International Journal of Arabic Language Teaching
P-ISSN: 2684-690X E-ISSN: 2686-214X
https://e-journal.metrouniv.ac.id/index.php/IJALT

Mustawâ Layâqah al-Ikhtibār li Mâdah al-Lughah al-'Arabiyyah

Yogia Prihartini[1]*, Wahyudi Buska[2], Faridl Hakim[3]
[123]Universitas Islam Negeri (UIN) Sulthan Thaha Siafuddin Jambi, Indonesia

Article History:
Received : 11 November 2022
Revised : 29 November 2022
Accepted : 30 November 2022
Published : 04 December 2022

Keywords:
Properness Level; Final Test; Arabic Material.

***Correspondence Address:**
wahyudi@uinjambi.ac.id

© 2022

DOI: 10.32332/ijalt.v4i02.5710

Abstract: This study aims to describe: The types of tests in the final exams of the first semester, the quality of the tests, and the reasons why teachers do not carry out this test analysis. The sample of this study was students majoring in Arabic using purposive sampling using the percentage formula for validity, Cronbach Alpha for reliability, difficulty formula, and differentiation formula. The results of this study were: (1) the forms of the tests were achievement tests (learning objectives), teacher-made tests (writers), summative tests (test time), written tests (implementation), and subjective tests (how-to scores according to point of view. (2) The quality of a test is: (a) The validity of the test is good because it fulfills content validity, although the test items in the first-semester test are based on language structure 41.67% and vocabulary aspects 25% and reading skills 33.35 (b) Reliability of the test was higher at 5.87 (c) The level of difficulty of the very easy test is 0.57%. (d) The differential power test is not good enough by 0.52%. (3) The reason the teacher does not carry out an analytical test is (a) Lack of understanding of problem analysis, (b) They did not experiment in carrying out analytical tests, (c) They did not understand the ins and outs of revaluation, (d) They did not understand problem analysis and, (e) There were teachers who had not received training or seminars in the field of evaluation and analysis.

مقدمة

إن الاختبار الجيد أساسي للتعليم الجيد والتعلم الجيد وأن شروط الاختبار الجيد هو الصدق والثبات وسهولة التدريج والتمثيل والتمييز والزمن.[1] كما قال كاسرام إن الطريقة الجيدة في التقييم هي ماكانت نتيجة التقييم وصفا واضحا عن عنوان التقييم. وأما لمعرفة الأداة الجيدة في التقييم فلابد من أن تتكون من الأمور الاتية: الصدق والثبات ومستوى الصعوبة والتمييز.[2] وقد تطورت بنود السؤال لايستطيع أن يقدر ما هو المنشود. وكانت بنود السؤال المتطورة غير صادقة. وهي لا تقيس الشيء الذي وضع من أجله، مع أن الاختبار

[1] Muhammad 'Alî Al Khûlî, *Asâlîbu Tadrîsi Al Lughati Al 'Arabiyati* (Riyadh: al Mamlakatu al 'Arabiyyatu al Su'ûdiyyah, 1986).

[2] Yogia Prihartini, Wahyudi Buska, and Nur Hasnah, 'Analysis of Test Item on The Final Test Semester Exam on Arabic Subjects', *Arabiyat: Jurnal Pendidikan Bahasa Arab Dan Kebahasaaraban*, 6.1 (2019), 71–92.

الجيد هو الذي يناسب بأهداف المادة التعليمية التى علمها المعلم.[٣] والمادة التعليمية نفسها متطورة على أساس الهدف. ومن ثم توجد العلاقة بين الأجزاء الثلاث كمثل الأهداف والمادة التعليمية الآت الاختبار.[٤]

وللوصول الى أدة التقييم الحسنة الموافقة بالأمور المتقدمة ذكرها، فهناك الخطوات التى لابد على المعلم بإستعدادها قبل بداية عملية التقويم وهي تعيين هدف وتعيين عنوان وتعيين طريقة وإعداد أداة.[٥] الامتحان النهائي ويسمى الاختبار الاجمالى. الإختبار الإجمالى (*Summative-test*) هو الاختبار الذى يقوم به المعلم فى نهاية برنامج الدراسة لإنتاج فعالية برنامجها والقيام بمعيارها، والنتيجة من هذا التقييم هى معرفة درجة قدرة الطالب فى إستيعاب المواد الدراسية المعينة.[٦] وهذا الاختبار عقد فى نهاية المرحلة من المراحل الدراسية، إما الإختبار لنصف السنة أو آخرها. ووقد الإختبار الإجمالى يسمى بالإختبار النهائي ويقصد به لمعرفة نجاح الطالب ومقياسه فى التعلم ولتعيين القرار عن نجاحه ورسوبه.

الامتحان النهائي للمرحلة الأولى لمادة اللغة العربية قيام فى كل آخر سنة. وتنفيذه ليس فى مرحلة الإبتدائية فحسب، ولكن تنفيذه كذلك فى مرحلة المتوسطة والثانوية والجامعة. وكآلة لقيمة وصل التعليمى، فواجب على الاختبار فى الامتحان النهائي لمادة للغة العربية إستوفى مواصفات الاختبار الجيد إما من ناحية الصدق والثبات وإما من ناحية مستوى الصعوبة والتمييز.

لقد تقدم الباحثون البحث المتقدم عن الأساتذة في شعبة اللغة العربية بجامعة سلطان طه سيف الدين الإسلامية الحكومية جامبى، الذى يكون موضوعا لهذا البحث. ومن نتيجة هذا البحث المتقدم هناك المظاهر التى تكون وصفا عن النقصان لدى المعلم أو الأساتذة الذى لابد على إصلاحه، خصوصا فيما يتعلق بتقييم نتيجة التعليم. وأما الوصف من تلك المظاهر هى: (١) كان الأساتذة في شعبة اللغة العربية بالجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى يقومون تحليل نموذج الاختبار في الامتحان النهائى بدون التخطيط ولايفهمون أو لا يستطيعون ان يستعيبوا كيفية كتابة أسئلته. (٢).كان الأساتذة بالجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى لم يفهمون تحليل نموذج أسئلة الإختبار وأسبابها الأتية: عدم إستيعاب عن تحليل نموذج أسئلة الإختبار، عدم الخبرة والمعلومات والمفاهيم عن تحليل نموذج الاختبار، نقصان معرفة عن التقييم وتحليل نموذج الإختبار، عدم الفهم الدقيق عن كيفية تحليل نموذج الاختبار، لم يشتركوا فى التدريب أو التمرين عن التقييم وتحليل نموذج الاختبار.

