

Item Analysis of the Final Semester Evaluation in Madrasah Ibtidaiyah Thotiqul Huda Malang

Faqih Zaenal Abidin^{1*}, Magfiroh^{2*} Nur Qomari^{3*}

¹ Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia

² Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia

³ Universitas Islam Negeri Maulana Malik Ibrahim Malang, Indonesia

*¹faqihzainal2@gmail.com, *²maghfirohqusyairi11@gmail.com, *³qomari@uin-malang.ac.id

ARTICLE INFO

Keywords:

Item analysis, evaluation, reliability, difficulty level, discrimination power, distractor quality, Arabic language assessment

Article history:

Received: 2025-06-30

Revised : 2025-07-02

Accepted :2025-08-01

ABSTRACT

This study aims to analyze the quality of the final semester evaluation items for the Arabic language subject of fifth-grade students at Thoriqul Huda Islamic Elementary School (Madrasah Ibtidaiyah) in Wagir. The study employed a quantitative descriptive method with documentation techniques as the data collection tool, which included the test items, answer keys, and student responses. The data were analyzed using "Anates" software to measure reliability, difficulty level, discrimination power, and distractor quality. The results showed that the test instrument had a high reliability level with a coefficient of 0.78, and most of the questions were at a moderate difficulty level. There were 22 items with good to excellent discrimination power, 14 items with poor discrimination, and 2 items with negative discrimination power. Regarding distractor quality, the analysis indicated that some answer choices still need improvement. Based on the analysis, 19 items can be reused in future tests, while 21 items require revision and further development. The study recommends that the item development team pay more attention to question quality standards to ensure effective measurement of students' learning achievements.

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



Corresponding Author:

Faqih Zaenal Abidin, etc Universitas Islam Negeri Maulana Malik Ibrahim Malang; e-mail@e-mail.com;
faqihzainal2@gmail.com

المقدمة

مادة اللغة العربية من المواد الدراسية المهمة في المدارس الإسلامية، خاصة في مرحلة المدرسة الابتدائية الإسلامية. إتقان أساس اللغة العربية منذ الصغار هو أساس للمتعلمين في فهم النصوص العربية وتعزيز هويتهم الإسلامية. لذلك، فإن نجاح تعليم اللغة العربية لا يعتمد فقط على عملية التدريس، بل يعتمد أيضًا على كيفية تطبيق عملية التقييم بشكل منهجي وذو جودة. لأن وظيفة التقييم كمقياس للنجاح تهدف إلى معرفة مدى نجاح تطبيق برنامج التعلم. (Muhammad, Afifullah; (Ridho, 2018(Siti, 2020; Muhammad Iqbal, 2021; Sholihatul Atik, 2020

في "المنهج المستقل (Kurikulum Merdeka) "يُعرف التقييم التعليمي بمصطلحي "التقييم التكويني" و"التقييم الختامي"، حيث لكل منهما معنى وهدف مختلف في التنفيذ. فالتقييم التكويني يهدف إلى تقديم معلومات أو تغذية راجعة للمعلمين والمتعلمين لتحسين عملية التعلم، ويُنفَّذ بعد عملية التعلم. أما التقييم الختامي فهو التقييم الذي يعمل للتأكد من تحقيق أهداف التعلم. يتم تنفيذ هذا التقييم في نهاية تعلم المادة الدراسية أو يعمل لقياس هدفين تعليميين أو أكثر، أو في نهاية الفصل الدراسي وفقًا لتقدير المعلم وسياسة المؤسسة التعليمية. بخلاف التقييم التكويني، فإن التقييم الختامي يدخل ضمن احتساب الدرجات في نهاية الفصل الدراسي، أو نهاية السنة الدراسية، أو نهاية المرحلة التعليمية. (Dion Ginanto et al., 2024; (Ghani, Ribahan, & Nasri, 2023; Rahman & Fuad, 2023; Yahrif & R. (Supardi, 2023

التقييم الختامي له دور استراتيجي في قياس مدى تحقيق كفاءات الطلاب. ونتائج هذا التقييم يتمنى للمعلمين على تحسين جودة تدريسهم، وتدفع المتعلمين على تحسين جودة تعلمهم. فالتقييم يزود الصف والمعلم بمعلومات لتحسين جودة عملية التعليم والتعلم. وينبغي أن تكون المعلومات المستخدمة في تقييم برنامج التعلم دقيقة بأقل نسبة من الخطأ. فالتقييم في جوهره هو إصدار حكم على نتائج التقييم، ولذلك يجب تقليل الخطأ في التقييم والقياس إلى أدنى حد ممكن. (Munip, 2017(Andrian,) (2024; EFIE, 2020) لذلك، فإن إعداد اختبار التقييم الختامي أمر بالغ الأهمية يجب أن يؤخذ بسرعة، لأنه يُعطي تصورًا حول فعالية أساليب التدريس التي يستخدمها المعلم. فليس الغرض من الاختبار الجيد أن يكون أداة قياس فحسب، بل يجب أن يعكس أيضًا جودة عملية التعلم نفسه

