الباب الثاني

الإطار النظري

أ. نظرة عن التقييم

أ. تعريف التقييم

التقييم هو أنشطة لتحليل الأسئلة في التعليم والتعلم ولمعرفة وتحديد او اتخاذ القرارات الى أي مدة تم تحقيق أهداف التدريس من قبل المتعلمين. سيتم تنفيذ هذه العملية بشكل مستمر بحيث إذا كان هناك خطأ يمكن مراجعتها. كرونباج (Cronbach) في الكتابه (Designing Evaluator of Educational and Social Program) الذي "التقييم هو عملية مستمرة بحيث يمكن في عملية أنشطتها مراجعة ما إذا كان هناك خطأ".[٢٧] رأي مهرين و ليهمان (Mehrens, Lehman) الذي "التقييم هو عملية تخطيط والحصول على وتوفير المعلومات التي لا غنى عنها لاتخاذ قرارات بديلة".[٢٨]

وكذلك فإن نشاط التقييم هو يتم التخطيط لها عمدا للحصول على معلومات أو بيانات، بناء على البيانات ثم محاولة اتخاذ قرار.

ب.أهداف ومهمّة التقييم

يتم استخدام أهداف التقييم التي تم تحديدها كمعيار لتحديد مدى تحقيق الهدف في البرنامج. المثال، ما يصل إلى ٢٠ طالبا مدرسا مستهدفا، هناك ١٥ طالبا يجيبون بقيمة ١٠، ولكن بعد التحقق من نتائج الاختبار من يجيب بقيمة ١٠ أو يصحح جميع الطلاب العشرة فقط، وبالتالي يمكن القول أن معدل نجاح المعلم يبلغ حوالي ٥٠٪. لذلك، لا يمكن إجراء

[27] H. Daryanto, Evaluasi Pendidikan. . . ,h. 30
[28] M. Ngalim Purwanto, Prinsip-Prinsip dan Teknik Evaluasi Pengajaran, (Bandung: PT Remaja Rosdakarya، 2008), h. 3.

التقييم بلا مبالاة. ومع ذلك، يجب أن يتم ذلك بشكل مستمر واستخدام أدوات قياس دقيقة. الهدف الرئيسي من إجراء تقييم في عملية التعليم والتعلم هو الحصول على معلومات دقيقة حول درجة تحقيق الأهداف التعليمية من قبل الطلاب بحيث يمكن متابعة"[٢٩]

تنقسم أهداف التقييم قسمين، التقييم عام والتقييم خاص :

أ. تقييم عام

- لجمع المواد الإعلامية التي سيتم استخدامها كدليل على درجة التطور أو درجة التقدم الذي يمر به الطلاب، بعد متابعتهم لعملية التعلم لفترة زمنية معينة.
- لتحديد درجة درجة طرق التدريس التي تم استخدامها في عملية التعلم لفترة زمنية معينة.

ب. تقييم خاص

- لتحفيز أنشطة الطلاب في أخذ البرامج التعليمية. بدون التقييم، لذلك من المستحيل إثارة الإثارة أو التحفيز لدى الطلاب لتحسين وتحسين إنجازاتهم.
- لإيجاد العوامل التي تسبب نجاح وفشل الطلاب في المشاركة في البرامج التعليمية، بحيث يمكن البحث عن حلول أو طرق للتحسين.[٣٠]

وظيفة التقييم في التعلم هي كإجراء يستخدم للتحسين من الغرض من التقييم كتطور وتقدم للطلاب. تنقسمها إلى أربع مجموعات:

[29] H. Daryanto, Evaluasi Pendidikan. . . ، h. ١١.
[30] Anas Sudijono, Pengantar Evaluasi Pendidikan. . . ، h. ١٧

- لمعرفة مدى تقدم وتطور ونجاح الطلاب بعد القيام بأنشطة التعلم لفترة زمنية معينة.
- لمعرفة نسبة نجاح البرنامج التدريسي.
- لأغراض التوجيه والإرشاد.
- لأغراض تطوير وتحسين المناهج الدراسية للمدرسة المعنية.[٣١]

لمّا وظيفة التقييم من مصالح كل طرف ، يمكن استنتاجها على النحو التالي:

أ. وظيفة التقييم للمعلمين

- لمعرفة تقدم الطلاب
- لمعرفة موقف كل الطلاب في المجموعته
- لمعرفة نقاط الضعف في التعليم
- لتحسين التعليم
- لتحديد تخرج الطلاب

ب. وظيفة التقييم للطلاب

- لمعرفة تقدم الطلاب
- لتحسين التعليم
- لتعزيز الدافع في التعليم

ج. وظيفة التقييم للمدرسة

- لقياس درجة التعليم
- لمعرفة تقدم وانتكاسات المدرسة
- لاتخاذ القرارات للطلاب
- لتحسين المناهج الدراسية

[31] M. Ngalim Purwanto, Prinsip-Prinsip dan Teknik Evaluasi Pengajaran. . . , h. 7