---

[3] James E Rebele and E Kent St Pierre, 'A Commentary on Learning Objectives for Accounting Education Programs: The Importance of Soft Skills and Technical Knowledge', *Journal of Accounting Education*, 48 (2019), 71–79.

[4] M Ngalim Purwanto, 'Prinsip-Prinsip Dan Teknik Evaluasi Pengajaran', 2019.

[5] Wayan Nurkancana and P P N Sumartana, 'Evaluasi Pendidikan', *Surabaya: Usaha Nasional*, 1986.

[6] Ricky Lam, 'Formative Use of Summative Tests: Using Test Preparation to Promote Performance and Self-Regulation', *The Asia-Pacific Education Researcher*, 22.1 (2013), 69–78.

وهذه كلها هى التى تجعل وظيفة تحليل الاختبار ونتيجته لم تكن شاملة حتسلم تعرف عن نوعية الاختبار بحق المعرفة.هل لها مواصفات الاختبار الجيد أم لا؟ أو هل لها ثباتا وصدقا فيه؟ أو كيف مستوي الصعوبة والتمييز فيه؟ مع أن هذه النشاطات مهمة لأن بها قد تخرج المعلومات عن نوعية الاختبار حتى يمكن الإصلاح فى أسئلته إما بضياعها أو بتغير بعضها.

وأما الفوائد من نتيجة تحليل الاختبار بجانب معرفة نوعية الاختبار، فإنها تعطى وصفا عن ثلاثة العوامل التعليمية وهى: ١) الهدف، إلى أى حد كان الهدف قد وصل، ٢) عملية التعليم، صحة الطريقة وصحة أسلوب فيها، ٣) التقييم، نوع الاختبار وطريقة تعيين البنود منه.وبإصلاح هذه العوامل الثلاثة فيه، يرجى منه أن تكون عملية التعليم أحسن النفذ والجذبة والقدرة للوصول إلى الغاية المطلوبة منها.[٧]

إن الاختبار النهائي لا بد تملك معايير الاختبار الجيد.أما الاختبار الجيد عند عبد الوهاب عبد السلام طويلة في كتابه" *التربية الإسلامية وفن التدريس*" أن الإختبار الجيد كما يلي:(١) الوضوح في المعنى والبعد عن الغموض.(٢) جودة الخط والسلامة من الأخطاء، فيجيب أن تكتب بخط واضح وتصاغ بعبارات خالية من الأخطاء اللغوية والإملائية، ويخلو محتواها عن الأخطاء العلمية والفنية.(٣) أن تكون الأسئلة ملائمة لمستوى الطلاب العلمي والزمني. (٤) الملائمة للزمن المقرر،فيقدر المدرس الزمن المتوقع للإجابة عن كل سؤال ثم عن جميع الأسئلة،بحيث لا تستغرق الوقت المقرر كاملا،ليتمكن الطالب من مراجعة إجابته،وبحيث لا ينتهي من الإجابة بعد بدء الإختبار بوقت قصير. التدرج في السهولة والصعوبة مراعاة للفروق الفردية،فيجيب أن تشمل على عناصر سهلة تمكننا من قياس الضعفاء وعناصر صعبة تمكننا من قياس الأشياء،وبذلك تظهر مستويات الطلاب بين ضعيف.[٨]

ذكر طعيمة أن هناك المواصفات الأساسية للاختبار الجيد وهى خمسة: (١) الصدق؛ ويقصد به أن الاختبار يقيس ما وضع لقياسه. (٢) الثبات؛ ويقصد به أن الاختبار يعطى نفس النتائج لنفس المجموعة من الأفراد إذا ما طبق مرة أخرى فى نفس الظروف، وبعد مسافة قصيرة. (٣) الموضوعية؛ ويقصد بها عدم تأثير شخصية المصحح على وضع أو تقدير علامات الطلاب فى الاختبار. (٤) العملية؛ ويقصد بها أن الاختبار لا يتطلب من المعلم جهدا كبيرا سواء فى وضعه أو تطبيقه أو تصحيحه. (٥) التمييز؛ ويقصد به أن الاختبار يستطيع أن يبرز الفروق بين الطلاب فيبين لنا الأقوياء من الضعاف.[٩] ومن أهم هذه الصفات هي الصدق و الثبات ومستوى الصعوبة والتمييز.قرّر اللغوى أربع مواصفات الاختبار الجيد وهى فيما يلى :أولا، الصـــدق (Validity) ، وهو من الاختبار هو علامة مهمة و أولا التى لابد أمتلاكها. وبها يمكن على معرفة

[7] Logan Fiorella and Richard E Mayer, 'The Relative Benefits of Learning by Teaching and Teaching Expectancy', *Contemporary Educational Psychology*, 38.4 (2013), 281–88.

