

Pemanfaatan Software SPSS untuk Analisis Instrument Penelitian Dalam Penelitian Tindakan Kelas

Deden Mutakin

Mualimin Muhammadiyah Yogyakarta, Yogyakarta, Indonesia

dedenmutakin@muallimin.sch.id

Received: May 30, 2023

Revised: June 23, 2023

Accepted: June 26, 2023

KATAKUNCI

Analisis;
Instrumen Penelitian;
Penelitian Tindakan Kelas;
SPSS

KEYWORDS

Question item analysis
SPSS 16.0

ABSTRAK

Dalam kegiatan PTK instrument tes (soal) yang digunakan sebagai alat ukur untuk mengukur keberhasilan suatu metode atau tindakan perlu diuji kualitasnya. Penelitian ini merupakan penelitian kuantitatif dengan metode deskriptif. Penelitian ini dimaksudkan untuk mencari informasi dan data yang dapat digunakan untuk mendeskripsikan kualitas instrumen tes secara empiris berdasarkan unsur validitas, reliabilitas, indeks kesukaran, dan daya pembeda soal yang diolah dengan bantuan software SPSS versi 16.0. Diagnostik soal dengan menguji cobakan soal kepada 42 siswa kelas VIII B Madarash Muallimin Muhammadiyah Yogyakarta. Hasil uji coba didapatkan 20 item soal baik instrument tes subtema 1 dan 2 kesemuanya valid dan reliable. Tingkat kesukaran instrument tes subtema 1 terdapat 4 soal kategori sedang dan 16 soal kategori mudah. Instrumen tes subtema 2 terdapat 10 soal kategori sedang dan 10 soal kategori mudah. Uji daya pembeda instrument tes subtema 1 kesemua soal dalam kategori baik. Instrument tes subtema 2 terdapat 19 soal dengan kategori baik dan 1 soal dengan kategori tidak baik. Manfaat setelah dilakukan analisis instrumen soal menggunakan SPSS ini adalah dapat membantu dalam evaluasi atas tes yang digunakan, mendukung penulisan butir soal yang efektif, menentukan apakah suatu fungsi butir soal sesuai dengan yang diharapkan, dan merevisi materi yang dinilai atau diukur.

Utilization of SPSS software for Research Instrument Analysis in Classroom Action Research

In Class-Room Action Research (CAR) activities, the test instrument (question) used as a measuring tool to measure the success of a method or action needs to be tested for quality. This research is a quantitative research with descriptive method. This research is intended to find information and data that can be used to describe the quality of test instruments empirically based on the elements of validity, reliability, difficulty index, and question differentiability processed with the help of SPSS 16.0 software. Diagnostic questions by testing questions to 42 students of class VIII B Madrasah Muallimin Muhammadiyah Yogyakarta. The test results obtained 20 items of questions both subtheme 1 and 2 test instruments are all valid and reliable. The level of difficulty of the subtheme 1 test instrument there are 4 questions in the medium category and 16 questions in the easy category. The subtheme 2 test instrument has 10 questions in the medium category and 10 questions in the easy

category. Test the distinguishing power of the subtheme 1 test instrument all questions in the good category. The subtheme 2 test instrument has 19 questions in the good category and 1 question in the bad category. The benefits after analyzing the question instrument using SPSS are that it can assist in evaluating the test used, support effective item writing, determine whether an item function is as expected, and revise the material being assessed or measured.

This is an open-access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Pendahuluan

Penelitian Tindakan Kelas (PTK) atau *Classroom Action Research* (CAR) pertama kali diperkenalkan oleh ahli psikologi sosial Amerika yang bernama Kurt Lewin pada tahun 1946 (Sukarsono, 2014; Gayatri dan Wirakusuma, 2016; Santoso et al., 2021). Penelitian tindakan kelas termasuk penelitian kualitatif, karena sejarah awalnya berupa penyelidikan kualitatif dalam ilmu sosial terapan (Adelman, 1993; Whitehead & McNiff, 2006). Meskipun data yang dikumpulkan bisa saja bersifat kuantitatif, dimana uraiannya bersikap deskriptif dalam bentuk kata-kata (Kunandar, 2011). Dalam PTK jenis data yang lazimnya dikumpulkan adalah data tentang proses pembelajaran, termasuk interaksi guru-siswa dan siswa-siswa jika relevan. Data yang benar akan membawa pada hasil sesuai dengan keadaan yang sebenarnya. Benar tidaknya data tergantung pada baik tidaknya instrumen pengumpul data atau pengukur objek dari suatu variabel penelitian (Arikunto, 2010). Baik tidaknya suatu instrument penelitian ditentukan oleh validitas dan reliabilitasnya (Yusup, 2018). Validitas instrument memperlmasalahkan sejauh mana pengukuran tepat dalam mengukur apa yang hendak diukur (Nurkancana dkk, 1992) dan menilai apa yang seharusnya dinilai (Sudjana, 2004). Sedangkan reliabilitas memperlmasalahkan sejauh mana suatu pengukuran dapat dipercaya karena keajegannya (Sudjana, 2004) dan mengungkap data yang bisa dipercaya (Arikunto, 2010).

Salah satu ciri penelitian kualitatif yaitu peneliti bertindak sebagai instrumen sekaligus pengumpul data. Instrumen selain manusia (seperti; angket, pedoman wawancara, pedoman observasi dan sebagainya) dapat pula digunakan, tetapi fungsinya terbatas sebagai pendukung tugas peneliti sebagai instrumen kunci. Begitu pula penelitian tindakan kelas, dimana penelitian ini termasuk penelitian kualitatif maka makna langsung dari tindakan yang dibatasi oleh sudut pandang peneliti (guru) itu sendiri terhadap proses penelitiannya (Sanjaya, 2012). Seperti yang dituliskan Hopkins (1993), terdapat 6 prinsip penelitian tindakan kelas salah satunya adalah teknik pengumpulan data tidak menuntut waktu dan cara yang berlebihan dan prosedur pengumpulan data yang dapat ditangani sendiri. Artinya adalah terdapat metodologi yang agak longgar dan lebih bersifat informal, meskipun dijaga keobjektifannya tetapi masih menimbulkan keraguan.

