

## Question Items Analysis of Nahwu Lesson based on Articulate Storyline Learning Media

Riza Nurlaila<sup>1\*</sup>, Muhammad Wahyudi<sup>2</sup>, Alfy Mamduh Nuruddin<sup>3</sup>, Hamdan Fauzan Al  
Anshori<sup>4</sup>

<sup>1234</sup>Universitas Darussalam Gontor, Indonesia

\*Corresponding E-mail: [rizanurlaila@unida.gontor.ac.id](mailto:rizanurlaila@unida.gontor.ac.id)

### Abstract (10pt)

This study aims to analyze the quality of question items of nahwu lesson, specifically utilizing Articulate Storyline for Class XI at Senior High School 1 Cimahi 2024/2025 Academic Year in terms of validity, reliability, difficulty level, and differentiating power. The research is a descriptive quantitative research with using data collection technique by documentation. The object of this study is the question items of Nahwu Lessons about Asma' after implementation of arabic learning media based on Articulate storyline for Class XI at Senior HighSchool 1 Cimahi. Based on the findings, then the results showed that: 1) The Validity test of 2 of 5 multiple choice test was declared valid (40%), 5 of 5 essay test was declared valid (100%), and 4 of 5 matching test they were declared valid (80%). 2) The test reliability result indicates that the reliability score of the question items of the multiple choice test is 0.06, which is very low, the reliability score of the question items of the essay test is 0.74, which is high, and the reliability score of the question items of the paired test is 0.60, which is moderate. 3) The difficulty level test of the multiple choice questions categorized as medium is 3 items (60%), and easy is 2 items (40%). The difficulty level test of the essay questions categorized as medium is 3 items (60%), and easy is 2 items (40%). And The difficulty level test of the matching test categorized as medium is 4 items (80%), and easy is 1 item (20%). 4) The differential power of the multiple choice test categorized as bad is 1 item (20%), good is 2 items (40%), very good is 2 items (40%). The differential power of the essay test categorized as enough is 1 item (20%), good is 2 items (40%), very good is 2 items (40%). And the differential power of the matching test categorized as good is 1 item (20%), and very good is 4 items (80%).

**Keywords:** Question Items Analysis, Nahwu Lesson, Articulate Storyline, Arabic Learning Media.

### مقدمة

تقوم الوسيلة التعليمية Articulate Storyline كإيجاد الحلول التكنولوجية لتحسين العملية التعليمية التفاعلية من خلال تقديم الصور والرسوم المتحركة والتدريبات الممتعة (Rahmawati, 2022). ولم يكف في تطويرها بأداء التجربة والتنفيذ لدى الطلاب، بل من حاجة المعلم كالمحور الرئيسي في الفصل عملية التقييم نحو الأسئلة التي تكون أداة لقياس نتائج تعلم الطلاب في تلك الوسيلة التعليمية.

إن فنون وزير التعليم والثقافة بجمهورية إندونيسيا رقم 23 سنة 2016 ينظم في معايير تقييم التربية والتعليم، وإحدى منها أداة تقييم نتائج الطلاب كالمعيار الأساسي في المدارس بإندونيسيا (Juandi, 2019). والتقييم هو عملية الإجراءات لتقيس مدى الهدف المعين في العملية التعليمية ثم استخلاص القرارات ومتابعتها. ولتحصل على ذلك يحتاج التقييم أداة وهو الاختبار أو غير الاختبار (Handriawan & Nurman, 2021). فالتقييم التربوي هو نشاط أو عملية لتحديد قيمة التعليم بحيث يمكن معرفة الجودة أو النتائج. أما نجاح عملية التقييم لنتائج التعلم معتمد على خصائص تصميم الأسئلة وأداء الامتحان وعملية تحليل النتائج فيه (Al Fathiyah, 2019). وبذلك مواصفات الاختبار الجيد هي إحدى العوامل المؤثرة في نجاح عملية التقييم في المدارس والمعاهد. فتعتبر التدريبات في وسائل التعليم جزءاً من محتوى تلك الوسائل نفسها. وتصميمها يناسب بمحتوى مضمون المادة حسب الأهداف المعينة في الوسيلة التعليمية.

تحليل بنود الأسئلة هو نشاط مراجعة أسئلة الاختبار من أجل الحصول على أسئلة ذات جودة مناسبة. الغرض من هذا النشاط هو مراجعة كل عنصر يهدف إلى المساعدة في تحسين الاختبار من خلال التصحيح أو التخلص من الأسئلة غير الفعالة (Afida, 2023). بنود الأسئلة الجيدة هي البنود التي تستوفي خصائصها الشروط كمعايير لخصائص العناصر الجيدة. سيتم إلغاء تحليل العناصر التي تم تحليلها لأن خصائصها لم تستوف متطلبات العناصر الجيدة بحيث لا تمتلك القدرة على قياس نتائج التعلم بشكل جيد. التحليل الذي يجب إجراؤه للمعلم هو تحليل بنود الأسئلة. تُعد أنشطة تحليل بنود الأسئلة من أهم أجزاء النشاط في إعداد الأسئلة من أجل الحصول على بنود الأسئلة ذات جودة عالية (Mahmudi & Ahmad, 2020).

