



## **PENGEMBANGAN INSTRUMEN ASSESSMENT OF LEARNING UNTUK MENGUKUR KEMAMPUAN LITERASI SAINS PADA MATERI VIRUS**

Ahmad Malik Akbarudin\*, Jurusan Pendidikan Biologi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Yogyakarta, Indonesia  
Atik Kurniawati, Jurusan Pendidikan Biologi, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Yogyakarta, Indonesia  
\*e-mail: [ahmadmalik.2018@student.uny.ac.id](mailto:ahmadmalik.2018@student.uny.ac.id)

**Abstrak.** Penelitian ini bertujuan untuk mengembangkan instrumen tes dan mengetahui kelayakan produk untuk mengukur kemampuan literasi sains peserta didik kelas X SMA pada materi virus. Penelitian ini merupakan penelitian pengembangan menggunakan metode *Research and Development* dengan model 4-D yang termodifikasi menjadi *Define, Design, dan Develop*. Subjek uji coba merupakan peserta didik kelas X SMAN 2 Bantul tahun 2022/2023 sebanyak 70 orang yang ditentukan dengan teknik random sampling. Instrumen penelitian terdiri dari instrumen tes sebanyak 50 *item* pilihan ganda dan lembar validasi. Analisis data dilakukan melalui validasi muka dan analisis validitas dan reliabilitas secara empiris menggunakan program QUEST dengan pendekatan *Item Response Theory* (IRT) model *Rasch*. Hasil penelitian ini meliputi: (1) pengembangan instrumen tes menggunakan model 4-D termodifikasi sampai *Develop* berdasarkan *framework* PISA 2018 dengan level kognitif HOTS dan (2) instrumen tes yang dikembangkan terbukti memiliki kualitas butir soal yang layak berdasarkan hasil validasi muka dengan kategori layak dan hasil penyelidikan validitas, reliabilitas, dan karakteristik butir pada tingkat kesukaran *item* secara empiris dinyatakan valid dan reliabel untuk mengukur kemampuan literasi sains pada materi virus. Berdasarkan hasil penelitian ini, disimpulkan bahwa instrumen tes kemampuan literasi sains yang dikembangkan dinyatakan layak berdasarkan hasil *expert judgement* dan hasil penyelidikan secara empiris sejumlah 44 *item*.

**Kata Kunci:** *Assessment of learning, Literasi sains, Materi virus*

### **PENDAHULUAN**

Perkembangan sains dan teknologi abad 21 memengaruhi perubahan di berbagai aspek kehidupan, termasuk kebutuhan penguasaan kemampuan 4C abad 21. Kemampuan ini berkaitan dengan kemampuan literasi sains. Rahayuni (2016: 144) menyebutkan bahwa terdapat hubungan positif antara kemampuan berpikir kritis dengan literasi sains. Penguasaan kemampuan literasi sains pada seseorang disebut juga dengan istilah “melek sains”. “Melek sains” adalah kemampuan literasi sains yang terdiri atas kemampuan untuk memahami, mengomunikasikan, serta menerapkan kemampuan sains untuk memecahkan masalah berdasarkan pertimbangan sains (Yuliati, 2017: 22). Literasi sains merupakan salah satu kemampuan literasi dasar yang harus dikuasai oleh setiap orang dalam upaya penguasaan kemampuan sesuai tuntutan abad 21 yang teridentifikasi oleh *World Economic Forum* (World Economic Forum, 2015: 3).

Literasi sains menurut *framework* PISA 2018 didefinisikan sebagai kemampuan untuk terlibat dengan isu-isu yang berhubungan dengan sains dengan melibatkan ide-ide sains sebagai bentuk warga negara yang reflektif. Kemampuan literasi sains membutuhkan tiga

kompetensi utama yaitu menjelaskan fenomena ilmiah, mengevaluasi dan merancang penyelidikan secara ilmiah, dan menafsirkan data dan bukti secara ilmiah (OECD, 2019: 100-101). Hasil nilai literasi sains PISA Indonesia masih tergolong rendah berada di bawah nilai rata-rata negara peserta PISA pada ketiga bidang penilaian, termasuk bidang sains. Nilai rata-rata Indonesia bidang sains PISA 2018 mengalami penurunan dibandingkan PISA 2015 (Kementerian Pendidikan dan Kebudayaan, 2019: 41-42). Hasil penilaian sains PISA Indonesia pada lima tahun terakhir pelaksanaan PISA juga masih berada di bawah rata-rata nilai literasi sains PISA dalam rentang 382 sampai 403. Hal ini menunjukkan bahwa ada kesenjangan dalam pelaksanaan pendidikan sains yang dilakukan di Indonesia dengan negara lainnya.

Salah satu faktor penyebab rendahnya nilai literasi sains pada PISA adalah tidak terbiasannya peserta didik mengerjakan soal yang menggunakan wacana seperti model soal PISA (Angraini, 2014: 167). Hal ini menunjukkan bahwa instrumen penilaian yang digunakan di Indonesia belum valid untuk mengukur kemampuan literasi sains. Sudijono (2011: 6) menyatakan bahwa pengukuran yang baik adalah proses melakukan pengukuran dengan alat ukur yang bersifat valid dan reliabel. Alat ukur bersifat valid dalam arti bahwa hasil pengukuran mencerminkan apa yang diukur dan bersifat reliabel atau konsisten yang diartikan bahwa pengukuran akan menghasilkan skor sama jika dilakukan berulang-ulang.

Proses penyempurnaan kurikulum terus dilakukan oleh Indonesia, diantaranya adalah penyempurnaan standar penilaian yang memfokuskan pada peningkatan kemampuan berpikir tingkat tinggi (HOTS) yang akan mendorong peserta didik untuk berpikir secara luas dan mendalam tentang materi pelajaran (Widana, 2017: 1). Penyempurnaan standar penilaian dilakukan dengan mengadaptasi model model penilaian yang berstandar internasional seperti PISA. Salah satu bentuk penilaian pembelajaran yang dapat dikembangkan adalah *assessment of learning*. Dalam konteks *assessment of learning* diartikan sebagai kegiatan penilaian pembelajaran yang dilakukan setelah selesai mengikuti suatu program pembelajaran yang bertujuan untuk mengetahui keberhasilan peserta didik (Subali, 2006: 164-165).