[8] 'Abdu al Wahhâbi 'Abdu al Salâmi Thuwailah, *Al Tarbiyyatu Al Islâmiyyatu Wa Fannu Al Tadrîsi*, (Riyadh: Dâru al Salâmi, 1992). h..٢٤٢-٢٤٣

[9] Rusydi Ahmad Thu'aimah, 'Ta'lim Al-Lugah Li Ghairi Nathiqina Biha: Mana> Hijuhu Wa Asa> Libuhu', *Ar-Ribat: UCESCO*, 1989. h. .٢٤٧

المعايير والتقييمية المستخدمة فى معيار عما وجب على معياره بالضبط.يقصد بالصادق هو الاختبار الذي يقيس ما يدعى أنه يقيسه.[١٠] وقال براون في كتابه"*أسس تعليم اللغة و تعليمها*"أن صدق الاختبار يعني أن الاختبار نجح فى قياس ما يهدف إلى قياسه.[١١] وهذا يعني إلى أي مدى يقيس الاختبار الشيء الذي وضع من أجله. وقد يعنى صدق الاختبار بالصحيح، وأما الثبات يعنى بالمنشود.[١٢] وتكون آداة المقياس صدقا إذا كان له القدرة فى تعيين الشيء أو معناها المناسبة بين آداة المقياس ووظيفته ووسائله.[١٣]

ينقسم صدق الاختبار إلى أربعة أقسام وهي: أ)- صدق المحتوى (*Content Validity*). إن الاختبار يكون صادقا من ناحية محتوى إذا قياس الهدف الخاص المعين مناسب بالمواد أو بمضمون الدراسية المكتوبة في المنهج،وقديسمى كذالك بصدق المنهج.[١٤] فى إثناء هذا، كأساس المشاهدة لاستعمال هذا الاختبار هو المنهج أو المحتوى من مادة الدروس التى قد تعلمه. من هذه معرفة هل ذالك الاختبار قد كان مناسبا بمادة الدروس المعطى أو بمنهج ثبّته، أو بالعكس. إذا كان ذلك الاختبار قد كان مناسبا بمادة الدروس، فيقول أن ذلك الاختبار له صدق المحتوى. ولكن إذا كان ذلك الاختبار لا يتصل بالدروس فيقول أن ذلك الاختبار ليس له صدق المحتوى. ب)- الصدق الظاهري (*Face Validit*)، ولا بد الاختبار من أن يكون مصوغا من حيث الظاهر بحيث يقنع الدارس بأن الاختبار يمكن أن يقيس قدراته.[١٥] وعلى سبيل المثال، لاختبار اللغة العربية الصدق الظاهري أذا كانت أسئلة فيه باللغة العرابية،وليست فيه اللغة الإندونيسية. ج)- صدق بنود الاختبار وكذلك د)- الصدق التجريبي (*Empirical Validity)*

ويسمى أحيانا الصدق الأحصائي.هذا النوع من الصدق عن طريق مقارنة نتائج الاختبار بنتائج معيارية أخرى يعتقد أنها صادقة،مثل إختبار أخر ثبت أنه صادق أو في ضوء تقديرات مدرس الصف التى تعطي في نهاية المقرر الدراسي أو أي مقياس أخر مستقل.

يقصد بالثبات (Reliability) عدم التذبذب فى الاختبار إذا ما قصد به أن يكون بمثابة المقياس. فالمقياس المترى مثلا يمكن أن تقيس به الطول والعرض لعدة أشياء، ويمكن بعد فترة أن تقيس الأشياء نفسها بالمقياس المترى نفسه وتحصل على النتائج نفسها دون تذبذب مادام الطول والعرض كما هما لم يتغيرا.[١٦]

---

[10] Howard Schuman, 'The Random Probe: A Technique for Evaluating the Validity of Closed Questions', *American Sociological Review*, 1966, 218–22.

[11] Douglas Brown, *Ususu Ta'lîmi Al Lughati Wa Ta'lîmuhâ (Tarjamahu 'Abdu Al Râjihi Wa 'Alî Ahmad Sya'Bânu)* (Bairut: Dâru al Nahdhati al 'Arabiyyati, 1994). h.٢٦٨

[12] Jack R Fraenkel, Norman E Wallen, and Helen H Hyun, *How to Design and Evaluate Research in Education* (McGraw-hill New York, 2012), VII.

[13] Roderick P McDonald, *Test Theory: A Unified Treatment* (psychology press, 2013).

[14] Suharsimi Arikunto, *Dasar-Dasar Evaluasi Pendidikan* (Jakarta: Bumi Aksara, 2021).

[15] Grant Wiggins, 'A True Test: Toward More Authentic and Equitable Assessment', *Phi Delta Kappan*, 92.7 (2011), 81–93.

[16] Vijay Govindarajan and Praveen K Kopalle, 'Disruptiveness of Innovations: Measurement and an Assessment of Reliability and Validity', *Strategic Management Journal*, 27.2 (2006), 189–99.

وأما الرمز المستخدم لتحليل مستوى الثبات(Reliability) هو:[١٧]

(١) حساب متغير كل البند باستخدام الرمز :

$$Sn^2 = \frac{\sum(n^2) - \frac{(\sum n)^2}{N}}{N}$$

حيث :

Sn^2 : متغير البند $(\sum n)^2$: مربعات مجموع درجات البند

$\sum(n^2)$: مجموع مربعات درجات البند N : عدد الطلاب المشتركين فى الإختبار

(٢) حساب مجموع متغيرات البنود باستخدام الرمز :

$$\sum Si^2 = S_1^2 + S_2^2 + \ldots\ldots\ldots$$

حيث :

Si^2 : مجموع درجات جميع البنود S_1^2 : متغير البند من الأسئلة ١

S_t^2 : متغير البند جميع البنود

(٣) حساب متغير الإختبار باستخدام الرمز :

$$S_x^2 = \frac{\sum(X^2) - \left(\frac{\sum X}{N}\right)^2}{}$$

حيث :

Sx^2 : متغير الاختبار $(\sum n)^2$: مربعات مجموع درجات الطلاب

$\sum(n^2)$: مجموع مربعات درجات الطلاب N : عدد الطلاب المشتركين فى الإختبار

(٤) حساب معامل ثبات الإختبار باستخدام الرمز كرونباخ ألفا :

$$a = \frac{K}{}\left(1 - \frac{\sum S^2}{}\right)$$

[17] Burhan Nurgiyantoro, *Penilaian Dalam Pengajaran Bahasa Dan Sastra* (BPFE-Yogyakarta, 2001).

