintah. Penilaian sikap yang telah dilakukan para guru di antaranya adalah dengan observasi atau pengamatan langsung, jurnal, penilaian antar teman, dan penilaian diri. Penilaian pengetahuan dan keterampilan dilakukan dengan tes dan tugas. Tes yang digunakan berbentuk uraian

dan pilihan ganda. Terkait penilaian, masalah yang dialami para guru di antaranya adalah: (a) model penilaian dirasakan cukup rumit karena para guru merasa banyak hal yang mesti dinilai, (b) terjadinya perubahan kebijakan menjadikan batasan materi terasa tidak jelas sehingga penilaiannya juga menjadi kurang terarah, (c) menyusun instrumen penilaian khususnya untuk aspek sikap dan keterampilan terasa sulit karena belum terbiasa, (d) referensi contoh instrumen penilaian berdasarkan Kurikulum 2013 dari pemerintah kalimatnya dirasakan sulit dipahami terutama untuk siswa dengan kemampuan sedang dan kurang, dan (e) panduan dari pemerintah banyak memuat soal berpikir tingkat tinggi sementara para guru masih memerlukan contoh soal untuk kemampuan berpikir tingkat menengah.

Terkait penentuan kualitas butir, selama ini dilakukan secara langsung menggunakan program yang ada di sekolah namun terkait pendekatan teori yang digunakan ketika melakukan analisis butir kurang diperhatikan para guru. Berdasarkan hasil wawancara, semua guru menyatakan masih sangat memerlukan referensi terkait contoh instrumen penilaian berdasarkan Kurikulum 2013 termasuk dalam materi statistika. Adapun terkait Kriteria Ketuntasan Minimal (KKM) yang digunakan, diperoleh informasi bahwa penetapan KKM tergantung kebijakan masing-masing sekolah. KKM yang digunakan ternyata bervariasi antar sekolah, ada yang menggunakan KKM 80, 75, 68, atau 66. Hal ini mengingat bahwa penentuan KKM dipengaruhi oleh daya dukung sekolah, intake siswa, dan kompleksitas materi.

Mencermati informasi yang diperoleh menunjukkan gambaran pelaksanaan penilaian matematika SMP berdasarkan Kurikulum 2013 yang terjadi hingga saat ini

dan memerlukan perhatian semua pihak baik yang terkait langsung maupun tidak langsung. Berdasarkan studi pendahuluan tersebut menunjukkan bahwa kegiatan pengembangan instrumen penilaian dan penambahan ketersediaan contoh instrumen penilaian sangat diperlukan para guru matematika SMP yang menerapkan Kurikulum 2013. Oleh sebab itu, pengembangan instrumen penilaian matematika perlu dilakukan guna mempermudah dan membantu para guru matematika dalam melakukan penilaian berdasarkan Kurikulum 2013. Perlunya pengembangan instrumen penilaian matematika utamanya didasarkan pada fakta bahwa secara eksplisit semua guru yang terlibat dalam penelitian ini menyatakan masih sangat memerlukan tambahan ketersediaan referensi terkait instrumen penilaian yang mempermudah dan membantu para guru dalam melakukan penilaian berdasarkan Kurikulum 2013.

Pengkajian pustaka dilakukan dengan melakukan penelusuran berbagai referensi yang terkait dengan penilaian pencapaian kompetensi siswa dalam pembelajaran matematika, langkah-langkah penyusunan instrumen penilaian pencapaian kompetensi sikap, pengetahuan, dan keterampilan dalam pembelajaran matematika, kualitas instrumen penilaian termasuk uji keselarasan antara penilaian dengan standar dalam kurikulum. Penelusuran kajian pustaka juga dilakukan dengan mengkaji berbagai hasil penelitian yang relevan. Berdasarkan kajian berbagai referensi serta memperhatikan karakteristik mata pelajaran matematika di sekolah, peneliti merumuskan langkah-langkah penyusunan instrumen penilaian aspek sikap, pengetahuan, dan keterampilan dalam pembelajaran matematika yang digunakan dalam penelitian ini.

Penyusunan instrumen penilaian aspek sikap dalam penelitian ini dilakukan melalui langkah-langkah: (1) menentukan tujuan penilaian sikap dan jenis sikap yang hendak dinilai, (2) melakukan kajian teori, (3) memilih teknik penilaian, bentuk instrumen, dan model skala pengukuran, (4) menyusun kisi-kisi, (5) menulis butir-butir instrumen, sistem penskoran, dan cara penafsirannya. Adapun penyusunan instrumen penilaian pengetahuan dan keterampilan dalam penelitian ini dirancang secara bersamaan. Hal ini mengingat bahwa pada pembelajaran matematika, aspek pengetahuan dan keterampilan bersifat saling melengkapi dan terintegrasi sehingga penilaiannya seharusnya tidak dipisah-pisahkan. Langkah-langkah menyusun instrumen penilaian aspek pengetahuan dan keterampilan dalam penelitian ini adalah: (1) menentukan tujuan penilaian dan teknik penilaian yang digunakan, (2) menentukan indikator dan merumuskan *abstract continuum*, (3) menyusun kisi-kisi, dan (4) menulis butir-butir instrumen dan pedoman penskorannya.

Kegiatan selanjutnya setelah melalui tahap I yakni studi pendahuluan sebagaimana diuraikan di atas adalah mengembangkan produk awal yang dalam penelitian ini merupakan tahap II. Pengembangan instrumen dilakukan pada materi statistika SMP berdasarkan Kurikulum 2013. Secara rinci, pengembangan produk awal dalam penelitian ini dilakukan dalam tiga kegiatan yaitu:

- 1) Penyusunan instrumen awal yakni menyusun instrumen penilaian pencapaian kompetensi siswa untuk aspek sikap, pengetahuan, dan keterampilan pada materi statistika SMP kelas VII berdasarkan Kurikulum 2013 menggunakan langkah-langkah penyusunan instrumen yang telah dirumuskan.

- 2) Melakukan *Focus Group Discussion* (FGD) dengan guru dan dosen pendidikan matematika untuk mendiskusikan instrumen penilaian yang telah disusun.
- 3) Melakukan validasi instrumen penilaian yang dikembangkan kepada para pakar yang terdiri dari dua pakar pendidikan matematika, dua pakar statistika, dan dua pakar penilaian hasil belajar matematika.

Hasil dari pelaksanaan ketiga kegiatan tersebut secara detail dijabarkan sebagai berikut.

1. Penyusunan Instrumen Awal

- a. Penilaian Aspek Sikap

Penyusunan instrumen penilaian aspek sikap untuk materi statistika SMP berdasarkan Kurikulum 2013 dilakukan melalui langkah-langkah sebagai berikut.

- 1) Menentukan tujuan penilaian sikap dan jenis sikap yang hendak dinilai. Langkah ini dilakukan dengan mengacu kerangka dasar dan struktur kurikulum yang digunakan khususnya Kompetensi Dasar (KD) dari materi statistika SMP kelas VII berdasarkan Kurikulum 2013. Berdasarkan acuan tersebut, ditentukan tujuan penilaian aspek sikap dan jenis sikap yang dinilai. Tujuan penilaian aspek sikap adalah untuk mengetahui pencapaian kompetensi siswa pada aspek sikap. Adapun jenis sikap yang dinilai yakni karakter jujur, karakter tanggung jawab, dan sikap matematis.
- 2) Melakukan kajian teori. Kajian teori dilakukan guna memperoleh definisi konseptual dan merumuskan definisi operasionalnya dengan mengkaji berbagai teori tentang karakter jujur, karakter tanggung jawab, dan sikap matematis.

Berdasarkan definisi operasional yang ditetapkan selanjutnya dirumuskan indikator untuk masing-masing faktornya.

- 3) Memilih teknik penilaian, bentuk instrumen, dan model skala pengukuran. Berdasarkan definisi operasional yang telah dirumuskan indikator-indikatornya, langkah selanjutnya adalah memilih teknik penilaian, bentuk instrumen, dan model skala pengukuran dengan memperhatikan level sikap yang dinilai. Teknik penilaian yang dipilih dalam penelitian ini adalah penilaian diri (*self assessment*) dengan bentuk instrumen penilaian berupa skala sikap. Hal ini mengingat penilaian terhadap karakter jujur, karakter tanggung jawab, dan sikap matematis lebih tepat dan memungkinkan dilakukan para guru matematika pada level *moral knowing*. Model skala pengukuran yang dipilih dalam penelitian ini adalah penskalaan subjek untuk skala pengukur karakter jujur dan skala pengukur karakter tanggung jawab. Adapun untuk skala sikap matematis berupa penskalaan respons menggunakan skala Likert.
- 4) Menyusun kisi-kisi. Kisi-kisi instrumen untuk skala pengukur karakter jujur, skala pengukur karakter tanggung jawab, dan skala sikap matematis disusun dengan memuat faktor masing-masing skala dan indikator-indikatornya. Kisi-kisi dibuat dalam bentuk tabel matrik dan dilengkapi dengan keterangan nomor butir yang mewakili indikator-indikator tersebut. Adapun khusus untuk model penskalaan respons dilengkapi dengan keterangan tentang sifat butir termasuk butir yang *favorable* atau *unfavorable*. Kisi-kisi instrumen penilaian aspek sikap untuk tiap skala adalah sebagai berikut.

Tabel 7. Kisi-Kisi Skala untuk Mengukur Karakter Jujur

Faktor	Indikator	Nomor Butir	Kode Butir
A. Berperilaku apa adanya sesuai kenyataan.	1. Menyampaikan informasi apa adanya tanpa menambah atau mengurangi.	1,2	J1, J2
	2. Mengatakan kebenaran apapun konsekuensinya.	3	J3
	3. Mengungkapkan keadaan apa adanya atau tidak berpura-pura.	4,5	J4,J5
	4. Menyusun laporan berdasarkan data apa adanya.	6	J6
B. Mengakui setiap perbuatan baik positif maupun negatif.	1. Mengakui kesalahan yang dilakukan.	7	J7
	2. Memberi penghargaan pada kelebihan yang dimiliki diri sendiri.	8	J8
	3. Tidak menutup-nutupi kesalahan orang lain.	9	J9
	4. Memberi penghargaan pada kelebihan orang lain.	10	J10
C. Mengerjakan tugas atau ujian sesuai kemampuan sendiri.	1. Tidak menyontek.	11	J11
	2. Tidak mengharapkan bantuan teman.	12	J12
	3. Tidak memanfaatkan kesempatan yang ada untuk menyontek.	13	J13
	4. Tidak saling tukar hasil pekerjaan dengan teman.	14	J14
	5. Tidak mengambil/menyalin karya teman atau orang lain tanpa menyebutkan sumber.	15	J15
D. Dapat dipercaya atas kepemilikan orang lain	1. Memberi pengakuan atas kepemilikan orang lain.	16	J16
	2. Tidak merusak atau merugikan kepemilikan orang lain.	17	J17

Tabel 8. Kisi-Kisi Skala untuk Mengukur Karakter Tanggung Jawab

Faktor	Indikator	Nomor Butir	Kode Butir
A. Komitmen pada tugas	1. Melaksanakan dengan sungguh-sungguh setiap tugas yang diberikan guru.	1	T1
	2. Senantiasa mengkondisikan diri untuk siap mengerjakan tugas dari guru.	2	T2
	3. Berperan aktif memecahkan permasalahan dalam setiap tugas yang diberikan guru.	3	T3
	4. Mengerjakan tugas tanpa disuruh.	4	T4
	5. Menampilkan prakarsa dalam penyelesaian tugas.	5	T5
B. Sportif	1. Berani mengakui kesalahan yang dilakukan.	6	T6
	2. Menolak perbuatan menyalahkan/menuduh orang lain tanpa bukti yang akurat.	7	T7
	3. Bersedia meminta maaf atas kesalahan yang dilakukan.	8	T8
	4. Bersedia mengundurkan diri karena gagal dalam melaksanakan tugas, jika hal itu merupakan jalan keluar yang terbaik bagi kepentingan umum.	9	T9
	5. Bersedia dikenai sanksi hukum yang berlaku apabila terbukti melanggar peraturan.	10	T10
C. Disiplin	1. Mematuhi peraturan yang ada di kelas.	11	T11
	2. Mematuhi peraturan dalam kelompok.	12	T12
	3. Menepati janji.	13	T13
	4. Tepat waktu dalam mengembalikan setiap barang yang dipinjam dari orang lain.	14	T14

Tabel 9. Kisi-Kisi Skala Sikap Matematis

Faktor	Indikator	Nomor Butir	Kode Butir
A. Mencoba memahami masalah atau tujuan atau substansi masalah dengan jelas secara mandiri	1. Berusaha untuk bertanya.	1(+), 2(-)	M1, M2
	2. Berusaha untuk memahami persoalan.	3(+), 4(+)	M3, M4
	3. Berusaha menemukan masalah matematika dari kehidupan sehari-hari.	5(-)	M5
B. Mencoba mengambil tindakan logis	1. Berusaha mencapai kompetensi matematika yang ditetapkan.	6(+), 7(-)	M6, M7
	2. Berusaha memahami sifat-sifat matematika.	8(+), 9(+)	M8, M9
	3. Berusaha berpikir berdasarkan data yang dapat digunakan, yang sebelumnya telah dipelajari, dan asumsi.	10(+)	M10
C. Mencoba mengekspresikan berbagai hal dengan jelas dan ringkas	1. Berusaha untuk merekam dan mengkomunikasikan masalah dan hasilnya jelas dan ringkas.	11(-)	M11
	2. Berusaha berpikir secara sistematis.	13(-)	M13
D. Mencoba mencari penyelesaian yang lebih baik	1. Berusaha untuk memahami matematika dari yang konkrit menuju abstrak.	12(+)	M12
	2. Berusaha berpikir secara objektif dan subjektif serta berpikir kritis.	14(+)	M14
	3. Berusaha memanfaatkan pikiran dan hasil usaha yang telah didapat.	15(+)	M15

5) Menulis butir-butir instrumen, sistem penskoran, dan cara penafsirannya. Penulisan butir-butir instrumen, sistem penskoran, dan cara penafsirannya mengacu pada skala pengukuran yang telah ditentukan dan kisi-kisi instrumen. Secara lengkap, sistem penskoran dan cara penafsiran masing-masing skala dapat dilihat pada Lampiran Produk Disertasi.

b. Penilaian Aspek Pengetahuan dan Keterampilan

Penyusunan instrumen penilaian pencapaian kompetensi untuk aspek pengetahuan dan keterampilan dalam penelitian ini dirancang secara bersamaan. Hal ini mengingat bahwa pada pembelajaran matematika, aspek pengetahuan dan keterampilan bersifat saling melengkapi dan terintegrasi sehingga penilaiannya seharusnya tidak dipisah-pisahkan. Adapun langkah-langkah menyusun instrumen penilaian pengetahuan dan keterampilan untuk materi statistika SMP kelas VII yang dilakukan dalam penelitian ini adalah sebagai berikut.

1) Menentukan tujuan penilaian dan teknik penilaian yang digunakan.

Langkah ini dilakukan dengan mencermati kerangka dasar dan struktur kurikulum yang digunakan dan melakukan analisis Kompetensi Dasar (KD) pada materi statistika SMP berdasarkan Kurikulum 2013. Materi statistika yang dipilih adalah materi statistika kelas VII berdasarkan peraturan pelaksanaan Kurikulum 2013 yang berlaku pada tahun pelajaran 2015/2016. Kompetensi Dasar dalam materi statistika kelas VII yang menjadi fokus penelitian ini adalah menganalisis hubungan antara data dua variabel dengan cara penyajiannya (tabel, grafik garis, diagram batang, dan diagram lingkaran). Adapun teknik penilaian yang dipilih

adalah tes uraian yang bersifat objektif dan tugas proyek individu. Bentuk tes uraian objektif dipilih dimaksudkan agar bisa mengungkap pencapaian kompetensi pengetahuan siswa secara lebih detail dan mendalam pada keseluruhan level kemampuan siswa. Tugas proyek berbentuk penugasan untuk dikerjakan siswa yang pengerjaannya dilakukan secara individu dimaksudkan agar dapat mengungkap pencapaian kompetensi keterampilan siswa secara perseorangan.

2) Menentukan indikator dan merumuskan *abstract continuum*.

Berdasarkan Kompetensi Dasar materi statistika kelas VII tahun pelajaran 2015/2016 dirumuskan indikator pencapaian kompetensi. Mengacu pada indikator-indikator tersebut dibuat rumusan *learning continuum* yakni semua indikator yang dihasilkan dipindahkan secara berurutan sebagai *abstract continuum*. Kompetensi Dasar, indikator, dan rumusan *abstract continuum* secara lengkap dapat dilihat pada Lampiran 4c.

3) Menyusun kisi-kisi.

Kisi-kisi instrumen penilaian pengetahuan dan keterampilan disusun secara bersamaan dan disajikan dalam satu tabel kisi-kisi. Kisi-kisi instrumen penilaian pengetahuan dan keterampilan merupakan tabel matrik yang memuat Kompetensi Dasar (KD), indikator, dimensi pengetahuan dan proses kognitif dari revisi taksonomi Bloom yang bersesuaian dengan indikator, bentuk butir, dan jumlah butir. Kisi-kisi ini disusun untuk memastikan butir-butir tes dan tugas telah mewakili apa yang seharusnya diukur secara proporsional.

Tabel 10. Kisi-Kisi Instrumen Penilaian Aspek Pengetahuan dan Keterampilan

Kompetensi Dasar	Indikator	Bentuk Instrumen	Nomor Butir Tes	Taksonomi Bloom yang Direvisi	Nomor Butir Tugas Proyek
Aspek Pengetahuan	1. Mengemukakan berbagai cara mengumpulkan data.	Tes Uraian dan Tugas Proyek	1(a)	<i>B3</i>	1
	2. Mengemukakan berbagai cara menyajikan data.	Tes Uraian	1(b)	<i>B3</i>	
Menganalisis hubungan antara data dengan cara penyajiannya (tabel, grafik garis, diagram batang, dan diagram lingkaran).	3. Menyajikan data menggunakan tabel.	Tes Uraian dan Tugas Proyek	2(a)	<i>C3</i>	2(a)
	4. Menganalisis hubungan antara data yang disajikan dalam bentuk tabel.	Tes	2(b)	<i>C4</i>	
Aspek Keterampilan	5. Menyajikan data menggunakan diagram batang.	Tes Uraian dan Tugas Proyek	3(a)	<i>C3</i>	2(a)
	6. Menyajikan data menggunakan diagram lingkaran.	Tes Uraian dan Tugas Proyek	3(b)	<i>C3</i>	2(b)
Menyajikan dan menafsirkan data dalam bentuk tabel, grafik garis, diagram batang dan diagram lingkaran	7. Menganalisis hubungan antara data yang disajikan dalam bentuk diagram batang dan lingkaran.	Tes	3(c)	<i>C4</i>	
	8. Menyajikan data menggunakan grafik garis.	Tes	4(a)	<i>C3</i>	
	9. Menganalisis hubungan antara data yang disajikan dalam bentuk grafik garis.	Tes	4(b)	<i>C5</i>	

4) Menulis butir-butir instrumen dan pedoman penskorannya.

Berdasarkan kisi-kisi yang telah disusun, dilakukan penulisan butir-butir instrumen tes dan tugas serta pedoman penskorannya menggunakan bentuk parsial untuk tes dan rubrik untuk tugas. Secara lengkap, pedoman penskoran tes dan rubrik tugas dapat dilihat pada Lampiran Produk Disertasi.

