المقارنة بين بعض النماذج الإحصائية لتحديد درجة القطع في اختبار مرجعى المحك "دراسة سيكومترية مقارنة"

إعـــــداد

د/ محمد بن حسن يحيى الزبيدي

أستاذ القياس والتقويم المشارك بقسم علم النفس

جامعة الطائف

المقارنة بين بعض النماذج الإحصائية لتحديد درجة القطع في اختبار مرجعى المحك "دراسة سيكومترية مقارنة"

د/ محمد بن حسن يحيى الزبيدي *

الملخص:

تشير نتائج الدراسة الحالية الى وجود فاعلية نماذج إحصائية في تحديد الدرجة الفاصلة لاختبار مرجعي المحك في مادة مقاييس الذكاء والقدرات .وقد قام الباحث بمقارنة بين أسلوبين، أحدهما يستخدم نماذج إحصائية منفصلة (مثل نموذج ذي الحدين) والآخر يستخدم نماذج إحصائية متصلة (مثل طريقة أنجوف)، حيث تم تطبيق الاختبارعلى طلاب كلية الآداب جامعة الطائف. بالمملكة العربية السعودية، واشتمل الاختبار على 20 سؤال من نوع الاختيار من المتعدد، وكانت من أهم النتائج التي توصل اليها البحث الى أهمية أسلوب أنجوف في تحديد درجة القطع عند مستوى % 66.73أى (13) سؤال تقريبا لتمكن الطالب، بينما نموذج ذى الحدين أسفرعن وجود فروق ذات دلالة إحصائية بين التوزيع الواقعى والتوزيع النظرى للنسب المئوية عند مستوى تمكن %0.70وهو المستوى المطلوب للتمكن لاجتياز اختبار مقاييس الذكاء والقدرات، أى أن الطالب المتمكن يستطيع الإجابة على 14 سؤال، وكانت من أهم النتائج التي توصل اليها الباحث أن نموذج ذى الحدين أدق فى حساب درجة القطع نظراً أنها أقل تحيز إحصائي.

الكلمات المفتاحية : درجة القطع، مسميات مستويات الأداء، محك تصحيح المفردة، طريقة أنجوف

* **د/ محمد بن حسن يحيى الزبيدي:** أستاذ القياس والتقويم المشارك بقسم علم النفس–جامعة الطائف.

Comparison between some statistical models to determine the cut-off score in a benchmark test "a comparative psychometric study."

Dr. Muhammad bin Hassan Yahya Al-Zubaidi

Associate Professor of Measurement and Evaluation Department of Psychology, Taif University

Abstract

Refers to the results of a current study on the effectiveness of statistical models in determining the cut-off score for a benchmark test in the subject of Intelligence and Abilities Scales. The researcher compared two methods, one of which uses discrete statistical models (such as the binomial model) and the other uses continuous statistical models (such as the Angoff method) , where the test was applied to students at the For students of the Faculty of Arts, Taif University of Saudi Arabia, The test contained 20 multiple-choice questions, and one of the most important results that the research reached was the importance of the Angoff method in determining the cut-off score at the level of 66.73%, or approximately (13) questions, to enable the student to master, while the binomial model resulted in significant differences. There is statistical significance between the realistic distribution and the theoretical distribution of percentages at a mastery level of 0.70%, which is the level required for mastery to pass the Intelligence and abilities test, meaning that a proficient student can answer 14 questions. One of the most important results reached by the researcher was that the binomial model is more accurate in calculating the score. cut off because it has less statistical bias

Keywords: cut-off score, nomenclature of performance levels, criterion for correcting vocabulary.

مقدمــة:

تنادى الاتجاهات الحديثة فى التعليم بجودة التعليم للوصول إلى التمكن والإتقان فى التعليم ومن ثم تنادى الاتجاهات الحديثة أن تصنف اختبارات الأفراد سواء كانت تحصيلية أو مهنية فى مجال التقدم لوظيفة مثلا إلى أفراد متقنيين وغير متقنيين، ولعل المثال الواقعى اليوم هو اختبار الرخصة الدولية لقيادة الحاسب الآلي International Computer Driving License حيث تقدر درجة نجاح الفرد بالمرور من الاختبار passأو عدم المرور ومن ثم الحكم على الممتحن بالتمكن أو عدم التمكن وكل اختبار من اختبارات الرخصة له درجة للنجاح للحكم على مدى تمكن الممتحن وهذه الدرجة ممكن أن نطلق عليها الدرجة الفاصلة cut of scour، حيث يشير جيمس وآخرون James Algine , Douglas Bill Coulson, (1978) إلى أنة لكى نصل إلى برامج تعليمية موضوعية يجب الاهتمام بالاختبارات مرجعية المحك حيث إن درجاتها تزودنا بمعلومات تمكننا من التعرف على الفروق الفردية بين الأفراد ومن ثم نستطيع اتخاذ قرارات موضوعية فى ضوء بيانات درجة الاختبار Test Score Informationويذكر أيضا أن الاختبار المرجع إلى معيار Norm Referencedاقل فى هذا الجانب من الاختبار المرجع إلى محك. ; .James Algine) (Douglas Bill Coulson: 1978 , 1- 47

حيث يصيغ النظام الذى ينسب أداء الفرد إلى محك Criterion – Referenced System إلى تفسير الدرجات المستمدة من الاختبارات والمقاييس بموازنة أداء الفرد بمحك أداء متوقع. ويصاغ الأداء عادة على صورة كفايات محددة أو نواتج متوقعة أو أهداف سلوكية مرتبة بحيث تصف مختلف مستويات الأداء. ولا تستند مرجعية تفسير الدرجة فى الاختبار أو المقياس مرجعى المحك إلى أداء الأقران أو معيار جماعة التقنين، وإنما إلى الأداء المتوقع أو المرجو تحققه والذى يحدد تحديداً دقيقا. (صلاح الدين محمود علام: 1995، 17)

وتذكر لورا جودوين (Laura D. Goodwin, 1996) إلى أن هناك طرق رئيسية لتحديد درجة الفصل لإعطاء شهادة أو رخصة امتحان وأن هذا يتحقق من خلال العديد من الاختبارات محكية المرجع بحيث تكون نتيجة الممتحن فى الاختبار (نجاح أو فشل /مرور أو رفض) وتصلح لوصف الأداء، وأن هناك حالات لمقاييس كثيرة تستخدم لحساب درجة الفصل لتحديد من سوف يمر (من ينجح فى الاختبار ومن يفشل) (who will pass a test and who will fail)، فضلا عن ذلك تحدد هذه الدرجة من يأخذ شهادة أو رخصة ومن يأخذ اجر ومن يقبل فى برامج الجامعة ومن يسمى متقن masterومن يسمى غير متقن Non

master أى أننا فى حاجة إلى درجات قطع كثيرة ومتنوعة (Laura,. Goodwin, 1996: 249 – 256)

وقد زاد الاهتمام ببناء الاختبارات المرجعة إلى محك فى الفترة الأخيرة والتى تم فيها تحديد المحكات بدرجات فاصلة cut of points ينبغى أن يصل إليها المفحوص حتى يجتاز الاختبار بنجاح، فمثلا إذا أردنا أن نعلم الأطفال جمع الأرقام عند مستوى إتقان 85%فأكثر، فيتم تحديد المحك فى ضوء ما يلى:

- هدف تعليمى معين.
- مستوى كفاءة محدد ينبغى أن يصل إلية كل متعلم.
- نتيجة محددة مطلوب تحقيقها بواسطة كل متعلم. وهذا النوع من الاختبارات لا يقتصر على قياس عينة من السلوك الإنسانى فحسب ولكنة يشتمل على جميع جوانب السلوك المراد قياسه ويستخدم هذا النوع من الاختبارات فى عدة استراتيجيات تربوية منها ما يلى:
  - استراتيجية التعلم للإتقان.
  - التعلم المبرمج.
  - تفريد التعليم.
  - التعليم القائم على الكفاءات. (محمود عبد الحليم منسى، 192 : 2003).

وقد أشار (محمود منسى 2003) أنة نظرا لزيادة أعداد المتقدمين لامتحانات الشهادات العامة وبخاصة شهادة الثانوية العامة، فأن هذا قد أدى إلى زيادة عدد المعلمين المشاركين فى أعمال تصحيح إجابات الطلاب فى هذه الامتحانات ومن ثم قد نتجت بعض الأخطاء فى تصحيح أوراق الإجابات وفى رصد درجاتها مما أدى إلى لجوء بعض الطلاب وأولياء الأمور إلى القضاء مشككين فى عدالة هذه الامتحانات وتنظر المحاكم كل عام عددا كبيرا من قضايا الامتحانات التى تركز على العدالة الاجتماعية لهذه الامتحانات وعلى مدى عدم عدالتها بالنسبة للشاكين وهذا يؤكد أن هناك بعض المآخذ القانونية على الامتحانات التى تعقد على مستوى قومى فى مصر. (محمود عبد الحليم منسى،119 :2003)

فضلا عن ذلك فأن كل أشكال الامتحانات التى تقدم للطلاب وخصوصا امتحانات الشهادات العامة تمثل شكلا من أشكال الإرهاب الاجتماعى، ودرجة الفصل هى محك جوهرى (أساسي) فى تحديد الكفاءة المعرفية للاختبار، ويتم الوصول إلى الدرجة الفاصلة من خلال اخذ آراء أهل الخبرة (Leon J. Gross: 1985, 469 – 493).

وفى هذا الصدد أشارت (دراسة منار أحمد طومان 2003) إلى أن الدراسات التى قارنت بين بعض طرق تحديد درجات الفصل للتعرف على مدى الفروق بينهما ما زالت محدودة حتى الآن، ولم يحسم الأمر بشكل نهائى لصالح إحدى الطرق. (منار أحمد طومان، 2003: 55)

وسوف يستخدم الباحث نموذجين لتحديد الدرجات الفاصلة أحدهما طريقة أنجوف كنموذج متصل ونموذج ذى الحدين كنموذج منفصل وذلك للمقارنة بين أفضلية أى النموذجين المتصل أم المنفصل لتحديد درجة الفصل.

حيث تشير سوميه (شكرى، 2006)إلى أن هناك ندرة فى الدراسات التى حاولت دراسة صدق درجات الفصل، بل ركزت معظم الدراسات على مقارنة الدرجات الناتجة عن استخدام طرق مختلفة فى حساب درجات الفصل وعند تدريس الوحدات أو الموضوعات المتتابعة تكمن خطورة خطأ القياس المرتبط بدرجات الفصل فى القرارات التى يجب أن يتخذها المربون بناء على هذه الدرجات، (حيث إنه فى حالة خطأ ألفا) ينتقل الطلاب غير المتقنين إلى مرحلة تعليمية تالية دون تمكن من المهارات السابقة، فلا يمكنهم استيعاب مهام هذه المرحلة الجديدة، وفى حالة خطأ (بيتا) يكلف الطلاب المتقنون بإعادة تعلم مهام قد أتقنوها وهم فى غنى عن تعلمها من جديد، فضلا عن الأثر النفسى السلبى الذى يلحق بهم.(سوميه شكرى، 7 :2006)

وفى هذا الصدد هدفت دراسة (Ricker, Kathryn L. 2006) إلى إعادة نقد أنجوف (1971) وتعديل طرق أنجوف لإعداد أو تحديد درجات الفصل Cut – of Scores وأن المحك المستخدم فى هذا التقرير ما اقترحه بيرك Berk 1986أن هناك الكثير من المزاعم حول طريقة أنجوف وأن هناك إصدارات مختلفة تناولت ونقدت طريقة أنجوف.

