

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

A. Hasil Penelitian

1. Gambaran Obyek Penelitian

Pengujian sampel penelitian dilakukan di dua madrasah. Pertama, di MA NU Al-Hidayah Kudus yang beralamat di Jl. Desa Getassrabi No. 1, Kecamatan Gebog, Kabupaten Kudus. MA NU Al-Hidayah merupakan salah satu MA swasta di Kudus yang dikelola oleh Yayasan Pendidikan Islam Manafiuul Ulum. MA NU Al-Hidayah didirikan pada tanggal 1 Juli 1986 oleh KH. Ali As'ad, K. Ali Muzammil, KH. Ibrahim Kholili., KH. Ahmad Hadi, dan Imam Supardi. MA NU Al-Hidayah berdiri diatas tanah seluas $\pm 1750 \text{ m}^2$ di desa Getassrabi, Gebog Kudus.

Kedua, pengambilan sampel dilaksanakan di MA NU Raudlatus Shibyan yang beralamat di Desa Peganjaran Gang 02, Kecamatan Bae, Kabupaten Kudus. MA NU Raudlatus Shibyan didirikan pada tanggal 28 April 2015 dan diresmikan oleh KH. Sya'roni Ahmadi dan H. Agus Sofwan Hadi, S.H.. MA NU Raudlatus Shibyan berdiri dibawah naungan BPPPMNU Raudlatus Shibyan dalam naungan LP Ma'arif NU. Kurikulum yang digunakan merupakan kurikulum terpadu dengan sistem manajemen modern yang disertai keterampilan kerja otomotif dan tata busana.

2. Analisis Data

a. Validitas

Analisis validitas soal pada penelitian ini meliputi validitas isi, validitas konstruk, dan validitas statistik. Perhitungan validitas isi dan validitas konstruk terdiri dari aspek materi, konstruksi, dan bahasa. Penilaian dilakukan oleh dua orang *judgement* ahli, yaitu Iseu Laelasari, M.Pd. dan Siti Rahmawati, S.Pd.. Hasil penilaian dihitung menggunakan rumus Aiken V sebagai berikut.

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

Keterangan :

s = $r - l_0$

l₀ = angka penilaian validitas terendah

c = angka penilaian validitas tertinggi

r = angka yang diberikan oleh penilai

n = jumlah ahli yang menilai

Hasil perhitungan untuk masing-masing tipe soal disajikan dalam tabel sebagai berikut.

Tabel 4. 1. Perhitungan Validitas Isi dan Konstruk Soal Tipe
A

No. Soal	$\sum s$	V	Kategori
1	37	0.68	Tinggi
2	34	0.62	Tinggi
3	36	0.66	Tinggi
4	40	0.74	Tinggi
5	38	0.70	Tinggi
6	41	0.75	Tinggi
7	38	0.70	Tinggi
8	41	0.75	Tinggi
9	31	0.57	Cukup
10	37	0.68	Tinggi
11	40	0.74	Tinggi
12	35	0.64	Tinggi
13	41	0.75	Tinggi
14	36	0.66	Tinggi
15	37	0.68	Tinggi
Jumlah	562	0.69	Tinggi

Tabel 4. 2. Perhitungan Validitas Isi dan Konstruk Soal Tipe
B

No. Soal	$\sum s$	V	Kategori
1	28	0.51	Cukup
2	36	0.66	Tinggi
3	28	0.51	Cukup
4	24	0.44	Cukup
5	36	0.66	Tinggi
6	35	0.64	Tinggi
7	41	0.75	Tinggi
8	36	0.66	Tinggi
9	36	0.66	Tinggi
10	36	0.66	Tinggi
11	42	0.77	Tinggi
12	33	0.61	Tinggi
13	37	0.68	Tinggi
14	39	0.72	Tinggi
15	37	0.68	Tinggi
16	38	0.70	Tinggi

17	42	0.77	Tinggi
18	41	0.75	Tinggi
19	38	0.70	Tinggi
20	32	0.59	Cukup
Jumlah	715	0.66	Tinggi

Tabel 4. 3. Perhitungan Validitas Isi dan Konstruk Soal Tipe C

No. Soal	Σs	V	Kategori
1	37	0.68	Tinggi
2	42	0.77	Tinggi
3	42	0.77	Tinggi
4	37	0.68	Tinggi
5	38	0.70	Tinggi
6	42	0.77	Tinggi
7	42	0.77	Tinggi
8	36	0.66	Tinggi
9	37	0.68	Cukup
10	38	0.70	Tinggi
11	35	0.64	Tinggi
12	33	0.61	Tinggi
13	42	0.77	Tinggi
14	29	0.53	Cukup
15	37	0.68	Tinggi
16	20	0.37	Rendah
17	42	0.77	Tinggi
Jumlah	629	0.68	Tinggi