2. Focus Group Discussion (FGD)

Instrumen penilaian aspek sikap, pengetahuan, dan keterampilan yang telah tersusun berdasarkan hasil kegiatan penyusunan instrumen awal, selanjutnya didiskusikan dalam forum *Focus Group Discussion* (FGD). Hasil FGD ditindaklanjuti peneliti dengan konsultasi pembimbing. Kegiatan FGD dilaksanakan pada hari Kamis tanggal 21 Juli 2016 di Ruang 1.15 Gedung Baru Program Pascasarjana Universitas Negeri Yogyakarta (UNY). Peserta FGD adalah 4 guru matematika dan 3 dosen dari Jurusan Pendidikan Matematika FMIPA UNY. Daftar hadir dan hasil pelaksanaan FGD secara lengkap dapat dilihat pada Lampiran 1.

3. Validasi Instrumen

Validasi instrumen dalam penelitian ini dilakukan dengan teknik *delphi* dan melibatkan enam pakar yang terdiri atas dua pakar pendidikan matematika, dua pakar statistika, dan dua pakar penilaian hasil belajar matematika sebagaimana tercantum pada Tabel 2. Para pakar melakukan telaah dan penilaian terhadap instrumen penilaian yang disusun peneliti. Telaah yang dilakukan para pakar pada instrumen penilaian aspek sikap meliputi: (a) kesesuaian butir dengan indikator, (b) bahasa yang digunakan, (c) pernyataan butir yang tidak bias, (d) kejelasan pilihan jawaban, dan

(e) ketepatan pedoman penskoran. Adapun untuk instrumen penilaian aspek pengetahuan dan keterampilan meliputi: (a) materi/substansi, (b) konstruksi, dan (c) bahasa. Hal-hal yang dinilai para pakar tersebut tersebut dituangkan pada lembar telaah sebagai acuan bagi para pakar dalam memberikan penilaian. Format lembar telaah secara lengkap dapat dilihat pada Lampiran 2a.

Hasil penilaian para pakar pada kegiatan validasi ini bersifat kualitatif dan kuantitatif. Hasil kualitatif berupa pernyataan para pakar yang menyatakan bahwa instrumen telah layak dan siap digunakan. Pernyataan ini diperoleh setelah peneliti melakukan revisi berdasarkan masukan dan saran yang diberikan para pakar. Secara lengkap, masukan dan saran para pakar serta tindak lanjut yang dilakukan peneliti dapat dilihat pada Lampiran 2b. Adapun hasil kuantitatif berupa penilaian para pakar terhadap kesesuaian butir dengan indikator menggunakan skala Likert yakni Sangat Tidak Sesuai (STS), Tidak Sesuai (TS), Kurang Sesuai (KS), Sesuai (S), dan Sangat Sesuai (SS). Khusus untuk instrumen tes, para pakar juga memberikan penilaian terkait muatan revisi taksonomi Bloom. Data hasil penilaian para pakar secara lengkap dapat dilihat pada Lampiran 2c. Berdasarkan hasil akhir validasi, semua pakar menyatakan bahwa instrumen yang telah disusun layak dan siap digunakan. Pernyataan para pakar yang menyatakan kelayakan instrumen dapat dilihat pada Lampiran 2d.

Kelayakan instrumen penilaian yang dikembangkan, didukung pula dengan besarnya indeks validitas isi yang dihitung menggunakan rumus V Aiken. Hasil perhitungan indeks V Aiken untuk kesesuaian antara butir dengan indikator

dibandingkan dengan nilai V_{tabel} yang tertera pada tabel V dari Aiken sebagaimana dapat dilihat pada Lampiran 3b. Nilai V_{tabel} yang digunakan dalam penelitian ini adalah 0,79. Hal ini karena validasi dalam penelitian ini melibatkan enam pakar dan banyaknya kategori pilihan jawaban ada lima yakni yakni STS, TS, KS, S, dan SS. Adapun skor untuk masing-masing kategori adalah STS=1, TS=2, KS=3, S=4, dan SS=5. Hasil perhitungan indeks V Aiken secara lengkap dapat dilihat pada Lampiran 3a. Adapun secara ringkas hasil perhitungan indeks V Aiken pada skala untuk mengukur karakter jujur, skala untuk mengukur karakter tanggung jawab, dan skala sikap matematis disajikan sebagai berikut.

Tabel 11. Hasil Perhitungan Indeks V Aiken pada Skala untuk Mengukur Karakter Jujur

Butir	s	n	$c-1$	V_{tabel}	V_{hitung}	Keterangan
1	22	6	4	0,79	0,92	Valid
2	21	6	4	0,79	0,88	Valid
3	22	6	4	0,79	0,92	Valid
4	22	6	4	0,79	0,92	Valid
5	22	6	4	0,79	0,92	Valid
6	21	6	4	0,79	0,88	Valid
7	22	6	4	0,79	0,92	Valid
8	22	6	4	0,79	0,92	Valid
9	21	6	4	0,79	0,88	Valid
10	21	6	4	0,79	0,88	Valid
11	22	6	4	0,79	0,92	Valid
12	22	6	4	0,79	0,92	Valid
13	21	6	4	0,79	0,88	Valid
14	22	6	4	0,79	0,92	Valid
15	22	6	4	0,79	0,92	Valid
16	22	6	4	0,79	0,92	Valid
17	22	6	4	0,79	0,92	Valid

Tabel 12. Hasil Perhitungan Indeks V Aiken pada Skala untuk Mengukur Karakter Tanggung Jawab

Butir	<i>s</i>	<i>n</i>	<i>c-1</i>	V_{tabel}	V_{hitung}	Keterangan
1	22	6	4	0,79	0,92	Valid
2	21	6	4	0,79	0,88	Valid
3	22	6	4	0,79	0,92	Valid
4	22	6	4	0,79	0,92	Valid
5	22	6	4	0,79	0,92	Valid
6	21	6	4	0,79	0,88	Valid
7	22	6	4	0,79	0,92	Valid
8	22	6	4	0,79	0,92	Valid
9	21	6	4	0,79	0,88	Valid
10	21	6	4	0,79	0,88	Valid
11	22	6	4	0,79	0,92	Valid
12	22	6	4	0,79	0,92	Valid
13	21	6	4	0,79	0,88	Valid
14	22	6	4	0,79	0,92	Valid

Tabel 13. Hasil Perhitungan Indeks V Aiken pada Skala Sikap Matematis

Butir	<i>s</i>	<i>n</i>	<i>c-1</i>	V_{tabel}	V_{hitung}	Keterangan
1	22	6	4	0,79	0,92	Valid
2	21	6	4	0,79	0,88	Valid
3	22	6	4	0,79	0,92	Valid
4	22	6	4	0,79	0,92	Valid
5	22	6	4	0,79	0,92	Valid
6	21	6	4	0,79	0,88	Valid
7	22	6	4	0,79	0,92	Valid
8	22	6	4	0,79	0,92	Valid
9	21	6	4	0,79	0,88	Valid
10	21	6	4	0,79	0,88	Valid
11	22	6	4	0,79	0,92	Valid
12	22	6	4	0,79	0,92	Valid
13	21	6	4	0,79	0,88	Valid
14	22	6	4	0,79	0,92	Valid
15	22	6	4	0,79	0,92	Valid

Berdasarkan Tabel 11 tampak bahwa nilai V_{hitung} butir-butir skala untuk mengukur karakter jujur sebesar 0,88-0,92. Berdasarkan Tabel 12, nilai V_{hitung} butir-butir skala untuk mengukur karakter tanggung jawab sebesar 0,88-0,92. Berdasarkan Tabel 13, nilai V_{hitung} butir-butir skala untuk mengukur sikap matematis sebesar 0,88-0,92. Hal ini menunjukkan bahwa seluruh butir pada skala untuk mengukur karakter jujur, skala untuk mengukur karakter tanggung jawab, dan skala sikap matematis memiliki validitas isi yang baik ditinjau dari kesesuaiannya dengan indikator. Artinya, skala untuk mengukur karakter jujur dapat digunakan untuk mengukur karakter jujur siswa, skala untuk mengukur tanggung jawab dapat digunakan untuk mengukur karakter tanggung jawab siswa, dan skala sikap matematis dapat digunakan untuk mengukur sikap matematis siswa.

Adapun untuk instrumen penilaian aspek pengetahuan dan keterampilan yang berupa tes dan tugas, hasil perhitungan indeks V Aiken untuk kesesuaian antara butir dengan indikator dapat dilihat pada Lampiran 3a. Secara ringkas disajikan sebagai berikut.

Tabel 14. Hasil Perhitungan Indeks V Aiken pada Tes

Butir	s	n	$c-1$	V_{tabel}	V_{hitung}	Keterangan
1a	21	6	4	0,79	0,88	Valid
1b	21	6	4	0,79	0,88	Valid
2a	24	6	4	0,79	1,00	Valid
2b	24	6	4	0,79	1,00	Valid
3a	22	6	4	0,79	0,92	Valid
3b	22	6	4	0,79	0,92	Valid
3c	22	6	4	0,79	0,92	Valid
4a	23	6	4	0,79	0,96	Valid
4b	23	6	4	0,79	0,96	Valid

Tabel 15. Hasil Perhitungan Indeks V Aiken pada Tugas

Butir	<i>s</i>	<i>n</i>	<i>c-l</i>	V_{tabel}	V_{hitung}	Keterangan
1a	21	6	4	0,79	0,88	Valid
2(a)	22	6	4	0,79	0,92	Valid
2(b)	22	6	4	0,79	0,92	Valid

Tabel 14 menunjukkan bahwa nilai V_{hitung} butir-butir tes sebesar 0,88-1,00. Adapun berdasarkan Tabel 15, nilai V_{hitung} butir-butir tugas sebesar 0,88-0,92. Hal ini menunjukkan bahwa seluruh butir pada tes dan tugas memiliki validitas isi yang baik ditinjau dari kesesuaiannya dengan indikator. Khusus untuk instrumen tes, dukungan bukti validitas isi juga diperoleh dari hasil perhitungan indeks keselarasan Porter yakni sebesar 0,93. Besarnya indeks keselarasan ini termasuk dalam kategori sangat baik sehingga dapat dikatakan bahwa tes yang dikembangkan memiliki keselarasan yang sangat baik dengan standar yang ditetapkan dalam Kurikulum 2013 untuk materi statistika SMP. Secara lengkap perhitungan indeks keselarasan Porter dapat dilihat pada Lampiran 3c.

Mengacu pada keseluruhan hasil validasi sebagaimana dikemukakan di atas, dapat disimpulkan bahwa instrumen penilaian aspek sikap, pengetahuan, dan keterampilan yang dikembangkan dalam penelitian ini secara teoretis terbukti memiliki validitas isi yang baik. Hal ini berarti bahwa keseluruhan instrumen penilaian yang dikembangkan dalam penelitian ini ditinjau dari isinya dapat mengukur apa yang seharusnya diukur. Oleh sebab itu, keseluruhan instrumen dapat dikatakan telah layak untuk digunakan atau diujicobakan ke sekolah dalam rangka memperoleh bukti empiris kualitas instrumen.

B. Hasil Uji Coba Produk

1. Uji Coba I

Pelaksanaan uji coba I terhadap produk yang dikembangkan dalam penelitian ini dilakukan pada bulan Agustus 2016 sampai dengan September 2016. Kegiatan uji coba dilakukan pada seluruh siswa kelas VIII di SMP N 5 Yogyakarta, SMPIT Abu Bakar Yogyakarta, dan SMP Muhammadiyah 2 Yogyakarta. Pemilihan siswa kelas VIII dilakukan mengingat materi statistika pada pelaksanaan Kurikulum 2013 tahun pelajaran 2015/2016 untuk kelas VII dipelajari di akhir semester 2 dan kondisi di lapangan tidak memungkinkan untuk diambil data karena sangat berdekatan dengan pelaksanaan ujian kenaikan kelas. Oleh sebab itu, data diambil ketika siswa masuk awal semester 1 kelas VIII. Hal ini masih sangat bersesuaian karena pada semester 1 kelas VIII ini materi statistika juga dipelajari siswa dengan mengacu pada peraturan Kurikulum 2013 yang berlaku pada tahun pelajaran tersebut.

Berdasarkan hasil uji coba I, sebanyak 688 responden mengisi dan mengerjakan keseluruhan instrumen penilaian yang dikembangkan. Data hasil uji coba secara lengkap dapat dilihat pada Lampiran 5 untuk instrumen penilaian aspek sikap dan Lampiran 6 untuk instrumen penilaian aspek pengetahuan dan keterampilan. Hasil pelaksanaan uji coba I secara lengkap diuraikan sebagai berikut.

a. Instrumen Penilaian Aspek Sikap

Data hasil uji coba instrumen penilaian aspek sikap dianalisis dengan bantuan program *Lisrel* versi 8.51 menggunakan *second order* CFA. Sebelum dilakukan analisis *second order* CFA terlebih dahulu dilakukan uji persyaratan analisis yakni uji

multivariate normality. Secara lengkap hasil uji *multivariate normality* dari skala untuk mengukur karakter jujur dapat dilihat pada Lampiran 7a, skala untuk mengukur karakter tanggung jawab pada Lampiran 7d, dan skala sikap matematis pada Lampiran 7g. Berdasarkan hasil analisis diperoleh bahwa data pada semua skala ternyata tidak berdistribusi normal multivariat. Mengacu pada Widhiarso (2012: 1-2) yang mengemukakan bahwa pendekatan yang dapat digunakan untuk mengatasi ketidaknormalan data adalah GLS dan WLS, dan mengingat data dari instrumen penilaian aspek sikap bersifat ordinal maka analisis *second order CFA* dilakukan dengan model estimasi *Weighted Least Square (WLS)*.

Analisis *second order CFA* selanjutnya dilakukan pada semua skala yang dikembangkan dalam penelitian ini. Secara lengkap, hasil analisis *second order CFA* berdasarkan data uji coba I dari skala untuk mengukur karakter jujur dapat dilihat pada Lampiran 7b, Skala untuk mengukur karakter tanggung jawab pada Lampiran 7e, dan skala sikap matematis pada Lampiran 7h.

Berdasarkan hasil analisis Lisrel pada data dari skala untuk mengukur karakter jujur pada uji coba I sebagaimana dapat dilihat pada Lampiran 7b, diperoleh nilai RMSEA sebesar 0,048, Chi square 298,98 dengan *p value* 0,00, GFI sebesar 0,98, dan AGFI sebesar 0,97. Walaupun *Chi square* tidak signifikan karena nilai *p value* kurang dari 0,05, namun kriteria lain telah terpenuhi. Berdasarkan hasil ini kecocokan model dapat dianggap telah terpenuhi. Hal ini mengacu pada Hooper, Coughlan, & Mullen (2008: 57) yang menyebutkan bahwa penggunaan *Chi square* bersifat sensitif terhadap ukuran sampel yang digunakan dan Hair et al. (2009: 645)

yang mengemukakan bahwa nilai *Chi square* yang tidak signifikan tetap memungkinkan peneliti untuk mengatakan bahwa kecocokan modelnya adalah baik dengan memperhatikan indeks kecocokan model yang lainnya. Adapun hasil *second order CFA* untuk nilai *t-value* dan *Standardized Loading Factor* (SLF) skala untuk mengukur karakter jujur disajikan sebagai berikut.

Tabel 16. Hasil *Second Order CFA* Skala untuk Mengukur Karakter Jujur pada Uji Coba I

Faktor	Indikator	Butir	<i>Second order CFA</i>		Keterangan
			<i>t-value</i>	SLF	
A. Berperilaku apa adanya sesuai kenyataan.	A1	J1	**	0,36	Baik
	A2	J2	2,61	0,14	Kurang Baik (Direvisi)
	A3	J3	5,03	0,29	Kurang Baik (Direvisi)
	A4	J4	6,77	0,58	Baik
	A5	J5	5,54	0,30	Baik
	A6	J6	0,21	0,01	Kurang Baik (Direvisi)
B. Mengakui setiap perbuatan baik positif atau negatif.	B1	J7	**	0,75	Baik
	B2	J8	16,22	0,69	Baik
	B3	J9	13,69	0,65	Baik
	B4	J10	5,02	0,24	Kurang Baik (Direvisi)
C. Mengerjakan tugas atau ujian sesuai kemampuan sendiri.	C1	J11	**	0,69	Baik
	C2	J12	13,70	0,57	Baik
	C3	J13	18,29	0,81	Baik
	C4	J14	17,64	0,80	Baik
	C5	J15	14,11	0,59	Baik
D. Dapat dipercaya atas kepemilikan orang lain.	D1	J16	**	0,68	Baik
	D2	J17	4,95	0,54	Baik

Tabel 16 menunjukkan bahwa ditinjau dari *t-value*, seluruh butir signifikan dalam mendukung konstruk jujur, namun ada empat butir yang memiliki nilai SLF kurang dari 0,3 yakni butir J2, J3, J6, dan J10. Mengingat bahwa tiap butir ini mewakili satu indikator maka keempat butir ini tidak dieliminasi melainkan direvisi berdasarkan masukan siswa dan diskusi dengan pembimbing.

Berdasarkan hasil analisis Lisrel pada data dari skala untuk mengukur karakter tanggung jawab pada uji coba I sebagaimana dapat dilihat pada Lampiran 7e, diperoleh nilai RMSEA sebesar 0,046, *Chi square* sebesar 181,91 dengan *p-value* sebesar 0,00, GFI sebesar 0,98, dan AGFI sebesar 0,98. Walaupun *Chi square* tidak signifikan karena *p-value* kurang dari 0,05, namun kriteria lain telah terpenuhi. Oleh sebab itu, mengacu pada Hooper, Coughlan, & Mullen (2008: 57) dan Hair et al. (2009: 645), kecocokan model dapat dianggap telah terpenuhi. Adapun hasil *second order* CFA untuk nilai *t-value* dan *Standardized Loading Factor* (SLF) skala untuk mengukur tanggung jawab disajikan sebagai berikut.

Tabel 17. Hasil *Second Order* CFA Skala untuk Mengukur Karakter Tanggung Jawab pada Uji Coba I

Faktor	Indikator	Butir	<i>Second order</i> CFA		Keterangan
			<i>t-value</i>	SLF	
A. Komitmen pada tugas.	A1	T1	**	0,58	Baik
	A2	T2	11,93	0,54	Baik
	A3	T3	12,98	0,69	Baik
	A4	T4	13,75	0,72	Baik
	A5	T5	14,10	0,71	Baik
B. Sportif.	B1	T6	**	0,50	Baik
	B2	T7	7,56	0,54	Baik
	B3	T8	8,30	0,60	Baik
	B4	T9	2,28	0,16	Kurang Baik (Direvisi)
	B5	T10	10,14	0,66	Baik
C. Disiplin.	C1	T11	**	0,42	Baik
	C2	T12	6,83	0,44	Baik
	C3	T13	6,22	0,38	Baik
	C4	T14	6,75	0,61	Baik

Tabel 17 menunjukkan bahwa ditinjau dari *t-value*, seluruh butir signifikan dalam mendukung konstruk tanggung jawab, namun ada satu butir yang memiliki nilai SLF

kurang dari 0,3 yakni butir T9. Mengingat butir ini mewakili satu indikator maka tidak dieliminasi melainkan dilakukan perbaikan atau direvisi.

Berdasarkan hasil analisis Lisrel untuk data skala sikap matematis pada uji coba I sebagaimana dapat dilihat pada Lampiran 7h, diperoleh nilai RMSEA sebesar 0,064, *chi square* sebesar 331,65 dengan *p-value* sebesar 0,00, GFI sebesar 0,97, dan AGFI sebesar 0,96. Walaupun *chi square* tidak signifikan karena nilai *p-value* kurang dari 0,05, namun kriteria lain telah terpenuhi. Oleh sebab itu, mengacu pada Hooper, Coughlan, & Mullen (2008: 57) dan Hair et al. (2009: 645), kecocokan model dapat dianggap telah terpenuhi.