عند جيواندونو أن مستوى الصعوبة هي عبارة عن مدى صعوبة بنود الاختبار أو سهولتها.فإن الاختبار الجيد هو الذي توسط بين السهل و الصعب.[١٨] إذا كان السؤال صعب جدا حتى لا يستطيع أحد أن يجيبه،فأن هذا السؤال غير جيد لأن لا يستطيع أن يفرق بين الذكي و الضعيف أو بين المتوفق وغير متوفق.كذلك العكس إذا كان السؤال سهل جدا حتى يستطيع جميع الطلاب أن يجيبوه،فأن هذا السؤال غير جيد.[١٩] فهذا لا ينبغي أن يكون موجودا في الاختبار.لأن من أهم الاختبار أنه يعطى المعلومات عن قدرة الطلاب ويخبار من هؤلاء المتفوقون وغير المتوفقين.أن معامل الصعوبة يوضع لنا مدى الصعوبة أو السهولة التى يتمتع بها بند ما في الإختبار،وهو عبارة عن النسبة المؤية من الطلاب الذين أجابوا عن البند إجابة صحيحة.

إن الدرجة التى تسير الى مستوى الصعوبة هي ما بين ٠،٠٠. إلى ١،٠٠ .فدرجة ٠،٠٠. تدل على أن الاختبار صعب جدا لا يوجد أحد أن يستطيع إجابة السؤال.ودرجة ١،٠٠ تدل على أن الاختبار سهل جدا حتى جميع الطلاب يستطيعون الإجابة عن السؤال.

الرمز المستخدم لتحليل مستوى الصعوبة هو:[٢٠]

$$P = (JJB : JPT) \times 100$$

حيث :

P :مستوى الصعوبة لبند الاختبار

JJB : عدد الطلاب الذين أجابوا إجابة صحيحة

JPT :عدد الطلاب المشتركين فى الاختبار

الدرجة الجيد لمستوى الصعوبة في الإختبار ما بين ٠،٢٠ إلى ٠،٨٠.إذا كانت الدرجة أعلى من ٠،٨٠ فذلك السؤال في الحقيقة سهل جدا يحتاج إلى التبديل أو التصحيح.وكذلك لو كانت الدرجة أقل ٠،٢٠.فذلك السؤال في الحقيقة صعب جدا يحتاج إلى التبديل أو التصحيح.

إن الاختبار يستطيع أن يبرز الفروق بين الطلاب،فيبين لنا أن الأقوياء من الضعاف. ويطلب هذا ان يكون هناك مدى واسع بين السهل والصعب من الأسئلة بحيث يؤدي هذا إلى توزيع معتدل بين أعلى

[18] Djiwandono, *Tes Bahasa Dalam Pengajaran* (Bandung: ITB, 1996).

[19] Alan H Schoenfeld, 'When Good Teaching Leads to Bad Results: The Disasters of'well-Taught'mathematics Courses', *Educational Psychologist*, 23.2 (1988), 145–66.

[20] Suharsimi Arikunto, *Prosedur Penelitian : Suatu Pendekatan Praktik* (Jakarta: Rineka Cipta, 2010).

وأقل الدرجات.[٢١] يرتبط معمل التمييز الى درجة كبيرة بمعامل الصـــعوبة.فإذ كان الغرض من الاختبار هو أن يفرق بين القادرين من الطلاب وأولئك الأقل قدرة فإن البند المميز هو ما يقود الى هذا الغرض.[٢٢]

إذن، الاختبار الجيد هو الذى يستطيع أن يفرق بين الطلاب المتفوقين والطلاب الضعاف. ومعامل تمييز بند الإختبار فى درجة الرقم ٠،٠٠-١،٠٠. فالرقم ٠،٠٠يدل على أن بند الاختبار لا يستطيع أن يفرق بين الطالب الماهر وغير الماهر. والرقم ١،٠٠يدل على أن بند الاختبار يســـتطيع أن يفرق بين الطالب الماهر وغير الماهر بصفة كاملة.[٢٣]

ومن معانى الموجودة في الاختبار الجيد من حيث التمييز هو الذي يســـتطيع أن يفرق بين الطلاب الماهر م غير الماهر.إن المقصـــود من هذا الاختبار هو الفصـــل بين الطلاب الذين يتعلمون عن المواد المعينة والطلاب الذين لم يتعلموا هنها،فأســـئلة الاختبار لهذتين التمير المختلفتين حســـب قدرتهم وطبقاتهم. وعلى هذا،يمكن المعلم التمييز عن مكان الطالب ودرجته في الفصـــل وهو التفريق بين القدرة لدى الطالب الماهر والطالب الجاهل.

إذا معامل التمييز ينبغي أن يوضـح لنا مدى فعالية بند ما في التمييز بين الطلاب ذي القدرة اللعالية وللطالب ضـعيف المستوى بالقدر نفسـه الذي يفرق الاختبار بصـورة عامة بينهما في الدرجة الهائية.أما الرمز المستخدم لتحليل التمييز هو:[٢٤]

$$D = (T - R) : N$$

حيث :

D : معامل التمييز لبند الاختبار

R : عدد الإجابات الصحيحة فى المجموعة الدنيا

T : عدد الإجابات الصحيحة فى المجموعة العليا

N : عدد الطلاب فى المجموعة العليا أو الدنيا

منهج البحث

البحث عن"مستوى لياقة الاختبار لمادة اللغة العربية بالجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى هو بحث يقصد به وصف نوع الاختبار في الامتحان النهائي للمرحلة الأولى الذي تستخدمه

[21] John T Cacioppo, Richard E Petty, and Katherine J Morris, 'Effects of Need for Cognition on Message Evaluation, Recall, and Persuasion.', *Journal of Personality and Social Psychology*, 45.4 (1983), 805.

[22] Muhammad, Muhammadu 'Abdu al Khâliqi, *Ikhtibârâtu Al Lughati*, (Riyadh: Mathâbi'u Jâmi'ati al Mulki Su'ûd., 1996).