Maka timbul pertanyaan dalam PTK, apakah kaidah-kaidah penelitian ilmiah mampu dijaga selama pengumpulan data? Bisa jadi terdapat potensi manipulasi yang dilakukan baik oleh guru maupun oleh siswa karena perintah guru. Misal, suatu contoh peneliti (guru) melakukan penelitian tindakan kelas menggunakan metode tertentu (X) untuk meningkatkan hasil belajar (Y). Untuk mengukur sejauh mana tingkat penguasaan siswa terhadap materi ajar yang telah disampaikan melalui penggunaan metode yang dipilih maka dilakukan evaluasi (test). Sebagai contoh bentuk manipulasi instrument tesnya seperti seluruh soal dibuat ranah kognitif L1/LOTS (*Low Order Thinking Skills*), proporsi jumlah tingkat kesukaran soal tidak memenuhi kaidah yang ditentukan/seluruh soal dibuat mudah, dan daya pembeda yang tidak baik/tidak efektif sehingga semua soal mampu dijawab benar oleh siswa kelompok tinggi maupun rendah. Dengan penerapan soal yang seperti itu maka sudah jelas pasti hasil evaluasi akan baik yang berarti metode yang digunakan dalam PTK dinilai “berhasil”. Atau dengan kata lain keshahihan, kehandalan, tingkat kesukaran, dan daya pembeda pada soal yang digunakan untuk evaluasi/test dalam PTK patut dipertanyakan.

Teknik pengumpulan data merupakan cara yang ditempuh peneliti untuk mengumpulkan data dengan alat yang cocok untuk digunakan dalam penelitian. Sedangkan alat pengumpul data yang digunakan haruslah disesuaikan dengan teknik pengumpul data, misal seperti salah satu contoh teknik pengukuran dengan alatnya berupa tes (instrument). Seperti yang diungkapkan oleh Tuckmaus (dalam Nurgiyantoro, 2010) bahwa alat tes harus dapat dipertanggungjawabkan dari segi kelayakan (*appropriateness*). Syarat tes (butir soal) yang baik harus shahih (*valid*) dan handal (*reliable*) (Gronlund, 2002; Suryabrata, 2007; Slameto, 2008; Arikunto, 2009). Kemudian memiliki keseimbangan antara butir soal sukar, sedang dan mudah (Nurkancana dkk, 1992). Soal yang baik adalah memiliki indeks kesukaran 0,3 – 0,7 (Fernandes, 2004), 0,25 – 0,75 (Sudjana, 2004). Lebih lanjut test juga memiliki daya pembeda soal. Agar dapat diterima maka nilai D (*discrimination*: daya pembeda soal) adalah 0,30 atau lebih (Naga, 2002), mempunyai nilai D 0,40 dianggap efektif (Hopkins et al., 1999).

Karena kelonggaran metodologi atau memang sudah menjadi kebiasaan dalam pelaksanaan PTK. Seringkali peneliti (guru) tidak melakukan uji validitas, reliabilitas, tingkat kesukaran, dan daya pembeda soal (alat test) terlebih dahulu. Padahal wajib hukumnya dalam kaidah metodologi ilmiah test (instrument) perlu dilakukan pengukuran. Tujuannya agar supaya dapat diketahui sejauh mana ketepatan dan kecermatan suatu alat ukur dalam melakukan fungsi ukurannya, mengukur apa yang seharusnya diukur, memberikan hasil ukur yang tepat dan akurat sesuai dengan maksud dikenakannya tes tersebut, pengukuran dapat dipercaya, dan pengukuran yang konsisten dan stabil dari waktu ke waktu. Hal inilah yang sering dipertanyakan oleh peneliti mengenai keshahihan PTK yang dilakukan guru sendiri di dalam kelasnya (Mc Taggart, 1991).

Beberapa literatur ditemukan bahwa pemaparan PTK yang dilakukan tidak mencantumkan kegiatan uji alat tes/instrument (soal). Seperti penelitian (Udzma & Farid, 2023; Mutakin, 2022a, 2022b; Rifanty, 2019; Yusuf, 2015). Pada penelitian-penelitian tersebut tidak dibahas bagaimana langkah dalam uji validitas, reliabilitas, uji indeks kesukaran dan daya pembeda soal sebagai instrument alat ukur untuk mengukur hasil capaian dalam penerapan suatu metode *Index Card Match* (Udzma & Farid, 2023), hasil belajar menggunakan *google form* (Mutakin, 2022), hasil belajar menggunakan metode *team quiz* (Mutakin, 2022), hasil belajar menggunakan model pembelajaran kooperatif tipe *make a match* (Rifanty, 2019), pembelajaran EMPECE (Yusup, 2018).

Untuk mengetahui validitas, reliabilitas, indeks kesukaran, dan daya pembeda soal suatu instrumen penelitian diperlukan analisis statistik. Padahal sebetulnya tidak terlalu sulit jika kaidah itu (uji validitas, reliabilitas, uji indeks kesukaran dan daya pembeda soal) dilakukan menggunakan *software* statistik, contohnya seperti SPSS. Seperti penggunaan IBM SPSS statistic 25 untuk mengetahui validitas reliabilitas tingkat kesukaran dan daya beda butir soal ujian akhir semester (Magdalena et al., 2021). Kemudian SPSS adalah program aplikasi yang memiliki analisis data statistik yang cukup tinggi (Tarigan et al., 2022). SPSS salah satu program aplikasi yang paling banyak diminati dan digunakan oleh para analis dan peneliti untuk mengolah data-data statistik (Machali, 2015). SPSS mampu melakukan analisis statistik yang akurat karena sudah disesuaikan dengan prinsip-prinsip statistika modern yang sudah diintegrasikan oleh para ahli statistik (Yuliardi & Nuraeni, 2016). Kelebihan dari *software* SPSS ini mampu mengakses data dari berbagai macam format data yang tersedia misalnya: *Excel, Open Acces, Text* bisa langsung dibaca *software* SPSS untuk dianalisis; *software* SPSS memberi tampilan data yang lebih informatif, memudahkan pengguna dengan simbol dan ikon yang menarik dan mudah dipahami, *software* SPSS memberikan informasi lebih akurat dengan memperlakukan *missing data* secara tepat, yaitu dengan memberi kode alasan mengapa terjadi *missing data* (Nuraeni, 2021).