Adapun hasil *second order* CFA untuk nilai *t-value* dan *Standardized Loading Factor* (SLF) skala sikap matematis secara ringkas disajikan sebagai berikut.

Tabel 18. Hasil *Second Order* CFA Skala Sikap Matematis pada Uji Coba I

Faktor	Indikator	Butir	<i>Second order</i> CFA		Keterangan
			<i>t-value</i>	SLF	
A. Mencoba memahami masalah atau tujuan atau substansi masalah dengan jelas secara mandiri.	A1	M1	**	0,56	Baik
		M2	6,98	0,30	Baik
	A2	M3	13,18	0,63	Baik
		M5	9,46	0,37	Baik
A3	M4	7,05	0,30	Baik	
B. Mencoba mengambil tindakan logis.	B1	M6	**	0,57	Baik
	B2	M7	8,95	0,35	Baik
		M8	16,91	0,76	Baik
	B3	M9	13,25	0,54	Baik
M10		15,46	0,74	Baik	
C. Mencoba mengekspresikan berbagai hal dengan jelas dan ringkas.	C1	M11	**	0,48	Baik
	C2	M13	5,79	0,54	Baik
D. Mencoba mencari penyelesaian yang lebih baik.	D1	M12	**	0,41	Baik
	D2	M14	3,83	0,15	Tidak Baik (Direvisi)
	D3	M15	9,64	0,64	Baik

Tabel 18 menunjukkan bahwa ditinjau dari *t-value*, seluruh butir signifikan dalam mendukung konstruk sikap matematis, namun ada satu butir yang memiliki nilai SLF kurang dari 0,3 yakni butir M14. Mengingat butir ini mewakili satu indikator maka butir ini tidak dieliminasi melainkan dilakukan perbaikan atau direvisi baik berdasarkan masukan dari siswa maupun diskusi dengan pembimbing.

b. Instrumen Penilaian Aspek Pengetahuan dan Keterampilan

1. Tes

Data hasil uji coba I pada tes dianalisis dengan bantuan program Lisrel menggunakan *first order* CFA. Hasil uji persyaratan analisis yakni uji *multivariate normality* untuk data tes secara lengkap dapat dilihat pada Lampiran 9a. Berdasarkan hasil analisis menunjukkan bahwa data tidak berdistribusi normal multivariat. Oleh sebab itu, analisis *first order* CFA dilakukan dengan model estimasi *Generalized Least Square* (GLS). Pemilihan model GLS karena data tes merupakan data yang bersifat kontinu. Secara lengkap hasil analisis *first order* CFA untuk tes pada uji coba I ini dapat dilihat pada Lampiran 9b. Adapun perhitungan koefisien reliabilitas tes untuk koefisien *Construct Reliability* (CR) dan koefisien Alpha dari Cronbach (*Cronbach's Alpha*) dapat dilihat pada Lampiran 10a.

Berdasarkan hasil analisis Lisrel sebagaimana dapat dilihat pada Lampiran 9b diperoleh nilai RMSEA sebesar 0,031, *chi Square* sebesar 20,11 dengan *p-value* 0,06, GFI sebesar 0,99, dan AGFI sebesar 0,98. Hasil ini menunjukkan bahwa kecocokan model telah terpenuhi dengan baik.

Adapun hasil *first order* CFA untuk nilai *t-value* dan *Standardized Loading Factor* (SLF) serta hasil perhitungan koefisien CR dan Alpha untuk tes pada kegiatan uji coba I secara ringkas disajikan sebagai berikut.

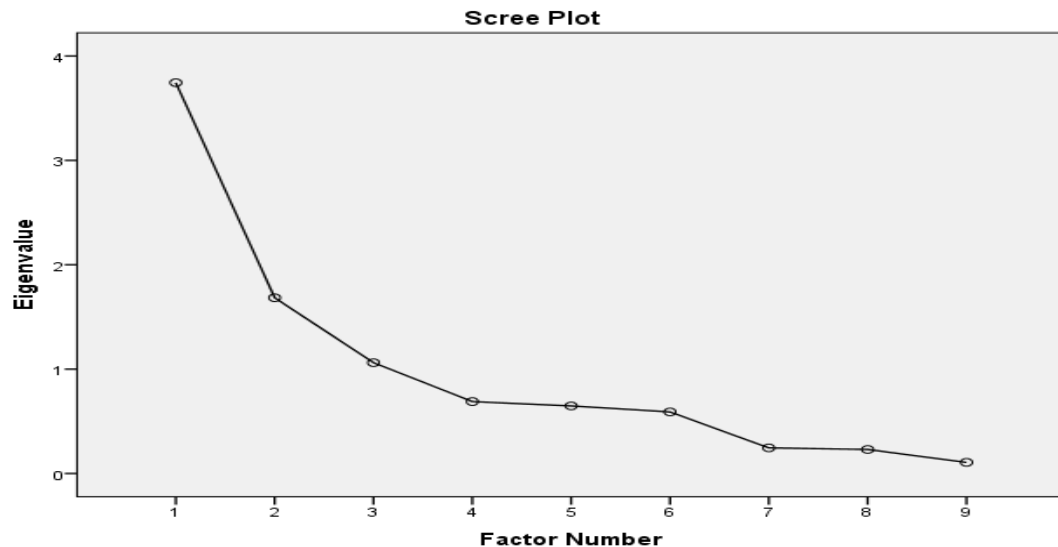
Tabel 19. Hasil *First Order* CFA Tes pada Uji Coba I

Butir	<i>First order</i> CFA				Validitas	Reliabilitas	
	<i>t-value</i>	Keterangan	SLF	<i>error</i>		CR	Alpha
1a	5,80	Signifikan	0,24	0,18	Cukup	0,88	0,78
1b	7,46	Signifikan	0,48	0,19	Baik		
2a	22,28	Signifikan	0,89	0,099	Baik		
2b	7,78	Signifikan	0,33	0,54	Baik		
3a	6,26	Signifikan	0,48	0,12	Baik		
3b	14,74	Signifikan	0,77	0,88	Baik		
3c	9,34	Signifikan	0,37	0,53	Baik		
4a	24,38	Signifikan	0,99	0,0067	Baik		
4b	16,58	Signifikan	0,44	0,84	Baik		

Tabel 19 menunjukkan bahwa ditinjau dari *t-value*, seluruh butir signifikan dalam mendukung konstruk tes dengan dukungan terendah oleh butir 1a dan tertinggi oleh butir 4a. Reliabilitas tes termasuk kategori baik dengan koefisien CR sebesar 0,88 dan koefisien Alpha 0,78. Hal ini berarti tes materi statistika ditinjau dari konstruksinya terdiri atas butir-butir yang dapat mengukur kemampuan statistika siswa sesuai dengan Kompetensi Dasar. Adapun ditinjau dari reliabilitasnya dapat dikatakan bahwa tes reliabel yang berarti bahwa hasil pengukuran yang diperoleh melalui tes ini bersifat ajeg atau konsisten.

Terbuktinya validitas konstruk yang baik pada tes melalui *first order* CFA ini sekaligus telah menunjukkan terpenuhinya asumsi unidimensi pada tes. Artinya, tes hanya mengukur satu kemampuan. Terpenuhinya asumsi unidimensi ini juga dapat

dilihat secara visual dalam grafik berupa *plot eigen value* yang diperoleh melalui analisis faktor dengan SPSS versi 20 sebagaimana disajikan pada Gambar 6 sebagai berikut.



Gambar 6. Grafik *Eigen Value* untuk Tes

Gambar 6 menunjukkan bahwa *eigen value* pada salah satu faktor bersifat dominan dibandingkan *eigen value* pada faktor lainnya. Hal ini berarti bahwa asumsi unidimensi dapat dikatakan telah terpenuhi, sehingga bisa dilakukan analisis menggunakan IRT politomus.

Karakteristik butir tes ditentukan menggunakan pendekatan IRT politomus model GPCM dan diperoleh melalui analisis dengan bantuan program *Parscale* versi 4.1. Berdasarkan hasil analisis diperoleh estimasi parameter butir, kemampuan atau theta, kurva karakteristik butir, kurva informasi butir dan tes serta kesalahan baku pengukurannya. Secara lengkap hasil analisis *Parscale* untuk tes dapat dilihat pada

Lampiran 11a. Adapun secara ringkas hasil estimasi parameter butir tes disajikan sebagai berikut.

Tabel 20. Hasil Estimasi Parameter Butir Tes pada Uji Coba I

Butir	a_j	b_j	d_k	b_{jk}	Keterangan
1a	1,216	-1,849	0,000	-1,849	Baik
1b	0,905	-2,179	0,183 -0,183	-2,362 -1,996	Baik
2a	1,371	0,475	0,716 -0,716	-0,241 1,191	Baik
2b	0,865	-1,086	0,171 -0,171	-1,257 -0,915	Baik
3a	2,008	-2,329	-0,059 0,059	-2,270 -2,388	Baik
3b	0,359	-1,691	-3,745 3,662 1,981 -3,214 1,316	2,054 -5,353 -3,672 1,523 -3,007	Baik
3c	0,813	-0,977	0,008 -0,008	-0,985 -0,969	Baik
4a	1,314	-0,715	-0,353 0,353	-0,362 -1,068	Baik
4b	0,310	0,010	0,575 0,545 -1,121	-0,565 -0,535 1,131	Baik

Berdasarkan Tabel 20 tampak bahwa seluruh butir tes termasuk kategori baik dengan indeks daya beda butir (a_j) secara keseluruhan terletak pada 0,310-2,008 dan indeks kesukaran butir (b_j) sebesar -2,329 sampai 0,475 pada skala logit. Hal ini sebagaimana mengacu pada kriteria butir yang baik dari Wells, Hambleton, & Purwono (2008) yakni butir termasuk kategori baik jika parameter daya beda (a) nilainya $>0,25$ pada skala logit dan parameter indeks kesukaran (b -global) nilainya

terletak pada selang -3 dan 3 pada skala logit. Adapun butir tes dengan daya beda paling tinggi adalah butir 3a dan paling rendah adalah butir 4b. Terkait dengan indeks kesukaran tiap langkah (b_{jk}), menunjukkan bahwa secara keseluruhan dapat dikatakan termasuk baik karena hampir semua langkah pada tiap butir indeks kesukarannya terletak pada selang -3 dan 3 pada skala logit kecuali butir 3a. Butir 3a merupakan butir dengan langkah terbanyak sehingga sangat memungkinkan ada langkah yang indeks kesukarannya lebih kecil dari -3. Namun ditinjau dari nilai b -global butir 3a yakni -1,691 sudah memenuhi kategori butir yang baik. Berdasarkan hasil analisis *Parscale* sebagaimana dapat dilihat pada Lampiran 11a, juga diperoleh bahwa tes memberikan informasi yang akurat pada siswa dengan estimasi kemampuan (θ) sebesar -3,0 sampai 0,8. Bahkan tes juga dapat memberikan informasi yang akurat untuk siswa dengan θ kurang dari -3,0. Tes memberikan informasi tertinggi pada siswa dengan θ sekitar -2,4. Hal ini berarti bahwa instrumen bentuk tes uraian yang dihasilkan dalam penelitian ini cocok untuk siswa dengan kemampuan rendah atau sedang.

2. Tugas

Data hasil uji coba tugas juga dianalisis dengan bantuan program Lisrel menggunakan *first order* CFA. Hasil uji persyaratan analisis yakni uji *multivariate normality* untuk tugas secara lengkap dapat dilihat pada Lampiran 9c. Hasil analisis menunjukkan bahwa data tidak berdistribusi normal multivariat. Oleh sebab itu, analisis *first order* CFA dilakukan dengan model estimasi *Generalized Least Square* (GLS). Hal ini karena data tugas bersifat kontinu dan mengacu pada Widhiarso

(2012: 1-2) bahwa pendekatan yang dapat digunakan untuk mengatasi ketidaknormalan data adalah GLS dan WLS. Secara lengkap hasil analisis *first order* CFA tugas pada uji coba I dapat dilihat pada Lampiran 9d. Adapun perhitungan indeks reliabilitas tugas untuk *Construct Reliability* (CR) dan Alpha dapat dilihat pada Lampiran 10b.

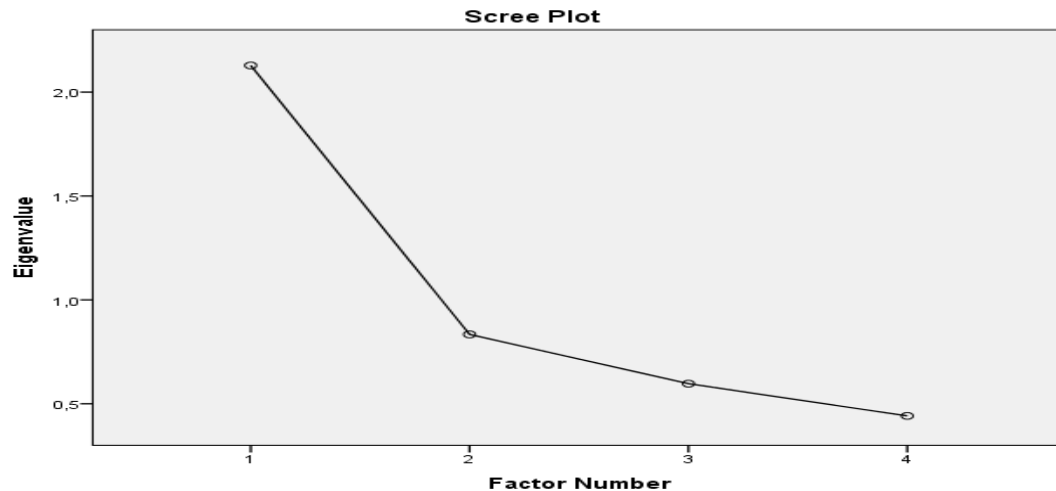
Berdasarkan hasil analisis Lisrel sebagaimana pada Lampiran 9d diperoleh nilai RMSEA sebesar 0,045, *chi Square* sebesar 4,77 dengan *p-value* sebesar 0,09, GFI sebesar 1,00, dan AGFI sebesar 0,99. Hasil ini menunjukkan bahwa kecocokan model terpenuhi dengan baik. Adapun hasil *first order* CFA untuk *t-value* dan *Standardized Loading Factor* (SLF) serta hasil perhitungan indeks CR dan Alpha disajikan sebagai berikut.

Tabel 21. Hasil *First Order* CFA Tugas pada Uji Coba I

Butir	<i>First order</i> CFA				Validitas	Reliabilitas	
	<i>t-value</i>	Keterangan	SLF	<i>Error</i>		CR	Alpha
1	14,83	Signifikan	0,64	0,59	Baik	0,82	0,70
2a	12,45	Signifikan	0,57	0,68	Baik		
2b	22,80	Signifikan	0,81	0,35	Baik		
3	23,95	Signifikan	0,89	0,20	Baik		

Berdasarkan Tabel 21 tampak bahwa *t-value* pada seluruh butir tugas signifikan dalam mendukung konstruk tugas dengan dukungan terendah oleh butir 2a dan tertinggi oleh butir 3. Nilai SLF untuk seluruh butir lebih besar dari 0,3. Hal ini berarti keempat butir memiliki validitas konstruk yang baik dalam membangun konstruk tugas. Terbuktinya validitas konstruk ini sekaligus membuktikan asumsi

unidimensi pada tugas yang secara visual dapat disajikan dalam grafik berupa *plot eigen value* yang diperoleh melalui analisis faktor dengan SPSS versi 20 berikut.



Gambar 7. Grafik *Eigen Value* untuk Tugas

Gambar 7 menunjukkan bahwa *eigen value* pada salah satu faktor bersifat dominan dibandingkan *eigen value* pada faktor lainnya. Hal ini berarti bahwa pada instrumen berupa tugas, asumsi unidimensi dapat dikatakan terpenuhi sehingga bisa dilakukan analisis menggunakan IRT politomus. Karakteristik butir tugas ditentukan menggunakan pendekatan IRT politomus GPCM dan diperoleh melalui analisis dengan bantuan program *Parscale* versi 4.1.

Berdasarkan hasil analisis diperoleh estimasi parameter butir, kemampuan atau theta, kurva karakteristik butir, kurva informasi butir dan tugas serta kesalahan baku pengukurannya. Secara lengkap hasil analisis *Parscale* untuk tugas dapat dilihat pada Lampiran 11b. Adapun secara ringkas hasil estimasi parameter butir tugas disajikan sebagai berikut.

Tabel 22. Hasil Estimasi Parameter Butir Tugas pada Uji Coba I

Butir	a_j	b_j	d_k	b_{jk}	Keterangan
1	0,615	-1,826	-0,094 0,094	-1,732 -1,920	Baik
2a	0,458	-2,012	0,886 -0,372 -0,514	-2,898 -1,640 -1,498	Baik
2b	0,956	-1,017	0,641 -0,219 -0,422	-1,658 -0,798 -0,595	Baik
3	1,113	-0,907	-0,752 0,752	-0,155 -1,659	Baik

Berdasarkan Tabel 22 tampak bahwa seluruh butir tugas termasuk kategori baik dengan indeks daya beda (a_j) secara keseluruhan terletak pada 0,458–1,113 dan indeks kesukaran (b_j) terletak pada -2,012 sampai -0,907. Hal ini mengacu pada kriteria butir yang baik menurut Wells, Hambleton, & Purwono (2008) yakni parameter daya beda (a) nilainya $>0,25$ pada skala logit dan parameter b -global nilainya terletak pada selang -3 dan 3 pada skala logit. Butir tugas yang memiliki daya beda paling tinggi adalah butir 3 dan paling rendah adalah butir 2a. Berdasarkan hasil analisis *Parscale* sebagaimana dapat dilihat pada Lampiran 11b, juga diperoleh bahwa tugas memberikan informasi yang akurat pada siswa dengan estimasi theta sebesar -3,0 sampai 0,5. Bahkan tugas juga dapat memberikan informasi yang akurat untuk siswa dengan theta kurang dari -3,0. Selain itu, tugas juga memberikan informasi tertinggi pada siswa dengan theta sekitar -1,2. Hal ini berarti bahwa instrumen bentuk tugas proyek individu yang dihasilkan dalam penelitian ini cocok untuk siswa dengan kemampuan rendah atau sedang.