لتحليل البنود بناءً على النظرية الكلاسيكية، هناك أربع خطوات على الأقل، وهي اختبار الصلاحية، والموثوقية، ومستوى الصعوبة، والتمييز (Farida et al., 2021). بناءً على الخلفية السابقة أكدت الباحثة البحث عن تحليل بنود الأسئلة في مادة النحو المؤسس على استخدام الوسيلة التعليمية Articulate Storyline للطلاب الصف الحادي عشر بمدرسة عالية حكومية 1 مدينة تسماحي.

## منهجية البحث

نوع البحث المستخدم في هذا البحث هو البحث الكمي الوصفي. ويستخدم البحث أسلوب جمع البيانات بطريقة الوثيقة وهي الأسئلة في مادة النحو عن موضوع الأسماء بالمدرسة العالية الحكومية الأولى تسماحي مؤسسا على استخدام الوسيلة التعليمية Articulate Storyline. الأسئلة المصممة تحتوي على 15 بنود الأسئلة وتفصيلها ما يلي: 5 بنود الأسئلة من اختبار اختيار المتعدد، و5 بنود الأسئلة من الاختبار الوصفي، و5 بنود الأسئلة من اختبار المزاجية. استخدمت الباحثة برنامج Anates 4.0.5 كأداة المساعدة لتحليل البيانات في هذا البحث. ويتم إجراء تحليل البيانات على صدق الاختبار أو صلاحية الاختبار وثباته ومستوى الصعوبة والقوة التمييزية.

## النتيجة والمناقشة

نتيجة تحليل بنود الأسئلة في مادة النحو المؤسس على استخدام الوسيلة التعليمية Articulate Storyline وهي:

### أ. صدق الاختبار

صدق الاختبار هو دقة القياس التي يمتلكها البند في قياس الأسئلة. إن بنود الأسئلة ذات صلاحية عالية إذا كان هناك ارتباط إيجابي كبير بين درجة البند والدرجة الكلية (Mahmudi & Ahmad, 2020). وبعد إتمام الإحصاء بمساعدة التطبيق Anates 4.0.5 فيحصل الاختبار لمادة النحو المؤسس على الوسيلة التعليمية Articulate Storyline على

قبل إجراء الاختبار إلى الطلاب،

وفقاً لعقد الاختبار في مادة النحو المؤسس على الوسيلة التعليمية Articulate Storyline للطلاب الصف الحادي عشر بمدرسة عالية حكومية 1 مدينة تسماحي، فحصلت نسبة صلاحية من خمسة عشر سؤالاً لهذا الاختبار على 73,3% وهي 11 أسئلة صادقة، أما نسبة غير صادقة حصلت على 26,67% وهي 4 أسئلة. نتيجة الصلاحية لخمس أسئلة من أسئلة اختبار المتعدد هي أسئلتان معقولان (40%)، أما بالنسبة للأسئلة غير صادقة فهناك ثلاثة أسئلة (60%). وكانت نتيجة الصلاحية لخمس أسئلة من بنود أسئلة الاختبار الوصفي معقولة (100%) ولم توجد الأسئلة غير معقولة. ونتيجة

الصلاحية لخمسة أسئلة من بنود أسئلة اختبار المزاوجة هي أربعة أسئلة معقولة (80%)، أما بالنسبة للأسئلة غير صادقة فهناك أسئلة واحدة (20%). ويوضح تحليل النتيجة حسب أنواع الاختبار إلى ما يلي:

1. درجة الصدق لاختبار الاختيار من متعدد بعد إجراء العملية الحسابية، تتم مطابقة النتيجة مع جدول  $r$  عند مستوى الدلالة 5% ومع خمس أسئلة في اختبار المتعدد من خمسة عشر طالبا، فيكون جدول  $r$  هو 0.576. وإذا كان جدول  $r \leq r$  حساب  $r$ ، فإن عناصر السؤال صادقة. ولكن إذا كان جدول  $r \geq r$  حساب  $r$ ، فإن عناصر السؤال غير صادقة. بشكل عام، نتيجة الصلاحية أو الصدق لخمسة بنود من أسئلة الاختبار المتعدد هي أن سؤالين صادقان (40%) وهي البنود رقم 1 و 5. أما بالنسبة للأسئلة غير صادقة فهناك ثلاثة أسئلة (60%) وهي البنود رقم 2، و3، و4.