Sejauh ini tidak banyak sumber artikel jurnal terbitan terbaru dari para peneliti (guru) dalam melakukan penelitian tindakan kelas melakukan serangkaian kegiatan analisis butir soal pada instrument testnya (menggunakan *software* SPSS atau sejenisnya misalnya). Bisa menggunakan *expert judgement* tapi hal ini tentu dibutuhkan biaya yang tidak sedikit bagi peneliti (guru). Penelitian ini berusaha membawa suatu semangat bahwa instrument yang valid dan reliabel pada PTK sangat perlu diperhatikan karena pada dasarnya penelitian tindakan kelas juga merupakan serangkaian kegiatan ilmiah dimana membutuhkan keketatan dalam setiap tahapan metodologinya. Suatu anggapan yang salah bahwa PTK tidak memerlukan analisis statistik yang rumit. Bukanlah suatu hal yang rumit dalam analisis statistik sebetulnya karena sudah ada macam-macam *software* statistik (SPSS 16.0 salah satunya) yang menawarkan

kemudahan dalam menganalisis butir soal secara kuantitatif untuk mengetahui kualitas soal sebagai instrument dalam penelitian tindakan kelas (PTK). Berdasarkan uraian diatas, maka sangat perlu dilakukan analisis butir soal instrument test dalam setiap penelitian tindakan kelas (PTK).

Metode

Penelitian ini merupakan penelitian kuantitatif dengan metode deskriptif. Penelitian ini dimaksudkan untuk mencari informasi dan data yang dapat digunakan untuk mendeskripsikan kualitas instrumen tes secara empiris berdasarkan unsur validitas, reliabilitas, indeks kesukaran, dan daya pembeda soal yang diolah dengan bantuan *software* SPSS versi 16.0. Penelitian ini menggunakan data sekunder dari soal pembelajaran IPS tema 3 “Keunggulan dan Keterbatasan antarruang serta Pengaruhnya terhadap Kegiatan Ekonomi, Sosial, Budaya di Indonesia dan ASEAN”. Subtema 1 “Keunggulan dan Keterbatasan antarruang serta Peran Pelaku Ekonomi dalam suatu Perekonomian” dan subtema 2 “Perdagangan Antardaerah atau Antarpulau dan Perdagangan Internasional”. Soal tersebut nantinya akan diujikan kepada 43 siswa kelas VIII A Madrasah Muallimin Muhammadiyah Yogyakarta. Maka sebelumnya dilakukan diagnostik soal dengan menguji cobakan soal kepada 42 siswa kelas VIII B Madrasah Muallimin Muhammadiyah Yogyakarta. Teknik pengambilan data dilakukan secara *purposive sampling* dengan asumsi kelas VIII B memiliki kemampuan yang sama dengan kelas VIII A yang ditunjukkan dengan kesamaan nilai rata-rata akhir semester I Tahun Ajaran 2021/2022. Data sekunder diperoleh dengan dokumen tertulis dengan empat opsi (a, b, c, dan d) sebanyak 20 butir soal.

Pengujian validitas dengan kriteria pengambilan keputusan jika r hitung $\geq r$ tabel maka soal “valid” dan jika r hitung $\leq r$ tabel maka soal “tidak valid”. Nilai r tabel yang menjadi acuan dengan derajat kepercayaan 99% $N = 42$ sebesar 0,393. Uji reliabilitas dengan kriteria pengambilan keputusan jika *Cronbach’s Alpha* hitung \geq *Cronbach’s Alpha* acuan maka instrument “reliable” dan jika *Cronbach’s Alpha* hitung \leq *Cronbach’s Alpha* acuan maka instrument “tidak reliable”. Adapun nilai *Cronbach’s Alpha* acuan yang dipakai sebesar $> 0,80$ dengan tingkat keandalan “sangat andal” (Hair et al, 2010). Uji taraf kesukaran dengan melihat mean (output SPSS), kriteria pengambilan keputusan:

- 0,00 – 0,25 Sukar
- 0,26 – 0,75 Sedang
- 0,76 – 1,00 Mudah (Zainul dan Nasoetion, 1997).

Sedangkan daya pembeda dengan melihat r hitung (Output SPSS), kriteria pengambilan keputusan:

- 0,40 – 1,00 Baik

- 0,30 – 0,39 Sedang (tidak perlu direvisi)
- 0,20 – 0,29 (perlu direvisi)
- -1,00 – 0,19 Tidak Baik (Dali, 1992).

Hasil dan Pembahasan

1. Uji Validitas

Hasil dan Pembahasan dapat disajikan dalam subbab. Membahas secara jelas pokok bahasan sesuai dengan masalah, tujuan penelitian, dan teori yang digunakan.

Berdasarkan hasil uji validitas instrument soal kepada siswa kelas VIII B dengan menggunakan program SPSS 16.0 *for windows* didapat hasil sebagai berikut.

Tabel 1. Rekap Data Hasil Uji Validitas Instrumen Tes Subtema 1

No Soal	r Hitung	r Tabel $\alpha = 0,01$ (99%) N = 42	Kriteria pengambilan keputusan	Hasil
1	0,911			
2	0,702			
3	0,865			
4	0,615			
5	0,805			
6	0,512			
7	0,432		Jika r Hitung \geq r Tabel maka soal	
8	0,640		"valid"	
9	0,834			
10	0,728	0,393	Jika r Hitung \leq r Tabel maka soal	valid
11	0,911		"tidak valid"	
12	0,702			
13	0,865			
14	0,615			
15	0,805			
16	0,512			
17	0,432			
18	0,640			
19	0,834			
20	0,728			

Sumber: Olah data

Tabel 2. Rekap Data Hasil Uji Validitas Instrumen Tes Subtema 2

No Soal	r Hitung	r Tabel $\alpha = 0,01$ (99%) N = 42	Kriteria pengambilan keputusan	Hasil
1	0,788			
2	0,650			
3	0,682			
4	0,431			
5	0,750			
6	0,760			
7	0,765		Jika r Hitung \geq r Tabel maka soal	
8	0,497		"valid"	

<i>No Soal</i>	<i>r Hitung</i>	<i>r Tabel $\alpha = 0,01$ (99%) N = 42</i>	<i>Kriteria pengambilan keputusan</i>	<i>Hasil</i>
9	0,619			valid
10	0,760	0,393	Jika $r \text{ Hitung} \leq r \text{ Tabel}$ maka soal	
11	0,788		"tidak valid"	
12	0,760			
13	0,682			
14	0,431			
15	0,750			
16	0,760			
17	0,703			
18	0,497			
19	0,619			
20	0,730			

Sumber: Olah Data

Dari tabel rekap data hasil uji validitas instrumen tes diatas disimpulkan bahwa uji coba ke-20 soal kepada 42 siswa kelas VIII B baik instrument tes subtema 1 dan 2 didapat r hitung $\geq r$ tabel maka soal dikatakan **valid**.