Instrumen penilaian PISA 2018 disusun berdasarkan *framework* PISA 2018. Kerangka penilaian PISA 2018 merumuskan tiga aspek, yaitu aspek konteks, aspek pengetahuan, dan aspek kompetensi, dengan tiga tuntutan kognitif. Aspek kompetensi yang diukur meliputi kompetensi menjelaskan fenomena secara ilmiah, mengevaluasi dan merancang penyelidikan secara ilmiah, dan menafsirkan data dan bukti secara ilmiah yang dikaitkan dengan aspek kontek dalam PISA meliputi personal, lokal/nasional, dan global mengenai suatu aspek pengetahuan, salah satunya adalah aspek pengetahuan konten (OECD, 2019: 100-112). Ketiga aspek ini dapat dijadikan pedoman perumusan instrumen penilaian kemampuan literasi sains dalam konteks *assessment of learning* dengan penyesuaian terhadap kompetensi dasar yang berlaku yang mengarah pada penilaian kemampuan kognitif menganalisis (C4), mengevaluasi (C5), dan mencipta (C6) atau penilaian HOTS (Krathwohl & Anderson, 2001: 215).

Instrumen objektif yang biasa digunakan dalam penilaian dan juga digunakan dalam instrumen PISA adalah tes objektif berbentuk pilihan ganda. Subali (2016: 79) menyebutkan bahwa terdapat beberapa hal yang perlu diperhatikan dalam menyusun item bentuk pilihan ganda yang efektif dan efisien, diantaranya aspek substansi, konstruksi, dan kebahasaan. Selain memenuhi ketiga aspek tersebut, instrumen tes dapat dikatakan baik apabila telah memenuhi validitas dan reliabilitas item tes melalui penyelidikan secara empiris menggunakan hasil analisis kuantitatif dengan menggunakan suatu program tertentu dan dilakukan setelah adanya uji coba lapangan.

Salah satu materi yang tercantum dalam kurikulum 2013 adalah materi virus. Materi virus adalah salah satu materi yang diujicobakan dalam PISA 2018, yaitu pada aspek pengetahuan konten tentang sistem kehidupan (OECD, 2019: 106-107). Materi virus

tercantum dalam KD Pengetahuan 3.4, yaitu menganalisis struktur, replikasi dan peran virus dalam kehidupan (Kementerian Pendidikan dan Kebudayaan, 2018: 50). KD ini selanjutnya dijabarkan menjadi indikator pembelajaran dan akhirnya menjadi pokok-pokok bahasan dalam tema pembelajaran virus yang mencakup struktur virus, replikasi virus, dan peran virus dalam kehidupan dan seringkali ditambah materi sejarah penemuan virus.

Berdasarkan hasil studi data yang dilakukan sebelumnya, belum ada data penelitian secara khusus menyebutkan tentang pengembangan instrumen penilaian kemampuan literasi sains peserta didik SMA pada pembelajaran biologi materi virus. Keberadaan dan pengaplikasian penilaian literasi sains materi virus menjadi salah satu hal yang penting untuk dikembangkan saat ini dan memiliki potensi untuk meningkatkan peserta terbiasa mengerjakan soal literasi sains biologi. Oleh karena itu, penelitian ini memiliki tujuan, yaitu mengetahui pengembangan model instrumen tes kemampuan literasi sains dan mengetahui kelayakan produk instrumen tes untuk mengukur kemampuan literasi sains peserta didik kelas X SMA materi virus.

## **METODE**

Penelitian ini merupakan penelitian pengembangan instrumen tes untuk mengukur kemampuan literasi sains pada materi virus. Metode dalam penelitian adalah Research and Development (R&D) dengan model 4-D yang dimodifikasi menjadi *Define*, *Design*, dan *Develop* (Thiagarajan et al., 1974: 5). Uji coba dilakukan melalui validasi dari ahli dan praktisi terhadap instrumen tes literasi sains yang dikembangkan kemudian dilanjutkan uji empiris pada subjek uji coba. Subjek uji coba dalam penelitian ini adalah peserta didik SMA N 2 Bantul tahun ajar 2022/2023 sejumlah 70 peserta didik yang sudah mendapatkan pembelajaran pada materi virus sebelum pelaksanaan uji coba. Teknik pengambilan sampel menggunakan random sampling. Pengumpulan data menggunakan instrumen penelitian yang terdiri dari instrumen tes dan lembar validasi. Teknik analisis data dilakukan melalui analisis penilaian kelayakan instrumen dari hasil validasi muka dan analisis data instrumen secara empiris dengan pendekatan *Item Response Theory* (IRT) model Rasch dengan bantuan program QUEST untuk menentukan butir soal yang layak dan tidak layak sesuai model yang digunakan.

## **HASIL DAN PEMBAHASAN**

### **1. Hasil Pengembangan Produk Awal**

Hasil pengembangan produk dari penelitian ini adalah sebuah produk instrumen tes untuk mengukur kemampuan literasi sains peserta didik kelas X SMA pada pembelajaran biologi materi virus. Model pengembangan instrumen tes ini menggunakan model 4-D dengan modifikasi menjadi tiga tahapan, yaitu tahap pendefinisian (*Define*) mencakup analisis awal, analisis aspek dan indikator kemampuan literasi sains, dan analisis kompetensi dan indikator materi virus, tahap perancangan (*Design*) mencakup pemilihan format, penyusunan kisi-kisi instrumen, dan penyusunan butir soal dan pedoman penskoran, dan tahap pengembangan (*Develop*) mencakup validasi muka oleh ahli dan praktisi, uji coba pengembangan produk, dan analisis validitas dan reliabilitas instrumen.

