البحث الأول

قرار التصنيف الناتج عن طريقتي أنجوف و ندلسكاى

في تحديد درجة القطع لاختبار محكي المرجع '' دراسة تجريبية

---

* سبق و أن تم نشر هذا البحث فى " مجلة البحث فى التربية و علم النفس" ، كلية التربية جامعة المنيا، المجلد السابع عشر، العدد الثالث ، يناير ٢٠٠٤ .

قرار التصنيف الناتج عن طريقتي أنجوف و ندلسكاى في تحديد درجة القطع لاختبار محكي المرجع '' دراسة تجريبية''

إعداد

دكتور/حجاج غانم أحمد على*

(مدرس علم النفس التربوى – كلية التربية بقنا)

مقدمة البحث:

يعتبر مجال القياس النفسي و التربوي من المجالات المهمة في علم الـنفس، لمـا له من دور كبير في إمكانية قياس الظواهر السلوكية مثل التحصيل و القدرات و الاستعدادات و الاتجاهات و الانفعالات المختلفة،و لأنه قدم الأسس و المبادئ التي على أساسها يمكن قياس هذه الظواهر ، و التي نحتاج إلى قياسها كأخصائيين و تربويين و مهتمين بالمجال.

و إذا كانـت الظواهـر الطبيعـية مثل الطول و الوزن و درجة الحرارة و الضغط يمكن قياسها بأدوات مقننة و ثابتة و بوحدات متدرجة علـى طول المقياس، فان الظواهر السلوكية أيضا يمكن قياسها بأدوات تعد لذلك يطلق عليها الاختبارات و المقاييس النفسية و التربوية.

و يعـتمد تفـسير الدرجات المستمدة من أدوات القياس النفسي و التـربوي علـى أنظمـة مـرجعية مختلفة،و يقصد بالنظام المرجعي *Referenced System* الأساس الذي يبنى عليه التفسير و تقارن في ضوئه الدرجات بحيث يمكن الإفادة من المعلومات المستمدة من هذه الأدوات في الأغراض المختلفة للتقويم (صلاح الدين علام،١٩٩٥،١٣).

و هـناك نظامـان مرجعيان أساسيان يعتمد عليهما تفسير الدرجات المـستمدة من الاختبارات النفسية و التربوية ،و يستخدمان فى صناعة قـرارات تربوية ، النوع الأول يسمى الاختبار المعياري المرجع *(NRT)*

---

*E-mail(hagaggg@yahoo.com)

Norm-Referenced Test و هـذا الـنوع من الاختبارات يمدنا بمعلومات عن الوضع النسبي للفرد بالنسبة لأقرانه الآخرين ، و ذلك بمقارنة أداء الفـرد (درجـته) بمعـيار مستمد من جماعة معيارية مشابهة، و تتحدد أفـضلية الفـرد فـي ضوء ترتيبه النسبي بين أقرانه بغض النظر عن كفائـته، أما النوع الثاني من الاختبارات يسمى الاختبار المحكي المرجع *Criterion –Referenced Test (CRT)* ، و هذا النوع من الاختبارات يمدنا بمعلـومات عـن مـستوى محدد من الكفاية و الإتقان لمهارة معينة أو مجمـوعة من المهارات لدى الطالب، و ذلك بمقارنة أداء الفرد (درجته) بمستوى معين يطلق عليه محك *Criterion (Kubiszyn&Borich,2000,26)*.

ويخـتلف الاختبار المحكي المرجع عن الاختبار المعياري المرجع ، فـي أن هـذا الـنوع من الاختبارات يعد اختبار تشخيصي و يدور حول مجمـوعة محـددة تحديداً جيداً من الأهداف التي يتم اختيارها من جانب المختصين لكي تنطبق في موقف محدد، و المفردات في هذا الاختبار يتم اختـيارها لكي تغطى أهدافاً تعليمية محددة ، غالباً (٣-٥)مفردات لكل هـدف و لهـذا الـسبب يطلق عليها أيضاً *Objective-Referenced Test (ORT)*، كمـا تخـتلف أيضاً عن الاختبارات المعيارية في أن قياس أداء الطالب ينسب إلى أهداف تعليمية محددة بدلاً من الاختبار ككل، و بالتالي يمكـن التعـرف علـى أي الأهداف التي حققها الطالب، و من ثم يمكن الاسـتفادة مـن نتائج الاختبار في توجيه التعليم المستقبلي، أو الأنشطة العلاجية*(Slavin,1997,535)*.

و يهـدف الاختبار المحكي المرجع إلى تحديد الفرد الذي اجتاز محك معين يتطلبه الاختبار من أجل اتخاذ قرار خاص بهذا الفرد سواء تسكينه في مرحلة تعليمية أعلى ،أو التحاقه بمهنة معينة ، وعلى ذلك فانه لا بد بعد تطبيق هذا النوع من الاختبارات أن يكون هناك درجة محددة مسبقاً على أساسها يتم الحكم على أن هذا الفرد اجتاز المحك أم لا ، بحيث إذا

حـــصل الفرد على هذه الدرجة أو ما يزيد عنها يعتبر اجتاز المحك ، أما إذا حـــصل علـــى درجة أقل من الدرجة المحددة يحكم عليه أنه لم يجتاز المحك،و يطلق على هذه الدرجة "درجة القطع" *(Cut Off Score)* .

و تعـــرف درجـــة القطع بأنها الدرجة التي تشير إلى الحد الأدنى اللازم لإتقان *Mastery* مادة معينة أو موضوع معين*(Slavin,1997,535)*.

كما تعرف على أنها الدرجة أو الدرجات التي تفصل بين المتمكنين و غير المتمكنين أو بين مستويات مختلفة من التمكن على متصل درجات الاختـــبار،و يطلق عليها عدة مصطلحات مثل مستويات التمكن، درجات النجاح أو الاجتياز ، الحد الأدنى للكفاية ، مستويات المحك

(صلاح الدين علام،١٩٩١،٧٧).

و هـــناك طرق عديدة لتحديد درجة القطع، و هي طرق إحصائية و لكنها تعتمد بدرجات متفاوتة على أحكام الخبراء *Judgments* في المجال الـــذي يقيسه الاختبار، و كذلك على الدرجات الفعلية للمفحوصين و من هـــذه الطرق طريقة ندلسكاى *Nedlesky* فى عام١٩٥٤،و طريقة أنجوف *Angoff* فى عام ١٩٧١،و طريقة ايبل *Ebel* فى عام ١٩٦٢، و طريقة جيجـــر *Jaeger* فى عام ١٩٨٢، و طريقة أنجوف المعدلة بواسطة بيرك *Berk* فـــى عـــام ١٩٨٤، و طـــريقة المجمـــوعات المحكـــة، و طريقة المجموعات المتناقضة، طريقة المجموعات الحدية

(*Bontempo et al. ,1998*; صلاح الدين علام،١٩٩٥،٢٣٣).

و ســـيتم إلقاء الضوء على طريقتي ندلسكاى و أنجوف لخضوعهما للدراسة فى البحث الحالى.

صـــممت طريقة ندلسكاى لكى تطبق على الاختبارات ذات الاختيار من متعدد لتحديد درجة قطع الاختبار محكى المرجع،و تعتمد هذه الطريقة علـــى الاسترشـــاد بجمع من الخبراء و المتخصصين *Judges*فى المجال الذى يقيسه الاختبار لكي يقوموا بثلاث خطوات هى:

١-تحديد عدد المشتتات فى كل مفردة من مفردات الاختبار و التى سيستبعدها الشخص ذو الحد الأدنى للكفاية من تفكيره عند اختيار البديل الصحيح.

٢-تسجيل مقلوب عدد البدائل المتبقية(فمثلاً لو تم استبعاد بديلين من خمسة بدائل مقلوب عدد البدائل المتبقية يساوى٠,٣٣

٣-تجميع مقلوبات مفردات الاختبار لكى ينتج درجة القطع للاختبار من وجهة نظر المحكم،و بأخذ متوسط هذه الدرجات على كل المحكمين نحصل على درجة القطع للاختبار (صلاح علام،١٩٩٥،٢٣٣).

و الفرد ذو الحد الأدنى للكفاية هو الفرد الذى يحصل على الحد الأدنى المقبول لدخوله ضمن مصاف المتمكنين ،و هو شخصية افتراضية يحددها المحكم و يطلق عليه باللغة الأجنبية *Hypothetical Minimally Acceptable Person* .*(Reckase,2000)*

أما طريقة أنجوف فصممت لكى تطبق على الاختبارات ذات الاختيار من متعدد أيضاً للحكم على درجة قطع الاختبار محكى المرجع،و تعتمد هذه الطريقة على الاسترشاد بجمع من المحكمين *Judges*فى المجال الذى يقيسه الاختبار و هى تقتضى الطلب من كل محكم تقدير احتمال (النسبة المئوية) لإجابة الفرد ذى الحد الأدنى من الكفاية على كل مفردة من مفردات الاختبار إجابة صحيحة، و يتم جمع هذه النسب المئوية لكل المفردات لكى نحصل على درجة القطع من وجهة نظر المحكم ، بعد ذلك يتم أخذ المتوسط لدرجات القطعِ لكل المحكمين لكى نحصل على درجة القطع للاختبار

(صلاح علام،١٩٩٥،٢٣٩؛*Wang,2003,231*).

و مـــن ثم فان هناك طرق عديدة لتحديد درجة القطع ، هذه الدرجة التـــى فـــى ضـــوئها يتم تصنيف *Classification* الأفراد إلى فئتين متقنين *Masters* و هم من يحصلون على درجة القطع أو ما يزيد عنها، و غير متقنـــين *Nonmasters* و هـــم مـــن يحصلون على درجة أقل من درجة القطع،و بالتالي فانه كلما كان هناك صدق فى تحديد درجة القطع ، كلما كان هناك دقة *Accuracy* فى قرار التصنيف.

و لكـــن يتأثـــر قـــرار التصنيف بنوعين من الخطأ يؤثران بشكل ملحـــوظ علـــى دقته، النوع الاول يسمى بالخطأ الموجب *False Positive Error*، الـــنوع الثانـــى يـــسمى بالخطـــأ الـــسالب *False Negative Error* *(Berk,1995,100-101).*

و لفهم هذين النوعين من الخطأ ينبغى معرفة أن هناك نوعان من التصنيف:

الـــنوع الاول:التـــصنيف المحك و فيه يصنف الفرد على أساس البرنامج التعليمـــي الـــذي مـــر به، بحيث إذا مر بالبرنامج التعليمي يحصل على الدرجة (١)، و إذا لم يمر بهذا البرنامج يحصل على الدرجة(٠).

النوع الثانى:التصنيف المنبأ و فيه يصنف الفرد على أساس درجة القطع التـــى تم تحديدها، بحيث إذا حصل الفرد على درجة تساوى أو أعلى من درجـــة القطع يحصل على الدرجة (١)، و إذا حصل على درجة أقل من درجة القطع يحصل على الدرجة(٠)

فالخطـــأ المـــوجب هـــو خطـــأ فـــى القـــرار المتعلق بالتصنيف *Misclassification* ، و يحـــدث عندما يحصل الطالب فى المجموعة غير المـــتعلمة *[أى يأخذ القيمة (٠) فى التصنيف المحك]* ، على درجة القطع أو مـــا يزيد عنه *[أى يأخذ القيمة (١) فى التصنيف المنبئ]* ، و يصنف خطـــأ علـــى أنه متقن،أما الخطأ السالب هو خطأ آخر فى القرار المتعلق بالتـــصنيف *misclassification* ، و يحـــدث عـــندما يحـــصل الطالب فى

المجمـوعة المـتعلمة [أى يأخذ القيمة (١) فى التصنيف المحك]، على درجـة دون درجة القطع [أى يأخذ القيمة (٠) فى التصنيف المنبئ]، و يـصنف خطأ على أنه غير متقن ،و يؤثر هذان النوعان من الخطأ فى صـحة قرار التصنيف *Decision Accuracy* ، و كذلك معامل صدق القرار *Validity Coefficient* ، و دوال المـنفعة *Utility Function* و الـضرر *Disutility Function* و المـنفعة العظمـى *Maximal Utility Function* الـناتجة عـن قـرار التـصنيف ، و يعرف صحة القرار بأنه هو نسبة القرارات الصحيحة *Probability Correct Decisions P(CD)* التي تم اتخاذها فـي التصنيف المنبئ(المبنى على درجة القطع)، أما معامل الصدق فهو عـبارة عـن معامـل الارتباط الرباعى (معامل ارتباط فاى) بين المتغير الثنائى (التصنيف المحك) (١ ، ٠) و المتغير الثنائى (التصنيف المنبأ) (١ ، ٠) ، و فـى ضوء القرارات الصحيحة يتم حساب دالة المنفعة ، كما أنه فى ضوء الأخطاء الموجبة و السالبة يتم حساب دالة الضرر، كما أن دالـة المـنفعة العظمـى تأتى من مزج التقديرات الناتجة عن دالتى الضرر و المنفعة خلال كل الطلاب الخاضعين لقرار التصنيف

. *(Berk,1976,5-8)*

و بـذلك يتـضح مما سبق أن قرار تصنيف الفرد على أنه متقن أو غيـر متقن له مؤشرات و هى (الخطأ الموجب و الخطأ السالب و صحة قـرار التصنيف و معامل الصدق و دوال ا لضرر و المنفعة و المنفعة العظمى).

و الدراسـة الحالية هى محاولة لمقارنة هذه المؤشرات بين طريقتى أنجـوف و ندلسكاى المستخدمتان فى تحديد درجة القطع لاختبار محكى المرجع فى التدريبات المعملية لعلم النفس .

مشكلة البحث:

يستخدم الاختبار محكى المرجع فى تصنيف الأفراد إلى متقنين و غير متقنين، و تعتمد عملية التصنيف على درجة القطع التى تم تحديدها، و التى على أساسها يصنف الفرد على أنه متقن أو غير متقن، بحيث إذا حصل الفرد على درجة القطع أو ما يزيد عنها يعتبر متقناً، أما إذا حصل على درجة أقل من درجة القطع يعتبر غير متقن، و بناءً عليه يتم اتخاذ القرار التربوى المناسب، و بالتالى فان التحديد الدقيق لدرجة القطع يحظى بأهمية كبيرة نظراً لتأثيره فى اتخاذ القرار.

فتحديد مستويات الأداء (درجات القطع) يحظى بأهمية كبيرة لما يتم فى ضوئه من اتخاذ القرارات التربوية المتعلقة بوضع التقديرات الدراسية للطلاب ، و تحديد الناجحين و الراسبين و المتمكنين من مهارات أو كفايات معينة و غير المتمكنين منهم ،و منح الشهادات الخاصة بالطلاب، و بالتالى فان تحديد درجات القطع تؤثر فى التوجهات المستقبلية للطلاب(صلاح الدين علام، ١٩٩١ ، ٧٧).

حيث أن درجات القطع يتم تحديدها لتضع قرارات مهمة جداً عن الأفراد، بالاعتماد على أدائهم على الاختبار،فعلى سبيل المثال ربما يحصل الأفراد على رخصة لمزاولة مهنة بناءً على هذه الدرجة، أو ينتقل الطلاب من مرحلة تعليمية إلى مرحلة أخرى ، و على ذلك فان الطرق المستخدمة فى تحديد درجات القطع ينبغي أن تخضع للتفحيص و التدقيق *(Plake,Impara&Irwin,1999)*.

و قد أثبتت العديد من الدراسات و البحوث أن درجة القطع تختلف باختلاف الطريقة المستخدمة، و من هذه الدراسات ، دراسة *(Berk , 1976)* التي استخدمت طريقة تجريبية فى تحديد درجة القطع تعتمد على رسم منحنيين أحدهما يمثل درجات المجموعة المتعلمة و المنحنى الآخر يمثل درجات المجموعة غير المتعلمة ، و يمثل نقطة تقاطع المنحيين درجة القطع ،و توصلت الدراسة إلى أن درجة القطع المثالية و التى

تقابـــل أفضل مؤشرات للتصنيف، كانت درجة القطع الناتجة عن الطريقة التجـــريبية، و التـــى كانت (٤,٥) و تم تقريبها إلى (٥) لاختبار درجته الكلـــية(٨) ، حـــيث كانـــت مؤشـــرات التـــصنيف هى(القرار الصحيح للتصنيف''٠,٧٤'' والخطأ ا لموجب''٠,١٢'' و الخطأ السالب ''٠,١٤'' و معامل الصدق ''٠,٤٨''و دالة المنفعة ''١,١'' و دالة الضرر''-٠,٣٨'' و دالـــة المنفعة العظمى''١٤٤'')، و هى أفضل من مؤشرات التصنيف التى تقابل درجات القطع و التى تم وضعها بطريقة تعسفية ، و دراسة (صلاح الدين علام،١٩٩١) التى توصلت إلى اختلاف درجة القطع ، باخـــتلاف الطريقة المستخدمة فى تحديد هذه الدرجة ، فقد كانت درجة القطع لاختبار درجته الكلية(٦٠) ، باستخدام طريقة ندلسكاى (٤٠)، و طـــريقة المجموعة الحدية (٤٤)، و طريقة أنجوف(٤٣)، و طريقة المجموعات المتناقضة (٤١) ، و دراسة(نادية عبد السلام ، ١٩٩١) التـــي استخدمت نفس طريقة *Berk* التجريبية فى تحديد درجة القطع ، و توصـــلت الدراســـة إلـــى عدة نتائج منها أن مؤشرات التصنيف (الخطأ الموجب، الخطأ السالب ، القرار الصحيح للتصنيف، معامل صدق القرار، دالـــة المـــنفعة ، دالة الضرر) الناتجة عن الطريقة التجريبية، أفضل من مؤشـــرات التـــصنيف الناتجة عن وضع درجات قطع بصورة تعسفية و ذلك بالنسبة للبعدين الاول و الثالث (مستويات القياس ،و مقاييس النزعة المركزية) من أبعاد الاختبار المحكي موضوع البحث، أما باقي الأبعاد فتم التوصـــل إلـــى عكـــس النتيجة ، و هذا إن دل على شئ فإنما يدل على القـــصور الـــذي يحـــدو الطريقة التجريبية المقترحة فى تحديدها لدرجة القطع ، و دراسة*(Chang,1996)* التى توصلت إلى اختلاف درجة القطع باخـــتلاف الطـــريقة المطبقة فكانت ٧١%، باستخدام طريقة أنجوف ، و ٥٧% باســـتخدام طريقة ندلسكاى ، كما توصلت الدراسة إلى أن طريقة ندلـــسكاى أعلـــى ثبات داخلى بين المحكمين مقارنة بطريقة أنجوف ، و

دراســة(*Donnoe&Amato,1997*) التى توصلت إلى اختلاف درجة القطع باختلاف الطريقة المطبقة فكانت ٥٠% استخدام طريقة أنجوف ، ٣٣% باستخدام طريقة ندلسكاى ، و ٦٠% باستخدام طريقة ايبل، كما توصلت الدراســة إلى وجود فروق ذات دلالة إحصائية بين الطرق الثلاثة ، حيث كانت قيمة ف تساوى ٢٧، عند درجات حرية (٢،٢٥)(عدد المحكمين= ٢٨)،كمــا لا تــوجد ارتــباطات بين طريقة ندلسكاى و كل من طريقتى أنجوف و ايبل، بينما يوجد ارتباط بين طريقتى أنجوف و ايبل ، و دراسة (*Bontempo et al.,1998*) التــى توصلت باستخدام التحليل البعدى إلى أن طريقتــى المجمــوعة الحدية و جيجر تؤديان إلى درجات قطع أعلى من طــريقة أنجوف ،بينما طريقة ندلسكاى تؤدى إلى درجة قطع أقل،كما أن عدد المحكمين يؤثر فى تباين درجات القطع. و دراسة(*Shen,2001*) التي توصلت إلى أن نسبة الأخطاء الناجمة عن استخدام طريقة أنجوف فى تحديــد درجة القطع كانت ٢١%، و هى أكثر من نسبة الأخطاء لطريقة خــريطة المفردة لراش و التى كانت ١٢,٥%، كما توصلت الدراسة إلى أن درجــة القطــع المحددة بواسطة طريقة أنجوف أكثر حدة من درجة القطــع المحددة بواسطة طريقة خريطة المفردة ، كما أن طريقة خريطة المفردة تعكس محك مرغوب من محتوى الاختبار مقارنة بطريقة أنجوف ، و دراسة (*Buckendahl ,et al.,2002*) التى توصلت إلى وجود اختلاف بــسيط بــين درجات القطع المحددة باستخدام طريقتى أنجوف و طريقة التــرتيب التــصاعدى لصعوبة المفردات (*Bookmark*) ، فكانت درجة القطــع لطــريقة أنجوف خلال محاولتين ٥١%، ٤٨%، و درجة القطع لطريقة الترتيب التصاعدى لصعوبة المفردات (*Bookmark*)٤٩%، ٥٢% ، كمــا أن الانحــراف المعــياري لــدرجات القطــع الناتجة عن طريقة (*Bookmark*) أقــل مــن الانحــراف المعياري لدرجات القطع الناتجة عن طريقة أنجوف مما يعطي الطريقة الأولى أفضلية،و دراسة(*Wang,2003*)