Sedangkan perhitungan validitas statistik dilakukan berdasarkan hasil yang didapat dari uji coba soal terhadap 100 siswa. Perhitungan ini dilakukan dengan bantuan aplikasi Anates 4.0.9. Hasil perhitungan dikategorikan ke dalam batas signifikansi dengan derajat kebebasan $N-2$.

Tipe soal A memiliki $N=15$. Jika dilihat pada r_{tabel} dengan taraf signifikansi 5%, maka diperoleh nilai 0.482. Persentase analisis validitas tipe soal A disajikan dalam tabel berikut.

Tabel 4. 4. Persentase Analisis Validitas Soal Tipe A

No	Kategori	No. Soal	%
1	Sangat rendah (0,00 – 0,20)	6, 13	13.3%

2	Rendah (0,21 – 0,40)	4, 7, 10, 11, 14	33,3%
3	Cukup (0,41 – 0,60)	1, 2, 3, 5, 8, 9, 12, 15	53,4%
4	Tinggi (0,61 – 0,80)	-	0%
5	Sangat tinggi (0,81 – 1)	-	0%

Tipe soal B memiliki $N=20$. Jika dilihat pada r_{tabel} dengan taraf signifikansi 5%, maka diperoleh nilai 0.423. Persentase analisis validitas tipe soal B disajikan dalam tabel berikut.

Tabel 4. 5. Persentase Analisis Validitas Soal Tipe B

No	Kategori	No. Soal	%
1	Sangat rendah (0,00 – 0,20)	3, 5, 11	15%
2	Rendah (0,21 – 0,40)	1, 4, 6, 14, 19	25%
3	Cukup (0,41 – 0,60)	2, 7, 8, 9, 10, 12, 13, 15, 18, 20	50%
4	Tinggi (0,61 – 0,80)	16, 17	10%
5	Sangat tinggi (0,81 – 1)	-	0%

Tipe soal C memiliki $N=17$. Jika dilihat pada r_{tabel} dengan taraf signifikansi 5%, maka diperoleh nilai 0.423. Persentase analisis validitas tipe soal C disajikan dalam tabel berikut.

Tabel 4. 6. Persentase Analisis Validitas Soal Tipe C

No	Kategori	No. Soal	%
1	Sangat rendah (0,00 – 0,20)	16	5,9%
2	Rendah (0,21 – 0,40)	4, 5, 8, 13, 14	29,4%
3	Cukup (0,41 – 0,60)	1, 2, 3, 6, 7, 9, 10, 11, 12, 15, 17	64,7%
4	Tinggi (0,61 – 0,80)	-	0%
5	Sangat tinggi (0,81 – 1)	-	0%

b. Reliabilitas

Analisis butir soal sistem reproduksi manusia ini dilakukan dengan menggunakan aplikasi Anates 4.0.9. Hasil perhitungan reliabilitas masing-masing tipe soal disajikan dalam tabel berikut.

Tabel 4. 7. Hasil Reliabilitas Soal Tipe A

Rata-rata	Simpangan Baku	Korelasi XY	Reliabilitas Tes
11.24	2.45	0.47	0.64

Tabel 4. 8. Hasil Reliabilitas Soal Tipe B

Rata-rata	Simpangan Baku	Korelasi XY	Reliabilitas Tes
15.29	3.47	0.61	0.75

Tabel 4. 9. Hasil Reliabilitas Soal Tipe C

Rata-rata	Simpangan Baku	Korelasi XY	Reliabilitas Tes
13.26	2.75	0.53	0.69

c. Tingkat Kesukaran

Tingkat kesukaran dari 3 tipe latihan soal sistem reproduksi manusia dianalisis menggunakan aplikasi Anates 4.0.9. Hasil persentase tingkat kesukaran masing-masing soal kemudian diinterpretasikan berdasarkan kriteria berikut.

Tabel 4. 10. Interval Tingkat Kesulitan

Interval p	Kriteria
0,0 – 0,30	Sulit
0,31 – 0,70	Sedang
0,71 – 1,00	Mudah

Persentase tingkat kesukaran masing-masing tipe soal disajikan dalam tabel berikut.