بناءً على نتائج الملاحظة الأولى في بعض المدارس الابتدائية الإسلامية في قريتي "واجير" و"باكيساجي"، تم انضمام بعض المدارس إلى مجموعة عمل المعلمين (KKG) لمناقشة الأمور المتعلقة بأنشطة التعليم والتعلم، بما في ذلك إعداد أسئلة التقييم الختامي. يقوم المعلمون المختارون كفريق صناعة أسئلة الاختبار بإعداد بنود أسئلة التقييم الختامي، وعندما يتم الانتهاء منها تُوزع إلى جميع المدارس التابعة لمجموعة عمل المدارس (KKM) في قريتي "باكيساجي" و"واجير".

يركز هذا البحث على تحليل بنود الأسئلة الاختيارية المتعدد باختبار التقييم الختامي لمادة اللغة العربية لطلاب الصف الخامس في المرحلة (C). لذلك، هناك حاجة ماسة إلى تحليل دقيق لجودة بنود أسئلة اختبار التقييم الختامي لمعرفة جودة الاختبار القائم، وكذلك لتقديم توصيات عملية لتحسين جودة تقييم اللغة العربية في مستوى المدرسة الابتدائية الإسلامية.

منهجية البحث

المنهج المستخدم في هذا البحث هو المنهج الوصفي الكمي. ووفقاً لما ذكره أريكنتو، فإن الهدف من البحث الوصفي هو البحث عن معلومات تتعلق بالظواهر القائمة كما هي في وقت إجراء البحث. (Arikunto, 1998). أما البحث الكمي فهو البحث الذي يهدف إلى الحصول على معلومات باستخدام البيانات الرقمية كأداة لمعرفة أو اكتشاف معلومة معينة. (Irfan Syahroni, 2022)

البيانات المستخدمة في هذا البحث هي أسئلة الاختبار الختامي للفصل الدراسي الثاني لمادة اللغة العربية للصف الخامس، ومفتاح الإجابة، وإجابات طلاب الصف الخامس في مدرسة "طريق الهدى" الابتدائية الإسلامية - واجير. أما مجتمع البحث فهو جميع استجابات الطلاب لاختبار التقييم الختامي للفصل الدراسي الثاني لمادة اللغة العربية للصف الخامس، بالإضافة إلى مفتاح الإجابة وإجابات طلاب الصف الخامس في نفس المدرسة. وقد تم اختيار العينة باستخدام أسلوب العينة العشوائية، حيث تم الحصول على 20 استجابة من الطلاب كعينة للبحث.

أما التقنية التي استخدمت لجمع البيانات في هذا البحث فهي تقنية التوثيق. والتوثيق هو عملية جمع البيانات من خلال الاطلاع المباشر على المصادر الوثائقية المرتبطة بالبحث. بمعنى آخر، التوثيق هو جمع البيانات من الوثائق المكتوبة أو الإلكترونية. (Arikunto, 1998) جمع البيانات باستخدام طريقة التوثيق وفقاً للدراسات الحديثة يُعتبر الطريقة الأكثر فعالية لتحليل بنود الأسئلة بدقة تصل إلى 95% (Dewi Ratih et al., n.d.). أداة البحث تتكون من 40 سؤالاً. وتشمل الوثائق جميع إجابات الطلاب التي تُستخدم كمادة للتحليل من أجل تقييم جودة كل بند من بنود الأسئلة. تم إجراء تحليل البيانات باستخدام "Anates" الإصدار 4.0، وهو برنامج مخصص لتحليل الاختبارات من نوع الاختبار الاختياري والأسئلة المقالية إحصائياً. (Interdiana et al., 2014)

تم إدخال بيانات إجابات الطلاب التي تم توثيقها إلى "Microsoft Excel"، ثم تم تحليلها باستخدام "Anates" لإجراء عدة أنواع من التحليل مثل: مستوى صعوبة السؤال، قدرة التمييز (القدرة التفرقية)، الصدق، والثبات. نتيجة عملية التحليل هذه الحصول على نتائج التقييم بشكل آلي وفقاً لمتطلبات البحث. وتُعرض نتائج التحليل المستخرجة من "Anates" في شكل نسب مئوية. ثم يتم وصف البيانات التي تم تحليلها بشكل مفصل لتقديم تصور حول جودة بنود الأسئلة. ومن خلال هذا الوصف يمكن تصنيف الأسئلة ضمن الفئات: جيدة، بحاجة إلى تحسين، أو غير صالحة للاستخدام.