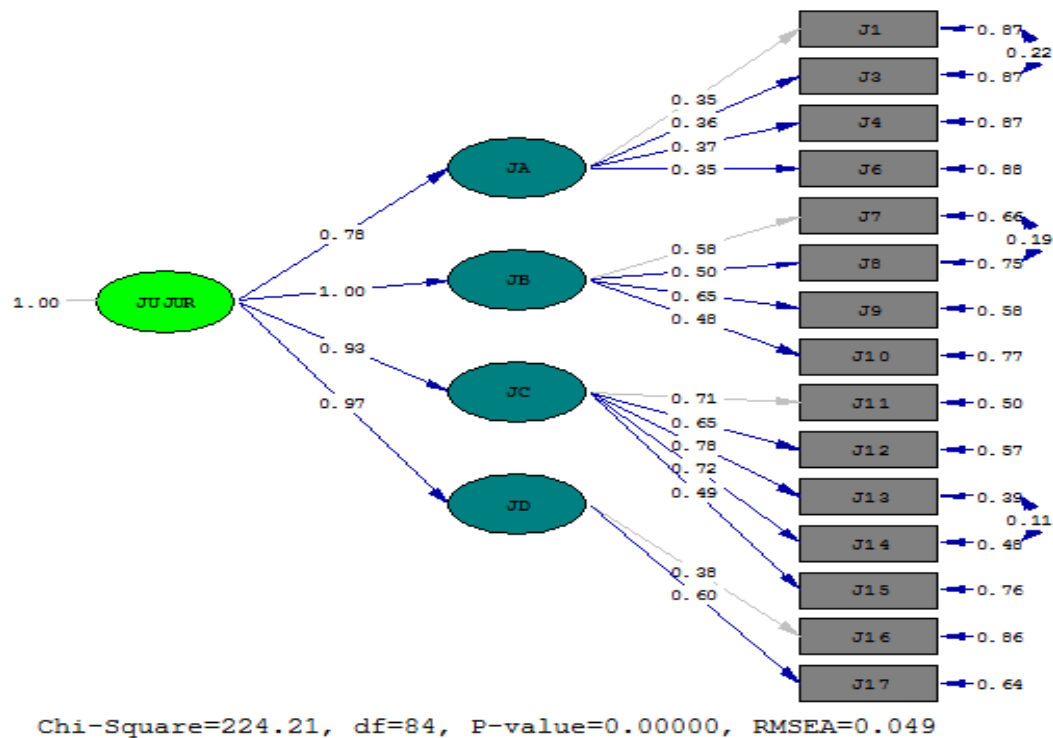
2. Uji Coba II

Berdasarkan hasil uji coba I diperoleh bahwa untuk instrumen penilaian aspek sikap ada beberapa butir yang perlu direvisi. Adapun keseluruhan butir pada instrumen penilaian aspek pengetahuan dan keterampilan dapat dikatakan termasuk butir yang baik sehingga tidak perlu ada yang direvisi dan diuji coba lagi. Oleh sebab itu, uji coba II hanya dilakukan pada instrumen penilaian aspek sikap yang telah direvisi. Kegiatan uji coba II pada penelitian ini juga dilakukan pada SMP N 5 Yogyakarta, SMP Muhammadiyah 2 Yogyakarta, dan SMPIT Abu Bakar Yogyakarta. Berdasarkan pelaksanaan uji coba II diperoleh hasil isian instrumen penilaian aspek sikap sebanyak 696 siswa.

Sebagaimana analisis pada uji coba I, analisis data uji coba II juga dilakukan dengan program Lisrel menggunakan *second order CFA*. Sebelum dilakukan analisis *second order CFA* dilakukan uji persyaratan analisis yakni uji *multivariate normality*. Secara lengkap hasil uji *multivariate normality* untuk data uji coba II dapat dilihat pada Lampiran 7a skala untuk mengukur karakter jujur, Lampiran 7d skala untuk mengukur karakter tanggung jawab, dan Lampiran 7g untuk skala sikap matematis. Hasil analisis menunjukkan bahwa data pada semua skala ternyata tidak berdistribusi normal multivariat. Oleh sebab itu, sebagaimana uji coba I, analisis *second order CFA* dilakukan dengan model estimasi *Weighted Least Square (WLS)*. Secara lengkap hasil analisis *second order CFA* berdasarkan data uji coba II dapat dilihat pada Lampiran 7c, Lampiran 7f, dan Lampiran 7i. Berikut hasil analisis data dari instrumen penilaian aspek sikap pada uji coba II.

1. Skala untuk Mengukur Karakter Jujur

Berdasarkan hasil analisis awal pada data Uji Coba II skala untuk mengukur karakter jujur, ada dua butir yang memiliki nilai SLF kurang dari 0,3 yaitu butir J2 dan J5. Apabila kedua butir ini dieliminasi tidak berpengaruh pada indikator. Artinya masih ada butir lain yang mewakili indikator yang diwakili butir J2 dan J5. Oleh sebab itu, analisis lanjut dilakukan setelah mengeliminasi butir J2 dan J5. Analisis lanjut juga memperhatikan modifikasi indeks yang disarankan Lisrel. Peneliti memanfaatkan saran tersebut guna meningkatkan kecocokan model. Berikut model pengukuran skala untuk mengukur karakter jujur menggunakan modifikasi sebagaimana disarankan Lisrel.



Gambar 8. Model Pengukuran Skala untuk Mengukur Karakter Jujur pada Uji Coba II

Sebagaimana disajikan pada Gambar 8, modifikasi yang dilakukan peneliti adalah dengan menambahkan kovariansi di antara dua kesalahan (*error*) pada J1 dan J2, J7 dan J8, J13 dan J14. Pertimbangan peneliti adalah karena masing-masing mewakili variabel laten yang sama dan menurunkan nilai *chi-square* serta dapat didukung oleh alasan yang kuat.

Indikator J1 adalah menyampaikan informasi apa adanya tanpa menambah atau mengurangi dan J2 adalah mengatakan kebenaran apapun konsekuensinya. Indikator J7 adalah mengakui kesalahan yang dilakukan dan J8 adalah memberi penghargaan pada kelebihan yang dimiliki diri sendiri. Adapun indikator J13 adalah tidak memanfaatkan kesempatan yang ada untuk menyontek dan J14 adalah tidak saling tukar hasil pekerjaan dengan teman. Memperhatikan indikator-indikator tersebut, *error* yang terjadi pada J1 jelas sangat berpengaruh pada indikator J2. Demikian pula untuk J7 dan J8 serta J13 dan J14. Hal ini menunjukkan bahwa cukup beralasan untuk menambahkan kovariansi di antara dua *error* pada J1 dan J2, J7 dan J8, serta J13 dan J14.

Berdasarkan hasil analisis akhir setelah dilakukan modifikasi sebagaimana dapat dilihat pada Lampiran 7c, diperoleh nilai RMSEA sebesar 0,049, *chi square* sebesar 221,21 dengan *p-value* sebesar 0,00, GFI sebesar 0,98, dan AGFI sebesar 0,97. Walaupun *chi square* tidak signifikan karena *p-value* kurang dari 0,05, namun kriteria lain telah terpenuhi. Oleh sebab itu, mengacu pada Hooper, Coughlan, & Mullen (2008: 57) dan Hair et al. (2009: 645), kecocokan model dapat dianggap telah terpenuhi.

Hasil *second order* CFA untuk *t-value* dan *Standardized Loading Factor* (SLF)

skala untuk mengukur karakter jujur siswa secara ringkas disajikan sebagai berikut.

Tabel 23. Hasil *Second Order* CFA Skala untuk Mengukur Karakter Jujur pada Uji Coba II

Faktor	Indikator	Butir	<i>Second order</i> CFA			Validitas
			<i>t-value</i>	Keterangan	SLF	
A. Berperilaku apa adanya sesuai kenyataan.	A1	J1	**	Signifikan	0,35	Baik
	A3	J3	5,60	Signifikan	0,36	Baik
	A4	J4	4,64	Signifikan	0,37	Baik
	A6	J6	4,11	Signifikan	0,35	Baik
B. Mengakui setiap perbuatan baik positif maupun negatif.	B1	J7	**	Signifikan	0,58	Baik
	B2	J8	10,26	Signifikan	0,50	Baik
	B3	J9	12,05	Signifikan	0,65	Baik
	B4	J10	8,56	Signifikan	0,48	Baik
C. Mengerjakan tugas atau ujian sesuai kemampuan sendiri.	C1	J11	**	Signifikan	0,71	Baik
	C2	J12	15,73	Signifikan	0,65	Baik
	C3	J13	15,58	Signifikan	0,78	Baik
	C4	J14	16,11	Signifikan	0,72	Baik
	C5	J15	11,36	Signifikan	0,49	Baik
D. Dapat dipercaya atas kepemilikan orang lain.	D1	J16	**	Signifikan	0,38	Baik
	D2	J17	6,42	Signifikan	0,60	Baik

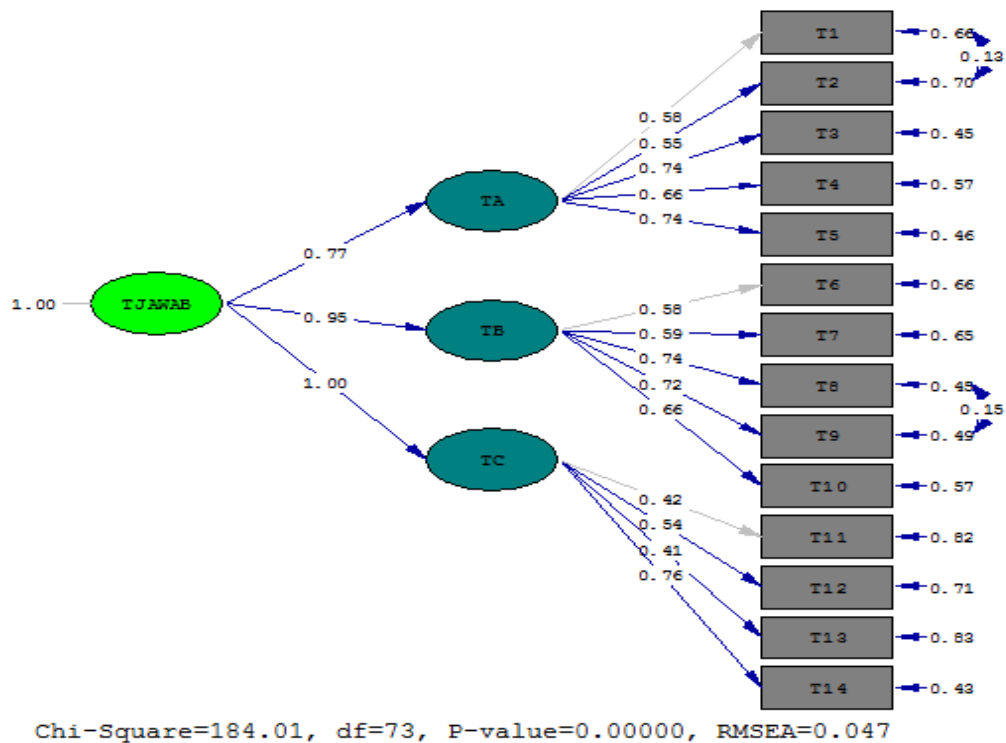
Tabel 23 menunjukkan bahwa ditinjau dari *t-value*, seluruh butir signifikan dalam mendukung konstruk jujur. Apabila ditinjau dari nilai SLF juga menunjukkan bahwa semua butir memenuhi kriteria validitas konstruk yang baik.

Adapun perhitungan koefisien reliabilitas skala untuk mengukur karakter jujur dapat dilihat pada Lampiran 8a. Koefisien reliabilitas yang digunakan adalah koefisien reliabilitas Omega karena konstruk jujur bersifat multidimensi. Berdasarkan hasil perhitungan diperoleh koefisien reliabilitas Omega sebesar 0,86. Hasil ini menunjukkan bahwa skala untuk mengukur karakter jujur memiliki reliabilitas yang baik sebagaimana mengacu pada kriteria dari Hair et al. (2009: 688) yakni koefisien

reliabilitas >0,7 tergolong baik. Hal ini berarti bahwa ditinjau dari konstruksya, skala untuk mengukur karakter jujur siswa dapat dikatakan bersifat reliabel atau konsisten.

2. Skala untuk Mengukur Karakter Tanggung Jawab

Berdasarkan hasil analisis awal pada data uji coba II skala untuk mengukur tanggung jawab, tidak ada butir yang memiliki nilai SLF kurang dari 0,3. Hal ini berarti tidak ada butir yang dieliminasi. Analisis lanjut dilakukan pada seluruh butir dengan memperhatikan modifikasi indeks yang disarankan Lisrel guna meningkatkan kecocokan model. Berikut model pengukuran skala untuk mengukur karakter tanggung jawab dengan modifikasi model sebagaimana disarankan Lisrel.



Gambar 9. Model Pengukuran Skala untuk Mengukur Karakter Tanggung Jawab pada Uji Coba II

Sebagaimana disajikan pada Gambar 9, modifikasi yang dilakukan peneliti adalah dengan menambahkan kovariansi di antara dua kesalahan (*error*) pada T1 dan T2, T8 dan T9. Pertimbangan peneliti adalah karena masing-masing mewakili variabel laten yang sama dan menurunkan nilai *chi-square*. Selain itu, penambahan kovariansi di antara dua kesalahan (*error*) pada T1 dan T2 serta T8 dan T9 cukup beralasan atau dapat didukung oleh alasan yang kuat.

Indikator T1 adalah melaksanakan dengan sungguh-sungguh setiap tugas yang diberikan guru dan T2 adalah senantiasa mengkondisikan diri untuk siap mengerjakan tugas dari guru. *Error* yang terjadi pada T1 jelas sangat berpengaruh pada indikator T2. Adapun indikator T8 adalah bersedia meminta maaf atas kesalahan yang dilakukan dan T9 adalah bersedia mengundurkan diri karena gagal dalam melaksanakan tugas, jika hal itu merupakan jalan keluar yang terbaik bagi kepentingan umum. *Error* yang terjadi pada T8 ini juga jelas sangat berpengaruh pada indikator T9. Hal ini menunjukkan bahwa cukup beralasan untuk menambahkan kovariansi di antara dua *error* pada T1 dan T2 serta T8 dan T9 tersebut.

Berdasarkan hasil analisis akhir setelah dilakukan modifikasi sebagaimana dapat dilihat pada Lampiran 7f, diperoleh nilai RMSEA sebesar 0,047, *chi square* sebesar 184,01 dengan *p-value* sebesar 0,00, GFI sebesar 0,98, dan AGFI sebesar 0,98. Walaupun *chi square* tidak signifikan karena *p-value* kurang dari 0,05, namun kriteria lain telah terpenuhi. Oleh sebab itu, mengacu pada Hooper, Coughlan, & Mullen (2008: 57) dan Hair et al. (2009: 645), kecocokan model dapat dianggap telah terpenuhi.

Hasil *second order* CFA untuk *t-value* dan *Standardized Loading Factor* (SLF)

secara ringkas disajikan sebagai berikut.

Tabel 24. Hasil *Second Order* CFA Skala untuk Mengukur Karakter Tanggung Jawab pada Uji Coba II

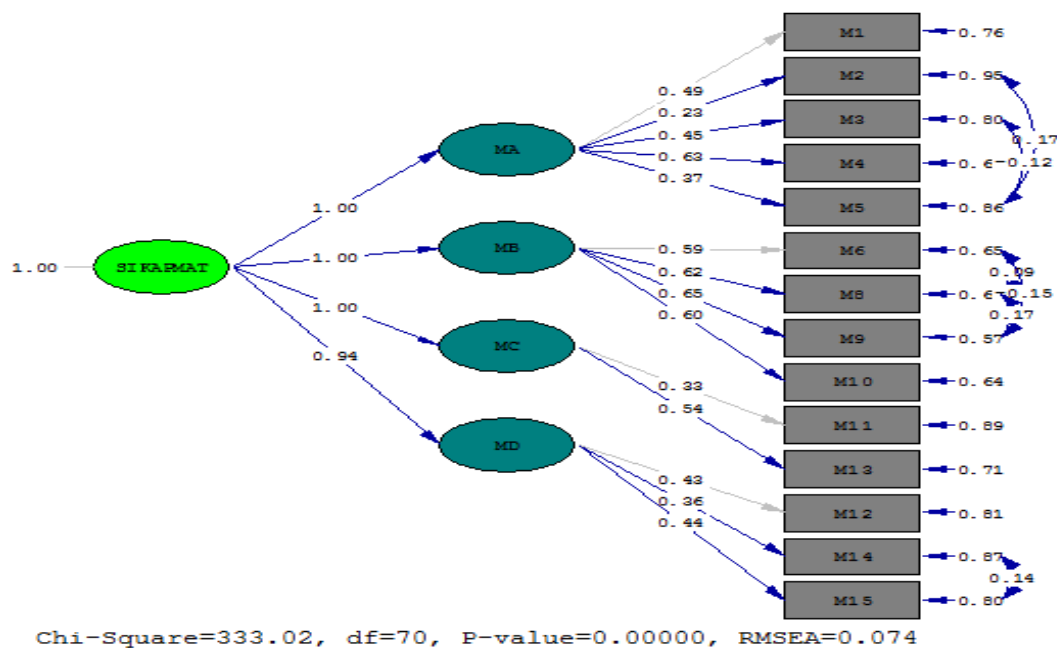
Faktor	Indikator	Butir	<i>Second order</i> CFA			Validitas
			SLF	<i>t-value</i>	Keterangan	
A. Komitmen pada Tugas	A1	T1	0,58	**	Signifikan	Baik
	A2	T2	0,55	12,13	Signifikan	Baik
	A3	T3	0,74	13,27	Signifikan	Baik
	A4	T4	0,66	12,30	Signifikan	Baik
	A5	T5	0,74	12,60	Signifikan	Baik
B. Sportif	B1	T6	0,58	**	Signifikan	Baik
	B2	T7	0,59	9,77	Signifikan	Baik
	B3	T8	0,74	12,25	Signifikan	Baik
	B4	T9	0,72	12,49	Signifikan	Baik
	B5	T10	0,66	11,93	Signifikan	Baik
C. Disiplin	C1	T11	0,42	**	Signifikan	Baik
	C2	T12	0,54	7,99	Signifikan	Baik
	C3	T13	0,41	6,98	Signifikan	Baik
	C4	T14	0,76	8,35	Signifikan	Baik

Tabel 24 menunjukkan bahwa ditinjau dari *t-value*, seluruh butir pada skala untuk mengukur karakter tanggung jawab signifikan dalam mendukung konstruk tanggung jawab. Ditinjau dari nilai SLF, tidak ada butir yang memiliki nilai SLF yang kurang dari 0,3. Hal ini berarti semua butir pada skala untuk mengukur karakter tanggung jawab memenuhi kriteria validitas konstruk yang baik. Adapun perhitungan koefisien reliabilitas skala untuk mengukur karakter tanggung jawab dapat dilihat pada Lampiran 8b. Koefisien reliabilitas yang digunakan adalah koefisien reliabilitas Omega karena konstruk tanggung jawab bersifat multidimensi. Berdasarkan hasil perhitungan diperoleh koefisien reliabilitas Omega sebesar 0,89. Hasil ini

menunjukkan bahwa skala untuk mengukur karakter tanggung jawab memiliki reliabilitas yang baik. Hal ini berarti bahwa ditinjau dari konstruksya, skala untuk mengukur tanggung jawab siswa dapat dikatakan bersifat reliabel atau konsisten.

3. Skala Sikap Matematis

Berdasarkan hasil analisis awal Lisrel pada data skala sikap matematis pada uji coba II diperoleh bahwa tidak ada butir yang memiliki nilai SLF kurang dari 0,3. Hal ini berarti tidak ada butir yang dieliminasi. Oleh sebab itu, analisis lanjut dilakukan pada seluruh butir dengan memperhatikan modifikasi indeks yang disarankan Lisrel. Peneliti memanfaatkan saran tersebut guna meningkatkan kecocokan model. Modifikasi model yang dilakukan peneliti sebagaimana disarankan dari software Lisrel dapat dilihat pada Gambar 10 berikut.



Gambar 10. Model Pengukuran Skala Sikap Matematis pada Uji Coba II

Sebagaimana disajikan pada Gambar 10, modifikasi yang dilakukan peneliti adalah dengan menambahkan kovariansi di antara dua kesalahan (*error*) pada M2 dan M5, M3 dan M5, M6 dan M8 serta M14 dan M15. Pertimbangan peneliti adalah karena masing-masing mewakili variabel laten yang sama dan menurunkan nilai *chi square*. Selain itu, penambahan kovariansi di antara dua kesalahan (*error*) pada M2 dan M5, M3 dan M5, M6 dan M8 serta M14 dan M15 cukup beralasan atau dapat didukung oleh alasan yang kuat.