و التـــى توصلت إلى اختلاف درجة القطع باختلاف الطريقة المطبقة و عـــدد المفردات و عدد المحكمين فكانت درجات القطع ٦٨%،٥١%(١٣ محكم و ٤٣ مفردة)؛ ٦٧%،٤٩%(٦ محكمين و ٧٣ مفردة)؛ ٦٧ %، ٤٨%(٦ محكمـــين و ٧٩ مفـــردة)؛ ٨١ %،٧٢%(٨ محكمـــين و ٥٣ مفردة)، و ذلك بالنسبة لطريقتى أنجوف و خريطة المفردة على الترتيب، كما توصلت الدراسة إلى أن طريقة خريطة المفردة تتيح أعلى معدل ثبات داخلى و أعلى موافقة بين المحكمين مقارنة بطريقة أنجوف

يتـــضح من الدراسات السابقة أن هناك طرق عديدة لتحديد درجات القطع، و تؤدى هذه الطرق إلى درجات قطع متباينة، كما يلاحظ أن معظم الدراسات التى اهتمت بالمقارنة بين طرق تحديد الدرجة الفاصلة لم تأخذ مؤشرات قرار التصنيف(الخطأ الموجب و الخطأ السالب و نسبة القرارات الصحيحة و معامل الصدق و دوال لضرر و المنفعة و المنفعة العظمى) كجوانب فى عملية المقارنة، و لعل هذا يمثل بعداً مهماً لمشكلة البحث ، أمـــا البعد الآخر لمشكلة البحث فيعتمد على الأهمية التربوية التى تحظى بهـــا طرق تحديد درجة القطع ،لأنه باختلاف الطريقة المستخدمة تختلف درجـــة القطع و من ثم تختلف عملية تصنيف الأفراد إلى متقنين و غير متقنين، و من ثم يختلف القرار التربوي المتخذ، و كلما توصلت الطريقة إلى درجة قطع صحيحة كلما أدى ذلك إلى تصنيف صحيح للأفراد المطبق عليهم الاختبار،أما إذا توصلت الطريقة إلى درجة قطع زائفة ستؤدى إلى تصنيف خاطئ للأفراد.

و تتوقف خطورة التصنيف الخاطئ بين المتقن و غير المتقن على مـــا سوف يترتب على ذلك من نتائج،فعندما يصنف المتقن خطأ على أنه غيـــر متقن(الخطأ السالب) سيترتب على ذلك تكليفه بمهارات جديدة كان فى غنى عنها فضلاً عن الأثر النفسى الذى يلحق به،فقد اعتبر غير متقن مـــع أنـــه فـــى الأصل متقن ، و فى حالة العكس أى تصنيف الفرد غير

المستقن على أنه متقن(الخطأ الموجب) ،سيترتب على ذلك أن يحتل هذا الفرد مكاناً أو مركزاً لا يستحقه ،فضلاً عن أنه سوف يغش فى الأغلب و الأعم فى المهارات المتدرجة التالية لأنها تفترض أنه أتقن المهارات التعليمية السابقة مع أنه لم يتقنها فى واقع الأمر

(نادية عبد السلام، ١٩٩٢ ، ٩٢).

يتضح من العرض السابق أن نوع الطريقة التى سيتم اختيارها لتحديد درجة القطع ،له أبعاد تربوية مهمة جداً نظراً لارتباطها بمدى صحة عملية التصنيف،و بالطبع فان عملية التصنيف تكون صحيحة عندما تقل الأخطاء الموجبة و السالبة و دوال الضرر و بالتبعية تزيد القرارات الصحيحة و يزيد معامل صدق قرار التصنيف و تزيد دوال المنفعة و المنفعة العظمى، و من ثم فان المقارنة بين طرق تحديد الدرجة الفاصلة تتطلب أخذ مؤشرات قرار التصنيف كجوانب مهمة فى عملية المقارنة.

لذلك تتحدد مشكلة البحث الحالى فى التعرف على الاختلاف فى مؤشرات قرار التصنيف(الخطأ الموجب و الخطأ السالب و القرار الصحيح و معامل الصدق و دوال الضرر و المنفعة و المنفعة العظمى) بين طريقتي(ندلسكاى،أنجوف) فى تحديد درجة القطع لاختبار محكى المرجع فى التدريبات المعملية لعلم النفس.

مسلمة البحث:

يستند البحث على مسلمة رئيسية و هى أن أى طالب يمر ببرنامج تعليمى فاعل (تابع للمجموعة المتعلمة) من المفترض أن يكون متقن لهذا المحتوى ، و العكس صحيح فان أى طالب لم يمر بهذا البرنامج التعليمي (تابع للمجموعة غير المتعلمة) من المفترض أن يكون غير متقن.

فيشير *berk* إلى أنه اعتماداً على أن اكتساب المعرفة عملية متصلة ، فمن المتوقع أن يكتسب الطلاب فى المجموعة المتعلمة المعرفة التى يعكسها الهدف التعليمى للبرنامج ، و ذلك بعكس المجموعة غير المتعلمة التى تكتسب معرفة أقل ، و لذلك يمكن وصف اكتساب أو عدم اكتساب المعرفة التى يعكسها الهدف التعليمى بمتغير ثنائى حقيقى *A true dichotomy* ، و هو ما يطلق عليه المتغير المحك*(berk , 1976 , 5)* .

تساؤلات البحث:

يحاول الباحث الإجابة على التساؤلات التالية:

١- ما مدى الاختلاف بين طريقتى ندلسكاى و أنجوف فى درجة القطع المحددة ؟.

٢-ما مدى الاختلاف بين طريقتى ندلسكاى و أنجوف فى مؤشرات قرار التصنيف(الخطأ الموجب و الخطأ السالب و القرار الصحيح و معامل الصدق و دوال الضرر و المنفعة و المنفعة العظمى) الناتجة عن كل طريقة؟

أهداف البحث:

يهدف البحث الحالى إلى التعرف على :

١- مدى اختلاف درجة القطع باختلاف الطريقة المستخدمة(ندلسكاى – أنجوف) ؟ .

٢- مدى الاختلاف فى مؤشرات قرار التصنيف(الخطأ الموجب و الخطأ السالب و القرار الصحيح و معامل الصدق و دوال الضرر و المنفعة و المنفعة العظمى) باختلاف الطريقة المستخدمة
(ندلسكاى – أنجوف) ؟ .

أهمية البحث:

١-يتم فى هذا البحث إعداد مقياس محكي المرجع فى التدريبات المعملية لعلم النفس ، مما قد يسهم فى الاستفادة منه فى البرامج التشخيصية أو الدراسات المستقبلية.

٢-يتصدى البحث الحالى لموضوع الدرجة الفاصلة و طرق تحديدها و مؤشرات قرارات التصنيف المتخذة بناءً على هذه الدرجة ، و التى فى ضوئها نتعرف على مدى صحة عملية التصنيف، و هى موضوعات غاية فى الأهمية بالنسبة لعملية تشخيص الأفراد و اتخاذ القرارات التربوية الخاصة بالطلاب.

٣-يتم فى هذا البحث المقارنة بين طريقتى ندلسكاى و أنجوف، و التعرف على الطريقة الأفضل يؤدى إلى دقة عملية التصنيف، مما يؤدى إلى صحة القرار التربوى المتخذ بشأنهم.

الإطار النظري

١-الاختبارات محكية المرجع و الاختبارات معيارية المرجع:

هناك نظامان مرجعيان أساسيان يعتمد عليهما تفسير الدرجات المستمدة من الاختبارات النفسية و التربوية ،و يستخدمان فى صناعة قرارات تربوية ، النوع الاول يسمى الاختبار المعياري المرجع(*NRT*) *Norm Referenced Test* و هذا النوع من الاختبارات يمدنا بمعلومات عن الوضع النسبى للفرد بالنسبة لأقرانه الآخرين ، و ذلك بمقارنة أداء الفرد (درجته) بمعيار مستمد من جماعة معيارية مشابهة، و تتحدد أفضلية الفرد فى ضوء ترتيبه النسبى بين أقرانه بغض النظر عن كفائته، أما النوع الثانى من الاختبارات يسمى الاختبار المحكى المرجع *Criterion –Reference Test (CRT)* ، و هذا النوع من الاختبارات يمدنا بمعلومات عن مستوى محدد من الكفاية و الإتقان لمهارة معينة أو مجموعة من المهارات لدى الطالب، و ذلك بمقارنة درجة الفرد بمستوى معين يطلق عليه محك *Criterion (Kubiszyn&Borich,2000,26)*.

فالاختبار معيارى المرجع(*NRT*) هو ذلك الاختبار الذى تفسر درجاته فى ضوء المعايير التى تستخدمها و التى ترد إليها هذه الدرجات، أى أن الدرجات التى نحصل عليها من هذا الاختبار نرجعها إلى أحد الجداول المعيارية و نستطيع أن نحصل من هذه الجداول على مستوى الأداء النسبى للفرد فى ضوء الدرجة التى يحصل عليها فى الاختبار بالنسبة للمجموعة التي ينتمى إليها هذا الفرد أو بالنسبة إلى جماعته المعيارية(أنور الشرقاوى،١٩٩٦ ، ٢٢).

فهذا النوع من الاختبارات يتم فيه مقارنة الدرجة الحقيقية للتلميذ بمعيار معين من خلاله يمكننا إن نحدد ما إذا كانت الدرجة أعلى أو أقل من متوسط المجموعة المعيارية *Norm* و قد تكون هذه المجموعة ، تلاميذ الفصل، أو تلاميذ المدرسة أو تلاميذ البلدة الذين فى نفس المرحلة العمرية للتلميذ ، و يمكن استخدام الاختبار معيارى المرجع فى أفضل صورة عند قياس قدرة عامة فى مواد معينة مثل اللغة الإنجليزية، الجبر، العلوم العامة،أو عند تقييم مدى القدرات بين مجموع كبيرة من الأفراد(المدى المطلق)، أو عند انتقاء أفضل المرشحين إذا كان المطلوب عدد منهم (*Woolfolk,1995,514-516*).

و بالرغم من أن القياس معيارى المرجع يزودنا بمعلومة خاصة بمقدرة الطالب و ذلك فى ضوء مقارنة أدائه بأداء زملائه، إلا أن هذا النوع من القياس لا يعطينا معلومات عن كيفية براعة هذه الطالب أو تفوقه فى أعمال الموضوع المتضمنة فى الاختبار

(نادية عبد السلام،١٩٩٦:ب، ٩٠).

أما الاختبار محكى المرجع (*CRT*) فهو مبنى بهدف إعطاء معلومات تفسر مباشرة بالنسبة لمحك مطلق للأداء ، و داخل هذا الاتجاه فان التأكيد ينصب على السؤال ، ماذا تعلمه الطالب؟، فالطالب إما أن يستطيع إظهار مهارة خاصة ، أو إعطاء ناتج معين، أو القيام بسلوك محدد أو لا

يستطيع، و بذلك فان الاتجاه محكى المرجع يخدم وظيفتين هامتين جداً داخل المضمون التعليمى هما :توجيه الانتباه إلى سلوك و إنجازات الطلبة ،و بناءً عليه إثابتهم على أساس تحصيلهم بالنسبة للإنجازات المحكية أكثر من إثابتهم على أساس تحصيلهم بالنسبة لزملائهم

(نادية عبد السلام،١٩٩٦:أ، ٤٣).

فميزة التقويم القائم على الاختبارات محكية المرجع هى أننا نعرف نظرياً على الأقل ما يمكن أن يقوم به الطلاب المتقنون ، و نوع المشكلات التى يستطيعون حلها أو المهام التى يمكن أن يقوموا بها

(رجاء أبو علام ، ٢٠٠٤ ، ٤٢٧) .

و تطبق الاختبارات محكية المرجع فى دول كثيرة ضمن البرامج الاختبارية و برامج التقويم التربوى المعاصر ، و يستخدم هذا النوع من الاختبارات فى تقدير أداء الفرد بالنسبة إلى نطاق سلوكى ، أى مجموعة من المعارف و المهارات المحددة تحديداً سلوكياً دقيقاً بصرف النظر عن علاقة أداء الفرد بأداء غيره من الأفراد الذين يطبق عليهم الاختبار نفسه ، أى أن الأداء فى هذا النوع من الاختبارات يكون مطلقاً و ليس نسبياً، و لذلك فإن هذا النوع من الاختبارات يناسب المواقف التعليمية التى تتطلب تحديد درجة اتقان الطالب لمجموعة من المهارات الأساسية المتعلقة بمادة دراسية أو محتوى دراسى معين

(صلاح الدين علام،٣،٢٠٠١).

و يطلق على الاختبارات محكية المرجع اختبارات التفوق أو الإتقان *Mastery* أو اختبارات المقدرة أو الكفاية أو اختبارات المهارات الأساسية و هى مكونة بحيث تسمح بتفسير درجات الاختبار للفرد بالنسبة لمجموعة من الأهداف أو الكفايات المحددة بوضوح

(نادية عبد السلام،١٩٩٦: ب ، ٩٢).

ويختلف الاختبار المحكى المرجع عن الاختبار المعيارى المرجع ، فى أن الاول يعد اختبار تشخيصى و يدور حول مجموعة محددة تحديداً جيداً من الأهداف التى يتم اختيارها من جانب المختصين لكى تنطبق فى موقف محدد، و المفردات فى هذا الاختبار يتم اختيارها لكى تغطى أهداف تعليمية محددة غالباً (٣-٥) مفردات لكل هدف و لهذا السبب يطلق عليها أيضاً *Objective Referenced Test(ORT)*، كما تختلف أيضاً عن الاختبارات المعيارية فى أن قياس أداء الطالب ينسب إلى أهداف تعليمية محددة بدلاً من الاختبار ككل، و بالتالى يمكن التعرف على أى الأهداف التى حققها الطالب، و من ثم يمكن الاستفادة من نتائج الاختبار فى توجيه التعليم المستقبلى، أو الانشطة العلاجية(*Slavin,1997,535*).

و يفيد نتائج الاختبار محكى المرجع المعلم فى قياس الإتقان و المهارات الأساسية، و تحديد ما إذا كان يتوفر لدى التلاميذ المتطلبات الرئيسية للبدء فى وحدة تعليمية جديدة،و تقييم الأهداف الانفعالية و النفس حركية،و تقسيم التلاميذ فى مجموعات للتعليم(*Woolfolk,1995,516*).

و يشير بابام *Popham* فى عام ١٩٧٨،و هدسون و لينخ *Hudson & Lynch* فى عام ١٩٨٤،و براون *Brown* فى عام ١٩٨٩ إلى انه يقصد بالمحك *Criterion* فى الاختبارات محكية المرجع هو السلوك أو المهارة المطلوب اختبارها أو تقييمها(*Davidson& Lynch,1998*) .

و يصاغ أداء المحك على صورة كفايات محددة أو نواتج متوقعة أو أهداف سلوكية مرتبة بحيث تصف مختلف مستويات الأداء

(صلاح الدين علام ، ١٩٩٥ ،١٧).

و بذلك فان الاختبارات هدفية المرجع هى إحدى صور الاختبارات محكية المرجع حيث تبنى على أهداف سلوكية مصاغة مسبقاً فى صورة نواتج تعليمية(صلاح الدين عَلام،١٩٩٥، ٢٤)

و يهـدف الاختـبار المحكى المرجع إلى تحديد الفرد الذى اجتاز محـك معين يتطلبه الاختبار من أجل اتخاذ قرار خاص بهذا الفرد سواء تسكينه فى مرحلة تعليمية أعلى ، أو التحاقه بمهنة معينة ، و على ذلك فانـه لا بـد بعد تطبيق هذا النوع من الاختبارات أن يكون هناك درجة محـددة مسبقاً على أساسها يتم الحكم على أن هذا الفرد اجتاز المحك أم لا ، بحيث إذا حصل الفرد على هذه الدرجة أو ما يزيد عنها يعتبر اجتاز المحك ، أما إذا حصل على درجة أقل من الدرجة المحددة يحكم عليه أنه لم يجتاز المحك،و يطلق على هذه الدرجة "درجة القطع" (*Cut Off Score*).

٢-درجة القطع(*Cut Off Score*):

تفـيد الاختـبارات محكية المرجع فى اتخاذ قرارات تعليمية تتعلق بدرجـة إتقـان الفرد لنطاق سلوكى من المعارف و المهارات المحددة و مـستوى تمكنه من محتوى هذا النطاق، كما تستخدم فى تصنيف الأفراد بحـسب مـستويات تمكـنهم أو إتقـانهم لهذا النطاق، و اتخاذ قرارات التصنيف ينبغى أن يعتمد على تحديد مسبق لمستوى أو مستويات الأداء فى الاختبار،(و هو ما يشير إلى درجة القطع) و هى الدرجة التى تفصل بـين المتمكنين و غير المتمكنين ، أو بين مستويات مختلفة من التمكن على متصل درجات النطاق السلوكى الشامل للاختبار

(صلاح الدين علام،١٩٩٥، ٢٢٦).

و يطلـق علـى درجة القطع مسميات عديدة منها مستوى أو مـستويات الأداء*Performance Standard*،درجات التمكن *Mastery Scores*،درجات النجاح أو الاجتياز *Passing Scores* ،الحد الأدنى للكفاية *Minimum Competency Level*،درجـات المحـك *Criterion Scores* (صـلاح الدين عـلام، ١٩٩١، ٧٧)، و تعـرف درجـة القطـع بأنها الدرجة التـى تشـير إلى الحد الأدنى اللازم لإتقان *Mastery* مادة معينة أو موضوع

معين (*Slavin,1997,535*)، كما تعرف بأنها الدرجة التى تفصل بين المتقنين و غير المتقنين(رجاء أبو علام ، ٢٠٠٤ ، ٤٢٧)، و تعرف أيضاً بأنها الحد الأدنى من الأداء فى الاختبارات محكية المرجع الذى ينبغى أن يحققه الطلاب من أجل رفع المستوى أو نيل الشهادات (صلاح الدين علام ، ٢٠٠٤ ، ١٦).

فمستوى المحك للأداء على الاختبار يمكن أن يعبر عنه بواسطة درجة(٤ من ٥ مفردات) ، أو يعبر عنه بواسطة نسبة مئوية ٨٠ % (و هى تقابل ٤ من ٥ مفردات)،و القيمة ٤ تعبر عن عدد المفردات التى ينبغى أن يجيب عليها المفحوص إجابة صحيحة، و هى فى الواقع تغطى أهداف تعليمية أو مهارات معينة،و يطلق عليها درجة القطع *cut-off score* و هى الدرجة التى تقطع توزيع درجات الأفراد إلى صنفين ، الصنف الاول يحتوى على الدرجات التى يحال أصحابها إلى مستوى التمكن أو الكفاءة *mastery/competency* ، أما الصنف الثانى يحتوى على الدرجات التى يحال أصحابها إلى مستوى عدم التمكن أو عدم الكفاءة *nonmastery/noncompetency*

(berk,1986,138)

كما يشير *Kane* فى عام ١٩٩٤إلى أن درجة القطع يطلق عليها درجة الاجتياز *Passing Score* و هى نقطة على تدريج درجات الاختبار و التى تعبر عن مقياس الكفاءة فى الاختبار (*In:Sizmur,1997,4*).