(Ricker, Kathryn L, 2006: 53 – 64)

أيضا فى إجراءات تحديد معيار أنجوف يتم تقدير أداء المحكمين باحتمال أن يكون اختيار المحكم عشوائيا واختير بطريقة كفء لأدني استجابة صحيحة لكل مفردة مكونة أو مؤلفة فى الاختبار. وفى حالات كثيرة يكون تقدير أداء المفردة يكون مرتين، بواسطة المعلومات المقسمة مع المحكمين بين التقديرات لاسيما فيما يتعلق بالاختبارات الطويلة، عمليات التقدير هذه تستهلك وقت وترهق fatiguing المشتركين وقد نادت هذه الدراسة بإمكانية اختزال عدد المفردات للمحكمين فى دراسة تحديد معيار أنجوف Angoff Standard – Setting ونتائج هذه الدراسة تقترح أن 50% من مفردات الاختبار ربما تكون كافية لتقدير درجة مرور مكافئة وتسجل دراسة أنجوف هذه النتيجة المقدرة بشكل جوهرى لادخار الوقت الكثير.(Ferdous, Abdullah. A.; Plake Barbara, 2005: 185 – 201)

أن إجراء الإطار المعيارى لكتاب العلامات الخاصة بالمحكمين يساعد فى التغلب على كثير من المشكلات غير المفهومة بالنسبة للمحكمين وأن اكثر الطرق انتشارا ومتعلقة بتعيين درجة الفصل بإجراءات أنجوف (1971) بحيث إن الغرض من هذه الدراسة مراجعة كتاب العلامات bookmark وتقويم البنود أو المفردات لبيرك (Berk, S., 1986) ومحك تقويم

درجة الفصل ونقاط القوى والضعف المتعلقة بكتاب العلامات وعموما فأن نقاط قوة Bookmark هى:

- تسوية بنية الاستجابة فى مفردات اختيار الاستجابة المختارة.
- تسوية كفاءة درجات الفصل متعددة خصوصا فى الاختيار من المتعدد.
- اختزال التعقيد المعرفى المتعلق بالمحكمين.

ومن ثم يظل كتاب العلامات إجراء قاطع لتحديد درجة الفصل ومعايير الأداء النهائية. (Lin , Jie, 2006: 36 – 52)

ويشير باول (Brandon, Paul R. ,2004) فى مقالة بعنوان "استنتاجات حول موضوعات أنجوف "إلى أن هذه المقالة تراجع (9) عناوين لتعديل طريقة لتحديد معيار أنجوف ويقول أننا يمكن أن نكون واثقين بشكل معقول حول اختيار عدد المحكمين المناسب لزيادة اتساع المحكمين لأنجوف لتعديل بنود أنجوف وعلى الرغم من القصور فى قدرة الطرق المتعلقة بالإصلاح أو معالجة ما يتعلق بتأثير أنشطة المحكم بين تجميع المعيار المحدد، تحتاج كثير من الدراسات إلى امتلاك المحكم مستوى من الخبرات وعمليات حول وصف الأداء للمستوى عند المشاركين لتحديد درجة الفصل (فئاتها)، أن التطور التجريبى لطريقة أنجوف غالبا به نقص الكفاءة Sufficientوهناك عدم استقرار حول الطرق الباقية. (Brandon, Paul R, 2004: 59 – 88)

وتشير دراسة جيراد وجيمس وبلاك Giraud , Gerald ; Impara , James c.; (Plake , Barbara S. 2005) حول إدراكات المعلمين للمفحوصين "المختبرين "فى طريقة تحديد معيار أنجوف إلى عملية تحديد درجة الفصل وأنة يجب النظر إلى موضوع القضية وإلى أهل الخبرة من المحكمين لصناعة الأحكام حول الأداء المفضل الذى يتعلق بالاختبارات فى مستوى المهارة المستخدم.وقد استخدمت درجة الفصل لتحديد 12 – Kالفصل الطلاب الذين لا يمتلكون إتقان مهارات محكية، وذلك للحكم على مدى تمكن الطالب، وتختبر هذه الدراسة مدى مشاركة المعلمين فى تحديد درجة الفصل وذلك من خلال ورش العمل أو المشاركة لعمل تصور حول هدف الاختبار حيث وصف غالبية المعلمين أهمية تحديد درجة الفصل وأهمية تفسيرها. (Giraud , Gerald. ; Impara , James c.; Plake , Barbara S., 2005: 223 – 232)

ومما سبق يحاول الباحث دراسة "مدى فاعلية بعض النماذج الإحصائية فى تحديد درجة الفصل لاختبار مرجعى المحك " لذلك تناول الباحث طريقة أنجوف وذى الحدين لمعرفة أى الأسلوبين افضل وأدق لتحديد درجة الفصل.

مشكلة البحث:

بالنسبة إلى أهمية وجود معيار محدد لتحديد درجة الفصل تشير دراسة ريكاس (Reckase, Mark D. 2006) إلى أن سكيولز (Schulz, 2006) أشار إلى التزويد بتصورات مختلفة عن وضع أو تحديد المعيار ويقترح تعديل كتاب العلامات الخاص بالمحكمين Book Mark وإجراءاتها وبعض المشاركين فى لجان التحكيم، وأنة يجب تعديل طريقة أنجوف باستخدام اقتراحات متعلقة بالمشاركين فى لجان التحكيم واحتمالات تصحيح الاستجابات وتقديرها وأن تكون قاطعة وأنة يجب أن يكون هناك درجة تعقيد معينة لتحديد المعيار ومن ثم مناقشة هذا المعيار. (Reckase, Mark D., 2006: 14 – 17)

وتشير دراسةريكاس (Reckas , Mark D. 2006) إلى اقترح عمل إطار مفاهيمى حول نظرية القياس النفسى متعلقة بوضع أو تحديد المعيار (Standard Setting)أو إطار أو هيكل الاقتراحات للمشاركين فى التحكيم فى وضع عملية أو تحديد المعيار (المشاركين فى لجان التحكيم) وأن تنمى بشكل داخلى، وأنة يجب أن يكون المحكمين على درجة من التدريب وخلفياتهم مرتبطة بالمجال (ذات صلة) والقياس النفسى متضمن فى هذه العملية لأن نقاط التقدير يتم تحديدها بواسطة المحكمين ويجب الاهتمام بالإطار المفاهيمى لتحديد المعيار المستخدم واشتقاق محكات لذلك متعلقة بعمليات التقويم و وضع المعيار وتغيير فى Book Mark كتاب العلامات وتعديل طريقة أنجوف. (Reckase, Mark D., 2006: 4- 18)

تشير دراسة (لى وليندر و هانس) Change, Lei.; Van Der Linder Wim J.; (Vos, Hans J. 2004)، إلى أنة يجب اكتشاف التعارض فى آراء المحكمين حول الاستجابات الخاصة بالمفردات الخاصة بالاختيار من المتعدد وأن التعارض يكمن فى الفرق بين النتائج التى نحصل عليها بالملاحظة وبين النتائج المحسوبة على أساس صيغة ما وأنة بتحليل الطرق القائمة بالفعل (أنجوف وندلسكى) وجد أن أنجوف افضل وأنة يجب تحديد مصدر التعارض بين المحكمين سواء كانت فى (بدائل الإجابة أو المحكمين) وأنه يجب أن يكون هناك تقويمات تكاملية ليس بها تعارض.

(Change , Lei.; Van Der Linder Wim J.; Vos , Hans J., 2004: 781–801)

وقد أجرى عصام الدسوقى (1998) دراسة لبحث مدى فاعلية طريقة أنجوف فى تحديد درجة الفصل، وذلك من خلال مقارنة نتائجها مع نتائج طريقة هوفستى واستخدم الباحث اختبارا من إعداده فى وحدة معادلات الدرجة الثانية من مقرر الجبر للصف الأول الثانوى، يتضمن كل منهما 20مفردة من نوع الاختيار من متعدد ذات الأربع استجابات.

وقد استعان الباحث بعدد 6 محكمين من موجهى الرياضيات بالإضافة إلى أربعة مدرسين أوائل فضلا عن ذلك تكونت عينة الدراسة من 65طالبا وقد استخدم الباحث رسم المنحنى فى طريقة هوفستى وقد توصلت الدراسة إلى فاعلية طريقة أنجوف فى تحديد درجة الفصل، حيث لم تختلف نتائجها مع نتائج طريقة هوفستى، حيث كانت درجة الفصل بطريقة أنجوف 74% بينما بلغت درجة الفصل باستخدام طريقة هوفستى 86% أى أن الاختلاف بينهما كأن فى مدى قليل. (عصام الدسوقى، 73 – 43 :1998)

وتشير سوميه شكرى (2006) فى دراستها عن "فاعلية بعض طرق تقدير درجة الفصل فى التنبؤ بالتحصيل اللاحق فى الهندسة للمرحلة الإعدادية "إلى اختلاف قيمة درجة الفصل تبعا لاختلاف الطريقة المستخدمة فى حسابها (أنجوف، المجموعات المتقابلة، المجموعة الحدية).

أيضا اختلفت فاعلية درجة الفصل فى التنبؤ بالتحصيل اللاحق فى الهندسة تبعاً لاختلاف الطريقة المستخدمة فى حسابها (أنجوف، المجموعات المتقابلة، المجموعة الحدية)، حيث إن طريقة المجموعات المتقابلة أكثر فاعلية فى التنبؤ بالتحصيل اللاحق، تليها طريقة المجموعة الحدية، ثم أخيراً طريقة أنجوف. (سوميه شكرى، 264 :2006)

أى أنة لا يوجد قرار واضح ومحدد حول أفضلية طريقة على أخرى لتحديد درجة الفصل ومن ثم يحاول الباحث المقارنة بين النماذج المنفصلة والمتصلة لمعرفة مدى فاعلية أى منهما لتحديد درجة الفصل.

وتشير دراسة عبد الله وباربارا(Ferdous, Abdullah A.; Plake Barbara S.2005) عن "فهم العوامل التى تؤثر على قرارات أعضاء اللجنة فى دراسة تحديد المعيار "إلى اختبار كيف يمكن وضع معيارا للمشاركين فى التفكير حول وضع تقدير للأداء على مفردة ما متعلقة بالطلاب الماهرين وهذه الدراسة امتداد لدراسات سابقة باعتبار أن هناك عوامل تأثير متعلقة بقرارات المشاركين فى التحكيم فى أنجوف 1971اعتمادا على وضع معيار كوظيفة أو دالة لتقدير الأداء على المفردة وقد قسم المحكمين إلى ثلاث مجموعات وذلك لتقدير الأداء على مفردة واحدة (منخفضة –متوسطة – مرتفعة) (low– moderate – high)أظهرت الدراسة إلى أن المشاركين فى المجموعات المنخفضة لم يظهروا التفكير فى التصور أو منظور مرجعى المعيار non- referenced بينما المجموعات المرتفعة والمتوسطة فكروا فى منظور محكى المرجع Criterion – Referenced.

(Ferdous , Abdullah A.; Plake Barbara S, 2005: 257 – 267)

وقد أوضحت دراسة باربارا(Plake, Barbra S., 2000) إلى أهمية اختبار التخمين الضمنى والاتساق المتعلق بالأداء على المفردة وتوقع ذلك من تحديد معيار أنجوف Angoff Standard Setting لمدة عامين وذلك بواسطة 29من المشاركين فى لجان التحكيم فى عام واحد و 30 محكم فى مرة ثانية. وقد أثبتت النتائج أن هناك تحسن فى تقديرات الأداء على البند وأن هناك اتساق داخلى وذلك عبر أو من خلال الهيئات وهذا قد يستغرق عامين.وهذه العوامل تكون ذات تأثير كبير على الدرجات العليا وثبات المفردات (Plake, Barbra S., .. 2000: 347- 355)

وفى دراسة لى شانج (Chang, Lei, 1999) التى قارنت بين نيدلسكى L.Nedelsky , (1954) أنجوف (Angoff ,W 1971) فى تحديد المعيار فى ثلاث دراسات تضمنت 80من المحكمين Judgesلدرجة الفصل لندلسكى ذات دلالة اقل عن درجة الفصل لدى أنجوف وأن هناك اقتراحات أخذت مميزات كل منهما لوضع إجراءات تحديد المعيار.
(Chang. Lei., 1999: 151 – 165)

وفى هذا الصدد تشير دراسة لورا جودوين (Goodwin, Laura D. 1999) إلى العلاقة بين تقديرات أنجوف Minimum Passing Levels مستويات الاجتياز البسيطة وقيمة pالحقيقية المتعلقة بالحدود الفاصلة للممتحنين borderline فى دراسة 115 من الممتحنين اخذوا اختبار التخطيط المالي المعتمد، وجدوا أنة لا توجد اقتراحات لمحكمين أنجوف لمهمة تكون مستحيلة جدا ولكن الاقتراحات تأخذ تحسن فى إجراءات تحديد المعيار.
(Goodwin , Laura D., 1999: 13 – 28)