وبناءً على ذلك، سيتوصل الباحث إلى استنتاج حول ما إذا كانت هذه الأسئلة يمكن إعادة استخدامها في الفصل الدراسي القادم أم أنها بحاجة إلى تعديل وتحسين.

النتائج والمناقشة

ثبات الاختبار (الموثوقية)

إن أداة الاختبار التي يُمكن الوثوق بها هي الأداة التي تعطي نتائج قياس يُعتمد عليها. ومن معايير الأداة الموثوقة أن تعطي نتائج ثابتة إذا استخدمت مراراً وتكراراً. يمكن القول إن الاختبار يُعد موثوقاً إذا طبق مراراً على نفس مجموعة الطلاب وأعطى نتائج متقاربة بشكل نسبي في كل مرة. يشير ثبات

الاختبار إلى مدى الاتساق أو الثبات لأداة القياس في تقديم نتائج القياس. ويُقال إن الاختبار يتمتع بدرجة موثوقية جيدة إذا كان نتيجته متسقة كلما استُخدم. وكلما ارتفع مستوى الثقة في الاختبار، زادت استقرارية ودقة نتائج القياس. (Asrul et al., 2014)

وفقًا لما ذكرته آني إنتر ديانا وميرنا هيرواتي في بحثهما، يمكن استخدام برنامج Anates لتحليل وتصنيف موثوقية الأسئلة بناءً على المعايير التالية: (Interdiana et al., 2014)

الجدول 1. تصنيف موثوقية الأسئلة

معامل الارتباط (معامل الثبات)	موثوقية الأسئلة
0,00 – 0,19	ارتباط منخفض جدًا
0,20 – 0,39	ارتباط منخفض
0,40 – 0,59	ارتباط متوسط
0,60 – 0,89	ارتباط مرتفع
0,90 – 1,00	ارتباط مرتفع جدًا

مستوى الصعوبة

يشير مستوى الصعوبة (مؤشر الصعوبة) إلى مقياس يُظهر درجة صعوبة السؤال بالنسبة للطلاب في حله. وبعبارة بسيطة، يمكننا أن نقول إن السؤال يُعتبر صعبًا إذا فشل معظم المختبرين في إجابته. وعلى العكس من ذلك، يُقال إن السؤال سهل إذا تمكن معظم المختبرين من حله. ومن إحدى الطرق البسيطة لمعرفة مؤشر مستوى صعوبة السؤال هو حساب نسبة الإجابات الصحيحة التي أجابها الطلاب عند حل الأسئلة. (Inanna et al., 2021)

معرفة جودة أو عدم جودة بنود أسئلة اختبار التحصيل الدراسي في البداية من خلال معرفة درجة الصعوبة أو السهولة التي يمتلكها كل بند من هذه البنود. وتُعتبر بنود أسئلة اختبار التحصيل

الدراسي جيدة إذا لم تكن صعبة جدًا ولا سهلة جدًا (Widodo, 2021) وقد استخدم الباحث Anates في هذا البحث، والذي يصنّف مستوى صعوبة الأسئلة إلى ثلاث فئات وهي:

الجدول 2: تصنيف مستوى الصعوبة

الوصف	معامل الارتباط
صعب	0,00 – 0,30
متوسط	0,31 – 0,70
سهل	0,71 – 1,00

القدرة التمييزية (التمييز بين الطلاب)

تشير القدرة التمييزية للسؤال إلى مدى قدرة السؤال على التمييز بين الطلاب الذين ينتمون إلى المجموعة المتفوقة (المجموعة العليا) والطلاب الذين ينتمون إلى المجموعة الضعيفة (المجموعة المنخفضة). الرقم الذي يوضح قيمة التمييز يُسمى "معامل التمييز" ويُرمز له بـ (D). كما هو الحال مع مؤشر الصعوبة، يتراوح معامل التمييز بين 0,00 إلى 1,00. لكن هناك الفرق أن مؤشر الصعوبة لا يستخدم الإشارة السالبة (-)، بينما في معامل التمييز قد تظهر الإشارة السالبة. تُستخدم الإشارة السالبة في معامل التمييز عندما يكون السؤال "معكوسًا" في إظهار جودة الطالب؛ أي أن الطالب الذكي يُعتبر ضعيفًا، والطالب الضعيف يُعتبر ذكيًا. وفقًا لما ذكرته آني إنتر ديانا وميرنا هيرواتي في بحثهما، يمكن استخدام برنامج Anates لتحليل وتصنيف القدرة التمييزية للأسئلة بناءً على المعايير التالية:

الجدول 3: تصنيف القدرة التمييزية

قيمة D	التصنيف	الشرح
$> 0,20$	ضعيف	قدرة تمييز ضعيفة
$0,20 - 0,40$	مقبول	قدرة تمييز متوسطة