No Butir Baru	No Butir Asli	Korelasi	Signifikansi
1	1	0.675	Signifikan
2	2	-0.131	-
3	3	0.476	-
4	4	0.564	-
5	5	0.647	Signifikan

الصورة 1: درجة الصدق لاختبار الاختيار من متعدد (مرجع: Anates 4.0.5)

الجدول 1: النسبة المئوية من اختبار الاختيار من متعدد

الرقم	درجة الصدق	بنود الأسئلة	مجموع الأسئلة	النسبة المئوية
1	أكبر من 0.576	1 و 5	2	40%
2	أصغر من 0.576	2، 3، 4	3	60%

2. درجة الصدق للاختبار الوصفي تتم مطابقة النتيجة مع جدول  $r$  عند مستوى الدلالة 5% ومع خمس أسئلة في الاختبار الوصفي من خمسة عشر طالبا، فيكون جدول  $r$  هو 0.576. وإذا كان

جدول  $r \leq$  حساب  $r$ ، فإن عناصر السؤال صادقة. ولكن إذا كان جدول  $r \geq$  حساب  $r$ ، فإن عناصر السؤال غير صادقة. بشكل عام، نتيجة الصلاحية لخمس بنود من أسئلة الاختبار الوصفي هي أن جميع البنود صادقة (100%) من رقم 1 إلى 5، لأن الحاصل من درجة الصدق أكبر من جدول  $r$ .

No Butir Baru	No Butir Asli	Korelasi	Signifikansi
1	1	0.686	Signifikan
2	2	0.747	Sangat Signifikan
3	3	0.717	Sangat Signifikan
4	4	0.622	Signifikan
5	5	0.637	Signifikan

الصورة 21: درجة الصدق للاختبار الوصفي (مرجع: Anates 4.0.5)

الجدول 21: النسبة المئوية من اختبار الاختيار الوصفي

الرقم	درجة الصدق	بنود الأسئلة	مجموع الأسئلة	النسبة المئوية
1	أكبر من 0.576	1، 2، 3، 4، 5	5	100%
	(صادق)			
2	أصغر من 0.576	-	0	0%
	(غير صادق)			

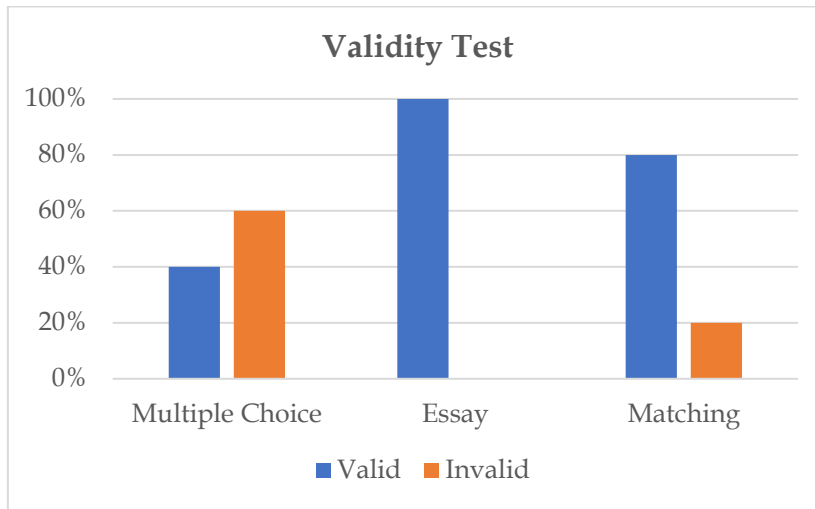
3. درجة الصدق لاختبار المزاوجة  
تتم مطابقة النتيجة مع جدول  $r$  عند مستوى الدلالة 5% ومع خمس أسئلة في الاختبار الوصفي من خمسة عشر طالبا، فيكون جدول  $r$  هو 0.576. وإذا كان جدول  $r \leq$  حساب  $r$ ، فإن عناصر السؤال صادقة. ولكن إذا كان جدول  $r \geq$  حساب  $r$ ، فإن عناصر السؤال غير صادقة. فنتيجة الصلاحية أو الصدق لخمس بنود من أسئلة اختبار المزاوجة هي أن أربعة أسئلة صادقة (80%) وهي البنود رقم 1، 3، 4، و5. أما بالنسبة للأسئلة غير صادقة فهناك أسئلة واحدة (20%) وهي البند رقم 2.