Indikator M2 adalah berusaha untuk bertanya dan M5 adalah berusaha menemukan masalah matematika dalam kehidupan sehari-hari. Indikator M3 adalah berusaha untuk memahami persoalan. *Error* yang terjadi pada M2 sangat berpengaruh pada indikator M5. Demikian pula *error* yang terjadi pada M3 tentu sangat berpengaruh pada M5. Adapun indikator M6 adalah berusaha mencapai kompetensi matematika yang ditetapkan dan M8 adalah berusaha memahami sifat-sifat matematika. Indikator M14 adalah berusaha berfikir secara objektif dan berpikir kritis. Indikator M15 adalah berusaha memanfaatkan pikiran dan hasil usaha yang telah didapat. *Error* yang terjadi pada M6 sangat berpengaruh pada indikator M8. Demikian pula *error* yang terjadi pada M14 tentu juga sangat berpengaruh pada M15. Hal ini menunjukkan bahwa cukup beralasan untuk menambahkan kovariansi di antara dua *error* pada M2 dan M5, M3 dan M5, M6 dan M8 serta M14 dan M15 tersebut.

Berdasarkan hasil analisis akhir setelah dilakukan modifikasi sebagaimana dapat dilihat pada Lampiran 7i, diperoleh nilai RMSEA sebesar 0,074, *chi square*

sebesar 333,02 dengan *p-value* sebesar 0,00, GFI sebesar 0,97, dan AGFI sebesar 0,95. Walaupun *Chi square* tidak signifikan karena *p-value* kurang dari 0,05, namun kriteria lain telah terpenuhi. Oleh sebab itu, mengacu pada Hooper, Coughlan, & Mullen (2008: 57) dan Hair et al. (2009: 645), kecocokan model dapat dianggap telah terpenuhi. Adapun hasil *second order CFA* untuk *t-value* dan *Standardized Loading Factor* (SLF) disajikan sebagai berikut.

Tabel 25. Hasil *Second Order CFA* Skala Sikap Matematis pada Uji Coba II

Faktor	Indikator	Butir	<i>Second order CFA</i>			Validitas
			SLF	<i>t-value</i>	Keterangan	
A. Mencoba memahami masalah atau substansi masalah dengan jelas secara mandiri.	A1	M1	0,49	**	Signifikan	Baik Kurang Baik
		M2	0,23	5,32	Signifikan	
	A2	M3	0,45	9,87	Signifikan	Baik Baik
		M5	0,63	12,6	Signifikan	
	A3	M4	0,37	8,16	Signifikan	Baik
	B. Mencoba mengambil tindakan logis.	B1	M6	0,59	**	Signifikan
M7			-	-	-	
B2		M8	0,62	15,67	Signifikan	Baik Baik
		M9	0,65	14,31	Signifikan	
B3		M10	0,60	12,71	Signifikan	Baik
C. Mencoba mengekspresikan berbagai hal dengan jelas dan ringkas.		C1	M11	0,33	**	Signifikan
	C2	M13	0,54	6,97	Signifikan	Baik
D. Mencoba mencari penyelesaian yang lebih baik.	D1	M12	0,43	**	Signifikan	Baik
	D2	M14	0,36	7,13	Signifikan	Baik
	D3	M15	0,44	7,72	Signifikan	Baik

Tabel 25 menunjukkan bahwa ditinjau dari *t-value*, seluruh butir signifikan dalam mendukung konstruk sikap matematis. Adapun ditinjau dari nilai SLF, ada satu butir yakni M2 yang memiliki nilai SLF kurang dari 0,3 sehingga validitasnya kurang baik

dalam mendukung konstruk sikap matematis. Apabila butir ini dieliminasi ternyata tidak berpengaruh pada indikator yang diwakilinya. Artinya masih ada butir lain yang mewakili indikator yakni butir M1. Oleh sebab itu tidak perlu dilakukan revisi butir dan uji coba kembali.

Adapun perhitungan koefisien reliabilitas skala sikap matematis dapat dilihat pada Lampiran 8c. Koefisien reliabilitas yang digunakan adalah koefisien reliabilitas Omega karena konstruk sikap matematis juga bersifat multidimensi. Berdasarkan hasil perhitungan diperoleh koefisien reliabilitas Omega sebesar 0,81. Hasil ini menunjukkan bahwa skala sikap matematis memiliki reliabilitas yang baik. Hal ini berarti bahwa ditinjau dari konstruksinya, skala sikap matematis dapat dikatakan bersifat reliabel atau konsisten.

3. Pencapaian Kompetensi Siswa

Pencapaian kompetensi siswa pada aspek sikap, pengetahuan, dan keterampilan dalam penelitian ini diperoleh dari penerapan instrumen yang secara empiris terbukti berkualitas baik. Instrumen penilaian aspek sikap yang secara empiris terbukti berkualitas baik diperoleh pada uji coba II. Instrumen penilaian pengetahuan dan keterampilan yang secara empiris berkualitas baik diperoleh pada uji coba I. Data dari penerapan instrumen yang secara empiris terbukti berkualitas baik tersebut dianalisis secara deskriptif dan ditindaklanjuti dengan penafsiran skor menggunakan cara penafsiran yang ditetapkan. Secara lengkap hasil analisis deskriptif pencapaian kompetensi siswa dapat dilihat pada Lampiran 12. Adapun secara ringkas disajikan sebagai berikut.

a. Pencapaian Kompetensi Aspek Sikap

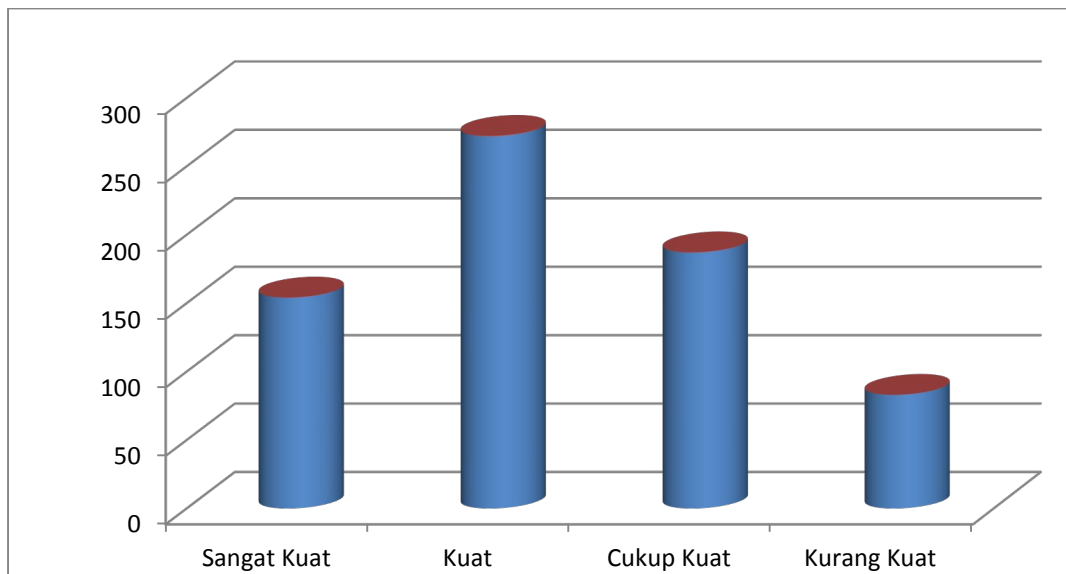
Pencapaian kompetensi aspek sikap dalam penelitian ini meliputi karakter jujur, karakter tanggung jawab, dan sikap matematis. Penafsiran skor untuk karakter jujur dikategorikan dalam empat kategori yakni sangat kuat, kuat, cukup kuat, dan kurang kuat. Kategori ini mencerminkan tingkat komitmen atau tekad dalam diri siswa untuk berbuat atau bertindak jujur. Penafsiran skor untuk karakter tanggung jawab juga dikategorikan dalam empat kategori yakni sangat kuat, kuat, cukup kuat, dan kurang kuat. Kategori ini mencerminkan tingkat komitmen atau tekad dalam diri siswa untuk memiliki sikap tanggung jawab dalam dirinya. Adapun untuk sikap matematis dikategorikan dalam empat kategori yakni sangat baik, baik, cukup baik, dan kurang baik. Kategori ini menunjukkan tingkat sikap sistematis dalam diri siswa pada pembelajaran matematika. Secara keseluruhan, hasil pencapaian kompetensi siswa dalam bentuk kategori untuk karakter jujur, karakter tanggung jawab, dan sikap matematis siswa dapat dilihat pada Lampiran 12a.

Secara ringkas, pencapaian kompetensi siswa untuk karakter jujur disajikan sebagai berikut.

Tabel 26. Pencapaian Kompetensi Siswa pada Karakter Jujur

Kategori	Banyaknya Siswa	Persentase (%)
Sangat Kuat	154	22,13
Kuat	272	39,08
Cukup Kuat	187	26,87
Kurang Kuat	83	11,93

Berdasarkan Tabel 26 menunjukkan bahwa tingkat komitmen atau tekad para siswa dalam dirinya untuk berbuat atau bertindak jujur adalah 22,13% sangat kuat, 39,08% kuat, 26,87% cukup kuat, dan 11,93% kurang kuat. Adapun secara visual, pencapaian kompetensi siswa untuk karakter jujur disajikan sebagai berikut.



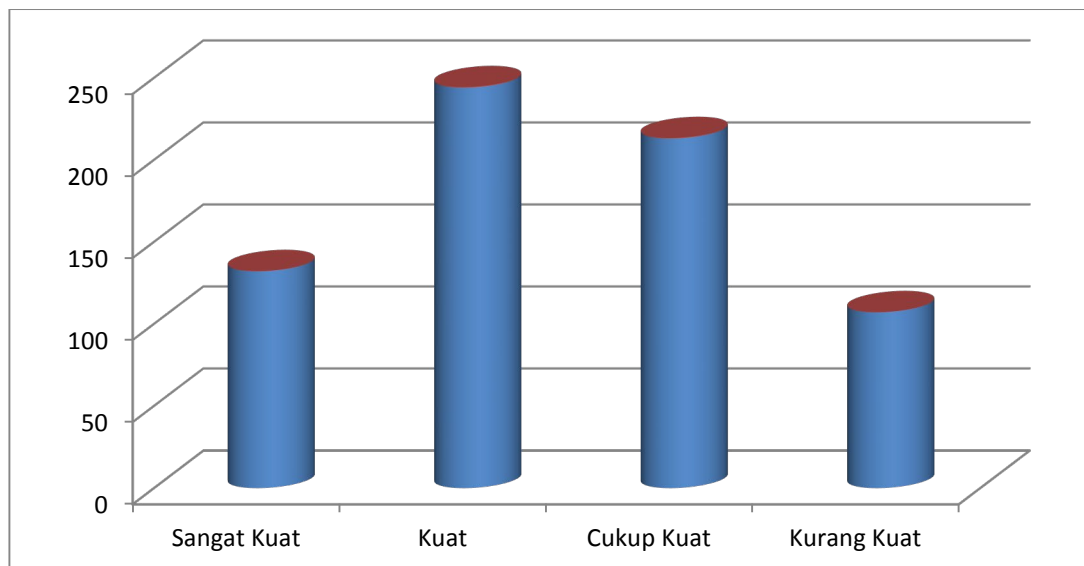
Gambar 11. Pencapaian Kompetensi Siswa pada Karakter Jujur

Pencapaian kompetensi siswa untuk karakter tanggung jawab secara ringkas disajikan sebagai berikut.

Tabel 27. Pencapaian Kompetensi Siswa pada Karakter Tanggung Jawab

Kategori	Banyaknya Siswa	Persentase (%)
Sangat Kuat	132	18,97
Kuat	244	35,06
Cukup Kuat	213	30,60
Kurang Kuat	107	15,37

Berdasarkan Tabel 27 menunjukkan bahwa tingkat komitmen atau tekad para siswa dalam dirinya untuk berbuat atau bertindak dengan penuh tanggung jawab adalah 18,97% sangat kuat, 35,06% kuat, 30,60% cukup kuat, dan 15,37% kurang kuat. Adapun secara visual, pencapaian kompetensi siswa untuk karakter tanggung jawab disajikan sebagai berikut.



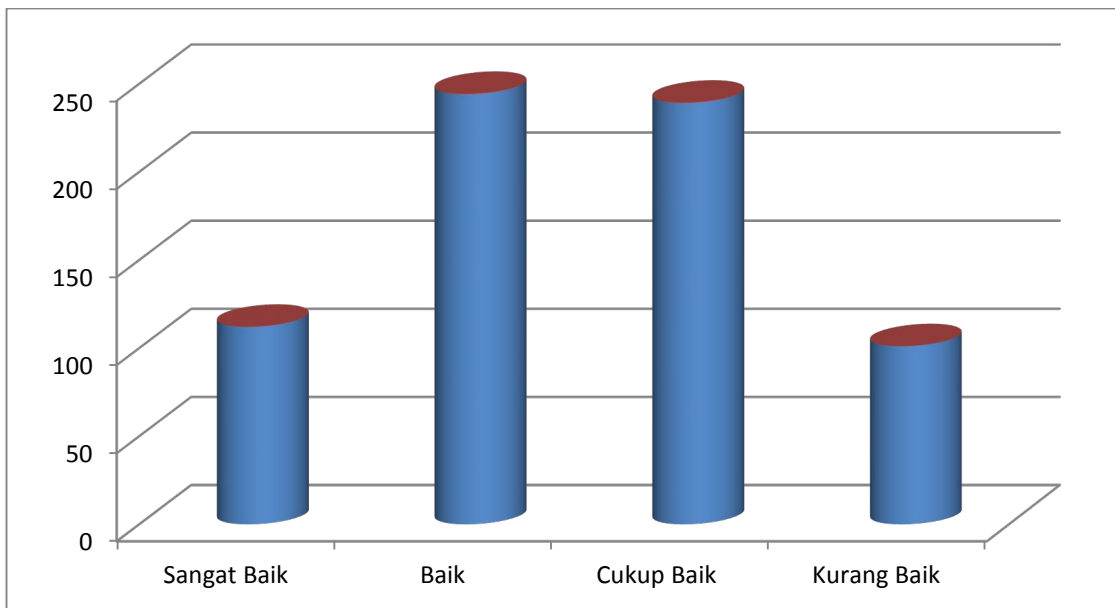
Gambar 12. Pencapaian Kompetensi Siswa pada Karakter Tanggung Jawab

Pencapaian kompetensi siswa untuk sikap matematis secara ringkas disajikan sebagai berikut.

Tabel 28. Pencapaian Kompetensi Siswa pada Sikap Matematis

Kategori	Banyaknya Siswa	Persentase (%)
Sangat Baik	112	16,09
Baik	244	35,06
Cukup Baik	239	34,34
Kurang Baik	101	14,51

Berdasarkan Tabel 28 menunjukkan bahwa sikap matematis siswa dalam pembelajaran matematika adalah 16,09% sangat baik, 35,06% baik, 34,34% cukup baik, dan 14,51% kurang baik. Adapun secara visual, pencapaian kompetensi siswa untuk sikap matematis disajikan sebagai berikut.

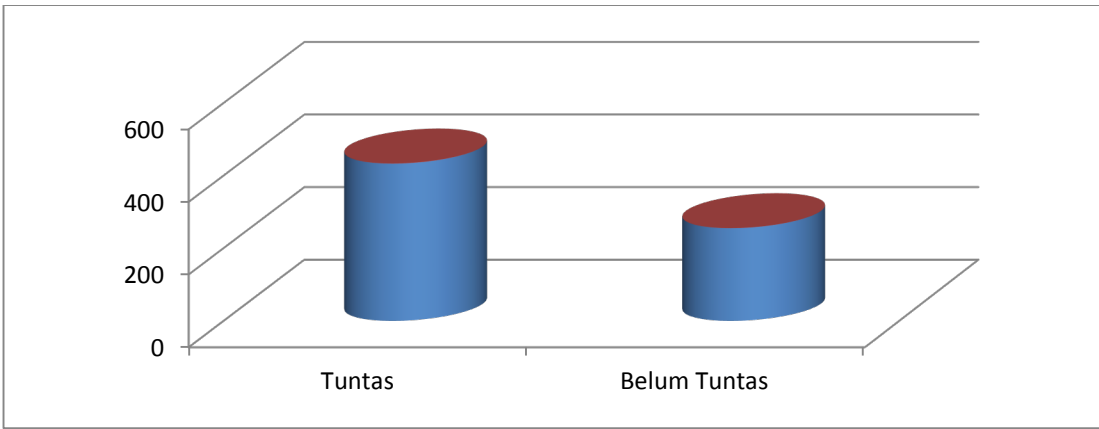


Gambar 13. Pencapaian Kompetensi Siswa pada Sikap Matematis

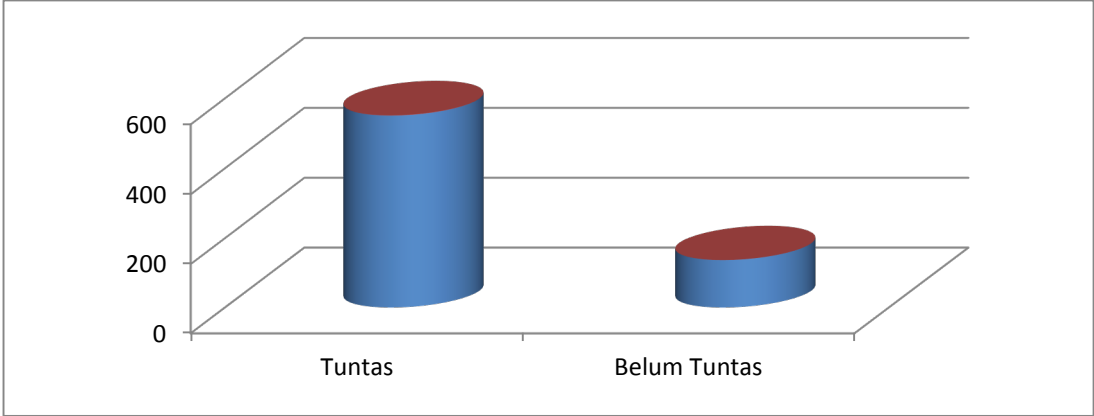
b) Pencapaian Kompetensi Aspek Pengetahuan dan Keterampilan

Pencapaian kompetensi siswa pada aspek pengetahuan diperoleh berdasarkan skor tes dan aspek keterampilan berdasarkan skor tugas. Secara lengkap hasil analisis deskriptif untuk pencapaian kompetensi aspek pengetahuan dan keterampilan dapat dilihat pada Lampiran 12b. Berdasarkan hasil analisis, untuk pencapaian kompetensi

pengetahuan diperoleh rata-rata 79,09 dan simpangan baku 14,80. Menggunakan KKM sebesar 75 diperoleh sebanyak 433 siswa atau 62,94% yang tuntas atau mampu mencapai KKM tersebut. Adapun berdasarkan hasil analisis untuk pencapaian kompetensi keterampilan diperoleh rata-rata 86,64 dan simpangan baku 16,30. Menggunakan KKM sebesar 75 diperoleh sebanyak 551 siswa atau 80,09% yang mampu mencapai KKM tersebut. Secara visual digambarkan sebagai berikut.



Gambar 14. Pencapaian Kompetensi Pengetahuan



Gambar 15. Pencapaian Kompetensi Keterampilan

E. Revisi Produk

Berdasarkan hasil analisis uji coba I, diperoleh bahwa instrumen penilaian aspek pengetahuan dan keterampilan secara keseluruhan telah dapat dikatakan memiliki kualitas yang baik sehingga tidak perlu dilakukan perbaikan atau revisi. Namun untuk instrumen penilaian aspek sikap perlu dilakukan perbaikan atau revisi pada beberapa butir yakni butir J2 dan J5 pada skala untuk mengukur karakter jujur, butir T9 pada skala untuk mengukur karakter tanggung jawab, dan butir M14 pada skala sikap matematis.