[23] Elan C Hope and others, 'Connecting Self-Esteem and Achievement: Diversity in Academic Identification and Dis-Identification Patterns among Black College Students', *American Educational Research Journal*, 50.5 (2013), 1122–51.

[24] Anne-Marie Brady, 'Assessment of Learning with Multiple-Choice Questions', *Nurse Education in Practice*, 5.4 (2005), 238–42.

بالجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى لمادة اللغة العربية من ناحية الصدق والثبات ومستوى الصعوبة ثم مستوى التمييز.بالإضافة إلى وصف كون الأساتذة في شعبة اللغة العربية بالجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى لا يفهمون تحليل نموذج الاختبار في الامتحان النهائي وأسبابها ولا يقدمونه كما هو مطلوب.

يستخدم الباحثون طريقة وصفية التحليلية وهي التى تقدم وصفا لظواهر وأحدث موضوع البحث ثم تسعى بتفسير الأحداث والظواهر وتحليلها والخروج بنظريات وقوانين التعييم والتنبؤ.[٢٥] أو هو تهدف الى وصف ظواهر او الأحداث أو الأشياء المعينة.وأما أهدف من البحث الوصفى هو تعيين الوصف عن الصور والرسم منظما وواقعيا ومناسبا بالعوامل والصفات وما يتعلق بالظواهر.[٢٦]

يقصد بمجتمع البحث هو جمع معلومات أو بيانات البحث.[٢٧] .أي كل فرد من أفراد المجتمع أم أي قضية أو أي موضوع الرموز لها رمزا ظاهرا وبينا.[٢٨] وأن المجتمع المقصود في هذا البحث هو جميع الطلاب في شعبة اللغة العربية بالجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى للمرحلة الأولى الذين اشتركوا للاختبار في الامتحان النهائي للمرحلة الأولى في كلية التربية لمادة اللغة العربية،عددهم ٢١٢طالب وهم في خمسة شعوب:في شعبة تعليم اللغة العربية عدد المشتركين٣٦ طالب،وفي شعبة التربية الدنية الإسلامية عدد المشتركين٣٧طالب، وفي شعبة التربي الإسلامي عدد المشتركين ٣٤ طالب، وفي شعبة اللغة الإنجيليزية عدد المشتركين٤٦طالب، وفي شعبة التعليم الرياضية عدد المشتركين ٤٩ طالب.بالنظر الى عدد الطلاب الكثير وبوجود الشعوب الموجودة فتأخذ الباحثة هذه الجامعة مجتمعا للبحث. وأما وسائل جمع البيانات لهذا البحث فهي المقابلة والوثائق

نتائج البحث ومناقشتها

إن أسئلة الاختبار في الإمتحان النهائي للمرحلة الأولى التى تستخدمه الجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى في شعبة اللغة العربية لمادة اللغة العربية تتكون من ٦٠ سؤالا.وبناء على النظريان السابق عن نوع الاختبار،يمكن تحليل نوع الاختبار من خمس نواحي وهي غرض الاختبار ومنتج الاختبار وساعة الاختبار وطرق عمل الاختبار ثم نوع التقدير

أما مواصفات الاختبار في الامتحان النهائى للمرحلة كما يلي: (أ) صدق الاختبار.من مواصفات الاختبار الجيد الصدق،يقصد بذلك أن الاختبار يقيس الشيء الذي وضع من أجله. وفي هذا البحث، الباحثة تبحث في صدق المحتوى للاختبار النهائي للمرحلة الأولى الذى تستخدمه الجامعة الإسلامية

[25] ʻAbdu al Rahmâni Ahmad Utsmân, *Manâhiju Al Bahtsi Al ʻIlmi Wa Thuruqu Kitâbati Al Rasâ'Ili Al Jâmi'ati* (Khartûm: Dâru Jâmi'ati Ifrîqiyâ al ʻAlamiyyati Linnasyari, 1995).

[26] Zainal Arifin, *Evaluasi Instruksional Prinsip Teknik Prosedur* (Bandung: Remaja Rosdakarya, 1991).

[27] Arikunto, *Prosedur Penelitian : Suatu Pendekatan Praktik*.

[28] Furchan, *Pengantar Penelitian Dalam Pendidikan* (Surabaya: Usaha Nasional, 1982).

الحكومية سلطان طه سيف الدين جامبى في شعبة اللغة العربية لمادة اللغة العربية. (ب) ثبات الاختبار. لمعرفة ثبات الاختبار فى الإمتحان النهائي للمرحلة الأولى الذى تستخدمه الجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى في شعبة اللغة العربية لمادة اللغة العربية، لأن هذا الاختبار هو الاختبار الذاتى فتستخدم الباحثة الرمز كرنباخ ألفا (Cronbach Alpha). (ج) مستوى صعوبة الاختبار. إن السؤال الجيد هو السؤال المتوسط بين السهل والصعب.إذا كان السؤال سهلا جدا حتى يستطيع جميع الطلاب أن يجيبوه فإن هذا السؤال غير جيد لأنه لا يستطيع أن يعطى المعلومات عن حقيقة قدرة الطلاب.[٢٩] وبالعكس،إذا كان السؤال صعبا جدا حتى لا يستطيع جميع الطلاب أن يجيبوه فإنه هذا السؤال غير جيد لأنه لا يستطيع أن يعطى المعلومات الصحيحة عن قدرة الطلاب.ولذلك كأنهم ضعفاء. ومن درجات مستوى الصعوبة (المذكورة في الباب الثالث)،تستطيع الباحثة أن ترى البنود الجيدة ة البنود غير الجيدة.إذا كان للبند مستوى الصعوبة ٠،٠٠-٠،٢٠ (صعب جدا) أو ٠،٨١-١،٠٠ (سهل جدا) فهذا البند لا ينبغي إبقائه و يحتاج إلى التصحيح.أما إذا كان للبند مستوى الصعوبة ٠،٢٠-٠،٨٠ (متوسط) فهذا البند ينبغي إبقائه دون التغيير من ناحية التركيب. (د) مستوى تمييز الاختبار. ويقصد بالتمييز أن بنود الاختبار تستطيع أن تبرز الفروق بين الطلاب المتفوقين والطلاب غير المتفوقين.البنود الجيدة من حيث التمييز هي التي يجيبها المتفوقون في حين لا يستطيع غير المتفوقين الاجابة عنها، والبنود غير جيدة هي التي لا يستطيع المتفوقون الإجابة عنها في حين يستطيع غير المتفوقين الإجابة عنها أو البنود لا تستطيع أن تفرق بين الطلاب المتفوقين وغير المتفوقين.