No Butir Baru	No Butir Asli	Korelasi	Signifikansi
1	1	0.630	Signifikan
2	2	0.537	-
3	3	0.739	Sangat Signifikan
4	4	0.600	Signifikan
5	5	0.577	Signifikan

الجدول 3: النسبة المئوية من اختبار المزاجية

الرقم	درجة الصدق	بنود الأسئلة	مجموع الأسئلة	النسبة المئوية
1	أكبر من 0،576 (صادق)	5 ،4 ،3 ،1	4	%80
2	أصغر من 0،576 (غير صادق)	2	1	%20

وفقاً للحاصل من صدق الاختبار سابقاً، يستتبط بأن 11 بنود من الأسئلة صادقة ويُستخدم لعقد الاختبار في مادة النحو المؤسس على الوسيلة التعليمية Articulate Storyline للطلاب الصف الحادي عشر بمدرسة عالية حكومية 1 مدينة تسماحي. أما الباقية هي 4 بنود من الأسئلة غير صادقة ولا يُستخدم في عقد الاختبار. وسيتم عرض النتائج من خلال الرسم البياني ما يلي:



الرسم البياني 1: صدق الاختبار

## ب. ثبات الاختبار

إن الاختبار الذي يصل إلى مستوى عالٍ من الثقة إذا يمكن له أن يقدم نتائج ثابتة. ثبات الاختبار هو دقة النتائج، فإذا إجراء الاختبار في بعض الأماكن فتبقى الثبات كما هو. أما إذا تغيرت النتائج، فيمكن القول إن التغييرات التي تحدث تكون غير مهمة (Farida et al., 2021). تتعلق المطالبة بضرورة أن تكون أساليب التقييم صالحة بتوقع

الحصول على بيانات صحيحة. في هذه الحالة، لا تختلف متطلبات الموثوقية كثيراً. فإذا كانت الصلاحية تتعلق بصحة الشيء، وهي ليست سوى عدم انحراف البيانات عن الواقع، بمعنى أن تكون البيانات صحيحة، والبيانات دقيقة، والبيانات موثوقة (Handriawan & Nurman, 2021). بالواقع، بمعنى أن البيانات صحيحة، فإن مفهوم الموثوقية يرتبط بتعدد التصوير عدة مرات. الأداة الجيدة هي الأداة التي يمكنها توفير بيانات تتوافق مع الواقع بشكل موثوق.

استخدمت الباحثة برنامج Anates 4.0.5 لتحليل درجة الثبات. والمعايير المستخدمة لحساب النتيجة الموثوقية أو درجة الثبات هي:

الجدول 3: معايير الثبات

بيان الدرجة	معيير الثبات
مرتفع جدا	1-0,90
مرتفع	0,89-0,70
متوسط	0,69-0,50
منخفض	0,49-0,30
منخفض جدا	0,30<

درجة الثبات لاختبار المتعدد هي 0,06، فنتضح من معايير الثبات سابقا بأن الحاصل منخفض جدا. أما درجة الثبات للاختبار الوصفي هي 0,74، فنتضح من معايير الثبات سابقا بأن درجته مرتفعة. وأما درجة الثبات لاختبار المزاوجة هي 0,60، فنتضح من معايير الثبات سابقا بأن الحاصل متوسط الثبات. وهذا البيان يأتي من نتيجة البيانات في ثبات الاختبار باستخدام البرنامج المذكور التي توضحها الصور التالية:

Rata2= 3.47  
Simpang Baku= 1.13  
KorelasiXY= 0.03  
Reliabilitas Tes= 0.06

الصورة 4: درجة الثبات لاختبار المتعدد (مرجع: Anates 4.0.5)

Rata2= 7.53  
Simpang Baku= 2.39  
KorelasiXY= 0.58  
Reliabilitas Tes= 0.74

الصورة 5: درجة الثبات للاختبار الوصفي (مرجع: Anates 4.0.5)

Rata2= 2.80  
Simpang Baku= 1.52  
KorelasiXY= 0.43  
Reliabilitas Tes= 0.60

الصورة 6: درجة الثبات لاختبار المزوجة (مرجع: Anates 4.0.5)

### ج. مستوى الصعوبة

مستوى الصعوبة لبنود الاختبار هو نسبة المتقدمين للاختبار الذين يجيبون بشكل صحيح على أحد بنود الاختبار. في حين أن الرقم الذي يشير إلى صعوبة بنود الاختبار يُسمى مؤشر الصعوبة والذي يُرمز له بالرمز  $p$  (نسبة الإجابة الصحيحة). كلما زادت قيمة  $p$  تعني أنه كلما زادت نسبة المتقدمين للاختبار الذين أجابوا بشكل صحيح على بنود الاختبار، كلما انخفض مستوى صعوبة بنود الاختبار، مما يعني أن بنود الاختبار أسهل. وعلى العكس من ذلك، كلما كانت قيمة  $p$  أصغر يعني أن نسبة من المتقدمين للاختبار الذين أجابوا بشكل صحيح على بنود الاختبار، كلما كان مستوى صعوبة بنود الاختبار أعلى. كلما زادت قيمة  $p$  تعني كلما ارتفع مستوى صعوبة بنود الاختبار، مما يعني أن بنود أكثر صعوبة (Widoyoko, 2018).

