المفاضلة بين طريقتي نيدلسكي و ايبيل في تحديد درجات القطع لاختبار محكي المرجع في مادة الرياضيات (دراسة مقارنة)

تريسكة عبدالحميد نوري
مديرية تربية اربيل
Shawbo.sulayman@su.edu.krd

أ.م.د.شەوبۆ شمس الدين سلاحشور
جامعة صلاح الدين–اربيل/كلية التربية
triskshahooyi@gmail.com

الكلمات المفتاحية: طريقة نيدلسكي – طريقة ايبيل– معامل ليفنجستون – مفاضلة بين درجات القطع– مصفوفة الارتباك – معيار بيرك.

كيفية اقتباس البحث

نوري، تريسكة عبدالحميد ، شەوبۆ شمس الدين سلاحشور، المفاضلة بين طريقتي نيدلسكي و ايبيل في تحديد درجات القطع لاختبار محكي المرجع في مادة الرياضيات (دراسة مقارنة)،مجلة مركز بابل للدراسات الانسانية، نيسان ٢٠٢٦، المجلد:١٦ ،العدد:٤ .

هذا البحث من نوع الوصول المفتوح مرخص بموجب رخصة المشاع الإبداعي لحقوق التأليف والنشر (Creative Commons Attribution) تتيح فقط للآخرين تحميل البحث ومشاركته مع الآخرين بشرط نَسب العمل الأصلي للمؤلف، ودون القيام بأي تعديل أو استخدامه لأغراض تجارية.

مفهرسة في Indexed
IASJ

مسجلة في Registered
ROAD

A Comparative Study between the Nedelsky and Ebel Methods for Determining Cut Scores in a Criterion-Referenced Mathematics Test (An Extracted Research Paper)

Triska Abdul Hamid Nouri
Erbil Education Directorate / University of Salahaddin-Erbi
triskshahooyi@gmail.com

Shawbo Sh. Selahshoor
AssisProf. l/College of Education
Shawbo.sulayman@su.edu.krd

Keywords : Nedelsky method – Ebel method – Livingston coefficient – Cut score comparison – Confusion matrix – Berk's classification index.

How To Cite This Article
Nouri , Triska Abdul Hamid , Shawbo Sh. Selahshoor, A Comparative Study between the Nedelsky and Ebel Methods for Determining Cut Scores in a Criterion-Referenced Mathematics Test (An Extracted Research Paper),Journal Of Babylon Center For Humanities Studies, April 2026,Volume:16,Issue 4.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

Abstract:

The purpose of this research was to compare the Nedelsky and Ebel methods for setting a cut score on a criterion-referenced test (CRT) in mathematics for seventh-grade students. The study aimed to determine which method yields a more valid and reliable classification of students' performance into mastery and non-mastery categories. The sample consisted of 440 seventh-grade students selected from various educational settings. To ensure the content validity of the test, a panel of subject matter experts reviewed the test items, and a detailed table of specifications was developed to align each item with specific instructional objectives.

In addition to expert validation, empirical evidence of validity was obtained. The reliability of the test was examined through multiple methods. Internal consistency was assessed using the Kuder–Richardson Formula 20 (KR-20), which yielded a reliability coefficient of 0.73. Furthermore, Livingston's coefficient was calculated to assess the consistency of classification decisions and was found to be 0.74.

To evaluate and compare the effectiveness of the Nedelsky and Ebel methods in determining cut scores, the researchers applied two decision-based evaluation criteria: the confusion matrix and Berk's classification index. These methods helped assess the accuracy and stability of mastery classifications based on the derived cut scores.

The findings revealed that the Ebel method outperformed the Nedelsky method in terms of both decision validity and classification reliability. Specifically, the Livingston coefficient for the Ebel method reached 0.98, significantly higher than the 0.74 recorded for the Nedelsky method. The study concluded with practical recommendations and proposed directions for future research focused on enhancing standard-setting practices and test development methodologies.

الملخص

هدفت هذه البحث إلى إجراء مقارنة منهجية بين طريقتي نيدلسكي وإيبل في تحديد درجة القطع لاختبار محكي المرجع (CRT) في مادة الرياضيات، والمُعد خصيصًا لطلبة الصف السابع الأساسي. وقد سعت الباحثتان من خلال هذه المقارنة إلى معرفة مدى فاعلية كل من الطريقتين في تقديم تصنيفات دقيقة وموثوقة لأداء الطلبة، بما يُسهم في دعم القرارات التعليمية المتعلقة بالإتقان الأكاديمي.

شملت عينة البحث (٤٤٠) طالبًا وطالبة من الصف السابع في المرحلة الاساسية تم اختيارهم من بيئات تعليمية متنوعة بهدف ضمان التمثيل المناسب للمجتمع التعليمي. وتم بناء الاختبار في ضوء مجموعة محددة من الأهداف التعليمية، حيث تم ربط كل فقرة من فقراته بهذه الأهداف من خلال جدول مواصفات محكم. كما تم عرض الفقرات على لجنة من المحكمين المتخصصين للتحقق من صدق المحتوى.

وللتحقق من صدق الاختبار أيضًا، تم جمع أدلة تجريبية داعمة. أما من حيث الثبات، فقد تم استخدام معادلة كودر–ريتشاردسون (KR–20) لقياس الاتساق الداخلي، وبلغ معامل الثبات الناتج (٠.٧٣). كما تم احتساب معامل ليفنجستون لقياس ثبات القرارات التصنيفية، وبلغ (0.74).

وفي اجراء المفاضلة بين كل من الطريقتين، تم تطبيق مصفوفة الارتباك ومعيار بيرك كمؤشرات تقييم كمية للحكم على جودة التصنيف. وأظهرت النتائج أن طريقة إيبل كانت أكثر دقة وثباتًا من طريقة نيدلسكي، حيث حققت معامل ليفنجستون (٠.٩٨) مقارنةً بـ (٠.٧٤) لطريقة نيدلسكي. وقد اختُتمت البحث بعدد من التوصيات للاستفادة من نتائج بحث و عددا من المقترحات العلمية وذلك بهدف تطوير اختبارات محكية المرجع، والتوسع في استخدام طرق علمية دقيقة لتحديد درجات القطع.

أولاً : مشكلة البحث

تستخدم الاختبارات فى تحديد مدى نجاح الطلبة وفقًا للهدف منها، ويتم إعداد هذه الاختبارات للكشف عن مدى امتلاك الفرد(الطالب) للسمة المقاسة (Gündeğer and Doğan, 2014:54). ويتم تفسير نتائج أي اختبار بعدما يحصل الطالب على درجة النجاح فى الاختبار. فمثلا في مادة مقسمة إلى وحدات من المتوقع أن يصل الطالب إلى مستوى معين من المعرفة حول الموضوعات الموجودة في الوحدة من أجل الانتقال من وحدة إلى أخرى. يتم تحديد هذا المستوى من المعرفة من خلال درجة النجاح أو القطع(Cut-off score). فاذا حصل الفرد على درجة أعلى من درجة القطع فيمكنه الانتقال إلى الوحدة التالية، وإذا حصل على درجة أقل من درجة القطع فمن المتوقع أن يعيد تلك الوحدة. ويلاحظ أنه في بعض الأحيان يتخذ المدرسون قرارات بناء على نتائج قياس غير موثوقة، وأحيانا يتم اختيار محكات غير صالحة أو يتم استخدام محكات بشكل غير صحيح، وأحيانا يتم اتخاذ القرارات دون اختيار أي محكات على الإطلاق. وبما أن مثل هذه الأخطاء تؤدي إلى قرارات غير دقيقة، فإن تحديد المحكات في التقييمات أمر مهم للغاية (Subkoviak et al., 2002:3).

جميع طرق وضع المعايير ذاتية إلى حد ما. "، إذ أنها جميعًا تنطوي على أحكام حول "ما هو القدر الكافي" أو "ما هو القدر الزائد". ولكن بمجرد وضع المعيار، يتم القضاء على الذاتية الاجرائية، أو على الأقل يتم تقليلها بشكل كبير" (Kane& Bridgeman,2017:490).

ومن منطلق أن البحث الحالى يهدف إلى المفاضلة بين طريقتي نيدلسكي و ايبل في تحديد درجة القطع لاختبار محكي المرجع في مادة الرياضيات، فإن الأسئلة التي يتم تناولها فى هذا البحث هي:

١)هل يتمتع الاختبار المحكى المرجع فى مادة الرياضيات لطلبة الصف السابع الاساسي بدرجة مقبولة من الصدق والثبات ؟

٢)أى درجات القطع لاختبار التحصيل فى الرياضيات أعلى(نيدلسكي- إيبل) ؟

٣)ما الفرق في مقارنة الأخطاء النوعية بين الطريقتين(نيدلسكي- إيبل)؟

٤)أى الطريقتان (نيدلسكي- إيبل) أفضل في معامل الصدق؟

٥)أى الطريقتان (نيدلسكي- إيبل) أفضل في معامل الثبات؟

ثانيا: أهمية البحث:

الاختبار محكي المرجع اختبار يقيس أداء المتقدم للاختبار وفقًا لمحك أو معيار معين تم الاتفاق عليه، بحيث يجب أن يصل المتقدم للاختبار إلى هذا المستوى من الأداء لاجتياز الاختبار، ويتم تفسير درجة المتقدم للاختبار بالإشارة إلى درجة المحك، وليس إلى درجات المتقدمين الآخرين للاختبار (Richard & Schmidt, 2002:66) . وبالتالي، فإن الاختبار المرجعي المعياري هو نهج للتقييم يتم من خلاله قياس أداء المتعلم فيما يتعلق بنفس المعيار في الفصل الدراسي (Mrunalini ,2013:39).

ويرى معظم خبراء القياس التربوي والنفسي أن الاختبارات محكية المرجع تتميز عن غيرها من الاختبارات التي أوضحناها فيما سبق بأنها تستخدم بدرجة أفضل في تحديد وكشف ما إذا كان الطالب حقق الأهداف التعليمية المرجوة ومعرفة الأهداف التي لم يستطع تحقيقها، وبالتالي فهي تقيس نتائج العملية التعليمية قياسا مباشراً (فتح الله،١٩٩٥: ٢٢).

ومن مميزات الاختبارات محكية المرجع أنها تساعد في تشخيص صعوبات التعلم، وسهولة وسرعة عملية التقويم، كما انها تحقق الموضوعية في تقويم التحصيل الدراسي للطلبة وتساعد في تطوير العملية التعليمية(السامرائي والخفاجي، ٢٠١٢: ٩٦٨).

ويُسهّل الاختبار محكي المرجع على المدرس تتبع مدى تقدم الطالب في تعلمه، وبناء تصور دقيق لاحتياجاته التعليمية المستقبلية. ويتم ذلك من خلال مقارنة أداء الطالب بالمستوى المتوقع من الإتقان للأهداف التعليمية المحددة سلفًا، وليس من خلال مقارنته بأداء زملائه. وعلى الرغم من أن الأساليب التقليدية التي تعتمد على مقارنة الطالب بالآخرين قد تعطي فكرة عن موقعه النسبي، إلا أنها غالبًا لا تزود المدرس بمعلومات كافية تساعده في تحديد الخطوات التعليمية التالية بشكل واضح(Esler & Dziuban,1974:173) .

وتتميّز الاختبارات محكية المرجع بدقة عالية، نظرًا لارتباط محتواها مباشرة بالأهداف التعليمية المراد قياسها، إذ تمثل فقراتها نماذج واقعية من الأداء المتوقع للمتعلّمين. وتركّز هذه الاختبارات على ما يستطيع الفرد فعله فعليًا، وما يمتلكه من معارف ومهارات محددة. وتُعد من أهم أنواع الاختبارات، خاصة في ضوء الانتقادات التي وُجّهت للاختبارات المعيارية التي تقارن أداء الطالب بأداء أقرانه، حيث تعتمد نتائجها على خصائص العينة المستخدمة سواء من حيث

الطلاب أو محتوى الأسئلة، مما يحدّ من صدق نتائجها ويجعل تعميمها أمرًا غير دقيق (علام، ١٩٩٥: ٢٢).

وتعد عملية تحديد درجة القطع Cutting score في الاختبارات محكية المرجع محور ارتكاز بناء تلك الاختبارات، لأنها تزود المدرس بتقديرات كمية لتحديد مدى تمكن الطالب من المهارات والمعارف والحكم على أدائه ومن ثم اتخاذ القرارات المناسبة (Hambleton et al., 1978:31). وذلك لأن الاختبارات المحكية لا تهتم بمقارنة أداء الفرد بأداء أقرانه، وإنما تهتم باتخاذ قرارات تعليمية تتعلق بمدى إتقان المتعلم للمحتوى الدراسي المحدد بدقة والذي يقيسه الاختبار، من خلال تصنيف المتعلمين إلى مجموعتين أحدهما متقنة لهذا المحتوى والأخرى غير متقنة، وبتشخيص جوانب الضعف عند المتعلمين غير المتقنين لتوجيههم إلى البرامج العلاجية المناسبة، إلا أن عملية التصنيف تتم بناءً على مقارنة أداء المتعلم بالنقطة المحددة مسبقاً على متصل التحصيل، والتي تقابل ما يسمى بدرجة القطع cutting score التي تفصل بين المتقنين وغير المتقنين للمحتوى الدراسي المقاس(علام، ٢٠٠٧: ٤١).

وقد أشار منشاوي (٢٠٢٢) إلى ندرة الدراسات التي تناولت صدق درجة القطع بشكل مباشر، إذ ركز أغلب الباحثين على مقارنة الدرجات التي تنتج عند استخدام طرق مختلفة لحسابها، دون التحقق الفعلي من دقتها (منشاوي، ٢٠٢٢: ٤٨٩).

وبناء على ما تم ذكره تتمثل أهمية البحث الحالي فى الاتى:

١)بناء اختبار محكى المرجع فى ماده الرياضيات للصف السابع الاساسى في اقليم كوردستان العراق وهي اول دراسة –على حد اطلاع الباحثتان – تجرى في اقليم كوردستان العراق.

٢)يبحث في المفاضلة بين طريقتي نيدلسكي و ايبل في تحديد درجة القطع لاختبار محكي المرجع في مادة الرياضيات، إذ تعتبر الاختبارات محكية المرجع أكثر انواع ا لاختبارات التحصيلية ملاءمة لقياس تحصيل الطلبة .

٣)يسهم هذا البحث في توضيح اسلوبين من اساليب تحديد درجة القطع للاختبارات محكية المرجع ، نيدلسكي و ايبل، وهذا من شأنه أن يساعد المدرس على اختيار الطريقة المناسبة لأداته ، وحيث أن صدق القرارات التربوية يعتمد على الاختيار والتطبيق المناسبين لطريقة تحديد درجة القطع.

ثالثا: أهداف البحث:

يهدف البحث الحالى الى تحقيق الهدفين الآتيين:

١)بناء اختبار محكى المرجع فى مادة الرياضيات لطلبة الصف السابع الاساسي ، وحساب صدقه وثباته وموصفاته التفصيلية.

٢)المفاضلة بين درجة القطع بطريقتى نيدلسكي وايبيل لاختبار محكي المرجع.

رابعا: حدود البحث:

يقتصر البحث الحالي على:

-**الحدود العلمية** : الكشف عن المفاضلة بين طريقتي نيدلسكي و ايبل في تحديد درجة القطع لاختبار محكي المرجع في مادة الرياضيات

-**الحدود البشرية** : طلبة الصف السابع الأساسى ومن كلا الجنسين الذكور والاناث.

-**الحدود الزمانية**: العام الدراسي ٢٠٢٤-٢٠٢٥

-**الحدود المكانية**: مدارس المرحلة الاساسية في اقليم كوردستان العراق.

خامسا: تحديد المصطلحات

أولا: الاختبار: (Tests)

•**تعريف علام(٢٠٠٣)**: أدوات قياس لجمع معلومات تتعلق بخصائص الطالب بأساليب منظمة .وهذه الخصائص تشتمل على : التحصيل الدراسي , القدرات العقلية , الاستعدادات الخاصة ,الميول , المهارات, الاتجاهات الذكاء, وسمات الشخصية , وغيرها(علام،٢٠٠٣ :٦٦)

•**تعريف براون(Brown 2003)**: طريقة لقياس قدرة الشخص أو معرفته أو أدائه في مجال معين (Brown, 2003:38)

ثانيا: الاختبار محكي المرجع: (Critertion referenced test)

•**تعريف علام (٢٠٠٧)** : ذلك الاختبار الذي يستخدم في تقييم أداء الفرد بالنسبة إلى محك مستوى أداء مطلق دون الحاجة إلى مقارنة أدائه بأداء الأفراد الآخرين (علام، ٢٠٠٧:١٢٩).

ويعرف اجرائيا بانه أداة تقويمية أُعدت خصيصًا لقياس مدى إتقان طلبة الصف السابع الأساسي لمجموعة محددة من المهارات والمعارف الواردة في مقرر الرياضيات، بناءً على معايير أداء واضحة ومحددة مسبقًا بهدف تصنيف الطلبة إلى فئتين: متقنين وغير متقنين، ووفقًا لدرجة قطع تم تحديدها باستخدام طريقة نيدلسكي.

ثالثا: درجة القطع:(Cut – of Scores)

•**تعريف هامبلتون،ونوفيك (Hambleton & Novick ,1983)** نقطة على متصل درجة الاختبار تستخدم لتصنيف الطلاب إلى فئتين تعكس مستويات الأداء المختلفة بالنسبة لهدف

معين (أو الأهداف) المراد قياسها في الاختبار Hambleton & Novick ,1983, 1983:160)

وتعرف اجرائيا بانها أقل درجة في اختبار الرياضيات المعد فى البحث الحالى يمكن ان يحصل عليها طالب الصف السابع والتى من خلال مقارنة درجته بها فيمكن اعتباره متمكنا أو غير متمكنا من مادة الرياضيات.