د. وظيفة التقييم للوالدين

- لمعرفة تقدم ابنها
- لزيادة الإشراف والتوجيه والمساعدة لإبنها في جهود التعلم
- لتوجيه اختيار التخصصات ، أو أنواع التعليم الإضافي لإبنها

هـ. وظيفة التقييم للمجتمع

- لمعرفة تقدم المدرسة
- للمشاركة في نقد واقتراح تحسينات على المناهج التعليمية في المدرسة
- لزيادة مشاركة المجتمع المحلي في الجهود الرامية إلى مساعدة المؤسسات التعليمية[٣٢]

ج. مبادئ التقييم

مبادئ التقييم التي يجب مراعاتها عند تجميع الاختبارات المستخدمة لقياس أهداف التعلم، لأنه إذا تم تجاهل مبادئ التقييم، فلا يمكن قياس نتائج الاختبار بدقة، وبعبارة أخرى، ستعمل أداة القياس بشكل صحيح إذا تم إجراء المعايرة أولا على الأداة قبل الاستخدام.

المبادئ الأساسية اللازمة لتجميع الاختبار، وهي:

أ. يجب أن يكون الاختبار قادرا على قياس نتائج التعليم بوضوح التي تم إنشاؤها وفقا للأغراض التعليمية.

ب. قياس عينة تمثيلية من نتائج التعلم والمواد التعليمية التي تم تدريسها

[32] M. Chabib Thoha, Teknik Evaluasi Pendidikan, (Jakarta: PT Raja Grafindo Persada، 2003), h. 1-11.

ج. يتضمن أشكالا مختلفة من الأسئلة المناسبة حقا لقياس النتائج المرجوة وفقا للأهداف.

د. مصممة وفقا لفائدتها للحصول على النتائج المرجوة.

هـ. تستخدم لتحسين طريقة تعلم الطلاب وكيفية تدريس المعلمين[٣٣]

التقييم يتم بشكل جيّد عند الالتزام بثلاثة مبادئ أساسية، وهي :

أ. المبدأ العام، يمكن القول أن المبدأ العام أو المبدأ الشامل هو أن تقييم نتائج التعلم يمكن أن يتم بشكل جيّد إذا تم إجراء التقييم بالإجماع وسليمة وشاملة دون أن يتم تنفيذها بشكل منفصل.

ب. المبدأ الاتتمرارية، يمكن القول أن تقييم نتائج التعليم الجيّدة هو تقييم يتم إجراؤه بانتظام وبشكل مستمر من وقت لآخر.

ج. المبدأ الموضوعية، بمعنى أن تقييم نتائج التعلم يمكن التعبير عنه كتقييم جيّد إذا كان يمكن أن يكون أوسع العوامل الذاتية.[٣٤]

د. خطوات التقييم

بشكل عام ، تشمل الخطوات الرئيسية للتقييم التربوي ثلاثة أنشطة رئيسية:

أ. التحضير

ب. التنفيذ

ج. التحجيز التنائج[٣٥]

[33] M. Ngalim Purwanto, Prinsip-Prinsip dan Teknik Evaluasi Pengajaran. . . ، h. 23-25
[34] Anas Sudijono, Pengantar Evaluasi Pendidikan. . . ,h. 31-33
[35] M. Chabib Thoha, Teknik Evaluasi Pendidikan. . . , h. 18

على الرغم من أنها ليست متشابهة للغاية ، بشكل عام ، يقسم الخبراء في مجال التقييم التربوي أنشطة تقييم نتائج التعلم إلى ست خطوات رئيسية ، بما في ذلك:

- وضع خطة تقييم نتائج التعليم، قبل إجراء تقييم نتائج التعليم، يجب إعداد تخطيط جيّد ودقيق.
- جمع البيانات، جمع البيانات هو إجراء قياسات ، على سبيل المثال عن طريق إجراء اختبارات نتائج التعلم أو المقابلات أو الاستبيانات باستخدام الأدوات.
- التحقق من البيانات، يجب تصفية البيانات التي تم جمعها بنجاح أولا قبل إجراء مزيد من المعالجة. تعرف عملية الفحص باسم البحث عن البيانات أو التحقق من البيانات.
- معالجة وتحليل البيانات، تتم معالجة وتحليل نتائج التقييم بهدف إعطاء معنى للبيانات التي تم جمعها بنجاح في أنشطة التقييم.
- تقديم التفسير واستخلاص النتائج، تفسير بيانات تقييم التعلم هو في الأساس تعبير لفظي للمعنى الوارد في البيانات التي خضعت للمعالجة والتحليل. على أساس تفسير بيانات التقييم، يمكن تقديم بعض الاستنتاجات في النهاية
- متابعة التقييم، بدءا من بيانات التقييم التي تم تجميعها وترتيبها ومعالجتها وتحليلها واستنتاجها بحيث يمكن معرفة المعنى الوارد فيها، في النهاية سيكون المقيم قادرا على اتخاذ القرارات أو صياغة السياسات التي تعتبر ضرورية كمتابعة لنشاط التقييم. يجب أن نتذكر دائما أن كل نشاط تقييم يتطلب متابعة ملموسة.[٣٦]