قدرة تمييز جيدة	جيد	0,40 – 0,70
قدرة تمييز ممتازة	ممتاز	0,70 – 1
قدرة تمييز سالبة / ضعيفة جدًا	–	(–)

جودة الإجابات المشتتة

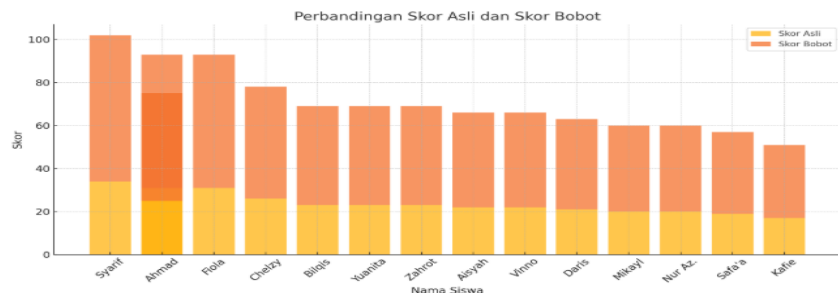
جودة الإجابات المشتتة هي عملية تقييم تهدف إلى دراسة كل خيار من خيارات الإجابة الجاهزة. ومن خلال هذا التحليل يمكن تحديد أي من خيارات الإجابة فعال ويستحق الإبقاء عليه، وأي منها غير مناسب ويحتاج إلى تعديل أو بدل. تُقاس فعالية الإجابات المشتتة في تشتيت انتباه الطلاب بناءً على معايير محددة. (دوي رتيه وآخرون) ولتحديد الإجابات المشتتة يتم استخدام المعايير التالية:

الجدول 4: مؤشر جودة الإجابات المشتتة

النسبة المئوية	مؤشر المشتت
76% – 125%	جيد جدًا
51%–75% atau 126%–15-%	جيد
26%–50% atau 151%–175%	متوسط
0%–25% atau 176% – 200%	ضعيف
<200%	ضعيف جدًا

المناقشة

تم تحليل بنود أسئلة اختبار التقييم الختامي للفصل الدراسي الثاني لعام 2025 في مادة اللغة العربية لطلاب الصف الخامس بمدرسة "طريق الهدى" الابتدائية الإسلامية - واجير - مالانج. وقد تم استخدام أداة تقييم على شكل أسئلة اختيار من متعدد، تتكون من أربعة خيارات للإجابة (أ، ب، ج، د) وتم تحديد وزن كل سؤال على النحو التالي: إذا كانت الإجابة صحيحة يحصل الطالب على 4 درجات، وإذا كانت الإجابة خاطئة يحصل على 0 درجة



من خلال الرسم البياني أعلاه، يتبين أن شريف حصل على أعلى درجة سواء في الدرجة الأصلية أو في الدرجة الموزونة. وتزداد الدرجة الموزونة بشكل يتناسب مع عدد الإجابات الصحيحة. ويلاحظ وجود توزيع متقارب لقدرات الطلاب بشكل عام، إلا أن هناك بعض الطلاب الذين حصلوا على نتائج منخفضة مثل كافي وسفعات.

ثبات الاختبار (موثوقية الاختبار)

• المتوسط الحسابي = 23,73

• الانحراف المعياري = 4,91

• معامل الارتب XY = 0,63

• ثبات الاختبار = 0,78

الصورة 1: نتيجة تحليل موثوقية الاختبار

No.Urut	No. Subyek	Kode>Nama Subyek	Skor Ganjil	Skor Genap	Skor Total
1	1	Ahmad Azizur ...	14	11	25
2	2	Ahmad Zidan A...	16	15	31
3	3	Aisyah Deris ...	10	12	22
4	4	Bilqis Latish...	11	12	23
5	5	Chelzy Putri ...	13	13	26
6	6	Daris Febian ...	12	9	21
7	7	Fiola Clara S...	14	17	31
8	8	Kafie Abiyu A...	8	9	17
9	9	Mikayla Rara ...	9	11	20
10	10	Nur Azizah Qa...	9	10	19
11	11	Safa'at Nur F...	11	8	19
12	12	Syarif Maulan...	19	15	34
13	13	Vinno Henri A	10	12	22
14	14	Yuanita Septi...	10	13	23
15	15	Zahrotur Rohmah	12	11	23

تشير نتائج اختبار الموثوقية إلى أن أداة الاختبار تتمتع بمستوى موثوقية عالٍ. وقد ظهر ذلك من خلال قيمة معامل الموثوقية التي بلغت **0,78** وباعتماد على القيم المحسوبة لمتوسط الدرجات، والانحراف المعياري، ومعامل الموثوقية، يمكن الاستنتاج أن أداة الاختبار هذه تحققت فيها معايير الثبات وتتميز بجودة جيدة.