رابعا: طريقة نيدلسكي(Nedlsky's Method): تتطلب هذه الطريقة من كل محكم وخبير في المادة العلمية أن يتنبأ بمستوى صعوبة فقرات الاختبار، ثم يحدد من بين البدائل الخاطئة لكل فقرة تلك التي يمكن للطالب تجنبها بأقل قدر من الكفاية. والاحتمال الأدنى للإجابة الصحيحة على الفقرة هو معكوس عدد البدائل المتبقية (Nedelsky,1954: 4).

طريقة إيبل(Ebel's Method): هي طريقة وضع معايير تتطلبمن المحكمين و الخبراء في المادة العلمية تصنيف كل فقرة في الامتحان وفقًا لأهميتها وصعوبتها، ثم تتم مقارنة النتيجة المحسوبة بمصفوفة لتحديد احتمالية حصول المرشح على الإجابة الصحيحة على هذا السؤال (Ebel,1972:27)

إطار نظري و دراسات سابقة:

أولا: إطار نظري:- يشير مصطلح القياس إلى العملية الكاملة لتجميع المعلومات حول شخص ما واستخدامها لاستنتاج خصائص الشخص أو للتنبؤ بسلوكه. ويشمل مفهوم القياس جمع المعلومات من مصادر متنوعة ومقارنتها معًا، مثل المقابلات، والملاحظات، والسجلات، ونتائج الاختبارات، بالإضافة إلى البيانات التي قد تُستمد من الأسرة أو الأصدقاء أو المختصين (Groth-Marnat & Wright, 2003, p. 4).

في علم النفس،تُعد الاختبارات والمقاييس الأدوات الأكثر شيوعًا لتطبيق عمليات القياس. ويُنظر إلى كل من الاختبار أو المقياس على أنه إجراء منظم ومقنن لقياس جانب معين من السلوك.قد يكون هذا الإجراء اختبارًا مدرسيًا، أو مجموعة من العبارات يُطلب من الفرد الإجابة عنها كما في الاستبيانات والمقابلات، أو حتى مقياسًا لزمن الاستجابة (Anastasi & Urbina, 1997, p. 14).

ويرى بعض الباحثين أن هناك فرقًا بين "الاختبار" و"المقياس" بناءً على ما إذا كانت هناك إجابات صحيحة أم لا كما في اختبارات التحصيل والمعرفة مقابل مقاييس الشخصية أو الاتجاهات ومع ذلك، يستخدم كثير من المتخصصين المصطلحين بشكل مترادف دون تمييز دقيق بينهما(Geisinger et al., 2013:3) .

الاختبارات التحصيلية(Achievement Tests)

يشير مصطلح الاختبارات التحصيلية إلى الاختبارات المصممة لقياس المعرفة والمهارات والقدرات التي اكتسبها المتقدم للاختبار في مجال أو مجال موضوع أو مجال محتوى تلقى المتقدم للاختبار فيه تدريبًا أو تعليمات. قبل القرن الحادي والعشرين، لم تكن اختبارات الإنجاز تتميز دائمًا عن اختبارات الاستعداد. وقد أكد ويليام إتش أنجوف أنه في التقييمات التعليمية، لا يوجد تمييز واضح جدًا بين الإنجاز والاستعداد ولا يوجد فرق حاد بين مقاييس التحصيل ومقاييس الاستعداد. والهدف من اختبارات الاستعداد هو الإشارة إلى استعداد المتقدم للاختبار للتعلم أو تطوير الكفاءة في مجال معين إذا تم توفير التعليم أو التدريب، ويمكن لاختبارات التحصيل أن تخدم نفس الغرض(Koretz& Hamilton, 2006: 538).

الاختبارات محكية المرجع(Critertion referenced test)

صمَّم الاختبارات محكية المرجع عادة لقياس أهداف تعليمية محددة بدقة، تكون غالبًا مرتبطة بمحتوى خاص ببرنامج دراسي معين وتشكل جزءًا أساسيًا من بنية المنهج. ولهذا السبب، من الضروري أن يكون كل من المدرسين والطلبة على دراية واضحة بهذه الأهداف، حتى يمكن توجيه جهود التعليم والتعلم نحو تحقيقها بفعالية.

تُفسّر نتائج هذا النوع من الاختبارات تفسيرًا مطلقًا؛ أي أن درجة الطالب تعكس مستوى إتقانه للمادة التعليمية بحد ذاته، دون الحاجة إلى مقارنة أدائه بأداء الآخرين. فدرجة الطالب في هدف معين تبيّن النسبة التي أتقن بها تلك المهارة أو اكتسب بها تلك المعرفة.

ومن الجدير بالذكر أن توزيع الدرجات في الاختبارات محكية المرجع لا يشترط أن يكون توزيعًا طبيعيًا، ففي حال أظهر جميع الطلبة إتقانًا كاملًا لجميع الأهداف، فمن المتوقع أن يحصلوا جميعًا على نفس الدرجة، دون وجود فروق بينهم.. (Brown,1989:80).

إعداد مواصفات الاختبار محكي المرجع :

ويقصد بها مجموعة القواعد التفصيلية المحددة التى تبنى علي أساسها الفقرات الاختبارية،وقد حدد (فتح الله ،١٩٩٥: 37-34)مكونات مواصفات الاختبار في عدة نقاط وهى:

(١) الوصف العام للهدف (٢) فقرات المثير (٣) فقرات الاستجابة (٤) عينة من الفقرات (٥) ملحق المواصفات (٦)درجة القطع- حد الإتقان ٧-زمن الإجابة

تحديد الخصائص السيكومترية للاختبار محكي المرجع

اولا:- صدق الاختبار محكي المرجع

•انواع صدق الاختبارات المحكية المرجع:

١.الصدق الوصفي(Descriptive Validity):

يُعدّ مفهوم **الصدق الوصفي (Descriptive Validity)** من الجوانب الهامة في تقييم الاختبارات، ويُشير إلى مدى قدرة الاختبار على قياس السلوك المراد قياسه بدقة وشمولية. وقد أشارت بعض الدراسات إلى أن الصدق الوصفي يُشابه **صدق المحتوى** في الاختبارات معيارية المرجع، مع الأخذ في الاعتبار أن الصدق الوصفي يتميز بكونه أكثر شمولية، لا سيما عند تقييم الجوانب الوجدانية والنفسحركية للسلوك. ويرى **بوبام (Popham)** أن الصدق الوصفي يمثل محاولة لتأكيد ما يقيسه الاختبار محكي المرجع بالفعل، ويُعتبر تحديد هذا النوع من الصدق شرطًا أساسيًا لبقية أنواع الصدق الأخرى. فالصدق الوصفي هو أفضل وسيلة لتفسير أداء الطالب في الاختبار، ويُعتبر الاختبار صادقًا إذا استطعنا من خلاله وصف أداء الفرد بدقة بناءً على السلوك المستهدف بالقياس ، ويُشار إلى أن هذا النوع من الصدق يُسمى أحيانًا "صدق المحتوى" في الاختبارات معيارية المرجع. وفي هذا السياق، يُعدّ الصدق الوصفي مؤشرًا لصدق كل فقرة من فقرات الاختبار على حدة، وكذلك لصدق الاختبار ككل. وهذا يعني أن كل فقرة تقيس ما وُضعت لقياسه، وأن الاختبار بأكمله يقيس ما صُمم لقياسه (علام، ٢٠٠٧: ٦٤).

٢.الصدق الوظيفي(Functional Validity):

أشار علام إلى مفهوم الصدق الوظيفي Functional Validity بوصفه مفهوما موازيا للصدق المرتبط بالمحك في الاختبارات معيارية المرجع (علام، ١٩٨٦: ٨٢)،كما أوضح بابام Popham أن المقصود بالصدق الوظيفي هو مدى دقة الاختبار محكي المرجع في أداء الوظيفة التي صُمم من أجلها وقد أُطلق عليه هذا الاسم لأن بعض الخصائص في الاختبارات محكية المرجع تستلزم وجود محك خارجي بينما خصائص أخرى لا تتطلب ذلك وفي هذا السياق يُستخدم التصنيف على أساس درجة القطع بحيث يُعد الطالب متقنا إذا تجاوز هذه الدرجة وغير متمكن إذا كانت درجته أقل منها كما ورد في و(علام، ٢٠٠٧: ٦٥)

٣.صدق انتقاء النطاق السلوكي(Domain Selection Validity):

يشير بابام (Popham)بأن صدق انتقاء النطاق السلوكي يتعلق بمدى دقة باني الاختبار في اختيار نطاق سلوكي محدد يمكن أن يستخدم كمؤشر لإيضاح مستوى الطالب بالنسبة إلى نطاق عام ،وقد اقتراح بابام(Popham)طريقتين لتحديد صدق انتقاء النطاق السلوكي وهما:

١ – تدريس الطلبة مجموعة من المهارات ، حتى يتكمنوا من اتقانها، ثم ينظر بعد ذلك في تقييم إمكانية تعميم قدرتهم على استيعاب والتمكن من هذه المهارات إلى نطاقات أخرى (اى مهارات اخرى).

٢ – تحديد مجموعتين الأولى تلقت التعليم والثانية لم تتلق التعليم ، وتطبيق عدة اختبارات عليهما، ثم القيام بتحديد النطاق السلوكي الذي يمكنه التمييز بين هاتين المجموعتين. (علام ، ٢٠٠٧: ٦٦).

٤– صدق البناء(Construct Validity):

على الرغم من أن الاعتماد على أدلة الصدق المرتبطة بالمحتوى أو المحك يُعد كافيًا في كثير من أنواع الاختبارات، إلا أن هناك بعض الحالات التي لا يمكن فيها استخدام هذا النوع من الأدلة بشكل فعّال، لا سيما عند قياس السمات النفسية المجردة مثل العدوانية أو الميل إلى الجدال، أو القدرات المعرفية العامة كعامل الذكاء. فهذه البنى لا يمكن تعريفها بدقة من خلال محتوى اختبار محدد أو محك خارجي واضح، ولمعالجة هذه الإشكالية قدم كل من كرونباخ وميل في عام(١٩٥٥)مفهوم "صدق البناء"، وهو يشير إلى مدى اتساق نتائج الاختبار مع **مجموعة** من المؤشرات السلوكية والنظرية المرتبطة بالبنية النفسية التي يُفترض أن يقيسها الاختبار. ويُعد هذا النوع من الصدق ضروريًا حين تكون الفرضيات المستندة إلى النظرية هي المرجع في تفسير دلالة نتائج الاختبار، ويُستخدم في تقييم مدى تمثيل الاختبار لتلك البنية النظرية المستهدفة.(Markus & Borsboom, 2013: 78)

استخراج الخصائص السيكومترية لفقرات اختبار محكي المرجع :

١–معامل حساسية الفقرة(Item Sensitivity):

وهي عبارة عن قدرة الفقرة على تحديد مستوى الفروق الفردية بين الأفراد المفحوصين الذين يمتلكون السمة أو يعرفون الإجابة الصحيحة وبين الذين لا يمتلكون السمة المقاسة أو لا يعرفون الإجابة الصحيحة لكل فقرة من فقرات الاختبار" وهو مدى فاعلية المقياس في التمييز بين الأفراد المتقنين والأفراد غير المتقنين في السمة أو الخاصية التي يهدف الاختبار إلى قياسها. (Hani et al., 2023:64).

ويمكن ان يتم التحقق من حساسية الفقرة من خلال استخراج معامل برينان وهو مشتق من الطريقة المعروفة لحساب معامل التمييز والذي يتم حسابه عن طريق حساب الفرق في صعوبة الفقرة بين أفراد المجموعة العليا وأفراد المجموعة الدنيا. (Lin,1988:34–35)إذ يستبدل مفهوم

المجموعة العليا بالمجموعة المتقنة، والمجموعة الدنيا بالمجموعة غير المتقنة وحدده برينان(Brennan) بالمعادلة الاتية والتي تعرف بمؤشر برينان (B–Index):

$$B = \frac{U}{n_1} - \frac{L}{n_2}$$

حيث ترمز:

B:الى معامل برينان.

U: الى عدد المتقنين الذين أجابوا عن الفقرة إجابة صحيحة.

L: الى عدد غير المتقنين الذين أجابوا عن الفقرة إجابة صحيحة.

n_1 : الى عدد المتقنين

n_2 : الى عدد غير المتقنين

وفى المجموعه (n_1) يمكن ان تكون عدد افرادها مساوية أو مختلفة عن عدد افراد المجموعه (n_2)(Lin,1988:34–35) ، وتتراوح قيمة هذا المعامل بين(+١،–١) كما تعد الفقرة جيدة وفق معامل التمييز (ب) إذا كانت قيمته أكبر من أو تساوي (٠.٢٠)

٢–معامل صعوبة الفقرة:Item Difficulty index

معامل صعوبة الفقرة يعد من ابسط الاحصاءات المستخدمة في تحليل الفقرات .وهو يعبر عن نسبة الافراد الذين اجابوا على الفقرة بشكل صحيح .يحسب هذا المعامل بقسمة عدد الافراد الذين اجابوا اجابة صحيحة على عدد الافراد الذين خضعوا للفقرة في الاختبارات محكية المرجع، يستخدم معامل الصعوبة لتحديد ما اذا كانت الفقرة تقيس مستوى الاداء المطلوب بدقة، اي ما إذا كانت غالبية الافراد قد تمكنوا من تحقيق الهدف المرتبط بالفقرة .اذا كانت نسبة الاجابة الصحيحة منخضة جدا فقد يدل ذلك على ان الفقرة لا تعبر بدقة عن الهدف، اوان المحتوى لم يغط بشكل كاف اثناء التدريس (Croker & Algina,1986:311) ، ويمكن التعبير عن معادله الصعوبه بما يلى :

$$P = \frac{n_u + n_l}{N}$$

حيث ترمز:

P:الى معامل صعوبة الفقرة.

n_u:الى عدد الافراد الذين اجابوا على الفقرة بشكل صحيح من المجموعة العليا

n_l : الى عدد الافراد الذين اجابوا على الفقرة بشكل صحيح من المجموعة الدنيا

N:الى العدد الكلي للافراد من المجموعتين العليا و الدنيا معاً.(عبابنة،٢٠٠٩:٧٧).

٣-صدق الفقرة باستخدام معامل الارتباط الثنائي التسلسل(Point-Biserial Correlation):

يستخدم لتحديد قوة واتجاه العلاقة بين متغيرين احدهما ثنائي والاخر كمي مستمر،كلما اقتربت قيمة المعامل من (+١) او (-١)زادت قوة العلاقة بين المتغيرين وتستخدم لمعرفة صدق الفقرة في الاختبارات وذلك كمؤشر لوجود ارتباط الفقرة بالدرجة الكلية للاختباروذلك عندما تكون الاجابة ثنائية الاجابة (٠-١). فاذا كانت القيمة الموجبة فهذا يشير الى ان القيمة الاكبر للمتغير المستمر تربط غالبا بالقيمة (١) للمتغير الثنائي ،اما اذا كانت سالبة فترتبط القيمة الكبرى للمتغير المستمر غالبا بالقيمة (٠).

$$r_{pb}\frac{\bar{x}_1-\bar{x}_0}{s}\times\sqrt{\frac{pq}{n}}$$

حيث ترمز:

r_{pb}: الى معامل الارتباط الثنائي التسلسل.

$\bar{x}_1$: الى المتوسط الحسابي الدرجات الكلية للذين اجابوا صحيحا

$\bar{x}_0$:الى المتوسط الحسابي الدرجات الكلية للذين اجابوا خطا

p:الى نسبة الذين اجابوا بشكل صحيح.

q:الى نسبة الذين اجابوا بشكل خاطيء. (Gravetter and Wallnau,2017:518)

صدق الفقرة باستخدام معامل ارتباط فاي(Phi -Correlation Coefficient):

يُستخدم معامل ارتباط فاي (ϕ) كأحد الأساليب الإحصائية المناسبة لقياس قوة العلاقة بين متغيرين ثنائيين (dichotomous)، وغالبًا ما يُستخدم في الجداول ذات البُعدين (٢×٢). وفي سياق الاختبارات محكية المرجع، يُعد معامل فاي أداة فعالة للتحقق من صدق الفقرة، وذلك من خلال فحص مدى الارتباط بين استجابة الطالب على فقرة محددة (صحيحة أو خاطئة) وانتمائه لفئة المتقنين أو غير المتقنين، استنادًا إلى درجة القطع (Norton, 1978, :4)، تُعد هذه الطريقة مناسبة لأن طبيعة البيانات في الاختبارات محكية المرجع تكون ثنائية غالبًا، حيث

يُصنّف الطلاب إلى فئتين (متقن / غير متقن)، وتُسجّل استجابتهم على الفقرات أيضًا بشكل ثنائي (إجابة صحيحة / غير صحيحة). لذلك، فإن استخدام معامل فاي يمكّن الباحث من فحص مدى قدرة الفقرة على التمييز بين الفئتين، أي ما إذا كانت الفقرة تعكس بالفعل الأداء المتوقع من الطالب المتقن مقارنةً بغير المتقن، مما يوفر مؤشراً على صدقها في قياس المهارة المستهدفة، وكلما اقتربت قيمة فاي من (±١) دلّ ذلك على وجود ارتباط قوي بين أداء الطالب على الفقرة وتصنيفه من حيث الإتقان، وهو ما يُعد مؤشراً إيجابياً على صدق الفقرة. أما إذا كانت القيمة قريبة من الصفر، فذلك قد يشير إلى أن الفقرة لا تميز بشكل كافٍ بين المتقنين وغير المتقنين، ما قد يستدعي مراجعتها أو استبعادها من الاختبار(Crocker & Algina, 2006:92-93).ويمكن حسابه على وفق المعادلة الاتية:

$$Phi\ Coefficient = \frac{AD+BC}{\sqrt{A+B}(C+D)(A+C)(B+D)}$$

.(Norton, 1978:6)

ويمكن أن تتراوح قيم معامل فاي من(٠ إلى +١.٠)، وتعني القيمة(١.٠)أن هناك ارتباطًا مثاليًا بنسبة (١٠٠%)بين المتغيرين. (Agresti,1996:101).