وتقترح دراسة كل من نيدلسكى (1954) وأنجوف (1971) إجراءات لتكوين أساس لدرجة الفصل (الدرجات –معدلات –النسب –القيم) للمحكمين حول الأداء المحتمل المتعلق بأدنى كفاءة للممتحنين على بنود الاختبار فى هذه الورقة نظرية القدرة المعممة Generalizability theory أو تعميم القدرة حيث تستخدم فى قياس التباين لنتائج درجة الفصل من بعض التخمينات وبيانات الاختبار التجريبى تستخدم فى تفسير هذا المدخل والمقارنة بين الإجراءين مع الأخذ فى الاعتبار تأثير (الدرجات –معدلات –النسب –القيم) وهناك تعارض أو بعض الاختلافات بين حساب الثبات والصدق بين النتائج المقترحة لكل من إجراءات نيدلسكى وأنجوف وأن هناك أهمية كبيرة تظهر فى التماثل فى البنود وطبيعة حدود مقياس الاستدلال لنيدلسكى لتكوين أساس الاستطلاع بجدية لملائمة الإجراءات.
(Robert L, Brennan, 1980, 218 – 240)

وتشير دراسة جرجورى و نورمان Hurtz , Gregory M.; Hertz, Norman R. (1999) إلى أن تقييم استخدام طريقة أنجوف لتقدير (8) امتحانات لتراخيص مهنية مختلفة من خلال النظرية المعممة أو العامة لتقدير العدد الأفضل المتعلق ب)الدرجة أو المعدل أو القيمة) حيث أشارت النتائج إلى أن من % 15 – 10من الدرجات حققت الهدف الأفضل.
(Hurtz , Gregory M.; Hertz , Norman R., 1999: 885-897)

ويذكر كل من (Impara , James C.; Plake, Barbara, 1998) أن معلموا الصف السادس وعددهم 26 قاموا بتقدير بنود الأداء المتعلقة بطلابهم والبالغ عددهم (724) على 50مفردة وقسموا على اختبار العلوم وحرص المعلمون على الدقة فى تقدير الأداء المتعلق بالمجموعة ككل من المجموعة الواقعة على الخط الفاصل Borderlinولكن وجدوا أنة لا توجد حالة واحدة ذات درجة عالية، وأن تقدير النسب الذى تستخدم فيه إجراءات أنجوف لوضع المعيار ربما تكون موضع شك :1998, Impara, James C.; Plake, Barbara) . (69 – 81

وتشير دراسة لى شانج (Chang , Lei. et al., 1996) إلى تأثير معلومات ومعارف المحكمين حول وضع المعيار المتعلق باختبارات الكفاءات وقد درست مع 17محكم على من يأخذ شهادة فى علم الاقتصاد يخضع لاختبار بينما وضع معايير الكفاءات competency يستخدم إجراءات أنجوف بالبنود والذين أجابوا إجابة صحيحة وفئة معايير منخفضة للذين أجابوها إجابة غير صحيحة (Chang , Lei. ; And Others: 1996, 161 – 173))

ولتقدير فنون ومهارات إتقان اللغة الإنجليزية فى مرحلة المراهقة المبكرة لتوضيح أو إظهار درجة تعقد القرارات الخاصة بالمشاركين فى لجنة التحكيم والذى يجب أن يفعلوه بإتقان ولقد تم عمل بروفيل لتحديد أو وضع معيار لتقدير الأداء المركب بصورة مفصلة وتم استخدام طريقة أنجوف وتم ذلك من خلال مجموعة من الخطوات وبعض الإجراءات العقلانية للمحكمين(Putnam, Sarah E et al., 1995: 57 – 83) .

وتذكر باربرا . (Plak, Barabara S, (1995) أن هناك ثلاث مداخل لتحديد أو وضع المعيار فى هذا الإصدار الخاص تم تلخيصه ومعرفة وجهة الاختلاف عند مقابلته بشيء آخر وهى:

- سياسة جذب أنتباة المحكمين Judgmental Policy Capturing.
- طريقة البروفيل السائد.
- العمل وفق طريقة أنجوف.

وهناك توصيات تالية بتعديل أو تطوير هذه الطرق.
(Plak, Barabara S., 1995: 85-92)

وبالنسبة لتوزيع ذى الحدين وجد أن هناك كثير من التطبيقات المتعلقة بذات الحدين Binomial Coefficients لأن معدل الاتساع الكبير المتعلق بالعلاقات المتداخلة تكون ممكنة لحساب كمية كبيرة متعلقة بالجهود الرياضية تكون ضائعة أو مبددة فى برهنة أو إثبات صيغ التكافؤ الجوهرية، وكذلك ثبات تطبيقات الطرق المستخدمة فى القاعدة خارج الحل المغلق لمجاميع محددة حيث وصفت هذه المقالة طرق متنوعة لإنتاج مجموع صفوف مثلث باسكال Pascal's triangle تم بحثه واستطلاعه وتم عمل جزء خاص به فى متسلسلات القوى. (Ollerton, Richard L., 2007: 124 – 127)

ويشير بيرتن ريتشارد (Burton, Richard F., (2006 إلى أن هناك اختبارات أكاديمية كثيرة (بمعنى إجابات قصيرة والاختيار من المتعدد) وعينة بسيطة من الأسئلة تقديراتها 1، (صفر) أو ما يسمى التقدير المنقسم إلى قسمين وأن هناك بعض الكتب المدرسية Text – booksتعطى إرشادات مفيدة على احتياج الاختبار الطويل وأنة يحتاج إلى درجة من الثبات ويوفر نموذج بوزيس لذات الحدين 1932) ,Posey's Binomial) للتزويد بنقطة البدء الأفضل ولكن هناك تباين خاص بالسؤال ومعرفة قوته التميزية وذلك لمعرفة توزيع معرفة الطلاب المتفاوتة حتى مع اخذ هذا فى الحساب ظهر فى الاختبارات التى تتراوح من – 30 60مفردة، ويجب أن يكون طول الاختبار مناسب حيث إن الاختبارات الطويلة والتى يبلغ عدد مفرداتها 300 مفردة يقلل من درجة ثباتها.

(Burton, Richard F, 2006: 569 – 582)

ولقد ذكر والتير 2007) ,Walters , Gleen,) إلى وجود أربع عينات استخدمت فى حساب الزيادة فى الصدق المتعلق بالقائمة النفسية لأساليب التفكير الجنائى (PICTS) والتفكير الإجرامي العام (GCT) وتم استخدام مجموعة مقاييس التفاعل وجها لوجه – VIS A – VIS واستخدام أسلوب الاستجابة فى التنبؤ بالتوافق الأنضباطى والعودة إلى الإجرام وتم استخدام أسلوب أو احتمال ذات الحدين Binomial لتحليل الانحدار واستخدام مقاييس Pحيث ظهر الصدق التجريبى التدريجى النسبى فى المجموعات الكبيرة المشاركة.

(Walters , Glenn , 2007: 35 – 43)

ويذكر ليفين دوجلاس (Levine, Douglas W., 2006) إلى أننا يجب أن نركز على المشكلة المتعلقة بإهمال الاستقلال الاحصائى وأن معادلة ذات الحدين تستخدم فى تحديد ما إذا أراد المحكمين أن يطابقوا استنادا إلى نظرتهم المنفردة للتصميم التجريبى وحساب احتمال إعطاء محكم المطابقة الصحيحة وتعقيب تغيراتهم من تحكيم إلى آخر أو من اختبار إلى آخر ونحن نعرض احتمال مطابقة النجاح لصاحب التعقيب. وهكذا يغير عدد الناجحين

المطابقين وتوقع تباين الكمية فى النهاية يتم عرض افتراض الفشل المتعلق باستقلال النتائج لاحتمال تغيرات الخطأ وتغيرات احتمال خطأ رفض الفرض الصفرى (بمعنى خطأ من النوع الأول (Type I Error.) (Levine , Douglas W., 2006: 44 – 48)

وقد أجرى صلاح الدين محمود علام (1985) دراسة عن استخدام النموذج ذى الحدين فى تقدير درجة القطع لاختبار محكى المرجع على أساس بيانات تجريبية، حيث اشتمل الاختبار المرجعى المحك الذى أعده على سبع مفردات من نوع الاختيار من المتعدد ويقيس مهارة صياغة الأهداف السلوكية، حيث كون توزيعا نظريا احتماليا للنسب المئوية لأعداد الطلاب الذين يمكن اعتبارهم متمكنين وفقا لدرجات قطع مختلفة، مستخدما فى ذلك نموذج ذى الحدين، وتكونت عينة البحث من 154 طالبا من طلاب الفرقة الرابعة بكلية التربية جامعة الأزهر واختبر الباحث مطابقة التوزيع الواقعى للنسب المئوية للاعتمادا على الدرجات التى حصلوا عليها بالفعل فى الاختبار بهذا التوزيع النظرى الاحتمالي عند مستوى 70%، ومن أهم نتائج الدراسة أنة لا توجدفروق ذات دلالة إحصائية بين كل من التوزيع النظرى الاحتمالي والتوزيع الواقعى . (صلاح الدين محمود علام، 43 – 26 :1985)

ومن ثم فأنة يمكن صياغة مشكلة البحث الحالي فى التساؤلات التالية:

- ما فاعلية طريقة أنجوف لتحديد درجة الفصل لاختبار تحصيلى فى مادة علم النفس التربوى؟
- ما فاعلية نموذج ذى الحدين لتحديد درجة الفصل لاختبار تحصيلى فى مادة علم النفس التربوى؟
- أى النموذجين افضل لتحديد درجة الفصل (النموذج المتصل أم المنفصل)؟

أهداف البحث:

- يهدف البحث الحالي إلى دراسة فاعلية كل من (طريقة أنجوف –ذى الحدين) فى تحديد درجة الفصل لاختبار مرجعى المحك فى مقرر مقاييس الذكاء والقدرات.
- يهدف البحث الحالي أيضا إلى المقارنة بين الطريقتين لمعرفة أى النموذجين افضل لتحديد درجة الفصل النموذج المتصل أم النموذج المنفصل.

أهمية البحث:

يستمد البحث الحالي أهميته من تناوله لموضوع تحديد درجات الفصل حيث يذكر محمود منسى (2013) أنه لتلافى مشكلات عدم التزام الامتحانات التى تعقد على المستوى المحلى أو على مستوى المدرسة بمستويات محددة للأداء يتساوى فيه جميع المتعلمين بجميع المناطق التعليمية فيمكن اقتراح إعداد اختبارات للاستعدادات المدرسية Scholastic

Aptitude Tests وتعقد على المستوى القومى لانتقاء الطلاب للالتحاق بمراحل التعليم المختلفة، فهذا النوع من الاختبارات هو من النوع المرجع إلى مستوى أو إلى محك. وإذا أحسن إعداد مثل هذه الاختبارات وتحديد درجات فاصلة لها Cut Off Point على المستوى القومى فأنها تناسب قياس مدى تحقيق أهداف التعليم بكل مرحلة دراسية.

ومعامل الصدق الأكثر ارتفاعا هو الأكثر دقة فى التنبؤ والدقة التبؤية للاختبار Predictive Accuracy تأثر أيضا بعدد من العوامل الأخرى متضمنة الأخطاء الزائفة الموجبة والسالبة واجتياز النسبة القاعدية وتحدث الأخطاء الزائفة الموجبة إذا كانت الدرجة الفاصلة للاختبار منخفضة للغاية وهنا يحدث قبول غير صحيح أو إيجابية زائفة أى أن بعض المفحوصين يتم اختيارهم للمهنة أو البرنامج الدراسة ولكنهم غير قادرين على النجاح فيها، والعكس عندما تكون الدرجة الفاصلة مرتفعة للغاية كما هو الحال الأخطاء الزائفة السالبة، فأن بعض الأفراد لا يقبلون فى المهنة أو برنامج الدراسة بالرغم من ارتفاع قدراتهم فى المهارات التى يقيسها الاختبار وفى هذه الحالة تكون الدرجة سلبية زائفة.