[36] Anas Sudijono, Pengantar Evaluasi Pendidikan. . . , h. 59-62

ب.نظرة عن الاختبار

أ. تعريف الاختبار

الاختبارات هو مجموعة من الأسئلة التي يجب الإجابة عليها، أو التي يجب الرد عليها، أو المهام التي يجب أن يقوم بها الشخص الذي يتم اختباره. تستخدم الاختبارات لقياس مدى إتقان المتعلم للدروس المقدمة، والتي تغطي بشكل أساسي جوانب المعرفة والمهارات.[٣٧]

الاختبارات نتائج التعلم هي اختبارات تستخدم لتقييم نتائج الدروس التي قدمها المعلمون لطلابهم، أو من قبل المحاضرين للطلاب، خلال فترة زمنية معينة.[٣٨]

ب.وظيفة الاختبار

بشكل عام، للاختبار وظيفتان، وهما:

أ. كجهاز قياس للطلاب. في هذه الحالة، يعمل الاختبار على قياس درجة التطور أو التقدم الذي حققه الطلاب بعد أن أخذوا عملية التعلم خلال فترة زمنية معينة.

ب. كمقياس لنجاح برنامج التدريس، لأنه من خلال الاختبار يمكن معرفة مدى تحقيق أهداف التعليم.[٣٩]

بناء على الشرح أعلاه، يمكن ملاحظة أن وظيفة الاختبار هي كأداة لقياس درجة التطور الذي يحققه الطلاب بعد أخذ عملية التعليم والتعلم.[٤٠]

ج. أنواع الاختبار

أ. اختبار الكتابة

---

[37] Sudaryono, Dasar-Dasar Evaluasi Pembelajaran. . . ,h. 101

[38] M. Ngalim Purwanto, Prinsip-Prinsip dan Teknik Evaluasi Program. . . ,h. 33.

[39] Joko Prasetyo, Evaluasi dan Remediasi Belajar. . . , h. 84

[40] Miftahul Jannah, "*Analisis Butir Soal UAS Fisika Pada Ujian Semester Genap Kelas XI Tahun Ajaran 2017/ 2018 di SMAN 16 Banda Aceh*", Banda Aceh, 2019 h. 33

اختبار يطرح أسئلة مع توقع إجابات مكتوبة. عادة ما يستخدم هذا الاختبار لقياس الجوانب المعرفية للطلاب.

ب. اختبار شفوي

اختبار يطرح أسئلة مع إجابات شفهية. يتم إجراء هذا الاختبار أيضا للجوانب المعرفية للطلاب.

ج. اختبار الأفعال

اختبار يطرح أسئلة تتطلب إجابات في شكل أفعال. يستخدم هذا الاختبار لتقييم الجوانب/ المهارات الحركية النفسية للمتعلمين.

وفقا لوظيفتها، ينقسم الاختبار إلى:

أ. الاختبارات التكوينية

اختبار يتم إجراؤه بعد الانتهاء من موضوع واحد. يعمل هذا الاختبار على تحديد ما إذا كان موضوع واحد مكتملا أم لا. المتابعة التي يمكن القيام بها بعد معرفة نتائج الاختبارات التكوينية للطلاب هي:

- إذا تم إتقان المادة التي تم اختبارها ، فسيستمر التعلم بموضوع جديد.
- إذا كانت هناك أجزاء لم يتقنها الطلاب، فقبل متابعة المادة الجديدة، كرر أولا أو اشرح مرة أخرى الأجزاء التي لم يتم إتقانها، بهدف تحسين درجة إتقان الطلاب.

ب. الاختبار الختامي

يتم إجراء الاختبارات بعد الانتهاء من مجموعة من وحدات برنامج التعليم. في المدرسة يعرف هذا الاختبار باسم الاختبار العام.

ج. الاختبارات التشخيصية

يتم إجراء الاختبارات لتحديد نوع الصعوبة التي يواجهها المتعلمون في موضوع معين بدقة.

وفقا لوقتها ، ينقسم الاختبار إلى:

أ. الاختبار المسبق

يتم إعطاء الاختبارات قبل عملية التعلم. يهدف هذا الاختبار إلى معرفة مدى إتقان الطلاب للمادة التي سيتم تدريسها.

ب.الاختبار النهائي

يتم إجراء الاختبار بعد تنفيذ عملية التعليم. يهدف الاختبار إلى تحديد درجة التقدم الفكري (درجة إتقان المادة) للطلاب. عادة ما يحتوي هذا الاختبار على نفس أسئلة الاختبار المسبق.