طرق تقدير ثبات الاختبارات محكية المرجع:

تُعدّ اختبارات محكية المرجع أدوات فعالة في إصدار أحكام دقيقة حول أداء الأفراد، ويُعدّ التحقق من ثبات هذه الأحكام أمرًا أساسيًا في عملية بنائها. فلا يكفي أن تُصمم لقياس محتوى تعليمي محدد، بل يجب أن تضمن اتساق نتائجها التصنيفية عند تطبيقها في مناسبات مختلفة. ويُقاس الثبات في هذا السياق ليس على أساس ترتيب الأفراد كما هو الحال في الاختبارات معيارية المرجع، بل من خلال مدى تكرار ودقة تصنيف الأفراد إلى فئات مثل "متقن" و"غير متقن" دون تأثر بالعوامل العشوائية أو التذبذب في الأحكام.(Cizek & Bunch, 2007: 130) .

ويمكننا ان نستعرض الطرق و المعادلات التي تستخدم للتحقق من ثبات الاختبارات المحكية وفق ما يلي:

١-استخدام معادلة كودر-ريتشاردسون (٢٠):$Kuder-Richardson_{20}$).

يُعد ثبات القرار (Decision Consistency) من أهم المعايير التي يُعتمد عليها عند تقييم موثوقية الاختبارات محكية المرجع، لما له من دور مباشر في ضمان دقة التصنيف بين المتقنين وغير المتقنين. إلى جانب ذلك، قد يُستخدم معامل كودر-ريتشاردسون ٢٠ (KR-20) كأداة

داعمة، خاصة في المراحل الأولى من بناء الاختبار، وذلك للتحقق من الاتساق الداخلي للفقرات ومدى تجانسها في قياس السمة أو المحتوى المحدد. يوفر هذا التقدير مؤشراً أولياً على جودة الفقرات وترابطها، مما يساهم في تعزيز الثقة في بنية الاختبار قبل استخدامه لأغراض التصنيف (Crocker & Algina, 1986: 139)، ويُعد تطبيق(KR–20) مناسبًا بشكل خاص للاختبارات التي تكون فقراتها ثنائية التصنيف (صواب/خطأ أو متقن/غير متقن)، وهي الصيغة الشائعة التي تستخدم في الاختبارات المحكية المرجع. كما أن هذه المعادلة لا تفترض وجود تجانس في مستوى صعوبة الفقرات، مما يجعلها عملية ومرنة في ظروف التصميم الواقعية. ويُشير معامل KR–20 المرتفع إلى أن الفقرات تتشارك في قياس بُعد معرفي واحد، وهو شرط ضروري لضمان دقة قرارات الإتقان لاحقًا. ولحساب معامل ثبات(كيودر ريتشاردسون–٢٠) نستخدم المعادلة الاتية:

$$KR-20=\frac{k}{k-1}\ (1-\frac{\sum piqi}{\delta^2})$$

اذ ترمز :

K : الى عدد الفقرات (الأسئلة)

pi: الى نسبة الافراد الذين أجابوا على الفقرة i إجابة صحيحة (أي أخذوا ١)

qi: الى نسبة الافراد الذين أجابوا على الفقرة i إجابة خاطئة (أي أخذوا ٠)

∑piqi : الى حاصل جمع ضرب pi و qi لكل الفقرات

σ2 : التباين الكلي لدرجات الاختبار الكاملة للطلبة (أي تباين مجموع الدرجات الكلية لكل فرد)

وتزيد قيم معامل الثبات وهي معاملات الاستقرار والتكافؤ والاتساق الداخلي بزيادة تباين الدرجات .(Kuder& Richardson, 1937:153)

٢–معامل ليفنجستون(Livingston Index):

وهذا المعامل يهتم بإيجاد انحراف درجات الأفراد عن درجة القطع، إضافة إلى اهتمامه بمتوسط مجموع مربعات انحرافات الدرجات الملاحظة المتوقعة وفق الصيغة الاتية:

$$k^2(X,T)=\frac{SD^2(KR20)+(M_x-C)^2}{SD^2+(M_x-C)^2}$$

إذ ترمز :

$K^2(X,T)$: الى معامل ليفنجستون

T: إلى متوسط مجموع مربعات انحرافات الدرجات الحقيقية عن درجة القطع (C)

X: الى متوسط مجموع مربعات انحرافات الدرجات الملاحظة حول درجة القطع (C)

M : إلى متوسط درجات الطلبة في المجال السلوكي الذي يقيسه الاختبار

Ni. : إلى عدد الأسئلة

C : الى درجة القطع (مجيد، ٢٠١٤: ٤٩):

وتنحصر قيم هذا المعامل بين(+١، – ١) (علام، ٢٠٠٧ : ٢٩٩).

٣-طريقة هاريس(Harris Method) :

هذه الطريقة لا تعتمد في تحديدها على متغير طول الاختبار (K) ، وانما تقوم على الارتباط بين متغير يمثل مجموع درجة الطالب في الاختبار ومتغير ثنائي آخر يمثل تصنيف الطالب إلى متقن أو غير متقن على ضوء درجة القطع ، لذا فأنه يتم تصنيف الطلبة إلى أربعة مجموعات في جدول رباعي وهي طلاب متقين فعلا ، وطلاب غير متقين فعلا . ومجموعة طلاب الخطأ الأول(α) ومجموعة طلاب الخطأ الثاني (β) ، ويتم تحديدهم بناء على درجة القطع للاختبار ويطلق على هذه الطريقة بمعامل هاريس . ويرمز له بالرمز (MC) ويتم تقديره بالصيغة التالية :

$$MC = \frac{SSB}{SSB+SSW}$$

إذ ترمز:

Mc: الى معامل ثبات هاريس

(SSW,SSb)= الى مجموع المربعات (داخل ، بين) المجموعات وتمتد قيمة المعامل بين قيمتي (٠-١) (مجيد، ٢٠١٤ : ٥٢).

٤-معامل كابا (Kappa):

يعد معامل كابا أحد المعاملات العديدة المستخدمة لتقدير ثبات المقدرين وأنواع مماثلة من الثبات، تم تطويره في عام(١٩٦٠)من قبل جاكوب كوهين(Jacob Cohen). في مفهومه الأصلي، كان معامل كابا، الذي يرمز له(κ)، مؤشرًا يستخدم لقياس مستوى الاتساق بين اثنين من المقيمين الذين يستخدمون معايير أو أدوات أخرى لوضع الأشخاص في إحدى فئات(κ)الاسمية. على سبيل المثال، قد يستخدم اثنان من المقيمين معيارًا لتصنيف الاستراتيجيات التعليمية المستخدمة في الفصل الدراسي كواحدة من فئتين اسميتين، مثل فعالة أو غير فعالة، أو

قد يستخدم اثنان من علماء النفس أداة لتحديد اكتئاب الشخص على أنه اكتئاب حاد أو اضطراب ثنائي القطب أو اكتئاب مستمر أو اكتئاب ذهاني. في كلتا الحالتين، قد تؤدي الذاتية إلى خلافات حول الفئة المخصصة وتثير مخاوف بشأن استخدام وتفسير الأداة. لهذا السبب، تم تطوير معاملات الاتفاق، مثل(κ). منذ نشأته، كان κ واحدًا من أكثر معاملات الاتفاق استخدامًا على نطاق واسع، وقد استُخدم في مجالات مثل التعليم والطب والعلوم الاجتماعية. وفي هذه المجالات وغيرها، يمكن استخدام κ ليس فقط لتقدير الثبات ولكن أيضًا لقياس التباين الذي يمكن أن يُعزى إلى عملية التصنيف(Cohen, 1960:38) .

وتتراوح قيمة معامل كابا بين (−١، +١) ، وتعني القيم السالبة عدم الاتساق أي عدم الثبات ، وكذلك تتأثر قيمته باختلاف درجة القطع وتكون معادلة معامل كابا مما يلي :

$$K = \frac{P-P_0}{1-P_0}$$

إذ ترمز :

K: الى معامل كابا

P: الى نسبة الاتفاق الملاحظ

Po: نالى سبة الاتفاق المتوقع(علام، ١٩٩٥ : ٤٢)

٥-**معامل كارفر(Caver)** تُستخدم لقياس **ثبات القرارات** المتعلقة بإتقان الطلبة عند تطبيق اختبارين متوازيين أو متماثلين على نفس المجموعة.

وتعتمد هذه الطريقة على مقارنة عدد الطلبة الذين تم تصنيفهم على أنهم **متقنون أو غير متقنين في كلا الاختبارين**، فإذا كان هناك تطابق في التصنيفات، يُعد الاختباران ثابتين.وتحسب درجة الثبات باستخدام المعادلة التالية:

$$\frac{A+C}{N}$$

إذ ترمز :

A: الى مجموع الطلبة المتقنين بالفعل

C : الى مجموع الطلبة غير المتقنين بالفعل

N : االى لعدد الكلي للطلاب الذين خضعوا للاختبارين.

وتُفسَّر القيمة الناتجة كنسبة مئوية لمدى **اتساق قرارات التصنيف** (إتقان/عدم إتقان)لكن يُنتقد هذا الأسلوب بأنه **غير حساس للفروقات الدقيقة في الأداء** بين الطلبة، لأنه لا يأخذ في الاعتبار درجة قرب أو بعد الطالب عن حد الإتقان، بل يكتفي بالتصنيف الثنائي فقط (Brown, 1989 : 87)

درجة القطع في الاختبارات المحكية:

تتعدد مسميات درجة قطع الاختبارات محكية المرجع بتعدد آراء علماء القياس والتقويم التربوي وخبراتهم في هذا المجال، ومن تلك المسميات وهي (درجة القطعCut off score) و (مستوى التمكنMastery Level) و (الحد الأدنى من الكفايةMinimum Competency) و (مستويات المحك Criterion Levels –) و (درجة النجاح – Passing Score) (عبابنة ، ٢٠٠٩: ١٢٧).

وهناك من ١٠ إلى ٢٠ طريقة رئيسة لتحديد درجة القطع، ويوجد بينها نقاط اختلاف كثيرة ، (Hambelton 2009)، وقد ذكر علام(٢٠٠١) طرق تحديد درجة القطع بانها تنقسم الى ثلاث مجموعات بشكل عام وهي :

- المجموعة الاولى : طرق تعتمد على التحكيم فقط.
- المجموعة الثانية :طرق تعتمد على التحكيم و تسترشد بالبيانات.
- المجموعة الثالثة : طرق تعتمد على البيانات التجريبية و تسترشد بالتحكيم.

(علام، ٢٠٠١: ٢٥٥-٢٥٦)

وسوف تقوم الباحثتان بتقسيم هذه الطرق الى طرق ترتكز على فحص الاختبار وفقراته و طرق تعتمد على تحديد درجة القطع بناء على تحديد مستوى الطالب ويتم عرض هذه الطرق وفق ما يلي :

اولا:- طرق تحديد درجه القطع التي ترتكز على فحص الاختبار وفقراته: تعتمد الأساليب التي ترتكز على الاختبار على رأي الخبراء فيما يتعلق بالفقرات المدرجة في الاختبار. يقوم الخبراء بتقييم كل فقرة وفقًا لمستوى الأداء المحدد وفي نهاية هذه العملية، يتم حساب النتيجة النهائية (Taşdelen, 2009, : 5). وهذه الطرق هى انجوف(Angoff وايبل Ebel ونيدليسكى Nedelsky)و نعم/ لا (**Angoff Yes/No**)وطريقة وضع العلامات(Marking) .

وسوف تركز الباحثتان على طريقتي نيدلسكي و ايبيل كونها اعتمدتا هاتين الطريقتين لتحديد درجة القطع في الاختبار المعد في البحث الحالي

١-. طريقة نيدلسكي (Nedlsky's Method)

تعد سهولة التطبيق سببًا مهمًا لاستمرار استخدام الطريقة التي اقترحها نيدلسكي في (عام ١٩٥٤) حتى اليوم. في السنوات التي تم فيها تقديم هذه الطريقة،. ويعتبر أسلوب نيدلسكي من الأساليب التي تعمل على تسريع الانتقال من مستويات الأداء المبنية على مقارنة الأشخاص فيما بينهم إلى أسلوب التقييم الذي يوضح مستويات الأداء الفعلية للأشخاص (Cizek and Bunch, 2007: 52).

تعتمد طريقة نيدلسكي في تحديد درجة القطع على تحليل كل فقرة في الاختبار، حيث يُقدّر الخبراء مدى قدرة الطالب غير المتقن على استبعاد الخيارات الخاطئة. وبعد تحديد عدد البدائل التي قد تبقى دون استبعاد، يُحسب احتمال اختيار الإجابة الصحيحة من بينها. هذا الاحتمال يمثل درجة الفقرة. تُجمع هذه التقديرات لكل الفقرات، ثم يُحسب المتوسط الحسابي لها لتحديد درجة القطع النهائية للاختبار. (Cizek and Bunch, 2007: 52). ويتم حساب احتمالية حل فقرة الاختبار من خلال مراجعة البدائل غير الصحيحة للفقرات الموجودة في الاختبار. مع الأخذ في الاعتبار السمة التي تم قياسها في الاختبار، يقرر الخبراء عدد الخيارات غير الصحيحة في كل فقرة من فقرات الاختبار يمكن للشخص غير المتمكن تخمينه بشكل صحيح. وبهذه الطريقة، هناك عدد من الخيارات لكل فقرة من فقرات الاختبار لا يمكن للأشخاص غير المتمكنين استبعادها. احتمال التعرف على الإجابة الصحيحة من بين الخيارات التي لا يمكن حذفها يعطي درجة الفقرة. على سبيل المثال، يتوقع أحد الخبراء أن خيارين في فقرة مكونة من خمسة خيارات سيتم حذفهما الأشخاص المؤهلين. في هذه الحالة، هناك(٣)خيارات متبقية. احتمال أن يجد الشخص الإجابة الصحيحة من بين الخيارات الثلاثة المتبقية هو واحد من ثلاثة. وبناء على ذلك، فإن احتمال حل الفقرة المعنية، الذي يقدره الخبير، يُحسب بـ(٠.٣٣). تتم إضافة كافة الاحتمالات المحسوبة معًا للحصول على النتيجة النهائية لكل خبير. يتم حساب درجة القطع للاختبار لطريقة نيدلسكي من خلال أخذ المتوسط الحسابي لدرجات القطع (Zieky,M &Perie, 2006, : 10-12). ويتم التعبير عن صيغة الحساب على النحو التالي (Tanrıverdi, 2006: 9).

$$T_{ci} = \sum_{i=1}^{n} (qi - k_{ii})^{-1}$$

حيث ترمز :

Tcj: الى درجة القطع التي يحددها الخبير

qi: الى عدد الخيارات في الفقرة (i).

kij: الى عدد الخيارات التي يعتقد الخبير(j) أنه من الممكن استبعاد طالب من الحد الأدنى من الكفاءة منها بشكل غير صحيح.

n: الى عدد فقرات الاختبار

٢.طريقة إيبل: (Ebel's Method)

في طريقة إيبل، يقوم الخبراء و المحكمين أولاً بفحص كل فقرة من فقرات الاختبار لتحديد مستوى الصعوبة (سهل، مناسب، وصعب) والأهمية (أساسي، مهم، مقبول، ومشكوك فيه) لكل فقرة. بعد ذلك، يتم تصنيف كل فقرة في جدول مصفوفة(٣ × ٤)وفقًا لمستوى الصعوبة والأهمية. بعد ذلك، تحدد اللجنة النسبة المئوية المتوقعة للإجابات الصحيحة على الفقرات في كل خلية من جدول المصفوفات، والذى يقوم بذلك شخص يتصف بالكفاءة. أخيرًا، يتم ضرب عدد الفقرات في كل فئة في النسبة المئوية المتوقعة للإجابات الصحيحة، ويتم إضافة النتائج الإجمالية لحساب درجة القطع. تتضمن طريقة إيبل عملية وضع معايير أكثر تعقيدًا من طرق وضع المعايير الأخرى، والتي تستند إلى تحليل محتوى أداة الاختبار، وبالتالي فهي تفرض عبئًا على المحكمين (Alsmadi, 2007:480).

اجراءات المفاضلة بين درجات القطع :

اولا:صدق درجة القطع: (Cut Score Validity)

هناك اختلاف بين صدق محتوى الاختبار ،وصدق درجة القطع ؛ إذ أن صدق المحتوى للاختبار يقصد به الى أى مدى تمثل الفقرات التي يشتمل عليها الاختبار المحتوى المطلوب قياسه ، في حين أن صدق درجة القطع يقوم على مدى دقة هذه الفقرات في تقسيم الأفراد(الممتحنين) إلى متقنين وغير متقنين ، وعادة ما يعتمد هذا الصدق على محك خارجي للأداء ، ويعتمد صدق درجة القطع لدرجات الاختبار في هذه الحالة على نجاحها في التمييز بين هاتين الفئتين من الممتحنين(المتقنين وغير المتقنين)(علام، ٢٠٠٠: ٣٢٧)، إذ يتم تصنيف الممتحنين إلى فئتين للإتقان في ضوء درجة القطع المراد حساب صدقها، والمحك الخارجي

،ويتمثل في وجود الشخص فى المجموعة ، والجدول التكرارى (الجدول١) الاتى يمثل هذا الاجراء:

جدول(١) تصنيف الممتحنين إلى فئتين للإتقان في ضوء درجة القطع المراد حساب صدقها، والمحك الخارجي

التصنيف في ضوء المحك / التصنيف في ضوء درجة القطع	متمكن	غير متمكن	المجموع
متمكن	a	b	a+b
عير متمكن	c	d	c+d
المجموع	a+c	b+c	n

ومعامل صدق درجة القطع هو معامل ارتباط بيرسون بين المتغيرين الاتيين:

الأول- هو التصنيف المتوقع: حيث تقدر درجة كل تلميذ حصل على درجة أعلى من درجة القطع أو تساويها القيمة (١)، بينما تقدر (صفر) للتلميذ الذي حصل على درجة أقل من درجة القطع.