Tabel 4. 11. Analisis Kesukaran Soal Tipe A

No.	Kategori	Nomor Soal	%
1	Sulit (0,0 – 0,30)	-	0%
2	Sedang (0,31 – 0,70)	5, 6, 13, 14	26,7%
3	Mudah (0,71 – 1,00)	1, 2, 3, 4, 7, 8, 9, 10, 11, 12, 15	73.3%

Tabel 4. 12. Analisis Kesukaran Soal Tipe B

No.	Kategori	Nomor Soal	%
1	Sulit (0,0 – 0,30)	-	0%
2	Sedang (0,31 – 0,70)	3, 5, 11, 19	20%
3	Mudah (0,71 – 1,00)	1, 2, 4, 6, 7, 8, 9, 10, 12, 13, 14, 15, 16,17,18, 20	80%

Tabel 4. 13. Analisis Kesukaran Soal Tipe C

No.	Kategori	Nomor Soal	%
1	Sulit (0,0 – 0,30)	-	0%
2	Sedang (0,31 – 0,70)	4, 5, 16	17.6%
3	Mudah (0,71 – 1,00)	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 17	82.4%

d. Daya Pembeda

Daya pembeda dari 3 tipe latihan soal sistem reproduksi manusia dianalisis menggunakan aplikasi Anates 4.0.9. Hasil persentase daya pembeda masing-masing soal kemudian diinterpretasikan berdasarkan kriteria berikut.

Tabel 4. 14. Interval Daya Pembeda

Interval Daya Pembeda	Kriteria
0,00 – 0,20	Jelek
0,21 – 0,40	Cukup
0,41 – 0,70	Baik
0,71 – 1,00	Baik sekali

Persentase daya pembeda masing-masing tipe soal disajikan dalam tabel berikut.

Tabel 4. 15. Persentasi Daya Pembeda Soal Tipe A

No.	Kategori	Nomor Soal	%
1	Jelek (0,00 – 0,20)	13	6.7%
2	Cukup (0,21 – 0,40)	1, 2, 3, 4, 6, 9, 10, 11, 14	60%
3	Baik (0,41 – 0,70)	5, 7, 8, 12, 15	33.3%
4	Baik Sekali (0,71 – 1,00)	-	0%

Tabel 4. 16. Persentasi Daya Pembeda Soal Tipe B

No.	Kategori	Nomor Soal	%
1	Jelek (0,00 – 0,20)	3, 5	10%
2	Cukup (0,21 – 0,40)	1, 2, 6, 7, 9, 11, 12, 14, 15, 20	50%
3	Baik (0,41 – 0,70)	4, 8, 10, 13, 16, 17, 18, 19	40%
4	Baik Sekali (0,71 – 1,00)	-	0%

Tabel 4. 17. Persentasi Daya Pembeda Soal Tipe C

No.	Kategori	Nomor Soal	%
1	Jelek (0,00 – 0,20)	8, 11, 14, 16	23.5%
2	Cukup (0,21 – 0,40)	2, 4, 6, 7, 12, 13, 15, 17	47.1%
3	Baik (0,41 – 0,70)	1, 3, 5, 9, 10,	29.4%
4	Baik Sekali (0,71 – 1,00)	-	0%

e. **Tingkat Pengecoh (*Distractor*)**

Perhitungan tingkat pengecoh (*distractor*) pada latihan soal sistem reproduksi manusia ini dilakukan dengan menggunakan aplikasi Anates 4.0.9. Hasil analisis pada setiap butir soal tipe A dan tipe B diinterpretasikan berdasarkan adaptasi dari skala *Likert* berikut.

- 1) Apabila ketiga jawaban memiliki pengecoh yang baik, maka soal dikategorikan memiliki efektivitas pengecoh yang baik
- 2) Apabila terdapat dua jawaban yang memiliki pengecoh yang baik, maka soal tersebut dikategorikan memiliki efektivitas pengecoh yang cukup baik.
- 3) Apabila hanya ada satu jawaban yang memiliki pengecoh yang baik, maka soal dikategorikan memiliki efektivitas pengecoh yang kurang baik.
- 4) Apabila semua jawaban tidak berfungsi sebagai pengecoh, maka soal dikategorikan memiliki efektivitas yang tidak baik.

Hasil analisis tingkat pengecoh tipe soal A dan tipe B disajikan dalam tabel berikut.