2. Uji Reliabilitas

Berdasarkan hasil uji reliabilitas soal dengan menggunakan program SPSS 16.0 *for windows* didapat hasil sebagai berikut.

Tabel 3. Rekap Data Hasil Uji Reliabilitas Instrumen Tes Subtema 1

<i>Cronbach's Alpha Hitung</i>	<i>Cronbach's Alpha Acuan</i>	<i>Kriteria Pengambilan Keputusan</i>	<i>Keputusan</i>
0,944	0,8	Jika Cronbach's Alpha Hitung \geq Cronbach's Alpha Acuan maka instrument "reliabel" Jika Cronbach's Alpha Hitung \leq Cronbach's Alpha Acuan maka instrument "tidak reliabel"	reliabel

Sumber: Olah Data

Tabel 4. Rekap Data Hasil Uji Reliabilitas Instrumen Tes Subtema 2

<i>Cronbach's Alpha Hitung</i>	<i>Cronbach's Alpha Acuan</i>	<i>Kriteria Pengambilan Keputusan</i>	<i>Keputusan</i>
0,937	0,8	Jika Cronbach's Alpha Hitung \geq Cronbach's Alpha Acuan maka instrument "reliabel" Jika Cronbach's Alpha Hitung \leq Cronbach's Alpha Acuan maka instrument "tidak reliabel"	reliabel

Sumber: Olah Data

Dari tabel rekap data hasil uji reliabilitas instrumen tes diatas disimpulkan bahwa uji coba ke-20 soal kepada 42 siswa kelas VIII B baik instrument tes subtema 1 dan 2 didapat Cronbach's Alpha Hitung \geq Cronbach's Alpha Acuan maka soal dikatakan **reliable**.

3. Tingkat Kesukaran

Berdasarkan hasil uji tingkat kesukaran soal dengan menggunakan program SPSS 16.0 for windows didapat hasil sebagai berikut.

Tabel 5. Rekap Data Hasil Uji Tingkat Kesukaran Instrumen Tes Subtema 1

<i>No. Soal</i>	<i>Mean (Output SPSS)</i>	<i>Kriteria pengambilan keputusan</i>	<i>Tingkat kesulitan</i>
1	0,81	0,00 – 0,25 Sukar	Mudah
2	0,90	0,26 – 0,75 Sedang	Mudah
3	0,81	0,76 – 1,00 Mudah	Mudah
4	0,83		Mudah
5	0,74		Sedang
6	0,93		Mudah
7	0,64		Sedang
8	0,83		Mudah
9	0,83		Mudah
10	0,86		Mudah
11	0,81		Mudah
12	0,90		Mudah
13	0,81		Mudah
14	0,83		Mudah
15	0,74		Sedang
16	0,93		Mudah
17	0,64		Sedang
18	0,83		Mudah
19	0,83		Mudah
20	0,86		Mudah

Sumber: Olah Data

Tabel 6. Rekap Data Hasil Uji Tingkat Kesukaran Instrumen Tes Subtema 2

<i>No. Soal</i>	<i>Mean (Output SPSS)</i>	<i>Kriteria pengambilan keputusan</i>	<i>Tingkat kesulitan</i>
1	0,81	0,00 – 0,25 Sukar	Mudah
2	0,69	0,26 – 0,75 Sedang	Sedang
3	0,81	0,76 – 1,00 Mudah	Mudah
4	0,83		Mudah
5	0,74		Sedang
6	0,67		Sedang
7	0,64		Sedang
8	0,83		Mudah
9	0,83		Mudah
10	0,67		Sedang
11	0,81		Mudah
12	0,67		Sedang
13	0,81		Mudah
14	0,83		Mudah
15	0,74		Sedang
16	0,67		Sedang
17	0,62		Sedang
18	0,83		Mudah
19	0,83		Mudah

<i>No. Soal</i>	<i>Mean (Output SPSS)</i>	<i>Kriteria pengambilan keputusan</i>	<i>Tingkat kesulitan</i>
20	0,64		Sedang

Sumber: Olah Data

Dari tabel rekap data hasil uji tingkat kesukaran instrumen tes diatas disimpulkan bahwa uji coba ke-20 soal kepada 42 siswa kelas VIII B instrument tes subtema 1 kategori sedang ada 4 soal dan kategori mudah ada 16 soal. Kemudian pada instrumen tes subtema 2 kategori sedang ada 10 soal dan kategori mudah ada 10 soal.

4. Daya Pembeda

Berdasarkan hasil uji daya pembeda soal dengan menggunakan program SPSS 16.0 *for windows* didapat hasil sebagai berikut.

Tabel 7. Rekap Data Hasil Uji Daya Pembeda Instrumen Tes Subtema 1

<i>No. Soal</i>	<i>r Hitung (Output SPSS)</i>	<i>Kriteria Pengambilan Keputusan</i>	<i>Daya Beda Butir Tes</i>
1	0,911	0,40 – 1,00 Baik	Baik
2	0,702	0,30 – 0,39 Sedang (tidak perlu direvisi)	Baik
3	0,865	0,20 – 0,29 (perlu direvisi)	Baik
4	0,615	-1,00 – 0,19 Tidak Baik (dibuang)	Baik
5	0,805		Baik
6	0,512		Baik
7	0,432		Baik
8	0,640		Baik
9	0,834		Baik
10	0,728		Baik
11	0,911		Baik
12	0,702		Baik
13	0,865		Baik
14	0,615		Baik
15	0,805		Baik
16	0,512		Baik
17	0,432		Baik
18	0,640		Baik
19	0,834		Baik
20	0,728		Baik