الثاني- هو المحك الخارجي: حيث يقدر كل تلميذ تم تصنيفه من قبل المحكمين على أنه متقن بالقيمة (١)، في حين يقدر (صفر) في حال تم تصنيفه في عينة غير المتقنين (عبد السلام، 1992 ب: 33).

أما علام (2000 : 328) يشير إلى أنه في حال كان المتغيران من النوع المتقطع تقطع ثنائي، فإنه يمكن استخدام معامل فاي (ϕ) المستنتج من معامل ارتباط بيرسون وفق المعادلة الاتية:

$$Phi\ Coefficient = \frac{AD+BC}{\sqrt{A+B}(C+D)(A+C)(B+D)}$$

حيث ترمز:

AD: الى نسبة التلاميذ الذين صنفوا متقنين قبل وبعد تطبيق الاختبار.

C : الى نسبة التلاميذ الذين صنفوا متقنين قبل تطبيق الاختبار.

B : الى نسبة التلاميذ الذين صنفوا متقنين بعد تطبيق الاختبار.

وبما أن (معامل ϕ)هو امتداد لمعامل ارتباط بيرسون فقيمته تكون محصورة بين (+١ – -١)، حيث يصل (معامل ϕ إ)لى أعلى قيمة له (+١) (علام، ٢٠٠٠ب: ٣٢٧).

ثبات درجة القطع(Cut Score Reliability):

يُعد معامل ليفنجستون Livingston's r من المؤشرات الإحصائية الجوهرية المستخدمة للتحقق من ثبات درجة القطع (cut score) في الاختبارات المحكية، حيث يوفر تقديرًا لمقدار الاتساق في القرارات التصنيفية التي تنتج عن استخدام درجة القطع عبر تطبيقات مختلفة للاختبار ذاته . وتكمن أهمية هذا المعامل في كونه يأخذ في الاعتبار التباين الكلي في الأداء، والاحتمال الإحصائي للخطأ في تصنيف الأفراد على جانبي درجة القطع، مما يعزز من مصداقية نتائج التقويم واتساقها عبر الزمن".(Livingston & Zieky, 1982: 33)

ويتم استخدام معادلة ليفنجستون والتي تكون بالصيغة الاتية لاستخراج معامل ثبات درجة القطع:

$$Rcrt = \frac{Rxx.SD^2(Mean - Criterion\ Score)^2}{SD^2 + (Mean - Criterion\ Score)^2}$$

إذ ترمز:

Rcrt الى ثبات القرار لدرجة القطع

R_{xx} الى ثبات الاختبار المعياري المرجع اباستخدام معادلة كيودر ريتشارسون-٢٠ او الفا كرونباخ

SD^2 الى تباين درجات الافراد على الاختبار

Mean Criterion Score **الى** المتوسط الحسابي للاختبار الى درجه القطع المستخدمة للاختبار (Berk,1986:67)

ثالثا: استخدام مصفوفة الارتباك (Confusion Matrix):

مصفوفة الارتباك كأداة للحكم على أفضلية التصنيف

تُعد مصفوفة الارتباك (Confusion Matrix) من الأدوات الإحصائية المهمة في تقييم أداء النماذج التصنيفية، تحديداً في مجال البحوث التربوية والقياس النفسي، لما توفره من معلومات دقيقة حول نوعية القرارات التصنيفية (الصحيحة والخاطئة) التي يتخذها النموذج أو الطريقة المستخدمة في التقدير.(Kubat, 2017 : 93)

تتكون مصفوفة الارتباك من أربعة فقرات أساسية:

- **True Positive (TP)**:حالات تم تصنيفها بشكل صحيح على أنها إيجابية (مثل: طالب متقن تم الحكم عليه بأنه متقن).
- **False Positive (FP)**:حالات تم تصنيفها خطأً على أنها إيجابية.
- **True Negative (TN)**:حالات تم تصنيفها بشكل صحيح على أنها سلبية.

- **False Negative (FN)**:حالات تم تصنيفها خطأً على أنها سلبية : (Fawcett, 2006 863).

من خلال هذه القيم، يمكن اشتقاق مجموعة من المؤشرات التي تعزز من دقة الحكم على أفضلية التصنيف:

١.**الدقة (Accuracy)** :النسبة الكلية للحالات المصنفة بشكل صحيح. لكنها قد تكون مضللة إذا كانت البيانات غير متوازنة.(Sokolova & Lapalme, 2009: 131)

٢.**الحساسية (Sensitivity) أو(True Positive Rate)**:تقيس قدرة النموذج على اكتشاف الإيجابيات الحقيقية، وهي مؤشر مهم في حالة الرغبة في التقليل من الأخطاء من النوع الثاني (Powers, 2011: 40).

٣.**النوعية (Specificity)**: تقيس قدرة النموذج على التعرف على السلبيات الحقيقية، وهي مفيدة في حالات الرغبة في تجنب الأخطاء من النوع الأول.(Tharwat, 2021: 2)

٤.**الدقة التنبؤية الموجبة (Precision)**: تقيس عدد الإيجابيات المصنفة بشكل صحيح من بين كل الإيجابيات التي تم التنبؤ بها. وهي مهمة في حالات وجود تكلفة عالية للأخطاء الإيجابية الكاذبة.(Han et al., 2011: 155)

٥.**معدل الخطأ من النوع الأول (Type I Error) والنوع الثاني (Type II Error)**: يُستخدمان للحكم على مدى موثوقية النموذج في اتخاذ قرارات دقيقة دون الوقوع في الأخطاء الحاسمة.

وفيما يلي الجدول (٢) و الذي يوضح فيه اسم كل خلية والتي تستخدم لاستخراج كل مؤشر في مصفوفة الارتباك والتي تم ذكرها.

جدول (٢)

الجدول التقاطعي بين تصنيف المحكمين والتصنيف وفق درجة القطع بطريقة محددة

		وفق درجة قطع معينة	
		متقن	غير متقن
وفق آراء المحكمون	متقن	True Positive (TP)	False Negative (FN)
	غير متقن	False Positive (FP)	True Negative (TN)

(Altman & Bland, 1994: 404)

وقد أشار (Fawcett, 2006) إلى أن استخدام مصفوفة الارتباك يوفّر نظرة متعددة الأبعاد حول أداء النموذج بخلاف الاعتماد على مؤشر الدقة فقط، خاصةً عند العمل مع مجموعات بيانات غير متوازنة أو عندما يكون تصنيف الفئات مختلفًا في الأهمية.

كما شدد (Powers, 2011)على أن الحكم على أفضلية أي نموذج تصنيفي يجب أن لا يعتمد على مؤشر واحد، بل على تحليل متكامل لمصفوفة الارتباك بما يشمل مقاييس الحساسية والنوعية والدقة التنبؤية، لضمان تقويم موضوعي وشامل.

وتتكون معادلات مصفوفة الارتباك كما يلي :

$$\text{الدقة (Accuracy)} = \frac{TP+TN}{TP+TN+FP+FN}$$

حيث ترمز :

Accuracy:الى الدقة

TP: الى المتقن وصنف كمتقن

TN:الى غير المتقن وصنف كغير المتقن

FP:الى غير المتقن وصنف كمتقن

FN:الى المتقن وصنف كغير المتقن

$$\text{الحساسية (Sensitivity)} = \frac{TP}{TP + FN}$$

حيث ترمز :

Sensitivity: الى الحساسية

TP:الى المتقن وصنف كمتقن

FN:المتقن وصنف كغير المتقن

$$\text{النوعية (Specificity)} = \frac{TN}{TN+FP}$$

حيث ترمز:

Specificity: الى النوعية

TN:الى غير المتقن وصنف كغير المتقن

FP:الى غير المتقن وصنف كمتقن (Powers ,2011: 39-40)

خطأ القياس:

فى القياس النفسى، لكل درجة على مقياس ما تتكون من درجتين هما الدرجة الحقيقية والدرجة التي ترجع إلى الخطأ. والخطأ يعد أحد مكونات الدرجة التي يحصل عليها الفرد على الاختبار ، أى أن : الدرجة الكلية = الدرجة الحقيقية + الدرجة التي ترجع إلى الخطأ(عبد الرحمن ، 1998: ٧٣).

يعتمد مفهوم خطأ القياس على فكرة مفادها أنه يمكن التمييز بين المفهوم النظري وقياسه. وخطأ القياس هو أي انحراف عن القيمة الحقيقية للمفهوم في القياس. وبالتالي، فإن أي ملاحظة علمية تمثل مزيجًا من الدرجة الحقيقية للمفهوم والخطأ؛ ويُنظر إلى خطأ القياس على أنه الفرق بين القيمة الحقيقية والدرجة المرصودة لمفهوم ما(Weber, 2017 :2) .

الخطأ المرتبط بتصنيف الطلبة إلى متقنين وغير متقنين:

١.**الخطأ الموجب** (α) (Positive Error): ويعود ذلك الى امكانية تصنيف طلاب أو ممتحنين غير متقنين على أنهم متقنون، فى حالة احتمالية أن يقدم هذا الطالب غير المتقن استجابة صحيحة لفقرة ما (Rudner, 2001 :1) ويقسم عدد الطلبة الذين لم يكونوا متقنين فعلاً ولكن صُنِّفوا متقنين على مجموع عدد الذين هم فعلاً غير متقنين ويتم حسابه وفق المعادلة الاتية:

$$\alpha = \frac{False\ Positives\ (FP)}{\text{عدد غير المتقنين الحقيقيين}}$$

٢. **الخطأ السالب** (β) (Negative Error): ويعود ذلك الى امكانية تصنيف طلاب أو ممتحنين متقنين على أنهم غير متقنين، فى حالة احتمالية أن يقدم هذا الطالب المتقن استجابة غير صحيحة لفقرة ما (Rudner, 2001:1)، ويقسم عدد الطلبة الذين هم فعلاً متقنين ولكن صُنِّفوا غير متقنين على مجموع عدد الذين هم فعلاً متقنين ويتم حسابه وفق المعادلة الاتية:

$$\beta = \frac{False\ Negatives\ (FN)}{\text{عدد المتقنين الحقيقيين}}$$

وفى ضوء هذين النوعين، يتم تقسيم الممتحنين وفقاً لأي درجة قطع ومحك خارجي إلى فئات أربع:

(١)متقن حقيقى: وهو الشخص الذي يصنف على أنه متقن وفقاً لدرجة القطع المحددة للاختبار والمحك الخارجي.

(٢)متقن غير حقيقى: وهو الشخص الذي يصنف على أنه متقن وفقاً لدرجة القطع المحددة للاختبار ، بينما يصنف على أنه غير متقنين وفقا للمحك الخارجي.

(٣)غير متقن حقيقى: وهو الشخص الذي يصنف على أنه غير متقن وفقاً لدرجة القطع المحددة للاختبار والمحك الخارجى

(٤)غير متقن غير حقيقي: هو الشخص الذي يصنف على أنه غير متقن وفقاً لدرجة القطع المحددة للاختبار، بينما يصنف على أنه متقنين وفقا للمحك الخارجي (Rudner, 2001:1).

ويحسب احتمال التصنيف الصحيح المترتب على درجة القطع كالاتى:

احتمال التصنيف الصحيح = $1-\alpha-\beta$ (Berk , 1976: 5)

ثانيا: دراسات سابقة

أ.دراسات عربية:

١. .دراسة علام(1991) "**دراسة مقارنة لبعض طرق تحديد مستويات الاداء في الاختبار مرجعي المحك**"

هدفت الدراسة الى الموازنة بين أربع طرق لتحديد مستويات الأداء في الاختبار الذي أعدم الباحث اثنتان منها تعتمد على تحكيم محتوى الاختبار وفقراته وبذلك تعدان من الطرق المطلقة، والاثنتان الأخريان يعتمدان على تحكيم أداء المختبرين وجمع بيانات تجريبية وهي (طرق أنجوف، نيدلسكي المجموعة الحدية، المجموعة المتناقضة)، وتألفت عينة الدراسة من (٣٨٢) طالباً من طلاب كلية التربية ممن يدرسون مقرراً سيكولوجياً في الفروق الفردية، أما عينة المحكمين فتكونت من (٢٦) محكماً من ثلاث مجموعات متباينة في مستوى تمكنها في مجال القياس والتقويم التربوي والنفسي، والأولى (١٤) مدرساً من مدرسي المدارس الثانوية العامة الحاصلين على دبلوم خاص في التربية ودورة تدريبية في مجال القياس والتقويم، والثانية (٦) مدرسين مساعدين بدرجة دكتوراه في علم النفس التعليمي ولديهم خبرة (٣) أعوام في تدريس المقاييس والاختبارات و (٦) أساتذة في علم النفس التعليمي.وكانت من نتائج الدراسة الحصول على درجات قطع مختلفة نسبياً بمتوسطات (٤٠،٤٣،٤١،٤٤) لكل من طريقة نيدلسكي وأنجوف والمجموعة المتناقضة والمجموعة الحدية على التوالي.

٢.دراسة العنبكي (٢٠٠٩) بعنوان "المفاضلة في تحديد درجات القطع لاختبار محكي المرجع"

هدفت الدراسة إلى بناء اختبار تحصيلي محكي المرجع في مادة الاحصاء الوصفي والاستدلالي للمرحلة الثانية قسم العلوم التربوية والنفسية كليات التربية الجامعات العراقية ، وتحديد درجات القطع له بنوعيها (المعتمدة على المحكمين ، والمعتمدة على البيانات التجريبية) والمفاضلة بين درجات القطع بنوعيها اعلاه، وهذه الطرق هي:

(١)طريقة جيجر((Jaeger) (الطرق المعتمدة على المحكمين (الخبراء)
(٢)طريقة أنجوف (Angoff) (الطرق المعتمدة على المحكمين (الخبراء)
(٣)طريقة المجموعة الحدية(Borderline Group) (الطرق المعتمدة على البيانات التجريبية)
(٤)طريقة المجموعات المتناقضة(.(Contrasting Groups) (الطرق المعتمدة على البيانات التجريبية)

وطبّقت على عينة مكونة من (٢٠٨)طالبًا وطالبة. وبعد تحكيم الفقرات من قبل الخبراء واستبعاد غير المناسبة، استقر الاختبار على(٩٨) فقرة.

بلغت قيمه درجات القطع وفق الطرق جيجر(٥٩.٥) ، أنجوف(٣٦.٢٠)، المجموعة الحدية(٤٥)، المجموعة المتناقضة(٣٦.٢٥) ،كما تم تقييم الخصائص السيكومترية لكل طريقة، من حيث الصدق، والثبات، والخطأ المعياري، وجاءت النتائج بان أعلى صدق كانتلطريقة المجموعة الحدية(0.674) و أعلى ثباتلطريقة جيجر(0.943) أقل خطأ معياريلطريقة جيجر (3.075) ، وقد توصل الباحث إلى أن طريقة المجموعة الحدية هي الأنسب لتحديد درجة القطع في اختبار الإحصاء، لما تمتاز به من توازن بين الدقة والثبات.

ب- دراسات اجنبية:

٣.دراسة كنج (Chang, 1999) عنوان الدراسة "**تحليل تقديري للفقرات وفق طريقتي نيدليسكي وانجوف في تحديد المعايير**"

هدفت الدراسة المقارنة بين طريقتين (نيدلسكي، وأنجوف) لتحديد درجة القطع وذلك لمعرفة درجة القطع الأكثر ملاءمة لاختبار مقرر طرق أصول البحث الذي يدرس لطلبة شهادة الماجستير والدكتوراه في جامعة العاصمة الكبرى في الولايات المتحدة، من خلال اخذ فقرات الاختبار من نوع الاختيار المتعدد وتألفت عينة الدراسة من (٨٠) طالباً، و اسفرت الدراسة عن ان درجات القطع الناتجة عن طريقة نيدلسكي،اقل من درجات القطع الناتجة عن طريقة

انجوف.طريقة نيدلسكي اعطت اختلافات معيارية ، اكبر من طريقة انجوف .طريقة نيدلسكي اكثر ملاءمة للاختبارات الصعبة.

٤ .**دراسة كوكلوك تاشدمير (Çokluk & TaşdemiR, 2013)بعنوان**" تحليل دقة تصنيف اختبار باستخدام طريقتي (0-1) Angoff، Nedelsky، وتحليل ROC وتقدير الفواصل الزمنية".

هدفت الدراسة إلى مقارنة دقة تحديد درجة القطع لنتائج اختبار مكون من أسئلة في اللغة التركية والرياضيات، باستخدام طريقتين مبنيتين على رأي المحكّمين ((0-1) Angoff & Nedelsky)،تكونت عينة الدراسة من ٢١١ طالبًا في السنة الثالثة من برنامج تعليم الصفوف بجامعتي أنقرة و(Hacettepe). وقد تم اختيار العينة باستخدام أسلوب العينة الغرضية، حيث شملت فقط الطلاب الذين اجتازوا مادتي اللغة التركية والرياضيات. تم استخدام اختبار مكون من ٣٠ سؤالًا مختارًا عشوائيًا من أسئلة اللغة التركية والرياضيات المطبقة بين عامي(٢٠٠٦ و ٢٠١٢)كأداة لجمع البيانات في البحث ،تم حصول طريقة Angoff (١-٠) حددت أعلى درجات قطع لكنها أعطت أقل نسبة نجاح. طريقة Nedelsky أعطت درجات قطع أدنى ونسبة نجاح أعلى، وحققت توافقًا عاليًا بين المحكّمين.وخلصت النتائج إلى وجود فرق دال بين النسب المئوية للطلاب الذين تم اعتبارهم ناجحين. كما تبين أن اختبار الرياضيات يمكنه تصنيف الطلبة الناجحين / غير الناجحين بشكل صحيح. وقد وجد أن درجات القطع التي تم تحديدها وفقًا لتقدير الفاصل الزمني لاختبار الرياضيات واللغة التركية متوافقة مع درجات القطع التي تم تحديدها في تحليل الخاصية الاجرائية للمستقبل (receiver operating characteristic) analysis (ROC).وقد توصلت الباحثتان باستخدام أكثر من طريقة لتحديد درجات القطع لضمان الدقة في التصنيف التربوي.