و هناك العديد من مواقف القياس التى تستلزم استخدام درجة القطع ، و التى فى ضوئها يتحدد من سيجتاز الاختبار ، و من سيفشل ،من سيؤهل أو يحصل على رخصة أو يقبل فى العمل ، أو من سيقبل فى برنامج للدراسات العليا بالجامعة، فهى تفصل بين المتمكن و غير المتمكن ،و هذا الموضوع يرتبط بالاختبارات محكية المرجع، مقارنة بالاختبارات معيارية المرجع. *(Goodwin,1996,249)*

إن درجـــات القطع تصنع قرارات مهمة جداً فى التربية لأنها تصنف الأفراد إلى ناجح/راسب، متمكن/غير متمكن، مؤهل/غير مؤهل،كما يمكن أن تصنفهم إلى مستويات متعددة ،فعلى سبيل المثال يصنف التلاميذ بناءً علـــى درجات القطع إلى مستويات مختلفة أ،ب،جـــ،د،هـــ، و فى ضوء درجـــة القطـــع ينتقل التلميذ إلى مرحلة تعليمية تالية،أو يتم انتقاؤه فى الكلية، أو يتم تسكينه فى فصل دراسى معين، كما أن درجة القطع تصنع قـــرارات مهمة فى سوق العمل ، حيث فى ضوء درجة القطع يتم تقرير ما إذا كان الطالب كفئ لدرجة تؤهله للعمل كمدرس،أو مدير مدرسة ،أو طبيب،أو أى نوع آخر من المهن(*Bontempo et al.,1998*).

٣-طرق تحديد درجة القطع *Standard Setting Methods*:

هـــناك العديد من الطرق التى تستخدم فى تحديد درجة القطع ، و كـــل طـــريقة تـــم مهاجمـــتها و الـــدفاع عنها، من الوجهة النظرية و الامبـــريقية،و لقـــد أثبتت الدراسات السابقة أن الطرق المختلفة لتحديد درجـــة القطع تؤدى إلى درجات قطع مختلفة، و من هذه الطرق طريقة ندلـــسكاى *Nedlesky* فـــى عام١٩٥٤،و طريقة أنجوف *Angoff* فى عام ١٩٧١،و طريقة ايبل *Ebel* فى عام ١٩٦٢، و طريقة جيجر *Jaeger* فى عـــام ١٩٨٢، و طـــريقة أنجـــوف المعدلة بواسطة بيرك *Berk* فى عام ١٩٨٤، و طـــريقة المجمـــوعات المحكـــة، و طـــريقة المجمـــوعات المتناقضة،و طريقة المجموعات الحدية(*Bontempo et al.,1998*) .

و سيتم إلقاء الضوء على طريقتى ندلسكاى و أنجوف المستخدمتين فى الدراسة الحالية:

أولاً:طريقة ندلسكاى *Nedlesky Method* :

سميت طريقة ندلسكاى بهذا الاسم لمؤسسها ندلسكاى و تم تأسيـــسها عـــام ١٩٥٤ و هى صممت لكى تطبق على الاختبارات ذات الاختيار من متعدد للحكم على درجة قطع الاختبار محكى المرجع،و تعتمد

هذه الطريقة على الاسترشاد بجمع من الخبراء و المتخصصين *Judges* فى المجال الذى يقيسه الاختبار لكي يقوموا بثلاث خطوات هى:

(١)-تحديد البدائل فى كل مفردة من مفردات الاختبار و التى سيستبعدها الشخص ذو الحد الأدنى للكفاية من تفكيره عند اختيار البديل الصحيح.

(٢) -تسجيل مقلوب عدد البدائل المتبقية(فمثلاً لو تم حذف بديلين من خمسة بدائل ، يكون مقلوب عدد البدائل المتبقية يساوى٠,٣٣

٣-تجميع مقلوبات مفردات الاختبار لكى ينتج درجة القطع للاختبار من وجهة نظر المحكم،و بأخذ متوسط هذه الدرجات على كل المحكمين نحصل على درجة القطع للاختبار

(صلاح علام، ١٩٩٥ ، ٢٣٣ *;Berk,1986,151)*

و الفرد ذو الحد الأدنى للكفاية هو الفرد الذي يحصل على الحد الأدنى المقبول لدخوله ضمن مصاف المتمكنين ،و هو شخصية افتراضية يحددها المحكم و يطلق عليه باللغة الاجنبية *(Reckase,2000).*

Hypothetical Minimally Acceptable Person، كما يطلق عليه أيضاً *(Plake&Impara,1996)Minimally Competent Candidate.*

و يمكن تعريفه فى البحث الحالى بأنه الطالب الذى يمتلك حد أدنى من المهارات و الكفايات (المعارف و المهارات اللازمة لممارسة تجربة التعلم،إعداد جدول البيانات، رسم منحنى التعلم، التعليق على منحنى التعلم، معرفة التطبيقات التربوية لتجربة التعلم) التى تؤهله لتطبيق مفهوم التعلم عملياً داخل معمل علم النفس.

و تفترض هذه الطريقة أن الأفراد المختبرين ذوى الحد الأدنى من الكفاية المقبولة يمتلكون مستويات مقبولة من المعرفة و المهارات و القدرات و يستطيعون استبعاد البدائل التى تتضح أنها غير صحيحة أثناء التفكير فى اختيار البديل الصحيح،و مهمة المحكمين هى تحديد هذه البدائل *(Donnoe & Amato,1997).*

و أشـــار ندلـــسكاى إلـــى أنـــه لتلافى مشكلة خطأ القياس، يمكن الحصول على الانحراف المعيارى لدرجات القطع الناتجة ، ثم يتم ضرب قـــيمة الانحـــراف المعيارى فى مقدار ثابت يتفق عليه المحكمون و هذا المقـــدار الـــثابت يعتمد على التوزيع الطبيعى لدرجات القطع للاختبار،ثم تضاف القيمة الأخيرة إلى درجة القطع التى تم تحديدها لنصل إلى درجة القطـــع بالطـــريقة المعدلـــة، و قـــيم الـــثابت التى يمكن أن يختار منها المحكمـــون هـــى(٢أو ١) و هى تقابل النسبتين المئويتين(٨٤،٩٨) علـــى الترتيب ، و هذه النسب تعنى نسبة فشل المجموعة الحدية (ذوى الحد الأدنى من الكفاية) فى الوصول لدرجة القطع

(صلاح الدين علام، ١٩٩٥ ،٢٣٤)

غيـــر أن الطـــريقة المعدلـــة تعتـــريها الصعوبة من ناحية فهم المحكمين لها بالإضافة إلى أن معظم الدراسات تستخدم الطريقة الأصلية لسهولة فهمها من جانب المحكمين .

و هـــناك العديـــد من الدراسات التى أجريت على طريقة ندلسكاى و معظـــم هذه الدراسات توصلت إلى انخفاض درجة القطع الناتجة عن هذه الطريقة مقارنة بالطرق الأخرى و من هذه الدراسات

(صلاح الدين علام،١٩٩١؛ *Chang,1996;Bontempo et al.,1998*)

ثانياً:طريقة أنجوف *Angoff Method*

ســـميت طريقة أنجوف بهذا الاسم لمؤسسها أنجوف*Angoff* فى عـــام ١٩٧١ و هى تعد من أشهر الطرق*Widely Used* التى تستخدم فى تحديد درجات القطع، و صممت لكى تطبق على الاختبارات ذات الاختيار مـــن مـــتعدد للحكم على درجة قطع الاختبار محكى المرجع،و تعتمد هذه الطـــريقة علـــى الاسترشاد بجمع من المحكمين *Judges*فى المجال الذى يقيسه الاختبار و هى تقتضى الطلب من كل محكم تقدير احتمال (النسبة المـــئوية) لإجابـــة الفرد ذى الحد الأدنى من الكفاية على كل مفردة من

مفـــردات الاختـــبار إجابة صحيحة، و يتم جمع هذه النسب المئوية لكل المفردات لكى نحصل على درجة القطع من وجهة نظر المحكم ، بعد ذلك يـــتم أخذ المتوسط لدرجات القطع لكل المحكمين لكى نحصل على درجة القطـــع للاختـــبار *(Fehrmann et al.,1991,858;Plake & Impara,1996; Hurtz& Hertz, 1999 , 885; Plake et al.,1999; Wang,2003,231; Ferdous & Plake,2003)*

و هـــناك عدة صور معدلة لطريقة أنجوف ، ففى عام ١٩٧٦ أجرى أول تعـــديل لطـــريقة أنجـــوف ، و فيها لا يطلب من المحكمين تقدير النسبة المـــئوية على كل مفردة من تلقاء أنفسهم، و لكن باختيار نسبة من (٧) اختـــيارات(٥%، ٢٠%، ٤٠%، ٦٠%، ٧٥%، ٩٠%، ٩٥%)، و لكن يعـــاب على هذا التعديل أنها تقيد المحكمين بالنسب المعطاة، و لا تعطى لهـــم الحـــرية لاختيار نسب أخرى،و هناك صورة أخرى لطريقة أنجوف ظهرت عام ١٩٧٨ و هى أيضاً تعتمد على اختيار المحكمين لنسبة مئوية ، و لكـــن يكـــون الاختـــيار من اختيارين و ليس سبعة كما فى الطريقة السابقة،حيث يتم سؤال المحكمين عما إذا كان الفرد ذا الحد الأدنى من الكفايـــة قـــادر علـــى إجابة المفردة بصورة صحيحة ، و بالتالى يختار المحكـــم(نعم ، أو لا)،و يتم تحديد درجة القطع فى ضوء نسبة الموافقة علـــى (نعم)، ثم يتم تحرير هذه الدرجة من أخطاء القياس، و لكن يعاب على هذا التعديل أيضاً أنه يقيد المحكمين بنسبتين فقط (١٠٠%''نعم''، صـــفر%''لا'')، ممـــا يـــزيد مـــن أخطـــاء القـــياس إلـــى أقصى حد *(Berk,1986,148)* .

و هـــناك تعـــديل آخر لطريقة أنجوف عام ١٩٨٦ حيث يسترشد الخبراء فى عملية التحكيم ببعض البيانات المستمدة من التجريب الميدانى لمفـــردات الاختبار، و فى هذه الطريقة يطلب من كل محكم تقدير احتمال إجابـــة الأفـــراد من ذوى الحد الأدنى من الكفاية المطلوبة عن مفردات الاختبار ككل إجابة صحيحة أي أن تقديراتهم تتعلق بالاختبار ككل و ليس

كـل مفردة كما سبق،و بعد الحصول على هذه التقديرات يقدم لكل منهم بـيانات تجـريبية عـن مفـردات الاختبار تتضمن متوسط صعوبة كل مفـردة،ثم يطلب منهم تعديل تقديراتهم الاحتمالية فى ضوء ذلك،و تكون درجة القطع هى متوسط تقديراتهم المعدلة

(صلاح الدين علام،١٩٩٥ ، ٢٤٣) .

ولقد أوضح *Ried* فى عام ١٩٨٤ أن هناك تعديل آخر لطريقة أنجـوف ، و هـو مـرتبط أيـضا بطريقة ندلسكاى ، و يعتمد على أخذ المتوسـط *average* لدرجتـى القطـع الناتجتـين عن طريقة أنجوف و ندلـسكاى ، و تتميـز هـذه الطـريقة بـسهولة تنفـيذها و فهمهـا *(berk,1986,152)* .

و جديـر بالذكر أن طريقة أنجوف و صورها المعدلة *modified* بالـرغم من النقد الموجه لها، إلا أنها حظيت باهتمام و تفضيل علماء القـياس النفـسى و الباحثـين لـسهولة فهم المحكمين للاجراءات التى تتطلبها هذه الطريقة*(Goodwin,1996,252)*.

و لقـد أجـريت العديـد من الدراسات على طريقة أنجوف سواء للتعـرف علـى مدى دقتها فى تحديد درجة القطع ، أو مقارنتها بالطرق الأخرى و أغلب الدراسات أثبتت ارتفاع درجة القطع الناتجة عن طريقة أنجـوف و ذلك مقارنة بالطرق الأخرى، كما أثبتت انخفاض معدل الثبات الداخلـى بين المحكمين و من هذه الدراسات*Fehrmann et Shen,2001;* *al.,1991; (Wang,2003 Chang,1996)* ، و مـن الدراسات العربية دراسة (صلاح الدين علام،١٩٩١) .

و لقـد توصـلت دراسـة*(Fehrmann et al.,1991)* إلـى أن إمـداد المحكمـين فى طريقة أنجوف سواء بصورة فردية أو بصورة جماعية بأطر مرجعية *Frame – Referenced* عن الفرد ذى الحد الأدنى من الكفاية

و تعريف دقيق له و أداء حقيقى عن أدائه يسهم فى زيادة معدلات الثبات بين المحكمين ، و كذلك قرارات تصنيف صحيحة.

فى ضوء العرض السابق، تم إخضاع الطريقة الأصلية لندلسكاى، و الطريقة الأصلية لأنجوف، للمقارنة فى البحث الحالى.

٤-مؤشرات قرار التصنيف : *Classification Decision*

بعد تحديد درجة القطع للاختبار محكي المرجع، يتم تطبيق هذا الاختبار على عينة من الأفراد المفحوصين، و يتم رصد درجاتهم على الاختبار ، و فى ضوء درجة القطع التي تم تحديدها ، يتم اتخاذ قرار تصنيفي *Classification Decision* خاص بكل مفحوص ، بحيث إذا حصل المفحوص على درجة القطع المحددة ، أو ما يزيد عنها، يتم اتخاذ القرار التصنيفي بأن هذا الفرد متمكن أو ناجح، أو حقق الأهداف المطلوب قياسها فى الاختبار أو اجتاز المحك، أما إذا حصل المفحوص على درجة أقل من درجة القطع يتم اتخاذ القرار التصنيفي بأن هذا الفرد غير متمكن أو راسب، أو فشل فى تحقيق الأهداف المطلوب قياسها فى الاختبار أو لم يجتاز المحك.

و لقد أشار *Berk* إلى أن درجة القطع هي نقطة اتخاذ القرار، حيث أن كل درجة يحصل عليها المفحوص تقارن بدرجة القطع، و هنا قرار يتخذ بتصنيف هذا المفحوص سواء متقن أو غير متقن *Mastery/Nonmastery* ، و حيث أن درجة القطع درجة ملاحظة، أى يمكن ملاحظتها ، لذلك هناك خطأ مرتبط بدرجة القطع، و خطأ مرتبط بالقرار المعتمد على هذه الدرجة، و يسمى الخطأ المرتبط بحساب درجة القطع بخطأ القياس، أما الخطأ المرتبط بالقرار يسمى بخطأ التصنيف *Classification Error* ، و يتأثر قرار التصنيف بنوعين من الخطأ يؤثران بشكل ملحوظ على دقته، أحدهما يسمى بالخطأ الموجب *False Positive Error* و فيه تصنيف خاطئ للطالب غير المتقن *Nonmaster* على أنه

متقن *Master* ، و الآخر يسمى بالخطأ السالب *False Negative Error* و فيه تصنيف خاطئ للطالب المتقن *master* على أنه غير متقن *nonmaster* . *(Berk,1986,138 ;Berk ,1995,100-101)*

و يؤثر هذان النوعان من الخطأ فى دقة قرار التصنيف *Decision Accuracy* ، و كذلك معامل صدق القرار، و دوال المنفعة و الضرر و المنفعة العظمى الناتجة عن قرار التصنيف.

و لفهم هذه المعاملات السبع(الخطأ الموجب و الخطأ السالب و دقة القرار(القرار الصحيح) و معامل الصدق و دالة لضرر و دالة المنفعة و المنفعة العظمى) الخاصة بقرار التصنيف ينبغى معرفة أن حساب هذه المعاملات يعتمد فى الأصل على اختيار مجموعتين محكيتين هما المجموعة المتعلمة *Instructional Group(I)* و هى تتكون من الطلاب الذين تلقوا تعليم فاعل *Effective Instruction* على الأهداف المطلوب تحقيقها، و يمثلون الطلاب فى هذه المجموعة الطلاب المتقنون*(M)* ، و المجموعة غير المتعلمة*Uninstructional Group(U)* و هى تتكون من الطلاب الذين لم يتلقوا تعليم على الأهداف المطلوب تحقيقها، و يمثلون الطلاب فى هذه المجموعة الطلاب غير المتقنين*(N)* .، و ينبغي أن تتساوى العينتان فى الحجم ، و كل مجموعة ينبغي أن تتكون من (١٠٠) طالب ، لكي نحصل على احتمالات مستقرة و ثابتة، و لو كانت هناك صعوبة فى اختيار المجموعتين المحكيتين ، يمكن اختيار مجموعة واحدة قبل تطبيق البرنامج التعليمي و بعده، ثم بعد ذلك يتم تطبيق الاختبار محكي المرجع (و الذي يعكس هدف أو عدة أهداف) على أفراد المجموعتين المحكيتين، وينحصر توزيع درجات الطلاب بين(صفر– ن) ، حيث (ن) عدد مفردات الاختبار و التى تمثل الدرجة الكلية للاختبار، و يتم تقسيم هذا التوزيع بواسطة درجة القطع إلى صنفين متضادين هما

المتقنون المتنبأ بهم ، و غير المتقنين المتنبأ بهم، و بالتالى يكون هناك نـــوعان من التصنيف هما:التصنيف المحك *Criterion Classification* ، و فـــيه يـــصنف الفـــرد على أساس البرنامج التعليمي الذي مر به ، حيث يحصل كل طالب فى المجموعة المتعلمة على القيمة (١) ، و يحصل كل طالـــب فى المجموعة غير المتعلمة على القيمة(٠) ، و التصنيف المنبئ *Predicted Classification* ، و فـــيه يصنف الفرد على أساس درجة القطع التـــى تم تحديدها، بحيث إذا حصل الفرد على درجة تساوى أو أعلى من درجـــة القطع يحصل على الدرجة (١)، و إذا حصل على درجة أقل من درجـــة القطـــع يحصل على الدرجة(٠) ، و تنتج التصنيفات المحكية و المتنبأ بها للطلاب بناءً على درجاتهم فى الاختبار أربعة نواتج ممكنة:-

١-متقـــنون حقيقـــيون*((True Masters(TM)* و هم الطلاب المتعلمون و الذين تقع درجاتهم عند أو أعلى درجة القطع .

٢-متقـــنون غيـــر حقيقيـــين*((False Masters(FM)* و هو الخطأ الموجب لـــسوء التصنيف و يحدث عندما تقع درجات الطلاب غير المتعلمين عند أو أعلى درجة القطع و يصنفون تبعاً لذلك على أنهم متقنون.

٣-غيـــر المتقنين الحقيقيون*((True Nonmasters(TN))* و هم الطلاب غير المتعلمين و الذين تقع درجاتهم على الاختبار تحت درجة القطع .