Tahap Desain akan menghasilkan instrumen awal tes literasi sains yang dinamai Draf I yang kemudian dilanjutkan pada tahapan pengembangan yang mencakup validasi ahli dan praktisi untuk mendapatkan masukan agar rancangan instrumen tes yang berkategori layak untuk diujicobakan kepada peserta didik kelas X sebagai subjek uji coba dalam penelitian ini. Hasil revisi dari hasil validasi ahli dinamai Draf II. Sedangkan hasil revisi dari hasil validasi praktisi/guru biologi dinamai Draf III. Konstruksi instrumen butir soal untuk mengukur kemampuan literasi sains pada pembelajaran biologi materi virus yang telah dikembangkan *Draft* I ditunjukkan Tabel 1. Instrumen soal kemampuan literasi sains yang

dikembangkan berbentuk tes pilihan ganda dengan sistem penskoran dikotomus (jika benar skor 1 dan jika salah skor 0).

**Tabel 1.** Distribusi Soal Berdasar Kompetensi, Konteks, dan Jenjang Proses Kognitif

Aspek Kompetensi	Konteks		
	Personal	Lokal/Nasional	Global
Menjelaskan fenomena secara ilmiah	11A, 33A, 34B, 36C, 47B	6A, 8C, 28C, 43B, 45A, 48A	1A, 3A, 10B, 19B, 21A, 23C, 32B
Mengevaluasi dan merancang penyelidikan ilmiah	12C, 35C, 38C	7C, 24C, 26C, 30C, 41C, 49C	2C, 4C, 15C, 16B, 18C
Menafsirkan data dan bukti ilmiah	13B, 37A, 39B, 42A	14C, 25B, 27B, 29B, 40A, 44B, 46A, 50A	5B, 9B, 17A, 20B, 22B, 31A

Keterangan: A: Jenjang proses kognitif C4; B: Jenjang proses kognitif C5, C: Jenjang proses kognitif C6

## 2. Hasil Uji Coba Produk

Tahap validasi ahli dan praktisi dibuktikan dengan penilaian instrumen yang dilakukan dengan kriteria layak digunakan untuk penelitian, layak digunakan dengan perbaikan, dan tidak layak digunakan untuk penelitian. Hasil dari validasi konstruk, materi, kebahasaan instrumen tes ini dinyatakan dalam Tabel 2.

**Tabel 2.** Hasil Validasi Konstruk, Materi, dan Kebahasaan

Instrumen	Konstruksi	Materi	Kebahasaan
Validasi Soal	Layak Digunakan dengan Perbaikan	Layak Digunakan dengan Perbaikan	Layak Digunakan dengan Perbaikan
Validasi Kisi-Kisi Soal	Layak Digunakan dengan Perbaikan	Layak Digunakan dengan Perbaikan	Layak Digunakan dengan Perbaikan
Validasi Pedoman Penskoran	Layak Digunakan dengan Perbaikan	Layak Digunakan dengan Perbaikan	Layak Digunakan dengan Perbaikan

Berdasarkan hasil validasi di atas, instrumen tes yang dikembangkan berdasarkan pemenuhan aspek konstruks, materi, dan kebahasaan terbukti layak digunakan pada semua butir soal dengan perbaikan-perbaikan yang telah dilakukan berdasarkan masukan dan saran dari masing-masing validator. Draf III dinyatakan layak setelah adanya perbaikan-perbaikan untuk digunakan dalam uji coba pengembangan produk. Draf III disusun atas dua tipe instrumen, yaitu instrumen tipe A dan tipe B disesuaikan dengan durasi pengerjaan soal.

### Analisis Hasil Uji Coba Pengembangan Instrumen

Analisis validitas, reliabilitas, dan tingkat kesukaran butir terhadap hasil uji coba pengembangan produk dilakukan menggunakan program QUEST dengan prinsip teori respon butir (Item Respons Theory/IRT) model 1 PL/model Rasch. Analisis validitas, reliabilitas, karakteristik butir hasil uji coba secara empiris dilakukan dua kali. Hal ini menyesuaikan dengan file data hasil uji coba produk yang akan dianalisis yang terdiri dari hasil uji coba instrumen tipe A dan hasil uji coba instrumen tipe B.

Analisis QUEST dilakukan untuk menganalisis tiga indikator, yaitu validitas instrumen berdasarkan besarnya nilai INFIT MNSQ dan Outfit t, reliabilitas instrumen berdasarkan besarnya nilai *item reliability* pada bagaian *Summary of item estimates*, dan karakteristik butir soal berdasarkan tingkat kesukaran butir soal yang ditunjukkan melalui besarnya nilai *Thresholds*. Indikator validitas dan karakteristik butir soal akan menentukan *fit*

atau tidak *fit*-nya butir soal yang dianalisis untuk menentukan butir soal yang valid/sahih. Sedangkan tingkat reliabilitas instrumen soal didasarkan pada besarnya nilai *reliability of item estimates*. Hasil analisis QUEST secara keseluruhan terhadap hasil uji coba pengembangan ditunjukkan dalam Tabel 3.

**Tabel 3.** Hasil Analisis QUEST Instrumen Tes Tipe A dan B secara keseluruhan

Indikator Nilai	Hasil Analisis Instrumen	
	Tipe A	Tipe B
Rentang INFIT MNSQ	0,89 - 1,11	0,85 - 1,11
Rentang Outfit t	-0,46 - 0,54	-0,58 - 0,94
Reliabilitas <i>Item</i>	0,86	0,86
Jumlah item yang berada rentang -2,0 sampai 2,0	23	23
Jumlah <i>item fit</i>	22	22
Jumlah <i>item tidak fit</i>	3	3

#### a. Pembuktian Validitas Instrumen Hasil Uji Coba

Pembuktian tingkat validitas butir soal instrumen tes dilakukan dengan pendekatan IRT dengan model Rasch menggunakan program QUEST. Analisis validitas instrumen tes dilakukan pada instrumen tes tipe A dan instrumen tes tipe B. Setiap butir dinyatakan *fit* dengan Rasch Model jika nilai INFIT MNSQ berada di antara 0,77 sampai dengan 1,30 dan nilai OUTFIT  $t \leq 2$  (Subali, 2006: 140-153).