Tabel 4. 18. Efektivitas Pengecoh Soal Tipe A

No	Kategori	Nomor Soal	%
1	Baik	1, 9	13.3%
2	Cukup Baik	11, 12, 15	20%
3	Kurang Baik	2, 3, 4, 5, 6, 7, 8, 10, 13, 14	66.7%
4	Tidak Baik	-	0%

Tabel 4. 19. Efektivitas Pengecoh Soal Tipe B

No	Kategori	Nomor Soal	%
1	Baik	6, 11, 13, 20	20%
2	Cukup Baik	1, 2, 8, 10, 17	25%
3	Kurang Baik	3, 4, 5, 7, 9, 12, 14, 15, 19	45%
4	Tidak Baik	16, 18	10%

Sedangkan hasil analisis tingkat pengecoh soal tipe C diinterpretasikan berdasarkan adaptasi dari skala *Likert* berikut.

- 1) Apabila keempat jawaban memiliki pengecoh yang baik, maka soal dikategorikan memiliki efektivitas pengecoh yang sangat baik
- 2) Apabila ketiga jawaban memiliki pengecoh yang baik, maka soal tersebut dikategorikan memiliki efektivitas pengecoh yang baik.
- 3) Apabila terdapat dua jawaban yang memiliki pengecoh yang baik, maka soal dikategorikan memiliki efektivitas pengecoh yang cukup baik.
- 4) Apabila hanya satu jawaban yang memiliki pengecoh yang baik, maka soal dikategorikan memiliki efektivitas pengecoh yang kurang baik.
- 5) Apabila semua jawaban tidak berfungsi sebagai pengecoh, maka soal dikategorikan memiliki efektivitas yang tidak baik.

Hasil analisis tingkat pengecoh tipe soal C disajikan dalam tabel berikut.

Tabel 4. 20. Efektivitas Pengecoh Soal Tipe C

No	Kategori	Nomor Soal	%
1	Sangat Baik	-	0%
2	Baik	5	5.9%
3	Cukup Baik	3, 4, 6, 8, 9, 10, 11, 12, 15	52.9%
4	Kurang Baik	1, 7, 14, 16, 17	29.4%
5	Tidak Baik	2, 13	11.8%

B. Pembahasan

Analisis butir soal sistem reproduksi manusia pada program Wordwall ini menggunakan bantuan aplikasi Microsoft Excel 2019 dan Anates 4.0.9.. Hasil analisis meliputi validitas, reliabilitas, tingkat kesukaran, daya pembeda, dan tingkat pengecoh akan dibahas sebagai berikut.

1. Validitas

Validitas merupakan ukuran untuk menentukan bahwa instrumen dapat mengukur apa yang seharusnya diukur dan dilakukan penilaian.¹ Analisis validitas ketiga tipe soal sistem reproduksi manusia ini meliputi validitas isi, validitas konstruk, dan validitas statistik.

Perhitungan validitas isi dan validitas konstruk terdiri dari aspek materi, konstruksi, dan bahasa. Hasil validitas isi dan konstruk merupakan hasil analisis logika dari *judgement* ahli.² Penilaian dilakukan oleh dua orang *judgement* ahli yang kemudian dihitung menggunakan rumus Aiken V.

Hasil perhitungan validitas soal tipe A menunjukkan nilai 0.69 yang berarti soal tersebut memiliki validitas tinggi. Dari 15 soal yang ada, hanya satu soal yang memiliki validitas dengan kategori cukup. Sedangkan 14 soal lainnya termasuk soal dengan kategori tinggi.

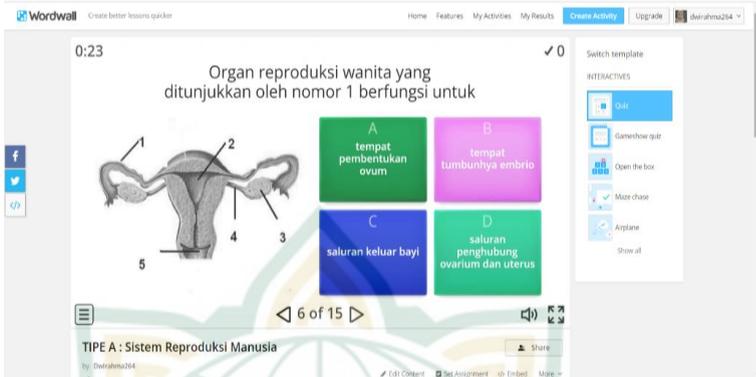
Soal tipe B memiliki nilai validitas 0.66 yang berarti soal tersebut termasuk soal dengan kategori validitas tinggi. Empat dari 20 soal tipe B memiliki validitas yang cukup, sedangkan 16 soal sisanya memiliki validitas yang tinggi.