٤-غير المتقنين غير الحقيقيين *((False Nonmasters(FN)* و هو الخطأ الـــسالب لـــسوء التصنيف و يحدث عندما تقع درجات الطلاب المتعلمين تحت درجة القطع و يصنفون تبعاً لذلك على أنهم غير متقنين ، و هذه النواتج معروضة فى الجدول التالى

جدول(١)

تفاعل متغيرى التصنيف المحك و التصنيف المنبئ

التصنيف المحك

(البرنامج التعليمى)

المجموعة المتعلمة(*I*) المجموعة غير المتعلمة(*U*)

متقنون حقيقيون (*TM*)	متقنون غير حقيقيين (*FM*) (الخطأ الموجب)
غير متقنون غير حقيقيين (*FN*) الخطأ الأول (الخطأ السالب)	غير متقنين حقيقيين (*TN*)

التصنيف المنبئ (درجة القطع)

المتقنون المتنبأ بهم *PM=TM+FM*

غير المتقنين المتنبأ بهم *PN=FN+TN*

المتقنون *M=TM+FN*

غير المتقنين *N= FM + TN*

(Berk , 1976 , 4-5)

و اعتماداً على ما سبق يمكن شرح معاملات قرارات التصنيف كالتالى:

(١)- **الخطأ الموجب** : *False Positive Error*

الخطأ الموجب هو خطأ فى القرار المتعلق بالتصنيف *Misclassification* ، و يحدث عندما يحصل الطالب فى المجموعة غير المتعلمة *(*أى يأخذ

القيمة (٠) فى التصنيف المحك[، على درجة القطع أو ما يزيد عنه[أى يأخذ القيمة (١) فى التصنيف المنبئ[، و يصنف خطأ على أنه متقن ، و يمكن حساب نسبة احتمال الخطأ الموجب من القانون $P(FM) = FM/M+N$ ، حيث FM هى عدد القرارات الخاطئة من النوع الموجب، أما المقام$(M+N)$ فيمثل العدد الكلى للمجموعتين(المتعلمة و غير المتعلمة) . *(Berk , 1976 , 5)*

و يمكن أن يرجع سبب الخطأ الموجب لقرار التصنيف إلى عوامل تؤدى إلى حصول هذا الطالب غير المتعلم على درجة أعلى من درجة القطع ، و من هذه العوامل التخمين الذكى ، ، و أحياناً الحظ و الصدفة ، و التى تجعله يختار بدائل صحيحة بدون وعى بذلك مما يجعله يجتاز درجة القطع ،و يصنف خطأ على أنه متقن .

و لقد أوضحت نادية عبد السلام أن خطورة الخطأ الموجب تكمن فيما يترتب عليه من وضع الطالب لدراسة وحدة تعليمية أعلى من مستواه، مما يؤدى إلى تأثير سلبى على كفاءة البرنامج التعليمى عندما تكون الأهداف التعليمية متدرجة تدرجاً هرمياً بطبيعتها

(نادية عبد السلام ، ١٩٩٢ ، ٩٦) .

(٢)- **الخطأ السالب** : *False Negative Error*

الخطأ السالب هو خطأ آخر فى القرار المتعلق بالتصنيف *Misclassification* ، و يحدث عندما يحصل الطالب فى المجموعة المتعلمة [أى يأخذ القيمة (١) فى التصنيف المحك[، على درجة دون درجة القطع [أى يأخذ القيمة (٠) فى التصنيف المنبئ[، و يصنف خطأ على أنه غير متقن ، و يمكن حساب نسبة احتمال الخطأ السالب من القانون $P(FN) = FN/M+N$ ، حيث FN هى عدد القرارات الخاطئة من النوع السالب *(Berk , 1976 , 5)* .

و يمكـــن أن يـــرجع سبب الخطأ السالب إلى بعض العوامل التى يتعرض لها الطالب (المتعلم) ، و التى تؤثر على استجابته بالسلب مما يجعلـــه يحـــصل علـــى درجة أقل من درجة القطع ، و من هذه العوامل الإرهـــاق أ و شرود الذهن أ و التعب أو التعرض لمؤثرات سلبية قبل الامتحان.

و لقد أوضحت نادية عبد السلام أن لكل من الخطأين أثره على عملـــية اتخـــاذ القرار التعليمى، و لكن ضرر الخطأ السالب حيث يدرس محـــتوى الأهداف مرة أخرى للطالب فى المجموعة المتعلمة بالرغم من سابق دراسته لها ، أقل من الخطأ الموجب و الذى فيه ينتقل الطالب إلى المـــرحلة التعليمـــية التالـــية بالرغم من عدم إتقانه للأهداف فى الوحدة السابقة(نادية عبد السلام ، ١٩٩١ ، ١٠٨) .

فبالرغم من المرغوب فيه تقليل احتمالات كلا النوعين من الخطأ عـــند اختـــيار درجـــة القطع ، فان هناك بعض المواقف التى يكون فيها اقتـــراف نـــوع معين من الخطأ أكثر حدة من النتائج المترتبة عن الخطأ الآخر، فعلى سبيل المثال النتائج المترتبة على تكليف تلميذ بمواد تعليمية جديـــدة و هـــذا التلمـــيذ فى الأصل غير قادر على دراسة هذه المواد و الـــتعامل معها(الخطأ الموجب)، هذا الخطأ أكثر حدة من النتائج المترتبة علـــى إعـــادة تكليف مواد تعليمية و أنشطة علاجية لتلميذ درس هذه المواد و الانشطة من قبل (الخطأ السالب) *(Berk , 1986 , 139)* .
ممـــا ســـبق يتضح أن أساس حدوث كلا النوعين من الخطأ هو اختلاف القـــيمة التـــى يحصل عليها الطالب فى التصنيف المحك عن القيمة التى يحصل عليها فى التصنيف المنبئ/(٠، ١) ؛ (١ ، ٠)/، أما اتفاق القيمة فـــى كلا التصنيفين المحك و المنبئ أى /(١، ١) ؛ (٠ ، ٠)/، فهذا يعنى أن قرار التصنيف صحيح ، و يمكن توضيح القرارات الناتجة عن تفاعل

المتغيرين الثنائيين التصنيف المحك(١، ٠)، و التصنيف المنبئ (١، ٠) ، فى الجدول التالى:

جدول(٢)

الاحتمالات المختلفة لقرارات التصنيف الخاصة بالطالب

التصنيف المحك	الوصف	التصنيف المنبئ	الوصف	القرار
١	متعلم	١	الحصول على درجة القطع أو تجاوزها	قرار صحيح
٠	غير متعلم	١	الحصول على درجة القطع أو تجاوزها	خطأ موجب
١	متعلم	٠	الحصول على درجة أقل من درجة القطع	خطأ سالب
٠	غير متعلم	٠	الحصول على درجة أقل من درجة القطع	قرار صحيح

(٣)- **القرار الصحيح للتصنيف**: *Correct Decisions ProbabilityP(CD)* هــو نسبة القرارات الصحيحة(*Correct Decisions ProbabilityP(CD* التي تم اتخاذها في التصنيف المنبئ(المبنى على درجة القطع)، و يحسب من القانــون *P(C D) = P(TM)+P(TN)* حيث *P(TM)=TM/M+N* ،و هى نسبة المتقنــين الدقيقيــين *(TM)* فــى الحجــم الكلــى للعيــنة*(M+N)* ، *P(TN)=TN/M+N*،و هــى نــسبة غير المتقنين الحقيقيين *(TN)*فى الحجم الكلى للعينة*(M+N)* (*Berk,1986,5-6*).

كمـــا أشار *Hambelton* و زملاؤه فى عام ١٩٧٦أن صحة القرار *Decision Accuracy* الخـــاص بالتـــصنيف يمكـــن حـــسابه من المعادلة:

$$*Decision\ Accuracy\ P(CD)=1-\Sigma E^2/N$$

حيث " E "هى خطأ التصنيف و يأخذ إحدى ثلاث قيم (−١) ، (٠) ، (١) ، و تـــتحدد قـــيمة E علـــى حسب نوع الخطأ ، فإذا صنف الطالب تصنيفاً صحيحاً ، فان E =(٠)، أما فى حالة الخطأ الموجب ،فان E =(١) ، و فـــى حالة الخطأ السالب فان E =(−١) ،أما N فهى العدد الكلى لأفراد المجموعتين المحكيتين

(فى:نادية عبد السلام، ١٩٩٢ ، ٩٨ −٩٩).

ممـــا سبق يتضح أن هناك طريقتين للتعرف على القرار الصحيح للتصنيف أحدهم بحساب نسبة القرارات الصحيحة مباشرة و هى طريقة *Berk* ،أمـــا الطـــريقة الثانـــية فتأتـــى عـــن طريق طرح نسبة القرارات الخاطـــئة(الخطـــأ المـــوجب و الخطأ السالب) من الواحد الصحيح ، فى ضوء المعادلة التى ذكرها *Hambelton* و زملاؤه.

(٤)-معامـــل صـــدق قـــرار التـــصنيف *Validity Coefficient of Classification Decision*

إن درجة القطع التي تحظى بأعلى معامل صدق سينتج عنها أعلى نـــسبة من القرارات الصحيحة، و هذا المعامل(ϕ) يعد مقياس لمدى قدرة المتغيـــر الثنائـــي التـــصنيف المنبئ (١، ٠)على التقدير الدقيق للمتغير الثنائـــي التصنيف المحك(١، ٠)،و يعتمد هذا المعامل على إيجاد معامل الارتباط الرباعي بين المتغيرين الثنائيين، باستخدام القانون التالي:

$$\phi = P(TM) - BR(SR) / [BR\ (1-BR)\ SR\ (1-SR)]^{1/2}$$ ، حـــيث BR ترمـــز إلى *Base Rate* و تعنـــى معـــدل الأصـــل و هـــي تشير إلى نسبة عدد أفراد

---

* هذه المعادلة كانت مكتوبة باللغة العربية فى المرجع الموضح ، و لكن حرصا على أن تكون كل المعادلات باللغة الأجنبية فقد تم استبدالها بالصورة الموضحة.

المجمـــوعة المـــتعلمة(M) إلى العدد الكلى لأفراد المجموعتين المحكيتين (M+N) ، حيث تأتى من القانون: $BR=P(FN)+P(TM) = M/M+N$

، أمـــا SR ترمز إلى Selection Ratio و تعنى نسبة الانتقاء و هى تشير إلىنـــسبة عـــدد الأفراد المتقنين المتنبأ بهم (PM) إلى العدد الكلى لأفراد المجموعتـــين المحكيتـــين(M+N) ، حـــيث تأتـــى مـــن القانـــون:

$SR=P(TM)+P(FM) = PM/M+N$ ، و تقـــيد كل من معدل الأصل و نسبة الانـــتقاء مـــدى القيم لمعامل الصدق، فعندما $BR=SR$ تكون أقصى قيمة لمعامـــل الـــصدق و تـــصل إلـــى (١) ، و كلمـــا بعـــدت النسبتين عن بعضهما،يتناقص معامل الصدق (Berk,1986,7) .

(٥)-**دالة الضرر**: Disutility Function

إن خطـــأي التـــصنيف(الخطـــأ الموجب و الخطأ السالب) يتضمنان بعض الخسارة و الفاقد، هذه الخسارة تعد دالة للخطورة النسبية الناتجة عـــن نوعـــى الخطـــأ، و عندما يتم استخدام نسبة القرارات الصحيحة و معامـــل الصدق فى اختيار درجة القطع ، يفترض أن هذين النوعين من الخطأ لهما نفس الخطورة، و لكن هذين النوعين من الخطأ لهما درجات مـــتفاوتة مـــن الخطـــورة، و من ثم فان احتمالات الخطأ تحتاج أن يتم تعـــديلها لكى تعكس الخسارة الحقيقية المرتبطة بقرار التصنيف المتخذ، فالخسارة المقدرة في إعادة الوقت التعليمي و إحباط الطالب المتقن غير الحقيقـــي، و الـــذي يحاول تحقيق أهداف تعليمية أصعب من الأهداف الـــسابقة (و التي لم يتقنها أصلاً(الخطأ الموجب) ، هذه الخسارة تكافئ مـــرتين الخـــسارة الناتجة عن تكرار الطالب غير المتقن غير الحقيقي لدراســـة المادة التعليمية لأنه حصل على درجة دون درجة القطع بالرغم مـــن دراسته للبرنامج التعليمي (الخطأ السالب)، و على ذلك فان تقدير الخسارة (الضرر) المتوقع و المرتبط بدرجة قطع معينة يعد أصلاً الضرر المـــتوقع لقـــرار التـــصنيف المـــتخذ، و تحـــسب دالة الضرر (δ) لقرار

وحـــيد(خـــاص بدرجـــة قطع) من مجموع نواتج احتمالات الخطأ(الخطأ الموجب ، و الخطأ السالب) كما في المعادلة التالية: $\delta_K = P(FN)\ (D_1) + P(FM)\ (D_2)$ ، حـــيث $D1$ ، و $D2$ تمـــثل القيمتان الوزنيتان لتعديل حجـــم الخـــسارة الناتجة عن أخطاء قرارات التصنيف، و في ضوء ما سبق تأخذ $D1$القيمة (-١) ، و تأخـــذ $D2$ القيمة (-٢) ، نظراً لأن $P(FN)$ ترمز إلى نسبة تكـــرار الأخطـــاء السالبة و التي فيها يحصل الطالب المتقن*Master* (مر ببـــرنامج تعليمي)على درجة دون درجة القطع ، و يصنف خطأ على أنه غيـــر مـــتقن و من ثم يعيد *Retain* دراسة المادة التعليمية ، و $P(FM)$ ترمـــز إلى نسبة تكرار الأخطاء الموجبة و التى يحصل فيها الطالب غير المـــتقن*Nonmaster* على درجة القطع أو ما يزيد عنها و من ثم يصنف خطأ على أنه متقن و من ثم ينتقل إلى مرحلة تعليمية تالية، و هذا الخطأ كما سلف ذكره يكافئ مرتين الخطأ السالب فى الضرر الناجم عنه، لذلك أخذ الخطأ الموجب القيمة الوزنية$(D_2 = -2)$ ، و أخذ الخطأ السالب القيمة الوزنية$(D_1 = -1)$ *(Berk,1986,7)* .

فالـــضرر الـــناتج عن سوء التصنيف هو فى الواقع دالة لأخطاء التـــصنيف الموجبة و السالبة ، و هذه الدالة يتم تحديدها بصورة ذاتية (فـــى ضـــوء اختـــيار القيم الوزنية)، و هى تعتمد على سلوك المخاطرة لصانع القرار و اتجاهه نحو نسبة العوامل المسببة للضرر.

و أشـــار *Vos* إلى أن العوامل المسببة للضرر يمكن أن تنحصر فى العواقب الاجتماعية و النفسية و الاقتصادية التى تصاحب القرار الخاطئ المتخذ*(Vos,2000,197)*.

(٦)-**دالة المنفعة** : *Utility Function*

إن دالـــة المنفعة تقدر بواسطة نسبة القرارات الصحيحة ، و التى تأخذ أيضاً قيماً وزنية مثل ما حدث فى دالة الضرر مع القرارات الخاطئة

، و لكن الفارق هنا أن القيم الوزنية تأخذ قيماً موجبة لأنها دالة منفعة، و يرمز لدالة المنفعة لقرار تصنيفي بالرمز (υ_K) ، أما القيمتين الوزنيتين فى دالة المنفعة فيرمز لهما U_1 ، و U_2 ، و قيمتهما تأتى من أن الفائدة أو المكسب فى الكفاءة و الفاعلية التعليمية الناتجة عن التصنيف الصحيح لفرد متقن(TM) ، يكافئ مرتين الفائدة الناتجة عن التصنيف الصحيح لفرد غير متقن (TN) ، و بالتالى فان القيمة الوزنية المقابلة لنسبة الإتقان الحقيقى(TM) تساوى +٢$(U_1 = +2)$ ، كما أن القيمة الوزنية المقابلة لنسبة عدم الإتقان الحقيقي(TN) تساوى +١$(U_2 = +1)$ ، و تكون الفائدة المتوقعة *Expected Utility* لقرار التصنيف K المبنى على درجة قطع معينة يساوى مجموع نواتج القرارات الصحيحة معدلة بواسطة القيمتين الوزنيتين $(U_1 = +2 ; U_2 = +1)$ و دالة الفائدة موضحة فى المعادلة التالية:

$$\upsilon_K = P(TM)\,(U_1) + P(TN)\,(U_2) \qquad (Berk,1986,7)$$

(٧)- **الفائدة العظمى المتوقعة** *The Expected Maximal Utility* :

و تأتى من مزج التقديرات الناتجة عن دالتى الضرر و المنفعة خلال كل الطلاب الخاضعين لقرار التصنيف من المعادلة: $\tau = \sum^{M+N}_{k=1} (\upsilon_K + \delta_K)$، و يعرف بأنه المجموع الموزون(المعدل) لكل الاحتمالات الخاصة بالقرارات الصحيحة و الخاطئة و المرتبطة بدرجة قطع معينة ، و درجة القطع التى تقابل أعلى فائدة عظمى (τ) يزيد الفائدة المتوقعة من القرار المستخدم$(Berk,1986,7\text{-}8)$.

الدراسات المرتبطة و فروض البحث الحالى:

١- **دراسة**$(Berk\ ,\ 1976)$ التي استخدمت طريقة تجريبية فى تحديد درجة القطع تعتمد على رسم منحنيين أحدهما يمثل درجات المجموعة المتعلمة و المنحنى الآخر يمثل درجات المجموعة غير المتعلمة ، و يمثل نقطة تقاطع المنحيين درجة القطع ، و كان المقياس المستخدم هو اختبار

محكى المرجع فى حساب رؤوس الأموال *Capitalization*، يتكون من (٨) مفـردات ، و لا توجد عينة للمحكمين نظراً لأن الطريقة المستخدمة فى تحديـد درجـة القطع لا تعتمد على المحكمين و لكن تعتمد على الرسم البيانـى، أمـا عيـنة المفحوصـين فتكونت من مجموعتين : مجموعة المتعلمـين و تكـونت من (١٠٠) طالب، و مجموعة غير المتعلمين و تكـونت أيـضاً مـن (١٠٠) طالب، و تعتبر هذه الدراسة من الدراسات الـنادرة التي تناولت مؤشرات التصنيف (الخطأ الموجب و الخطأ السالب و القرار الصحيح للتصنيف و معامل الصدق و دوال المنفعة و الضرر و المنفعة العظمى) ، حيث قارنت الدراسة بين مؤشرات التصنيف الناتجة عن الطريقة التجريبية ، و مؤشرات التصنيف الناتجة عن وضع درجات قطع بصورة تعسفية ، و توصلت الدراسة إلى أن درجة القطع المثالية و التـى تقابـل أفـضل مؤشرات للتصنيف، كانت درجة القطع الناتجة عن الطـريقة التجـريبية، و التى كانت (٤,٥) و تم تقريبها إلى (٥) ، حيث كانت مؤشرات التصنيف هى(القرار الصحيح للتصنيف''٠,٧٤'' والخطأ ا لمـوجب''٠,١٢'' و الخطـأ الـسالب ''٠,١٤'' و معامل الصدق '' ٠,٤٨''و دالـة المـنفعة ''١,١'' و دالـة الضرر''-٠,٣٨'' و دالة المـنفعة العظمـى''١٤٤'')، و هى أفضل من مؤشرات التصنيف التى تقابل درجات القطع و التى تم وضعها بطريقة تعسفية .

٢- **دراسة***(Brennan & Lockwood ,1980)* التـى هدفت إلى المقارنة بين طريقتى ندلسكاى و أنجوف ، و تركزت المقارنة على قيمة درجة القطع ، و كـان الاختبار المستخدم يتكون من (١٢٦) مفردة من نوع الاختيار مـن مـتعدد (رباعية البدائل) فى مجال الصحة العامة، و تكونت عينة المحكمين من (٥) محكمين فى المجال الذى يقيسه الاختبار ، و توصلت الدراسـة إلـى عـدة نتائج منها اختلاف درجة القطع باختلاف الطريقة المطبقـة فكانت(٨٤) باسـتخدام طريقة أنجوف باتحراف معيارى قدره

(٤,٧) ، و (٧٠) باستخدام طريقة ندلسكاى بانحراف معيارى قدره (٩,٠٣).