Sumber: Olah Data

Tabel 8. Rekap Data Hasil Uji Daya Pembeda Instrumen Tes Subtema 2

<i>No. Soal</i>	<i>r Hitung (Output SPSS)</i>	<i>Kriteria Pengambilan Keputusan</i>	<i>Daya Beda Butir Tes</i>
1	0,788	0,40 – 1,00 Baik	Baik
2	0,650	0,30 – 0,39 Sedang (tidak perlu direvisi)	Baik
3	0,682	0,20 – 0,29 (perlu direvisi)	Baik
4	0,431	-1,00 – 0,19 Tidak Baik	Baik
5	0,750	(hasil negatife dibuang)	Baik
6	000		Tidak Baik
7	0,765		Baik

<i>No. Soal</i>	<i>r Hitung (Output SPSS)</i>	<i>Kriteria Pengambilan Keputusan</i>	<i>Daya Beda Butir Tes</i>
8	0,497		Baik
9	0,619		Baik
10	0,760		Baik
11	0,788		Baik
12	0,760		Baik
13	0,682		Baik
14	0,431		Baik
15	0,750		Baik
16	0,760		Baik
17	0,703		Baik
18	0,497		Baik
19	0,619		Baik
20	0,730		Baik

Sumber: Olah Data

Dari tabel rekap data hasil uji daya beda instrumen tes diatas disimpulkan bahwa uji coba ke-20 soal kepada 42 siswa kelas VIII B instrument tes subtema 1 kesemua soal dalam kategori baik. Kemudian pada instrument tes subtema 2 terdapat 19 soal dengan kategori baik dan 1 soal dengan kategori tidak baik. Pada soal nomor 6 instrumen tes subtema 2 memiliki r hitung atau indeks diskriminasi 0 yang berarti mempunyai daya pembeda yang rendah sekali. Butir soal tersebut dapat dijawab benar oleh semua kelompok atas maupun semua kelompok bawah, sehingga dilihat dari daya Pembedanya soal tersebut dikatakan tidak baik, karena tidak mempunyai daya pembeda.

Berdasarkan hasil uji validitas dan reabilitas instrument tes sebanyak 20 soal pilihan ganda baik pada soal subtema 1 dan 2 soal tersebut dikatakan valid dan reliable. Validitas adalah suatu derajat ketepatan/kelayakan instrumen yang digunakan untuk mengukur apa yang akan diukur (Ariffin, 2012). Menurut Sukardi (2013) validitas adalah derajat yang menunjukkan dimana suatu tes mengukur apa yang hendak diukur. Sedangkan menurut Azwar (2014) bahwa validitas mengacu sejauh mana akurasi suatu tes atau skala dalam menjalankan fungsi pengukurannya. Dari ketiga pendapat diatas dapat disimpulkan bahwa soal (sebagai alat ukur) yang valid artinya soal tersebut memiliki derajat keetepatan/kelayakan soal yang baik untuk digunakan dalam mengukur apa yang akan diukur (Magdalena dkk., 2021) serta soal tersebut menjalankan fungsi pengukurannya dengan baik (Neti, 2020). Validitas merupakan produk dari validasi. Soal yang valid maknanya bahwa soal tersebut sudah tervalidasi, artinya soal tersebut sudah menjalani suatu proses yang dilakukan oleh penyusun atau pengguna instrumen untuk mengumpulkan data secara empiris guna mendukung kesimpulan yang dihasilkan oleh skor instrumen. Berarti bisa dibayangkan jika dalam kegiatan PTK terdapat soal sebagai instrument tidak melalui sarangkaian validasi, lalu terdapat klaim keberhasilan maka bisa diragukan klaim tersebut. Hal ini senada dengan penelitian sebelumnya bahwa pentingnya kevalidan suatu instrumen seperti pada

penelitian (Musyaroh, 2021) yang mengatakan bahwa instrumen tes literasi kimia berbasis HOTS yang valid mampu memberikan gambaran tingkatan literasi kimia berdasarkan levelnya dan memberikan gambaran kemampuan literasi kimia dalam menjawab soal khususnya soal HOTS. Begitu juga dengan Utomo (2019); Sari dkk. (2019) yang berpendapat bahwa soal yang dijadikan alat evaluasi harus memiliki ciri-ciri soal yang baik salah satunya yaitu terkait kevalidan soal.

Kemudian reliabilitas, reliabilitas diterjemahkan dari kata *reliability* yang berarti hal yang dapat dipercaya (tahan uji). Sebuah tes dikatakan mempunyai reliabilitas yang tinggi jika tes tersebut memberikan data hasil yang ajeg (tetap) (Arikunto, 2021) walaupun diberikan pada waktu yang berbeda kepada siswa yang sama. Hasil tes yang tetap atau seandainya berubah maka perubahan itu tidak signifikan maka tes tersebut dikatakan reliabel. Oleh karena itu reliabilitas sering disebut dengan keterpercayaan, keterandalan, keajegan, konsistensi, dan kestabilan (Yusup, 2018). Reliabilitas menyangkut masalah ketepatan alat ukur (Budiastusi dan Agustinus, 2018). Beberapa ahli memberikan penekanan pada pengertian reliabilitas sebagai konsistensi tes. Yaitu, seberapa konsisten skor tes dari satu pengukuran ke pengukuran berikutnya. Reliabilitas merujuk pada ketetapan/keajegan soal (sebagai alat ukur) dalam menilai apa yang diinginkan, artinya kemampuan soal tersebut digunakan akan memberikan hasil yang relatif sama. Sehingga berdasarkan pernyataan-pernyataan diatas dapat disimpulkan bahwa, soal yang reliabel diartikan dengan soal yang memiliki keajegan (konsistensi) bila mana soal tersebut diuji berkali-kali hasilnya relatif sama, artinya setelah hasil tes yang pertama dengan tes yang berikutnya dikorelasikan terdapat hasil korelasi yang signifikan. Dalam PTK, kegiatan pengukuran lebih sering bersifat non fisik, seperti intelegensi, bakat dan minat, perilaku, persepsi siswa, atau hasil belajar siswa. Untuk mengukur dimensi tersebut diperlukan soal (instrumen tes) yang benar-benar valid dan reliabel. Kaitannya dengan kegiatan PTK dimana dalam penelitian menggunakan tindakan dalam bentuk pengaplikasian suatu metode tertentu untuk peningkatan belajar. Maka sudah tentu instrumen (soal) untuk mengukur peningkatan belajar haruslah diuji kevalidan dan reliabilitasnya terlebih dahulu sebelum diterapkan pada kelas tindakan. Seperti pada penelitian Fatayah (2022); Farida dan Musyarofah (2021); Tarigan dkk. (2022) bahwa pentingnya menganalisis butir soal untuk mengetahui sejauh mana kualitas soal, minimal kriteria yang digunakan valid dan reliabel.