Berdasarkan hasil analisis QUEST terhadap instrumen tes literasi sains tipe A secara keseluruhan diketahui rentang nilai INFIT MNSQ antara 0,89 sampai 1,11 dan rentang nilai Outfit t antara -0,46 sampai 0,54 sehingga secara keseluruhan item dapat dinyatakan fit dengan model yang digunakan atau fit dengan model Rasch. Namun apabila analisis butir soal didasarkan pada kriteria tingkat validitas dan tingkat kesukaran soal tiap *item* berdasarkan nilai INFIT MNSQ, Outfit t, dan nilai Threshold dihasilkan sejumlah 22 *item fit* dan 3 item tidak fit, yaitu nomor 8 (nilai threshold sebesar -3,23), nomor 20 (nilai threshold sebesar 4,12), dan nomor 16 (nilai INFIT MNSQ sebesar 0,71).

Sedangkan hasil analisis QUEST instrumen tes tipe B menunjukkan bahwa nilai INFIT MNSQ terletak pada rentang 0,85 sampai 1,11 dan rentang nilai Outfit t sebesar -0,58 sampai 0,94 sehingga secara keseluruhan item instrumen tes dinyatakan fit dengan model Rasch. Namun apabila analisis butir soal didasarkan pada tingkat validitas dan tingkat kesukaran item menghasilkan 22 *item fit* dan 3 item sisanya tidak fit, yaitu nomor 7 (nilai Outfit t sebesar 2,4), nomor 19 (nilai threshold sebesar -2,30), dan nomor 23 (nilai threshold sebesar -2,30).

Berdasarkan hasil analisis pada kedua instrumen tes kemampuan literasi sains (Tipe A dan Tipe B), diketahui bahwa sebagian butir soal yang diujicobakan dinyatakan fit dan sebagian lainnya dinyatakan belum fit dengan model Rasch. Hasil analisis QUEST instrumen tes tipe A dan instrumen tipe B dari jumlah keseluruhan item sebanyak 25 soal yang diujicoba masing-masing menghasilkan 22 item fit. Oleh karena itu, jumlah total item yang telah dinyatakan fit dengan model Rasch, yaitu sebanyak 44 butir soal yang selanjutnya dapat dikemas menjadi produk akhir instrumen tes kemampuan literasi sains peserta didik kelas X SMA pada pembelajaran biologi materi virus.

#### b. Estimasi Reliabilitas Instrumen Hasil Uji Coba

Pembuktian nilai estimasi reliabilitas dilakukan dengan menganalisis instrumen tes menggunakan pendekatan IRT dengan model Rasch. Kriteria pemenuhan nilai estimasi reliabilitas instrumen dilakukan berdasarkan koefisien reliabilitas. Menurut Istiyono (2018: 336), yang menyatakan beberapa rentang nilai estimasi reliabilitas ( $r_{11}$ ) beserta kriteria, yaitu

$r_{11} < 0,20$  (sangat rendah),  $0,20 < r_{11} < 0,40$  (rendah),  $0,40 < r_{11} < 0,60$  (sedang),  $0,60 < r_{11} < 0,80$  (tinggi), dan  $0,80 < r_{11} < 1,00$  (sangat Tinggi). Instrumen tes dapat dikatakan semakin baik/reliabel jika nilai estimasi reliabilitas mendekati nilai 1.

Hasil analisis QUEST instrumen tipe A menunjukkan bahwa nilai item reliability sebesar 0,86 (reliabilitas sangat tinggi) dan nilai person reliability sebesar 0,65 (reliabilitas sedang). Hal ini menunjukkan bahwa reliabilitas instrumen tes yang dikembangkan dilihat dari reliabilitas item dan reliabilitas subjek uji coba memiliki kategori sangat tinggi dan sedang secara berurutan sehingga baik item maupun subjek uji coba dinyatakan reliabel. Sedangkan, hasil analisis QUEST instrumen tipe B menunjukkan bahwa nilai item reliability sebesar 0,86 (reliabilitas sangat tinggi) dan nilai person reliability sebesar 0,61 (reliabilitas tinggi). Hal ini menunjukkan bahwa reliabilitas instrumen tes yang dikembangkan dilihat dari reliabilitas item dan reliabilitas subjek uji coba memiliki kategori sangat tinggi dan sedang secara berurutan sehingga baik item maupun subjek uji coba dinyatakan reliabel.

Berdasarkan hasil analisis pada kedua instrumen tes kemampuan literasi sains (Tipe A dan Tipe B), diketahui bahwa nilai reliabilitas instrumen tes yang dikembangkan termasuk dalam kategori sangat tinggi pada aspek reliabilitas item dan apabila dilihat dari nilai reliabilitas subjek uji coba menunjukkan bahwa sampel penelitian yang mengerjakan soal tipe A maupun soal tipe B memiliki konsistensi jawaban berkategori sedang dan sangat tinggi. Oleh karena itu, instrumen tes kemampuan literasi sains peserta didik kelas X SMA pada pembelajaran biologi materi virus memiliki reliabilitas item yang sangat tinggi baik yang menunjukkan bahwa instrumen tes dapat digunakan berulang-ulang kali.

### **c. Karakteristik Butir Soal Berdasar Tingkat Kesukaran Butir**

Analisis instrumen tes yang digunakan dalam penelitian ini adalah analisis butir soal menggunakan pendekatan IRT dengan model 1 Parameter Logistik (1PL) atau model Rasch. Model Rasch merupakan model analisis butir soal menggunakan satu parameter logistik, yaitu parameter tingkat kesukaran soal (item difficulty), sedangkan untuk parameter lainnya seperti parameter daya pembeda soal dianggap sama dan tebakan semu = 0 (Andayani et al., 2019: 39). Model karakteristik butir berbentuk satu parameter (1P) dirumuskan, yaitu  $1P: P(\theta) = f(b, \theta)$  di mana parameter  $\theta$  adalah parameter kemampuan responden. Parameter  $b$  adalah parameter taraf sukar butir (Sudaryono, 2011: 723). Berdasarkan pendekatan IRT model Rasch, maka karakteristik butir soal ditentukan melalui besarnya tingkat kesukaran butir soal/item ( $b$ ).

Suatu item dikatakan baik apabila memiliki tingkat kesukaran item antara -2,0 sampai +2,0 (Pratama, 2020: 64). Apabila terdapat item yang memiliki tingkat kesukaran  $< -2,0$ , maka tingkat kesukaran item tersebut dinyatakan sangat mudah dijawab benar dan jika terdapat item yang memiliki tingkat kesukaran item  $> 2,0$ , maka butir soal tersebut sangat sulit dijawab dengan benar.