Nilai validitas soal tipe C adalah sebesar 0.68 dengan kategori validitas soal tinggi. Dibandingkan dua tipe lainnya, soal tipe C memiliki validitas butir soal dengan kategori yang lebih beragam. Dari keseluruhan 17 soal, satu soal memiliki validitas yang rendah, dua soal bervaliditas cukup, dan 14 soal berkategori tinggi.

¹ Didi Nur Jamaludin, *Pengembangan Evaluasi Pembelajaran* (Kudus: IAIN Kudus, 2020), 190.

² Sudaryono *et al.*, "Validity of Test Instrument," *Journal of Physics: Conference Series* 1364 012050 (2019): 1-11.

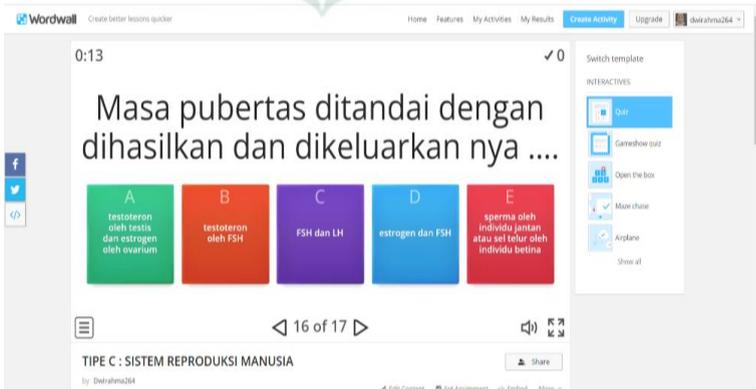
Gambar 4. 1. Soal dengan Validitas Isi dan Konstruk Tinggi (0.75)



Gambar 4. 2. Soal dengan Validitas Isi dan Konstruk Cukup (0.57)



Gambar 4. 3. Soal dengan Validitas Isi dan Konstruk Rendah (0.37)



Soal yang memiliki validitas isi dan konstruk tinggi dapat mengukur kesesuaian isi materi pembelajaran dengan kebenaran konsep. Selain itu soal tersebut juga dapat mengukur aspek berpikir siswa sesuai dengan kompetensi dasar atau tujuan pembelajaran. Ketiga tipe soal sistem reproduksi manusia tersebut memiliki validitas dengan kategori tinggi dari segi isi dan konstruk. Dengan demikian, ketiga tipe soal tersebut isi dan konstruknya adalah valid.

Sedangkan perhitungan validitas statistik dilakukan berdasarkan hasil yang didapat dari uji coba soal terhadap 100 siswa. Perhitungan ini dilakukan dengan bantuan aplikasi Anates 4.0.9. Hasil perhitungan dikategorikan ke dalam batas signifikansi dengan derajat kebebasan $N-2$. Tipe soal A memiliki $N=15$ yang jika dilihat pada r_{tabel} dengan taraf signifikansi 5%, maka diperoleh nilai 0.482. Tipe soal B memiliki $N=20$ dan jika dilihat pada r_{tabel} dengan taraf signifikansi 5%, maka diperoleh nilai 0.423. sedangkan tipe soal C memiliki $N=17$. Jika dilihat pada r_{tabel} dengan taraf signifikansi 5%, maka diperoleh nilai 0.423.

Hasil analisis validitas soal tipe A menunjukkan 2 soal (13.3%) memiliki validitas yang sangat rendah, 5 soal (33.3%) memiliki validitas rendah, dan 8 soal (53.4%) memiliki validitas yang cukup. Hal ini menunjukkan bahwa hampir sebagian butir soal tipe A memiliki validitas soal yang kurang. Meskipun Sebagian besar soal lainnya memiliki validitas yang cukup, akan tetapi instrumen soal tersebut tidak akan maksimal dalam mengukur kemampuan siswa.

Soal tipe B memiliki 3 soal (15%) dengan kategori sangat rendah, 5 soal (25%) berkategori rendah, 10 soal (50%) berkategori cukup, dan 2 soal (10%) berkategori tinggi. Soal tipe B lebih unggul dari soal tipe A karena memiliki soal dengan kategori tinggi. Meskipun demikian, kedua tipe soal didominasi oleh soal dengan kategori cukup.