هناك عدة أنواع من الاختبارات التي يمكن استخدامها في إجراء التقييمات، أحدها اختبار كتابي يستخدم غالبا في المدارس. مثل امتحانات الفصل الدراسي والامتحانات الوطنية. من أجل الحصول على القيمة التي تريد تحقيقها، من الضروري عمل مجموعة اختبار جيّدة.[٤١]

د. خطوات التحضير للاختبار

سيقول أي معلم بسهولة أي جزء من الدرس سيتم تغطيته في الاختبار إذا كان الغرض معروفا. تسلسل الخطوات التي يتم تنفيذها هو كما يلي:

- تحديد الغرض من إجراء الاختبار
- فرض قيود على المواد المراد اختبارها
- صياغة أهداف تعليمية محددة لكل جزء من المادة.
- صف جميع الدرجةات في جدول الإعداد الذي يحتوي أيضا على جوانب السلوك الواردة في الدرجةات

[41] Miftahul Jannah, "*Analisis Butir Soal UAS Fisika Pada Ujian Semester Genap Kelas XI Tahun Ajaran 2017/ 2018 di SMAN 16 Banda Aceh*"، Banda Aceh 2019 h. 36-37.

- قم بتجميع جدول مواصفات يحتوي على الموضوع وجوانب التفكير التي يتم قياسها جنبا إلى جنب مع التوازن بين الأمرين.
- اكتب بنود السؤال ، بناء على الدرجةات التي تمت كتابتها في جدول الدرجةات وجوانب السلوك التي تمت تغطيتها[٤٢]

يجب إجراء كل خطوة من خطوات التحضير للاختبار بانتظام، من تحديد الغرض من إجراء الاختبار إلى إعداد بنود الأسئلة. إذا تم تنفيذ خطوات التحضير مباشرة لجعل بنود السؤال، فقد تكون الأسئلة المقدمة ليست جيّدة تماما من حيث التوافق مع درجةات المواد ودرجة صلاحية السؤال أم لا.

ت. خصائص الاختبار الجيّد

أ. صحة

يقال إن الاختبار يكون صالحا إذا كان الاختبار دقيقا أو صحيحا صالحا أو قادرا على التعبير أو قياس ما يجب الكشف عنه أو قياسه من خلال الاختبار.[٤٣]

ب. موثوقيه

الموثوقية أو الاستقرار هي قياس يتم إجراؤه بشكل متكرر لمعرفة مدى الاختلاف في التغييرات في النتائج التي تم الحصول عليها. إذا كانت النتائج التي تم الحصول عليها هي نفسها دائما أو قريبة، فيمكن القول أن جهاز القياس في شكل الاختبار يتمتع بالفعل بموثوقية عالية. يتطلب مبدأ الموثوقية اتساق القياسات المتكررة لنفس الموضوع ولا تتغير النتائج التي تم الحصول عليها.[٤٤]

---

[42] Suharsimi Arikunto, Dasar-Dasar Evaluasi Pendidikan, h. 167-168

[43] Miftahul Jannah, "*Analisis Butir Soal UAS Fisika Pada Ujian Semester Genap Kelas XI Tahun Ajaran 2017/2018 di SMAN 16 Banda Aceh*"، Banda Aceh, 2019 h. 38

[44] Ibid.

ج. الهدف

يمكن القول أن اختبار نتائج التعلم هو اختبار موضوعي لنتائج التعليم، إذا تم تنظيم الاختبار وتنفيذه وفقا ل "كما هو". من حيث المحتوى أو مادة الاختبار، يحتوي مصطلح على فهم أن مادة الاختبار مأخوذة أو مصدرها من مواد أو مواد تعليمية تم تقديمها وفقا أو بما يتماشى مع أهداف تعليمية محددة تم تحديدها.[٤٥]

د. عملي

الاختبارات العملية تعني أنه يمكن إجراء اختبار نتائج التعلم بسهولة، لأن الاختبار بسيط وكامل. يمكن القول أن الاختبار يتميز بشكل جيّد إذا كان الاختبار يفي بالمتطلبات بما في ذلك الصلاحية والموثوقية ودرجة الصعوبة والقوة التمييزية والخداع وفقا لأحكام الاختبار.[٤٦]

ج. نظرة عن تحليل بنود أسئلة

أ. تعريف تحليل بنود أسئلة

تحليل بنود السؤال هو نشاط يتم إجراؤه للحصول على معلومات حول درجة بنود السؤال. يجب المعلم تحليل بنود السؤال حتى تتمكن الاختبارات المقدمة للطلاب من قياس درجة فهم الطلاب للمادة بشكل جيّد ولا تكون غامضة في توفير قيمة للطلاب. واحدة من أكثر الطرق درجة لتحسين عملية التعليم والتعلم هي تقييم اختبارات نتائج التعلم التي تم الحصول عليها من عملية التعلم نفسها. يمكن معالجة اختبارات نتائج التعلم من أجل تحسين عملية التعليم والتعلم عن طريق إجراء تحليل الأسئلة وحساب صحة وموثوقية الاختبار.[٤٧]

[45] Ibid.
[46] Ibid.
[47] M. Ngalim Purwanto, Prinsip-Prinsip danTeknik Evaluasi Program. . . , h. 118.