Kemudian dari hasil uji tingkat kesukaran instrumen tes bahwa uji coba soal subtema 1 kategori sedang ada 4 soal dan kategori mudah ada 16 soal. Kemudian pada soal subtema 2 kategori sedang ada 10 soal dan kategori mudah ada 10 soal. Menurut Arifin (2009) sebaiknya proporsi antara tingkat kesukaran soal tersebar secara normal. Perhitungan proporsi tersebut dapat diatur sebagai berikut.

- Soal sukar 25%, soal sedang 50%, soal mudah 25%, atau

- Soal sukar 20%, soal sedang 60%, soal mudah 20%, atau
- Soal sukar 15%, soal sedang 70%, soal mudah 15%.

Padahal berdasarkan kondisi uji soal pada subtema 1 proporsinya adalah 20% soal sedang dan 80% soal mudah. Maka jika mengikuti salah satu proporsi soal yang baik maka harus merubah item soal. Dengan mengganti 1 soal kategori mudah menjadi sedang dan mengganti 5 soal kategori mudah menjadi sukar. Kemudian pada uji soal subtema 2 terdapat proporsi soal 50% sedang dan 50% mudah. Maka harus mengganti 5 soal kategori mudah menjadi sukar. Seperti merubah/mengganti soal mudah (LOTS) yang hanya mengetahui (mengingat kembali) menjadi soal sedang (MOTS) menjadi soal yang bermuatan pemahaman (menjelaskan ide/konsep) dan pengaplikasian (menggunakan informasi domain yang berbeda). Kemudian menjadi soal sukar (HOTS) menjadi soal yang bermuatan menganalisis (menspesifikasi aspek/elemen), mengevaluasi (menggambil keputusan sendiri), dan mengkreasi (mengkreasikan ide/gagasan sendiri). Kemudian bisa dikembangkan bentuk soal pilihan ganda tidak hanya pilihan ganda biasa tetapi, menjadi pilihan ganda kompleks (sebab-akibat, benar-salah, pilihan ganda lebih dari satu jawaban benar).

Selanjutnya dari hasil uji daya pembeda soal subtema 1 kesemua soal dalam kategori baik dan soal subtema 2 terdapat 19 soal dengan kategori baik dan 1 soal dengan kategori tidak baik. Pada uji kesukaran soal subtema 1 terdapat proporsi tingkat kesukarannya 20% (4 soal) sedang dan 80% (16 soal) mudah. Kemudian pada uji kesukaran soal subtema 2 terdapat proporsi tingkat kesukaran soal 50% (20 soal) sedang dan 50% (20 soal) mudah. Walaupun dalam soal subtema 1 dan 2 tidak terdapat soal sukar atau tidak mengikuti kaidah soal yang baik menurut Arifin (2009) yaitu soal sukar 25%, soal sedang 50%, soal mudah 25%. Tetapi uji daya pembeda menunjukkan kategori baik. Berarti artinya, baik pada soal subtema 1 dan 2 walaupun soal tersebut hanya ada kategori sedang dan mudah tetapi soal-soal tersebut hanya dapat dijawab dengan benar oleh siswa yang pandai saja. Berarti kelompok anak yang berprestasi tinggi dapat menjawab soal dengan benar dan seluruh atau hampir suatu kelompok yang berprestasi rendah menjawab soal dengan salah (meskipun semua soal tersebut kategori sedang dan mudah). Kemudian pada soal subtema 2 terdapat 1 soal uji daya pembeda tidak baik. Artinya antara kedua kelompok sama-sama menjawab dengan benar, berarti indeks diskriminasi (D)/ r hitung (output SPSS) soal tersebut 0,00 atau tidak memiliki daya pembeda. Dalam Sudjiono (2016), besaran 0,00 – 0,20 digolongkan pada klasifikasi *poor* Daya pembedanya lemah sekali (jelek) tidak memiliki daya pembeda sama sekali (jelek). Maka perlakuan soal tersebut adalah; (1) ditelusuri untuk kemudian diperbaiki, dan setelah diperbaiki dapat diajukan lagi dalam tes hasil belajar, yang kemudian dianalisis lagi apakah meningkat atau tidak dan (2) dibuang atau didrop dan tidak dikeluarkan lagi untuk tes hasil belajar.

Kegiatan menganalisis butir soal merupakan kegiatan yang wajib yang dilakukan guru untuk mengetahui tingkat hasil belajar peserta didik dan untuk meningkatkan mutu soal yang telah disusun (Fatayah, 2022; Nuraini dkk., 2021; Magdalena dkk., 2021; Farida dan Musyarofah, 2021). Dalam PTK, langkah ini bisa dimasukkan dalam tahap perencanaan (*plan*) yaitu pada mensimulasikan pelaksanaan tindakan dan menguji keterlaksanaannya sebelum diimplementasikan. Kegiatan ini merupakan proses pengumpulan, peringkasan, dan penggunaan informasi dari jawaban dari kelas uji coba untuk membuat keputusan tentang setiap penilaian yang nantinya akan diimplementasikan kepada kelas tindakan. Tujuan penelaahan adalah untuk mengkaji dan menelaah setiap butir soal agar diperoleh soal yang bermutu sebelum soal digunakan (Magdalena dkk., 2021). Di samping itu, tujuan analisis butir soal juga untuk membantu meningkatkan tes melalui revisi atau membuang soal yang tidak efektif, serta untuk mengetahui informasi diagnostik pada peserta didik apakah mereka sudah/belum memahami materi yang telah diajarkan (Farida dan Musyarofah, 2021) melalui pemilihan/penggunaan metode tertentu dalam kegiatan PTK.