مستوى الصعوبة:

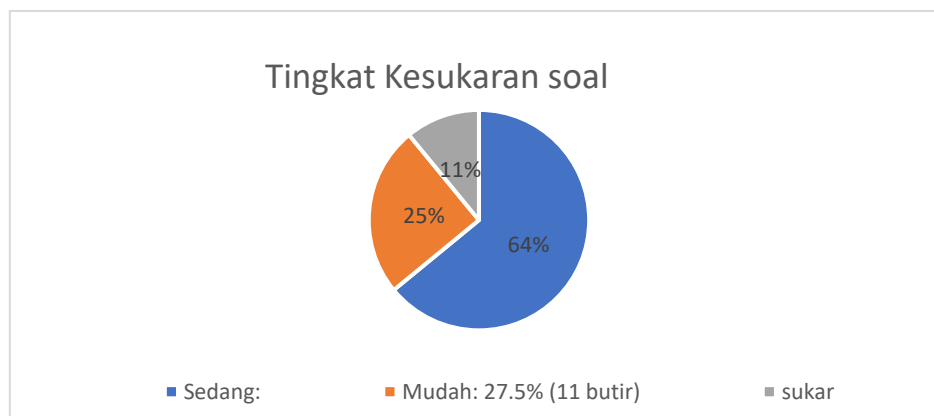
- عدد المشاركين = 15
- عدد بنود الأسئلة = 40

الصورة 2: مستوى الصعوبة

No Butir Baru	No Butir Asli	Jml Betul	Tkt. Kesuklanan(%)	Tafsiran
1	1	12	80,00	Mudah
2	2	12	80,00	Mudah
3	3	11	73,33	Mudah
4	4	10	66,67	Sedang
5	5	5	33,33	Sedang
6	6	8	53,33	Sedang
7	7	6	40,00	Sedang
8	8	6	40,00	Sedang
9	9	3	20,00	Sulkar
10	10	7	46,67	Sedang
11	11	5	33,33	Sedang
12	12	3	20,00	Sulkar
13	13	5	33,33	Sedang
14	14	5	33,33	Sedang
15	15	5	33,33	Sedang
16	16	5	33,33	Sedang
17	17	5	33,33	Sedang
18	18	4	26,67	Sulkar
19	19	10	66,67	Sedang
20	20	6	40,00	Sedang
21	21	6	40,00	Sedang
22	22	6	40,00	Sedang
23	23	9	60,00	Sedang
24	24	11	73,33	Mudah
25	25	10	66,67	Sedang
26	26	12	80,00	Mudah
27	27	12	80,00	Mudah
28	28	10	66,67	Sedang
29	29	10	66,67	Sedang
30	30	9	60,00	Sedang
31	31	14	93,33	Sangat Mudah
32	32	13	86,67	Sangat Mudah
33	33	13	86,67	Sangat Mudah
34	34	14	93,33	Sangat Mudah
35	35	14	93,33	Sangat Mudah
36	36	14	93,33	Sangat Mudah
37	37	14	93,33	Sangat Mudah
38	38	12	80,00	Mudah
39	39	9	60,00	Sedang
40	40	11	73,33	Mudah

من خلال نتائج تحليل برنامج Anates ، تم الحصول على بيانات تفيد بأن معظم الأسئلة

تقع في مستوى صعوبة متوسط، مما يدل على أن الأسئلة تعتبر ممثلة بشكل كافٍ لقياس قدرات الطلاب بشكل عام.



القدرة التمييزية

الصورة 3: القدرة التمييزية

Jumlah Subyek= 15
Klp atas/bawah(n)= 4
Butir Soal= 40
Nama berkas: BELUM_ADA_NAMA.ANA

No Butir Baru	No Butir Asli	Kel. Atas	Kel. Bawah	Beda	Indeks DP (%)
1	1	3	3	0	0,00
2	2	4	2	2	50,00
3	3	4	2	2	50,00
4	4	3	3	0	0,00
5	5	3	0	3	75,00
6	6	3	2	1	25,00
7	7	3	1	2	50,00
8	8	3	1	2	50,00
9	9	2	1	1	25,00
10	10	2	2	0	0,00
11	11	1	0	1	25,00
12	12	0	0	0	0,00
13	13	2	2	0	0,00
14	14	2	0	2	50,00
15	15	1	1	0	0,00
16	16	2	2	0	0,00
17	17	3	1	2	50,00
18	18	2	2	0	0,00
19	19	4	2	2	50,00
20	20	4	1	3	75,00
21	21	3	0	3	75,00
22	22	3	0	3	75,00
23	23	3	1	2	50,00
24	24	4	0	4	100,00
25	25	4	2	2	50,00
26	26	4	2	2	50,00
27	27	4	2	2	50,00
28	28	4	1	3	75,00
29	29	4	1	3	75,00
30	30	2	2	0	0,00
31	31	4	4	0	0,00
32	32	4	4	0	0,00
33	33	4	4	0	0,00
34	34	4	4	0	0,00
35	35	4	4	0	0,00
36	36	4	4	0	0,00
37	37	4	4	0	0,00
38	38	4	2	2	50,00
39	39	2	2	0	0,00
40	40	2	4	-2	-50,00