Sedangkan soal tipe C memiliki 1 soal (5.9%) dengan kategori sangat rendah, 5 soal (39.4%) dengan kategori rendah, dan 11 soal (64.7%) berkategori cukup. Tidak ditemukan soal dengan kategori tinggi dan sangat tinggi pada tipe soal ini. Sama halnya dengan kedua tipe soal lainnya, soal tipe C ini juga didominasi oleh soal dengan kategori validitas cukup.

Gambar 4. 4. Soal dengan Validitas Statistik Tinggi (0.70)



Gambar 4. 5. Soal dengan Validitas Statistik Sangat Rendah (0.12)



2. Reliabilitas

Reliabilitas berasal dari bahasa Inggris “*reliability*” yang berarti dapat dipercaya. Instrumen tes dikatakan reliabel jika hasil penilaian yang didapat memberikan hasil yang tetap atau konsisten.³ Validitas reliabilitas memiliki keterkaitan. Jika suatu instrumen soal dinyatakan valid, maka instrumen soal tersebut sudah pasti reliabel. Akan tetapi, instrumen soal yang reliabel belum tentu valid. Secara umum, mengukur kevalidan soal lebih susah dibandingkan mengukur reliabilitas soal. Hal ini dikarenakan validitas merupakan pengukuran data yang berkaitan dengan pengetahuan, sedangkan reliabilitas menyangkut konsistensi skor.⁴

³ Eko Putro Widoyoko, *Penilaian Hasil Pembelajaran di Sekolah*, (Yogyakarta : Pustaka Belajar, 2018), 252.

⁴ Nor Hasnida M.G., “A Reliability and Validity of an Instrument to Evaluate the School-Based Assessment System: A Pilot Study,” *International Journal of Evaluation and Research in Education (IJERE)* 5, no.2 (2016): 148-157.

Analisis butir soal sistem reproduksi manusia pada program *Wordwall* ini dilakukan dengan menggunakan aplikasi Anates 4.0.9. Hasil perhitungan reliabilitas masing-masing tipe soal kemudian diinterpretasikan berdasarkan interval r_{11} .

Koefisien reliabilitas soal sistem reproduksi manusia tipe A adalah sebesar 0.64. Hal itu berarti soal tipe A memiliki tingkat reliabilitas yang tinggi. Pada soal tipe B, didapat koefisien reliabilitas soal sebesar 0.75. Artinya, soal tipe B memiliki reliabilitas yang tinggi. Koefisien reliabilitas soal tipe C adalah sebesar 0.69, dimana nilai tersebut dapat diinterpretasikan bahwa soal tipe C memiliki tingkat reliabilitas yang tinggi. Dengan demikian, dapat dilihat bahwa ketiga tipe soal sistem reproduksi manusia yang dianalisis adalah reliabel. Hal ini juga membuktikan bahwa ketiga tipe soal sistem reproduksi yang valid, juga sudah tentu reliabel.

3. Tingkat Kesukaran

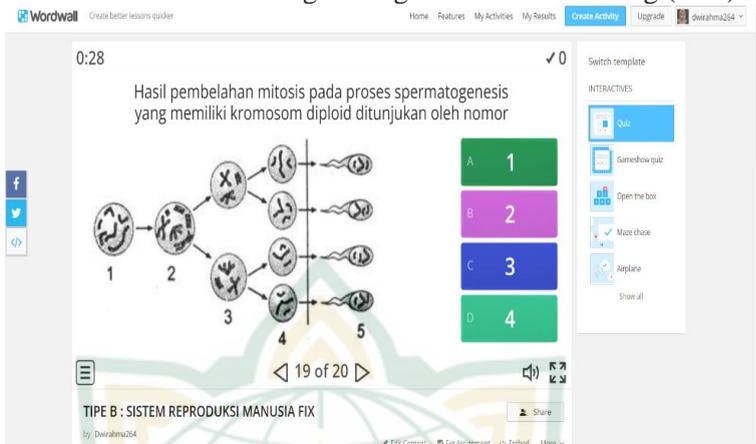
Soal yang baik merupakan soal yang tidak terlalu mudah sehingga dapat merangsang kemampuan berpikir tinggi, serta tidak terlalu sulit sehingga tidak menyebabkan siswa berputus asa. Keseimbangan tingkat kesulitan soal mengacu pada kurva normal dimana instrumen soal terdiri dari 25% mudah, 50% sedang, dan 25% sulit.⁵