وعندما تكون الدرجة الفاصلة التى تستخدم للاختبار أو الانتقاء تمثل صعوبة فى الحكم على المتقدمين فأن الدرجة الفاصلة المتعددة تكون فائدة فى الاختبار بصفة عامة لأن الدرجة الفاصلة تتأثر بأى أدوات أو معلومات أخرى متاحة مرتبطة بدرجات الاختبار والدرجات الفاصلة المتعددة هى مجموعة من درجات المقاييس أو الاختبارات المختلفة تستخدم فى اتخاذ قرار التصنيف أو الانتقاء وهى عبارة عن درجات فاصلة منفصلة لكل مجموعة مقاييس وعلى المتقدم أن يصل إلى الدرجة الفاصلة فى كل مقياس أو اختبار على حدة يحصل على العمل أو حتى يمكن اختياره، وطريقة الدرجات الفاصلة المتعددة اكثر ملائمة فى مواقف التنبؤ عندما تكون الدرجة المرتفعة فى أحد المتغيرات لا تعوض الدرجة المنخفضة فى متغير آخر، وعلى سبيل المثال فأن القدرة على السمع جيدا تكون مهمة جدا ومؤثرة على أداء العازف فى فرقة موسيقية محددة، وبغض النظر عن قدرات المرشح العقلية والمهارية فأن الشخص ضعيف السمع لن يصلح للعمل كعازف فى فرقة موسيقية.) محمود منسى، (125 – 118 :2003)

ولقد اقترح نيدلسكى (1954) وأنجوف (1971) إجراءات متعلقة بإنشاء أو تأسيس درجة قطع رئيسية على تخمينات المحكمين حول الأداء المفضل لأدنى كفاءة للممتحنين على كل بند من بنود الاختبار أو كل مفردة من مفردات الاختبار وتفى هذه الورقة نظرية تعميم القدرة generalizability وتستخدم فى وصف أو تمييز توقع قياس التباين فى نتائج درجة الفصل الناتجة من كل إجراء وتستخدم بيانات الاختبار التجريبى فى توضيح هذا المدخل ومقارنة الإجراءين على اعتبار هذا أيضا يعطى تأثير عدم قبول تخمين بعض الأشياء المتعلقة بثبات

القياس واستقلاليته والنتائج تقترح الفروق بين إجراءات نيدلسكى وأنجوف ربما تكون أنجوف افضل من التناظر الظاهر بصفة خاصة وتقترح نيدلسكى جدية استطلاع المحك Robert . (L. Brennan, 1980: 219–240)

ويشير جيرى وآخرون (Gary A., et al. 1986) إلى أهمية ثبات الاختبارات مرجعية المحك تستخدم تصنيفات الإجادة وعدم الإجادة وأن مقياس الخطأ المعيارى (B) بيتا (Standard Error Of Measurement فى حساب فترات الثقة المتعلقة بتقديرات درجة الفصل. (Gary A., et al., 1986115–125)

وتقترح دراسة سوامى ناثان (Swaminathan, H. 1974) فرض مبدئى متعلق بالاختبارات مرجعية المحك فى الأساس الموضوعى للبرامج التعليمية، حيث قسمت الممتحنين إلى حالات التمكن أو فئات على الغايات أو الأهداف المتضمنة فى الاختبار وتقترح الدراسة الثبات لدرجات اختبار مرجعى المحك يكون معرف فى البنود والتجانس (الاتساق) لصناعة القرار وعملياته من خلال تكرار تطبيق الاختبار على وجه الخصوص والثبات عرف على أنة مقياس قبول مرتفع أو منخفض لتوقع الصدفة بين صنع القرارات حول الممتحنين لحالات المتقنين منهم فى إعادة الاختبار لكل مقياس موضوعى بواسطة اختبار مرجعى المحك.
(Swaminathan H et al., 1974: 263)

وتقترح دراسة ميشيل وسبكوفيك (Michael J. S., 1988) أنة من وصف المعلمون وصناع القرار فى الولاية أن الطرق الحالية للحصول على الثبات ومؤشرات اختبار التمكن ومعامل الاتفاق ومعامل Kappa كابا يكون شاق تماما (أى إجراءات وطرق إحصائية معقدة) فعلى سبيل المثال بعض الطرق تتطلب تطبيق أو استخدام اختبارين والآخر يتطلب تطبيق فردى يتضمن مدخل معقد بإجراءاته الإحصائية ويتطلب الوصول إلى برامج كمبيوتر مناسبة. (Michael J. Subkoviak, 1988: 47)

وتصف مقالة برادلى وروبيرت (Bradley. A Hanson; Robert L. Brennan 2002) إجراءات لتقدير تصنيفات متعددة لتصنيف الثبات والدقة لفئات متعددة وتصنيفاتها لاستخدام البيانات من التطبيق الفردى للاختبار وتقدير تصنيفات الثبات والدقة وتم المقارنة بين ثلاث نماذج للقياس مختلفة وهى:

- متغيرين لنموذج بيتا ذى الحدين.
- أربعة متغيرات لنموذج بيتا ذى الحدين.
- ثلاث متغيرات منطقية (نظرية الاستجابة المفردة).

ويتم استخدام بيانات حقيقية توضح إجراءات التقدير.

(Bradley. A. Chanson.; Robert L. Brennan: 2002, 412 – 432)

ويشير ويليام وروى (William E. Donnoe, Roy P. Amato, 1997) إلى أنة بالنقاش حول طريقة كل من أنجوف ونيدلسكى وايبل Angoff, Ebel ,Nedlsky حول تحديد كل منهم لدرجة الفصل أن طريقة أنجوف تحتاج إلى البيانات التجريبية وتعتبر الطريقة التى تعطى الرخصة للاختبارات المهنية وهناك بعض المقترحات وهى:

- طريقة أنجوف تكون افضل طريقة وهى اجدر بالتفضيل عن طرق نيدلسكى وايبل.
- لا توجد فروق دالة بين إحراز تقدم أو نجاح باستخدام مدخل تقليدى ومعدل لطريقة أنجوف .
- أن التماثل الإحصائي بين المدخل التقليدى والمعدل لطريقة أنجوف دال إحصائياً
- أن الطرق الثلاث يقارن بينها لتحديد درجة المرور فى اختبارات الاختيار من المتعدد Multiple Choiceحيث إن هناك استنتاجات جوهرية توضح لماذا أن محتوى هذه الطرق ينتفع به لوضع نقاط المرور وهى
- نقطة المرور تساعدنا فى تحديد الفروق الفردية لمن يجتاز متطلبات (المعرفة –المهارات –القدرات) (KSAS) (Knowledge – Skills – Abilities) وذلك للوصول نحو أفضل مهام للوظيفة.
- هذه الطرق تستطيع تحديد اقل معيار (الحد الأدنى) الذى يتعلق بالكفاءات والمهارات المرتبطة بالوظيفة.
- تصلح هذه الطرق لاختيار افضل الأفراد المرشحين لوظيفة ما لكى يجتازوا شروط تلك الوظيفة.
- تتطلب عملية الاختيار باستخدام تلك الطرق لتحديد اقل كفاءة مطلوبة إعطاء استجابات صحيحة على مفردات الاختبار.

(William E. Donnoe.; Roy P. Amato, 1997: 1 – 14)

مصطلحات البحث:

١–طريقة أنجوف:

هى طريقة مماثلة لطريقة نيدلسكى لتحديد درجة الفصل حيث تحسب الدرجة المتوقعة للإجابة على كل مفردة من مفردات الاختبار (الاختيار من المتعدد) والفرق هنا أن طريقة أنجوف لا تتطلب من المحكم تحديد الاختيار غير الصحيح بوضوح عن كل مفردة من مفردات الاختبار بل يتطلب من المحكم تحديد إجابة المفردة الصحيحة)الاختيار الصحيح) أى أن المحكم يحدد النسب المئوية للإجابة عن كل مفردة وتكون مجموع النسب المئوية تكون درجة الفصل. (William E, Donnoe, Roy P. Amato, 1997: 1 – 14)

وقد أصبحت طريقة أنجوف من افضل الطرق المعروفة وهى غالبا تستخدم كمدخل لتحديد المعيار أو وضع المعيار. , Berk , 1986 , Kane , 1994 ; Van der Linden (1982)، وقد تم اخذ رأى كل محكم عن تصوره عن الممتحنين test – takerوهى طريقة يتم من خلالها الحكم على قوة السؤال أو السؤال المفاهيمى وذلك لكل مفردة والتعرف على نسبة الممتحنين الذين يعطوا إجابة صحيحة على السؤال ويتم حساب ذلك بالنسبة لكل محكم وتكون درجة الفصل هى مجموع الاحتمالات)احتمال الإجابة على السؤال) أى أن درجة الفصل هى متوسط المحكمين على نتائجهم الفردية، وهناك تعديلات لأنجوف تتضمن تغاير فى تعليمات جديدة للمحكمين، ومنها أنة على المحكم تقدير Estimate النسبة المئوية المتعلقة بأقل الأفراد كفء ممن يعرفون الإجابة (Bemknof, Curry & Bashaw, 1979; 1976)

بالإضافة إلى عملية الحكم على بناء المفردة وذلك عن طريق أن يعطى كل محكم إشارة إلى أى مدى يكون الممتحن قادر على الإجابة الصحيحة على مفردات الاختبار وأيضا يحدد المحكمين الحقيقى للمفردات (P- values) لكل مفردة ويتم حساب الوسيط المتعلق بقيم المحكمين وتكون هى درجة الفصل. , Bushch & Jaeger , 1990 ; Cross , Impara) Frary ,& Jaeger , 1984 ; Jager , 1982 ; Norcini, Lipner , Longdon & (Strecker , 1987 ; Norcini,Shea ,& Kanya , 1988

(Laura D. Goodwin, 1996: 249 – 256)

إجراءات طريقة أنجوف:

- تكتب مفردات الاختبار فى قائمة.
- يتم اختيار مجموعة من المحكمين فى مجال الاختبار للقيام بعملية التحكيم.
- يطلب من كل محكم أن يقرأ كل مفردة باهتمام وأن يحدد قيمة احتمالية لها، وهذا الاحتمال يمثل تقدير أن الممتحن ذو اقل مستوى مقبول من الكفاية سوف يجيب على المفردة إجابة صحيحة، أو أن يتخيل المحكم مجموعة كبيرة من الأفراد ذوى اقل مستوى مقبول من الكفاية، ثم يحدد النسبة التى يمكن أن تجيب على المفردة من هذه المجموعة إجابة صحيحة) (176 – 175 :1984 , Shepard.

وقد أجرى أنجوف 1967)) تعديلات على طريقته بحيث تيسر تقدير القيم الاحتمالية لمفردات الاختيار من المتعدد، لذلك يمكن تصميم استمارة تشتمل على التعليمات، وأرقام مفردات الاختبار، وتوضع أمام كل منها الحرف المناظر للإجابة الصحيحة للمفردة داخل مربع. وكذلك ميزان يشتمل على عشر فترات متساوية تقريبا =1)صفر 0.10 -، – 0.11 0.20، 0.30 -0.21،.....، 0.100 – 0.91)، ثم يطلب من كل محكم تقدير احتمال أن يعرف

الفرد الذى يمتلك الحد الأدنى من الكفاية المطلوبة إجابة المفردة دون أن يلجأ إلى التخمين، وأن يضع دائرة حول الرقم الذى يحتوى هذه القيمة على الميزان المعطى. أى أن هذه الطريقة تتطلب من كل محكم التنبؤ بمستوى صعوبة كل مفردة فى الاختبار لمثل هذا الفرد.
(صلاح الدين محمود علام، 240 -239 :1995)

ولتقدير درجة الفصل بطريقة أنجوف يقوم الباحث بتجميع تقديرات المحكمين لجميع مفردات الاختبار مرجعى المحك والتى تمثل درجة الفصل التى حددها كل محكم للاختبار ككل، ومن ثم يتم إيجاد متوسط درجات الفصل لجميع المحكمين ويكون هذا المتوسط هو درجة الفصل النهائية للاختبار التحصيلى.

٢- نموذج ذى الحدين:

هو من ابسط النماذج الرياضية المنفصلة ويمكن استخدامه للحصول على تقديرات لدرجات التمكن بعد تصحيحها من الأخطاء الناجمة عن التخمين. (صلاح الدين محمود علام، 31 :1985)

ويتطلب استخدام هذا النموذج تحقق بعض الشروط فى البيانات المستمدة من الاختبار وهى:

- أن يشتمل الاختبار على عينة عشوائية من المفردات الممثلة للنطاق السلوكى المطلوب قياسه.
- أن تكون درجة كل مفردة من النوع الثنائى)صفر، 1) .
- أن تكون كل مفردة من الاختبار مستقلة إحصائيا عن غيرها من المفردات، بمعنى أن إجابة أى مفردة لا تؤثر فى احتمال إجابة مفردة أخرى بالاختبار.