بنود الأسئلة الجيدة هي التي لا يكون سهلاً ولا صعباً للغاية. إذا كان العنصر سهلاً أو صعباً جداً، لا يمكن أن يعكس البند تحقيق نتائج التعلم. الأسئلة السهلة جداً لا تحفز الطلاب على زيادة جهودهم لحلها. وعلى العكس من ذلك، فإن الأسئلة الصعبة للغاية ستؤدي إلى تثبيط عزيمة الطلاب وعدم تحليهم بالروح للمحاولة مرة أخرى لأنها خارجة عن متناولهم (Mahmudi et al., 2023). قائمة درجة الصعوبة الجيدة للأسئلة هي على نسبة 2.5:5:2.5 يعني أن درجة الصعوبة الجيدة هي إذا بلغت 25% من السؤال السهل، و50% من السؤال المتوسط، و25% من السؤال الصعب (Firmansyah, 2018). والمعايير المستخدمة لتفسير الموثوقية هي كالتالي:



الجدول 4: معايير مستوى الصعوبة

بيان الدرجة	معيان الدرجة
صعب	0,30-0,00
متوسط	0,70-0,31
سهل	1-0,71

### 1. مستوى الصعوبة لاختبار المتعدد

حصلت نتائج اختبار مستوى الصعوبة لأسئلة اختبار المتعدد على 0 بنود الأسئلة الصعبة (0%)، وحصلت على 3 بنود الأسئلة المتوسطة (60) وهي البنود رقم 3 و4 و5، وحصلت على بندي الأسئلة السهلة (40%) وهما البنود رقم 1 و2. وهذا البيان يأتي من نتيجة البيانات في مستوى الصعوبة باستخدام البرنامج Anates التي توضحها الجدول والصورة التالية:

الجدول 5: النسبة المئوية في مستوى الصعوبة لاختبار المتعدد

معيان درجة مستوى الصعوبة	بنود الأسئلة	عدد البنود	النسبة المئوية
0,30-0,00 (صعب)	-	-	0%
0,70-0,31 (متوسط)	5، 4، 3	3	60%
1-0,71 (سهل)	2، 1	2	40%

No Butir Baru	No Butir Asli	Jml Betul	Tkt. Kesukaran(%)
1	1	11	73.33
2	2	14	93.33
3	3	9	60.00
4	4	10	66.67
5	5	8	53.33

الصورة 7: مستوى الصعوبة لاختبار المتعدد (مرجع: Anates 4.0.5)

## 2. مستوى الصعوبة للاختبار الوصفي

حصلت نتائج اختبار مستوى الصعوبة لأسئلة الاختبار الوصفي على 0 بنود الأسئلة الصعبة (0%)، وحصلت على 3 بنود الأسئلة المتوسطة (60%) وهي البنود رقم 3 و4 و5، وحصلت على بندي الأسئلة السهلة (40%) وهما البنود رقم 1 و2. وهذا البيان يأتي من نتيجة البيانات في مستوى الصعوبة باستخدام البرنامج Anates التي توضحها الجدول والصورة التالية:

الجدول 6: النسبة المئوية في مستوى الصعوبة للاختبار الوصفي

النسبة المئوية	عدد البنود	بنود الأسئلة	معيار درجة مستوى الصعوبة
0%	-	-	0,30-0,00 (صعب)
60%	3	5, 4, 3	0,70-0,31 (متوسط)
40%	2	2, 1	1-0,71 (سهل)

No Butir Baru	No Butir Asli	Tkt. Kesukaran(%)
1	1	83.33
2	2	70.83
3	3	68.75
4	4	62.50
5	5	62.50

الصورة 8: مستوى الصعوبة للاختبار الوصفي (مرجع: Anates 4.0.5)

## 3. مستوى الصعوبة للاختبار المزوجة

حصلت نتائج اختبار مستوى الصعوبة لأسئلة اختبار المزوجة على 0 بنود الأسئلة الصعبة (0%)، وحصلت على 4 بنود الأسئلة المتوسطة (80%) وهي البنود رقم 1 و2 و4 و5، وحصلت على بند الأسئلة السهلة (20%) وهي البند رقم 3. وهذا البيان يأتي من نتيجة البيانات في مستوى الصعوبة باستخدام البرنامج Anates التي توضحها الجدول والصورة التالية:

الجدول 7: النسبة المئوية في مستوى الصعوبة للاختبار المزوجة

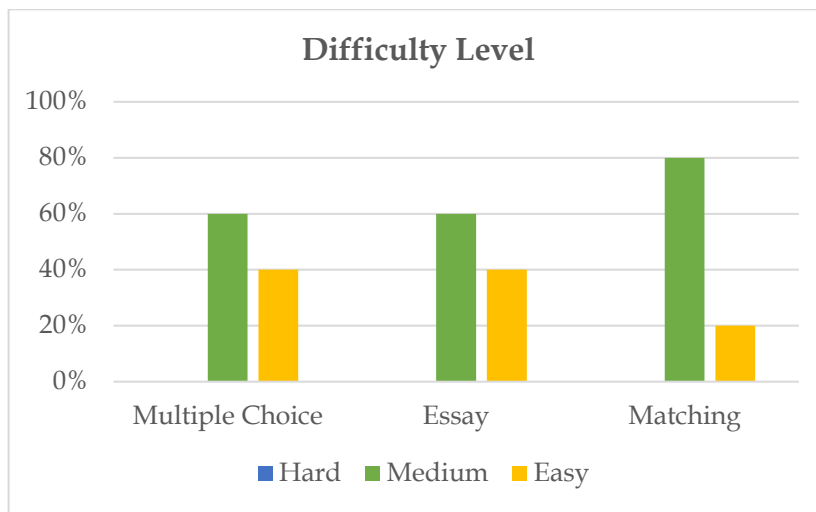
النسبة المئوية	عدد البنود	بنود الأسئلة	معيار درجة مستوى الصعوبة
----------------	------------	--------------	--------------------------

0%	-	-	0,30-0,00 (صعب)
80%	4	5,4,2,1	0,70-0,31 (متوسط)
20%	1	3	1-0,71 (سهل)

No Butir Baru	No Butir Asli	Jml Betul	Tkt. Kesukaran(%)
1	1	9	60.00
2	2	9	60.00
3	3	11	73.33
4	4	8	53.33
5	5	5	33.33

الصورة 9: مستوى الصعوبة لاختبار المزوجة (مرجع: Anates 4.0.5)

وفقاً للحاصل من درجة مستوى الصعوبة سابقاً، يستنتج بأن 5 بنود من الأسئلة سهلة و 10 بنود من الأسئلة متوسطة ولم يوجد بنود الأسئلة صعبة. وسيتم عرض النتائج من خلال الرسم البياني ما يلي:



الرسم البياني 2: درجة مستوى الصعوبة

د. القوة التمييزية هي قدرة السؤال على التمييز بين الطلاب الأنكياء (ذوي القدرات العالية) والطلاب الأغبياء (ذوي القدرات المنخفضة). تُفيد أسئلة تحليل القوة التمييزية في معرفة الطلاب الذين فهموا المادة والطلاب الذين لم يفهموا المادة وكذلك

معرفة الطالب الذكي والطالب الضعيف (Aryobimo, 2021). والمعايير المستخدمة لتفسير القوة التفاضلية هي كما يلي:

الجدول 8: معايير القوة التمييزية

بيان الدرجة	معيان الدرجة
ضعيف	0,20-0,00
مقبول	0,40-0,21
جيد	0,70-0,41
جيد جدا	1-0,71

وبالنظر إلى المعايير السابقة فقد بلغت نسبة الأسئلة 80%، وبذلك تكون القوة التمييزية لهذا الاختبار جيدة وقادرة على التفريق بين الطلاب من حيث استيعابهم للمادة والتفريق بين الطلاب الأذكياء والطلاب الضعفاء. أما بالنسبة لهذه القيم بناءً على نتائج التحليل أدناه:

#### 1. القوة التمييزية لاختبار المتعدد

حصلت نتيجة القوة التمييزية لاختبار المتعدد على أنها ضعيفة هو عنصر واحد (20%) وهو رقم 2، وتدل على أنها جيدة هو عنصرين (40%) وهما رقم 3 و5، أما التي تدل على أنها جيدة جدا هو عنصرين (40%) هما رقم 1 و4. وتوضح النتيجة السابقة في الصورة والجدول التالية:

No Butir Baru	No Butir Asli	Kel. Atas	Kel. Bawah	Beda	Indeks DP (%)
1	1	4	1	3	75.00
2	2	4	4	0	0.00
3	3	3	1	2	50.00
4	4	4	1	3	75.00
5	5	3	1	2	50.00

الصورة 10: القوة التمييزية لاختبار المتعدد (مرجع: Anates 4.0.5)

الجدول 9: النسبة المئوية في القوة التمييزية لاختبار المتعدد

معيان درجة مستوى الصعوبة	بنود الأسئلة	عدد البنود	النسبة المئوية
--------------------------	--------------	------------	----------------

0,20-0,00 (ضعيف)	2	1	%20
0,40-0,21 (مقبول)	-	-	%0
0,70-0,41 (جيد)	3 و 5	2	%40
1-0,71 (جيد جدا)	1 و 4	2	%40