Berdasarkan hasil analisis QUEST instrumen tes tipe A yang didasarkan pada tingkat kesukaran *item*, dapat diketahui bahwa *item* yang sesuai kriteria rentang dihasilkan sebanyak 23 *item*. Apabila penentuan *item fit* didasarkan pada hasil analisis tingkat kesukaran item dan validitas item, maka jumlah item yang fit dengan model Rasch sebanyak 22 *item fit*. Sedangkan, hasil analisis instrumen tes tipe B yang didasarkan pada kriteria rentang tingkat kesukaran item dihasilkan sebanyak 23 *item fit*. Jika penentuan item fit didasarkan pada tingkat kesukaran item dan validitas item, maka dihasilkan sebanyak 22 *item fit*. Oleh karena itu berdasarkan jumlah *item fit* secara keseluruhan akan dihasilkan sebanyak 44 *item fit* yang akan dilanjutkan pada tahap perakitan instrumen tes kemampuan literasi sains.

## **3. Revisi Produk**

Revisi produk dalam penelitian ini dilakukan melalui dua tahapan, yaitu tahap revisi instrumen tes berdasarkan hasil validasi isi/muka secara rasional oleh para validator dan

tahap revisi instrumen tes berdasarkan hasil analisis uji coba pengembangan produk. Revisi produk bertujuan untuk menghasilkan produk akhir instrumen tes yang valid dan reliabel.

### **1. Revisi Instrumen Tes Berdasarkan Hasil Validasi Muka**

Instrumen tes kemampuan literasi sains (Draft I) dilakukan proses validasi muka oleh empat ahli dan dua guru biologi SMA. Hasil revisi produk yang dilakukan berdasarkan masukan para ahli disebut Draft II dan dilanjutkan validasi guru yang akan dilakukan revisi menghasilkan instrumen tes yang layak untuk digunakan pada uji coba produk kepada peserta didik disebut Draft III.

Secara keseluruhan hasil validasi dari para ahli (ahli penilaian dan ahli ilmu biologi) dan praktisi/guru biologi SMA menyatakan bahwa instrumen tes yang dikembangkan (Draft I) telah layak diujicobakan kepada peserta didik tetapi diperlukan adanya beberapa revisi. Komentar dan masukan yang diberikan oleh validator baik ahli maupun praktisi secara keseluruhan berkaitan tentang aspek materi item, aspek konstruksi item, dan aspek kebahasaan pada seluruh item yang telah dikembangkan. Beberapa masukan dan saran yang diberikan oleh validator, meliputi:

- a. Terdapat lebih dari satu jawaban benar pada butir soal nomor 20, 22, 24, 38, 39, 41, 48.
- b. Ketidaksesuaian butir soal dengan indikator butir soal pada butir soal nomor 5, 28.
- c. Ketidaksesuaian butir soal dengan indikator kompetensi sains pada butir soal nomor 16.
- d. Rumusan soal dan pilihan jawaban tidak dirumuskan dengan benar pada butir soal nomor 4, 7, 10, 11, 12, 14, 16, 19, 24, 25, 27, 30, 32, 33, 34, 42, 43, 44, 45, 49, dan 50.
- e. Isi grafik tidak berkaitan dengan rumusan soal pada butir soal nomor 42.
- f. Penggunaan bahasa tidak menggunakan bahasa baku pada butir soal nomor 3, 34, 37, 45.
- g. Penggunaan kalimat yang tidak komunikatif pada butir soal nomor 15, 17, 20, 21.

Revisi item dilakukan berulang kali mulai dari validasi ahli kemudian dilanjutkan validasi guru biologi SMA. Revisi instrumen tes dilakukan berdasarkan masukan-masukan dari validator yang mencakup isi pada aspek materi, aspek konstruksi, dan aspek kebahasaan.

### **2. Revisi Instrumen Tes Berdasarkan Analisis Hasil Uji Coba**

Hasil revisi dari tahapan validasi muka menghasilkan instrumen tes kemampuan literasi sains yang layak digunakan dalam uji coba pengembangan produk berjumlah 50 item berupa Draft III. Kegiatan uji coba pengembangan produk dilakukan kepada peserta didik kelas X SMAN 2 Bantul sejumlah 70 orang dengan memmbagi instrumen tes menjadi dua tipe, yaitu tipe A berjumlah 25 item dan tipe B berjumlah 25 item. Setiap tipe instrumen tes dikerjakan minimal 30 peserta didik, yaitu instrumen tes tipe A dikerjakan 36 peserta didik dan instrumen tes tipe B dikerjakan 34 peserta didik. Hal ini sesuai dengan pernyataan Gay (2012: 139) bahwa ukuran sampel minimal adalah 30 peserta dan Sugiyono (2017: 91) menyatakan bahwa ukuran sampel yang layak dalam penelitian adalah antara 30 sampai 50 responden.

Hasil uji coba pengembangan produk ini dilakukan pendataan setiap jawaban testi yang disandingkan dengan masing-masing kunci jawaban. Hasil jawaban peserta didik tersebut dilakukan proses analisis validitas, reliabilitas, dan tingkat kesukaran item secara empiris dengan pendekatan IRT menggunakan model Rasch dengan bantuan program QUEST. Berdasarkan hasil analisis QUEST dapat diketahui bahwa instrumen tes yang dikembangkan memiliki reliabilitas item termasuk dalam kriteria sangat baik. Item yang tidak lolos atau tidak memenuhi standar validitas dan tingkat kesukaran item dilakukan modifikasi dengan cara dibuang kemudian dilakukan perakitan produk instrumen tes dengan menghimpun item yang memenuhi standar validitas, reliabilitas, dan tingkat kesukaran item baik dari instrumen tes tipe A maupun instrumen tes tipe B. Hasil penyusunan instrumen tes

yang telah dinyatakan valid dan reliabel berdasarkan hasil validitas muka dan hasil penyelidikan validitas, reliabilitas, dan tingkat kesukaran item secara empiris merupakan produk akhir instrumen tes yang dinyatakan valid dan reliabel untuk mengukur kemampuan literasi sains peserta didik kelas X SMA pada pembelajaran biologi materi virus berupa butir soal pilihan ganda dengan lima pilihan jawaban sejumlah 44 item.