٣-**دراسة**(*Livingston & Zieky ,1989*) التى هدفت أيضاً إلى المقارنة بين طريقتى ندلسكاى و أنجوف ، و كان هناك اختباران فى الدراسة أحدهما فى الرياضيات و يتكون من (٦٥) مفردة ، أما الاختبار الآخر فكان فى القراءة و يتكون من (٧٠) مفردة من نوع الاختيار من متعدد (رباعية البدائل) ، أما بالنسبة للمحكمين فى الاختبارين فتكونت عينتهم من عدد يتراوح بين (١٢ -٢٠) محكم فى المجال الذى يقيسه الاختبار ، تم اختيارهم من أربعة مدارس مختلفة و توصلت الدراسة إلى نتيجتين متناقضتين ،فبينما توصلت الدراسة فى اختبار الرياضيات إلى أن درجة القطع المحددة بطريقة أنجوف (٣٦)* ، أعلى من درجة القطع المحددة بطريقة ندلسكاى (٣١)، نجد أنه فى اختبار القراءة و بصورة مجملة توصلت إلى العكس، أنجوف (٣٤)، ندلسكاى (٤٢)

٤-**دراسة**(*DeMauro & Powers,1990*) هدفت الدراسة إلى معرفة الاتساق المنطقي لطريقة أنجوف ، و تكونت عينة المحكمين من (١٩) محكم و هم أعضاء المؤسسة الدولية لعلم النفس المدرسي في ذلك الوقت، و الاختبار المستخدم هو اختبار محكي المرجع في علم النفس المدرسي و يتكون من (١٣٠) مفردة ، و عينة المفحوصين عددها (٦٣٧٣) من الطلاب ، و توصلت الدراسة إلى عدة نتائج منها أن طريقة أنجوف تحظى بدرجة عالية نسبياً من الثبات المنطقي، حيث يوجد اتساق بين تقديرات المحكمين و الأداء الفعلي للطلاب قدره(٠,٧١) ،كما يوجد ارتباط بين تقديرات المحكمين على المفردات و الصعوبات الحقيقية

* تم أخذ متوسط درجات القطع المحددة فى الدراسة على الأربعة مدارس فى كل طريقة (أنجوف - ندلسكاى)، و كل اختبار(الرياضيات –القراءة).

للمفـــردات تـــراوح بين (٠,٢٥) إلى (٠,٥٦) بوسيط قدره (٠,٤٢)، و أوصـــى الباحث باستخدام طريقة أنجوف في تحديد درجة القطع ، و لكن أشار إلى أن نتائجه محكومة بالاختبار المستخدم و العينة المستخدمة.

٥-دراسة *(Fehrmann et al.,1991)* هدفت الدراسة إلى التعرف على تأثير إســـتراتجيتين مقدمتين للمحكمين على ثبات و دقة القرارات الناتجة عن طـــريقة أنجـــوف ،و كانـــت عينة المحكمين عبارة عن (٢١) محكم من الخبـــراء الـــذين لهـــم درايـــة تامـــة بعلم النفس العام، و لقد تم تقسيم المحكمـــين إلى ثلاث مجموعات بالتساوى بحيث تقوم المجموعة الأولى بتقديـــر المفـــردات بـــدون استراتيجية مقدمة، بينما تقدم للمجموعتين الأخـــريين الاستراتيجيتان موضوع البحث،و كان الاختبار المستخدم، هو اختبار الكفاءة فى علم النفس العام و يتكون من (٣٠) مفردة، و توصلت الدراسة إلى عدة نتائج منها أن طريقة أنجوف العادية (بدون تطبيق أى من الاستراتيجيتين) تتسم بانخفاض معدلات القرارات الصحيحة(٠,٥٤) ، و كذلك ارتفاع نسبة الاخطاء الموجبة(٠,٠٥) و السالبة(٠,٤١) لقرار التـــصنيف الناتج عنها ، و هذا ما حدا بالباحث إلى تقديم استراتيجيتين تعـــتمدان على إمداد المحكمين فى طريقة أنجوف سواء بصورة فردية(الاســـتراتيجية الأولـــى) أو بصورة جماعية(الاستراتيجية الثانية) بأطر مـــرجعية *Frame – Referenced* عـــن الفرد ذى الحد الأدنى من الكفاية و تعريف دقيق له و أداء حقيقى عن أدائه و الذى أدى إلى ارتفاع معدلات القـــرارات الـــصحيحة إلى (٠,٧٤؛ ٠,٦٩) ، و انخفاض نسبة الاخطاء الـــسالبة إلى (٠,١٤ ؛ ٠,٢٣) للاستراتيجيتين على الترتيب ، و لكن فـــى المقابل نجد أن نسبة الأخطاء الموجبة تزايدت و لكن بمقدار ضئيل حيث وصلت النسبة إلى (٠,١٢ ؛ ٠,٠٨) ،و ذلك للاستراتيجيتين على الترتيب .

٦-**دراسة (صلاح الدين علام،١٩٩١)** هدفت هذه الدراسة إلى المقارنة بين أربع طرق لتحديد درجة القطع و هى طريقة أنجوف و طريقة ندلسكاى و طريقة المجموعات الحدية و طريقة المجموعات المتناقضة، و كان المقياس المستخدم هو اختبار محكى المرجع فى أساسيات القياس و التقويم التربوى و النفسى لطلاب كلية التربية(٦٠ مفردة رباعية البدائل) ، و تكونت عينة المحكمين من (٢٦) محكم(١٤ من معلمى المدرسة الثانوية الحاصلين على الدبلوم الخاص فى التربية،و (٦) مدرسين مساعدين ، و (٦) من أساتذة علم النفس التعليمى، أما عينة المفحوصين فتكونت من (٣٨٢) طالب من طلاب كلية التربية بجامعة الأزهر، و تمت المقارنة بين الطرق الأربعة على درجات القطع نفسها التى توصل إليها المحكمون باستخدام الأساليب الإحصائية الاستدلالية(تحليل التباين ذى القياسات المتكررة)، و توصلت الدراسة إلى عدة نتائج منها أن الطرق الأربعة التى استخدمها المحكمون فى تقدير درجات القطع أدت إلى درجات قطع متباينة نسبيا فقد كانت درجة القطع باستخدام طريقة ندلسكاى (٤٠)، و طريقة المجموعة الحدية (٤٤)،و طريقة أنجوف (٤٣)،و طريقة المجموعات المتناقضة (٤١).

٧- **دراسة(نادية عبد السلام ، ١٩٩١)** التي استخدمت طريقة تجريبية فى تحديد درجة القطع تعتمد على رسم منحنيين أحدهما يمثل درجات المجموعة المتعلمة و المنحنى الآخر يمثل درجات المجموعة غير المتعلمة ، و يمثل نقطة تقاطع المنحيين درجة القطع ، و كان المقياس المستخدم هو اختبار محكى المرجع فى منهج الإحصاء الوصفى،و لا توجد عينة للمحكمين نظراً لأن الطريقة المستخدمة فى تحديد درجة القطع لا تعتمد على المحكمين و لكن تعتمد على الرسم البيانى، أما عينة المفحوصين فتكونت من مجموعتين : مجموعة المتعلمين و تكونت من (٧٥) طالب، و مجموعة غير المتعلمين و تكونت أيضاً من (٧٥) طالب

،و تناولت الدراسة مؤشرات التصنيف (قرار الإتقان و خطئا ألفا و بيتا و معامل الصدق و دالتا المنفعة و الضرر)، حيث قارنت الدراسة بين مؤشرات التصنيف الناتجة عن الطريقة التجريبية ، و مؤشرات التصنيف الناتجة عن وضع درجات قطع بصورة تعسفية ، و توصلت الدراسة إلى عدة نتائج منها أن مؤشرات التصنيف الناتجة عن الطريقة التجريبية، أفضل من مؤشرات التصنيف الناتجة عن وضع درجات قطع بصورة تعسفية و ذلك بالنسبة للبعدين الاول و الثالث(مستويات القياس، و مقاييس النزعة المركزية) من أبعاد الاختبار المحكي موضوع البحث(٥١ مفردة من نوع الاختيار من متعدد تتوزع على (٦) أبعاد فى محتوى الإحصاء الوصفى)، أما باقي الأبعاد فتم التوصل إلى عكس النتيجة ، و هذا إن دل على شئ فإنما يدل على القصور الذي يحدو الطريقة التجريبية موضوع البحث فى تحديدها لدرجة القطع .

٨-**دراسة**(*Chang,1996*) التى هدفت إلى المقارنة بين طريقتى ندلسكاى و أنجوف ، و تركزت المقارنة على الثبات الداخلى بين المحكمين *Intrajudge Consistency* فى كل طريقة، و أيضاً قيمة درجة القطع ، و تكونت عينة المحكمين من (٢٢) محكم ممن حصلوا على برنامج فى الدراسات العليا بكلية التربية، و كان الاختبار المستخدم يتكون من (٩) مفردات من نوع الاختيار من متعدد (رباعية البدائل) مأخوذ من الاختبار النهائى لهؤلاء المحكمين فى برنامج الدراسات العليا ، أما عينة المختبرين فتكونت من (٢٧٤)طالب بالدراسات العليا ليس من بينهم المحكمين، و توصلت الدراسة إلى عدة نتائج منها اختلاف درجة القطع باختلاف الطريقة المطبقة فكانت ٧١%، باستخدام طريقة أنجوف ، و ٥٧% باستخدام طريقة ندلسكاى ، كما توصلت الدراسة إلى أن طريقة ندلسكاى أعلى ثبات داخلى بين المحكمين مقارنة بطريقة أنجوف.

٩- **دراسة**(*Donnoe&Amato,1997*) التـــى هدفت إلى المقارنة بين طرق أنجـــوف و ندلـــسكاى و ايـــبل ،و تركـــزت المقارنة على درجات القطع باســـتخدام الطرق الإحصائية مثل تحليل التباين أحادى الاتجاه، و معامل الارتباط،و الدراسة عبارة عن تحليل بيانات تجريبية مستمدة من تطبيق اختبارين فى الرخصة المهنية*Occupational Licensing Examinations* ، و تكونت عينة المحكمين من (٢٩) محكم ،تم اختيارهم من عينة كلية من المحكمين بلغ قوامها(٥٥٠) محكم من المتخصصين فى الانتقاء المهنى ، و توصـــلت الدراســـة إلى عدة نتائج منها اختلاف درجة القطع باختلاف الطريقة المطبقة فكانت ٥٠% استخدام طريقة أنجوف ، ٣٣% باستخدام طـــريقة ندلسكاى ، و ٦٠% باستخدام طريقة ايبل، كما توصلت الدراسة إلـــى وجـــود فروق ذات دلالة احصائية بين الطرق الثلاثة ، حيث كانت قـــيمة ف تساوى ٢٧، عند درجات حرية (٢،٢٥)(عدد المحكمين=٢٨)،كما لا توجد ارتباطات بين طريقة ندلسكاى و كل من طريقتى أنجوف و ايبل، بينما يوجد ارتباط بين طريقتى أنجوف و ايبل.

١٠- **دراسة**(*Bontempo et al.,1998*) أجـــرت هـــذه الدراسة تحليل بعدى *Meta-Analytic* لطـــرق تحديـــد درجة القطع ،باستخدام حجم التأثير،حيث تمـــت مقارنـــة طـــريقة أنجـــوف المعدلة بأربع طرق أخرى هي(طريقة المجمـــوعة الحدية،طـــريقة المجموعات المتناقضة،طريقة ايبل، طريقة جيجـــر)،و توصـــلت الدراسة إلى عدة نتائج منها أن طريقتي المجموعة الحديـــة و جيجر يؤديان إلى درجات قطع أعلى من طريقة أنجوف ،بينما طـــريقة ندلسكاى تؤدى إلى درجة القطع أقل،كما أن عدد المحكمين يؤثر فى تباين درجات القطع.

١١-**دراسة**(*Impara & Plake,2000*) قارنت الدراسة بين عدة طرق لتحديد درجة القطع منها طريقتي أنجوف و المجموعة الحدية ، و تكونت عينة المحكمين من (٢٢) معلم للصف الرابع بإحدى المدارس، و كان الاختبار

المستخدم اختبار فى الرياضيات تم إعداده بواسطة النظام المدرسي ليغطى (٦٠) مفردة من نوع الاختيار من متعدد، و توصلت الدراسة إلى عدة نتائج منها أن درجة القطع الناتجة عن طريقة أنجوف(٤١,٥) أقل من درجة القطع الناتجة عن طريقة المجموعة الحدية(٤٨).

١٢- **دراسة**(*Shen,2001*) التي هدفت إلى المقارنة بين طريقتين من طرق تحديد درجة القطع هما طريقة أنجوف و طريقة خريطة المفردة لراش *Rasch Item Map* ، وقد تركزت عملية المقارنة على ما تعكسه درجة القطع فى كل طريقة من محتوى سلوكى، و كان المقياس المستخدم هو مقياس الرخصة لمزاولة مهنة الطب *Medical Licensure Examination* ، و تكونت عينة المحكمين من (٢٣) محكم من المختصين الذين يتلقون برنامج للدراسات العليا فى كلية الطب و توصلت الدراسة إلى عدة نتائج منها أن درجة القطع التى تم تحديدها بواسطة طريقة أنجوف كانت ٦٤ %،ومعدل الفشل ٢١%، و درجة القطع التى تم تحديدها بواسطة طريقة خريطة المفردة لراش كانت ٧٧%،ومعدل الفشل ١٢,٥%، كما توصلت الدراسة إلى أن درجة القطع المحددة بواسطة طريقة أنجوف أكثر حدة من درجة القطع المحددة بواسطة طريقة خريطة المفردة ، و لكن طريقة خريطة المفردة تعكس محك مرغوب من محتوى الاختبار مقارنة بطريقة أنجوف .

١٣- **دراسة**(*Buckendahl et al.,2002*) قارنت هذه الدراسة بين طريقتى أنجوف و طريقة الترتيب التصاعدى لصعوبة البنود " *Bookmark* "، و كان وجه المقارنة بين الطريقتين عبارة عن متوسط درجات القطع عبر المحكمين،و الانحراف المعيارى لهذه الدرجات ، و كان المقياس المستخدم هو مقياس تقييم الرياضيات*Mathematics Assessment* و يتكون من (٦٩) مفردة، و تكونت عينة المحكمين من (٢٦) محكم، أما عينة المفحوصين فتكونت من (٤٤٨) طالب و توصلت الدراسة إلى عدة نتائج

منها وجود اختلاف بسيط بين درجات القطع المحددة باستخدام الطريقتين فكانت درجة القطع لطريقة أنجوف خلال محاولتين ٥١%، ٤٨%، و درجة القطع لطريقة *Bookmark* ٤٩%، ٥٢% ، كما أن الانحراف المعياري لدرجات القطع الناتجة عن طريقة *Bookmark* أقل من الانحراف المعياري لدرجات القطع الناتجة عن طريقة أنجوف مما يعطي الطريقة الأولى أفضلية.

١٤- **دراسة**(*Wang,2003*) و تمت فى هذه الدراسة المقارنة بين طريقتى أنجوف و خريطة المفردة مع تغير فى عدد مفردات الاختبار و عدد المحكمين باستخدام نظرية إمكانية التعميم *Generalizability Theory* ، حيث تم اختبار متغيرين مستقلين يؤثران على معامل ثبات و موافقة المحكمين، و كان المقياس المستخدم هو مقياس الرخصة و التأهل المهني *Examination Licensure & Certification* ،و توصلت الدراسة إلى عدة نتائج منها اختلاف درجة القطع باختلاف الطريقة المطبقة و عدد المفردات و عدد المحكمين فكانت درجات القطع ٦٨%،٥١%(١٣ محكم و ٤٣ مفردة)؛ ٦٧%،٤٩%(٦ محكمين و ٧٣ مفردة)؛٦٧ %، ٤٨%(٦ محكمين و ٧٩ مفردة)؛ ٨١ %،٧٢%(٨ محكمين و ٥٣ مفردة)، و ذلك بالنسبة لطريقتى أنجوف و خريطة المفردة على الترتيب، كما توصلت الدراسة إلى أن طريقة خريطة المفردة تتيح أعلى معدل ثبات داخلى و أعلى موافقة بين المحكمين مقارنة بطريقة أنجوف.

يتضح من العرض السابق للدراسات السابقة ما يلي:

(١)-اهتمت معظم الدراسات السابقة بالمقارنة بين الطرق المختلفة لتحديد درجات القطع من خلال المقارنة بين درجات القطع نفسها باستخدام الأساليب الإحصائية المختلفة (تحليل التباين ،معامل الارتباط،الانحراف المعياري،معامل الثبات الداخلي)، و العدد القليل من هذه الدراسات قارن بين هذه الطرق فى ضوء مؤشرات التصنيف(الخطأ

الموجب و الخطأ السالب و القرار الصحيح للتصنيف و معامل الصدق و دوال المنفعة و الضرر و المنفعة العظمى) مثل دراسة

(Berk , 1976) ، ودراسة(نادية عبد السلام، ١٩٩٠) بالرغم من أنهما استخدما طريقة فى تحديد درجة القطع لا تعتمد على المحكمين، و دراسة (Fehrmann et al.,1991) و التى استخدمت القرار الصحيح للتصنيف و الخطأ الموجب و الخطأ السالب ، كما أن هذه الدراسة استخدمت طريقة واحدة فى تحديد درجة القطع ، و هى طريقة أنجوف و درست تأثير متغيرات تجريبية معينة على قرار التصنيف الناتج عن هذه الطريقة، و من ثم فهناك ندرة فى الدراسات التى قارنت بين الطرق المختلفة لتحديد درجة القطع من حيث مؤشرات التصنيف (الخطأ الموجب و الخطأ السالب و القرار الصحيح للتصنيف و معامل الصدق و دوال المنفعة و الضرر و المنفعة العظمى).

(٢)- يلاحظ من خلال المقاييس المستخدمة فى الدراسات السابقة أنها تغطى مجالات متعدد فى النواحى الأكاديمية،و الانتقاء المهنى،و الرخصة فى مجال الطب، مما يعطى أهمية للطرق المستخدمة فى تحديد درجة القطع ، كما يلاحظ أن معظم هذه المقاييس من نوع الاختيار من متعدد رباعى البدائل، و هذا ما اتبعه الباحث فى إعداد المقياس المستخدم فى البحث الحالى.

(٣)-من خلال نتائج الدراسات السابقة يمكن صياغة الفرضين الرئيسيين التاليين:

أ-تختلف درجة القطع باختلاف الطريقة المستخدمة في تحديد هذه الدرجة(طريقة ندلسكاى،طريقة أنجوف).

ب-تختلف مؤشرات قرار التصنيف(الخطأ الموجب و الخطأ السالب و القرار الصحيح و معامل الصدق و دوال المنفعة و الضرر والمنفعة

العظمـــى) باختلاف الطريقة المستخدمة فى تحديد درجة القطع (طريقة ندلسكاى،طريقة أنجوف).

إجراءات البحث

أولاً :عينة البحث:

هناك نوعان من الأفراد الذين يمثلون عينة البحث كالتالي:

الـــنوع الأول:المحكمـــون *Judges* و هـــم الأفراد الذين يقومون بتقدير مفـــردات الاختبار موضوع البحث بطريقتى أنجوف و ندلسكاى للحصول على درجة القطع الناتجة عن كل طريقة ، و بلغ عددهم (١٢) محكماً، تم اختـــيارهم من بين المعيدين و المدرسين المساعدين و المدرسين الذين قامـــوا بتدريس محتوى التدريبات المعملية(سكاشن علم النفس) بمعامل كلـــيات التـــربية (كلية التربية بقنا،كلية التربية النوعية بقنا،كلية التربية بالغردقة) ،و ذلك فى ضوء التوزيع المبين بالجدول التالى:

جدول(٣)

توزيع عينة المحكمين فى البحث

البيان / الكلية	عدد المحكمين
التربية بقنا	٨
التربية بالغردقة	٢
التربية النوعية بقنا	٢
المجموع	١٢

النوع الثانى:الأفراد المفحوصون *Examiners* و هم الأفراد الذين يقومون بالاستجابة لمفردات الاختبار المحكي المرجع في التدريبات المعملية لعلم الـــنفس ، و حـــيث أن مؤشرات التصنيف تتطلب مجموعتين من الأفراد

إحـــداهما متعلمة و الأخرى غير متعلمة، لذلك فإن الأفراد ينقسمون إلى مجموعتين:المجموعة المتعلمة:و هم الطلاب الذين مروا ببرنامج تعليمى فـــى التدريـــبات المعملـــية لعلم النفس(سكاشن علم النفس) خلال الفصل الدراســـى الأول مـــن العـــام الجامعى٢٠٠٤-٢٠٠٥ ،و بلغ قوامهم فى البداية (١٠٠) طالب من طلاب كلية التربية بقنا شعبة التعليم العام للقسم العلمـــى، ولكن تم استبعاد(٧) طلاب منهم لأسباب تتعلق بعدم الجدية فى الإجابـــة ، أو تطـــرف مستوى الذكاء(خارج الفئة ٩٠ -١١٠ %) نسبة ذكاء، و بذلك وصل عددهم النهائى إلى (٩٣)طالب،

والمجمـــوعة غيـــر المـــتعلمة و هـــم الطـــلاب الذين لم يمروا ببرنامج تعليمى(سكاشن علم النفس)،خلال نفس الفصل الدراسى، وبلغ قوامهم فى البداية(٨٨)طالب تم اختيارهم من بين طلاب الفرقة الثالثة بكلية التربية شـــعبة التـــربية الرياضية خلال نفس الفصل الدراسى،و تم استبعاد(٣) طلاب منهم، لنفس الأسباب السابقة ، و بذلك وصل عددهم النهائى إلى (٨٥)طالب.