الموازنة بين الدراسات السابقة و البحث الحالي :

سوف تقوم الباحثتان بالموازنة بين الدراسات السابقة وبين البحث الحالي في المحاور الاتية:

اولا :- **الاهداف** : سعت أغلب الدراسات إلى مقارنة بين طرق متعددة لتحديد درجة القطع. فمثلًا، هدفت دراسة **علام (١٩٩١)** التي هدفت إلى مقارنة أربع طرق لتحديد مستويات الأداء. في المقابل، ركزت دراسة **كنج (١٩٩٩)** تحليل مقارن لطريقتي نيدلسكي وأنجوف ، بينما هدفت دراسة **العنبكي (٢٠٠٩)** على بناء اختبار وتحديد درجة القطع بأربع طرق، مع تحليل الثبات والدقة . وركزت دراسة **كوكلوك تاشدمير (٢٠١٣)** فحص دقة تصنيف نجاح الطلبة وفق

لطريقتي انجوف(٠-١) وندلسكي.اما البحث الحالي فيهدف الى المفاضلة بين طريقتي نيدلسكي و ايبيل لتحديد درجة القطع من خلال تحليل الصدق و الثبات لكلا الطريقتين ، اما البحث الحالي فيهدف الى المفاضلة بين طريقتي نيدلسكي و ايبيل لتحديد درجة القطع لاختبار محكي المرجع في مادة الرياضيات .

ثانيا: نوع العينة وحجمها :- تباينت الدراسات في ذلك بشكل واضح. فبعضها استخدم طلابًا جامعيين فقط مثل دراسة **كنج(١٩٩٩)** (٨٠ طالبًا) ،دراسة **العنبكي(٢٠٠٩)** (٢٠٨ طالبًا وطالبة) ودراسة **كوكلوك و تاشتدمير(٢٠١٣)** (٢١١طالبا) ، في حين جمعت دراسات أخرى بين طلاب ومحكمين مثل دراسة **علام(١٩٩١)** التي شملت(٣٨٢)طالبًا و(٢٦)محكمًا، اما عينة البحث الحالي فيتكون من (٤٢٠) طالبا و طالبة من طلبة الصف السابع الاساسي.

ثالثا :- أدوات جمع المعلومات: اعتمدت جميع الدراسات على اختبارات مقننة أو مصممة خصيصًا. فمثلًا، استخدم دراسة **علام(١٩٩١)** اختبار مقرر الفروق الفردية، ودراسة **كنج(١٩٩٩)** استخدم اختبار اختيار من متعدد، و دراسة **كوكلوك و تاشتدمير(٢٠١٣)** استخدمت اختبار مكون من (٣٠) سؤلا تم اختيارها عشوائيا من اسئلة اللغة التركية والرياضيات . اما في البحث الحالي فسوف يتم بناء اختبار محكي في مادة الرياضيات وسوف يتم ذكر تفاصيل خطوات بناء الاختبار في فصل الاجراءات بالتفصيل.

رابعا:- الوسائل الإحصائية المستخدمة: استخدمت الدراسات عدة أساليب تحليلية متنوعة. فدراسة **كنج(١٩٩٩)** استخدمت تحليل الفروق والانحراف المعياري، بينما استخدمت **دراسة العنبكي(٢٠٠٩)** نماذج متقدمة مثل نموذج راش،(KR-20) ،والخطأ المعياري.و دراسة **علام(١٩٩١)** تحليل فروق الدرجات بين الطرق المستخدمة أما دراسات **كوكلوك و تاشتدمير(٢٠١٣)** استخدمت تحليل الخاصية الاجرائية للمستقبل(ROC) . اما في البحث الحالي فسيتم ذكر الوسائل الاحصائية المناسبة في فصل الاجراءات بالتفصيل.

خامسا:- نتائج الدراسات: أظهرت نتائج معظم الدراسات أن طريقة نيدلسكي تعطي عادةً أقل درجة قطع، كما في دراسة **كنج(١٩٩٩)** فقد اظهرت النتائج ان طريقة نيدلسكي أعطت درجات قطع أقل من أنجوف كما ان طريقة نيدلسكي كانت أكثر ملاءمة للاختبارات الصعبة بسبب اختلافاتها المعيارية العالية ،ودراسة **كوكلوك و تاشتدمير(٢٠١٣)** اظهرت ان طريقة انجوف(٠-١)أعطت أعلى درجة قطع مقارنة بطريقة نيدلسكي، ويشير ذلك إلى أن طريقة أنجوف (١-٠) قد تكون أكثر تحفظًا في تحديد درجة النجاح، فيما أظهرت **دراسة علام(١٩٩١)** تباينًا واضحًا

بين الطرق الأربع، حيث كانت أعلى درجة حسب طريقة نيدلسكي (٤٤) وأدناها حسب الطريقة الحدية (٤٠). أما دراسة **العنبكي**(٢٠٠٩) فقد بيّنت أن طريقة جيجر أعطت أعلى درجة قطع (٥٩.٥)، في حين أعطت أنجوف أقل درجة.(36.2) / اما فيما يتعلق بنتائج البحث الحالي فسوف تقوم الباحثتان بعرض النتائج بالتفصيل في فصل النتائج فيما يتعلق بالمفاضلة بين طريقتي تحديد درجة القطع .

ان ما يميز هذا البحث هو استخدام الباحثتان معيارا جديد لم تستخدم في اغلب الدراسات المحلية و العربية –حسب علم الباحثتان – وذلك للمفاضلة بين طريقتين لتحديد درجة القطع للاختبار المحكي المرجع وهذا المحك هو استخدام معادلة الارتباك (Confusion Matrix) والتي تم الاشارة اليها مفصلا في الاطار النظري وسيتم تناولها و تطبيق مؤشراتها في فصل الاجراءات وذلك لبيان المفاضلة بين طريقتي نيدلسكي و ايبيل لتحديد درجة القطع للاختبارالمعد في البحث الحالي و سوف يتم القاء الضوء على نتائجها و مخرجاتها في فصل النتائج بالتفصيل.

أوجه الاستفادة من عرض الدراسات السابقة :

١.ساعدت الدراسات السابقة في تدعيم الإطار النظري وتوضيح المفاهيم الأساسية مثل درجة القطع، الأخطاء التصنيفية، ونماذج الاختبارات المحكية.ومن حيث العينية، فقد افادت الباحثة من عرض الدراسات السابقة فى تحديد نوع العينة (طلبة السابع للمرحلة الاساسية.

٢.ساعدت الدراسات السابقة على توضيح الدور الحاسم لدرجة القطع في الحكم على مستوى إتقان المتعلم، وأبرزت ضرورة تحديدها وفق أسس علمية دقيقة.

٣.أوضحت الدراسات مزايا كل طريقة من طريقتي نيدلسكي و ايبيل من حيث الدقة والمرونة، وكذلك كشفت عن بعض التحديات المرتبطة بتطبيقها، الأمر الذي أسهم في تعميق فهم الباحثتان للفروق بين الطريقتين.

٤.استفادت الباحثتان من الأدوات المستخدمة في الدراسات السابقة، وخاصة فيما يتعلق بقياس صدق وثبات الاختبار، وتحليل نتائج درجات القطع، مما ساعد في تطوير منهجية البحث.

٥.ومن حيث الوسائل الاحصائية المستخدمة كمحكات المفاضلة بين الطريقتين، فقد حددت الباحثة تلك الوسائل التى تتناسب مع بحثها، متمثلة فى معامل فاي (Ø) لمعرفة الخصائص السيكومترية لدرجات القطع كصدق درجات القطع وفق طريقتي نيدليسكى وإيبل، و استخدام معامل ليفنجستون لحساب معامل ثبات درجات القطع بكلا الطريقتين.

اجراءات البحث

يتناول هذا الجزء جراءات البحث والتي يتم عرضها بشكل متتابع، حيث يُبيّن منهجية البحث والتي تتحدد بالمنهج الوصفي المقارن كأساس لمفاضلة طرائق تحديد درجة القطع، ثم يعرّف مجتمع الدراسة وعينته وفق اجراءات البحث، بعد ذلك يُستعرض اختبار محكي المرجع في مادة الرياضيات كأداة رئيسة، موضحًا مراحل بنائه من إعداد المواصفات التفصيلية وصياغة الفقرات إلى التحقق من صدقها وثباتها باستخدام عدة طرق ، ويتم عرض كيفية تحديد درجة القطع للاختبار بطريقتي نيدلسكس و ايبيل، ويتبع ذلك وصف إجراءات جمع البيانات الميدانية وتطبيق الاختبار على افراد عينة البحث، و كذلك تُعرض خطوات التحقق من الخصائص السيكومترية لدرجة القطع للاختبار بكلا الطريقتين المذكورتين واجراءاتها، وفي نهاية الفصل تُعرَض الوسائل الإحصائية المعتمدة لتحليل البيانات . وسيتم عرض الموضوعات في هذا الفصل وفق التسلسل الاتي:

اولاً:-منهجية البحث: تم استخدام المنهج الوصفي المقارن لأنها تسعى الى وصف طرق تحديد درجة القطع المختلفة و ركزت على مقارنة درجات القطع للاختبار والناتجة عن تطبيق اختبار محكي المرجع باستخدام طريقتين مختلفتين لتحديد درجة القطع و المفاضلة بينهما (علام، ٢٠٠٥: ١٢٨).

ثانياً:-مجتمع البحث: يتحدد مجتمع البحث بطلبة المدارس الاساسية في الصف السابع الأساسي ومن كلا الجنسين الذكور والاناث و بلغ حجم المجتمع (١٦٩٦٥) طالب وطالبة وبواقع (٨٧١٢) من الذكور و (٨٢٥٣) من الاناث بنسبة (٥١%) من الذكور و(٤٩%) من الاناث، وكذلك من المتخصصين في الرياضيات من المدرسين في مدارس محافظة اربيل/ في اقليم كوردستان العراق

ثالثاً:-عينة البحث:تالفت عينة البحث الحالي من (٤٤٠) فردا من طلبة الصف السابع في المرحلة الاساسية اختيرت بالاسلوب العشوائي من المدارس الاساسية لكلا الجنسين في اقليم كوردستان العراق للعام الدراسي (٢٠٢٤/٢٠٢٥)، حيث كان الصف الدراسي(السابع) هو وحدة الاختيار ، وقد تم اختيار خمسة مدارس بالاسلوب العشوائي من بين المدارس التي تحوي الصف السابع. وقد بلغ العدد الكلي للطلبة (٤٤٠) طالبا وطالبة وبواقع (٢١٠) طالب و (٢١٠) طالبة،

وصف عينات البحث و اعدادها وفق مراحل اجراءات البحث : يتكون اعداد العينات للبحث الحالي وفق ما هو مبين في الجدول(٤).

الجدول(٤):وصف عينات البحث موزعة حسب اجراءات بناء الاختبار المحكي المرجع في مادة الرياضيات و اجراءات المفاضلة بين طريقتي نيدلسكي و ايبيل لتحديد درجة القطع للاختبار

وصف العينة	عدد افراد العينة	الغرض من التطبيق
عينة تجريب الاختبار	٣٠	للتأكد من وضوح فقرات الاختبار و معرفة الوقت المستغرق
عينة ثبات الاختبار	٦٠	للتحقق من ثبات الاختبار باستخدام معادلة كودر-ريتشاردسون (٢٠) ، و كذلك باستخدام معادلة ليفنجستون
عينة التحليل الاحصائي للاختبار	١٨٠	للتحقق من صدق الفقرات و ثباتها و تمييزها و استخراج معاملات الصعوبة لها
عينة التحقق من ثبات التصنيف للاختبار	٧٠	للتحقق من ثبات تصنيف المتقنين وغير المتقنين على الاختبار في مرتي تطبيق الاختبار باستخدام معامل كابا
عينة المفاضلة بين طريقتي درجة القطع	١٠٠	بهدف التحقق من الخصائص السيكومترية لدرجة القطع بطريقتي نيدلسكي و ايبيل
المجموع الكلي	٤٤٠	

رابعاً:-أداة البحث:

تتكون اداة البحث الحالي من اختبار محكي المرجع في مادة الرياضيات ، و قد قامت الباحثتان ببناء اختبار محكي المرجع بالاعتماد على تحليل كتاب الرياضيات والتي تدرس في الصف السابع في المرحلة الاساسية في مدارس اقليم كوردستان العراق ، وقد قامت الباحثتان باتباع الخطوات الاتية لبناء الاختبار في الرياضيات :

أ-تحليل محتوى الفصول الثلاثة الاولى في كتاب الرياضيات :

لتحقيق أهداف هذه البحث قامت الباحثتان بتحليل محتوى كتاب الرياضيات والتي تدرس في الصف السابع الاساسي في اقليم كوردستان العراق وتحديدا للفصول (الاول ، الثاني الثالث) تحليلا دقيقا بالاستعانة بدليل المدرس ، والمنهاج ، و مدرسي و مدرسات المادة بهدف تحديد

الاهداف السلوكية و المهارات الرياضية بشكل دقيق وذلك بهدف بناء اختبار محكي المرجع للفصول الثلاثة المذورة والتي تتكون من الموضوعات الاتية:

الفصل الاول: نظرية الأعداد– المواضيع: (الحساب الذهني، القوى، النظام المتري، الصور العلمية، تراتب العمليات، الضرب و القسمة، القوى والجذور، قابلية القسمة، تحليل الأعداد، القاسم المشترك الأكبر، المضاعف المشترك الأصغر،24سؤالا).

الفصل الثاني: مدخل إلى الجبر–المواضيع (المقدار الجبري، القوى ،بين النص والمقدار، الحدود المتشابهة ، جمع المقادير الجبرية ، ضرب عدد في مقدار جبري ، المعادلات وحلولها، تمثيل المعادلات، حل المعادلات بالجمع والطرح ، حل المعادلات بالضرب و القسمة،22سؤالاً).

الفصل الثالث: الأعداد الصحيحة والأعداد النسبية– المواضيع: (الأعداد الصحيحة، المستوى الاحداثي، جمع الأعداد الصحيحة، طرح الأعداد الصحيحة، ضرب الأعداد الصحيحة وقسمتها، الأعداد النسبية، مقارنة الأعداد النسبية، العمليات على الأعداد النسبية، المعادلات والأعداد النسبية،24سؤالا).

ب–بناء جدول المواصفات (الخارطة الاختبارية) لمادة الرياضيات:

يُعد تصميم جدول المواصفات (الخارطة الاختبارية) من الخطوات الجوهرية في إعداد الاختبار، إذ يساعد على تنظيم العمل بشكل منهجي، ويُسهم في ضمان شمول الفقرات لمحتوى المادة التعليمية بشكل متوازن. كما يُراعي هذا التصميم توزيع الفقرات على المستويات المعرفية المختلفة، مما يُعزز من عدالة القياس ، وقد قامت الباحثتان بتصميم جدول المواصفات (الخارطة الاختبارية) لكتاب الرياضيات التي تدرس في الصف السابع الاساسي وكما هو مبين في الجدول(٥).

الجدول (٥): جدول مواصفات الاختبار (الخارطة الاختبارية) لمادة الرياضيات للصف السابع الاساسي في مبيناً فيه عدد الفقرات و الاهمية النسبية لكل فصل

الاهداف التعليمية / الفصل (المحتوى)	عدد الاهداف السلوكية	الاهمية النسبية للفصل (المحتوى)	الاهداف التعليمية				المجموع
			التذكر %20	الفهم %30	التطبيق %40	التحليل %10	
الفصل الاول	24	%34	5	7	10	2	24
الفصل الثاني	22	%31	4	7	9	2	22
الفصل الثالث	24	%35	5	7	10	2	24
المجمـــوع	70	%100	14	21	29	6	٧٠

ووفقاً لقائمة الأهداف التي أعدتها الباحثتان الخاصة بالفصول(الاول، الثانى، الثالث) المذكورة إذ قامتا بتحليل الفصول تحليلاً دقيقاً وقد تم اعداد قائمة مكونة من سبعين هدفاً تفصيلياً شاملة لموضوعات الفصول الثلاثة

ج-إعداد المواصفات التفصيلية للاختبار المحكي المرجع في مادة الرياضيات:

وفي هذه الخطوة قامت الباحثتان بإعداد المواصفات التفصيلية للاختبار باللغة الكوردية بالاعتماد على تحليل كتاب الرياضيات الذي يدرس في الصف السابع للمرحلة الاساسية وتحديدا للفصول الثلاثة الاولى، ويُقصد بالمواصفات التفصيلية للاختبار تلك القواعد التفصيلية الدقيقة التي ترشد مُعِدّ فقرات الاختبار وتمكنه من صياغة فقرات متكافئة تقيس الهدف نفسه(فتح الله، ٢٠٠٥ : ١٥٨)، وذلك استرشاداً بالمكونات التي حددها بوفام (Popham , 1975 , 141-146): وهي الوصف العام للهدف، وعينة من الفقرات ، وعناصر المثيرات، وعناصر الاستجابات ، ثم قامت الباحثتان ببناء الفقرات لكل هدف سلوكي بشكل منفصل، إذ تقابل كل هدف فقرة (سؤال) من نوع الاختيار المتعدد ولكل فقرة اربعة بدائل واحدة من هذه البدائل تمثل الاجابة الصحيحة و البدائل المتبقية تُعد خاطئة، وتم عرض هذه المواصفات والاختبار على عينة المحكمين، وقد أعدت الباحثتان استبانة للحكم على الاختبار من حيث الاتى :

١ – مدى ارتباط السؤال بالهدف الذي تقيسه من حيث المحتوى والمستوى المعرفي.