من خلال نتائج تحليل برنامج Anates ، تبين أن هناك 22 سؤالاً وله قدرة تمييز تصنف في المرتبة الجيدة والجيدة جداً، مما يدل على أن هذه الأسئلة فعالة في التمييز بين الطلاب ذوي القدرات العالية والمنخفضة. كما وُجد أن هناك سؤالين يعتبرين من الأسئلة المتوسطة. بالإضافة إلى ذلك، هناك 14 سؤالاً تعتبر "ضعيف"، مما يعني أن هذه الأسئلة غير قادرة بشكل كافٍ على التمييز بين الطلاب المتفوقين والطلاب الضعفاء. وأخيراً، وُجد أن هناك سؤالين يتمتعان بقدرة تمييز سلبية، مما

يعني أن هذه الأسئلة غير صالحة، لأن الطلاب ذوي القدرات المنخفضة أجابوا عليها بشكل صحيح أكثر من الطلاب ذوي القدرات العالية.

جودة الإجابات المشتتة

No Butir Baru	No Butir Asli	a	b	c	d	*
1	1	12**	1++	1++	1++	0
2	2	12**	2--	0--	1++	0
3	3	0--	11**	2+	2+	0
4	4	2++	1+	10**	2++	0
5	5	5**	3++	2+	5+	0
6	6	0--	8**	2++	5--	0
7	7	2+	3++	6**	4+	0
8	8	6**	3++	3++	3++	0
9	9	2-	6+	4++	3**	0
10	10	1-	7**	3++	4+	0
11	11	5**	2+	3++	5+	0
12	12	4++	6+	3**	2-	0
13	13	3++	5**	5+	2+	0
14	14	1-	5**	4++	5+	0
15	15	2+	6--	2+	5**	0
16	16	5**	3++	2+	5+	0
17	17	5**	6--	3++	1-	0
18	18	3++	3++	5+	4**	0
19	19	3--	1+	1+	10**	0
20	20	1-	6--	6**	2+	0
21	21	4+	0--	5-	6**	0
22	22	6**	2+	4+	3++	0
23	23	4--	9**	1-	1-	0
24	24	1+	1+	2+	11**	0
25	25	0--	10**	3--	2++	0
26	26	1++	12**	1++	1++	0
27	27	12**	0--	0--	3--	0
28	28	3--	1+	1+	10**	0
29	29	0--	5--	0--	10**	0
30	30	9**	2++	2++	2++	0
31	31	14**	0--	1--	0--	0
32	32	0--	1+	13**	1+	0
33	33	1+	13**	1+	0--	0
34	34	14**	0--	0--	1--	0
35	35	1--	14**	0--	0--	0
36	36	0--	1--	0--	14**	0
37	37	14**	0--	0--	1--	0
38	38	0--	2--	1++	12**	0
39	39	0--	9**	1-	5--	0
40	40	4--	0--	0--	11**	0

من خلال نتائج تحليل برنامج Anates ، تبين أن عدد الأسئلة هو 40 سؤالاً. من بين هذه الأسئلة، وُجد أن 19 سؤالاً (أي ما يعادل 47.5%) تحتوي على الإجابات المشتتة تعمل بشكل جيد وجيد جداً، مما يدل على فعاليتها في تشتيت انتباه الطلاب. بينما وُجد أن حوالي 8 أسئلة (20%) تحتوي على الإجابات المشتتة ذات أداء ضعيف ولا تعمل بكفاءة كما هو مطلوب. أما بالنسبة إلى حوالي 13 سؤالاً (32.5%) فقد تبين أن الإجابات المشتتة غير فعالة تماماً، بل تُصنف في مرتبة "سيئة إلى سيئة جداً". ومن الملاحظ أن في بعض الأسئلة مثل السؤالين رقم 39 و 40، قام معظم الطلاب باختيار خيارات ليست هي الإجابة الصحيحة، بل إن الإجابات المشتتة تم اختيارها من قبل الطلاب بنسبة أعلى بكثير من اختيار الإجابة الصحيحة.