يمكن الحصول على معلومات الاختبارات المقدمة للطلاب لفحص مدى وظيفة الاختبار المعطى. بحيث إذا كان هناك إصلاح أو فشل، فيم كن القيام بذلك عن طريق تحليل الاختبار. يهدف تحليل نقاط الاستفهام، من بين أمور أخرى، إلى تحديد الأسئلة الجيّدة والسيئة والسيئة. مع تحليل السؤال يمكن الحصول على معلومات حول قبح المشكلة و "تعليمات" لإجراء تحسينات.[٤٨]

تحليل بنود أسئلة مهمة جدا للمعلمين للقيام بها في تقييم الطلاب. إذا لم يتم تحليل مجموعة الاختبار تلقائيا، فإن حالة الاختبار غير معروفة ما إذا كانت ستكون جيّدة أم لا. بحيث لا يمكن للنتائج التي تم الحصول عليها من التقييم تحقيق الأهداف المرجوة.

أ. درجة الصحيح

يقال إن الاختبار صحيح إذا كان الاختبار يقيس ما يريد قياسه. في الواقع، لا يتم التأكيد على مناقشة الصلاحية هذه على الاختبار نفسه ولكن على نتائج الاختبار أو الدرجات. الصلاحية هي جانب من جوانب دقة القياس. أداة القياس الصالحة ليست قادرة فقط على إنتاج بيانات دقيقة ولكن يجب أن توفر أيضا صورة دقيقة للبيانات.[٤٩]

الصلاحية هي أهم شرط في أداة التقييم. يقال إن تقنية التقييم لها صلاحية عالية إذا كانت تقنية التقييم أو الاختبار يمكنها قياس ما سيتم قياسه بالفعل.[٥٠]

بشكل عام ، هناك نوعان من الصلاحية ، وهما الصلاحية المنطقية والتجريبية. وفقا سوديجونو (Sudijono) ، هناك أنواع مختلفة من الصلاحية ، بما في ذلك:

أ. اختبار صلاحية الاختبار بعقلانية

[48] H. Daryanto, Evaluasi Pendidikan. . . , h. 179.
[49] Joko Prasetyo, *Evaluasi dan Remediasi Belajar* …, h. 65-66
[50] M. Ngalim Purwanto, *Prinsip-Prinsip dan Teknik Evaluasi Program* … h. 137

الصلاحية المنطقية تعني التفكير. لذلك بالنسبة لتقييم الأداة ، يظهر شروط الأداة التي تلبي المتطلبات الصالحة بناء على نتائج التفكير.

- صلاحية المحتوى

يقال إن الاختبار له صلاحية المحتوى عندما يقيس أهدافا محددة معينة تتوافق مع مادة أو محتوى الدرس المعطى. لأن المواد التي يتم تدريسها مدرجة في المناهج الدراسية.

- صلاحية البناء

يقال إن الاختبار له صلاحية بناءة إذا كانت بنود السؤال التي يتكون منها الاختبار تقيس كل جانب من جوانب التفكير كما هو مذكور في غرض تعليمي محدد. بمعنى آخر ، إذا كانت بنود السؤال تقيس جانب التفكير وفقا لجانب التفكير الذي هو الهدف التعليمي.

ب. اختبار صلاحية الاختبار تجريبيا

يمكن القول أن الأداة لها صلاحية تجريبية إذا تم اختبارها من التجربة. الصلاحية التجريبية هي دقة القياس بناء على نتائج التحليل التجريبي.

- صلاحية "الآن"

يقال إن الاختبار له صلاحية تجريبية إذا كانت النتائج تتوافق مع التجربة. في هذه الحالة يتم إقران النتائج بنتائج التجربة. تدور التجربة دائما حول الأشياء التي كانت في الماضي بحيث تتوفر بيانات التجربة الآن. عند مقارنة نتائج الاختبار ، يتم التعامل مع معيار أو أداة مقارنة. ثم نتائج الاختبار هي شيء يتم مقارنته.