Hasil perbaikan instrumen penilaian aspek sikap selanjutnya diujicobakan kembali pada kegiatan uji coba II. Berdasarkan hasil uji coba II diperoleh bahwa instrumen penilaian aspek sikap yang berupa skala untuk mengukur karakter tanggung jawab pada keseluruhan butir terbukti secara empiris memiliki kualitas yang baik. Namun pada skala untuk mengukur karakter jujur perlu mengeliminasi butir J2 dan J5 serta untuk skala sikap matematis perlu mengeliminasi satu butir yakni M2. Eliminasi terhadap butir J2 dan J5 tidak mempengaruhi indikator yang membangun konstruk jujur. Artinya, semua indikator yang telah ditetapkan semuanya masih terwakili butir yang berkualitas baik. Demikian juga untuk eliminasi butir M2 tidak mempengaruhi indikator yang membangun konstruk sikap matematis. Oleh sebab itu tidak perlu dilakukan revisi dan uji coba kembali.

Produk akhir hasil pengembangan setelah dilakukan revisi adalah: (a) skala untuk mengukur karakter jujur yang terdiri atas 15 butir berbentuk penskalaan subjek, (b) skala untuk mengukur karakter tanggung jawab yang terdiri atas 14 butir

berbentuk penskalaan subjek, (c) skala sikap matematis yang terdiri atas 14 butir berbentuk penskalaan respons, (d) tes uraian, dan (e) tugas proyek individu. Secara lengkap produk akhir berupa instrumen penilaian aspek sikap, pengetahuan, dan keterampilan yang dihasilkan dalam penelitian dapat dilihat pada Lampiran Produk Disertasi yang disajikan secara terpisah.

F. Kajian Produk Akhir

Produk akhir yang dihasilkan dari penelitian ini adalah instrumen penilaian pencapaian kompetensi matematika siswa SMP pada materi statistika berdasarkan Kurikulum 2013 untuk aspek sikap, pengetahuan, dan keterampilan yang terbukti secara teoretis dan empiris berkualitas baik. Produk akhir yang dihasilkan meliputi: (a) instrumen penilaian aspek sikap berupa lembar penilaian diri yang terdiri atas skala untuk mengukur karakter jujur siswa, skala untuk mengukur karakter tanggung jawab siswa, dan skala sikap matematis, (b) instrumen penilaian aspek pengetahuan dan keterampilan berupa tes uraian materi statistika dan tugas proyek individu materi statistika.

Kualitas keseluruhan instrumen penilaian aspek sikap yang dihasilkan, ditinjau dari validitas isinya adalah baik dan seluruh butir memiliki kesesuaian yang baik dengan indikatornya. Hal ini ditunjukkan dengan besarnya indeks V Aiken sebesar 0,88-0,92. Ditinjau dari validitas konstruksinya, seluruh butir instrumen memiliki dukungan yang baik dalam membangun konstruk yang telah ditetapkan.

Adapun ditinjau dari reliabilitas konstruksya, ketiga skala yang dikembangkan juga termasuk kategori baik. Hal ini ditunjukkan dengan besarnya koefisien reliabilitas Omega pada skala untuk mengukur karakter jujur, skala untuk mengukur karakter tanggung jawab, dan skala sikap matematis berturut-turut adalah sebesar 0,85, 0,89, dan 0,81.

Mengacu pada produk akhir yang telah dihasilkan dengan dukungan bukti kualitas sebagaimana disebutkan di atas, tentunya sangat membantu menambah referensi bagi para guru terkait contoh instrumen penilaian yang dapat digunakan dalam melakukan penilaian berdasarkan Kurikulum 2013. Selain dalam hal penilaian, bentuk instrumen penilaian aspek sikap yang berupa skala baik berbentuk penskalaan subjek maupun penskalaan respons yang telah terbukti secara teoretis dan empiris berkualitas baik dapat digunakan pula oleh para guru sebagai salah satu referensi untuk melakukan penelitian seperti penelitian tindakan kelas atau eksperimen sebagai salah satu hal yang harus dipenuhi untuk meningkatkan kompetensi pedagogiknya. Adapun bagi peneliti lain dapat menjadi referensi tentang pengembangan instrumen penilaian khususnya terkait instrumen yang berbentuk penskalaan subjek karena masih sangat jarang dikembangkan khususnya dalam pembelajaran matematika.

Instrumen penilaian aspek sikap berupa skala untuk mengukur karakter jujur dan skala untuk mengukur karakter tanggung jawab yang dihasilkan dalam penelitian ini menggunakan model penskalaan subjek. Penulisan butir-butir instrumen dengan model penskalaan subjek memang tidaklah mudah. Namun model penskalaan subjek mampu mengungkap kondisi psikologis siswa secara lebih mendalam. Hal ini

sebagaimana dikemukakan Azwar (2005: 29-32) yakni penskalaan subjek bertujuan meletakkan individu-individu pada suatu kontinum penilaian sehingga kedudukan relatif individu menurut suatu atribut yang diukur dapat diperoleh. Model penskalaan subjek memiliki nilai praktis yang tinggi dan banyak digunakan para perancang skala psikologi.

Butir dalam bentuk penskalaan subjek yang diberikan kepada siswa melalui teknik penilaian diri menjadikan siswa lebih memahami tentang kondisi psikologis yang hendak diketahui dari dalam dirinya. Hal ini mengingat bahwa pada bentuk penskalaan subjek, soal dan pilihan jawaban disajikan dalam bentuk pernyataan atau bersifat naratif. Sebagai contoh, butir pada skala untuk mengukur karakter jujur yang memberi dukungan paling besar dalam membangun konstruk jujur yang dikembangkan dalam penelitian ini yaitu butir J13 yang mewakili indikator C3 dengan nilai SLF sebesar 0,78. Butir J13 adalah sebagai berikut.

Ketika sedang berlangsung ujian akhir mata pelajaran matematika, tiba-tiba pengawas keluar ruangan dan ternyata agak lama belum juga kembali ke kelas, maka kamu:

- a. *Memanfaatkan kesempatan yang ada tersebut untuk bertanya kepada teman sekelas mencocokkan jawaban agar lebih yakin.*
- b. *Tetap konsentrasi mengerjakan ujian walaupun teman-teman mulai berisik karena saling berusaha bertanya dengan teman lainnya.*
- c. *Melihat-lihat teman lain yang saling bertanya dan ikut menyimaknya.*

Butir J13 mewakili indikator C3 yakni tidak memanfaatkan kesempatan yang ada untuk menyontek. Adapun pedoman penskorannya adalah skor 0 untuk pilihan a, skor 2 untuk pilihan b, dan skor 1 untuk pilihan c. Selain butir J13, kelebihan bentuk penskalaan subjek juga dapat dilihat pada butir-butir lainnya, misalnya butir yang

memberikan dukungan terbesar pada skala untuk mengukur karakter tanggung jawab yaitu butir T14 yang mewakili indikator C4. Butir tersebut adalah sebagai berikut.

Ketika pelajaran matematika yang membahas tentang diagram lingkaran, kamu lupa membawa jangka. Kamu melihat temanmu membawa dua jangka yang sangat bagus dan tampak masih baru. Kamu diijinkan temanmu meminjam salah satu jangka tersebut. Ketika pelajaran matematika telah selesai, maka kamu:

- a. *Tetap meminjam jangka tersebut karena kamu merasa masih membutuhkan dan akan mengembalikan apabila diminta.*
- b. *Tidak segera mengembalikan jangka tersebut dan berharap agar jangka tersebut dikasihkan kepadamu karena temanmu punya dua jangka.*
- c. *Segera mengembalikan jangka tersebut dan berterima kasih karena telah dibantu.*

Butir T14 mewakili indikator C4 yakni tepat waktu dalam mengembalikan barang yang dipinjam dari orang lain. Adapun pedoman penskorannya adalah skor 1 untuk pilihan a, skor 0 untuk pilihan b, dan skor 2 untuk pilihan c. Melalui butir-butir yang disajikan seperti butir J13 dan T14 tersebut, siswa diharapkan memberikan jawaban yang benar-benar menunjukkan kondisi psikologis terkait karakter siswa yang sesungguhnya ketika menghadapi situasi tertentu sebagaimana tercantum pada butir-butir tersebut.

Penilaian aspek sikap dalam penelitian ini menggunakan teknik penilaian diri yang memberikan penekanan kepada siswa bahwa penilaian tidak berpengaruh pada nilai siswa melainkan untuk meningkatkan kualitas pembelajaran menjadikan siswa bersungguh-sungguh membaca dan mengisi instrumen yang diberikan. Hal ini dilakukan berdasarkan tujuan penilaian diri menurut Shepard (Brennan, 2006: 632) yang mengemukakan bahwa penilaian diri bukan untuk memberi nilai angka, namun

untuk memperoleh pengetahuan yang bisa digunakan dalam pembelajaran selanjutnya. Penggunaan teknik penilaian diri dalam penelitian ini dilakukan dengan hati-hati terutama terkait pelaksanaannya agar diperoleh hasil yang benar-benar objektif. Hal ini di antaranya dilakukan dengan memberikan pemahaman mendalam kepada siswa mengenai hakikat dan makna dari hasil penilaian diri yang dilakukan. Selain itu, pengawasan yang lebih juga diterapkan ketika siswa melakukan penilaian diri dengan mengisi skala sikap yang dikembangkan dalam penelitian ini.

Bentuk penskalaan respons yang diterapkan pada skala sikap matematis dilakukan mengingat bahwa secara teoretis sikap matematis bersifat lebih umum dan bisa berlaku untuk semua materi matematika. Skala sikap matematis yang dikembangkan dalam penelitian ini menggunakan skala Likert yakni Sangat Tidak Sesuai (STS), Tidak Sesuai (TS), Kurang Sesuai (KS), Sesuai (S) dan Sangat Sesuai (SS). Berdasarkan hasil analisis Lisrel pada data uji coba II, butir dengan dukungan terbesar pada konstruk sikap matematis adalah butir M9 yang mewakili indikator B3. Butir tersebut adalah sebagai berikut: “Saya akan mencari contoh penyelesaian soal yang sejenis apabila mengalami kesulitan menyelesaikan suatu soal matematika”. Butir ini mewakili indikator B3 yakni berusaha memahami sifat-sifat matematika.

Aspek sikap yang diungkap melalui instrumen yang dikembangkan dalam penelitian ini mengacu pada pengelompokan pendidikan karakter yang baik menurut Lickona (1992: 12) yakni *moral knowing*, *moral feeling*, dan *moral action*. Instrumen penilaian aspek sikap yang dikembangkan dalam penelitian ini difokuskan pada karakter mulia (*good character*) siswa melalui teknik penilaian diri sehingga kondisi

yang diungkap masih terbatas pada level *moral feeling*. Artinya, kondisi psikologis terkait karakter yang diungkap baru mencerminkan komitmen atau tekad siswa untuk bertindak baik atau berkarakter baik yakni jujur dan bertanggung jawab. Hal ini bersesuaian dengan taksonomi domain afektif dari Krathwohl yang menjadi acuan dalam Kurikulum 2013 yakni pada tahap *receiving*, *responding*, *valuing*, dan *organization*. Oleh sebab itu dapat dikatakan bahwa instrumen penilaian aspek sikap yang dihasilkan dalam penelitian ini cukup relevan dengan Kurikulum 2013.

Produk akhir instrumen penilaian aspek sikap yang dihasilkan penelitian ini terbukti secara teoretis dan empiris berkualitas baik dan relevan dengan Kurikulum 2013. Namun, memperhatikan bentuk instrumennya, membutuhkan waktu lama dalam penskoran apabila dilakukan secara manual. Oleh sebab itu, perkembangan teknologi sekarang seperti *google drive* dapat dimanfaatkan guna mempermudah dan mempercepat penskoran yang dilakukan. Hal ini sebagaimana Leachy (2012) yang mengemukakan bahwa perkembangan teknologi juga semestinya dapat dimanfaatkan guru untuk bidang penilaian karena teknologi dapat didesain dan dimanfaatkan guna memberikan respons cepat terhadap hasil kerja siswa. Upaya penggunaan teknologi yang sangat menunjang kelancaran penilaian aspek sikap memang perlu dilakukan mengingat pentingnya dilakukan penilaian afeksi siswa di sekolah. Hal ini sebagaimana dikemukakan Popham (2003: 107) bahwa penilaian afeksi di sekolah sebaiknya dilakukan karena penilaian afeksi yang dilakukan selama kegiatan pembelajaran di sekolah berkontribusi terhadap perilaku siswa setelah lulus.

Sebagaimana instrumen penilaian aspek sikap, instrumen penilaian aspek pengetahuan dan keterampilan yang dihasilkan penelitian ini juga terbukti berkualitas baik. Bahkan, berdasarkan hasil uji coba I, sudah memenuhi bukti secara empiris bahwa keseluruhan butir tes dan tugas dapat dikatakan sudah baik sehingga tidak perlu dilakukan revisi dan uji coba II. Ditinjau dari validitas isi dan validitas konstruksinya, seluruh butir tes dan tugas berkualitas baik. Indeks validitas isi V Aiken pada keseluruhan butir tes sebesar 0,88-1,00 dan pada tugas sebesar 0,88-0,92. Selain bukti indeks validitas isi dengan V Aiken, pada instrumen penilaian aspek pengetahuan yang berbentuk tes uraian juga didukung bukti validitas isi ditinjau dari keselarasannya dengan Kurikulum 2013 menggunakan metode revisi taksonomi Bloom. Indeks keselarasan yang diperoleh melalui perhitungan dengan menggunakan rumus Porter adalah sebesar 0,93.

Keselarasan yang termasuk dalam kategori tinggi tersebut di antaranya disebabkan karena instrumen penilaian aspek pengetahuan berupa tes uraian materi statistika yang dihasilkan dalam penelitian ini disusun mengacu pada revisi taksonomi Bloom. Dimensi pengetahuan dan proses kognitif yang diukur meliputi B3, C3, C4, dan C5. B3 berarti pengetahuan konseptual dan proses kognitif menerapkan. C3 berarti pengetahuan prosedural dan proses kognitif menerapkan. C4 berarti pengetahuan prosedural dan proses kognitif menganalisis. C5 berarti pengetahuan prosedural dan proses kognitif mengevaluasi. Hal ini menunjukkan bahwa tes uraian yang dihasilkan memuat butir-butir yang mengukur kemampuan

berpikir sedang dan tinggi sehingga memiliki relevansi yang tinggi dengan penekanan penguasaan kompetensi pengetahuan siswa berdasarkan Kurikulum 2013.

Adanya uji keselarasan dalam penelitian ini dapat menambah khasanah referensi terkait pembuktian validitas isi suatu instrumen penilaian. Hal ini mengingatkan bahwa studi tentang uji keselarasan antara penilaian dengan standar dalam kurikulum masih sangat jarang dilakukan di Indonesia. Padahal, studi ini penting untuk dilakukan sebagaimana diungkapkan Tindal (2005: 87) yang menyatakan bahwa hasil dari studi keselarasan dapat digunakan untuk menentukan apakah restrukturisasi penilaian diperlukan atau tidak. Jika restrukturisasi diperlukan, hasil uji keselarasan akan membantu untuk mengidentifikasi perubahan yang diperlukan dalam penilaian.

Ditinjau dari validitas konstraknya, seluruh butir tes dan tugas juga berkualitas baik. Seluruh butir tes mendukung konstruk tes dan konstruk tugas juga didukung oleh seluruh butir tugas. Analisis pada instrumen penilaian keterampilan berupa tugas proyek pada penelitian ini juga dilakukan sebagaimana pada tes mengingat kedua instrumen memang dirancang secara bersamaan dan keduanya saling terkait. Hal ini sebagaimana pendapat Fan (2004: 11) yang mengemukakan bahwa penilaian proyek dapat menilai domain kognitif, kemampuan pemecahan masalah siswa, dan keterampilan berpikir kreatif.

Adapun ditinjau dari koefisien reliabilitasnya baik reliabilitas konstruk maupun reliabilitas berdasarkan konsistensi internalnya, reliabilitas tes dan tugas juga termasuk kategori baik. Tes memiliki reliabilitas yang baik dengan koefisien CR sebesar 0,88 dan koefisien Alpha sebesar 0,78. Adapun tugas juga memiliki

reliabilitas yang baik dengan koefisien CR sebesar 0,82 dan koefisien Alpha sebesar 0,70. Hal ini menunjukkan bahwa tes dan tugas memiliki konsistensi hasil pengukuran yang tinggi. Namun demikian, dalam implementasi penggunaan instrumen ini untuk skala luas harus memperhatikan berbagai faktor yang mempengaruhi reliabilitas. Hal ini mengingat sebagaimana disebutkan Ebel & Frisbie (1991: 86) bahwa apabila tes digunakan sebagai tes standar maka koefisien reliabilitas sebaiknya antara 0,85-0,95 sedangkan bila digunakan untuk kelas minimal 0,65. Artinya, apabila tes digunakan dalam skala besar dan dimaksudkan untuk pengambilan keputusan yang sangat krusial sebaiknya terstandar dan memiliki koefisien reliabilitas antara 0,85-0,95. Oleh sebab itu tes yang dihasilkan dalam penelitian ini lebih cocok digunakan untuk penilaian dalam lingkup kelas.

Kualitas instrumen tes dan tugas ditinjau dari karakteristik butirnya termasuk baik. Seluruh butir tes berkategori baik dengan indeks daya beda 0,31-2,01 dan indeks kesukaran -2,33 sampai -0,47 dalam skala logit. Adapun untuk instrumen tugas, seluruh butir tugas berkategori baik dengan indeks daya beda 0,46-1,11 dan indeks kesukaran sebesar -2,01 sampai -0,91 dalam skala logit. Ditinjau dari fungsi informasi tes dan kesalahan baku pengukurannya, tes memberikan informasi akurat pada theta -3,0 sampai 0,8 dan informasi tertinggi pada theta sekitar -2,4. Tugas memberikan informasi akurat pada theta -3,0 sampai 0,5 dan informasi tertinggi pada theta sekitar -1,2. Hal ini berarti bahwa instrumen penilaian pengetahuan dan keterampilan dalam penelitian ini tepat digunakan pada siswa dengan kemampuan sedang dan kurang. Hasil ini menunjukkan bahwa instrumen yang dikembangkan

dalam penelitian ini sesuai apabila diterapkan pada siswa secara umum bukan pada siswa yang berada di sekolah atau kelas unggulan.

Penggunaan lebih lanjut dari instrumen penilaian aspek pengetahuan dan keterampilan yang dihasilkan penelitian ini harus memperhatikan dan disesuaikan dengan Kompetensi Dasar yang tertuang dalam peraturan terbaru Kurikulum 2013 yang berlaku. Hal ini mengingat bahwa rumusan Standar Kompetensi Lulusan (SKL) pada tingkat satuan pendidikan SMP juga terus mengalami perbaikan namun tetap meliputi aspek sikap, pengetahuan, dan keterampilan. Rumusan SKL SMP yang terbaru tertuang dalam Permendikbud Nomor 20 Tahun 2016 tentang Standar Kompetensi Lulusan Pendidikan Dasar dan Menengah. Adapun Kompetensi Inti dan Kompetensi Dasar untuk mata pelajaran matematika SMP yang berlaku saat ini tertuang dalam Permendikbud Nomor 24 Tahun 2016 tentang Kompetensi Inti dan Kompetensi Dasar Pelajaran pada Kurikulum 2013.