(صلاح الدين محمود علام، 34 :1985)

٣- درجة الفصل:

وقد عرفها الباحث على أنها "تلك الدرجة التى يحدد عندها الحد الأدنى من مستوى الأداء المقبول لاجتياز الفرد مهمة تعليمية معينة والانتقال إلى المهمة التالية لها".

- **الاختبار مرجعى المحك:**

"يرى جليزر Glaser ونيتكو Nitko أن مفهوم المحك يعنى نطاق شامل من المعارف والمهارات المحددة تحديدا بحيث يمكن نتيجة لموازنة أداء الفرد فى الاختبار بهذا النطاق أن نعرف ما يستطيع أن يؤديه هذا الفرد وما لا يستطيع أن يؤديه" (صلاح الدين محمود علام، 23 :1995)

عينة الدراســة:

يتمثل مجتمع الدراسة من طلاب درسوا مقرر الذكاء والقدرات فى الفصل الدراسى الثاني (2023/2022) من العام الهجرى 1443هـ/ 1444 هـ حيث بلغ قوام العينة 100 طالب وطالبة وقد قام الباحث بتدريس بعض موضوعات المقرر لمدة ست أسابيع وهى (مفاهيم اساسية في القياس النفسي، الفروق الفردية في القدرات العقلية،انواع القياس النفسي ومجالات القياس النفسي،الاساس النظري للقياس النفسي، المناحي النظرية لعملية القياس، اخلاقيات استخدام الاختبارات النفسية، اختبار بينية،اختبار وكسلر، اختبارات الشخصية، التطبيقات التربوية للاختبارات النفسية) بواقع ثلاث ساعات أسبوعيا.

-أداة القياس:

تتمثل اداة الدراسة فى الصورة الأولية للأسئلة التى تقيس الأهداف السلوكية لمقرر مقاييس الذكاء والقدرات المقرر على الطلاب بقسم علم النفس بكلية الآداب بجامعه الطائف، وسوف يقوم الباحث بتحديد درجة الفصل باستخدام كل من (طريقة أنجوف –نموذج ذى الحدين) وذلك لكى يحدد الباحث المستويات المختلفة للإتقان باعتبار أن الطرق الإحصائية هى محكات نحدد فى ضوئها مستوى الإتقان هذا.

وقد اتبع الباحث عند صياغته للصورة الأولية للأسئلة الخطوات الآتية:

- تحديد النطاق السلوكى التى تقيسه الأسئلة وهذا يتضمن:
  - تحديد محتوى المقرر الدراسى المراد قياسه وتحليله إلى عناصره الأساسية.
  - تحديد الأهداف السلوكية لعناصر المقرر الدراسى.
- بناء مفردات الاختبار التحصيلى وهذا أيضا يتضمن:
  - إعداد جدول مواصفات.
  - صياغة المفردات التى تقيس الأهداف السلوكية.
  - صياغة تعليمات الاختبار.
  - تحديد صدق الاختبار.
  - حساب الثبات للاختبار التحصيلى.

وقد اتبع الباحث الطرق التالية لبناء الاختبار التحصيلى وهى:

١- تحديد النطاق السلوكى التى تقيسه الأسئلة وذلك عن طريق تحديد محتوى المقرر الدراسى المراد قياسه وتحليله إلى عناصره الأساسية، حيث أنحصر مجال اهتمام الباحث فى محتوى مقرر مقاييس الذكاء والقدرات والذى يدرس بنظام الساعات المعتمدة لطلاب قسم علم النفس –كلية الآداب بجامعة الطائف، حيث قام الباحث بتحليل محتوى هذا

المقرر وتحديد عناصره الأساسية التى يتكون منها ووجد الباحث أن الموضوعات المقررة على الطلاب للعام الدراسى 1443هـ- 1444 هـ فى الفصل الدراسى الثأنى هى:

الموضوع الأول:

مفاهيم في القياس والتقويم:

- تعريف التقويم، تعريف التقييم، تعريف القياس، العلاقة بين القياس والتقييم والتقويم.

الموضوع الثاني:

–مجاالت القياس النفسي –غايا ت القياس النفسي –خطوات القياس النفسي.

–أنواع القياس مستويات القياس النفسي –الاختبارات –مناح التعامل مع الظاهرة النفسية – المجالات التي تقيسها الظاهرة النفسية –أغراض القياس.

–مشكلات تواجهنا في القياس تطوير الاختبارات والمقاييس.

الموضوع الثالث:

–أخطاء القياس – طبيعة القياس النفسي والتربوي –مبادئ القياس النفسي.

–الأساس النظري للقياس النفسي.

–المناحي النظرية لعمليه القياس –منطق القياس – أخلاقيات استخدام الاختبارات والمقاييس.

الموضوع الرابع:

الفروق الفردية والذكاء.

مقاييس الذكاءوالقدرات اختبار ستانفوردبنيه.

الموضوع الخامس:

– اختبار وكسلر للذكاء.

– اختبارات القدرات العقلية والاستعداد.

الموضوع السادس:

– اختبارات الشخصية.

– اختبارمينيسوتامتعدد الألوجه.

الموضوع السابع :

–المقاييس الاسقاطية.

– التطبيقات التربوية لمقاييس الذكاء والقدرات.

وقد ساعد تحديد النطاق السلوكى الذى يغطيه الاختبار التحصيلى فى تحديد الأهداف السلوكية لعناصر المقرر.

- تحديد الأهداف السلوكية لعناصر مقرر مقاييس الذكاء والقدرات الأساسية حيث قام الباحث بتحديد الأهداف السلوكية
- حيث حدد صلاح علام 1986)) بعض الملاحظات الخاصة بكتابة المفردات وهى:

١-**اختيار نوع المفردات** المناسبة لقياس الأهداف، مفردات الاختيار من المتعدد والصواب والخطأ والمزاوجة تصلح لقياس الأهداف ذات المستويات المعرفية الدنيا مثل (التذكر-الفهم-التطبيق) بينما تصلح مفردات المقال لقياس المستويات المعرفية العليا.

علما بأن:

- **التذكر**: قدرة الطالب على استدعاء)استرجاع) المعلومات أو يتعرف عليها.
- **الفهم:** أن يعيد الطالب صياغة المعلومات أو يقدم لها وصفا باستخدام ألفاظ من عنده.
- **التطبيق:** أن يطبق الطالب قاعدة أو مبدأ فى حل مشكلة معينة.) ناجى محمد قاسم الدمنهورى، 3 :2000)

٢- **تحديد عدد مناسب من المفردات،** حيث يمثل ذلك متغيرا مهما يؤثر فى ثبات الاختبار وقدرته على التطبيق، فكلما زاد عدد مفردات الاختبار زاد معامل ثبات الاختبار وفى الوقت نفسه يؤدى طول الاختبار إلى ملل المختبرين مما يؤثر على أدائهم فى الاختبار. (صلاح الدين محمود علام، 40 :1986)

٣-بناء مفردات الاختبار التحصيلى:

ويتضمن الخطوات التالية:

- إعداد جدول المواصفات.
- صياغة المفردات التى تقيس الأهداف السلوكية.
- صياغة التعليمات.

أولاً- إعداد جدول المواصفات:

يتم إعداد هذا الجدول بعد الفحص الدقيق لمحتوى المقرر الدراسى المراد قياسه وذلك بعد تحليل المحتوى إلى عناصره الأساسية من ناحية وتحديد مستويات الأهداف السلوكية من ناحية أخري ومن ثم تحديد الأهمية النسبية للموضوعات والأهداف.

وقد قام الباحث بإعداد جدول المواصفات فى ضوء توجيهات الخطوات السابقة وهذا يتضح من خلال استعراض نتائج الجدول التالي:

جدول (١) جدول المواصفات

م	موضوعات المقرر		المستويات المعرفية للأهداف			المجموع
			تذكر	فهم	تطبيق	
1	المفاهيم الاساسية في القياس النفسي	عدد الأهداف النسبة المئوية	1 0.05	–	–	1 0.05
2	مجاالت القياس النفسى –غايا ت القياس النفسى –خطوات القياس النفسى	عدد الأهداف النسبة المئوية	–	1 0.05	1 0.05	2 0.10
3	طبيعة القياس النفسى والتربوي –مبادئ القياس النفسى .الأساس النظري للقياس النفسى المناحى النظرية لعمليه القياس أخلاقيات استخدام الاختبارا ت والمقاييس.	عدد الأهداف النسبة المئوية	1 0.05	2 0.10	–	3 0.15
4	مقاييس الذكاءوالقدرات اختبار ستانفوردبنيه	عدد الأهداف النسبة المئوية	1 0.05	1 0.05	2 0.20	4 0.15
5	اختبار وكسلر	عدد الأهداف النسبة المئوية	–	1 0.05	2 0.10	3 0.15
6	اختبار منسوتا متعدد الاوجهMMPI	عدد الأهداف النسبة المئوية	1 0.05	2 0.10	2 0.10	5 0.25
7	الاختبارات الاسقاطية والتطبيقات التربوية	عدد الأهداف النسبة المئوية	–	1 0.05	1 0.05	2 0.10

وبعد أن قام الباحث بإعداد جدول المواصفات قام بتحديد عدد المفردات التى ترتبط بالأهداف فى موضوعات المقرر موضوع الدراسة الحالية.

ويوضح الباحث فى الجدول رقم (2) عدد الأسئلة أو المفردات المرتبطة بالأهداف والتى سيتكون منها الاختبار الحالي.

جدول (٢) عدد الأسئلة المرتبطة بالأهداف التى يتكون منها الاختبار

م	موضوعات المقرر		المستويات المعرفية للأهداف			المجموع
			تذكر	فهم	تطبيق	
1	الموضزع الاول	مجموع الأسئلة	1			1
2	الموضوع الثاني	مجموع الأسئلة		1	1	2
3	المزضوع الثالث	مجموع الأسئلة	1	2		3
4	الموضوع الرابع	مجموع الأسئلة	1	1	2	4
5	الموضوع الخامس	مجموع الأسئلة		1	2	3
6	الموضوع السادس	مجموع الأسئلة	1	2	2	5
7	الموضوع السابع	مجموع الاسئلة		1	1	2
	المجموع	مجموع الأسئلة	4	8	8	20

***صياغة مفردات الأسئلة التى تقيس الأهداف السلوكية:**

- قام الباحث باختيار الشكل المناسب لأسئلة الاختبار وهى الاختيار من المتعدد.

- وبالرجوع إلى جدول المواصفات السابق والاسترشاد به قام الباحث بصياغة أسئلة الاختبار بحيث تكون مطابقة للأهداف السلوكية التى تقيسها.
- وبعد أن قام الباحث بصياغة أسئلة الاختبار قام بعرض الصورة الأولية على مجموعة من المحكمين المتخصصين فى الميدان للحكم على مدى صحتها العلمية ومن أسلوب صياغتها ومدى تغطيتها للأهداف السلوكية من حيث مستوياتها المعرفية المختلفة بحيث تأكد السادة المتخصصين مما يلى:
  - هل تم صياغة أسئلة الاختبار بشكل علمى وأكاديمي مناسب؟
  - هل شملت أسئلة الاختبار الأهداف السلوكية بمستوياتها المعرفية المختلفة؟

وقد وصل العد النهائى لأسئلة الاختبار إلى (٢٠) مفردة وجميعها من نوع الاختيار من المتعدد، وتشتمل كل مفردة على أربعة اختيارات أو بدائل أحدها هى الإجابة الصحيحة، عدد المفردات التى تقيس كل هدف غير محددة. فبعض الأهداف يتطلب قياسها اكثر من مفردة واحدة.

والجدول التالي يوضح أرقام الأهداف السلوكية بمستوياتها المعرفية المختلفة)تذكر – فهم –تطبيق) وأرقام الأسئلة المقابلة لها.

جدول (٣) مستويات الأهداف السلوكية

رقم السؤال	تذكر	فهم	تطبيق	رقم السؤال	تذكر	فهم	تطبيق
1				11			
2				12			
3				13			
4				14			
5				15			
6				16			
7				17			
8				18			
9				19			
10				20			
				المجموع	4	8	8

فروض البحث:

من خلال استعراض ما سبق من دراسات سابقة وإطار نظرى يمكن صياغة الفرض الرئيسي للبحث وهو "تختلف قيمة درجة الفصل تبعا لاختلاف الطريقة المستخدمة فى حسابها (ذى الحدين، أنجوف).