Oleh karena itu berdasarkan manfaat ini dapat terlihat bahwa analisis butir soal ini memberikan peran dalam kualitas instrument tes yang akan digunakan. Hal ini dapat dilihat dari kualitas instrument tes baik segi tingkat validitas soal yang valid secara isi, tingkat kesukaran yang mempunyai tingkat proposional antara soal berkategori sedang, rendah dan tinggi, serta mampu membedakan peserta didik yang kurang dan peserta didik yang pandai. Ketika penyusunan instrument tes sudah memenuhi kriteria maka kualitas instrument tes pun akan meningkat sehingga sebuah instrument tes tidak hanya sebagai alat ukur yang mengukur apa yang hendak diukur akan tetapi setiap ranah dari siswa atau hal yang harus dievaluasi dari siswa dapat tergambar dari instrument tes yang telah disusun yang nantinya akan banyak temuan yang bisa terdiskripsikan dalam tahap refleksi.

Simpulan

Melalui penggunaan SPSS 16.0 untuk uji kualitas instrumen tes secara empiris berdasarkan unsur validitas, reliabilitas, indeks kesukaran, dan daya pembeda soal yang diujicobakan kepada siswa kelas VIII B. Didapatkan 20 item soal baik instrument tes subtema 1 dan 2 kesemuanya valid dan reliable. Kemudian uji coba tingkat kesukaran instrument tes subtema 1 terdapat 4 soal kategori sedang dan 16 soal kategori mudah. Kemudian pada instrumen tes subtema 2 terdapat 10 soal kategori sedang dan 10 soal kategori mudah. Kemudian uji daya pembeda instrument tes subtema 1 kesemua soal dalam kategori baik. Kemudian pada instrument tes subtema 2 terdapat 19 soal dengan kategori baik dan 1 soal dengan kategori tidak baik. Manfaat setelah dilakukan analisis instrumen soal menggunakan SPSS ini adalah dapat membantu dalam evaluasi atas tes

yang digunakan, mendukung penulisan butir soal yang efektif, menentukan apakah suatu fungsi butir soal sesuai dengan yang diharapkan, dan merevisi materi yang dinilai atau diukur. Dari hasil uji coba analisis butir soal ini, maka bisa dilakukan tindakan perbaikan yang nantinya digunakan untuk keperluan instrument PTK untuk kelas VIII A. Dengan adanya uji coba analisis butir soal ini, maka bisa didapatkan hasil pengukuran nantinya yang benar-benar berkualitas dan dapat dipertanggungjawabkan.

Ucapan Terimakasih

Penelitian ini merupakan hibah penelitian dari Madrasah Muallimin Muhammadiyah Yogyakarta. Penulis mengucapkan terimakasih kepada Direktur beserta jajaran Madrasah Muallimin Muhammadiyah Yogyakarta yang telah mendukung dan memberikan motivasi dalam kegiatan penelitian serta publikasi. Diucapkan terimakasih kepada Pengampu Jurnal Inovasi dan Manajemen Pendidikan yang telah memberikan kesempatan kepada saya untuk mengirimkan karya ini.

Daftar Pustaka

- Adelman, C. (1993). Kurt Lewin and the Origins of Action Research. *Educational Action Research*, 1(1), 7-24. <https://doi.org/10.1080/0965079930010102> Gayatri, wirakusuma, 2016.pdf. (n.d.).
- Arifin, Z. (2009). Evaluasi Pembelajaran Prinsip, Teknik, Prosedur. Bandung: PT Remaja Rosdakarya.
- Arikunto, S. (2021). Dasar-Dasar Evaluasi Pendidikan Edisi 3. Bumi Aksara.
- Azwar S. (2014). Psikologi Inteligensi. Yogyakarta.
- Budiastuti, D., dan Agustinus, B. (2018). Validitas dan Reliabilitas Penelitian. Jakarta: Mitra Wacana Media.
- Dali, N. S. (1992). Pengantar Teori Sekor pada Pengukuran Pendidikan. Jakarta: Besbats.
- Fatayah, F., Yuliana, I. F., dan Muf'idah L. (2022). Analisis Validitas Dan Reliabilitas Dalam Mendukung Ketuntasan Belajar Model STEM. *Jurnal Buana Pendidikan*. 18(1): 49-60.
- Farida dan Musyarofah, A. (2021). Validitas dan Reliabilitas dalam Analisis Butir Soal. *Al-Mu'arrib: Jurnal Pendidikan Bahasa Arab*. 1(1): 34-44.
- Fernandes, H.J.X. (2004). Testing and Measurement. Jakarta: Depdikbud.
- Gayatri dan Wirakusuma. (2016). Penelitian Tindakan Kelas Untuk Meningkatkan Keterampilan Pembuatan Proporsal Penelitian Mahasiswa. *E-Jurnal Akuntansi Universitas Udayana*. 14(2): 1539-1554.
- Gronlund, Norman E. (2002). *Construction Achievement Test*. Englewood Chiffs: Prentise-Hall.