ومن درجات معامل التمييز المذكورة في الباب الثالث تستطيع الباحثة أن ترى إذا كان للبند معامل التمييز ٠،٥٠-١،٠٠ فهذا البند ينبغي ابقائه دون التغيير من ناحية التركيب لأنه البند يستطيع أن يفرق جيدا بين الذكي والضعيف.وكذلك البند الذي له معامل تمييز ٠،٢٠-٠،٤٩ فلابأس ببقائه لأن يستطيع أن يفرق بين الذكي والضعيف.أما البند الذي له معامل التمييز ٠،١٠-٠،١٩ أو ٠ فلا ينبغي إبقائه لأنه لا يؤثر كثيرا عند التفريق بين الذكي والضعيف أو بين المتفوقين وغير المتفوقين ولا بد منه التصحيح.

تحليل الاختبار هو أمر مهم ومحتاج بعمله، لأن بتحليل الاختبار سوف تنال المعلومات عن جودة الاختبار حتى تعرف هل جودة ذلك الاختبار جيد أم غير جيد. ولكن من المدرسين وعامل الاختبار نادر منهم عمل فى تحليل بنود الاختبار.

للاختبار فى الإمتحان النهائي للمرحلة الأولى لمادة اللغة العربية بالجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى.،من البيانات التى حصلتها الباحثة، تعرف أن هناك أسباب عدم قيام كون الأساتذة في شعبة اللغة العربية لمادة اللغة العربية الجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى بتحليل

[29] Muhammad - ‘Abd al ‘Azîzi, *Mafâhîmu Al Taqwîmu Wa Ususuhu Wa Wazhâ’ifuhu* (Riyad: Maktabatu al Tarbiyyati al ‘Arabiyyati li Duwali al Khalîji, 1983).

نموذج الاختبار. وهى فيما يلى: أ)- جميع الأساتذة في شعبة اللغة العربية لمادة اللغة العربية بالجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى. ب)- عدم إستيعاب عن تحليل نموذج الاختبار. ج)- عدم الخبر والمعلومات والمفاهيم عن تحليل نموذج الاختبار. د)- نقصان معرفة عن التقييم و تحليل نموذج الاختبار. هـ)- عدم الفهم الدقيق عن كيفية تحليل نموذج الاختبار. و)- لم يشتركوا في التدريب أو التمرين عن التقييم و تحليل نموذج الاختبار.

مناقشة نتائج البحث

إن نوعية الاختبار فى الامتحان النهائي للمرحلة الأولى الذى تستخدمه في شعبة اللغة العربية لمادة اللغة العربية الجامعة الإسلامية بالحكومية سلطان طه سيف الدين جامبى يعتبر من الاختبار التحصيل (من ناحية غرض الاختبار) والاختبار المعلم (من ناحية منتج الاختبار) والاختبار الإجمالى (من ناحية ساعة الاختبار) والاختبار التحريرى أو الاختبار الكتابى (من ناحية طرق عمل الاختبار) والاختبار الذاتى أو المقال (من ناحية نوع التقدير).

إن هذه البنود لم تتوافر الأهداف المثبتة كما المذكور في المنهج التعليمي.هناك لا توجد البنود التي تشملها مهارة الإستماع والكلام.كما لا تلاحظ الباحثة أن هذه الاحتبار تعتبر من الاختبارات الكتابية،لأن فيه يقدم الطلاب الاجابات مكتوبة على أوراق الاجابة.ولا يستخدم فيه الأسئلة بطريقة الإستماع والكلام.لذا، من أجل تكميل هذا الاختبار، رأى الباحثون أن الاختبار لمادة اللغة وخاصة اللغة العربية يؤدي تحريريا وشفهيا. الاختبار التحريري يستخدم لقياس مهارة الطلاب في الكتابة والقراءة. والاختبار الشفهي يستخدم لقياس مهارة الطلاب في الإستماع والكلام.و كما عرف الباحثة من المنهج في تكوين أسئلة الاختبار النهائي أن نسبة الاسئلة للاختبار النهائي هي القواعد هي٢٥ بندا أو ٤١،٦٧% ونسبة بحث المفردات ١٥ بندا أو ٢٥% ومهارة القراءة ٢٠ بندا ٣٣،٣٣%.

تبين الاختبار النهائي للمرحلة الاولى له ٣٤ بندا (٥٧،٠%)تصل إلى قدر سهل جدا والبنود الأخرى ٢٦بندا (٤٣،٠%) تصل إلى قدر جيد وأما تنفى بنود تصل إلى قدر صعب.بمعنى ذلك أن البنود التى قدرها صعب إن معظم الطلاب لا يستطيعون الاجابة الصحيحة عن السؤال لذلك فتحتاج هذه إلى التصحيح. وكذالك البنود التى قدرها سهل تحتاج إلى التصحيح لأنه سهل أمام الطلاب بمعنى أن معظم الطلاب يستطيعون الاجابة الصحيحة عن السؤال فلا يستطيع أن تفرق بين الطلاب الماهر والطلاب غير الماهر.وأما البنود الجيد مناسبة لبقائها لأنها ليست صعبا ولا سهلا وتستطيع أن تفرق جيدا بين من يعلم ومن لم يعلم.إذن أن مستوى صعوبة الاختبار النهائي تقديره سهل جدا، فبنود السؤال غير جيد تقديره ٣٤ بندا أو٥٧،٠ % بمعنى ذلك أن البنود التى قدرها سهل جدا إن معظم الطلاب يستطيعون الاجابة الصحيحة عن السؤال لذلك فتحتاج هذه إلى التصحيح.