٢ – هل الفقرة مصاغة بلغة واضحة وسليمة ومفهومة .

٣ – هل الفقرة مستقلة عن غيرها من فقرات الاختبار.

خامساً:-التحقق من الخصائص السيكومترية للاختبار

١- صدق الاختبار: (Test Validity)

أ-الصدق الوصفي(Descriptive Validity):

تم عرض الاختبار على (٢٢) محكماً وخبيراً في تخصص العلوم التربوية و النفسية ، وطلب منهم تحكيم فقرات الاختبار و بدائله ومدى تطابق الفقرات مع الهدف و المستوى الذي تقيسه الفقرة في هرم بلوم المعرفي، و اجراء التعديلات المناسبة و ابداء آرائهم عليها، ومن خلال ملاحظاتهم و آرائهم قامت الباحثتان بإجراء بعض التعديلات الطفيفة على بعض الفقرات وبعض البدائل، وقد وافق الخبراء على جميع الفقرات في الاختبار، لذا اعتبرت الباحثتان أن ملاحظات المحكمين والقيام بالتعديل المطلوب تُعد صدقًا ظاهريًا لأداة الدراسة، وبناءً على ذلك اعتُبِر أن الاختبار صالح وملائم لقياس ما وضع له و يمكن تطبيقه على عينة التحليل الاحصائي.

ب-صدق المحتوى(Content Validity):

تم التحقق من صدق المحتوى عندما قامت الباحثتان ببناء جدول المواصفات للفصول الثلاثة الاولى لكتاب الرياضيات و التي تدرس في المرحلة السابعة في المدارس الاساسية وقد تم ذكر الاجراء بالتفصيل، وتم تصميم جدول المواصفات (الخارطة الاختبارية) وكما هو مبين في الجدول (٥)

ج-صدق البناء(صدق الاتساق الداخلي للاختبار(Construct Validity):

تم التأكد من صدق الاختبار من خلال الأسلوب التجريبي، إذ جرى تطبيقه على عينة التحليل الإحصائي المكوّنة من (١٨٠) طالبًا وطالبة من طلبة المدارس الأساسية. وقد تم اعتماد معامل الارتباط الثنائي المتسلسل لحساب العلاقة بين درجات كل فقرة ودرجة الفرد الكلية في الاختبار، وقد كانت قيم معامل الارتباط للفقرات الاختبارية تراوحت بين (٠.٢٠٣ – ٠.٦٦) وهي قييم دالة عند مستويي الدلالة (٠.٠١) مما يدل على تحقيق دلالة صدق البناء لفقرات الاختبار.

٢-:- ثبات الاختبار(Reliability Test) :

يُعد الثبات أحد أهم الخصائص السيكومترية التي يُبنى عليها الحكم على جودة الاختبارات، وقد تم التحقق من ثبات الاختبار المحكي في هذا البحث باستخدام أكثر من أسلوب، منها طريقة كودر-ريتشاردسون (KR-20) لقياس الاتساق الداخلي، بالإضافة إلى استخدام معامل ليفنجستون لقياس دقة القرار عند درجة القطع، وكذلك استخدام معادلة كابا لقياس اتساق قرار

تصنيف الطلبة الى متقنين و غير متقنين ، وفيما يلي ذكر الطرق المستخدمة للتحقق من ثبات الاختبار:

أ-استخدام معادلة كودر- ريتشاردسون(٢٠):

للتحقق من ثبات الاختبار، قامت الباحثتان باستخدام معادلة كودر-ريتشاردسون(٢٠) وصيغة معادلتها كالاتي وذلك للتحقق من ثبات الاختبار :

$$KR-20=\frac{k}{k-1}\left(1-\frac{\sum piqi}{\delta^2}\right)$$

(Kuder & Richardson,1973:151)

و تستخدم المعادلة في حساب ثبات الاختبارات ثنائية التدريج(٠،١) و بالاعتماد على القيم المذكورة في الجدول(٦)بلغت قيمة الثبات (٠.٧٣).

الجدول(٦):القيم المستخدمة لاستخراج قيمة معامل الثبات للاختبار المحكي المرجع في الرياضيات من خلال استخدام معادلة كودر-ريتشارسون(٢٠) و باستخدام درجة قطع(٣٠) وبطريقة نيدلسكي

المتوسط الحسابي	التباين الكلي على الاختبار	عدد الافراد	درجة القطع بطريقة نيدلسكي		قيمة معادلة(كيودر-ريتشاردسون (٢٠):	قيمة ثبات الاختبار باستخدام معادلة ثبات القرار لمعامل ليفنجستون باستخدام درجة قطع معينة
			الدرجة	النسبة		
٢٨.٧٢	٢٢.٥	٧٠	٣٠	٠.٤٣	٠.٧٣	٠.٧٤٨

وبملاحظة الجدول(٦)نجد ان قيمة معامل ليفنجستون تبلغ (٠.٧٤٨)، وتشير هذه النتيجة الى أن الاختبار يُظهر مستوى جيدًا من الثبات في اتخاذ القرار عند درجة القطع المحددة (٣٠).

ب-معامل ثبات ليفنجستون(Livingston Index):

تم استخدام معادلة ليفنجستون الاساسية لمعرفة ثبات القرار بالاعتماد على درجة قطع معينة درجة قطع معينة، وقد أظهرت النتائج كما مبينة في الجدول(5)على أن قيمة معامل ليفنجستون بلغت (٠.٧٤) وهي تُعد قيمة جيدة ومؤشراً على ثبات الاختبار.

ج – استخدام معامل كابا(Kappa) :

لغرض التحقق من اتساق تصنيف الطلبة إلى متمكنين أو غير متمكنين تم اختيار عينة لغرض تطبيق الاختبار عليهم ، فقد تم إعادة تطبيق الاختبار على عينة مكونة من (٧٠) طالب و طالبة بعد مرور أسبوعين على التطبيق ، ولتقدير الثبات قامت الباحثتان بمقارنة نسبة عدد الأفراد الذين يتم تصنيفهم على أنهم متمكنين في مرتي التطبيق ،عند درجة قطع محددة بطريقة نيدلسكي والتي تكونت من نسبة (٠.٤٣) من فقرات الاختبار التي تفترض ان يجاوب عليها الطالب الذي يمتلك الحد الادنى من الكفاية والتي كانت مساوية للدرجة (٣٠)، وذلك باستخدام معامل كابا (Kappa Coefficient)، كونه الأسلوب الإحصائي الذي يأخذ أخطاء التصنيف بعين الاعتبار، إذ كلما زاد اتساق قرار التصنيف في مرتي التطبيق دل ذلك على ارتفاع درجة الثبات للاختبار (علام ، ٢٠٠٠: ١٢٢) ، وقد تم تطبيق الاختبار مرتين على العينة مرتين وكانت النتائج كما مبينة في الجدول (٧).

الجدول (٧): يبين تصنيف الطلبة وفقا لدرجة القطع المستخدم بطريقة نيدلسكي(٣٠ درجة) على الاختبار والتي تمثل نسبة (٠.٤٣) من فقرات الاختبار الكلية لتحديد معامل ثبات التصنيف

باستخدام معامل كابا

التطبيق(٢) / التطبيق (١)	متقن	غير متقن
متقن	٣٩	٠
غير متقن	٩	٢٢

وقد تم ايجاد النسب المئوية الهامشية للتكرارات بالاعتماد على البيانات في الجدول (٧) وكما هو مبين في الجدول (٨).

الجدول (٨) النسب المئوية الهامشية لتكرارات الطلبة

التطبيق(٢) / التطبيق (١)		المجموعات		النسب الهامشية
		متقنة	غير متقنة	
متقنة	متقنة	٠.٥٦	٠.٠٠٠	٠.٥٦

	غير متقنة	٠.١٣	٠.٣١	٠.٤٤
النسب الهامشية		٠.٦٩	٠.٣١	

وبالاعتماد على البيانات المدونة في الجدول (٨) بلغت قيمة معامل ثبات كابا لتصنيف المتقنين وغير المتقنين على الاختبار في مرتي تطبيق الاختبار (٠.٧٣) وهي قيمة ثبات جيدة و حقيقية لذلك يمكن القول بان الاختبار يتمتع بدرجة جيدة من ثبات التصنيف.

سادساً:- تحديد درجة القطع للاختبار المحكي المرجع: تُعد درجة القطع عنصرًا جوهريًا في بناء الاختبارات المحكية المرجع، إذ تمثل الحد الذي يُقارن به أداء الفرد لتحديد مدى إتقانه للمهارات أو المعارف المطلوبة وقد تم اتباع ما يلي لتحديد درجة القطع بكلا الطريقتين ووفق ما يلي:

اولا:-طريقة نيدلسكي(Nedlsky's Method) :

تم عرض الاختبار وفقراته على (٢٢) خبيرا طُلب منهم ان يحددوا عدد البدائل التي يستبعدها الطالب الذي يمتلك الاحد الادنى من الكفاية للاجابة على كل فقرة ، وبعد القيام بالاجراءات الخاصة بهذه الطريقة كانت النتائج كما مبينةفي الجدول (٩).

الجدول (٩): تحديد درجة القطع باستخدام طريقة نيدلسكي(Nedelsky)(N=22)

المحكم	مجموع تقدير المحكم على جميع الفقرات	المحكم	مجموع تقدير المحكم على جميع الفقرات
١	37.14	١٢	47.86
٢	36.62	١٣	49.36
٣	30.16	١٤	31.94
٤	31.97	١٥	40.71
٥	37.57	١٦	39.23
٦	47.31	١٧	54.79
٧	41.48	١٨	40.91
٨	38.86	١٩	38.16
٩	53.08	٢٠	41.99

١٠	57.56	٢١	47.51
١١	49.45	٢٢	52.15
مجموع الدرجات لكل المحكمين مقسوما على عدد الفقرات = ٩٤٥.٨١ ÷ ٢٢ × ١٠٠ = ٤٢.٤٣ و بالتقريب يساوي (٠.٤٣)، معدل درجات المحكمين= ٤٣%			

ثانياً:-طريقة إيبل(Ebel's Method): تم عرض فقرات الاختبار على (٢٢) خبيرا طُلب منهم ان يقيموا الفقرة من حيث اهميتها و مدى سهولتها ، وبعد القيام بالاجراءات الخاصة بهذه الطريقة كانت النتائج كما مبينةفي الجدول (١٠).

الجدول (١٠) :حساب درجة القطع باستخدام طريقة إيبل

اهمية الفقرة	عدد الفقرات	نسبة توقع النجاح	حاصل ضرب عدد الاسئلة في توقع النجاح
مهمة جدا (١٥ فقرة)			
سهلة	–	–	–
متوسطة الصعوبة	٩	٦٠	٥٤٠
صعبة	٦	٥٠	٣٠٠
مهمة (٢١ فقرة)			
سهلة	١٠	١٠٠	١٠٠٠
متوسطة الصعوبة	١١	٧٠	٧٧٠
صعبة	–	–	ـ
مقبولة(١٩ فقرة)			
سهلة	١٠	١٠٠	١٠٠٠
متوسطة الصعوبة	٩	٦٠	٥٤٠
صعبة	–	–	
غير مهمة (١٥ فقرة)			
سهلة			
متوسطة الصعوبة	٨	٦٠	٤٨٠
صعبة	٧	٤٠	٢٨٠
	٧٠ فقرة	٧٠.١٤	تمثل درجة القطع على

اهمية الفقرة	عدد الفقرات	نسبة توقع النجاح	حاصل ضرب عدد الاسئلة في توقع النجاح
	٤٩١٠	٠.٧٠*٧٠= ٤٩ درجة	الاختبار

سابعا:-استخراج الخصائص السيكومترية لفقرات الاختبا المحكي المرجع:

أ-:صدق فقرات الاختبار :

للتحقق من صدق فقرات الاختبار تم استخدام معادلة فاي وكانت النتائج كما مبينة في الجدول () ، وقد تراوحت قيم معامل فاي للفقرات بين قيمتي (٠.٣١ – ٠.٦٧) وهي تعد قيم مقبولة وجيدة كمؤشر لصدق الفقرات .

ب-: حساسية فقرات الاختبار :

قامت الباحثتان بالتحقق من خاصية حساسية فقرات الاختبار من خلال استخدام مؤشر برينان (Brennan-Index) وكانت القيم تتراوح بين (٠.٢٠٣ – ٠.٦٦) وهي قيم مقبولة والنتائج كما مبينة في الجدول(١٧).

ج- : صعوبة فقرات الاختبار :

تم التحقق من صعوبة فقرات الاختبار فكانت قيم معامل الصعوبة تتراوح بين(٠.٤٩ – ٠.٦٩)وتعد القييم مقبولة

ثامناً:-التحقق من الخصائص السيكومترية لدرجات القطع

١-معامل صدق درجة القطع:

بما إن الاختبارات المحكية تحتاج إلى تحديد درجة قطع ليتم من خلالها التمييز بين المتقنين وغير المتقنين ، أذن لابد من ان تتوافر في هذه الدرجةخاصية الصدق ، وبعد ان تم حساب درجات القطع بإتباع طرقتى نيدلسكي وإيبل ،فلابد من التأكد من صدقها والذي يعني بمدى نجاح هذه الدرجة في التمييز بين المتقنين وغير المتقنين (عبد السلام ، 1992 أ ، :91)، ويشير علام (٢٠٠٠) إلى إن معامل صدق درجات القطع يستخرج من خلال معامل ارتباط فاي (Ø) (علام، ٢٠٠٠:٣٢٧).

ولغرض حساب صدق درجات القطع وفق الطريقتين المعتمدتين، تم اتباع الإجراءات الآتية:

مجلة مركز بابل للدراسات الإنسانية ٢٠٢٦ المجلد ١٦ / العدد ٤

١.تم اختيار عشرة محكمين من مدرسي مادة الرياضيات، وطُلب من كل محكم تصنيف عشرة طلاب تم اختيارهم **عشوائيًا** من بين طلبته الفعليين، بحيث يُحدد ما إذا كان الطالب "متقنًا" أو "غير متقن"، بناءً على تقديره المهني.

٢.بعد ذلك، تم تطبيق الاختبار التحصيلي في مادة الرياضيات للصف السابع الأساسي، والذي يتكوّن من (٧٠) فقرة من نوع الاختيار من متعدد، وتم إعداد فقراته وتحليلها وفق اجراءات بناء الاختبار المحكي واستخدام معادلات الخاصة ببناء الاختبارات المحكية وتم تحديد درجة القطع له مرة باتباع اجراءات طريقة :نيدلسكي ومرة اخرى باتباع اجراءات طريقة **إيبيل**.

تم تطبيق الاختبار على عينة مكوّنة من (١٠٠) طالب وطالبة، ثم تصنيفهم إلى فئتين (متقن / غير متقن) بناءً على درجة القطع المستخلصة من كل طريقة.

وبعد تطبيق الاختبار على عينة البحث المكونة من (١٠٠) طالب وطالبة، نتج عن ذلك متغيران متقطعان، يمثلان أساس المقارنة في تحليل صدق درجة القطع و هما:

١.متغير تصنيف المحكمين، حيث قام المحكمون بتصنيف الطلبة إلى فئتين هما: "متقن" و"غير متقن"، بناءً على حكمهم التربوي والخبرة الصفية المباشرة.

٢.متغير التصنيف الناتج عن تطبيق درجة القطع المستخرجة وفقًا لطريقة نيدلسكي أو إيبل، والذي صنّف الطلبة أيضًا إلى "متقنين" و"غير متقنين"، بناءً على أدائهم الفعلي في الاختبار".

وتُعرض النتائج التفصيلية للتقاطع بين تصنيف المحكمين و درجة القطع في الجدولين (١٢) و(١٣) في أدناه ووفقا لطريقتي نيدلسكي و ايبيل على التوالي.

جدول (١٢)الجدول التقاطعي بين تصنيف المحكمين والتصنيف وفق درجة القطع بطريقة نيدلسكي

		وفق درجة القطع بطريقة نيدلسكي		المجموع
		متقن	غير متقن	
تقدير المحكمون	متقن	True Positive (TP) ٢١	False Negative (FN) ٢٤	٤٥

	غير متقن	False Positive (FP) ٣٥	True Negative (TN) ٢٠	٥٥
المجموع	٥٦		٤٤	١٠٠

الجدول (١٣) الجدول التقاطعي بين تصنيف المحكمين والتصنيف وفق درجة القطع بطريقة ايبيل

		وفق درجة القطع بطريقة ايبيل		المجموع
		متقن	غير متقن	
وفق آراء المحكمين	متقن	True Positive (TP) ٤٧	False Negative (FN) ٤	51
	غير متقن	False Positive (FP) ٧	True Negative (TN) ٤٢	49
المجموع		54	46	100

وقد تم استخدم معامل فاي (Ø) لمعرفة صدق درجات القطع وفق طريقتي نيدلسكى وإيبل على وفق القييم و البيانات المذكورة في الجدولين (١٣) ،(١٤) ،والجدول (١٥) يبين ذلك.