إجراءات البحث:

قام الباحث ببناء اختبار مرجعى المحك مكون من 20مفردة وقام بعرضه على 9من السادة المحكمين لمعرفة مدى اتفاق المفردة مع الهدف السلوكى أو المجال التى تقيسه (تجانس المفردات) وقد قام بوضع جدول يقوم المحكم بوضع علامة (نعم –لا) أمام كل مفردة وذلك بعد اطلاع المحكم عليه موضحا به توزيع المفردات على الأهداف التى تقيسها ويطلب من السادة المحكمين بإعطاء تقدير (1+) للمفردة إذا كأن متأكداً من أن المفردة تناسب الهدف الموضوعة لقياسه، وأن يعطى تقدير (صفر) للمفردة إذا كأن متأكدا من أنها غير مناسبة لقياس الهدف. ثم قام الباحث بتطبيق الاختبار مرجعى المحك على العينة وهو من نوع الاختيار من المتعدد (مفردة وعليها أربعة بدائل) وقام بتصحيح الاختبار

صدق الاختبار:

استخدم الباحث صدق المحتوى والذى يعتمد على آراء المحكمين وذلك من خلال تحديدهم التجانس والاتساق بين المفردة والهدف الذى يقيس المفردة، وفيما يلى جدول يوضح نسبة اتفاق المحكمين على تطابق المفردة مع الهدف.

جدول (٤) نسبة اتفاق المحكمين على تطابق المفردة مع الهدف

رقم المفردة	عدد مرات الاتفاق	عدد مرات الاختلاف	نسبة الاتفاق (%)	رقم المفردة	عدد مرات الاتفاق	عدد مرات الاختلاف	نسبة الاتفاق (%)
1	9	–	100%	11	9	–	100%
2	9	–	100%	12	8	1	0.888
3	9	–	100%	13	9	–	100%
4	9	–	100%	14	9	–	100%
5	9	–	100%	15	9	–	100%
6	9	–	100%	16	9	–	100%
7	9	–	100%	17	8	1	0.888
8	9	–	100%	18	8	1	0.888
9	8	1	0.888	19	9	–	100%
10	9	–	100%	20	9	–	100%

ومن خلال استعراض نتائج الجدول السابق يتضح تطابق المفردة مع الهدف السلوكى التى تقيسه بدرجة تتراوح بين (100% إلى 88%) .

وعلية بعد تحديد درجة الفصل سوف يقوم الباحث بحساب الثبات بإحدى الطرق التى تناسب تطبيق الاختبار مرة واحدة وهى طريقة معامل الاتفاق لسابكوفياك (1976)

(Subkoviak Coefficient Of Agreement (Pc. صلاح الدين محمود علام، (1995: 285

نتائج البحث:

نتائج الفرض الرئيسى "تختلف قيمة درجة الفصل تبعا لاختلاف الطريقة المستخدمة فى حسابها (ذى الحدين، أنجوف)"، ولاختبار هذا الفرض سوف يقوم الباحث بعرض درجة الفصل الناتجة عن كل طريقة على حدة.

أولاً- طريقة أنجوف:

قام الباحث بحساب متوسط تقديرات المحكمين للمفردات حسب ما هو متبع فى طريقة أنجوف وهذا يتضح من خلال استعراض نتائج الجدول التالي:

جدول (٥) متوسط تقديرات المحكمين للمفردات باستخدام طريقة أنجوف

رقم المفردة	متوسط تقديرات المحكمين للمفردة	رقم المفردة	متوسط تقديرات المحكمين للمفردة
1	68.33	12	72.77
2	70.55	13	72.77
3	67.22	14	60.55
4	63.33	15	63.77
5	71.66	16	62.77
6	73.88	17	59.44
7	66.66	18	68.33
8	67.22	19	62.77
9	68.33	20	59.44
10	62.22	المجموع	1334.788
11	72.77		

ومن ثم تكون درجة الفصل لأنجوف هى % 66.73 = 20÷1334.788 =

أى أنها تقابل عددا من الأسئلة يساوى 13 13.34788سؤال تقريبا.

ثبات الاختبار:

بعد أن حددت طريقة أنجوف درجة الفصل عند 66.73% سوف يستخدم الباحث طريقة معامل الاتفاق لسابكوفياك (Subkoviak Coefficient Of Agreement (Pc), (1976) وهى من طرق تقدير ثبات الاختبارات مرجعية المحك والتى تتطلب تطبيق الاختبار مرة واحدة حيث تستخدم هذه الطريقة نتائج تطبيق الاختبار مرة واحدة ويستخدم سابكوفياك توزيع ذى الحدين فى حساب الثبات وتتلخص الخطوات فيما يلى:

حساب قيمة معامل الثبات بطريقة كيودر ريتشاردسون K21-، وذلك لأن هذة المعادلة فى حالة الحيرة وفى حالة تجنب إشكال معامل الارتباط على أن يكون تصحيح اختبارات هذة

الطريقة 1)، صفر) حيث تعطى درجة واحدة للإجابة الصحيحة وصفر للإجابة الخاطئة. (أحمد الطبيب، 193 :1999)

وبلغت قيمة معامل الثبات (0.512)، حيث بلغ متوسط الدرجات للعينة والتباين على الترتيب (14.37)، 7.872).

وجدير بالذكر أن الطريقة السابقة لحساب الثبات بالنسبة للاختبارات جماعية المرجع لا تصلح تماماً لكى تستخدم مع الاختبارات محكية –المرجع وذلك لأن التباين بين الدرجات على هذه الاختبارات يكون قليلاً وأن هذا النقص فى التباين ينزع إلى تقديم معامل ثبات منخفض. (Thorndike R.M.et al., 1991: 115)

- باستخدام صيغة ذى الحدين يتم حساب مقدار احتمال حصول كل فرد على درجة تساوى أو اكبر من درجة الفصل.
- إيجاد معامل الاتفاق الخاص بتصنيف كل فرد على حدة (ل س ف).

نوجد معامل الاتفاق للمجموعة ككل (ل س) وذلك بحساب متوسط القيم للمجموعة ككل.

قام الباحث بترتيب درجات الأفراد فى الاختبار تنازلياً ثم حدد مستوى 66.73% (0.66 درجة قطع لاجتياز الاختبار أى بعدد من الأسئلة بلغ (13 سؤال أى 20 × 0.667) 13.4 =أى 13سؤال) وذلك بعد نتائج طريقة أنجوف.

ثم قام الباحث بحساب احتمال النجاح P والفشل qلكل طالب مع العلم أن متوسط درجات الطلاب فى الاختبار بلغت قيمته 14.37، التباين بلغت قيمته 7.872وبعد حساب ذلك قام الباحث بحساب مقدار احتمال حصول كل فرد من أفراد العينة على درجة تساوى أو أكبر من درجة الفصل وهى س 13 =سؤال.

وفيما يلى جدول موضح به تسلسل الطلاب وقيمة p , q والقيمة الاحتمالية باستخدام توزيع ذى الحدين

جدول (٦) التوزيع النظرى الاحتمالي بذى الحدين

ترتيب تسلسل الدرجات تنازليا	قيمة p	قيمة q	القيمة الاحتمالية بذى الحدين	ترتيب تسلسل الدرجات تنازليا	قيمة p	قيمة q	القيمة الاحتمالية بذى الحدين
5	0.47	0.53	0.081	15	0.73	0.27	0.854
6	0.50	0.50	0.130	15	0.73	0.27	0.854
7	0.52	0.48	0.173	15	0.73	0.27	0.854
7	0.52	0.48	0.173	15	0.73	0.27	0.854
8	0.55	0.45	0.251	15	0.73	0.27	0.854
8	0.55	0.45	0.251	15	0.73	0.27	0.854
9	0.58	0.42	0.345	15	0.73	0.27	0.854
9	0.58	0.42	0.345	15	0.73	0.27	0.854

ترتيب تسلسل الدرجات تنازليا	قيمة p	قيمة q	القيمة الاحتمالية بذى الحدين	ترتيب تسلسل الدرجات تنازليا	قيمة p	قيمة q	القيمة الاحتمالية بذى الحدين
10	0.60	0.40	0.414	15	0.73	0.27	0.854
10	0.60	0.40	0.414	15	0.73	0.27	0.854
11	0.63	0.37	0.526	15	0.73	0.27	0.854
11	0.63	0.37	0.600	15	0.73	0.27	0.854
12	0.65	0.35	0.600	16	0.75	0.25	0.898
12	0.65	0.35	0.600	16	0.75	0.25	0.898
12	0.65	0.35	0.600	16	0.75	0.25	0.898
12	0.65	0.35	0.600	16	0.75	0.25	0.898
12	0.65	0.35	0.600	16	0.75	0.25	0.898
12	0.65	0.35	0.600	16	0.75	0.25	0.898
12	0.65	0.35	0.600	16	0.75	0.25	0.898
13	0.68	0.32	0.708	16	0.75	0.25	0.898
13	0.68	0.32	0.708	16	0.75	0.25	0.898
13	0.68	0.32	0.708	16	0.75	0.25	0.898
13	0.68	0.32	0.708	16	0.75	0.25	0.898
13	0.68	0.32	0.708	16	0.75	0.25	0.898
13	0.68	0.32	0.708	16	0.75	0.25	0.898
13	0.68	0.32	0.708	16	0.75	0.25	0.898
14	0.70	0.30	0.784	16	0.75	0.25	0.898
14	0.70	0.30	0.784	16	0.75	0.25	0.898
14	0.70	0.30	0.784	16	0.75	0.25	0.898
14	0.70	0.30	0.784	16	0.75	0.25	0.898
14	0.70	0.30	0.784	17	0.78	0.22	0.946
14	0.70	0.30	0.784	17	0.78	0.22	0.946
14	0.70	0.30	0.784	17	0.78	0.22	0.946
14	0.70	0.30	0.784	17	0.78	0.22	0.946
14	0.70	0.30	0.784	17	0.78	0.22	0.946
14	0.70	0.30	0.784	17	0.78	0.22	0.946
14	0.70	0.30	0.784	17	0.78	0.22	0.946
14	0.70	0.30	0.784	17	0.78	0.22	0.946
14	0.70	0.30	0.784	17	0.78	0.22	0.946
14	0.70	0.30	0.784	17	0.78	0.22	0.946
14	0.70	0.30	0.784	17	0.78	0.22	0.946
15	0.73	0.27	0.854	17	0.78	0.22	0.946
15	0.73	0.27	0.854	17	0.78	0.22	0.946
15	0.73	0.27	0.854	18	0.81	0.19	0.975
15	0.73	0.27	0.854	18	0.81	0.19	0.975
15	0.73	0.27	0.854	18	0.81	0.19	0.975
15	0.73	0.27	0.854	18	0.81	0.19	0.975

ترتيب تسلسل الدرجات تنازليا	قيمة p	قيمة q	القيمة الاحتمالية بذى الحدين	ترتيب تسلسل الدرجات تنازليا	قيمة p	قيمة q	القيمة الاحتمالية بذى الحدين
15	0.73	0.27	0.854	19	0.83	0.17	0.987
15	0.73	0.27	0.854	20	0.86	0.14	0.989

وقد بلغت قيمة معامل الاتفاق (0.7376الثبات) وهذه القيمة تساوى تقريبا قيمة المعامل (ل.) الذى اقترحه هامبلتون ونوفيك لتقدير اتساق التطبيق إذا طبق اختباران متوازيات على عينة الأفراد نفسها، والذى استخدم فى تقدير قيمة معامل كابا لسواميناثأن وهامبلتون وألجاينا. (صلاح الدين محمود علام، 293 – 285 :1995)

ثانيًا– نموذج ذى الحدين:

هذا النموذج الإحصائي يمثل التوزيعات الاحتمالية المنفصلة ويعتمد على حساب القيمة التقديرية لدرجة الفرد فى النطاق السلوكى الذى يقيسه الاختبار (أى احتمال نجاح الفرد فى الإجابة على المفردة ويرمز لها بالرمز (p)، وحساب احتمال فشل الفرد فى الإجابة على المفردة ويرمز لها بالرمز (q) مع العلم أن (p = q −1) ويتم حساب قيمة كل من (p , q) وهما بارا مترات توزيع ذى الحدين عن طريق حساب قيمة معامل ثبات كيودر ريتشاردسون K–21 نظرا لأن عدد مفردات الاختبار (20 مفردة).