- Hair, J. F., Anderson, R. E., Tatham, R. L., & Black, W. C., (1998). *Multivariate Data Analysis, fifth edition*. New Jersey: Prentice-Hall International, Inc.
- Hopkins, D.. (1993). *A Teacher's Guide to Classroom Research*. Philadelphia: Open University Press.
- Hopkins, Charles, D. and Antes, Richard L.. (1999). *Classroom Measurement and Evaluation*. Illionis, F.E. Peacock.
- Kunandar. (2011). *Langkah Mudah Penelitian Tindakan Kelas sebagai pengembang Profesi Guru*. Jakarta: PT Raja Grafindo Persada.
- Machali, I. 2015. *Statistik Itu Mudah: Menggunakan SPSS Sebagai Alat bantu Statistik*. Yogyakarta : Ladang Kata.
- McTaggart, R.. (1991). *Action Research: A Short Modern History*. Geelong: Deakin University Press.
- Magdalena, I., Fauziah, S. N., Faziah, S. N., & Nupus, F. S. (2021). Analisis Validitas, Reliabilitas, Tingkat Kesulitan Dan Daya Beda Butir Soal Ujian Akhir Semester Tema 7 Kelas Iii Sdn Karet 1 Sepatan. 3.
- Musayoroh, T. (2021). HOTS-Based Chemical Literacy Test Instrument Validation in Terms of Content Validity by Experts. *UNESA Journal of Chemical Education*, 10(3): 243–251.
- Mutakin, M. D. (2022a). Peningkatan Keaktifan dan Hasil Belajar IPS Menggunakan Aplikasi *Google Form* di Kelas 1G Madrasah Muallimin Muhammadiyah Yogyakarta. *Jurnal Kiprah*, 10(1), 12–22. <https://doi.org/10.31629/kiprah.v10i1.4179>
- Mutakin, M. D. (2022b). Peningkatan Keaktifan dan Hasil Belajar Ilmu Pengetahuan Sosial Siswa menggunakan Metode *Team Quiz* pada Pembelajaran Daring. *Jurnal Penelitian Ilmu-Ilmu Sosial*. 23(1): 1-18.
- Naga, Dali. S. (2002). *Pengantar Teori Skor pada Pengukuran Pendidikan*. Jakarta: Gunadarma.
- Neti. (2020). Analisis Validitas Soal Pilihan Ganda Penilaian Akhir Semester Genap Pelajaran Fisika Kelas XI MIPA 1 SMA Negeri 1 Kepahiang. *Jurnal Perspektif Pendidikan*. LP4MK STKIP PGRI Lubuklinggau. 14(02): 89-98. DOI: <https://doi.org/10.31540/jpp.v14i2.1049>.
- Nuraeni, Z., Simarmata, R. H., Sukmaningthias, N., dan Sari, N. (2021). Pelatihan Software SPSS untuk Menghitung Validitas, Reliabilitas, dan Analisis Butir Soal bagi Mahasiswa Calon Guru di Palembang. *Jurnal Anugerah*, 3(1), 15–23. <https://doi.org/10.31629/anugerah.v3i1.3383>
- Nurgiyantoro, B.. (2010). *Penilaian Pembelajaran Bahasa Berbasis Kompetensi*. Yogyakarta: BPFE.
- Nurkencana, Wayan, Sunartana. (1992). *Evaluasi Hasil Belajar*, Surabaya: Usaha Nasional.

- Rifanty, E. (2019). Peningkatan Keaktifan Belajar Menggunakan Model Pembelajaran Kooperatif Tipe Make A Match Pada Peserta Didik Kelas VB SD Muhammadiyah Condongcatur. *JURNAL JPSD*. 10(10): 1-6. DOI: <http://dx.doi.org/10.26555/jpsd>
- Sanjaya, W.. (2012). *Penelitian Tindakan Kelas*. Jakarta: Kencana Prenada Media Group.
- Santoso, E., Kania, N., Nurhikmayati, I., Jatisunda, M. G., & Suciawati, V. (2021). Pelatihan Penelitian Tindakan Kelas Sebagai Bentuk Pengembangan Profesionalisme Guru. *Jurnal Pengabdian Kepada Masyarakat*, 2(2). DOI: [10.31949/jb.v2i2.832](https://doi.org/10.31949/jb.v2i2.832)
- Sari, W. M., Riswanto, R., & Partono, P. (2019). Validitas mobile pocket book berbasis android menggunakan adobe flash pada materi suhu dan kalor. *Berkala Ilmiah Pendidikan Fisika*, 7(1): 35–42.
- Slameto. (2008). *Evaluasi pendidikan*. Jakarta: Bina Aksara.
- Sudjana,. (2004). *Penilaian Hasil Proses Belajar Mengajar*, Bandung: Remaja Rosdakarya.
- Sudjiono, A. (2016). *Pengantar Evaluasi Pendidikan*, Jakarta: Rajawali Press.
- Sukarsono, Fauzan, Arief N. F., dan Taufik M. (2014). *Modul Pengantar Penulisan Penelitian Tindakan Kelas (PTK). Pendidikan dan Latihan Profesi Guru Panitia Sertifikasi Guru (PSG) Rayon 144*. Diselenggarakan oleh Universitas Muhammadiyah Malang, Universitas Islam Malang, Universitas Wisnuwardhana Malang, dan SKIP PGRI Blitar.
- Sukardi, (2009). *Metodologi penelitian pendidikan: kompetensi dan praktiknya* Jakarta: Bumi Aksara.
- Suryabrata, S. (2007). *Pengembangan Tes Hasil Belajar*. Jakarta: Rajawali.
- Tarigan, E. F., Nilmarito, S., Islamiyah, K., Darmana, A., & Suyanti, R. D. (2022). Analisis Instrumen Tes Menggunakan Rasch Model dan Software SPSS 22.0. *Jurnal Inovasi Pendidikan Kimia*, 16(2), 92–96. <https://doi.org/10.15294/jipk.v16i2.30530>
- Udzma, N. S., & Farid, E. K. (2023). *1,2 Universitas Islam Zainul Hasan Genggong*.
- Utomo, B. (2019). Analisis validitas isi butir soal sebagai salah satu upaya peningkatan kualitas pembelajaran di madrasah berbasis nilai-nilai islam. *Jurnal Pendidikan Matematika (Kudus)*, 1(2).
- Whitehead, J., & McNiff, J. (2006). *Action research living theory*. SAGE Publications.
- Yuliardi, R., & Nuraeni, Z. (2017). *Statiska Penelitian plus tutorial software SPSS*. Yogyakarta: Graha Ilmu.
- Yusuf, I. (2015). Peningkatan Aktivitas dan Hasil Belajar Fisika Melalui Pembelajaran *EMPECE* pada Siswa Kelas XI-IPA 4 SMA Negeri 5 Yogyakarta tahun Pelajaran 2012/2013. *Jurnal Ilmiah Guru "COPE"*. 19(1): 71-78. DOI: <https://doi.org/10.21831/jig%20cope.v19i01.5643>

-
- Yusup, F. (2018). Uji Validitas dan Reliabilitas Instrumen Penelitian Kuantitatif. *Jurnal Tarbiyah : Jurnal Ilmiah Kependidikan*, 7(1). <https://doi.org/10.18592/tarbiyah.v7i1.2100>
- Zainul, A. dan Nasoetion, N. (1997). *Penilaian Hasil Belajar*. Pusat Antar Universitas, Direktorat Jenderal Pendidikan Tinggi: Departemen Pendidikan Dan kebudayaan.