و بالتالـــى فـــان المجمـــوعة المـــتعلمة تدرس سكاشن علم النفس ، و المجمـــوعة غيـــر المـــتعلمة لا تدرس هذه السكاشن بالرغم من أن كلا المجموعتين تدرس مقرراً نظرياً فى علم النفس التعلم .

و توزيع عينة الأفراد المفحوصين مبين فى الجدول التالى:

---

* تنص اللائحة الداخلية لكلية التربية بقنا خلال العام الدراسى ٢٠٠٤-٢٠٠٥ على أن يدرس طلاب كلية التربية شعبة التعليم العام في الفصل الدراسي الأول مقرراً نظرياً في علم نفس التعلم و يصاحبه مقرراً عملياً في التدريبات المعملية لعلم النفس، بينما يدرس طلاب الفرقة الثالثة شعبة التربية الرياضية نفس المقرر النظري في نفس الفصل الدراسي و لكن بدون المقرر العملي.

جدول(٤)

توزيع عينة الأفراد المختبرين فى البحث

المجموعة	المتعلمة (Masters)			غير المتعلمة (Nonmasters)
الشعبة	الرياضيات	الأحياء	الطبيعة و الكيماء	التربية الرياضية
العدد	٥٩	٢٤	١٠	٨٥
المجموع	٩٣			٨٥
المجموع الكلى	١٧٨			

ثانياً : أداتا البحث:

الأداة الأولى :الاختبار المحكي المرجع في التدريبات المعملية لعلم النفس (إعداد الباحث)

١-بناء الاختبار:

أوضح صلاح الدين علام(١٩٩٥، ٣٩-٩١) ، أن هناك خمس خطوات لبناء الاختبارات محكية المرجع هى:

أ-تحديد الكفايات الرئيسية المكونة لمحتوى ا لاختبار.

ب-تحليل الكفايات الرئيسية إلى مكوناتها الفرعية.

جـ-صياغة الأهداف السلوكية التى تعكس الكفايات الفرعية.

• ملحق البحث

د- بناء المفردات الاختبارية.

هـ-التحقق من صدق نواتج تحليل الكفايات .

و لقد اتبع الباحث الخطوات الأربع الأولى و الاستعاضة كبديل للخطوة الأخيرة(التحقق من صدق نواتج تحليل الكفايات)، بإجراء صدق و ثبات للاختبار ككل، حيث تم إجراء الصدق بطريقة الصدق الوصفى ، و إجراء الثبات بطريقة معامل سابكوفياك، كما سيلى شرحه.

و فيما يلى عرض هذه الخطوات على الاختبار موضوع البحث:

الخطوة الأولى- تحديد الكفايات الرئيسية المكونة لمحتوى الاختبار(التدريبات المعملية لعلم النفس):إن التدريبات المعملية لعلم النفس من صميم اختصاص الباحث ، نظراً لقيامه بتدريس هذا المحتوى المعملي للعديد من الطلاب عبر سنوات متعاقبة ، لذا فقد قام الباحث بتحديد خمس كفايات رئيسية تعبر عن أساسيات التدريبات المعملية لعلم النفس ، و هي كالتالي:

(١)- المعارف و المهارات اللازمة لممارسة تجربة التعلم.

(٢)- إعداد جدول البيانات.

(٣)- رسم منحنى التعلم.

(٤)- التعليق على منحنى التعلم.

(٥)- معرفة التطبيقات التربوية لتجربة التعلم.

الخطوة الثانية- تحليل الكفايات الرئيسية إلى مكوناتها الفرعية:و فى هذه الخطوة يتم تحليل الخمس كفايات السابقة إلى مهارات(كفايات) بسيطة يمكن قياسها، و لقد اتبع الباحث أسلوب العلاقات الهرمية بين المهارات*Hierarchical Skills Relations* ، و الذي يعنى تنظيم المهارات تنظيماً هرمياً بحيث أن إتقان المتعلم لمهارة أو كفاية معينة يعد متطلباً أساسياً لإتقانه المهارات التالية، و هذه ما ينطبق على المهارات المتطلبة فى معمل علم النفس، و لقد قام الباحث بتحليل الكفايات الخمس الرئيسية إلى (٢٥) كفاية فرعية، فعلى سبيل المثال،الكفاية الأولى (المعارف و

المهـــارات اللازمـــة لممارســـة تجربة التعلم)، هذه الكفاية لرئيسية تم تحلـيلها إلـــى خمـــس كفايات أو مهارات فرعية و هى:١-معرفة شكل التجربة،٢- تصنيف تجربة التعلم فى ضوء النظرية التابع لها،٣-تحديد متغيـرات التجـربة،٤-تحديـد منحنـيات التعلم فى التجربة،٥-معرفة الخطـوات الإجـرائية للتجربة،و هذه المهارات الفرعية تسير فى ترتب منطقـى،حيث أن معـرفة شـكل التجربة يؤدى بالمفحوص إلى تصنيف التجـربة فـى ضـوء النظرية التابع لها،مما يؤدى إلى تحديد متغيرات التجربة(المستقلة و التابعة)،و الذى يؤدى به تلقائيا إلى تحديد منحنيات الـتعلم فى التجربة ،و كل هذه المعلومات تؤدى به إلى معرفة الخطوات الصحيحة لإجراء التجربة.

و هكذا بالنسبة لباقى الكفايات ، و الذى يوضحه الجدول التالى:

جدول(٥)

الكفايات الرئيسية و الكفايات المتفرعة منها

الكفايـــــة الرئيسية	المعـــارف و المهـــارات اللازمـــة لممارســـة تجربة التعلم	إعـداد جـدول البيانات	رسـم منحنـى التعلم	التطـبيق علـى منحنى التعلم	معرفة التطبيقات التربوية لتجربة التعلم
الكفايـــات الفرعية	معـرفة شـكل التجـربة- تصنيف تجربة التعلم فى ضوء النظرية التابع لهـا- تحديـد متغيـرات التجـربة- تحديد منحنيات الـتعلم فـى التجـربة- معـرفة الخطـوات الإجـرائية للتجربة.	تحديـد عـدد صـفوف الجدول- تحديد عـدد أعـمدة الجـدول-رسم جدول البيانات- تـدوين بيانات التجـربة فـى الجدول.	معـرفة العلاقة بـين متغيـرات التجـربة و المحاور الأفقية و الرأسية-رسم المحور الأفقي و المحـور الرأسـي-تحيد مقـياس الرسم المناسـب-الدقة فـى وضع نقاط المنحنى-توصيل نقـاط المنحنى- توضـيح جـزء المنحنـى الذى يمـثل حـدوث التعلم.	التمييـز بـين المنحنـى الايجابـي و المنحنـى السـلبي-تحديد مـستوى التعلم مـن المنحنى- تحديـد سرعة الـتعلم مـن المنحنـى-تحيد الذبـذبات و الهـضاب على المنحنى-معرفة حـدود أقـصى الـتعلم-التمييز بـين المنحنـى المنـتظم و غير المنتظم.	استخلاص أهـداف تربوية لتجـربة التعلم- تحديـد العوامل العقـلية و غير العقـلية التى من الممكن أن تؤثر على التعلم أثناء ممارســة التجـربة-تقديم أمـثلة تـربوية علـى شـكل التجـربة-تقديم أمـثلة تـربوية علـى قوانـين الـتعلم للنظرية الـتابعة لهـا التجربة.

الخطوة الثالثة: صياغة الأهداف السلوكية التى تعكس الكفايات الفرعية: بعد أن تم تحليل الكفايات الرئيسية المكونة لمحتوى الاختبار إلى كفايات(مهارات) فرعية، فإن الخطوة التالية هى صياغة أهداف سلوكية تعكس هذه الكفايات ، و لقد قام الباحث بصياغة (٥٧) هدفاً سلوكياً تغطى هذه الكفايات الفرعية ، بحيث أن الكفاية الأولى(المعارف و المهارات اللازمة لممارسة تجربة التعلم) يعبر عنها (١٧) هدف سلوكى، و الكفاية الثانية(إعداد جدول البيانات) يعبر عنها (٦) أهداف سلوكية، و الكفاية الثالثة(رسم منحنى التعلم) يعبر عنها(١٣) هدف سلوكى، و الكفاية الرابعة (التعليق على منحنى التعلم) يعبر عنها (١٥) هدف سلوكى، ، و الكفاية الخامسة(معرفة التطبيقات التربوية لتجربة التعلم) يعبر عنها (٦) أهداف سلوكية.

الخطوة الرابعة: بناء المفردات الاختبارية:قام الباحث فى هذه الخطوة بكتابة المفردات الاختبارية، و التى تمثل انعكاس للأهداف السلوكية التى تم وضعها، بحيث أن كل هدف سلوكى من الأهداف السلوكية التى تم وضعها (٥٧) هدف سلوكى يعبر عنه بمفردة اختبارية أو أكثر من نوع الاختيار من متعدد ذى الأربعة بدائل ، و لقد راعى الباحث الأسس العلمية فى صياغة المفردات ذات الاختيار من متعدد و التى منها:

١-كل مفردة ينبغى أن تقيس ناتج مهم أو هدف للتعلم.

٢-ينبغى أن يقدم رأس السؤال فكرة واضحة فى لغة بسيطة.

٣-تأكد من أن كل البدائل متفقة نحوياً مع رأس السؤال و ليس البديل الصحيح فقط.

٤-لا تستخدم العبارات المطلقة(عادة-كل-أبداً -مطلقاً).

٥-احرص على أن تكون البدائل قابلة للتفكير فيها كبديل صحيح.

٦-لا تستخدم(كل ما سبق) كبديل من البدائل.

٧-لا تجعل البديل الصحيح يأخذ نفس الرقم فى كل المفردات.

٨-اجعل المفردات مستقلة بقدر الإمكان عن بعضها .

(Davis,1983,570-571)

و الجدول التالى يوضح العلاقة بين الكفايات الرئيسية و الأهداف السلوكية و المفردات الاختبارية:

جدول(٦) الكفايات الرئيسية و الأهداف السلوكية و المفردات الاختبارية

البيان / الكفاية	الأهداف السلوكية	العدد	أرقام المفردات المقابلة*	العدد
المعارف و المهارات اللازمة لممارسة تجربة التعلم	(يسمى تجربة تعلم موضحة بالشكل- يتعرف على مكونات تجربة معطاة- يميز بين تجربة التعلم التى تعتمد على مصدر تيار كهربائى و التى لاتعتمد على مصدر تيار كهربائى - يذكر اسم النظرية التى تتبعها التجربة المعطاة- يذكر نوع التعلم السائد لتجربة تعلم معطاة- يصنف مجموعة من تجارب تعلم معطاة فى ضوء نوع التعلم السائد- يصنف مجموعة من تجارب تعلم معطاة فى ضوء نظرية التعلم التابعة لها- يسمى المتغير المستقل الأساسى فى تجارب التعلم- يذكر عدد المتغيرات التابعة لتجربة تعلم معطاة- يسمى المتغير التابع (المتغيرات التابعة) لتجربة تعلم معطاة- يحدد مجموعة تجارب التعلم التى تشترك فى متغير تابع او أكثر- يحدد المتغير التابع (المتغيرات التابعة) التى يشترك بين مجموعة من التجارب المعطاة- يذكر عدد منحنيات التعلم لتجربة معطاة- يتعرف على العلاقة بين عدد منحنيات التعلم و عدد المتغيرات التابعة- يسمى منحنى(منحنيات) التعلم لتجربة معطاة- يتعرف على الترتيب المنطقى لخطوات إجراء تجربة تعلم معطاة- يميز بين دور الفاحص و دور المفحوص أثناء أداء التجربة	١٧	١-٣٢	٣٢
إعداد جدول البيانات	يتعرف على عدد صفوف جدول البيانات لتجربة معطاة- يتعرف على عدد أعمدة جدول البيانات لتجربة معطاة- يدون البيانات فى الجدول بوحدات مناسبة- يدون بيانات المتغيرات فى الجدول بترتيب منطقى- يميز بين دور الفاحص و دور المفحوص أثناء تدوين بيانات التجربة- يكتشف رقم المحاولة التى عندها يتم إيقاف التجربة	٦	٣٣-٤٢	١٠
رسم منحنى التعلم	يتعرف على العلاقة بين المتغير المستقل و المحور الأفقى--يتعرف على العلاقة بين المتغير التابع و المحور الرأسى- يرسم المحور الأفقى فى منحنى التعلم- يرسم المحور الرأسى فى منحنى التعلم- يعرف الأساس الذى عليه يتم اختيار مقياس الرسم- يقترح مقياس رسم مناسب فى ضوء بيانات معطاة- يفهم معنى بعد نقطة المنحنى عن المحور الأفقى- يفهم معنى بعد نقطة المنحنى عن المحور الرأسى- يعرف القيمتين اللتين على أساسهما يتم وضع نقطة المنحنى- يدقق فى وضع النقطة على المنحنى- يكتشف الشكل الصحيح لمنحنى التعلم- يستخرج الجزء من المنحنى الذى يمثل حدوث التعلم- يميز بين حدود أقصى التعلم	١٣	٤٣-٥٨	١٦
التعليق على منحنى التعلم	يعرف بداية المنحنى الايجابى- يعرف بداية المنحنى السلبى- يميز بين المنحنى الايجابى و المنحنى السلبى إذا علم بدايته- يحدد مستوى التعلم من منحنى تعلم معطى- يحدد سرعة التعلم من منحنى تعلم معطى- يقارن بين سرعة التعلم لدى مفحوصين من بيانات معطاة- يقارن بين مستوى التعلم لدى مفحوصين من بيانات معطاة- يعرف ذبذبة التعلم- يعرف هضبة التعلم- يحدد ذبذبة(أو ذبذبات) التعلم على منحنى معطى- يحدد هضبة (أو هضاب) التعلم على منحنى معطى- يدرك عدد النقاط اللازمة لتمثيل ذبذبة التعلم- يدرك عدد النقاط اللازمة لتمثيل هضبة التعلم- يتعرف على خصائص المنحنى المنتظم- يتعرف على خصائص المنحنى غير المنتظم	١٥	٥٩-٧٣	١٥
معرفة التطبيقات التربوية لتجربة التعلم	يستخرج هدف تعليمى من إجراء التجربة- يسرد العوامل العقلية التى من الممكن أن تؤثر على التعلم فى تجربة معطاة- يسرد العوامل غير العقلية التى من الممكن أن تؤثر على التعلم فى تجربة معطاة- يحدد قانون أو أكثر لنظرية التعلم التابع لها التجربة- يستخلص تطبيقات تربوية على قوانين التعلم للنظرية التى تتبعها التجربة- يستخلص تطبيقات تربوية من شكل التجربة	٦	٧٤-٨٧	١٤
المجموع		٥٧		٨٧

* انظر ملحق البحث للتعرف على مفردات الاختبار.

الخطوة الخامسة (ثبات و صدق الاختبار):

(١)-**صدق الاختبار**:تم استخدام أسلوب الصدق الوصفى *Descriptive Validity* للتحقق من صدق الاختبار، و يعنى الصدق الوصفى للاختبار محكى المرجع قدرة الاختبار على وصف أداء الفرد بالنسبة لنطاق سلوكى يقيسه الاختبار و يكون محدد تحديداً دقيقاً، و هذا النوع من الصدق يعد أساساً لأنواع الصدق الأخرى، فالاختبار الذى يفتقر إلى الصدق الوصفى لا يصلح للغرض المشار إليه،و بالتالى لا يكون هناك ثقة فى نتائجه(صلاح الدين علام، ١٩٩٥، ٢٥٥).

و لقد اتبع الباحث الخطوات التالية للتحقق من الصدق الوصفى فى ضوء ما أوضحه صلاح الدين علام (١٩٩٠، ١٤٣-١٤٥) كالتالى: تم عرض استمارة على المحكمين (الذين قاموا بتدريس التدريبات المعملية لعلم النفس) مبين عليها الهدف العام من الاختبار، و كفاياته الرئيسية الخمس، و الأهداف السلوكية التى تنطوى تحت كل كفاية، و كذلك المفردات الاختبارية التابعة لكل كفاية و كل هدف سلوكى، ثم تم الطلب منهم الإجابة على السؤالين التاليين:

(١)-ما مدى انتماء كل هدف سلوكي للكفاية الرئيسية التابع لها.

(٢)-)- ما مدى انطباق كل مفردة اختبارية على الهدف السلوكى التابعة له.