#### **4. Kajian Produk Akhir**

Penelitian ini menghasilkan produk akhir berupa instrumen tes untuk mengukur kemampuan literasi sains peserta didik kelas X SMA pada pembelajaran biologi materi virus. Instrumen tes yang dimaksud meliputi instrumen soal, kisi-kisi soal, kunci jawaban, dan pedoman penskoran. Instrumen soal ini telah dilakukan dua tahap pengembangan, yaitu tahap validasi muka oleh para validator ahli dan praktisi dan tahap penyelidikan validitas, reliabilitas, dan tingkat kesukaran item secara empiris. Instrumen soal juga telah dilakukan uji coba kepada 70 peserta didik kelas X SMAN 2 Bantul.

##### **Validitas Muka**

Instrumen tes dilakukan uji coba apabila telah dinyatakan layak untuk digunakan peserta didik saat tahapan validasi oleh expert judgement dari ahli maupun guru biologi SMA. Analisis hasil validasi ahli dan guru dilakukan dengan cara menganalisis melalui analisis deskriptif kualitatif berdasarkan masukan-masukan validator dan kesimpulan validasi instrumen tes tersebut. Berdasarkan analisis hasil validasi muka dari keseluruhan validator menyatakan bahwa instrumen tes yang telah dikembangkan layak digunakan uji coba dengan beberapa perbaikan di dalamnya, yang mencakup perbaikan dari aspek materi, aspek konstruksi, maupun kebahasaan.

##### **Validitas Item**

Proses penyelidikan validitas item dilakukan berdasarkan hasil jawaban yang didapatkan dari keseluruhan sampel penelitian sejumlah 70 peserta didik saat uji coba pengembangan produk dilakukan. Pembuktian tingkat validitas butir soal terhadap instrumen tes yang telah dikembangkan dan diujicoba dilakukan melalui penyelidikan validitas item secara empiris dengan pendekatan IRT dengan model Rasch menggunakan bantuan program QUEST. Nilai validitas item pada hasil analisis QUEST akan menunjukkan item mana saja yang memenuhi standar (item fit) dan item yang belum memenuhi standar (item belum fit).

Penentuan valid tidaknya suatu item dilakukan berdasarkan rentang nilai validitas yang disajikan dalam nilai INFIT MNSQ dan Outfit t. Setiap butir dinyatakan fit (item fit) sesuai dengan Rasch Model jika nilai INFIT MNSQ berada di antara 0,77 sampai dengan 1,30 dan nilai OUTFIT  $t \leq 2$  (Subali, 2006: 140). Apabila suatu item memenuhi standar tersebut, maka item telah dinyatakan valid/sahih. Item fit merupakan item yang telah sesuai model Rasch yang digunakan atau dengan kata lain item tersebut telah sesuai dengan variabel kemampuan yang seharusnya diukur.

##### **Estimasi Reliabilitas Instrumen Hasil Uji Coba**

Penentuan kategori reliabilitas berpedoman pada kriteria koefisien reliabilitas (Istiyono, 2018: 336), yang menyatakan beberapa rentang nilai estimasi reliabilitas ( $r_{11}$ ) beserta kriteria masing-masing, yaitu  $r_{11} < 0,20$  (sangat rendah),  $0,20 < r_{11} < 0,40$  (rendah),  $0,40 < r_{11} < 0,60$  (sedang),  $0,60 < r_{11} < 0,80$  (tinggi), dan  $0,80 < r_{11} < 1,00$  (sangat tinggi). Hasil analisis QUEST instrumen soal menunjukkan bahwa rata-rata nilai item reliability sebesar 0,87 (reliabilitas sangat tinggi). Hal ini menunjukkan bahwa estimasi reliabilitas instrumen soal yang dikembangkan dinyatakan reliabel. Instrumen soal yang reliabel akan menunjukkan hasil yang sama dalam pengukuran walaupun dilakukan berulang-ulang kali.

##### **Karakteristik Butir Soal**

Suatu item dikatakan baik apabila memiliki tingkat kesukaran item antara -2,0 sampai +2,0 dengan ketentuan:  $b > 2$  (Sangat Sukar),  $1 < b \leq 2$  (Sukar),  $-1 \leq b \leq 1$  (Sedang),  $-2 \leq b < -1$  (Mudah), dan  $b < -2$  (Sangat Mudah) (Pratama, 2020: 64). Item yang memiliki kriteria sangat sulit maupun sangat mudah akan dibuang atau tidak digunakan menjadi produk akhir instrumen tes kemampuan literasi sains. Berdasarkan hasil analisis tingkat kesukaran item dapat diketahui bahwa item yang sesuai standar tingkat kesukaran item, yaitu sejumlah 23 item tipe A dan 23 item tipe B sehingga dihasilkan item yang baik sejumlah 46 item yang lolos dengan kriteria item yang termasuk sukar, item yang termasuk sedang, dan item yang termasuk mudah.

Contoh *item* yang memiliki tingkat kesukaran *item* yang tergolong sukar yang berada pada rentang nilai 1 sampai 2 dapat dilihat pada Tabel 4 berikut ini.