Hasil penilaian pencapaian kompetensi siswa pada karakter jujur dan tanggung jawab adalah: (a) tingkat komitmen atau tekad para siswa dalam dirinya untuk berbuat atau bertindak jujur adalah 22,13% sangat kuat, 39,08% kuat, 26,87% cukup kuat, dan 11,93% kurang kuat, dan (b) tingkat komitmen atau tekad para siswa dalam dirinya untuk menjadi seorang yang bertanggung jawab adalah 18,97% sangat kuat, 35,06% kuat, 30,60% cukup kuat, dan 15,37% kurang kuat. Hasil ini menunjukkan bahwa karakter jujur dan tanggung jawab secara keseluruhan dapat dikatakan telah tertanam dengan cukup kuat pada siswa di kota Yogyakarta namun masih harus terus ditingkatkan dan diperkuat agar siswa benar-benar memiliki

karakter tersebut dalam dirinya. Guru perlu melakukan penguatan karakter jujur dan tanggung jawab melalui kegiatan pembelajaran yang dilakukan. Integrasi penguatan karakter jujur dan tanggung jawab dalam pembelajaran pada akhirnya juga dapat meningkatkan karakter baik lainnya seperti disiplin, kreatif, kritis, dan lain sebagainya. Adapun untuk penilaian aspek sikap matematis siswa diperoleh bahwa 16,09% sangat baik, 35,06% baik, 34,34% cukup baik, dan 14,51% kurang baik. Hal ini menunjukkan bahwa sikap matematis siswa SMP di kota Yogyakarta yang menerapkan Kurikulum 2013 secara umum juga dapat dikatakan cukup baik tapi tetap masih harus ditingkatkan. Hal ini karena sikap matematis yang dimiliki siswa sangat berpengaruh pada hasil belajar siswa dalam pembelajaran matematika. Melalui pembelajaran matematika yang menyenangkan dan metode pembelajaran yang bervariasi dapat menjadi salah satu cara meningkatkan sikap matematis siswa.

Berdasarkan hasil penilaian terhadap pencapaian kompetensi pengetahuan diperoleh rata-rata 79,09 dan simpangan baku 14,80. Menggunakan KKM sebesar 75 terdapat sebanyak 433 siswa atau 62,94% siswa yang tuntas atau mampu mencapai KKM tersebut. Hal ini menunjukkan bahwa penguasaan materi statistika siswa sudah cukup baik namun masih harus terus ditingkatkan. Adapun pencapaian kompetensi keterampilan siswa dalam pembelajaran matematika materi statistika diperoleh rata-rata 86,64 dan simpangan baku 16,30. Menggunakan KKM sebesar 75 terdapat 551 siswa adalah 80,09% siswa yang tuntas atau mampu mencapai KKM tersebut. Hasil ini menunjukkan bahwa pencapaian kompetensi keterampilan siswa pada materi statistika telah baik. Hal ini sangat dimungkinkan di antaranya karena instrumennya

berupa tugas sehingga siswa punya cukup waktu untuk mengerjakan dan menunjukkan performan yang lebih baik daripada ketika mengerjakan tes. Penguasaan kompetensi keterampilan dalam pembelajaran matematika sangat dipengaruhi oleh penguasaan kompetensi pengetahuan. Demikian pula sebaliknya, pencapaian kompetensi pengetahuan juga dipengaruhi oleh penguasaan kompetensi keterampilan siswa. Oleh sebab itu, penilaian aspek pengetahuan dan keterampilan dalam pembelajaran matematika bersifat saling melengkapi sehingga para guru matematika sebaiknya terus meningkatkan kualitas kedua penilaian tersebut dalam kegiatan penilaian yang dilakukannya. Selain itu, berdasarkan hasil yang diperoleh dalam penelitian ini, guru juga dapat merancang tindakan yang lebih tepat terhadap siswa yang telah tuntas maupun belum tuntas. Bagi siswa yang telah tuntas, guru dapat memberikan soal atau tugas bersifat pengayaan yang lebih tinggi levelnya dan meningkatkan kreativitas siswa. Adapun bagi siswa yang belum tuntas, guru dapat merancang berbagai program perbaikan seperti remedial secara klasikal maupun latihan individu secara intensif.

Hasil penilaian pencapaian kompetensi yang dipaparkan dalam penelitian ini merupakan pencapaian kompetensi yang dicapai para siswa SMP di kota Yogyakarta yang telah menerapkan Kurikulum 2013 sejak tahun pelajaran 2013/2014. Hasil ini dapat menjadi informasi yang sangat bermanfaat bagi para guru matematika khususnya di kota Yogyakarta dalam rangka meningkatkan proses pembelajaran yang dilakukan. Apalagi mengingat bahwa tahun 2016 pemerintah mengambil kebijakan untuk secara bertahap melaksanakan Kurikulum 2013 secara nasional menggunakan

Revisi Kurikulum 2013 yang ditetapkan pada tahun 2016 (Asyari, Jawa Pos: 21 Maret 2016). Hasil penelitian ini juga dapat menjadi masukan bagi pemerintah DIY yang mulai tahun 2016 menargetkan 25% sekolah di DIY telah menerapkan Kurikulum 2013 yang telah direvisi dan terus disempurnakan (Maharani, Republika: 12 Juni 2016).

Terjadinya dinamika dalam pemerintahan dengan adanya pergantian Menteri Pendidikan dan Kebudayaan (Mendikbud) ternyata tidak mempengaruhi kegiatan revisi Kurikulum 2013. Hal ini sebagaimana dikemukakan Menteri Pendidikan dan Kebudayaan Muhadjir Effendy bahwa Kurikulum 2013 tidak akan dirombak melainkan akan terus disempurnakan (Maharani, Republika: 6 September 2016). Bahkan, Menteri Pendidikan dan Kebudayaan Muhadjir Effendy menyebutkan kalau untuk pendidikan tingkat SMP porsi untuk ilmu pengetahuan sebesar 40% dan pendidikan karakter sebesar 60% (Maharani, Republika: 6 September 2016). Oleh sebab itu, produk hasil penelitian disertasi ini tentunya menjadi sangat bermanfaat bagi para guru yang menerapkan Kurikulum 2013. Terlebih, pemerintah juga sudah meluncurkan adanya program Penguatan Pendidikan Karakter di sekolah.

Produk penelitian berupa instrumen penilaian aspek sikap, pengetahuan, dan keterampilan yang dihasilkan penelitian ini, diperoleh berdasarkan langkah-langkah penyusunan instrumen yang dirumuskan melalui kajian analitis berbagai teori. Langkah-langkah penyusunan instrumen penilaian aspek sikap meliputi: (1) menentukan tujuan penilaian dan jenis sikap yang hendak dinilai, (2) melakukan kajian teoritik, (3) memilih teknik penilaian, bentuk instrumen, dan model skala

pengukuran, (4) menyusun kisi-kisi, dan (5) menulis butir-butir instrumen, sistem penskoran, dan cara penafsirannya. Adapun langkah-langkah penyusunan instrumen penilaian aspek pengetahuan dan keterampilan meliputi: (1) menentukan tujuan penilaian dan teknik penilaian yang digunakan, (2) menentukan indikator dan merumuskan *abstract continuum*, (3) menyusun kisi-kisi instrumen penilaian pengetahuan dan keterampilan, dan (4) menulis butir-butir instrumen dan pedoman atau rubrik penskorannya. Penerapan langkah-langkah tersebut dalam penelitian ini ternyata menghasilkan instrumen penilaian aspek sikap, pengetahuan, dan keterampilan yang terbukti secara teoretis dan empiris berkualitas baik. Hal ini juga menunjukkan bahwa langkah-langkah penyusunan instrumen yang digunakan dalam penelitian ini juga dapat menjadi panduan dan tambahan referensi terkait kegiatan penyusunan instrumen penilaian dalam pembelajaran matematika.

G. Keterbatasan Penelitian

Melalui penelitian ini telah dihasilkan instrumen penilaian pencapaian kompetensi matematika materi statistika siswa SMP berdasarkan Kurikulum 2013 untuk aspek sikap, pengetahuan, dan keterampilan yang secara teoretis dan empiris terbukti berkualitas baik. Instrumen yang dihasilkan dirancang untuk dapat digunakan sebagai instrumen yang bersifat standar dalam kegiatan penilaian pencapaian kompetensi matematika siswa pada materi statistika SMP berdasarkan Kurikulum 2013. Standarisasi instrumen telah diupayakan semaksimal mungkin, namun dalam pelaksanaannya terdapat berbagai kekurangan. Hal ini tentunya mengakibatkan

adanya keterbatasan dalam implikasi dan penerapan instrumen yang dihasilkan dari penelitian ini. Keterbatasan yang terjadi terutama terkait dengan sampel uji coba dan pelaksanaan penelitian.

- a) Sampel uji coba dalam penelitian ini masih belum sesuai dengan target siswa yang akan diuji. Walaupun teknik pengambilan sampel telah dilakukan dengan memperhatikan karakteristik sebaran sampel ditinjau dari kualitas sekolah namun belum sepenuhnya mewakili siswa yang ditargetkan. Berbagai karakteristik lain seperti letak sekolah, kondisi demografis, status sosial ekonomi, karakteristik siswa, dan variabel lainnya terkait heterogenitas siswa belum sepenuhnya diperhatikan. Penentuan sampel terbatas pada SMP yang berada di kota Yogyakarta yang telah menerapkan Kurikulum 2013 dengan mengacu pada kualitas sekolah yang mewakili sedang berkembang, baik, dan tinggi. Kondisi ini tentunya belum memenuhi syarat standar siswa pada umumnya. Implikasi dari hal ini di antaranya adalah terkait penggunaan kriteria kategorisasi dalam skala untuk mengukur karakter jujur, karakter tanggung jawab, dan sikap matematis dalam penelitian. Kriteria kategorisasi yang digunakan dalam penelitian ini tentunya belum bisa diterapkan secara umum melainkan hanya relevan digunakan pada sekolah dengan kondisi dan karakteristik siswa yang memiliki kemiripan atau hampir sama dengan sekolah tempat uji coba. Oleh sebab itu, perlu dilakukan penelitian lanjutan melibatkan sekolah-sekolah lain dengan berbagai karakteristik agar instrumen yang dihasilkan dapat digunakan pada skala yang lebih luas.

b) Pelaksanaan uji coba instrumen belum sepenuhnya memenuhi kriteria dalam proses standarisasi instrumen. Beberapa kekurangan terkait belum terpenuhinya kriteria pelaksanaan uji coba di antaranya: (1) waktu yang diperlukan untuk menyelesaikan semestinya sesuai dengan ketentuan, namun ketika uji coba berlangsung mengalami perubahan karena tingkat kemampuan siswa di lapangan sangat beragam; (2) pelaksanaan uji coba masih belum bebas dari kemungkinan timbulnya bias yang di antaranya disebabkan oleh: (a) tata tertib dan tata ruang yang semestinya diatur sedemikian sehingga siswa benar-benar menunjukkan pencapaian kompetensinya masing-masing, namun dalam penelitian ini kurang terkondisikan dengan baik, (b) setiap siswa dalam ruang yang sama mengerjakan instrumen yang sama secara serentak, satu bangku untuk dua siswa, dan pengawasan cenderung kurang ketat sehingga adanya siswa yang saling bekerja sama tidak dapat dihindarkan.

BAB V

SIMPULAN DAN SARAN

A. Simpulan tentang Produk

Penelitian ini menghasilkan produk berupa instrumen penilaian pencapaian kompetensi matematika siswa untuk aspek sikap, pengetahuan, dan keterampilan pada materi statistika SMP berdasarkan Kurikulum 2013 yang secara teoretis dan empiris terbukti berkualitas baik. Berdasarkan hasil penelitian, dapat dikemukakan simpulan tentang produk sebagai berikut.

1. Instrumen penilaian yang dihasilkan meliputi:
 - a. Instrumen penilaian aspek sikap berupa skala untuk mengukur karakter jujur siswa yang terdiri atas 15 butir berbentuk penskalaan subjek, skala untuk mengukur karakter tanggung jawab siswa yang terdiri atas 14 butir berbentuk penskalaan subjek, dan skala sikap matematis yang terdiri atas 14 butir berbentuk penskalaan respons.
 - b. Instrumen penilaian aspek pengetahuan dan keterampilan berupa tes bentuk uraian dan tugas proyek yang bersifat individu pada materi statistika.
2. Kualitas instrumen penilaian yang dihasilkan adalah sebagai berikut.
 - a. Instrumen Penilaian Aspek Sikap
 - 1) Validitas isi keseluruhan instrumen adalah baik dengan didukung kesesuaian butir dengan indikator pada keseluruhan butir adalah valid dengan indeks *V Aiken* sebesar 0,88–0,92.

- 2) Butir-butir pada skala untuk mengukur karakter jujur, skala untuk mengukur karakter tanggung jawab, dan skala sikap matematis secara keseluruhan memiliki validitas konstruk yang baik.
 - 3) Reliabilitas dari skala untuk mengukur karakter jujur, skala untuk mengukur karakter tanggung jawab, dan skala sikap matematis adalah baik dengan koefisien Omega berturut-turut sebesar 0,85, 0,89, dan 0,81.
- b. Instrumen Penilaian Aspek Pengetahuan dan Keterampilan
- 1) Validitas isi tes adalah baik dengan didukung kesesuaian butir dengan indikator pada keseluruhan butir adalah valid dengan indeks *V Aiken* sebesar 0,88–0,92 dan indeks keselarasan Porter sebesar 0,93. Validitas isi tugas adalah baik dengan didukung kesesuaian butir dengan indikator pada keseluruhan butir adalah valid dengan indeks *V Aiken* sebesar 0,88–1,00.
 - 2) Butir-butir tes dan tugas secara keseluruhan memiliki validitas konstruk yang baik.
 - 3) Reliabilitas tes adalah baik dengan koefisien CR sebesar 0,88 dan koefisien Alpha sebesar 0,778. Reliabilitas tugas adalah baik dengan koefisien CR sebesar 0,82 dan koefisien Alpha sebesar 0,70.
 - 4) Seluruh butir tes berkategori baik dengan indeks daya beda 0,31-2,01 dan indeks kesukaran -2,33 sampai -0,47 dalam skala logit. Seluruh

butir tugas berkategori baik dengan indeks daya beda 0,46–1,11 dan indeks kesukaran -2,01 sampai -0,91 skala logit.

- 5) Tes memberikan informasi akurat pada theta -3,0 sampai 0,8 dan informasi tertinggi pada theta sekitar -2,4. Tugas memberikan informasi akurat pada theta -3,0 sampai 0,5 dan informasi tertinggi pada theta sekitar -1,2.
3. Pencapaian kompetensi matematika materi statistika pada siswa SMP di Kota Yogyakarta berdasarkan penggunaan instrumen penilaian yang dihasilkan adalah sebagai berikut.
 - 1) Pencapaian Kompetensi Aspek Sikap
 - a) Hasil penilaian terhadap pencapaian kompetensi siswa pada karakter jujur adalah 22,13% sangat kuat, 39,08% kuat, 26,87% cukup kuat, dan 11,93% kurang kuat.
 - b) Hasil penilaian terhadap pencapaian kompetensi siswa pada karakter tanggung jawab adalah 18,97% sangat kuat, 35,06% kuat, 30,60% cukup kuat, dan 15,37% kurang kuat.
 - c) Hasil penilaian terhadap sikap matematis siswa adalah 16,09% sangat baik, 35,06% baik, 34,34% cukup baik, dan 14,51% kurang baik.

- 2) Pencapaian Kompetensi Aspek Pengetahuan dan Keterampilan
 - a) Hasil penilaian terhadap pencapaian kompetensi aspek pengetahuan diperoleh rata-rata sebesar 79,09 dan simpangan baku sebesar 14,80. Menggunakan KKM sebesar 75 terdapat sebanyak 433 siswa atau 62,94% yang tuntas atau mampu memenuhi KKM.
 - b) Hasil penilaian terhadap pencapaian kompetensi aspek keterampilan diperoleh rata-rata sebesar 86,64 dan simpangan baku sebesar 16,30. Menggunakan KKM sebesar 75 terdapat sebanyak 511 siswa atau 80,09% yang tuntas atau mampu memenuhi KKM.

B. Saran Pemanfaatan Produk

Saran yang diajukan terkait pemanfaatan produk adalah:

1. Secara keseluruhan, instrumen penilaian yang dihasilkan dapat langsung digunakan terutama oleh para guru matematika SMP dalam rangka melakukan penilaian pencapaian kompetensi aspek sikap, pengetahuan, dan keterampilan pada materi statistika. Namun untuk instrumen penilaian aspek pengetahuan dan keterampilan harus memperhatikan dan disesuaikan dengan Kompetensi Dasar yang tertuang dalam peraturan terbaru Kurikulum 2013 yang berlaku.
2. Skala untuk mengukur karakter jujur dan tanggung jawab yang dihasilkan penelitian ini masih terbatas pada mengukur karakter siswa pada level *moral feeling* atau tingkat komitmen siswa. Oleh sebab itu, dalam implementasinya

sebaiknya dilengkapi dengan pengamatan langsung pada siswa. Selain itu, apabila penerapan skala ini dilakukan pada sekolah dengan karakteristik hampir sama dengan subjek uji coba dalam penelitian ini, maka dapat menggunakan kategorisasi yang sama. Namun, apabila kondisi sekolahnya memiliki karakteristik yang berbeda perlu diperhatikan dengan cermat penentuan kategorisasinya.

3. Implementasi instrumen yang dihasilkan dari penelitian ini lebih cocok digunakan dalam penilaian yang bersifat sebagai *assessment of learning* baik pada siswa secara individual, klasikal, maupun institusional.
4. Produk berupa instrumen penilaian aspek sikap, pengetahuan, dan keterampilan yang dihasilkan pada penelitian ini juga dapat diadaptasi dan dimanfaatkan untuk kegiatan penelitian guru seperti penelitian tindakan kelas tentang upaya peningkatan hasil belajar atau karakter siswa terkait karakter jujur, karakter tanggung jawab, dan sikap matematis siswa. Hal ini karena skala untuk mengukur karakter jujur, skala untuk mengukur karakter tanggung jawab, skala sikap matematis, tes, dan tugas yang dihasilkan dalam penelitian ini telah teruji kualitasnya baik secara teoretis maupun empiris.

C. Diseminasi dan Pengembangan Produk Lebih Lanjut

1. Instrumen penilaian aspek sikap yang dihasilkan penelitian ini masih bersifat manual, sehingga penskoran memerlukan waktu yang relatif lama. Pengembangan lebih lanjut dapat dibuat dalam bentuk *software* atau disajikan secara online misalnya melalui *google drive* sehingga penskoran menjadi lebih mudah dan cepat. Hal ini mengingat bahwa penggunaan skala psikologis yang disajikan secara online memungkinkan untuk dilakukan pada sekolah-sekolah yang telah memiliki fasilitas pendukung laboratorium komputer yang lengkap.
2. Analisis hasil uji coba untuk pembuktian validitas konstruk tidak hanya dapat dilakukan dengan Lisrel, pengembangan lebih lanjut bisa juga dilakukan dengan *software* lain seperti Amos atau Mplus.
3. Kegiatan uji coba lebih lanjut perlu dilakukan di wilayah lain dengan responden yang lebih besar dan beragam sehingga semakin memperluas generalisasi dan menambah bukti kualitas instrumen yang dihasilkan dalam penelitian ini.
4. Pengembangan tes dapat dikembangkan lagi dalam bentuk tes uraian non objektif yang memuat *open ended problem* serta mengukur kemampuan metakognitif dan intuitif siswa. Adapun untuk instrumen penilaian aspek sikap dapat dikembangkan lagi untuk karakter lain seperti mandiri, percaya diri, dan lain sebagainya. Selain itu, pengembangan instrumen juga dapat dikembangkan

dalam bentuk instrumen yang mencakup ketiga aspek secara holistik dan benar-benar cocok digunakan dalam kegiatan penilaian yang bersifat *single event*.