خلاصة تحليل بنود الأسئلة

Btr Baru	Btr Asli	D.Pembeda(%)	T. Kesukaran	Korelasi	Sign. Korelasi
1	1	0,00	Mudah	-0,098	-
2	2	50,00	Mudah	0,324	Signifikan
3	3	50,00	Mudah	0,348	Signifikan
4	4	0,00	Sedang	-0,099	-
5	5	75,00	Sedang	0,547	Sangat Signifikan
6	6	25,00	Sedang	0,145	-
7	7	50,00	Sedang	0,505	Sangat Signifikan
8	8	50,00	Sedang	0,534	Sangat Signifikan
9	9	25,00	Sukar	0,450	Sangat Signifikan
10	10	0,00	Sedang	0,194	-
11	11	25,00	Sedang	0,129	-
12	12	0,00	Sukar	-0,148	-
13	13	0,00	Sedang	0,219	-
14	14	50,00	Sedang	0,458	Sangat Signifikan
15	15	0,00	Sedang	0,099	-
16	16	0,00	Sedang	0,219	-
17	17	50,00	Sedang	0,547	Sangat Signifikan
18	18	0,00	Sukar	0,193	-
19	19	50,00	Sedang	0,467	Sangat Signifikan
20	20	75,00	Sedang	0,649	Sangat Signifikan
21	21	75,00	Sedang	0,448	Sangat Signifikan
22	22	75,00	Sedang	0,419	Sangat Signifikan
23	23	50,00	Sedang	0,299	-
24	24	100,00	Mudah	0,634	Sangat Signifikan
25	25	50,00	Sedang	0,438	Sangat Signifikan
26	26	50,00	Mudah	0,324	Signifikan
27	27	50,00	Mudah	0,359	Signifikan
28	28	75,00	Sedang	0,467	Sangat Signifikan
29	29	75,00	Sedang	0,467	Sangat Signifikan
30	30	0,00	Sedang	-0,247	-
31	31	0,00	Sangat Mudah	0,098	-
32	32	0,00	Sangat Mudah	0,185	-
33	33	0,00	Sangat Mudah	0,185	-
34	34	0,00	Sangat Mudah	0,154	-
35	35	0,00	Sangat Mudah	0,154	-
36	36	0,00	Sangat Mudah	0,154	-
37	37	0,00	Sangat Mudah	0,154	-
38	38	50,00	Mudah	0,499	Sangat Signifikan
39	39	0,00	Sedang	0,069	-
40	40	-50,00	Mudah	-0,257	-

من خلال النتائج المكتوبة، يتبين أن هناك تسعة عشر سؤالاً يمكن استخدامها مرة أخرى في الاختبار القادم، بينما تحتاج واحد وعشرين سؤالاً أخرى إلى مراجعة وتعديل قبل إعادة استخدامها.

الخلاصة

نتائج التحليل باستخدام برنامج ANATES أن أسئلة اختبار التقييم الختامي للفصل الدراسي الثاني لعام 2025 للصف الخامس بمدرسة ابتدائية طريق الهدى لها جودة جيدة 0,75، ويمكن الاستنتاج أن أداة الاختبار هذه تُعتبر أداة موثوقة ذات جودة مناسبة حيث بلغت قيمة معامل الموثوقية. كما تبين أن مستوى صعوبة الأسئلة يقع في المستوى المتوسط، وهو مناسب لتقييم قدرات الطلاب بشكل متوازن. ومع ذلك، لا بد من إعادة النظر في الأسئلة السهلة جداً والصعبة جداً لتحقيق توزيع أكثر عدلاً بين المستويات المختلفة. أما بالنسبة لقدرة التمييز، فقد تبين أن معظم الأسئلة له قدرة تمييز جيدة، إلا أن هناك أربعة عشر سؤالاً تُصنّف ضمن المرتبة الضعيفة (Poor)، وسؤالين له قدرة تمييز

سلبية، مما يعني أن هذه الأسئلة قد تسببت في تشويش الطلاب وهي غير صالحة للاستخدام. كما أن بعض الإجابات المشتتة لا تزال بحاجة إلى تحسين لزيادة فعاليتها في التقييم.

من خلال خلاصة تحليل بنود الأسئلة، يمكن الاستنتاج أن هناك 19 سؤالاً صالحة للاستخدام في الاختبارات القادمة، بينما هناك 21 سؤالاً تحتاج إلى مراجعة وتقييم جديد. يوصى لفريق إعداد أسئلة اختبار التقييم الختامي للفصل الدراسي الثاني لعام 2025 أن يكثر من الاهتمام لمعايير جودة الأسئلة، لتكون فعالة في قياس درجة إنجاز الطلاب في عملي