2. القوة التمييزية للاختبار الوصفي  
 حصلت نتيجة القوة التمييزية للاختبار الوصفي على أنها مقبولة هو  
 عنصر واحد (20%) وهو رقم 1، وتدل على أنها جيدة هو عنصرين (40%) وهما  
 رقم 2 و 3، أما التي تدل على أنها جيدة جدا هو عنصرين (40%) هما رقم 4 و 5.  
 وتوضح النتيجة السابقة في الصورة والجدول التالية:

No	No Btr Asli	Rata2Un	Rata2As	Beda	SB Un	SB As	SB Gab	t	DP(%)
1	1	3.00	2.00	1.00	0.00	1.15	0.58	1.73	33.33
2	2	3.00	1.25	1.75	0.00	1.26	0.63	2.78	58.33
3	3	2.00	0.75	1.25	0.00	0.50	0.25	5.00	62.50
4	4	1.00	0.25	0.75	0.00	0.50	0.25	3.00	75.00
5	5	1.00	0.25	0.75	0.00	0.50	0.25	3.00	75.00

الصورة 11: القوة التمييزية للاختبار الوصفي (مرجع: Anates 4.0.5)

الجدول 10: النسبة المئوية في القوة التمييزية للاختبار الوصفي

النسبة المئوية	عدد البنود	بنود الأسئلة	معيار درجة مستوى الصعوبة
%0	-	-	0,20-0,00 (ضعيف)
%20	1	1	0,40-0,21 (مقبول)
%40	2	3 و 2	0,70-0,41 (جيد)
%40	2	5 و 4	1-0,71 (جيد جدا)

3. القوة التمييزية لاختبار المزاجية  
 حصلت نتيجة القوة التمييزية لاختبار المزاجية على أنها جيدة هو عنصر  
 واحد (20%) وهما رقم 2، أما التي تدل على أنها جيدة جدا هو أربعة عناصر

80%) هما رقم 1، و3، و4، و5. وتوضح النتيجة السابقة في الصورة والجدول التالية:

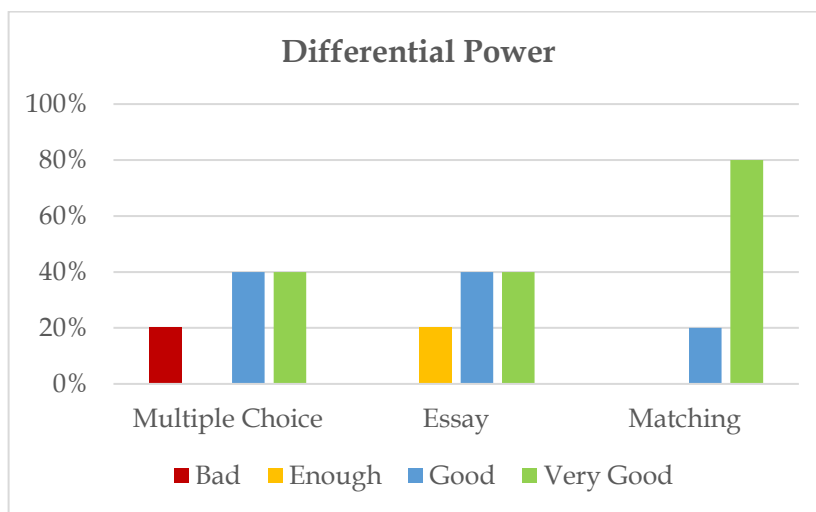
No Butir Baru	No Butir Asli	Kel. Atas	Kel. Bawah	Beda	Indeks DP (%)
1	1	4	1	3	75.00
2	2	3	1	2	50.00
3	3	4	0	4	100.00
4	4	4	1	3	75.00
5	5	4	1	3	75.00

الصورة 12: القوة التمييزية لاختبار المزوجة (مرجع: Anates 4.0.5)

الجدول 11: النسبة المئوية في القوة التمييزية لاختبار المزوجة

النسبة التمييزية	عدد البنود	بنود الأسئلة	معيار درجة مستوى الصعوبة
%0	-	-	0،20-0،00 (ضعيف)
%0	-	-	0،40-0،21 (مقبول)
%20	1	2	0،70-0،41 (جيد)
%80	4	5 و3 و4 و1	1-0،71 (جيد جدا)

وفقاً للحاصل من درجة القوة التمييزية سابقاً، يستنتج بأن بند واحد من الأسئلة ضعيف، وبند واحد من الأسئلة مقبول، و5 بنود الأسئلة جيدة، و8 بنود الأسئلة جيدة جداً. وسيتم عرض النتائج من خلال الرسم البياني ما يلي:



الرسم البياني 3: درجة القوة التمييزية

الخلاصة

بناء على النتيجة ومناقشة تحليل البيانات في بنود الأسئلة لمادة النحو المؤسس على استخدام الوسيلة التعليمية Articulate Storyline للطلاب الصف الحادي عشر بمدرسة عالية حكومية 1 مدينة تسماعي، تستنبط الباحثة في أربعة أمور، هي:

1. تشير نتيجة صدق الاختبار إلى أن بندين من الأسئلة للاختبار المتعدد صادقان (40%) و3 بنود الأسئلة غير صادقة (60%). وأن 5 بنود الأسئلة للاختبار الوصفي صادقة (100%) ولم توجد الأسئلة غير صادقة (0%). أما 4 بنود الأسئلة للاختبار المزوجة صادقة (80%) وبند واحد من الأسئلة غير صادق (20%).
2. وتشير نتيجة ثبات الاختبار إلى أن درجة الثبات من بنود الأسئلة للاختبار المتعدد تحصل على 0,06 وهي منخفضة جداً، ودرجة الثبات من بنود الأسئلة للاختبار الوصفي تحصل على 0,74 وهي مرتفعة، أما درجة الثبات من بنود الأسئلة للاختبار المزوجة تحصل على 0,60 وهي متوسطة.
3. تشير درجة مستوى الصعوبة إلى أن 3 بنود الأسئلة للاختبار المتعدد متوسطة (60%) وبندان من الأسئلة سهلة (40%)، أن 3 بنود الأسئلة للاختبار الوصفي متوسطة (60%) وبندان من الأسئلة سهلة (40%)، أما 4 بنود الأسئلة للاختبار المزوجة متوسطة (80%) وبند واحد من الأسئلة سهلة (20%).
4. تشير درجة القوة التمييزية إلى أن بند واحد من الأسئلة للاختبار المتعدد ضعيف (20%) وبندان من الأسئلة جيدة (40%) وبندان من الأسئلة جيدة جداً (40%)، وأن بند واحد من الأسئلة للاختبار الوصفي مقبول (20%) وبندان من الأسئلة جيدة (40%) وبندان من الأسئلة جيدة جداً (40%)، أما بند واحد من الأسئلة للاختبار المزوجة جيد (20%) و4 بنود الأسئلة جيدة جداً (80%).

وفقاً للخلاصة السابقة اقترحت الباحثة بأن تعقد المدرسة عملية تقييمية نحو الاختبارات خاصة في تحليل بنود الأسئلة حتى يتمكن المعلم كمصمم الاختبار من ملاحظات بنود الأسئلة حيث صدقها وثباتها ومستوى صعوبتها وقوتها التمييزية. ويجب المعلم كمطور وسيلة التعليم أن يناسب أداة التقييم بالمواد التي تعالج مشكلة تعلم الطلاب كما يجب المعلم أن يلاحظ مدى ملاءمة أسئلة الاختبار وإثارتها لأنها تساعد على جذب انتباه الطلاب عند الإجابة عن السؤال في تحليل عناصر الأسئلة.

- Afida, L. R. (2023). تحليل بنود أسئلة الاختبار النهائي الأول لمادة اللغة العربية في الفصل التاسع بمدرسة صراط. *الفصل التاسع بمدرسة صراط*. UIN Maulana Malik Ibrahim.
- Al Fathiyah, Si. F. (2019). ANALISIS BUTIR SOAL PELAJARAN BAHASA ARAB DI MA ROUDLOTUL ULUM PAGAK MALANG. *Tarbiyatuna*, 4(1), 76–100.
- Aryobimo, B. (2021). Analisis Butir Soal Latihan pada Buku *al-Naḥw al-Wāḍiḥ* untuk Siswa Madrasah Aliyah. 18(2), 127–138.
- Farida, Syarifah, & Muhammad. (2021). تحليل بنود الأسئلة في الإمتحان النهائي لمادة اللغة العربية في (2) الفصل العاشر بالمدرسة العالية الإسلامية "الإسلام" كموجا. 1.
- Handriawan, D., & Nurman, M. (2021). *Evaluasi Pembelajaran Bahasa Arab* (1st ed.). Sanabil.
- Juandi, A. (2019). *Standar Penilaian Pendidikan*.  
<https://osf.io/preprints/inarxiv/munp2>
- Mahmudi, I., & Ahmad, R. R. F. (2020). Item Analysis of Islamic Education For Class VIII at Junior High School 1 Jetis Ponorogo. *Educan*, 4(1).
- Mahmudi, I., Nurwardah, A., Rochma, S. N., & Nurcholis, A. (2023). Item Analysis Of Arabic Language Examination. *IJAZ ARABI*, 6(3), 563–573.  
<https://doi.org/10.18860/ijazarabi.v6i3.19821>
- Rahmawati, T. (2022). *Articulate Storyline* في تعليم اللغة العربية استخدام وسيلة التعليم على أساس. UIN Sunan Gunung Djati Bandung.
- Widoyoko, E. P. (2018). *Penilaian Hasil Belajar di Sekolah*. Pustaka Belajar.