Ketiga tipe soal sistem reproduksi manusia yang dianalisis memiliki tingkat kesukaran soal yang kurang baik. Hal ini ditunjukkan dengan tidak adanya soal yang berkategori sulit dan didominasi soal dengan kategori mudah. Soal tipe A hanya memiliki 4 (26.7%) soal dengan kategori sedang dan 11 soal (73.3%) berkategori mudah. Pada tipe B, 4 soal (20%) memiliki kategori sedang dan 16 soal (80%) berkategori mudah. Begitu pula dengan soal tipe C yang hanya memiliki 3 soal (17.6%) berkategori sedang dan 14 soal (82.4%) berkategori mudah.

Persentase diatas menunjukkan bahwa keseimbangan tingkat kesulitan ketiga tipe soal tidak mengacu pada kurva normal. Ketiga tipe soal didominasi soal dengan kategori mudah. Selain itu, tidak ditemukan adanya soal dengan kategori sulit.

⁵ Didi Nur Jamaludin, *Pengembangan Evaluasi Pembelajaran* (Kudus: IAIN Kudus, 2020), 207

Gambar 4. 6. Soal dengan Tingkat Kesulitan Sedang (0.70)



Gambar 4. 7. Soal dengan Tingkat Kesulitan Mudah (0.85)



4. Daya Pembeda

Daya pembeda dalam suatu analisis butir soal adalah kemampuan suatu butir soal untuk membedakan kelompok siswa yang memiliki kemampuan tinggi dan kelompok siswa yang berkemampuan rendah.⁶ Analisis daya pembeda pada ketiga tipe soal sistem reproduksi manusia ini dilakukan dengan mengelompokkan siswa kedalam kelompok unggul dan kelompok asor (bawah).

Hasil analisis daya pembeda soal sistem reproduksi manusia tipe A menunjukkan bahwa 1 soal (6.7%) memiliki daya

⁶ Bagiyono, “Analisis Tingkat Kesukaran dan Daya Pembeda Butir Soal Ujian Pelatihan Radiografi Tingkat I,” *Jurnal Widyanuklida* 16, no.1 (2017): 1-12.

pembeda yang jelek, 9 soal (60%) berkategori cukup, dan 5 soal (33.3%) berkategori baik. Soal sistem reproduksi tipe B memiliki 2 soal (10%) berdaya pembeda jelek, 10 soal (50%) berkategori cukup, dan 8 soal (40%) berkategori baik. Sedangkan soal tipe C memiliki 4 soal (23.5%) dengan kategori daya pembeda jelek, 8 soal (47.1%) berkategori cukup, dan 5 soal (29.4%) berkategori baik. Ketiga tipe soal yang dianalisis tidak memiliki soal yang berdaya pembeda baik sekali. Soal-soal tersebut hanya memiliki daya pembeda dengan kategori jelek, cukup, dan baik.

Gambar 4. 8. Soal dengan Daya Pembeda Baik (0.48)



Gambar 4. 9. Soal dengan Daya Pembeda Jelek (0.14)



5. Tingkat Pengecoh (*Distractor*)

Tingkat pengecoh merupakan distribusi jawaban siswa peserta tes. Butir soal yang baik memiliki pengecoh yang dipilih secara merata oleh siswa yang menjawab salah.⁷ Pengecoh yang baik setidaknya harus dipilih oleh minimal 5% dari jumlah responden. Analisis tingkat pengecoh memiliki beberapa fungsi, seperti memahami tingkat kognitif siswa, memilah jawaban

⁷ Didi Nur Jamaludin, *Pengembangan Evaluasi Pembelajaran* (Kudus: IAIN Kudus, 2020), 208-209

pengecoh yang baik, dan dapat meningkatkan hasil belajar siswa.⁸

Analisis efektivitas pengecoh pada soal tipe A dan tipe B dibagi menjadi empat kategori, yaitu baik, cukup baik, kurang baik, dan tidak baik. Hal ini dikarenakan kedua tipe soal tersebut hanya memiliki empat pilihan jawaban. Hasil analisis menunjukkan 2 soal (13.3%) tipe A memiliki efektivitas pengecoh yang baik, 3 soal (20%) soal berkategori cukup baik, dan 10 soal (66.7%) berkategori kurang baik. Pada soal tipe B, 4 soal (20%) memiliki efektivitas pengecoh yang baik, 5 soal (25%) dalam kategori cukup baik, 9 soal (45%) berkategori kurang baik, dan 2 soal (10%) memiliki efektivitas pengecoh yang tidak baik.