الجدول (١٥) مقارنة بين طريقتي نيدلسكي وإيبيل في تحديد درجة القطع ومعامل صدقها باستخدام معامل فاي(φ)

م	الطريقة	نسبة درجة القطع من الدرجة الكلية على الاختبار	درجة القطع	معامل صدق درجة القطع باستخدام معامل فاي
١	نيدلسكى	٠.٤٣	٣٠	٠.١٧
٢	إيبل	٠.٧٠	٤٩	٠.٧٨

٢-معامل ثبات القرار لدرجة القطع:

يُعدّ معامل الثبات أحد المؤشرات الأساسية في جودة القياس، إذ يُعبر عن استقرار أداء الفقرة في قياس السمة المستهدفة دون تأثر بالعوامل العارضة .ويؤكد المتخصصون في القياس النفسي على أهمية تحقق الثبات في فقرات الاختبارات النفسية والتحصيلية لضمان الاتساق والدقة في التقدير، كما أشار إلى ذلك السيد (السيد، ١٩٧٩،: ٦٥٠).

واستنادًا إلى هذا المبدأ، اعتمدت الباحثتان معادلة ليفنجستون (Livingston Index, 1972) لاشتقاق معامل الثبات في سياق الاختبارات المرجعية إلى المحك (CRT)، حيث ترتكز هذه المعادلة على مفهوم درجة القطع التي تُعدّ حدًا فاصلاً بين الأفراد المتقنين وغير المتقنين، وتم التعبير عن هذا المعامل بالرمز (rent). تم تطبيق معادلة ليفينجستون لدرجات القطع وفق طريقتى(نيدلسكى وإيبل)،كما تم استخراج قيمة الخطأ المعياري وفق قيمة معامل الثبات لكل درجة قطع ،إذ يُعد الخطأ المعياري لدرجة القطع من المؤشرات المهمة التي تُظهر مدى حساسية قرار الإتقان للتقلبات المحتملة، وبالتالي تساعد على الحكم على دقة التصنيف (Livingston & Zieky, 1982: 24)، فكانت قيمة معامل ثبات القرار لدرجة القطع و قيمة الخطأ المعياري له كما مبين في الجدول(١٦).

الجدول (١٦):قيم معامل الثبات باستخدام معادلة ليفينجستون لدرجات القطع وفق طريقتى نيدلسكى وإيبيل وقيم الخطأ المعياري المقابل لكل قيمة ثبات و قيمة نسبة الخطأ المعياري المرتبط بدرجة القرار(درجة القطع)♣

الطريقة	درجة القطع		قيمة التباين الكلي للاختبار	قيمة معامل الثبات للاختبار باستخدام معادلة كودر – ريتشاردسون(٢٠)	معامل ثبات ليفينجستون	قيمة الخطأ المعياري لمعامل الثبات	نسبه الخطأ المعياري المرتبط بدرجة القرار(درجة المحك)
	النسبة	الدرجة					
نيدلسكى	٠.٤٣	٣٠	٢٢.٥	٠.٧٣	٠.٧٤	٢.٤٦٤	%٨.٢١
إيبل	٠.٧٠	٤٩			٠.٩٨	٢.٤٦٤	%٥.٠٣

♣ تم استخدام معادلة الخطأ المعياري $SE = SD\sqrt{1 - Relibility}$، علماً ان قيمة الانحراف المعياري (SD) لدرجات استجابات الافراد على الاختبار تساوي (٤.٧٣٤). وقيمة الثبات للاختبار باستخدام معادلة كودر-ريتشاردسون(٢٠) = ٠.٧٣ .

وكما هو مبين في نتائج الجدول(١٦) فقد تم حساب الخطأ المعياري للتقدير (SEM) للاختبار باستخدام التباين الكلي (٢٢.٥) ومعامل الثبات(KR−20 = 0.73) ، وبلغت قيمته 2.46. ، وبحساب نسبة الخطأ المعياري إلى درجة القطع، تبين انه عند درجة القطع بطريقة نيدلسكي (٣٠ من ٧٠)، بلغت نسبة الخطأ المعياري ٨.٢١%، وانه عند درجة القطع بطريقة إيبيل (٤٩ من ٧٠)، بلغت النسبة٥.٠٣% ، ويشير ذلك إلى أن القرار المتخذ عند درجة القطع بطريقة إيبيل أكثر استقرارًا وأقل تأثرًا بالخطأ المعياري مقارنة بطريقة نيدلسكي، ويرجع ذلك إلى بُعد درجة القطع عن المتوسط (٢٨.٧٢) مما يقلل من احتمالية تداخل الأداء حول نقطة القرار.

وهذا يشير إلى أن طريقة إيبيل توفر نتائج أكثر دقة واستقرارًا في تقدير معامل الثبات، مما يجعلها خيارًا مفضلًا في هذا السياق.

الوسائل الاحصائية :

اعتمدت الباحثتان في تحليل البيانات على مجموعة من الأساليب الإحصائية، التي تم تطبيقها باستخدام برنامج الحزمة الإحصائية للعلوم الاجتماعية (SPSS –v.24) ، بالإضافة إلى استخدام برمجية (Excel)، وذلك لتحقيق اهداف البحث، وتمثلت هذه الأساليب فيما يأتي:

١ – **اختبار كاي لحسن المطابقة** وذلك لمعرفة الفروق بين آراء الموافقين وغير الموافقين على فقرات الاختبار المحكي المرجع وذلك للتحقق من الصدق الوصفي للاختبار.

٢ – **حساب متوسط تقديرات** المحكمين على كل فقرة من فقرات الاختبار بعد استبعاد البدائل التي يستبعدها الطالب الذي يمتلك الكفاية الادنى من المهارة، بغية تحديد درجة القطع وفق طريقة نيدلسكي.

٣ – **استخراج نسبة الإجابات الصحيحة لكل فقرة**، بهدف التوصل إلى معاملات الصعوبة الخاصة بفقرات الاختبار.

٤ – **معامل الارتباط الثنائي المتسلسل** بهدف قياس صدق الاتساق الداخلي، وذلك من خلال معرفة العلاقة بين درجات الأفراد على الفقرات ودرجاتهم الكلية في الاختبار.

٥ – **معادلة كودر – ريتشاردسون ٢٠** (KR−20) وذلك لاحتساب معامل الثبات للاختبار.

٦ – **معادلة ليفنجستون** وذلك لحساب ما يلي:

أ – معادلة ليفنجستون الاساسية للتحقق من ثبات الاختبار المحكي المرجع بعد بنائه بمعرفة درجة القطع .

ب – للتحقق من ثبات درجة القطع المستخرج بكلا الطريقتين نيدلسكي و ايبيل للاختبار المحكي المرجع من خلال معرفة الخطأ المعياري.

٧ – **معامل كابا** لتقدير ثبات الاختبار من خلال المقارنة بين تطبيقين متتاليين للاختبار على افراد عينة الثبات.

٨ – **استخدام مؤشر برينان (Brennan– Index)** لتحليل مدى حساسية فقرات الاختبار للتمييز بين الأفراد المتقنين وغير المتقنين، اعتمادًا على نسب الإجابات الصحيحة في كل مجموعة(المتقنين) و (غير المتقنين).

٩ – **تطبيق معامل ارتباط فاي (Phi)**وذلك لقياس مايلي:

ا.لقياس قدرة الفقرات على التمييز بين المفحوصين، وفقًا لدرجة القطع المعتمدة في الاختبار.

ب.للتحقق من معامل الصدق لدرجة القطع المستخرج بطريقتي نيدلسكي و ايبيل لاختبار الرياضيات في البحث الحالي.

١٠ – معادلة الخطأ المعياري لمعرفة قيمة الخطأ المعياري لثبات درجات القطع باستخدام التباين الكلي للاختبار باستخدام معادلة كيودر – ريتشاردسون (٢٠).

١١ – معادلة نسبة الخطأ المعياري المرتبط بدرجة القرار(درجة القطع) لمعرفة نسبة الخطأ المعياري وفقا لدرجة القطع بكلا الطريقتين (نيدلسكي و ايبيل).

١٢ – **مصفوفة الارتباك (Confusion Matrix)**:تم استخدام مصفوفة الارتباك لتحليل دقة تصنيف كل من طريقتي نيدلسكي وإيبل في تحديد المتقنين وغير المتقنين، وذلك من خلال حساب المؤشرات الدقة(Accuracy) ، الحساسية(Sensitivity) ، النوعية(Specificity) ، والقيمة التنبؤية الإيجابية والسلبية لتصنيف المتقنين و غير المتقنين على الاختبار.

١٣ – **النتائج و الاستنتاجات و التوصيات و المقترحات**

أولا. المفاضلة بين درجات القطع:

تم تحقيق هدف الهدف الاول والذي ينص على (بناء أختبار تحصيلي محكي المرجع في مادة الرياضيات للصف السابع الاساسي) وذلك باتباع الخطوات المنهجية لبناء الاختبارات المحكية و التحقق من الخصائص القياسية للاختبار و فقراتها باستخدام المعادلات الخاصة بمنهجية بناء الاختبارات المحكية، أما الهدف الثاني قامت الباحثتان بالاعتماد على معيارين للمفاضلة بين درجتي القطع وخصائص الصدق والثبات لدرجتي القطع بالاعتماد على استخدام مؤشرات مصفوفة الارتباك ،وكذلك بالاعتماد على معيار (بيرك) و اجراءات مستخدمة في دراسة (العنبكي،٢٠٠٩) ، وسوف يتم عرض هذه المعايير للمفاضلة وكما يلي:

اولاً: استخدام مؤشرات مصفوفة الارتباك (Confusion Matrix)

قامت الباحثتان باستخدام معادلات مصفوفة الارتباك فكانت النتائج كما مبينة في الجدول(١٧)

الجدول(١٧)خلاصة المفاضلة الكمية بين طريقتي نيدليسكى وايبيل في تحديد درجة القطع وفق مؤشرات الصدق ،والثبات ،ودقة التصنيف وفق نتائج تطبيق معادلة الارتباك

المؤشر		طريقة ندليسكى	طريقة ايبل	الافضل
درجة القطع المحددة.		٠.٤٣ (٧٠/٣٠)	٠.٧٠ (٧٠/٤٩)	طريقة نيدلسكي تميل للخفض (تساهل نسبي)، وطريقة ايبيل تميل للرفع(تشدد نسبي).
معامل صدق درجة القطع باستخدام معادلة معامل فاي(ø).		٠.١٧	٠.٧٨	طريقة ايبيل افضل
معامل ثبات درجة القطع باستخدام معادلة ليفنجستون.		٠.٧٤	٠.٩٨	تفوق واضح لطريقة إيبيل من حيث الصدق والثبات، مما يعني أن الدرجة التي تحددها أكثر اتساقًا ومصداقية في التمييز بين المتقنين وغير المتقنين.
نسبة الخطأ المعياري		٢.٤٦	٢.٤٦	متساويتان لكلا الطريقتين اذ تعتمدان فقط على قيمة التباين و الثبات للاختبار
نسبة الخطا المعياري المرتبط بدرجة القرار(القطع)		%٨.٢٢	%٥.٠٣	منخفضة لطريقة ايبيل لذلك تعد هي الافضل.
نتائج تطبيق معادلات مصفوفة الارتباك (Confusion	الدقة(Accuracy)	٠.٤١	٠.٨٩	الافضلية لطريقة ايبيل.
	الحساسية (Sensitivity)	0.47	٠.٩٢	الافضلية لطريقة ايبيل.
	النوعية (Specificity)	٠.٣٦	٠.٨٦	الافضلية لطريقة ايبيل.
	الدقة التنبؤية الموجبة	0.38	٠.٨٧	الافضلية لطريقة ايبيل.

	(Precision)			
	الدقة التنبؤية السالبة.	٠.٤٥	٠.٩١	الافضلية لطريقة ايبيل.
	معدل الخطأ من النوع الأول(FPR)	0.64	٠.١٤	الخطأ من النوع الأول (اعتبار غير المتقن متقنًا) والخطأ من النوع الثاني (اعتبار المتقن غير متقن)، كلاهما أقل بكثير في طريقة إيبيل.
	معدل الخطأ من النوع الثاني(FNR)	0.53	٠.٠٨	

وبملاحظة النتائج في الجدول (١٧) نجد ان جميع المؤشرات الإحصائية الواردة (الصدق، الثبات، الدقة، الحساسية، النوعية، التنبؤ، ومعدلات الخطأ) لطريقة إيبيل تتفوق بشكل واضح على طريقة نيدلسكي في تحديد درجة القطع، إذ تقدم نتائج أكثر صدقًا وثباتًا، ودقة تصنيف أعلى، مع تقليل كبير في نوعي الخطأ.

ثانياً: استخدام معيار بيرك (Berk) :

أ-معايير المفاضلة بين درجتي القطع (نيدلسكي - ايبيل) ي معامل الصدق لدرجةالقطع:

اعتمدت الباحثتان معيار بيرك (Berk) للمفاضلة في معامل الصدق لدرجات القطع والذي ينص على أنه (كلما اقترب قيمة معامل فاي (ø) المستخدم في ايجاد معامل صدق درجة القطع من الواحد الصحيح كانت درجة القطع أكثر ملائمة للاختبار وأنها نجحت في التمييز بين المتقنين وغير المتقنين ، (Berk, 1976: 5)، واستفادت الباحثتان من اجراءات دراسة (العنبكي،٢٠٠٩) لهذا الغرض إذ قامتا بما يلي :

تحديد مدى درجة القطع بطريقتي نيدلسكي و ايبيل ، وقد انحصرت معامل الصدق بين (٠.١٥ -٠.٧٩) لذا فإن معايير الصدق هي:

١. إذا كان معامل صدق درجة القطع يقع في مستوى يتراوح بين (٠.١٥-٠.٣١) تحصل درجة القطع على درجة واحدة للمفاضلة لكي تكون ضمن درجات القطع الملائمة المعد ويميز بين المتقنين وغير المتقنين.

٢.إذا كان معامل صدق درجة القطع يقع في مستوى يتراوح بين (٠.٣٢-٠.٤٧) تحصل درجة القطع على درجتين للمفاضلة لكي تكون ضمن درجات القطع الملائمة لاختبار الإحصاء المعد وتمييز بين المتقنين وغير المتقنين.

٣. إذا كان معامل صدق درجة القطع يقع في مستوى يتراوح بين (٠.٤٨-٠.٦٣) تحصل درجة القطع على ثلاث درجات للمفاضلة لكي تكون ضمن درجات القطع الملائمة لاختبار الإحصاء المعد وتمييز بين المتقنين وغير المتقنين.

٤.إذا كان معامل صدق درجة القطع يقع في مستوى يتراوح بين (٠.٦٤-٠.٧٩) تحصل درجة القطع على أربع درجات للمفاضلة لكي ضمن درجات القطع الملائمة لاختبار الاحصاء المعد وتميز بين المتقنين وغير المتقنين. والجدول رقم (١٨) يوضح ذلك.

جدول رقم (١٨): نتائج المفاضلة بين درجات القطع في معامل صدق درجات القطع

ت	الطريقة	درجة القطع	قيمة معامل فاى(ø) لدرجة القطع	درجة المفاضلة
١.	نيدلسكي	٣٠	٠.١٧	١
٢.	ايبيل	٤٩	٠.٧٨	٤

ويتضح من الجدول (١٨) إن المفاضلة في معامل صدق درجات القطع تشير إلى ان أعلى درجة للمفاضلة كانت لطريقة ايبيل والبالغة قيمتها (٤) درجات ، أما طريقة نيدلسكي فكانت قيمة درجة المفاضلة لها (١) درجة. مما يشير إلى أفضلية معامل صدق درجة القطع لطريقة ايبيل.

ب-معايير المفاضلة بين درجات القطع بطريقتي نيدلسكي و ايبيل في معامل الثبات لدرجة القطع باستخدام معامل ليفنجستون:

أعتمدت الباحثتان مستوى معامل ثبات درجة الثبات للاختبار للمفاضلة في معامل ثبات درجة القطع، وبما إن قيمة معامل الثبات انحصرت بين (٠.٧٠-٠.٩٩) لذا فإن معايير الثبات هي:

١.إذا كان معامل ثبات درجة القطع يقع في مستوى يتراوح بين (٠.٧٠-٠.٧٧) تحصل درجة القطع على درجة واحدة للمفاضلة لكي تكون ضمن درجات القطع الملائمة لاختبار الإحصاء المعد وتمييز بين المتقنين وغير المتقنين.

٢.إذا كان معامل ثبات درجة القطع يقع في مستوى يتراوح بين (٠.٧٨-٠.٨٤)تحصل درجة القطع على درجتين للمفاضلة لكي تكون ضمن درجات القطع الملائمة لاختبار الإحصاء المعد وتمييز بين المتقنين وغير المتقنين.

٣.إذا كان معامل ثبات درجة القطع يقع في مستوى يتراوح بين(٠.٨٥ – ٠.٩١) تحصل درجة القطع على ثلاث درجات للمفاضلة لكي تكون ضمن درجات القطع الملائمة لاختبار الإحصاء المعد وتمييز بين المتقنين وغير المتقنين.

٤.إذا كان معامل ثبات درجة القطع يقع في مستوى يتراوح بين (٠.٩٢–٠.٩٩) تحصل درجة القطع على أربع درجات للمفاضلة لكي تكون ضمن درجات القطع الملائمة لاختبار الإحصاء المعد وتمييز بين المتقنين وغير المتقنين. والجدول رقم (١٩) يوضح ذلك:

جدول رقم (١٩)نتائج المفاضلة بين درجات القطع في معامل ثبات ليفينجستون لدرجات القطع

ت	الطريقة	درجة القطع	معامل ثبات ليفينجستون	درجة المفاضلة
١.	نيدلسكي	٣٠	٠.٧٤	١
٢.	ايبيل	٤٩	٠.٩٨	٤

ويتضح من الجدول (١٩) أن المفاضلة في معامل ثبات ليفينجستون لدرجتي القطع تشير إلى إن أعلى درجة للمفاضلة كانت لطريقة ايبيل والبالغة قيمتها (٤) درجات.