- صلاحية التنبؤات

يقال إن الاختبار له صلاحية تنبؤية أو صلاحية تنبؤية إذا كان لديه القدرة على التنبؤ بما سيحدث في المستقبل. لمعرفة ما إذا كان يمكن التعبير عن اختبار نتائج التعلم كاختبار له بالفعل صلاحية تنبؤية أم لا ، يمكن إجراؤه من خلال: البحث عن الارتباطات بين نتائج التعلم التي يتم اختبارها للتأكد من صحة التنبؤات مع المعايير الحالية. إذا كان هناك ارتباط إيجابي كبير بين المتغيرين ، فيمكن التعبير عن نتائج التعلم التي يتم اختبارها للتأكد من صحة التنبؤ كاختبار لنتائج التعلم التي لديها بالفعل القدرة على التنبؤ الصحيحة ، بمعنى: ما تم التنبؤ به ، حقا ما حدث في الممارسة.[٥١]

كيفية معرفة صلاحية أدوات القياس باستخدام تقنية ارتباط لحظة المنتج التي اقترحها بيرسون (Pearson)[٥٢]:

$$r_{xy} = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{\left(N\sum X^2 - (\sum X)^2\right)\left(N\sum Y^2 - (\sum Y)^2\right)}}$$

معلومات	
r_{xy}	معامل الارتباط بين المتغيرات X و Y
$\sum X$	عدد نقاط العنصر
$\sum Y$	إجمالي عدد الدرجات
n	عدد الإجابات البديلة

[51] Sudaryono, *Dasar-Dasar Evaluasi* … h. 146
[52] Ibid

بناء على الشرح أعلاه، لتحديد صلاحية كل عنصر، يمكن حساب عنصر السؤال إذا كان له محاذاة أو ارتباط بين درجة العنصر والنتيجة الإجمالية. لحساب صلاحية كل عنصر يمكن استخدام صيغة الارتباط فيرسون (Pearson).

ب. درجة الموثوقية

يمكن القول أن الاختبار يتمتع بمستوى عال من الثقة إذا كان الاختبار يمكن أن يوفر نتائج ثابتة. لذا فإن فهم موثوقية الاختبار ، يتعلق بمشكلة تحديد نتائج الاختبار. أو إذا تغيرت النتائج ، يمكن القول أن التغييرات التي تحدث لا معنى لها.[٥٣]

أما بالنسبة لهذه الباحثة لحساب موثوقية أسئلة الاختيار من متعدد باستخدام برنامج (SPSS) بالصيغة كرونباخ ألفا (Alpha Cronbach)[٥٤]

ج. تقنيات تحليل بنود أسئلة

تقنية تحليل نقطة الاستفهام هي طريقة أو إجراء يستخدم لتحليل بنود السؤال. هناك العديد من تقنيات تحليل نقطة المشكلة. ومع ذلك، فإن تقنيات تحليل نقطة الاستفهام التي يستخدمها الباحثة هي كما يلي:

أ. درجة الصعوبة

هو عادة ما يتم التعبير عن فرصة الإجابة بشكل صحيح على سؤال عند درجة معين من القدرة في شكل فهرس. يتم حساب درجة الصعوبة هذا لكل رقم سؤال.[٥٥]

خصائص او درجةات عنصر السؤال الجيّد هي أسئلة ليست سهلة للغاية وليست صعبة للغاية. الأسئلة السهلة للغاية لا يمكن أن تحفز الطلاب على

[53] Anas Sudijono, *Pengantar Evaluasi Pendidikan* …, h. 184
[54] Ibid.
[55] H. Daryanto, Evaluasi Pendidikan. . . , h. 179.

تطوير قدرتهم على حل هذه المشكلات. الأسئلة الصعبة للغاية ستجعل الطلاب يائسين للمحاولة مرة أخرى لأنها تتجاوز قدرة الطلاب.[٥٦]

وفقا لأريكونتو، فإن الصيغة المستخدمة لإيجاد درجة الصعوبة هي كما يلي[٥٧]:

$$P = B/JS$$

معلومات:

P = درجة الصعوبة

B = عدد الطلاب الذين يجيبون على الأسئلة بشكل صحيح

JS = إجمالي عدد الطلاب الذين خضعوا للاختبار

لتحديد درجة صعوبة عنصر السؤال، يتم استخدام شكل الوصف

ميان (mean) **=**

جدول درجات الطلاب المتقدمين للاختبار على سؤال

إجمالي عدد الطلاب الذين خضعوا للاختبار

درجة الصعوبة :

ميان (mean)

تعيين الحد الأقصى للنقاط

[56] Suharsimi Arikunto, Dasar-Dasar, h. 122.
[57] Ibid.

الرقم الذي يشير إلى صعوبة وبساطة المشكلة يسمى درجة الصعوبة. يتراوح حجم درجة الصعوبة بين ٠,٠٠ و ١,٠٠. يشير السؤال الذي يحتوي على درجة صعوبة ٠,٠٠ إلى أن السؤال صعب للغاية، بينما يشير الدرجة ١,٠٠ إلى أن السؤال سهل للغاية.[٥٨]

كلما زاد درجة الصعوبة الذي تم الحصول عليه من نتائج الحساب، كانت المشكلة أسهل. كلما كان درجة الصعوبة أصغر سيشير إلى أن درجة بنود السؤال صعبة إلى حد ما.[٥٩]

[58] Suharsimi Arikunto, Dasar-Dasar, h. 223.
[59] Sudaryono, Dasar-Dasar., h. 128.