بعد تبين الباحثون مستوى التمييز لهذا الاختبار النهائي،إن البنود التى تستطيع أن تفرق بين الطلاب القادرين وغير القادرين على إجابة الأسئلة المعروضة (٢٩ بنود أو ٠،٤٨ %) من جميع الأسئلة. وهذه البنود مناسبة لبقائها أي لابأس بقائها. وأما البنود الباقية (٣١ بنود أو ٠،٥٢ %) تحتاج إلى التصحيح أو التبديل لأنه لا تفرق جيدا بين الطلاب الماهر والطلاب غير الماهر.إذن أن مستوى تمييز الاختبار النهائي تقديره الناقص لأن بنود السؤال غير جيد تقديره ٣١ بندا أو ٠،٥٢% .

عدم إستعاب كون الأساتذة في شعبة اللغة العربية لمادة اللغة العربية الجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى فى تحليل نموذج الاختبار وأسبابها هى فيمايلى :١) جميع الأساتذة في شعبة اللغة العربية لمادة اللغة العربية الجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى خريج من قسم تربية اللغة العربية. ٢) عدم إستيعاب عن تحليل نموذج الاختبار. ٣) عدم الخبر والمعلومات والمفاهيم عن تحليل نموذج الاختبار. ٤) نقصان معرفة عن التقييم وتحليل نموذج الاختبار. ٥) عدم الفهم الدقيق عن كيفية تحليل نموذج الاختبار. ٦) لم يشتركوا في التدريب أو التمرين عن التقييم وتحليل نموذج الاختبار.

الخلاصة

نوعية الاختبار فى الامتحان النهائي للمرحلة الأولى الذى تستخدمه الجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى في شعبة اللغة العربية لمادة اللغة العربية هى :أ) صدق محتوى الاختبار جيد، لأنه قد ملأ شروط صدق المحتوى، وأما بنود الاختبار النهائي كثير تتكون من بحث القواعد هي ٤١،٦٧% ونسبة بحث المفردات ٢٥% ومهارة القراءة ٣٣،٣٣%. ب) أن مستوى ثبات الاختبار تقديره عال وهو ٠،٨٦٩٣٦٦١٢٧ أو ٠،٨٦٩٤. ج) أن مستوى صعوبة الاختبار النهائي تقديره سهل جدا، فبنود السؤال غير جيد تقديره ٣٤ بندا أو 0,57% .من ناحية مستوى الصعوبة للاختبار: ١) ٢٦ بندا تقع فى مستوى متوسط أو ٠،٤٣% ، فهذه البنود جيدة وينبغى إبقائها لأنها متوسطة بين السهل والصعب. ٢) ٣٤ بندا تقع فى مستوى سهل جدا 0,57 %. ٣) وأما فى مستوى صعب جدا. إذن، البنود التى لابأس بإبقائها من حيث مستوى الصعوبة ٢٦ بندا أو٠،٤٣% من جميع الأسئلة.وعدد البنود التى تحتاج إلى التصحيح من حيث مستوى الصعوبة ٣٤ بنود (٣٤ بنود من مستوى سهل جدا، و. بنود من مستوى صعب جدا) أو ٥٧ % من اسئلة الاختبار غير جيدة ولا تصلح إبقائها للاختبار فى الإمتحان النهائي بالجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى في شعبة اللغة العربية لمادة اللغة العربية هو القادم.

إن مستوى تمييز الاختبار النهائي تقديره الناقص لأن فبنود السؤال غير جيد تقديره ٣١ بندا أو ٠،٥٢%. من ناحية تمييز للاختبار : ١)- ٥ بندا تقع فى مستوى جيد أو ٠،٠٨% فهذه البنود جيدة وينبغى إبقائها. ٢)- ٢٤ بندا تقع فى مستوى متوسط أو ٠،٠٤% فهذه البنود ينبغى إبقائها أيضا، لأنها تستطيع أن تفرق بين الطلاب الماهرون والطلاب غير الماهرون. ٣)- ٢٥ بندا تقع فى مستوى ناقص أو ٠،٤٢% من اسئلة الاختبار،

وهى تحتاج إلى التصحيح من حيث التمييز ولا تصلح ابقائها للاختبارات القادمة كما هو مطلوب. ٤)- ٤ بندا تقع فى مستوى عدم التمييز أو٠،٠٧% من اسئلة الاختبار عدم معامل التمييز وهى تحتاج إلى التبديل. ٥)- ٢ بندا تقع فى مستوى السلبى أو ٠،٠٣ اسئلة الاختبار سلبيا وهى توجب إلى التنفى. إذن، البنود التى لابأس بإبقائها للاختبار فى الإمتحان النهائي للمرحلة الأولى الذى تستخدمه الجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى في شعبة اللغة العربية لمادة اللغة العربية هو من حيث التمييز ٢٩ بندا أو ٠،٤٨% (٥ بنود تقع فى مستوى جيد،و ٢٤ بندا تقع فى مستوى متوسط) من جميع الأسئلة. إذن تمييز الاختبار هو ناقص.

كون الأساتذة في شعبة اللغة العربية لمادة اللغة العربية بالجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى هو لايفهمون تحليل نموذج الاختبار في الامتحان النهائي ولا يقدمونه كما هو مطلوب وهذا للأسباب هى فيمايلى: ١)- عدم إستيعاب عن تحليل نموذج الاختبار. ٢)- عدم الخبر والمعلومات والمفاهيم عن تحليل نموذج الاختبار. ٣)- نقصان معرفة عن التقييم و تحليل نموذج الاختبار. ٤)- عدم الفهم الدقيق عن كيفية تحليل نموذج الاختبار. ٥)- لم يشتركوا في التدريب أو التمرين عن التقييم و تحليل نموذج الاختبار.