وقد كون الباحث توزيعا نظريا احتماليا للنسب المئوية للطلاب الذين يمكن اعتبارهم متمكنين وفقا لدرجات قطع 0.70، 0.80، 0.90 ولقد اتبع الباحث بعض الخطوات لبناء هذا التوزيع النظرى الاحتمالي لدرجات الاختبار الحالي والتى بلغ عددها (20) مفردة وهذه الخطوات هى:

- الحصول على توزيع نظرى لاحتمالات حصول الطلاب ذوى مستويات التمكن المختلفة على درجة معينة فى الاختبار دون اعتبار لعامل التخمين.
- الحصول على توزيع نظرى لاحتمالات الإجابة عن عدد محدد من مفردات الاختبار عن طريق التخمين وحده من بين المفردات التى لا يعرف الطالب إجابتها.
- الحصول على توزيع نظرى لاحتمالات حصول الطلاب ذوى مستويات التمكن المختلفة على درجة معينة فى الاختبار بعد تصحيحها من أثر التخمين.

وفيما يلى عرض تفصيلى لهذه الخطوات:

أولاً– توزيع نظرى لاحتمالات حصول الطلاب ذوى مستويات التمكن المختلفة على درجة معينة فى الاختبار دون اعتبار لعامل التخمين.

قد قام الباحث ببناء مصفوفة تمثل صفوفها مستويات التمكن المختلفة وقد نتج عن ذلك مصفوفة رتبتها (21 × 3) حيث يمثل الرقم (3) عدد الصفوف والرقم (21) عدد الأعمدة الخاصة بالمصفوفة، وفيما يلى جدول لتوضيح ذلك :

جدول (٧) توزيع نظرى لاحتمالات حصول الطلاب ذوى مستويات التمكن المختلفة على درجة معينة فى الاختبار دون اعتبار لعامل التخمين

عدد المفردات التى يحتمل إجابتها إجابة صحيحة	مستويات التمكن			عدد المفردات التى يحتمل إجابتها إجابة صحيحة	مستويات التمكن		
	م0.70=	م0.80 =	م 0.90 =				
صفر	0.0	0.0	0.0	11	0.056	0.007	0.0
1	0.0	0.0	0.0	12	0.114	0.022	0.0
2	0.0	0.0	0.0	13	0.164	0.045	0.001
3	0.0	0.0	0.0	14	0.194	0.004	0.008
4	0.0	0.0	0.0	15	0.178	0.174	0.031
5	0.0	0.0	0.0	16	0.130	0.218	0.089
6	0.0	0.0	0.0	17	0.071	0.205	0.190
7	0.001	0.0	0.0	18	0.027	0.136	0.285
8	0.003	0.0	0.0	19	0.006	0.057	0.270
9	0.012	0.0	0.0	20	0.001	0.011	0.121
10	0.031	0.002	0.0	11	0.056	0.007	0.0

وهذه المصفوفة رتبتها 21 × 21)) حيث يمثل العدد الأول عدد الصفوف بينما العدد الثاني عدد الأعمدة ولكى نحصل على توزيع نظرى لاحتمالات حصول الطلاب ذوى مستويات التمكن المختلفة على درجة معينة فى الاختبار بعد تصحيحها من أثر التخمين سوف يقوم الباحث بضرب المصفوفة فى جدول (7) فى المصفوفة فى جدول (8) والضرب هنا تحقق شرطه وذلك لأن شرط الضرب بالنسبة للمصفوفات تحقق وهو أن عدد أعمدة المصفوفة الأولى (21) يساوى عدد صفوف المصفوفة الثانية (21) وناتج الضرب هو مصفوفة عدد صفوفها هو عدد صفوف المصفوفة الأولى وعدد أعمدتها هو عدد أعمدة الثانية وهذا يتضح من خلال استعراض نتائج الجدول السابق. وقد حصل الباحث على المصفوفة السابقة بإستخدام نموذج ذى الحدين على أن تكون القيمة الاحتمالية .

جدول (٩) توزيع نظرى لاحتمالات حصول الطلاب ذوى مستويات التمكن المختلفة على درجة معينة فى الاختبار بعد تصحيحها من أثر التخمين

عدد المفردات التى يحتمل إجابتها إجابة صحيحة	مستويات التمكن		
	م0.70=	م0.80 =	م 0.90 =

عدد المفردات التى يحتمل إجابتها إجابة صحيحة	مستويات التمكن		
	م0.70=	م0.80 =	م 0.90 =
صفر	0.0	0.0	0.0
1	0.0	0.0	0.0
2	0.0	0.0	0.0
3	0.0	0.0	0.0
4	0.0	0.0	0.0
5	0.0	0.0	0.0
6	0.0	0.0	0.0
7	0.001	0.0	0.0
8	0.0002	0.009	0.0
9	0.0011	0.003	0.0
10	0.0045	0.004	0.0
11	0.0141	0.080	0.0
12	0.0307	0.041	0.0
13	0.0748	0.151	0.001
14	0.1408	0.263	0.002
15	0.194	0.065	0.011
16	0.248	0.163	0.043
17	0.131	0.193	0.121
18	0.0960	0.225	0.262
19	0.0304	0.126	0.332
20	0.0056	0.370	0.209

جدول (١٠) توزيع نظرى احتمالي للنسب المئوية لأعداد الطلاب الذين يمكن اعتبارهم متمكنين وفقا لدرجات قطع مختلفة للاختبار ومستويات تمكن(0.70) ، 0.80، 0.90)

عدد مفردات الاختبار	درجات الاختبار	درجة الفصل)النسبة المئوية لعدد المفردات المطلوب إجابتها إجابة صحيحة)	مستويات التمكن					
			م 0.70 =		م 0.80 =		م 0.90 =	
			متمكن	غير متمكن	متمكن	غير متمكن	متمكن	غير متمكن
20	20	100%	0.005	0.995	0.037	0.963	0.209	0.791
	19	95%	0.036	0.964	0.163	0.837	0.541	0.459
	18	90%	0.132	0.868	0.388	0.612	0.803	0.197
	17	85%	0.263	0.737	0.581	0.419	0.924	0.076
	16	80%	0.511	0.489	0.744	0.256	0.967	0.033
	15	75%	0.705	0.295	0.809	0.191	0.978	0.022
	14	70%	0.846	0.154	0.835	0.165	0.98	0.02
	13	65%	0.902	0.098	0.85	0.15	0.981	0.019
	12	60%	0.952	0.048	0.876	0.124	0.981	0.019

عدد مفردات الاختبار	درجات الاختبار	درجة الفصل)النسبة المئوية لعدد المفردات المطلوب إجابتها إجابة صحيحة)	مستويات التمكن					
			م 0.70 =		م 0.80 =		م 0.90 =	
			متمكن	غير متمكن	متمكن	غير متمكن	متمكن	غير متمكن
	11	55%	0.966	0.034	0.884	0.116	0.981	0.019
	10	50%	0.97	0.03	0.888	0.112	0.981	0.019
	9	45%	0.971	0.029	0.891	0.109	0.981	0.019
	8	40%	0.971	0.029	0.891	0.109	0.981	0.019
	7	35%	0.971	0.029	0.891	0.109	0.981	0.019
	6	30%	0.971	0.029	0.891	0.109	0.981	0.019
	5	25%	0.971	0.029	0.891	0.109	0.981	0.019
	4	20%	0.971	0.029	0.891	0.109	0.981	0.019
	3	15%	0.971	0.029	0.891	0.109	0.981	0.019
	2	10%	0.971	0.029	0.891	0.109	0.981	0.019
	1	5%	0.971	0.029	0.891	0.109	0.981	0.019

وهذا الجدول قد حصل الباحث علية من خلال الجمع المتتالي للقيم المناظرة للاحتمالات المختلفة لكل مستوى من مستويات التمكن فى جدول (9).

أى إذا كأن مستوى التمكن 0.90 وحددنا درجة الفصل للاختبار بالنسبة 85%، فأننا نتوقع أن تحقق نسبة 92.4% من الطلاب هذا المستوى، بينما إذا خفضنا درجة الفصل إلى 65% فأن النسبة المتوقعة ستكون 98.1%.

ولكى يختبر الباحث حسن المطابقة للتوزيع الواقعى للنسب المئوية لأعداد طلاب العينة اعتمادا على الدرجات التى حصلوا عليها بالفعل فى الاختبار بالتوزيع النظرى الاحتمالي عند مستوى التمكن 0.70 ولإجراء ذلك استخدم الباحث مقياس الدلالة الإحصائية للفرق بين نسبتين متناظرتين مستقلتين عند درجات الفصل المختلفة للاختبا (فؤاد أبو حطب، آمال صادق، 817 – 815 :1996)

وهذا يتضح من خلال استعراض نتائج الجدول التالي:

جدول (١١) نتائج حسن المطابقة للتوزيع الواقعى للنسب المئوية لأعداد الطلاب بالتوزيع النظرى الاحتمالي عند درجات قطع مختلفة (مستوى التمكن 0.70 =)

عدد مفردات الاختبار	درجات الاختبار	تكرار الدرجات الواقعية	النسبة المئوية لتكرار الدرجات الواقعية	درجات الفصل	مستوى التمكن م 0.70 =		الفرق المعيارى
					التوزيع الواقعى	التوزيع النظرى	
20	20	1	0.01	100%	0.01	0.005	0.35
	19	1	0.01	95%	0.02	0.036	0.68
	18	4	0.04	90%	0.06	0.132	1.73

عدد مفردات الاختبار	درجات الاختبار	تكرار الدرجات الواقعية	النسبة المئوية لتكرار الدرجات الواقعية	درجات الفصل	مستوى التمكن م = 0.70		الفرق المعيارى
					التوزيع الواقعى	التوزيع النظرى	
	17	14	0.14	85%	0.20	0.263	1.06
	16	18	0.18	80%	0.38	0.511	1.87
	15	20	0.20	75%	0.58	0.705	1.86
	14	16	0.16	70%	0.74	0.846	1.85
	13	7	0.07	65%	0.81	0.902	1.87
	12	7	0.07	60%	0.88	0.952	1.84
	11	2	0.02	55%	0.90	0.966	1.88
	10	2	0.02	50%	0.92	0.97	1.56
	9	2	0.02	45%	0.94	0.971	1.06
	8	2	0.02	40%	0.96	0.971	0.44
	7	2	0.02	35%	0.98	0.971	0.43
	6	1	0.01	30%	0.99	0.971	1.00
	5	1	0.01	25%	0.100	0.971	1.81
	4	0	0.00	20%	0.100	0.971	1.81
	3	0	0.00	15%	0.100	0.971	1.81
	2	0	0.00	10%	0.100	0.971	1.81
	1	0	0.00	5%	0.100	0.971	1.81
		100					

ويتضح من خلال استعراض نتائج الجدول السابق أن جميع الفروق المعيارية ليس لها دلالة إحصائية عند مستوى دلالة 0.05.

أى عدم وجود فروق ذات دلالة إحصائية بين التوزيع الواقعى والتوزيع النظرى للنسب المئوية للطلاب عند مستوى تمكن 0.70حسب نموذج ذى الحدين وهذا مؤشر على أن مستوى التمكن م 0.70 =هو المستوى المطلوب لاجتياز اختبار مقاييس الذكاء والقدرات.

مناقشة النتائج:

يتضح من نتيجة الفرض السابق أن درجة الفصل الناتجة بكل من طريقة أنجوف ونموذج ذى الحدين مختلفتان وبالتالي يتم قبول الفرض السابق وهذه النتيجة تتفق مع نتائج الدراسات السابقة (دراسة صلاح الدين محمود علام، 1991)، (دراسة عصام الدسوقى 1998)، دراسة سمية شكرى (2006)، (Chang, Lei (1999) والتى وجدت اختلاف درجة الفصل باختلاف الطريقة المستخدمة فى حسابها ولكن هذه الدراسات قارنت الطرق المختلفة ذات التوزيع المتصل ولم تقارن دراسة فى حدود علم الباحث بين طريقة تمثل التوزيع الاحتمالي المتصل (طريقة أنجوف) وبين طريقة تمثل التوزيع الاحتمالي المتقطع أو المنفصل (نموذج ذى الحدين) وعلى الرغم من اختلاف درجة الفصل الناتجة بالطريقتين السابقتين لكن يرى الباحث

أن نموذج ذى الحدين أدق فى تحديد درجة الفصل نظرا لدقة وصعوبة العمليات الحسابية المستخدمة فى حساب وتحديد درجة الفصل بشكل أدق من أنجوف وذلك لأن طريقة أنجوف تتطلب لدقتها توفر ما يلى:

- مدى دقة وخبرة المحكم.
- دقة ووضوح كتاب العلامات المقدم للمحكمين.
- أن يكون المحكم على وعى وإدراك بمعنى الطالب ذى آدني مستوى من الأداء.
- فهم المحكم لطبيعة المقرر.
- فهم المحكم لطبيعة الطالب.