**Tabel 4.** Contoh Soal dengan Tingkat Kesukaran Sukar

Contoh soal dengan tingkat kesukaran sukar	Penjelasan
<p><b>Wacana 11 digunakan untuk soal nomor 21 dan 22</b></p> <p><b>Kasus HIV/AIDS di Indonesia sampai Tahun 2019</b></p> <p>Meskipun jumlah kasus HIV/AIDS di Indonesia fluktuatif, data kasus HIV/AIDS di Indonesia cenderung meningkat dari tahun ke tahun. HIV dapat ditularkan melalui pertukaran berbagai cairan tubuh dari orang yang terinfeksi, seperti darah, Air Susu Ibu (ASI), semen dan cairan vagina. HIV juga dapat ditularkan dari seorang ibu ke anaknya selama kehamilan dan persalinan. Virus dapat memasuki tubuh melalui pelapis mukus dari vagina, vulva, penis, atau rektum selama berhubungan intim atau melalui mulut selama seks oral. Hubungan seks yang tak terlindungi (tanpa kondom) dan transmisi melalui jarum yang terinfeksi-HIV bertanggung jawab terhadap hampir semua infeksi HIV. Sosialisasi tentang pengetahuan transmisi HIV terus digalakkan, salah satunya melalui beberapa kampanye pencegahan virus HIV/AIDS untuk meningkatkan kesadaran masyarakat terkait transmisi HIV.</p> <p><b>Grafik Jumlah Kasus HIV/AIDS di Indonesia</b></p>  <p>(Diện P2P (Sistem Informasi HIV/AIDS dan IMS (SIHA), 2019)</p> <p>21. Manakah pertanyaan yang mungkin diselidiki berdasarkan perolehan data pada grafik?</p> <p>A. Bagaimana pengaruh perilaku seks menyimpang terhadap tingginya kasus HIV?</p> <p>B. Apa faktor yang paling dominan yang memengaruhi tingginya kasus HIV/AIDS?</p> <p>C. Bagaimana pengaruh sosialisasi pencegahan HIV terhadap penurunan kasus HIV?</p> <p>D. Apa saja faktor yang memengaruhi terjadinya kecukupan jumlah kasus HIV/AIDS?</p> <p>E. Bagaimana keterkaitan antara jumlah kasus HIV dengan jumlah kasus AIDS?</p>	<p>Butir soal nomor 21 produk akhir merupakan soal pilihan ganda dengan konteks nasional. Pada <i>item</i> ini disajikan sebuah wacana tentang kasus HIV/AIDS di Indonesia tahun 2019 dengan rumusan soal untuk mengukur kompetensi mengevaluasi dan merancang penyelidikan ilmiah</p>

Contoh *item* yang memiliki tingkat kesukaran *item* yang tergolong sedang yang berada pada rentang nilai -1 sampai 1 dapat dilihat pada Tabel 5 berikut ini.

**Tabel 5.** Contoh Soal dengan Tingkat Kesukaran Sedang

Contoh soal dengan tingkat kesukaran sukar	Penjelasan
<p><b>Wacana 5 digunakan untuk soal nomor 11 dan 12</b></p> <p><b>Penyakit Rabies</b></p> <p>Penyakit rabies merupakan satu dari sekian penyakit virus manusia yang diakibatkan oleh peristiwa zoonosis. Rabies adalah nama penyakit yang menyebabkan peradangan progresif dan fatal pada otak dan sumsum tulang belakang. Transmisi virus ini banyak diakibatkan oleh anjing dan kelelawar melalui proses zoonosis. Zoonosis merupakan penyakit infeksi yang penularannya dari hewan ke manusia. Peneliti Ahli Muda, Pusat Riset Kesehatan Masyarakat dan Giri, Badan Riset dan Inovasi Nasional (BRIN), Pandji Wibawa Dheewantara menjelaskan bahwa penularan zoonosis bisa melalui kontak langsung atau tidak langsung dengan hewan tersebut. Kementerian Kesehatan mencatat saat ini terdapat sekitar 150 penyakit zoonosis yang mengancam Indonesia, seperti <i>Avian Influenza (AI)</i> atau flu burung, rabies, dan antraks. Kemunculan tiga penyakit tersebut dipicu berbagai faktor, seperti arus globalisasi, deforestasi (pengundulan hutan), perubahan iklim, serta peningkatan arus urbanisasi dan populasi manusia. (Sumber: <a href="https://www.wika.int/news-room/fact-sheets/detail/rabies">https://www.wika.int/news-room/fact-sheets/detail/rabies</a>)</p> <p>11. Berikut ini adalah perilaku manusia yang mendorong transmisi virus rabies semakin meluas, kecuali ...</p> <p>A. memperjualbelikan hasil perburuan hewan liar</p> <p>B. memelihara berbagai jenis hewan ternak</p> <p>C. melakukan eksploitasi hutan-butan alami</p> <p>D. melakukan pembakaran ladang dan hutan</p> <p>E. melepaskan anjing tanpa pengawasan</p>	<p>Butir soal nomor 11 produk akhir merupakan soal dengan konteks global. Pada <i>item</i> ini disajikan teks tentang penyakit rabies dengan tujuan untuk mengukur kompetensi menjelaskan fenomena secara ilmiah dengan membuat dan memilih pernyataan yang tepat</p>

Beberapa contoh *item* yang memiliki tingkat kesukaran *item* yang tergolong mudah yang berada pada rentang nilai -2 sampai -1 dapat dilihat pada Tabel 6 berikut ini.

**Tabel 6.** Contoh Soal dengan Tingkat Kesukaran Mudah

Contoh soal dengan tingkat kesukaran sukar	Penjelasan
<p>Wacana 12 digunakan untuk soal nomor 23 dan 24</p> <p style="text-align: center;"><b>Percobaan Penemuan Virus</b></p> <p>Konsep infeksi virus mulai diperkenalkan setelah munculnya penyakit mosaik pada tembakau yang menghambat pertumbuhan dan daun menjadi bercak-bercak. Salah satu peneliti yang melakukan percobaan tentang penyakit mosaik tembakau adalah Martinus Beijerinck. Pada tahun 1898 Martinus Beijerinck melakukan penelitian mengenai <i>Tobacco Mosaic Virus</i> (TMV). Eksperimen yang dilakukan Beijerinck merupakan serangkaian percobaan klasik dengan proses penyaringan ekstrak daun tembakau yang terinfeksi secara berulang-ulang menggunakan sebuah filter yang didesain untuk menangkap agen penyakit mosaik yang diduga adalah bakteri. Hasil eksperimennya menunjukkan bahwa agen penginfeksi mosaik dianggap sebagai partikel yang dapat bereplikasi dan setiap penyemprotan hasil filter tersebut berdampak pada munculnya mosaik pada tembakau dengan tingkat keparahan yang sama. Selain itu, agen penginfeksi ini tidak dapat dibiakkan pada media nutrisi. Hasil eksperimen ini menjadi dasar pengembangan konsep virus. (Sumber: Reece, dkk. 2010: 413).</p> <p>23. Manakah pernyataan yang sesuai dengan teks di atas?</p> <p>A. Hasil penelitian Beijerinck menjadi solusi pencegahan infeksi penyakit mosaik tembakau. B. Hasil penelitian Beijerinck menjadi awal mula berkembangnya penelitian tentang virus. C. Hasil penelitian Beijerinck menjadi awal mula berkembangnya konsep kultur virus. D. Hasil penelitian Beijerinck menjadi perintis berkembangnya konsep replikasi patogen. E. Hasil penelitian Beijerinck menjadi awal mula adanya konsep replikasi partikel virus.</p>	<p>Butir soal nomor 23 produk akhir merupakan soal dengan konteks global. Pada <i>item</i> ini disajikan wacana tentang percobaan penemuan virus dengan tujuan untuk mengukur kompetensi menjelaskan fenomena secara ilmiah dengan menerapkan pengetahuan ilmiah yang sesuai dengan teks yang diberikan</p>