ومن نتائج البحث المذكورة أن مستوى لياقة الاختبار فى الامتحان النهائي للمرحلة الأولى الذى تستخدمه بالجامعة الإسلامية الحكومية سلطان طه سيف الدين جامبى في شعبة اللغة العربية لمادة اللغة العربية ليست لها نوعية الاختبار من جهة التقييم الاربع أو غير المناسب بمواصفات الاختبارات الجيدة من ناحية صدق والثبات ومستوى الصعوبة والتمييز.

References / المراجع

'Abd al 'Azîzi, Muhammad, *Mafâhîmu Al Taqwîmu Wa Ususuhu Wa Wazhâ'ifuhu* (Riyad: Maktabatu al Tarbiyyati al 'Arabiyyati li Duwali al Khalîji, 1983)

Arifin, Zainal, *Evaluasi Instruksional Prinsip Teknik Prosedur* (Bandung: Remaja Rosdakarya, 1991)

Arikunto, Suharsimi, *Dasar-Dasar Evaluasi Pendidikan* (Jakarta: Bumi Aksara, 2021)

———, *Prosedur Penelitian : Suatu Pendekatan Praktik* (Jakarta: Rineka Cipta, 2010)

Brady, Anne-Marie, 'Assessment of Learning with Multiple-Choice Questions', *Nurse Education in Practice*, 5.4 (2005), 238–42

Brown, Douglas, *Ususu Ta'lîmi Al Lughati Wa Ta'lîmuhâ (Tarjamahu 'Abdu Al Râjihi Wa 'Alî Ahmad Sya'Bânu)* (Bairut: Dâru al Nahdhati al 'Arabiyyati, 1994)

Cacioppo, John T, Richard E Petty, and Katherine J Morris, 'Effects of Need for Cognition on Message Evaluation, Recall, and Persuasion.', *Journal of Personality and Social Psychology*, 45.4 (1983), 805

Djiwandono, *Tes Bahasa Dalam Pengajaran* (Bandung: ITB, 1996)

Fiorella, Logan, and Richard E Mayer, 'The Relative Benefits of Learning by Teaching and Teaching Expectancy', *Contemporary Educational Psychology*, 38.4 (2013), 281–88

Fraenkel, Jack R, Norman E Wallen, and Helen H Hyun, *How to Design and Evaluate Research in Education* (McGraw-hill New York, 2012), VII

Furchan, *Pengantar Penelitian Dalam Pendidikan* (Surabaya: Usaha Nasional, 1982)

Govindarajan, Vijay, and Praveen K Kopalle, 'Disruptiveness of Innovations: Measurement and an Assessment of Reliability and Validity', *Strategic Management Journal*, 27.2 (2006), 189–99

Hope, Elan C, Tabbye M Chavous, Robert J Jagers, and Robert M Sellers, 'Connecting Self-Esteem and Achievement: Diversity in Academic Identification and Dis-Identification Patterns among Black College Students', *American Educational Research Journal*, 50.5 (2013), 1122–51

Al Khûlî, Muhammad 'Alî, *Asâlîbu Tadrîsi Al Lughati Al 'Arabiyati* (Riyadh: al Mamlakatu al 'Arabiyyatu al Su'ûdiyyah, 1986)

Lam, Ricky, 'Formative Use of Summative Tests: Using Test Preparation to Promote Performance and Self-Regulation', *The Asia-Pacific Education Researcher*, 22.1 (2013), 69–78

McDonald, Roderick P, *Test Theory: A Unified Treatment* (psychology press, 2013)

Muhammad, Muhammadu 'Abdu al Khâliqi, *Ikhtibârâtu Al Lughati,* (Riyadh: Mathâbi'u Jâmi'ati al Mulki Su'ûd., 1996)

Nurgiyantoro, Burhan, *Penilaian Dalam Pengajaran Bahasa Dan Sastra* (BPFE-Yogyakarta, 2001)

Nurkancana, Wayan, and P P N Sumartana, 'Evaluasi Pendidikan', *Surabaya: Usaha Nasional*, 1986

Prihartini, Yogia, Wahyudi Buska, and Nur Hasnah, 'Analysis of Test Item on The Final Test Semester Exam on Arabic Subjects', *Arabiyat: Jurnal Pendidikan Bahasa Arab Dan Kebahasaaraban*, 6.1 (2019), 71–92

Purwanto, M Ngalim, 'Prinsip-Prinsip Dan Teknik Evaluasi Pengajaran', 2019

Rebele, James E, and E Kent St Pierre, 'A Commentary on Learning Objectives for Accounting Education Programs: The Importance of Soft Skills and Technical Knowledge', *Journal of Accounting Education*, 48 (2019), 71–79

Schoenfeld, Alan H, 'When Good Teaching Leads to Bad Results: The Disasters of'well-Taught'mathematics Courses', *Educational Psychologist*, 23.2 (1988), 145–66

Schuman, Howard, 'The Random Probe: A Technique for Evaluating the Validity of Closed Questions', *American Sociological Review*, 1966, 218–22

Thu'aimah, Rusydi Ahmad, 'Ta'lim Al-Lugah Li Ghairi Nathiqina Biha: Mana> Hijuhu Wa Asa> Libuhu', *Ar-Ribat: UCESCO*, 1989

Thuwailah, 'Abdu al Wahhâbi 'Abdu al Salâmi, *Al Tarbiyyatu Al Islâmiyyatu Wa Fannu Al Tadrîsi,* (Riyadh: Dâru al Salâmi, 1992)

Utsmân, 'Abdu al Rahmâni Ahmad, *Manâhiju Al Bahtsi Al 'Ilmi Wa Thuruqu Kitâbati Al Rasâ'Ili Al Jâmi'ati* (Khartûm: Dâru Jâmi'ati Ifrîqiyâ al 'Alamiyyati Linnasyari, 1995)

Wiggins, Grant, 'A True Test: Toward More Authentic and Equitable Assessment', *Phi Delta Kappan*, 92.7 (2011), 81–93