وبالتالي إذا لم يراعى المحكم كل ما سبق يكون التقدير غير موضوعى ومن الممكن التغلب على مشكلة عدم الموضوعية بأن يحكم الفرد كل مفردة مرتين لضمان الاتساق والتوافق بين مرتي التحكيم، على عكس نموذج ذى الحدين الذى يعتمد اعتمادا كليا على تطبيق معادلات رياضية فهى تضمن الحيادية وبالتالي نضمن صدق الدرجة الفاصلة ويتطلب تطبيق المعادلات الرياضة مهارة عالية فى التعامل مع برنامج SPSS وفى إجراء الاختصارات الرياضية المعقدة باستخدام برنامج Derive – 6 Algebra.

التوصيات والبحوث المقترحة:

توصى الدراسة الحالية بعدم الاقتصار على الاختبارات التحصيلية فى حساب درجات الفصل بل يجب الاهتمام بباقى الاختبارات الأخرى لحساب درجة الفصل الخاصة بها مثل اختبارات الشخصية واختبارات القدرات مثال التفكير الناقد والابتكارى، كما توصى الدراسة الحالية بإجراء بعض الدراسات والبحوث وهى:

- فاعلية بعض الطرق الإحصائية لتحديد درجة الفصل فى ضوء متغير سنوات خبرة المحكم.
- فاعلية بعض الطرق الاحصائية لتحديد درجة الفصل فى ضوء متغير عدد مفردات الاختبار.
- فاعلية بعض الطرق الإحصائية لتحديد درجة الفصل فى ضوء متغير حجم العينة.
- المقارنة بين الطرق الإحصائية المختلفة لتحديد درجة الفصل فى حساب الثبات والصدق.

المراجـع

أحمد الطبيب (1999): "التقويم والقياس النفسى والتربوى" الإسكندرية، المكتب الجامعى الحديث.

سوميه شكرى (2006): فاعلية بعض طرق تقدير درجة الفصل فى التنبؤ بالتحصيل اللاحق فى الهندسة للمرحلة الإعدادية "رسالة ماجستير غير منشورة، كلية التربية، جامعة المنيا.

صلاح الدين محمود علام (1985): "استخدام النموذج ذى الحدين فى تقدير درجة القطع لاختبار محكى المرجع، دراسة إحصائية وتجريبية"، المجلة العربية للعلوم الإنسانية، تصدر عن جامعة الكويت، العدد 19، المجلد الخامس، ص ص 43 – 26.

صلاح الدين محمود علام (1986): "تطورات معاصرة فى القياس النفسى والتربوى" إدارة التأليف والترجمة والنشر بجامعة الكويت، الكويت.

صلاح الدين محمود علام (1991): "دراسة مقارنة لبعض طرق تحديد مستويات الأداء فى اختبار مرجعى المحك"، المجلة المصرية للدراسات النفسية، تصدرها الجمعية المصرية للدراسات النفسية، ع1، سبتمبر، ص ص 96 – 77.

صلاح الدين محمود علام (1995): "الاختبارات التشخيصية مرجعية المحك فى المجالات التربوية والنفسية والتدريبية "، القاهرة، دار الفكر العربى.

عصام الدسوقى (1998): "مدى فاعلية نموذج أنجوف فى تحديد المستوى لاختبار محكى المرجع"، مجلة كلية التربية جامعة المنصورة، العدد 36، يناير، ص ص 73 – 43.

عماد عبد المسيح يوسف(1982): "دراسة لبناء مقياس مرجعى الميزان فى مادة الطبيعة بالصف الأول الثانوي"، رسالة ماجستير غير منشورة، كلية التربية جامعة المنيا.

فؤاد أبو حطب و آمال صادق (1996): "مناهج البحث وطرق التحليل الإحصائي فى العلوم النفسية والتربوية والاجتماعية "ط 2، القاهرة، الأنجلو المصرية.

محمود عبد الحليم منسى (2003): "التقويم التربوى "الإسكندرية، دار المعرفة الجامعية.

منار حمد طومأن: (2003): طرق حساب معامل ثبات الاختبار المرجع إلى المحك، دراسة إحصائية مقارنة"، رسالة ماجستير غير منشورة، معهد الدراسات والبحوث، جامعة القاهرة.

ناجى محمد قاسم الدمنهورى (2000) : "فعالية استخدام كل من استراتيجية التعلم التعاوني والتنافسي فى التحصيل الدراسى والاتجاه نحو دراسة مادة الرياضيات لدى طلاب المدرسة الثانوية الصناعية " مجلة كلية الآداب، جامعة المنوفية، العدد 41، ص ص 1 – 45.

Bradley.A.Hanson.; Robert L.Brennan (2002) : "Estimating Consistency And Accuracy Indices For Multiple Classifications",Applied Psychological Measurement ,v26,n4,p412-432

Brandon, Paul R.(2004) : "Conclusions About Frequently Studied Modified Angoff Standard Setting Topics" Applied Measurement In Education.v17 n1 pp 59-88.

Burton,Richard F.(2006) : "Sampling Knowledge And Understanding: How Long Should A Test Be?" Assessment Of Evaluation In Higher Education,v31,n5,pp 569-582.

Chang. Lei (1999) : "Judgment Item Analysis Of The Angoff Standard Settings Methods" Applied Measurement In Education,v12,n2.pp 151-165.

Chang. Lei.; et al., (1996) : "Does A Standard Reflect Minimal Competency Of Examinees Or Judge Competency?" Journal of Applied Measurement in Education, v9, n2, p161-173.

Chang, Lei.; Van Der Linden Wim J.; Vos, Hans J.(2004) : "Setting Standards And Detecting Intrajudge Inconsistency Using Interdependent Evaluation Of Response & Alternatives" Educational And Psychological Measurement ,v64,n5,p781-801.

Ferdous,Abdullah.A.; Plake Barbara S.(2005) : "The Use Of Subsets Of Test Questions In Angoff Standard-Setting Method" Educational And Psychological Measurement, v65, n2 ,p185-201.

20- Ferdous,Abdullah.A.; Plake Barbara S.(2005) : "Understanding The Factors That Influence Decisions Of Panelists In A Standard-

Setting Study" Applied Measurement In Education, v18,n3 ,p257-267.

Gary A. Schaeffer.; Robert E.Carlson,Robin L. Matas(1986) : "Assessing The Reliability Of Criterian-Referenced Measures Used To Evaluate Health-Education Programs." Evaluation. Review, v10, n1, 115-125.

Giraud,Gerald; Impara ,James C.;Plake ,Barbara S.(2005) : "Teachers Conceptions of The Target Examinee In Angoff Standard Setting" Journal Of Applied Measurement In Education,v18,n3, pp 223-232.

Goodwin , Laura D.(1999) : "Relations Between Observed Item Difficulty Levels And Angoff Minimum Passing Levels For A Group Borderline Examinees" Applied Measurement In Education ,v12,n1,p13-28.

Hurtz ,Gregory M.; Hertz , Norman R.(1999) : "How Many Rates Should Be Used For Establishing Cut-Off Scores With The Angoff Method?" h generalizability theory study. ,Journal Of Educational And Psychological Measurement , v59 n6 , p 885-897.

Impara , James C.; Plake, Barbara S.(1998) : "Teachers Ability To Estimate Item Difficulty: A Test Of The Assumptions In The Angoff Standard Setting Method." Journal Of Educational Measurement ,v35 n1,p 69-81.

James Algina.; Douglas Bill Coulson (1978) : "Criterion- Referenced Testing And Measurement: A Review Of Technical Issues And Developments" Review Of Educational Research,v48 n1, p1-47.

Laura d. Goodwin(1996) : "Focus On Quantitative Methods Determining Cut-Off Scores" Research In Nursing & Health, 19,249-256.

28-Leon J.Gross(1985) : "Setting Cut-Off Scores On Credentialing Examinations" Journal Of Evaluation & Health Professions v8,n4,p 469-493.

Levine ,Douglas W.;Rockhill Beverly(2006) : "I Remember You: Independence And The Binomial Model" teaching statistics,

International Journal For Teachers Of Pupils Aged Up To 9,v28n2, p44-48.

Lin , Jie.(2006) : "The Bookmark Procedure For Setting Cut-Scores And Finalizing Performance Standards: Strengths And Weaknesses" Journal of Education Research ,v52,n1, pp36-52.

Michel J. Subkoviak (1988) : "A Practitioner`S Guide To Computation And Interpretation Of Reliability Indices For Mastery Tests" Journal Of Educational Measurement v25, issue1,p47.

Ollerton,Richard L. (2007) : "Partial Row-Sums Of Pascal`S Triangle " International Journal Of Mathematical Education In Science & Technology,v38n1, p124-127.

Plake, Barbara s.(1995) : "An Integration And Reprise: What We Think We Have Learned" Journal Of Applied Measurement In Education, v8n1 ,p85-92.

Plake ,Barbara S.;ImparaJamesC.;IrwinPatrick M.(2000) "Consistency Of Angoff-Based Predictions Of Item Performance: Evidence Of Technical Quality Of Reform Of The Angoff Standard Setting Method" Journal Of Educational Measurement ,v37,n4 pp 347-355.

Putnam , Sarah E.; And Others(1995) : "A Multi-Stages Dominant Profile Method For Setting Standards On Complex Performance Assessment", Applied Measurement In Education ,v8,n1, p57-83.

Reckase, Mark D(2006) : "A Conceptual Framework For A Psychometric Theoy For Standard Setting With Examples Of Its Use For Evaluating The Functioning Of Two Standard Setting Methods" Educational Measurement: Issues And Practice ,v25 n2,p4-18.

Reckase, Mark D(2006) : "Rejoinder: Evaluating Standard Setting Methods Using Error Models Proposed By Schulz" Educational Measurement Issues And Practice ,v25,n3,p 14-17.

Ricker ,Kathryn L. (2006) : "Setting Cut-Scores; A Critical Review Of The Angoff And Modified Angoff" Alberta Journal Of Educational Research ,v52,n1. p53-64.

RobertA.Schaeffer.;Robert E.Corlson.;RobinL.Matas(1986) : "Assessing The Reliability Of Criterion-Referenced Measures

Used To Evaluate Health-Education Programs" Evaluation Review ,v10,n1 ,p115-125.

Robert L.Brennan (1980) ." A Comparison Of The Nedelsky And Angoff Cutting Score Procedures Using Generalizability Theory" Applied Psychological Measurement v4, n2,219-240.

Swaminathan, Ronald H K.Hambleton, JamesAlgina (1974) : "Reliability of Criterion-Referenced Tests: A Decision-Theoretic Formulation", Journal Of Educational Measurement , v11,issue 4, p263.

Shepard, L., (1984): " Setting Performance Standards , IN: Berk, R. (Ed): AGuide To Criterion – Referenced Test Construction, The Johns Hopkins University press, London , pp. 169 – 198.

Thorndike,R.M..; Cuningham,G.K.and, Throndike,R.L.and Hagen ,E.P.(1991) : Measurement and Evaluation in Psychology and Education. Fifth Edition. Maxwell Macmillan International Publishing Group,New York.

Walters, Glenn(2007) : "Reponse Style Versus Crime-Specific Cognition: Predicting Disciplinary Adjustment And Recidivism In Male And Female Offenders With The (PICTS) " Assessment ,v14 ,n1 p35-43.

William E. Donnoe.; Roy P. Amoto(1997) : " Supportive Data & Guidelines For Using The Angoff, Ebel And Nedelesky Cut-Off Score Methods" draft copy to be presented at the international personal management association assessment council (IPMAAC) conference , new port beach. pp 1- 14.