Berdasarkan keseluruhan hasil analisis yang meliputi validasi muka dan pembuktian validitas *item*, reliabilitas *item*, dan tingkat kesukaran *item*, maka dihasilkan *item fit* sejumlah 44 *item*. Produk akhir ini tersusun dari 22 *item fit* instrumen tipe A dan 22 *item fit* instrumen tipe B.

## SIMPULAN

Pengembangan instrumen tes untuk mengukur kemampuan literasi sains kelas X SMA materi virus dilakukan dengan menggunakan model 4-D dengan modifikasi yang terdiri dari tahap pendefinisian (analisis awal, analisis aspek dan indikator kemampuan literasi sains, dan analisis kompetensi dasar dan indikator materi virus), tahap perancangan (pemilihan format, penyusunan kisi-kisi instrumen tes, penyusunan butir soal dan pedoman penskoran), dan tahap pengembangan (validasi ahli dan praktisi, uji coba pengembangan produk, serta analisis validitas dan reliabilitas instrumen secara empiris). Instrumen tes yang dikembangkan disusun berdasarkan aspek konten, aspek konteks, dan aspek kompetensi literasi sains dengan menggunakan jenjang proses kognitif HOTS. Instrumen tes yang dikembangkan terbukti memiliki kualitas butir soal yang layak berdasarkan hasil validasi muka dan hasil penyelidikan validitas, reliabilitas, dan karakteristik butir pada tingkat kesukaran item secara empiris. Berdasarkan karakteristik butir soal, validitas, dan reliabilitas tersebut menghasilkan produk akhir sejumlah 44 *item fit* berbentuk pilihan ganda dengan lima pilihan jawaban.

## DAFTAR PUSTAKA

- Andayani, A., Purwanto, & Ramalis, T. R. (2019). Kajian Implementasi Teori Respon Butir dalam Menganalisis Instrumen Tes Materi Fisika. *Prosiding Seminar Nasional Fisika* 5.0, 1(1), 37–42.
- Angraini, G. (2014). Analisis Kemampuan Literasi Sains Siswa SMA Kelas X di Kota Solok. *Prosiding Mathematics and Sciences Forum 2014*, 161–169.
- Gay, L. R., Mills, Geoffrey E., & Airasian, P. W. (2012). *Educational Research Competencies for Analysis and Applications*. United States of America: Pearson.
- Istiyono, Edi. (2018). *Pengembangan Instrumen Penilaian dan Analisis Hasil Belajar Fisika dengan Teori Klasik dan Modern*. Yogyakarta: Universitas Negeri Yogyakarta Press.
- Kementerian Pendidikan dan Kebudayaan. (2018). *Peraturan Menteri Pendidikan dan*

*Kebudayaan RI Nomor 37 tahun 2018 tentang Perubahan atas Permendikbud Nomor 24 tahun 2016 tentang Kompetensi Inti dan Kompetensi Dasar Pelajaran pada Kurikulum 2013 pada Pendidikan Dasar dan Pendidikan Menengah.*

- Kementerian Pendidikan dan Kebudayaan. (2019). *Pendidikan di Indonesia Belajar dari Hasil PISA 2018*. Jakarta: Pusat Penilaian Pendidikan Balitbang Kemendikbud.
- Krathwohl, D. R., & Anderson, L. W. (2001). A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives. *Theory into Practice*, 41(4), 212–219..
- OECD. (2019). *PISA 2018 Assessment and Analytical Framework*. Paris: OECD Publishing.
- Pratama, D. (2020). Analisis Kualitas Tes Buatan Guru Melalui Pendekatan Item Response Theory (IRT) Model Rasch. *Tarbawy : Jurnal Pendidikan Islam*, 7(1), 61–70
- Rahayuni, G. (2016). Hubungan Keterampilan Berpikir Kritis dan Literasi Sains pada Pembelajaran IPA Terpadu dengan Model PBM dan STM. *Jurnal Penelitian dan Pembelajaran IPA*, 2(2), 131.
- Subali, Bambang. (2016). *Prinsip Asesmen dan Evaluasi Pembelajaran*. (2<sup>nd</sup> ed.). Yogyakarta: Universitas Negeri Yogyakarta Press.
- Sudaryono. (2011). Implementasi Teori Responsi Butir (Item Response Theory) Pada Penilaian Hasil Belajar Akhir di Sekolah. *Jurnal Pendidikan dan Kebudayaan*, 17(6), 719–732.
- Thiagarajan, S., Semmel, D. S., & Semmel, M. I. (1974). *Instructional Development for Training Teachers of Exceptional Children*. Indiana: Indiana University.
- Widana, W. (2017). *Modul Penyusunan Soal High Order Thinking Skill (HOTS)*. Jakarta: Direktorat Pembinaan Sekolah Menengah Atas Direktorat Jenderal Pendidikan Dasar dan Menengah Kementerian Pendidikan dan Kebudayaan.
- World Economic Forum. (2015). Unlocking The Potential of The Smart Grid. *AIP Conference Proceedings* (Vol. 1702).
- Yuyu Yuliati. (2017). Literasi Sains dalam Pembelajaran IPA. *Jurnal Cakrawala Pendas*, 3(2), 21–28.