المراجع

- Arikunto, S. (1998). *Prosedur penelitian suatu pendekatan praktek* (Edisi revisi 4). Rineka Cipta.
<https://cir.nii.ac.jp/crid/1130000795354347648.bib?lang=en>
- Asrul, Ananda, R., & Roenita. (2014). *Evaluasi Pembelajaran* (1st ed.). Citapustaka Media.
- Dewi Ratih, P., Aulia Pristanti, V., Hakim, L., Pratiwi, V., Akuntansi, P., & Negeri Surabaya, U. (n.d.). *Analisis Soal Pilihan Ganda HOTS Menggunakan Anates 4.0 Pada Etika Profesi Akuntansi SMK Kelas X*.
- Dion Ginanto, Ameliasari T, auresia Kesuma, & Dwi Setiyowati. (2024). *Panduan Pembelajaran dan Asesmen Pendidikan Anak Usia Dini, Pendidikan Dasar, dan Pendidikan Menengah*.
- Inanna, Rahmatullah, & Hasan, M. (2021). *Evaluasi Pembelajaran : Teori dan Praktek*. Tahta Media Grup.
- Interdiana, A., Sari, C., & Herawati, M. (2014). Aplikasi Anates Veris 4 dalam Menganalisis Butir Soal. *Faktor Jurnal Ilmiah Kependidikan*, 1(2).
- Irfan Syahroni, M. (2022). Prosedur Penelitian Kuantitatif. *Jurnal Al-Musthafa STIT Al-Aziziyah Lombok Barat*, 2(3).
- Munip, A. (2017). *Penilaian Pembelajaran Bahasa Arab*. Fakultas Ilmu Tarbiyah dan Keguruan UIN Sunan Kalijaga Yogyakarta. <https://www.researchgate.net/publication/332978391>
- Ridho, U. (2018). *Evaluasi dalam Pembelajaran Bahasa Arab*. 20(01).
- Widodo, H. (2021). *Evaluasi Pendidikan* (B. Ashari, Ed.; 1st ed.). UAD Press.
- Andrian, D. (2024). Implementasi Formatif Dan Sumatif Assesmen Dalam Meningkatkan Kualitas Pembelajaran Di Kelas. *Jurnal Pengabdian Masyarakat Bangsa*, 2(3), 479–485.
<https://doi.org/10.59837/jpmba.v2i3.845>
- EFIE, R. D. (2020). Implementasi unit kegiatan belajar mandiri (ukbm) pada kurikulum 2013 dalam kegiatan saintifik mata pelajaran fikih di madrasah tsanawiyah negeri 2 ponorogo. INSTITUT AGAMA ISLAM NEGERI PONOROGO.
- Ghani, A., Ribahan, R., & Nasri, U. (2023). Paradigma Diferensiasi Dalam Implementasi Kurikulum Merdeka: Konteks Pembelajaran Pendidikan Agama Islam Di Sekolah Dan Madrasah. *El-Hikmah Jurnal Kajian Dan Penelitian Pendidikan Islam*, 17(2), 169–179.
<https://doi.org/10.20414/elhikmah.v17i2.8867>
- Muhammad, Afifullah; Siti, M. (2020). EFEKTIFITAS PEMBELAJARAN BAHASA ARAB DARING PADA MASA PANDEMI COVID-19 DI FAKULTAS AGAMA ISLAM, UNIVERSITAS ISLAM MALANG. *VICRATINA : Jurnal Pendidikan*, 5(2), 65–71.
- Muhammad Iqbal, A. F. (2021). EFEKTIVITAS PEMBELAJARAN BAHASA ARAB MENGGUNAKAN MEDIA ONLINE SELAMA PANDEMI COVID-19 DI JURUSAN SASTRA ARAB UNIVERSITAS NEGERI MALANG. *Seminar Nasional Bahasa Arab Mahasiswa V Tahun*
- Faqih Zaenal Abidin / Item Analysis of the Final Semester Evaluation in Madrasah Ibtidaiyah Thotiqul Huda Malang

2021, 699–708.

Rahman, R., & Fuad, M. (2023). Implementasi Kurikulum Merdeka Belajar Dalam Pembelajaran Ips Di Sekolah Dasar. *DISCOURSE: Indonesian Journal of Social Studies and Education*, 1(1), 75–80. <https://doi.org/10.69875/djosse.v1i1.103>

Sholihatul Atik, H. (2020). EFEKTIFITAS MODEL PEMBELAJARAN DARING DALAM PERKULIAHAN BAHASA ARAB Institut Agama Islam Sunan Kalijogo Malang Volume 2 Nomor 1 Juni 2020 I . PENDAHULUAN Berdasarkan obsevasi di lapangan , terdapat perbedaan sikap mahasiswa terhadap perkuliahan Bahasa Ara. *Muhadasah*, 2(1), 77–89.

Yahrif, M., & R. Supardi. (2023). Pendampingan Implementasi Kurikulum Meredeka Melalui Pelatihan Komite Pembelajaran Sekolah Penggerak. *ABDI SAMULANG: Jurnal Pengabdian Kepada Masyarakat*, 2(2), 90–99. <https://doi.org/10.61477/abdisamulang.v2i2.23>