5. Uji keselarasan perlu dilakukan lebih lanjut dengan metode lain seperti Webb (1997), SEC (2001), dan lain sebagainya sehingga dapat dibandingkan hasilnya serta dilakukan tidak terbatas pada penilaian aspek pengetahuan saja. Selain itu, format telaah untuk uji keselarasan dalam penelitian ini masih terbatas pada revisi taksonomi Bloom secara umum menggunakan skala Likert. Pengembangan lebih lanjut dapat menggunakan metode dari Anderson et al. (2015) yakni dengan format yang lebih sederhana namun mendalam dan kegiatan validasi atau telaah oleh para pakar dilakukan dengan menggunakan sistem online.

DAFTAR PUSTAKA

- Airasian, P. W. & Miranda, H. (2002). The role of assessment in the revised taxonomy. *Theory into Practice*, 41(4), 249–254.
- Aiken, L. R. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement*, 40, 955-959.
- Ajzen, I. (1989). Attitude structure and behavior. Dalam Pratkanis, A. R., Beckler, S. J., & Greenwald, A. G. (Eds.), *Attitude structure and function*. Hillsdale, NJ: Lawrence Erlbaum.
- Allen, M. J. & Yen, W. M. (1979). *Introduction to measurement theory*. Monterey: Cole Publishing Company.
- Ananda, S. (2003). *Rethinking issues of alignment under No Child Left Behind*. San Francisco: WestEd.
- Anderson, D., Irvin, S., Alonzo, J., & Tindal, G. A. (2015). Gauging item alignment through online systems while controlling for rater effects. *Educational Measurement: Issues and Practice*, 34, 22-33.
- Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing. A revision of Bloom's taxonomy of educational objectives*. New York: Addison Wesley Longman.
- Anderson, L. W. (2002). Curricular alignment: A re-examination. *Theory into Practice*, 41(4), 255-260.
- Anderson, P. & Morgan, G. (2008). *Developing tests and questionnaires for a national assessment of educational achievement*. Washington, DC: The World Bank.

- Arcavi, A. & Schoenfeld, A. H. (2006) "Using the unfamiliar to problematize the familiar" (in Portuguese). Dalam Borba, M. (Ed.) *Tendencias Internacionais em Formação de Professores de Matemática (International Perspectives in Mathematics Teacher Education)*, 87-111 (English version available).
- Asmani, J. M. (2011). *Buku panduan internalisasi pendidikan karakter di sekolah*. Yogyakarta: Diva Press.
- Asyari, Y. (21 Maret 2016). Juli, Kurikulum tigabelas diberlakukan secara nasional. *Jawapos.com*
- Azwar, S. (2005). *Penyusunan skala psikologi*. Edisi 1. Yogyakarta : Pustaka Pelajar.
- Bahar, A. & Maker, C. J. (2015). Cognitif backgrounds of problem solving: A comparison of open-ended vs. closed mathematics problems. *Eurasia Journal of Mathematics, Science & Technology Education*, 11(6), 1531-1546.
- Baron, R. A. & Byrne, D. (1987). *Social psychology: understanding human interaction*, 5th. Ed. Boston, MA: Allyn and Bacon.
- Bhola, D. S., Impara, J. C., & Buchendahl, C. W. (2003). Aligning tests with states' content standards: Methods and issues. *Educational Measurement: Issues and Practice*, 22(3), 21–29.
- Biggs, J. (2003). *Teaching for quality learning at university*. Glasgow: The Society for Research into Higher Education & Open University Press.
- Bisri. (2009). *Akhlaq*. Jakarta: Direktorat Jenderal Pendidikan Islam Agama RI.
- Conner, C. (1991). *Assessment and testing in the primary school*. London: The Falmer Press.

- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Belmont: Wadsworth Group.
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of test. *Psychometrika*, 16, 297-334.
- De Ayala, R. J. (1993). An introduction to polytomous item response theory models. *Measurement and Evaluation in Counseling and Development*, 25, 172-189.
- de Lange, J. (2000). *Assessment: No change without problems*. Utrecht: Freudenthal Institute.
- Depdikbud. (2003). *Undang-Undang Republik Indonesia Nomor 20, Tahun 2003, tentang Sistem Pendidikan Nasional*.
- Dunn, W. N. (1994). *Pengantar analisis kebijakan public* (Terjemahan Samodra Wibawa, Diah Asitadani, Agus Heruanto Hadna, dkk.). Yogyakarta: Gadjah Mada University Press.
- Du Toit, M. (2003). *IRT from SSi: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood: SSi.
- Dyer, J. H., Gregersen, H. B., & Christensen, C. M. (2009). *The Innovator's DNA*. <http://www.barnesandnoble.com/w/innovators-dna-jeff-dyer/1100744423>.
- Ebel, R. L. & Frisbie, D. A. (1991). *Essentials of educational measurement*. USA: Prentice-Hall Inc.
- Emosda. (2011). Penanaman nilai-nilai kejujuran dalam menyiapkan karakter bangsa. *Innovatio*, X(1), 151-166.

- Ercegovic, Z., & Richardson, J. V. (2004). Academic dishonesty, plagiarism included, in the digital Age: A Literature Review. *College & Research Libraries*, 301-318.
- Gable, R. K. (1986). *Instrument development in the affective domain*. Boston: Klenver-Nijhoff Publishing.
- Gagne, R. M. & Briggs, L. J. (1979). *Principles of instructional design*. New York: Holt, Rinehart, and Winston.
- Gamoran, A., Porter, A. C., Smithson, J., & White P. A. (1997). Upgrading high school mathematics instruction: improving learning opportunities for low-achieving, low-income youth. *Edu Eval Policy Anal.* 19, 325–338.
- Garson, G. D. (2009). *Overview structural equation modeling*, <http://faculty.chass.ncsu.edu/garson/PA765/structur.htm>.
- Gronlund, N. E. & Linn, R. L. (1990). *Measurement and evaluation in teaching (6th ed)*. New York: Collier Macmillan Publishers.
- Hadi, S. (2007). *Konstruk kinerja kepala sekolah dan variabel yang mempengaruhinya*. Disertasi Doktor, tidak diterbitkan, Universitas Negeri Yogyakarta, Yogyakarta.
- Hair, J. F., Anderson, R. F., Tatham, R. L., & Black, W. C., (1998). *Multivariate data analysis (5thed)*. Prentice Hall [versi elektronik].
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2009). *Multivariate data analysis (7thed)*. Prentice Hall [versi elektronik].
- Hambleton, R.K. & Swaminathan, H. (1985). *Item response theory*. Boston, MA: Kluwer Inc.

- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage Publication Inc.
- Herman, J. L., Webb, N. M., & Zuniga, S. A. (2007). Measurement issues in the alignment of standards and assessments: A case study. *Applied Measurement in Education*, 20(1), 101-176.
- Herman, J. L. & Webb, N. M. (2007). Alignment methodologies. *Applied Measurement in Education*, 20, 1-5.
- Herman, T. (2001). *Asesmen portofolio dalam pembelajaran matematika*. Prosiding Seminar Nasional Matematika Universitas Gajah Mada, Yogyakarta 14 Juli 2001.
- Hooper, D., Coughlan, J., Mullen, M. R. (2008). Structural equation modelling: guidelines for determining model fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.
- Hopkins, C. D. & Antes, R. L. (1990). *Classroom measurement and evaluation*. Itasca, Illinois: F.E. Peacock Publisher, Inc.
- Igbaria, M., Zinatelli, N., Cragg, P., & Cavaye, A. L. M. (1997). Personal computing acceptable factors in small firms: A structural equation model. *MIS Quarterly*, September, 279-299.
- Jöreskog, K. & Sörbom, D. (1996). *Lisrel 8: User's reference guide*. Scientific Software Internasional.
- Jurich, D. & Goodman, J. (2009). *A comparison of IRT parameter recovery in mixed format examinations using PARSCALE and ICL*. Presented at the Annual meeting of Northeastern Educational Research Association, Oktober 2009.

- Katagiri, S. (2006). *Mathematical thinking and how to teach it*. Tokyo: Meijitosyo Publishers (CRICED, University of Tsukuba).
- Kemendikbud. (2013). *Modul pelatihan implementasi Kurikulum 2013: Materi pelatihan guru implementasi Kurikulum 2013 SMP/MTs*. Jakarta: Badan Pengembangan Sumber Daya Manusia Pendidikan dan Kebudayaan dan Penjaminan Mutu Pendidikan.
- Kemendikbud. (2014a). *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 104, Tahun 2014, tentang Penilaian Hasil Belajar oleh Pendidik pada Pendidikan Dasar dan Menengah*.
- Kemendikbud. (2014b). *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 160, Tahun 2014, tentang Pemberlakuan Kurikulum Tahun 2006 dan Kurikulum 2013*.
- Kemendikbud (2014c). *Refleksi pelaksanaan Kurikulum 2013*. Jakarta: Direktorat Pembinaan Sekolah Menengah Pertama.
- Kemendikbud. (2015). *Panduan Penilaian untuk Sekolah Menengah Pertama*. Jakarta: Direktorat Pembinaan Sekolah Menengah Pertama.
- Kemendikbud. (2016a). *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 23, Tahun 2016, tentang Standar Penilaian Pendidikan*.
- Kemendikbud. (2016b). *Peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 24, Tahun 2016, tentang Kompetensi Inti dan Kompetensi Dasar Pelajaran pada Kurikulum 2013 pada Pendidikan Dasar dan Pendidikan Menengah*.
- Kerlinger, F. N. (1986). *Foundations of behavioral research*. New York: Holt, Rinehart and Winston Inc.

- Kizlik, B. (2009). *Measurement, assessment, and evaluation in education*. Diakses tanggal 22 Oktober 2013 dari <http://www.adprima.com/measurement.html>.
- Koellhoffer, T. (2009). *Character Education: Being fair and honest*. New York: Infobase Publishing.
- Koss, J. (2011). Academic dishonesty among adolescents. *American Psychological Association*, 11(2), 38-46.
- Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). *Taxonomy of educational objectives: Handbook II: Affective domain*. New York: David McKay.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41, 212-218.
- Krueger, R.A. & Casey, M.A. (2000). *Focus groups: A practical guide for applied research* (3rd Edition). Thousand Oaks: Sage Publications.
- Kumaidi. (Maret 2014). Implementasi penilaian autentik dalam pembelajaran di kelas. *Makalah* disajikan dalam Seminar Nasional Implementasi Penilaian dan Pelaksanaan Kurikulum 2013, di Universitas Negeri Jakarta.
- La Marca, P. M. (2001). Alignment of standards and assessments as an accountability criterion. *Practical Assessment, Research & Evaluation*, 7(21). Diambil pada tanggal 14 Oktober 2013 dari <http://PAREonline.net/getvn.asp?v=7&n=21>.
- La Marca, P. M., Redfield, D., Winter, P. C., et al. (2000). *State standards and state assessment systems: A guide to alignment*. Washington, DC: Council of Chief State Officers.
- Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology*, 28, 563-575.

- Leachy, G. (2012). 'QR Code in mathematics classrooms'. Dalam *Mathematics Teaching Issue*, 235, 27-29. Derby UK: The Association of Teacher of Mathematics.
- Lickona, T. (1992). *Educating for Character*. New York: Bantam Books
- Linn, R. L. (1998). Validating inferences from National Assessment of Educational Progress achievement-level reporting. *Applied Measurement in Education*, 11, 23-47.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84-99.
- Maharani, E. (12 Juni 2016). DIY targetkan 25 persen sekolah terapkan K-13, *Republika.co.id*.
- Maharani, E. (6 September 2016). Mendikbud ingin ubah kurikulum 2013. *Republika.co.id*.
- Mardapi, D. (2008). *Teknik penyusunan instrumen tes dan non tes*. Yogyakarta: Mitra Cendekia Press.
- Mardapi, D. (2016). *Pengukuran, penilaian, dan evaluasi pendidikan*. Yogyakarta: Parama Publishing.
- Mehrens, W. A. & Lehmann, I. J. (1973). *Measurement and evaluation in education and psychology*. New York: Holt, Rinehart and Winston. Inc.
- Muraki, E. (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement*, 17(4), 351-393.

- Muraki, E. & Bock, R. D. (1997). *Parscale 3: IRT based test scoring and item analysis for graded items and rating scales*. Chicago: Scientific Software Inc.
- Muraki, E. & Bock, D. (2002) PARSCALE 4.1 Computer program. Chicago: Scientific Software International, Inc.
- Nasstrom, G. & Henriksson, W. (2008). Alignment of standards and assessment: A theoretical and empirical study of methods for alignment. *Electronic Journal of Research in Educational Psychology*, 6(3), 667-690.
- Nasstrom, G. (2009). Interpretation of standards with Bloom's revised taxonomy: A comparison of teachers and assessment experts. *International Journal of Research & Method in Education*, 32(1), 39-51.
- National Council of Teacher of Mathematics. (1995). *Assessment standards for school mathematics*, Reston, VA: Author
- National Council of Teacher of Mathematics. (2000). *Principles and standards for school mathematics*, Reston, VA: Author
- Ndlovu, M. & Mji, A. (2012). Alignment between south african mathematics assessment standards and the TIMSS assessment frameworks. *Pythagoras*, 33(3), Art. #182, 9 pages. <http://dx.doi.org/10.4102/pythagoras.v33i3.182>
- Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Upper Saddle River, NJ: Merrill.
- Nunnally, J. C. (1981). *Psychometric theory* (2nd ed). New Delhi: McGraw-Hill Publishing Company Limited.
- Oriondo, L. L. & Antonio, D. E. M. (1998). *Evaluation educational outcomes*. Manila: Rex Printing Compagny

- Ornstein, A.C. (1992). Essay tests: use, development, and grading. *ProQuest Education Journals*, 65(3).
- Overton, T. (2008). *Assessing learners with special needs: An applied approach (7th Edition)*. University of Texas – Brownsville.
- Palomba, C. A. & Banta, T. W. (1999). *Assessment essentials: planning, implementing, improving*. San Francisco: Jossey-Bass.
- Popham, W. J. (2009). *Test better, teach better: the instructional role of assessment*. United States of America: ASCD (*Association for Supervision and Curriculum Development*)
- Porter, A. C. & Smithson, J. (2001). *Defining, developing, and using curriculum indicators*. Philadelphia, PA: Consortium for Policy Research in Education, University of Pennsylvania.
- Porter, A. C. (2004). *Curriculum assessment*. Nashville, TN: Vanderbilt University.
- Reeve, B. B., & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. Dalam Fayers, P. & Hays, E.D. (Eds), *Assessing quality of life in clinical trials: Methods of practice* (2nd ed). New York: Oxford University Press.
- Retnawati, H. (2015). Hambatan guru matematika sekolah menengah pertama dalam menerapkan kurikulum baru. *Cakrawala Pendidikan*, 3, 390-403.
- Ross, J. A. (2006). The reliability, validity, and utility of self assessment. *Journal Practical Assessment, Research & Evaluation*, 11, 1-3.
- Rusyan, A. T. (2006). *Pendidikan budi pekerti*. Jakarta: Inti Media Cipta Nusantara.

- Sa'aduddin, I. A. (2006). *Meneladani akhlak nabi membangun kepribadian muslim*. Bandung: Rosda Karya
- Seels & Glasgow (1990). *Exercises in instructional design*. Columbus OH: Merrill Publishing Company.
- Shepard, L. A. (2006). Classroom assessment. Dalam Brennan, R. L. (Eds). *Educational measurement*. Washington: American Council Educational and Praeger Publisher.
- Sireci, S. & Bond, M. F. (2014). Validity evidence based on test content. *Psicothema*, 26(1), 100-107.
- Smith, M. S. & O'Day, J. (1990). Systematic school reform. Dalam Furhman, S.H. & Malen, B. (Eds.), *The politics of curriculum and testing* (pp.233-267). The 1990 yearbook of the Politics of Education Association. London: The Falmer Press.
- Soedjadi, R. (2000). *Kiat pendidikan matematika di Indoneasia (Konstantasi keadaan masa kini menuju harapan masa depan)*. Jakarta: Dirjen Dikti Depdiknas.
- Squire, P. J. (2001). Cognitive levels of testing agricultural science in senior secondary schools in Botswana. *Education*, 121(3), 597-603.
- Stern, L. & Ahlgren, A. (2002) Analysis of students' assessments in middle school curriculum materials: Aiming precisely at benchmarks and standards. *Journal of Research in Science Education*, 39, 889 – 910.
- Subali, B. (2016). Pengembangan tes beserta penyelidikan validitas dan reliabilitas secara empiris. Yogyakarta: UNY Press

- Sujono. (1988). *Pengajaran matematika untuk sekolah menengah*. Jakarta: Ditjen Dikti.
- Suranto, Muhyadi, & Mardapi, D. (2014). Pengembangan instrumen evaluasi Uji Kompetensi Keahlian (UKK) administrasi perkantoran di SMK. *Jurnal Penelitian dan Evaluasi Pendidikan*, 18(1), 98-114.
- Suurtamm, C. Koch, M., & Arden, A. (2010). Teacher's assessment practices in mathematics: classroom in the context of reform. *Assessment in Education: Principles, Policy, & Practice*, 17(4), 399-417.
- Thissen, D. & Wainer, H. (2001). *Test scoring*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Thorpe, G. L. & Favia A. (2012). *Data analysis using item response theory methodology: An introduction to selected programs and applications*. Psychology Faculty Scholarship. Paper 20. http://digitalcommons.library.umaine.edu/psy_facpub/20
- Tindal, G. (2005). *Alignment of alternate assessments using the webb system*. Washington, DC: Council of Chief State Officers.
- Umar, J. (2014). Kerancuan dalam penggunaan istilah 'construct reliability'. *JP3I*, III(4), 393-400.
- Van der Linden, W. J. & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer-Verlag.
- Van De Walle, J. A. (2007). *Elementary and middle school mathematics* (6th Ed.), Boston: Pearson Education Inc.

- Walstad, W. B. (2006). Testing for depth of understanding in economics using essay questions. *Journal of Economic Education*. Washington: Winter.
- Webb, N. L. (1997). *Criteria for alignment of expectations and assessments in mathematics and science education (Research monograph No. 6)*. Washington, DC: Council of Chief State School Officers.
- Weeden, P., Winter, J. & Broadfoot, P. (2002). *Assessment: what's in it for school*. New York: Routledge Falmer.
- Wells, C. S., Hambleton, R. K. & Purwono, U. (Juni 2008). Item response theory. Polytomus response IRT models and aplicatios. *Handout* disampaikan pada Pelatihan Asesmen Pendidikan dan Psikologi (Psikometri), di Universitas Negeri Yogyakarta.
- Widhiarso, W. (2009). Koefisien reliabilitas pada pengukuran kepribadian yang bersifat multidimensi. *Psikobuana*, 1(1), 39-48.
- Wijanto, S. H. (2008). *Structural equation modeling dengan Lisrel 8.8*. Yogyakarta: Graha Ilmu.
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Mahwah: Lawrence Erlbaum Associates, Inc.
- Yamin, S. & Kurniawan, H. (2009). *Structural equation modeling: Belajar lebih mudah teknik analisis data kuesioner dengan Lisrel-PLS*. Jakarta: Salemba Infotek.
- Zuchdi, D. (2016). *Pendidikan karakter dalam perspektif teori dan praktik*. Yogyakarta: UNY Press.