جدول (٢,١) درجة الصعوبة

مقرر	نوعي	P
مرفوض	صعب	P < ٠.٢٥
مقبول	متوسطة	٠.٢٥ ≤ P ≤ ٠ .٧٥
مرفوض	سهل	P>٠.٧٥

وفيما يلي متابعة نتائج تحليل درجة الصعوبة لهذا البند من السؤال:

- سجل بنود الأسئلة الجيّدة في دفتر بنك الأسئلة
- بالنسبة للأسئلة الصعبة للغاية والسهلة للغاية ، هناك احتمالان ، وهما التخلص منها أو إعادة فحصها (المراجعة).[٦٠]

ب. درجة التميزية

الدرجة المميزة للسؤال هي قدرة المشكلة على التميزية بين الطلاب الأذكياء (القدرة العالية) والطلاب الأغبياء (القدرة المنخفضة)[٦١]

الدرجة المميزة للأسئلة هي قدرة السؤال على التميزية بين الطلاب الذين أتقنوا المادة المطروحة والطلاب الذين لم يتقنوا المادة التي تم اختبارها.[٦٢]

الرقم الذي يشير إلى حجم القوة التميزيةية يسمى درجة التميز، والميةختصر D. كما هو الحال مع درجة الصعوبة، يتراوح درجة التميزية هذا من ٠,٠٠ إلى ١,٠٠. الفرق الوحيد هو أن درجة الصعوبة لا يتعرف على علامة سلبية (–) ، ولكن في درجة التميزية هناك علامة سلبية. يتم استخدام علامة سلبية على درجة التميزية إذا تم "عكس" السؤال الذي يشير إلى درجة

[60] Elis Ratnawulan & A. Rusdiana, Evaluasi Pembelajaran, h. 165.
[61] Suharsimi Arikunto, Dasar-Dasar., h. 232.
[62] Ibid. h. 233.

الشخص الذي تم اختباره. أي أن الأطفال الأذكياء يطلق عليهم اسم الأطفال الأغبياء ويطلق على الأطفال الأغبياء اسم أذكياء.[٦٣]

كلما ارتفع درجة قوة تمييز السؤال يعني أنه كلما كانت المشكلة أكثر قدرة، يميز الطلاب الذين فهموا المادة عن الطلاب الذين لا يفهمون المادة أو كلما كان السؤال أكثر قدرة، يميز الطلاب الأذكياء والطلاب الأقل ذكاء. وفقا لأريكونتو، فإن عنصر السؤال الجيّد هو عنصر سؤال يحتوي على درجة تمييز من ٠,٤١ إلى ٠,٧٠.[٦٤]

يتم تمييز حساب قوة التمايز بين المجموعات الصغيرة والمجموعات الكبيرة. المجموعات الصغيرة هي مجموعات تتكون من أقل من ١٠٠ (مائة) شخص، بينما المجموعات الكبيرة هي مجموعات تتكون من أكثر من ١٠٠ (مائة) شخص.

- للمجموعات الصغيرة

المشاركون للاختبار مقسمة إلى قسمين بالتساوي،٪ ٥٠ من المجموعة العليا والمجموعة السفلية. كل المشاركون يصطف من أعلى درجة إلى أسفل ثم ينقسم إلى نصفين.[٦٥] المثال :

مجموعة	نقاط	طلاب
مجموعة العليا (J_A)	٩	A
	٨	B
	٨	C
	٧	D

[63] Ibid.
[64] Ibid. h. 227.
[65] Ibid.

مجموعة السفلية (J_B)	٦	E
	٥	F
	٤	G
	٣	H

●للمجموعات الكبيرة

بالنظر إلى التكلفة والوقت المستغرقين للتحليل، عادة ما تأخذ المجموعات الكبيرة كلا القطبين فقط، وهي أعلى ٪ ٢٧ من النقاط كأفضل مجموعة وأدنى ٪ ٢٧ يسجلون كمجموعة دنيا.

الصيغة التي يمكن استخدامها لحساب القوة التميزيةية لأسئلة الاختيار من متعدد هي كما يلي[٦٦]:

$$\mathbf{D = B_A / J_A – B_B / J_B = P_A - P_B}$$

معلومات :

D = درجة التميزية

J^A = عدد المشاركين في المجموعة العليا

J^B = عدد المشاركين في المجموعة السفلية

B^A = عدد المشاركين في المجموعة العليا الذين أجابوا بشكل صحيح

B^B = عدد المشاركين في المجموعة السفلية الذين أجابوا بشكل صحيح

P^A = نسبة المشاركين في المجموعة العليا الذين أجابوا بشكل صحيح

[66] Suharsimi Arikunto, Dasar-Dasar, h. 223

P^B = نسبة المشاركين في المجموعة السفلية الذين أجابوا بشكل صحيح

الصيغة التي يمكن استخدامها لحساب درجة التميزيةية لمشكلة الوصف هي كما يلي :

$$DP = \frac{\bar{X}KA - \bar{X}KB}{\text{أقصى درجة}}$$

معلومات:

D = درجة التميزية

$\bar{X}$KA = متوسط المجموعة العليا

$\bar{X}$KB = متوسط المجموعة السفلية

الأقصى للدرجات = الأقصى للدرجات

يمكن استخدام تفسير نتائج حساب قوة التمايز المعايير التالي :

جدول (٢,٢) درجة التميزية

مقرر	نوعي	D
مرفوض	قبيح	٠,٠٠ — ٠,٢٠
مرفوض	كفي	٠,٢١ — ٠,٤٠
مقبول	جيّد	٠,٤١ — ٠,٧٠
مقبول	جيّد جدًا	٠,٧١ — ١,٠٠

مرفوض/ مراجع	كل شيء ليس جيّدا	سالب

العلاقة بين درجة الصعوبة والقوة التميزيةية في تجميع الأسئلة في بنك الأسئلة هي كما يلي:

- إذا كانت أسئلة صعبة للغاية وسهلة، ولكن لها قوة مميزة تفي بالمعايير، فيمكن اختيار السؤال وقبوله كأحد البدائل التي سيتم تخزينها في بنك الأسئلة.
- إذا كانت أسئلة صعبة للغاية وسهلة ثم لم تستوف القوة المميزة المعايير، فيجب مراجعة الأسئلة واختبارها مرة أخرى.[٦٧]

ج. درجة الخادعه

يجب أن يكون مخادع المشكلة معروفا بأنه ناجح أم لا. أن يكون معروفا بالعمل بشكل صحيح أم لا.صيغتها[٦٨] :

$$\mathrm{IP} = \frac{P}{\frac{N-B}{n-1}} \times 100\%$$

معلومات :

IP = درجة الخادعه

P = عدد الطلاب الذين اختاروا المشتت

N = إجمالي عدد الطلاب الذين خضعوا للاختبار

B = عدد الطلاب الذين يجيبون على الأسئلة بشكل صحيح

n = عدد الإجابات البديلة

[67] Zainal Arifin، Evaluasi Pembelajaran, h. 349.

[68] Fitriani، Analisis Butir Soal Akhir Sekolah (UAS) Mata Pelajaran Matematika Pada Tahun Ajaran 2015/ 2016 SMAN 1 Pitumpanua Kecamatan Pitumpanua Kabupaten Wajo, Skripsi, (Program Studi Pendidikan Matematika, UIN Alauddin Makassar, 2017), h. 46-47.

١ = رقم ثابت

لدرجة المخادعين على أساس مؤشر المخادعين هي:

جدول (٢,٣) الجادعه

جيّد جدًّا	IP	٢١٥٪ - ٧٦٪
جيّد	IP	١٥٠٪ - ١٢٦٪ أو ٧٥٪ - ٥١٪
غير جيّد	IP	١٧٥٪ - ١٥١٪ أو ٥٠٪ - ٢٦٪
قبيح	IP	٢٠٠٪ - ١٧٦٪ أو ٢٥٪ - ٠٪
قبيح جدًّا	IP	>٢٠٠٪

د. إطار التفكير

التقييم نتائج تعلم الطلاب في شكل اختبارات. الاختبارات هي أدوات تستخدم لقياس نتائج تعلم الطلاب. الإختبار النهائي هو أدوات الإختبار التي يستخدمها المعلمون لتحديد نتائج التعلم للطلاب خلال فصل دراسي واحد، لذلك من الضروري تحديد درجة الاختبار مسبقا.

تحليل بنود أسئلة لدرجة بعد الإختبار النهائي. تحليل الأسئلة من درجة الصلاحية والموثوقية درجة الصعوبة والتميزية والخادعه. لكن، الباحثة في الباحث تحليل الأسئلة من درجة الصعوبة والتميزية والخادعه يتم تحليل الأسئلة من درجة والصعوبة لمعرفة بنود السؤال المصنفة على أنها سهلة ومتوسطة وصعبة. ودرجة التميزية هي القياس الذين أتقنوا المادة مع أولئك الذين لم يتقنوا أو يفتقرون إلى إتقان المادة.

تنقسم نتائج تحليل درجة الأسئلة حول كل من بنود الأسئلة إلى عدة فئات، أي بناء على درجة الصعوبة والتميزية والخادعه الفئة الأولى هي نوعية جيّدة (تفسير متوسط) ونوعية رديئة (تفسير سهل وصعب) بناء على درجة الصعوبة. الفئة الثانية قبيحة، كافية، جيّدة، جيّدة جدا، وكل شيء ليس جيّدا بناء على درجة التميزية. سيتم تخزين الأسئلة ذات النوعية الجيّدة والممتازة في بنك الأسئلة. أما بالنسبة للأسئلة ذات النوعية الرديئة وكلها ليست جيّدة ، فسوف تتطلب مراجعة كبيرة حتى يتم التخلص منها بشكل أفضل.

جدول ٢,٤ مخطط إطار التفكير