Berbeda dengan soal tipe A dan B, soal tipe C memiliki 5 kategori efektivitas pengecoh, yaitu sangat baik, baik, cukup baik, kurang baik, dan tidak baik. Kategori ini dihitung berdasarkan jumlah pilihan jawaban pada soal tipe C, yaitu lima. Hasil analisis menunjukkan tidak ditemukan soal yang memiliki efektivitas pengecoh dengan kategori sangat baik. Soal tipe C memiliki 1 soal (5.9%) dengan pengecoh yang berkategori baik, 9 soal (52.9%) memiliki pengecoh cukup baik, 5 soal (29.4%) berkategori kurang baik, dan 2 soal (11.8%) memiliki pengecoh yang tidak baik.

Gambar 4. 10. Soal dengan *Distractor* Baik



⁸ Amalia Vidya M., "Item Analysis of English Final Semester Test," *Indonesian Journal of EFL and Linguistic* 5, no.2 (2020): 491-504.

Gambar 4. 11. Soal dengan *Distractor* Tidak Baik



6. Analisis Butir Soal Berdasarkan Validitas, Reliabilitas, Tingkat Kesukaran, Daya Pembeda, dan Tingkat Pengecoh (*Distractor*)

Butir-butir soal dari masing-masing tipe soal yang telah dianalisis kemudian dikelompokkan menjadi tiga kategori. Kategori tersebut adalah soal yang baik, cukup baik, dan tidak baik. Pengelompokan tersebut didasarkan pada analisis validitas, tingkat kesukaran, daya pembeda, dan *distractor*. Sedangkan analisis reliabilitas termasuk kedalam nilai perangkat soal secara keseluruhan, bukan pada setiap butir soal.

Ketiga tipe soal sistem reproduksi manusia yang dianalisis memiliki reliabilitas yang berkategori tinggi. Sedangkan untuk penilaian masing-masing butir soal dari setiap tipe dikategorikan berdasarkan kriteria berikut.

- a. Butir soal dinyatakan memiliki kualitas yang baik apabila soal tersebut memenuhi keempat kriteria yang ada, yaitu validitas tes dengan kategori valid; tingkat kesukaran berkategori sedang; daya pembeda memiliki kategori sangat baik, baik, dan cukup; serta *distractor* yang berfungsi dengan sangat baik, baik, atau cukup baik.
- b. Butir soal dinyatakan memiliki kualitas yang cukup baik apabila soal tersebut memenuhi 2 atau 3 kriteria dari 4 kriteria yang ada.
- c. Butir soal dinyatakan memiliki kualitas yang tidak baik apabila soal tersebut hanya memiliki 1 kriteria yang ada.

Hasil analisis setiap butir soal dari masing-masing tipe dapat dilihat pada tabel berikut.

Tabel 4. 21. Hasil Analisis Butir Soal Tipe A

Kategori	Nomor Soal	Jumlah	Persentase
Baik	-	0	0%
Cukup Baik	1, 5, 6, 8, 9, 12, 15, 11, 14	9	60%
Tidak Baik	2, 3, 4, 7, 10, 13	6	40%

Tabel 4.22. Hasil Analisis Butir Soal Tipe B

Kategori	Nomor Soal	Jumlah	Persentase
Baik	-	0	0%
Cukup Baik	1, 2, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20	16	80%
Tidak Baik	3, 4, 5, 14	4	20%

Tabel 4.23. Hasil Analisis Butir Soal Tipe C

Kategori	Nomor Soal	Jumlah	Persentase
Baik		0	0%
Cukup Baik	1, 3, 4, 5, 6, 7, 9, 10, 11, 12, 15, 17	12	70.6%
Tidak Baik	2, 8, 13, 14, 16	5	29.4%

Berdasarkan tabel diatas, tidak ditemukan soal yang memiliki kualitas baik. Ketiga tipe soal sistem reproduksi manusia yang dianalisis rata-rata memiliki kualitas butir soal yang cukup baik. Soal tipe A memiliki 60% soal dengan kualitas cukup baik, dan 40% soal memiliki kualitas soal yang tidak baik. Soal tipe B memiliki 80% soal dengan kategori cukup baik, dan 20% soal dengan kategori tidak baik. Sedangkan, soal tipe C memiliki 70.6% soal dengan kualitas cukup baik, dan 29.4% soal dengan kualitas tidak baik.