و يقوم المحكمون بالاستجابة للسؤالين السابقين بنعم(√) أو لا(x) أمام كل هدف سلوكى و كل مفردة ، و فيما يلى جدولان يوضحان نتائج السؤالين الأول و الثانى:

جدول(٧)

نتائج السؤال الاول

مدى انتماء الهدف السلوكي للكفاية الرئيسية التابع لها

رقم الهدف	الاتفاق(نعم)	رقم الهدف	الاتفاق(نعم)	رقم الهدف	الاتفاق(نعم)	رقم الهدف	الاتفاق(نعم)	رقم الهدف	الاتفاق(نعم)
١	٩٢	١٣	٩٢	٢٥	٩٢	٣٧	٩٢	٤٩	٨٣
٢	٩٢	١٤	٧٥	٢٦	٩٢	٣٨	١٠٠	٥٠	٨٣
٣	٨٣	١٥	٩٢	٢٧	٩٢	٣٩	٩٢	٥١	٩٢
٤	٩٢	١٦	١٠٠	٢٨	٨٣	٤٠	١٠٠	٥٢	٩٢
٥	١٠٠	١٧	١٠٠	٢٩	٩٢	٤١	١٠٠	٥٣	١٠٠
٦	٩٢	١٨	٨٣	٣٠	٨٣	٤٢	٩٢	٥٤	٩٢
٧	١٠٠	١٩	٨٣	٣١	٩٢	٤٣	٧٥	٥٥	٩٢
٨	٩٢	٢٠	٩٢	٣٢	٩٢	٤٤	٩٢	٥٦	١٠٠
٩	٩٢	٢١	٨٣	٣٣	٩٢	٤٥	١٠٠	٥٧	٩٢
١٠	٩٢	٢٢	٥٨	٣٤	١٠٠	٤٦	١٠٠		
١١	٩٢	٢٣	١٠٠	٣٥	١٠٠	٤٧	١٠٠		
١٢	٩٢	٢٤	٩٢	٣٦	٧٥	٤٨	١٠٠		

جدول(٨)

نتائج السؤال الثانى

انطباق كل مفردة اختبارية على الهدف السلوكى التابعة له

رقم المفردة	الاتفاق(نعم)	رقم المفردة	الاتفاق(نعم)	رقم المفردة	الاتفاق(نعم)	رقم المفردة	الاتفاق(نعم)	رقم المفردة	الاتفاق(نعم) %
١	١٠٠	١٩	٩٢	٣٧	٩٢	٥٥	١٠٠	٧٣	١٠٠
٢	١٠٠	٢٠	٩٢	٣٨	٩٢	٥٦	١٠٠	٧٤	١٠٠
٣	١٠٠	٢١	١٠٠	٣٩	٩٢	٥٧	٩٢	٧٥	١٠٠
٤	١٠٠	٢٢	١٠٠	٤٠	١٠٠	٥٨	٩٢	٧٦	١٠٠
٥	٨٣	٢٣	١٠٠	٤١	٩٢	٥٩	٩٢	٧٧	١٠٠
٦	٨٣	٢٤	١٠٠	٤٢	٩٢	٦٠	٩٢	٧٨	١٠٠
٧	١٠٠	٢٥	٩٢	٤٣	١٠٠	٦١	١٠٠	٧٩	٩٢
٨	١٠٠	٢٦	١٠٠	٤٤	١٠٠	٦٢	١٠٠	٨٠	١٠٠
٩	٩٢	٢٧	١٠٠	٤٥	٩٢	٦٣	١٠٠	٨١	٩٢
١٠	١٠٠	٢٨	١٠٠	٤٦	٨٣	٦٤	١٠٠	٨٢	٩٢
١١	٩٢	٢٩	١٠٠	٤٧	١٠٠	٦٥	١٠٠	٨٣	١٠٠
١٢	١٠٠	٣٠	١٠٠	٤٨	١٠٠	٦٦	١٠٠	٨٤	١٠٠
١٣	١٠٠	٣١	٩٢	٤٩	١٠٠	٦٧	١٠٠	٨٥	١٠٠
١٤	٩٢	٣٢	٩٢	٥٠	١٠٠	٦٨	١٠٠	٨٦	١٠٠
١٥	١٠٠	٣٣	٨٣	٥١	٩٢	٦٩	١٠٠	٨٧	١٠٠
١٦	٩٢	٣٤	٨٣	٥٢	١٠٠	٧٠	١٠٠		
١٧	٩٢	٣٥	٩٢	٥٣	٩٢	٧١	٩٢		
١٨	٩٢	٣٦	٩٢	٥٤	١٠٠	٧٢	١٠٠		

يلاحظ من الجدولين السابقين أن الأهداف السلوكية و المفردات الاختبارية تحظى بنسب اتفاق مرتفعة بين المحكمين، و أغلبها تصل إلى نسبة الاتفاق التام(١٠٠%)، و هذا يدل على صدق المفردات فى تمثيلها للنطاق السلوكى الذى يقيسه الاختبار، كما طلب الباحث من المحكمين إجراء أي تعديل أو حذف أو إضافة للمفردات الاختبارية بما يتوافق مع الأهداف السلوكية فى ضوء خبرتهم فى تدريس هذه المجال السلوكي، و لقد قام الباحث بإجراء بعض التعديلات على بعض المفردات الاختبارية في ضوء آراء المحكمين،و بذلك تم التحقق من الصدق الو صفى للاختبار.

(٢)- **ثبات المقياس**:

قام الباحث بالتحقق من ثبات المقياس عن طريق معامل الاتفاق لسابكوفياك*Subkoviak Coefficient of Agreement* ، و هى إحدى طرق ثبات الاختبارات محكية المرجع، و التى تقتضى تطبيق الاختبار مرة واحدة فقط،و لقد قام الباحث باتباع الخطوات التى أوضحها صلاح الدين علام(١٩٩٥، ٢٨٨-٢٩٤)، و التى تعتمد فى الأصل على التوزيع ذى الحدين*Binomial Distribution* و هى كما يلى:

١- نحصل على تقدير لنسبة عدد مفردات النطاق الشامل التى أجاب عنها كل فرد من أفراد المجموعة إجابة صحيحة، و ذلك باستخدام الصيغة ل ن= س ن/ن ، حيث س ن عدد المفردات التى أجاب عليها المفحوص(ف) إجابة صحيحة، ن العدد الكلى لمفردات الاختبار.

٢- نحصل على مقدار احتمال حصول الفرد على درجة أكبر من أو تساوى درجة القطع(س.) باستخدام الصيغة الرياضية لتوزيع ذات الحدين.

٣-نحصل على معامل سابكوفياك الخاص بتصنيف كل فرد (ف)على حدة باستخدام القيمة الناتجة عن الخطوة الثانية فى الصيغة:

ل س.(ف)= [ل(س ف ≤ س ٠)]٢+[١ - ل(س ف ≤ س ٠)]٢ ،حيث أن ل(س ف ≤ س.) هى توزيع ذات الحدين.

٤- نحصل على معامل سابكوفياك الخاص بالمجموعة ككل ، باستخدام الصيغة

ل س.= مجـــ ل س. (ف)ان، و هى تعنى متوسط معاملات الثبات لكل فرد على حدة ،حيث أن ل س. تمثل معامل سابكوفياك للثبات،و لقد تم استخدام برنامج *Excel* فى إجراء هذه الخطوات ، و تم التوصل إلى معامل ثبات قدره ٠,٩٦٧ عند درجة القطع (٧٩ %) (باستخدام طريقة أنجوف)، و معامل ثبات قدره ٠,٧٩٨ عند درجة القطع (٥٤ %) باستخدام طريقة ندلسكاى ، و هى معاملات ثبات مرتفعة و تشير إلى ثبات الاختبار.

مما سبق يتضح أن الاختبار محكى المرجع فى التدريبات المعملية لعلم النفس يحظى بدرجة مقبولة من الصدق و الثبات تجعلنا نثق فى نتائجه.

٤-الصورة النهائية للاختبار:

يتكون الاختبار فى صورته النهائية من (٨٧) مفردة رباعية البدائل ، و يحصل الطالب على الدرجة(١) فى حالة اختيار البديل الصحيح للمفردة ، و يحصل على الدرجة(٠) فى حالة اختيار بديل خاطئ للمفردة ، و بذلك تتراوح درجة الاختبار بين (٠) ، و (٨٧) و التى تمثل العدد الكلى للمفردات (أقصى درجة يمكن الوصول إليها) ، و الاختبار غير مقيد بزمن .

الأداة الثانية :اختبار ذكاء الشباب اللفظى

(إعداد:حامد زهران:١٩٧٧)

يتكون هذا الاختبار من (١٠٠) مفردة من نوع الإكمال، و هو من الاختبارات اللفظية، و يطبق بصورة جماعية، و يتطلب الاستجابة له (٣٠) دقيقة ، بالإضافة إلى (١٥) دقيقة للتعليمات ، و لقد تم استخدام هذا

الاختبار فى البحث الحالى لكى يتم ضبط متغير الذكاء بين المجموعتين المتعلمة و غير المتعلمة ، نظراً لإمكانية دخول متغير الذكاء كعامل مؤثر فى عملية تخمين الإجابة ، و خاصة فى المجموعة غير المتعلمة،و لقد تم التحقق من صدق الاختبار عن طريق الصدق المرتبط بمحك على عينة من الأفراد بلغ قوامها (١٩) طالب، و كان المحك هو اختبار تونى لقياس الذكاء(إعداد:محمد شوكت، ١٩٨٧)، وتوصل الباحث إلى معامل صدق قدره(٠,٥٧٩) و هى قيمة دالة احصائياً عند مستوى٠,٠١ ،كما تم التحقق من ثبات المقياس بطريقة التجزئة النصفية لسبيرمان-براون و تم التوصل إلى معامل ثبات قدره(٠,٨٤) ، و من ثم يتميز الاختبار بمعاملات صدق و ثبات مقبولة.

ثالثاً :إجراءات التطبيق:

١-تم عرض الاختبار المحكي المرجع في التدريبات المعملية لعلم النفس بعد التحقق من صدقه و ثباته على مجموعة المحكمين البالغ عددهم(١٢) محكم لتقدير مفرداته بطريقتى أنجوف و ندلسكاى ، ولقد تم فى هذه الخطوة شرح لطريقتى أنجوف و ندلسكاى للمحكمين مدعماً بمثال تطبيقى، كما تم شرح لهم مفهوم (الطالب ذى الحد الأدنى من الكفاية)، و هو الطالب الذى يمتلك حد أدنى من المهارات و الكفايات(المعارف و المهارات اللازمة لممارسة تجربة التعلم،إعداد جدول البيانات، رسم منحنى التعلم، التعليق على منحنى التعلم، معرفة التطبيقات التربوية لتجربة التعلم) التى تؤهله لتطبيق مفهوم التعلم عملياً داخل معمل علم النفس.

٢- تم الطلب من المحكمين تقدير كل مفردة كالاتى:

*تحديد النسبة المئوية لاحتمال إجابة(الطالب ذى الحد الأدنى من الكفاية) إجابة صحيحة على كل مفردة،و بذلك تصبح درجة القطع المحددة بواسطة كل محكم هى مجموع هذه النسب المئوية(فى صورة

كسر عشرى) خلال جميع المفردات، أما درجة القطع النهائية فهى متوسط درجات القطع المحددة بواسطة المحكمين(طريقة أنجوف).

*تحديد عدد المشتتات (البدائل الخاطئة) التى سيستبعدها (الطالب ذو الحد الأدنى من الكفاية) تلقائياً،عندما يفكر فى اختيار بديل من البدائل الأربعة، و لقد تم تمييز البديل الصحيح عن المشتتات بخط مغاير، ،و بذلك تصبح درجة القطع المحددة بواسطة كل محكم هى مجموع مقلوبات عدد البدائل المتبقية ، أما درجة القطع النهائية فهى متوسط درجات القطع المحددة بواسطة المحكمين(طريقة ندلسكاى).

٣-تم تقدير درجات القطع على الاختبار و ذلك بالنسبة لكل محكم، و كذلك تم حساب درجتى القطع النهائيتين للاختبار، بطريقتى أنجوف ، و ندلسكاى .

٤-تم تطبيق اختبار ذكاء الشباب الفظي على أفراد عينة المفحوصين (المجموعة المتعلمة''١٠٠'' طالب-المجموعة غير المتعلمة''٨٨'' طالب)،و ذلك لضبط متغير الذكاء بين المجموعتين، و بعد تطبيق الاختبار و تصحيحه و جد الباحث أن طلاب المجموعتين يدخلون فى فئة نسبة الذكاء(٩٠-١١٠) ، باستثناء (٦) طلاب ، منهم (٤) طلاب من المجموعة المتعلمة، و طالبين من المجموعة غير المتعلمة ، و ذلك لكى يكون هناك تقارب فى مستوى الذكاء بين المجموعتين، كما تم استبعاد (٤) طلاب آخرين من المجموعتين لعدم جديتهم فى الاستجابة لمفردات اختبار الذكاء، و بذلك تصبح المجموعتان متقاربتين إلى حد ما الذكاء، و مع تقارب المجموعتين فى العمر بحكم أنهما من فرقة واحدة ،و كذلك فى الخلفية العلمية فكلاهما من التخصص العلمى، بذلك فهم يختلفان فقط فى المتغير التجريبي(البرنامج التعليمى(سكاشن معمل علم النفس)، و الذى يميز مجموعة طلاب التعليم العام(المجموعة المتعلمة).

٥-بعد استبعاد الحالات التى لا تخضع للدراسة التجريبية موضوع البحث، تم تحديد أسماء مجموعة المتعلمين و البالغ قوامهم(٩٣) طالب، من طلاب الفرقة الثالثة شعبة التعليم لعام''القسم العلمى''، و كذلك تم تحديد أسماء مجموعة غير المتعلمين و البالغ قوامهم(٨٥) طالب من طلاب الفرقة الثالثة شعبة التربية الرياضية.

٦-تم إعطاء الدرجة(١) لكل فرد فى المجموعة المتعلمة، و كذلك الدرجة(٠) لكل فرد فى المجموعة غير المتعلمة، و هو ما يعبر عن متغير التصنيف المحك.

٧-تم تطبيق الاختبار المحكي المرجع في التدريبات المعملية لعلم النفس على أفراد المجموعتين، و تم رصد درجاتهم على الاختبار.

٨-تم إعطاء الدرجة (١) لأى فرد فى المجموعتين(المتعلمة أو غير المتعلمة) يحصل على درجة القطع النهائية أو ما يزيد عنها ، كما تم إعطاء الدرجة(٠) لأى فرد فى المجموعتين(المتعلمة أو غير المتعلمة) يحصل على درجة أقل من درجة القطع، و هو ما يعبر عن متغير التصنيف المنبئ، و ذلك لكل طريقة على حدة(أنجوف و ندلسكاى).

٩-تم معالجة البيانات إحصائياً للتحقق من صحة الفروض الموضوعة.

نتائج البحث

نتيجة الفرض الأول :

ينص هذا الفرض على الآتى: تختلف درجة القطع باختلاف الطريقة المستخدمة فى تحديد هذه الدرجة ''أنجوف - ندلسكاى '' .

للتحقق من صحة هذا الفرض تم عرض الاختبار محكى المرجع فى التدريبات المعملية لعلم النفس على المحكمين (ن = ١٢) ، لكى يقوموا بتقدير مفرداته بطريقتى أنجوف و ندلسكاى ، ثم تم تحليل استجابات المحكمين و التوصل إلى درجة قطع خاصة بكل محكم ، و كذلك متوسط

لـــدرجات القطع على كل المحكمين و هذا بالنسبة لكل طريقة على حدة ، و هذه البيانات معرضة فى الجدول الآتى:

جدول (٩)

درجات القطع المحددة بواسطة المحكمين

البيان / المحكم	درجة القطع	
	أنجوف	ندلسكاى
١	٧٢	٦٠
٢	٦٦	٦٢
٣	٧٢	٣٩
٤	٧٩	٧١
٥	٦٩	٢٧
٦	٧٧	٢٣
٧	٧٥	٤٠
٨	٧١	٤٤
٩	٦٠	٥٠
١٠	٦٨	٥٢
١١	٦٤	٥١
١٢	٥٧	٤٣
المتوسط	٦٩	٤٧
الانحراف المعيارى	٦,٦٢	١٣,٩٣
التباين	٤٣,٨	١٩٣,٩٧

و لقد تمت المقارنة بين درجة القطع الناتجة عن طريقة أنجوف و درجة القطع الناتجة عن طريقة ندلسكاى عن طريق الأسلوب الإحصائى اللابارامترى اختبار ولكوكسن *Wilcoxon Test* ، لأن حجم العينة صغير (ن = ١٢)، حيث أشار صلاح الدين علام أنه فى حالة العينات الصغيرة التى يتراوح حجمها بين (٥ – ٥٠) ، يتم استخدام هذا الاختبار و الذى يعد بديلاً لا بارامترياً للنسبة التائية، و نظراً لأن المجموعتين مرتبطتين ، و بالتالى يصبح بديلاً بارامترياً للنسبة التائية لمجموعتين مرتبطتين عندما لا تتحقق شروط الاختبار البارامترى(صلاح الدين علام ، ٢٠٠٥ ، ٢٥٨ –٢٦٠) ، و لقد تم استخدام حزمة البرامج الإحصائية للعلوم الاجتماعية *SPSS* فى إجراء هذا التحليل ، و فيما يلى جدول يوضح نتيجة هذا الفرض:

جدول(١٠)

نتيجة الفرض الاول

مجموع الرتب	متوسط الرتب	حجم العينة(ن)	النسبة الحرجة	الدلالة عند مستوى الطرفين
٧٨	٦,٥	١٢	٣,٠٦	٠,٠١

و بذلك يتضح صحة هذا الفرض لأن قيمة النسبة الحرجة (٣,٠٦) دالة إحصائياً عند مستوى

(٠,٠١) ،و بالتالى تختلف درجة القطع الناتجة عن طريقة أنجوف عن درجة القطع الناتجة عن طريقة ندلسكاى ، و ذلك لصالح طريقة أنجوف ، و هذا يعنى أن استخدام طريقة أنجوف يؤدى إلى تحديد درجات قطع أعلى من درجات القطع المحددة بواسطة طريقة ندلسكاى .

و يفسر الباحث هذه النتيجة على أساس أنه إذا كانت طريقة ندلسكاى تتطلب من المحكم تحديد عدد المشتتات التى سيستبعدها الطالب ذو الحد الأدنى من الكفاية تلقائيا أثناء التفكير فى اختيار البديل الصحيح و بالتالى تتطلب من المحكم فهم لمفهوم الطالب ذى الحد الأدنى من الكفاية ، فان الفـرق بين الطريقتين و الذى جعل طريقة أنجوف تؤدى إلى درجة قطع مـرتفعة مقارنـة بطـريقة ندلسكاى ، هو أن طريقة أنجوف تتطلب من المحكمـين تقديـر المفردة دون التفحص الدقيق لبدائل هذا المفردة. أما طـريقة ندلـسكاى و خاصـة فى الاختبار ذات المفردات رباعية البدائل فهـناك أربعة احتمالات لتحديد عدد المشتتات لكل مفردة و يقابله أربع قـيم احتمالـية على المفردة بناءً على عدد المشتتات ، فعندما لا يحدد المحكـم مـشتتات تكون عدد البدائل المتبقية (٤) ، و مقلوبها (٠,٢٥) يمـثل القيمة الاحتمالية على هذا المفردة ، أما عندما يحدد مشتت واحد تكـون القيمة الاحتمالية (٠,٣٣) ، و عندما يحدد مشتتين تكون القيمة الاحتمالية (٠,٥) ، و عندما يحدد ثلاثة مشتتات تكون القيمة الاحتمالية (١)*(chang ,1996)* .

و يـشير *Shepard* فى عام ١٩٨٤ إلى أن غالبية المحكمين يميلون إلى اختـيار مشتتين و الذى يقابله قيمة احتمالية على المفردة قيمتها(٠,٥)، و الـذى يؤدى بدوره إلى انخفاض درجة القطع المحددة بواسطة طريقة ندلـسكاى *(In:Berk , 1986 , 152)*، فطـريقة ندلـسكاى تسمح فقط بقيم احتمالـية منفـصلة و هى(٠,٢٥ ، و ٠,٣٣ ، و ٠,٥ ، و ١) بحسب عـدد بدائل المفردة ، و لا تسمح بقيم تنحصر بين (٠,٥) و (١) ، و بالتبعـية لا يميل معظم المحكمين فى كثير من الأحيان إلى تعيين القيمة الاحتمالية(١) (أى مستوى الإتقان ١٠٠%) ، و لذلك يضطرون لتعيين القـيمة الاحتمالية (٠,٥) لمعظم المفردات ، و حيث أن درجة القطع هى فـى الواقع مجموع هذه القيم الاحتمالية على كل المفردات ، لذلك تكون

درجـــة القطـــع المحددة بواسطة ندلسكاى أقل من درجة القطع المحددة بواسطة الطرق الأخرى(صلاح الدين علام ، ١٩٩٥ ، ٢٣٦) .

و ربما يكون الفرق فى درجة القطع بين طريقتى أنجوف و ندلسكاى دالة لمعـــرفة المحكمين بالمفردة و محتواها ، و قدرتهم فى الإجابة عليها ، حـــيث يـــضع المحكمـــون قـــيماً احتمالية أعلى للمفردات الملمين بها و بمحـــتواها ، و فـــى المقابل يضعون قيماً احتمالية أقل للمفردات التى لا يعـــرفون الإجابة عليها و تمثل مفردات صعبة بالنسبة لهم ، و التفحص الكامـــل للبدائل فى طريقة ندلسكاى و ملاحظة المحكم مدى التشابه بين الـــبدائل و منطقية كل بديل كإجابة محتملة تخلق تحدى لدى المحكم ، و الـــذى يـــؤدى بـــدوره إلى اقتناعه بعدم سهولة المفردة نظراً للمشتتات المتشابهة و التى تحير المفحوص، و بالتبعية يحدد المحكم قيماً احتمالية أقل و الذى يؤدى إلى انخفاض درجة القطع المحددة بواسطة ندلسكاى ، و العكس يحدث فى طريقة أنجوف ، حيث أن عدم تفحص المحكم لبدائل المفـــردة و ما بينها من فروق طفيفة و حضور الإجابة له أثناء تقدير المفـــردة ، تجعـــل المحكـــم يقدر تقدير مبالغ فيه *overestimate* لسهولة المفـــردة ، و بالتبعية يحدد المحكم قيماً احتمالية أعلى و الذى يؤدى إلى ارتفـــاع درجـــة القطع المحددة بواسطة أنجوف، و لذلك نجد أن نقص خبـــرة المحكـــم بالمفردة و ما تحويه من مفاهيم له تأثير أقل على قرار التـــصنيف المـــتخذ فى ضوء طريقة أنجوف ، بعكس طريقة ندلسكاى و التـــى نجد فيها أن الصعوبة التى ربما يجدها المحكم فى مفردة ما تؤثر فى قراره و الذى يؤدى إلى انخفاض درجة القطع *(Chang , 1996)* .