ثانياً. الاستنتاجات: توصلت الباحثتان الى الاستنتاجات الاتية:

١ – درجة القطع التي تحددها طريقة إيبل أكثر ملاءمة للاختبار وأكثر قدرة على التمييز الصحيح بين المتقنين وغير المتقنين.

٢ – طريقة إيبل توفر اتساقاً أكبر في تحديد درجة القطع عبر عينات متكررة، مما يزيد من اعتمادية القرارات المبنية عليها.

٣ – أن الخطأ المعياري العام ثابت لكلا الطريقتين لارتباطه بالتباين والثبات العام للاختبار، إلا أن نسبة تأثير هذا الخطأ على دقة اتخاذ القرار تختلف حسب موقع درجة القطع، وهو ما يؤكد أهمية اختيار درجة القطع بناءً على تحليل سيكومتري دقيق، وليس فقط على أسس نظرية أو اجتهادية.

٤ – من خلال نتائج مصفوفة الارتباك، يمكننا أن نستنتج أن طريقة إيبل كانت الأفضل في التمييز بين الأفراد بدقة. فهي قللت من الأخطاء في التصنيف، سواء من ناحية عدم قبول من لم يتقنوا، أو من ناحية عدم رفض من أتقنوا.

ثالثاً. التوصيات:

في ضوء نتائج البحث واستنتاجاته توصي الباحثتان بما يلي:

١ – نظرًا لتفوق طريقة إيبل في جميع المؤشرات (صدق، ثبات، دقة تصنيف، وتقليل الأخطاء)، يُنصح بتبنيها كطريقة رئيسية لتحديد درجة القطع في الاختبارات المماثلة.

٢ – يمكن اللجوء إلى استخدام طريقة نيدلسكي في الحالات التي يُراد فيها اعتماد مستوى أكثر تساهلاً أو عندما تكون طبيعة الاختبار تستدعي تخفيض درجة القطع.

٣ – ينبغي تنظيم تدريب مكثف للمدرسين والخبراء على إجراءات تحديد درجة القطع بطريقةايبيل عبر عقد ورش عمل تُعرّفهم بمراحل بناء المصفوفة ونسب الأوزان، يتبعها جلسات عملية على عينات تمثيلية لاختبار مدى فهمهم لآلية التصنيف.

رابعا: المقترحات:

للمزيد من الفائدة وتطوير آفاق البحث الحالي، تقترح الباحثتان اجراء البحوث المستقبلية الاتية:

١ – إجراء دراسة للمفاضلة بين عدة طرق لتحديد درجات قطع مثل طريقة ايبيل و طريقة جيجر.

٢ – دراسة أثر حجم وتنوّع لجنة الخبراء (عدد وتخصّصات) على صدق وثبات درجة القطع للاختبار.

٣ – إجراء دراسة للمفاضلة بين درجات القطع التي تركز على الاختبار و الطرق التي تعتمد على مستوى الطالب وباستخدام مصفوفة الارتباك

٤ – إجراء دراسة مقارنة بين معيار بيرك و مصفوفة الارتباك لتحديد درجة القطع للاختبارات المحكية المرجع.

المصادر العربية:

السامرائي، محمد أنور محمود، والخفاجي، أحمد محمد شاكر (٢٠١٢) بناء اختبار تحصيلي محكي المرجع في مادة علم نفس الخواص لطلبة أقسام العلوم التربوية والنفسية .مجلة الأستاذ للعلوم الإنسانية والاجتماعية،(٢٠٣): ١٠٠٢-٩٦٤

عبابنة، عماد. غصاب. (٢٠٠٩) الاختبارات محكية المرجع. ط١. عمّان: دار المسيرة للنشر والتوزيع

عبد الرحمن، سعد (1998) القياس النفسي " النظرية والتطبيق " ، دار الفكر العربي ، القاهرة ، ط٣.

عبد السلام، نادية محمد (1992 أ). ثلاث محكات ئيسية لتحديد مستوى الإتقان في الاختبارات محكية المرجع ، دراسة أمبريقية " ، المجلة المصرية للدراسات النفسية ، تصدرها الجمعية المصرية للدراسات النفسية ، العدد ٣ ، ديسمبر ، ص ص ٨٨ – ١٠٦

عبد السلام، نادية محمد (١٩٩٢ب). مشكلات عند بناء الاختبارات محكية المرجع ، تحليل وتقويم " ، مجلة علم النفس ، تصدرها الهيئة المصرية العامة للكتاب ، العدد ٢٣ ، يوليه / أغسطس / سبتمبر ، .السنة السادسة ، ص ص ٣٠ – ٣٩

علام ، صلاح الدين محمود. (١٩٨٦) تطورات معاصرة فى القياس النفسى والتربوى، الكويت، جامعة الكويت، إدارة التأليف والترجمة والنشر.

ـــــــــــ (١٩٩١) دراسة مقارنة لبعض طرق تحديد مستويات الأداء في الاختبار مرجعي المحك، المجلة المصرية للدراسات النفسية،الجمعية المصرية للدراسات النفسية، العدد (١)، ص ٧٧-٩٦.

ـــــــــــ (١٩٩٥) الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية . القاهرة : دار الفكر العربي

ـــــــــــ(٢٠٠٠) تحليل بيانات البحوث النفسية والتربوية والاجتماعية ، دار الفكر العربي ، القاهرة.

ـــــــــــ(٢٠٠١). القياس والتقويم التربوي والنفسي، القاهرة: دار الفكر العربي.

ـــــــــــ(٢٠٠٣) . التقويم التربوي المؤسس، القاهرة : دار الفكر العربي

ـــــــــــ(٢٠٠٧). الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية . القاهرة . دار الفكر العربي .

العنبكي، حيدر. جليل. عباس. (٢٠٠٩) المفاضلة في تحديد درجات القطع لاختبار محكي المرجع. أطروحة دكتوراه،(غير منشورة)كلية التربية – ابن رشد، جامعة بغداد.

فتح الله، محمد محمد (١٩٩٥). بناء اختبار محكى المرجع فى العلوم لتلاميذ نهاية مرحلة التعليم الأساسى. رسالة ماجستير (غير منشورة)، جامعة الأزهر، كلية التربية، قسم علم النفس التعليمى.

مجيد، سوسن شاكر (٢٠١٤)أسس بناء الاختبارات والمقاييس النفسية والتربوية ط.٣ ،مركز ديبونو لتعليم التفكير

منشاوي، أحمد عباس (٢٠٢٢) تحديد درجة القطع في مقياس الاحتراق الأكاديمي باستخدام طريقة أنجوف الموسعة، والتنبؤ به في ضوء متغيرات الكمالية الأكاديمية والشغف الأكاديمي والصمود الأكاديمي لدى طلاب الجامعة. مجلة الإرشاد النفسي، العدد ٧١ ج٢ ،أغسطس، ص ص ٤٨٥- ٥٦٤

المصادر الأجنبية:

Agresti, A. (1996). *Introduction to categorical data analysis*. New York, NY: Wiley.

Alsmadi, A. A. (2007). A comparative study of two standard setting techniques. *Social Behaviour and Personality*, 35 (4), pp. 479-486.

ALTMAN, Douglas G.; BLAND, J. Martin. Diagnostic tests. 1: Sensitivity and specificity. *BMJ: British Medical Journal*, 1994, 308.6943: 1552.

Anastasi, A., Urbina, S., 1997. Psychological Testing, 7ed ed. Dorling Kindersley, India.

Berk R. A. (1986) Consumer's guide to setting performance standards on criterion referenced tests. *Review of Educational Research*, 56 (1), pp. 137-172. (Available

Brennan (Ed.), Educational measurement (4th ed., pp. 531–578). Westport, CT: American Council on Education/Praeger.

Brennan, R.L. and Kane, M.T., 1977. *An index of dependability for mastery tests*. *Journal of Educational Measurement*, 14(3), pp.277–289. https://onlinelibrary.wiley.com/doi/10.1111/j.1745-3984.1977.tb00045.x

Brown, Frederick.Gramm(1989*). Principles of Educational and Psychological Testing* (3rd ed.). New YorkHolt, Rinehart and Winston

Brown, DH. (2003). Language Assessment: *Principles and Classroom Practices*. San Francisco: Longman

Chang, L. (1999).Judgmental Item Analysis of the Nedelsky and Angoff Standard-Setting methods .*Applied Measurement in Education*, 12 (2): pp. 151–165.

Cizek, G. J. & Bunch, M. B. (2007). Standard setting: *A guide to establishing and evaluating performance standards on tests*. Thousands Oaks, CA: Sage Publications.
Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), pp. 37–46. doi:10.1177/001316446002000104
Crocker L. & Algina, J., 1986. *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston
https://archive.org/details/introductiontoclassicalandmoderntesttheory
Crocker L, L. & Algina, J. (2006). *Introduction to Classical and Modern Test Theory*. Mason, OH: Cengage Learnin
Ebel, R. L. (1972). *Essentials of Educational Measurement* (1st ed.). Upper Saddle River, NJ: Prentice Hall.
Esler, W. K., & Dziuban, C. D. (1974). Criterion referenced test: Some advantages and disadvantages for science instruction. *Science Education*, 58(2), pp. 171–174. doi:10.1002/sce.3730580205
Fawcett, Tom. "An introduction to ROC analysis." *Pattern recognition letters* 27.8 (2006): 861-874.
Geisinger, K. F., Bracken, B. A., Carlson, J. F., Hansen, J.-I. C., Kuncel, N. R., Reise, S. P., & Rodriguez, M. C. (Eds.). (2013). *APA handbook of testing and assessment in psychology, Vol. 1. Test theory and testing and assessment in industrial and organizational psychology*. American Psychological Association. https://psycnet.apa.org/record/2012-22485-000
Gravetter,F.J. and Wallnau,L.B.,(2017).*Statistics for the Behavioral Scieces,10th ed .Bostonm, MA: Cengage Learning.*
Groth-Marnat, G. and Wright, A.J., 2003. Handbook of psychological assessment. 7th ed. Hoboken, NJ: Wiley.

Gündeğer, C. and Doğan, N. (2014). Comparison of Angoff, yes/no and Ebel standard setting methods. *Journal of Measurement and Evaluation in Education and Psychology*, 5(1), 53-60.
Çokluk, Ö. and TaşdemiR, F. (2013) Angoff (1-0), Nedelsky ve Sınır Değerleri Saptama Yöntemleri ile Bir Testin Sınıflama Doğruluklarının İncelenmesi. *Mediterranean Journal of Humanities*, 3 (2), pp. 241–241.
Hambleton, R.K.(1978). O*n the use of Cut Off Score with criterion referenced test in instructional setting jem.*
Hambleton, R.K & Novick ,M.R(1983).*Toward an integration of theory and method for criterion referenced tests* .JEM .10: pp. 159-170
Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
Hani ,A., Sami A., &Hussein H.(2023). Building a Criterion-Referenced Test in Measurement and Evaluation and Determining Its Cut-Off Score by Several Methods. Journal of Education and Practice,14(1): pp. 61-68
Kane, M., & Bridgeman, B. (2017). Research on validity theory and practice at ETS. In R. E. Bennett & M. von Davier (Eds.), *Advancing human assessment: The methodological, psychological and policy contributions of ETS* (pp. 489–552). Springer Science + Business Media. https://doi.org/10.1007/978-3-319-58689-2_16
Koretz, D., & Hamilton, L. S. (2006). *Testing for accountability* in K–12. In R. L.
Kubat, M. (2017). *An Introduction to Machine Learning* (2nd ed.). Springer.
Kuder, G.F. and Richardson, M.W., 1937. The theory of the estimation of test reliability. *Psychometrika*, 2(3), pp.151–160. Available at: https://doi.org/10.1007/BF02288391

Lin, Hui-Fen(1988):((*Acomparison of Three Item Selection Methods in Criterion Referenced Tests*)), Dissertation(Doctor of Philosophy), University of North Texas, USA

Livingston, S. A. (1972). Criterion-referenced applica-tions of Classical Test Theory. *Journal of Educational Measurement*,9(1), pp. 13-26

Livingston, S.A. & Zieky, M.J., 1982. *Passing scores: A manual for setting standards of performance on educational and occupational tests*. Princeton, NJ: Educational Testing Service. p. 24. https://www.scirp.org/reference/referencespapers?referenceid=830530

Markus, K. A., & Borsboom, D. (2013). Frontiers of test validity theory: *Measurement, causation*, and meaning. Routledge

Mrunalini, T. (2013). *Educational Evaluation.* 5th edition. New Delhi: Neelkamal Publications.

Nedelsky, L. (1954). Absolute grading standards for objective tests. *Educational and Psychological Measurement*, 14, pp. 1–19.

Norton, B. T. (1978, February). Karl Pearson and statistics: *The social origins of scientific innovation. Social Studies of Science*, 8(1), 3–34. https://faculty.fiu.edu/~blissl/Pearson1.pdf

Popham, W.J. (1975) *Educational Evaluation. Hall Press*, New York

Popham, W. J. (1978). *Criterion- Referenced Measurement*. New Jersy: Prentice-Hall.

Powers, D. M. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1), 37–63.

Richards, JC & Schmidt, R.S. (2002). *Approaches and Methods in Language Teaching*. New York: Cambridge University Press.

Rudner, M. (2001). Computing the expected proportions of misclassified examinees. Practical Assessment, *Research & Evaluation*, 7(14). Retrieved August 18, 2006 from http://edresearch.org/pare/getvn.asp?v=7&n=14 . This paper has been viewed 3,351 times since 3/30/01

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, 45(4), 427–437.

Subkoviak, Michael J.; Kane, Michael T.; and Duncan &Patrick H. (2002) A Comparative Study of the Angoff and Nedelsky Methods*: Implications for Validity*," Mid-Western Educational Researcher15(2):1-6. Available at: https://scholarworks.bgsu.edu/mwer/vol15/iss2/2

Tanrıverdi, S. (2006). *The effects of standard setting methods on passing scores*. Master's Thesis (Hacettepe University, Institute of Social Sciences).

Taşdelen, G. (2009). *A study on the comparison of Nedelsky and Angoff standard setting methods with generalizability theory*. Unpublished Master's Thesis. Hacettepe University, Ankara.

Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics*, 17(1), 168–192.

Weber, P. (2017). Measurement Error. In. J. Matthes (Ed.), The International Encyclopedia of Communication Research Methods. Hoboken, NJ: Wiley-Blackwell. doi: 10.1002/9781118901731.iecrm0138

Zieky, M. and Perie, M. (2006). A primer on setting cut scores on tests of educational achievement. Princeton, NJ: Educational Testing Service (ETS).

المصادر العربيه المترجمه للغه الانكليزيه

Al-Samarrai, Muhammad Anwar Mahmoud, & Al-Khafaji, Ahmed Muhammad Shaker (2012) Construction test of achievement in the subject of psychological properties for students of the departments of educational and psychological sciences. Journal of the Humanities and Social Sciences, (203): 1002-964

Ababneh, Imad. Ghassab. (2009) Spoken Tests Reference. I1. Amman: Dar Al-Masirah for Publishing and Distribution Abdel Rahman, Saad (1998) Psychometrics "Theory and Practice", Dar Al-Fikr Al-Arabi, Cairo, 3rd Edition.

Abdel Salam, Nadia Mohamed (1992a). Three main criteria to determine the level of proficiency in spoken tests reference, empirical study", Egyptian Journal of Psychological Studies, issued by the Egyptian Society for Psychological Studies, Issue 3, December, pp. 88-106

Abdeslam, Nadia Mohamed (1992b). Problems when Construction spoken tests reference, analysis and evaluation", Journal of Psychology, issued by the Egyptian General Book Authority, Issue 23, July / August / September, . Sixth Year, pp. 30-39

Allam, Salahuddin Mahmoud. (1986) Contemporary Developments in Psychological and Educational Measurement, Kuwait, Kuwait University, Department of Authorship, Translation and Publishing

Allam (1991) A comparative study of some methods of determining performance levels in the test reference test, Egyptian Journal of Psychological Studies, Egyptian Society for Psychological Studies, Issue (1), pp. 77-96.

Allam (1995) Diagnostic tests reference test in the fields of educational, psychological and training. Cairo : Dar Al-Fikr Al-Arabi

Allam (2000) Analysis of psychological, educational and social research data, Dar Al-Fikr Al-Arabi, Cairo.

Allam (2001). Educational and psychological measurement and evaluation, Cairo: Dar Al-Fikr Al-Arabi.

Allam, S. M. (2003*).Institutional Educational Evaluation: Its Foundations, Methodologies, and Applications in Teacher Evaluation*.Cairo: Dar Al-Fikr Al-Arabi.

Allam, S. M. (2007).Criterion-Referenced Diagnostic Tests in Educational, Psychological, and Training Contexts.Cairo: Dar Al-Fikr Al-Arabi.

Anbaki, H. J. A. *(2009).Differentiation in Determining Cut-off Scores for Criterion-Referenced Tests*. , (unpublished), College of Education - Ibn Rushd, University of Baghdad.

Fathallah, M.M., 1995. *Construction of a criterion-referenced test in science for students at the end of the basic education stage*. Unpublished Master's thesis, Al-Azhar University, Faculty of Education, Department of Educational Psychology.

Majeed, S.S., 2014. *Foundations of constructing psychological and educational tests and scales*. 3rd ed. Amman: Debono Center for Teaching Thinking.

Minshawi, A.A., 2022. Determining the cut-off score in the academic burnout scale using the extended Angoff method, and predicting it in light of academic perfectionism, academic passion, and academic resilience among university students. *Journal of Psychological Counseling*, 71(2), pp.485–564.

Al-Nabhan, M., 2004. *Fundamentals of measurement and behavioral sciences*. Amman: Dar Al-Shorouk for Publishing and Distribution.

.