كمـــا أشار *(Abbott , 2003)* إلى أن درجة القطع المحددة بواسطة طريقة أنجـــوف صعب اجتيازها لارتفاعها ، و هذا السبب جعل طريقة أنجوف تحتاج إلى بيانات معيارية أو امبريقية عن المفحوصين لكى يسترشد بها المحكمون فى حكمهم ، فعندما تتاح هذه البيانات للمحكمين ربما يحددون درجة قطع بصورة أكثر موضوعية.

و لقـد اتفقـت نتائج العديد من الدراسات مع نتيجة الفرض الحالى فى اخـتلاف درجة القطع المحددة بواسطة طريقة أنجوف عن درجة القطع المحـددة بواسطة طريقة ندلسكاى ،و هنا انقسمت النتائج إلى جانبين ، الأول و هـو الغالـب، و يشير إلى أن الفرق بين درجتى القطع المحدد بالطـريقتين كان لصالح طريقة أنجوف ،و هذا القسم من الدراسات مبين فى الجدول التالى

جدول (١١)

بعض الدراسات التى تشير إلى ارتفاع درجة القطع المحددة بواسطة طريقة أنجوف مقارنة بندلسكاى

البيان / الدراسة	عدد مفردات الاختبار	أنجوف		ندلسكاى	
		درجة القطع (نسبة مئوية)	درجة القطع (درجة)	درجة القطع (نسبة مئوية)	درجة القطع (درجة)
(صلاح الدين علام،١٩٩١)	٦٠	٧١,٧ %	٤٣	٦٦,٧	٤٠
(Chang,1996)	٩	٧١ %	٦ (بالتقريب)	%٥٧	٥ (بالتقريب)
(Donnoe&Amato,1997)	مسحوب من بنك أسئلة	٥٠ %	--	%٣٣	--
(brennan & lockwood ,1980)	١٢٦	%٦٧	٨٤	%٥٦	٧٠
(Livingston & zieky ,1989)	اختبار الرياضيات (٧٠)	%٥١	٣٦	%٤٤	٣١
الدراسة الحالية	٨٧	٧٩ %	٦٩	٥٤ %	٤٧

و بالـــرغم من أن هذه الدراسات تشير إلى ارتفاع درجة القطع المحددة بطـــريقة أنجوف مقارنة بدرجة القطع المحددة بطريقة ندلسكاى ، إلا أن هناك جانب اخرمن الدراسات توصل إلى انخفاض درجة القطع المحددة بطـــريقة أنجـــوف مقارنة بطريقة ندلسكاى ، منها دراسة *(Livingston &Zieky ,1989)* التى توصلت إلى انخفاض درجة القطع المحددة بطريقة أنجـــوف مقارنة بطريقة ندلسكاى و ذلك فى اختبار القراءة فقط ، و لكن فـــى اختـــبار الرياضيات توصلت إلى عكس النتيجة كما هو موضح فى الجـــدول ،فبالنـــسبة لاختبار القراءة توصلت إلى درجة القطع باستخدام طـــريقة أنجـــوف قدرها(٣٤) ، و باستخدام طريقة ندلسكاى كانت درجة القطـــع(٤٢) .و دراسة *smith & smith* فى عام ١٩٨٨التى توصلت إلى أن درجة القطع المحددة بواسطة طريقة أنجوف قدرها (٦٩ %)، و باســـتخدام طريقة ندلسكاى قدرها(٧١%)*(in: chang, 1996)*. و قد يرجع اخـــتلاف نتيجة هاتين الدراستين مع نتيجة الدراسة الحالية إلى اختلاف محتوى الاختبار، و اختلاف طريقة القاء التعليمات على المحكمين .

و مجمـــل مـــا ســـبق هو أن الفرض الذى تم وضعه و الذى يشير إلى اختلاف درجة القطع الناتجة عن طريقة أنجوف عن درجة القطع الناتجة عـــن طريقة ندلسكاى صحيح ، و جميع نتائج الدراسات التى تم عرضها تـــؤيد هـــذه الفـــرض، ومعظمها يشير إلى أن هذا الفرق لصالح طريقة أنجـــوف ، أما العدد القليل منها فيشير إلى أن هذا الفرق لصالح طريقة ندلسكاى .

نتيجة الفرض الثانى:

و الذى ينص على أنه: تختلف مؤشرات قرار التصنيف'' الخطأ الموجب و الخطـــأ السالب و القرار الصحيح و معامل الصدق و دوال الضرر و المـــنفعة و المـــنفعة العظمـــى'' باختلاف الطريقة المستخدمة فى تحديد درجة القطع ''طريقة أنجوف – طريقة ندلسكاى''.

أشـــارت نتيجة الفرض الأول إلى أن طريقة أنجوف تؤدى إلى درجة قطع قيمتها(٦٩) و هى أعلى من درجة القطع المحددة بطريقة ندلسكاى و التى قيمتها (٤٧) ، و نتيجة الفرض الأول ترتبط ارتباط كلى بنتيجة الفـــرض الثانـــى ، بمعنـــى أن ارتفاع درجة القطع الناتجة عن طريقة أنجوف ، و ما يقابلها من انخفاض فى درجة القطع فى طريقة ندلسكاى ، هو الذى سيجعل هناك اختلاف فى مؤشرات قرار تصنيف الطلاب فى الطريقتين ، و الذى يصب فى صالح طريقة ندلسكاى ، و فى الواقع فان نتـــيجة الفرض الثانى تشير إلى اختلاف مؤشرات قرار تصنيف الطلاب الـــناتج عـــن الطريقتين ، باستثناء الخطأ الموجب للقرار ، و الذى كان معدوماً فى الطريقتين و فيما يلى تحليل للفرض الثانى الرئيسى إلى سبعة فروض فرعية و التحقق من صحتها كالتالي :

الفرض الفرعى الأول:

يخـــتلف الخطأ الموجب لقرار التصنيف باختلاف الطريقة المستخدمة فى تحديد درجة القطع ''طريقة أنجوف – طريقة ندلسكاى '' ؟.

للتحقق من صحة هذا الفرض تم استخدام المعادلة $P(FM) = FM/M+N$ ، حـــيث FM هـــى عدد القرارات الخاطئة من النوع الموجب، أما المقام $(M+N)$ فيمثل العدد الكلى للمجموعتين(المتعلمة و غير المتعلمة)، حيث $M = 93$ (عـــدد أفـــراد المجمـــوعة المـــتعلمة) ، $N = 85$ (عـــدد أفراد المجموعة غير المتعلمة) ، و بالتالى يكون العدد الكلى $(M+N) = 93 + 85 = 178$ ، و هـــذا العـــدد الكلى ثابت فى الطريقتين المستخدمتين، و الجدول التالى يوضح نتيجة هذا الفرض:

جدول (١٢)

نتيجة الفرض الفرعى الأول

البيان / الطريقة	عدد الأخطاء الموجبة *(FM)*	نسبة احتمال الخطأ الموجب لقرار التصنيف *p(FM)*
ندلسكاى	صفر	صفر
أنجوف	صفر	صفر

يتضح من الجدول السابق أنه لا توجد أخطاء موجبة فى الطريقتين المستخدمتين ، فكلا الطريقتين لا ينتج عنهما متقنون غير حقيقيين ، و هم الأفراد الذين لم يدرسوا البرنامج التعليمى(غير متعلمين) ، و بالرغم من ذلك يحصلون على درجة القطع أو ما يزيد عنها.

و يفسر الباحث هذه النتيجة على أنها راجعة إلى فعالية البرنامج التعليمى المطبق على المجموعة المتعلمة ، و الذى لم يطبق على المجموعة غير المتعلمة ، بالإضافة إلى الدقة فى تصميم الاختبار موضوع البحث ، و ما يحويه من مفردات ذات بدائل موضوعة على أسس علمية تمنع فرصة الطالب فى المجموعة غير المتعلمة فى أن يخمن البديل الصحيح على كل مفردة لأن الطالب فى المجموعة غير المتعلمة ستكون إجابته ليست اعتماداً على دراسة للمحتوى الذى يقيسه الاختبار ، و لكن بصورة عشوائية معتمدة على الحظ و الصدفة ، و من ثم تمنع فرصة حصول الطالب فى المجموعة غير المتعلمة على درجة القطع أو ما يزيد عنها ، أى تمنع حدوث الخطأ الموجب للقرار، لذلك نجد الخطأ الموجب منعدم فى طريقتى أنجوف و ندلسكاى .

الفرض الفرعى الثانى:

يختلف الخطأ السالب لقرار التصنيف باختلاف الطريقة المستخدمة فى تحديد درجة القطع ''طريقة ندلسكاى – طريقة أنجوف '' ؟.

للتحقق من صحة هذا الفرض تم استخدام المعادلة $P(FN) = FN/M+N$ ، حيث FN هى عدد القرارات الخاطئة من النوع السالب، و الجدول التالى يوضح نتيجة هذا الفرض:

جدول (١٣)

نتيجة الفرض الفرعى الثانى

البيان / الطريقة	عدد الأخطاء السالبة (FN)	نسبة احتمال الخطأ السالب لقرار التصنيف $P(FN)$
ندلسكاى	١٤	٠,٠٨
أنجوف	٨٣	٠,٤٧

يتضح من الجدول السابق أن احتمال الخطأ السالب لقرار التصنيف الناتج عن طريقة ندلسكاى

$[P(FN) = 0.08]$ ، أقل من احتمال الخطأ السالب لقرار التصنيف الناتج عن طريقة أنجوف

$[P(FN) = 0.47]$ ، و بالتالى فإن طريقة أنجوف ينتج عنها عدد أكبر من غير المتقنين غير الحقيقيين ، و هم الأفراد الذين درسوا البرنامج التعليمى(مجموعة متعلمة) ، و بالرغم من ذلك يحصلون على درجة دون درجة القطع .

و يفسر الباحث هذه النتيجة بأنها ترجع إلى انخفاض درجة القطع الملحوظ فى طريقة ندلسكاى (٤٧) ، مقارنة بطريقة أنجوف (٦٩) ، و الذى يعبر عنه نتيجة الفرض الأول ، هذا الانخفاض فى درجة القطع ،

جعل احتمالية حصول الطالب فى المجموعة المتعلمة (أى درس البرنامج التعليمى) على درجة أقل من درجة القطع صغير، لأن درجة القطع فى الأصل منخفضة ، وهذا ما جعل الخطأ السالب فى طريقة ندلسكاى منخفض (٠,٠٨) ، و العكس نجده فى طريقة أنجوف ، فنجد أن ارتفاع درجة القطع فى هذه الطريقة جعل احتمالية حصول الطالب فى المجموعة المتعلمة (أى درس البرنامج التعليمى) على درجة أقل من درجة القطع كبير، لأن درجة القطع مرتفعة و هناك صعوبة فى اجتيازها ، و هذا ما جعل عدد كبير من الطلاب المتعلمين، بالرغم من مرورهم بالبرنامج التعليمى يفشلون فى تعدى درجة القطع ، و الذى أدى بدوره إلى ارتفاع الخطأ السالب فى طريقة أنجوف ، و فى هذا الصدد يشير *(Abbott , 2003)* إلى أن درجة القطع المحددة بواسطة طريقة أنجوف صعب اجتيازها لارتفاعها ، و لذلك ينصح بإمداد المحكمين ببيانات تجريبية عن أداء المفحوصين حتى تكون درجة القطع أكثر موضوعية.

الفرض الفرعى الثالث:

هل يختلف القرار الصحيح للتصنيف باختلاف الطريقة المستخدمة فى تحديد درجة القطع ''طريقة أنجوف – طريقة ندلسكاى '' ؟.

للتحقق من صحة هذا الفرض تم استخدام المعادلة $P(CD)=P(TM)+P(TN)$ ، حيث $P(CD)$ ترمز إلى نسبة القرارات الصحيحة، $P(TM)=TM/M+N$ ، و هى نسبة المتقنين الحقيقيين (TM) فى الحجم الكلى للعينة$(M+N)$، $P(TN)=TN/M+N$ ،و هى نسبة غير المتقنين الحقيقيين (TN)فى الحجم الكلى للعينة$(M+N)$ ،و نتيجة الفرض الفرعى الثالث موضحة فى الجدول التالى:

جدول (١٤)

نتيجة الفرض الفرعى الثالث

البيان / الطريقة	عدد المتقنين الحقيقيين (TM)	نسبة المتقنين الحقيقيين P(TM)	عدد غير المتقنين الحقيقيين (TN)	نسبة غير المتقنين الحقيقيين P(TN)	نسبة احتمال القرار الصحيح P(CD)
ندلسكاى	٧٩	٠,٤٤	٨٥	٠,٤٨	٠,٩٢
أنجوف	١٠	٠,٠٦	٨٥	٠,٤٨	٠,٥٤

يتضح من الجدول السابق أن احتمال صحة قرار التصنيف الناتج عن طريقة ندلسكاى[$P(CD) = 0.92$] ، أكبر من احتمال صحة قرار التصنيف الناتج عن طريقة أنجوف [$P(CD) = 0.54$] ، و هذا يعنى أن درجة القطع الناتجة عن طريقة ندلسكاى تصنف الطلاب تصنيفاً صحيحاً مقارنة بطريقة أنجوف ، و يفسر الباحث هذه النتيجة بأنها مرتبطة بنتيجة الفرض الأول و التى تشير إلى انخفاض درجة القطع الملحوظ فى طريقة ندلسكاى (٤٧) ، مقارنة بطريقة أنجوف (٦٩)، هذا الانخفاض الذى أدى إلى انخفاض حدوث الخطأ السالب فى طريقة ندلسكاى ، و الذى تم توضيحه فى نتيجة الفرض الفرعى الثانى يقابله منطقياً صحة فى القرار لطريقة ندلسكاى ، لأن أخطاء التصنيف تتناسب عكسياً مع القرارات الصحيحة للتصنيف ، و فى المقابل نجد العكس فى طريقة أنجوف ،فارتفاع درجة القطع المحدد بواسطة هذه الطريقة ، يقابله ارتفاع حدوث الخطأ السالب و الذى يقابله منطقياً انخفاض صحة قرار التصنيف المبنى على درجة القطع المحددة بواسطة طريقة أنجوف .

الفرض الفرعى الرابع:

يخـــتلف معامل صدق قرار التصنيف باختلاف الطريقة المستخدمة فى تحديد درجة القطع ''طريقة أنجوف – طريقة ندلسكاى '' ؟.

للتحقق من صحة هذا الفرض تم استخدام معامل الارتباط الرباعى بين المتغيرين الثنائيين التصنيف المحك(١، ٠) ، و التصنيف المنبئ (١، ٠)، باستخدام القانون التالي:

$\phi = P(TM) - BR(SR) / [BR\ (1-BR)\ SR\ (1-SR)]^{1/2}$ ، و الذى سبق شرحه، و فيما يلى جدول يبين إجابة هذا السؤال:

جدول(١٥)

نتيجة الفرض الفرعى الرابع

البيان / الطريقة	نسبة المتقنين الحقيقيين $P(TM)$	معدل الاصل BR	نسبة الانتقاء SR	معامل صدق قرار التصنيف ϕ
ندلسكاى	٠,٤٤	٠,٥٢	٠,٤٤	٠,٨٥
أنجوف	٠,٠٦	٠,٥٢	٠,٠٦	٠,٢٤

يلاحـــظ مـــن الجدول السابق أن معامل صدق قرار التصنيف الناتج عن طريقة ندلسكاى $[\ \phi = 0.85\ ]$ ، أعلى من معامل صدق قرار التصنيف الـــناتج عن طريقة أنجوف $[\ \phi = 0.24\ ]$.، و تعنى هذه النتيجة قدرة درجـــة القطع المحددة بطريقة ندلسكاى على التنبؤ بالمتقنين الحقيقيين *(TM)* ، و غير المتقنين الحقيقيين *(TN)* ، أى أن درجة القطع صادقة فى الفـــصل بـــين المتقنين و غير المتقنين ، مقارنة بأنجوف ، و هى نتيجة

منطقية و تتفق مع النتائج السابقة لها ، فانخفاض معدل الخطأ فى قرار التصنيف الناتج عن طريقة ندلسكاى ، و الذى أدى إلى صحة القرار ، أدى ذلك إلى وجود صدق لدرجة القطع المبنى عليها هذا القرار، و على العكس نجد أن ارتفاع معدل الخطأ فى طريقة أنجوف ، و الذى أدى إلى انخفاض نسبة القرار الصحيح للتصنيف ، أدى إلى انخفاض معدل الصدق لدرجة القطع الناتج عن هذه الطريقة.

الفرض الفرعى الخامس:

تختلف دالة الضرر الناتجة عن قرار التصنيف باختلاف الطريقة المستخدمة فى تحديد درجة القطع''طريقة أنجوف– طريقة ندلسكاى'' ؟. للتحقق من صحة هذا الفرض تم استخدام المعادلة $\delta_K = P(FN)\,(D_1) + P(FM)\,(D_2)$ ، و لقد تم شرح هذه المعادلة ، و التى يتم فيها وضع $D_1 = -1$ ، و $D_2 = -2$ ، لأن الخطورة الناتجة عن الخطأ الموجب أكبر من الخطورة الناتجة عن الخطأ السالب ، و بذلك تصبح المعادلة $\delta_K = -P(FN) - 2\,P(FM)$ ، حيث $p(FN)$ هى نسبة احتمالية الأخطاء السالبة ، $p(FM)$ هى نسبة احتمال الأخطاء الموجبة ، و فيما يلى الجدول الذى يبين نتيجة الفرعى الخامس

جدول(١٦)

نتيجة الفرض الفرعي الخامس

البيان / الطريقة	نسبة احتمالية الأخطاء السالبة $P(FN)$	نسبة احتمالية الأخطاء الموجبة $P(FM)$	دالة الضرر δ_K
ندلسكاى	٠,٠٨	صفر	-٠,٠٨
أنجوف	٠,٤٧	صفر	- ٠,٤٧

يلاحظ من الجدول السابق أن دالة الضرر لقرار التصنيف الناتج عن طريقة ندلسكاى[δ_K = -0.08] ، أقل من دالة الضرر لقرار التصنيف الناتج عن طريقة أنجوف [δ_K = -0.47] .، ويفسر الباحث هذه النتيجة بأنها منطقية و تتفق مع النتائج الفرعية السابقة ، حيث أنه مع تثبيت الخطأ الموجب لقرار التصنيف و المعدوم فى الطريقتين ، نجد أن انخفاض الخطأ السالب للتصنيف و الذى يميز طريقة ندلسكاى يقابله انخفاض فى معدلات الأضرار لقرار التصنيف إذا استخدمنا درجة القطع المحددة بواسطة هذه الطريقة ، و فى المقابل نجد أن ارتفاع الخطأ السالب للتصنيف و الذى تتسم به طريقة أنجوف يقابله ارتفاع فى معدلات الأضرار لقرار التصنيف إذا استخدمنا درجة القطع المحددة بطريقة أنجوف. كما أن ارتفاع معامل الصدق(٠,٨٥) لقرار التصنيف الناتج عن طريقة ندلسكاى ، يقابله انخفاض فى معدلات الأضرار ، و العكس يحدث فى طريقة أنجوف ، حيث أن انخفاض معامل صدق القرار(٠,٢٤) طبيعى أن يقابله ارتفاع فى معدلات الأضرار الناتجة عن قرار التصنيف المبنى فى ضوء درجة القطع المحددة بواسطة طريقة أنجوف .

الفرض الفرعى السادس:

تختلف دالة المنفعة الناتجة عن قرار التصنيف باختلاف الطريقة المستخدمة فى تحديد درجة القطع''طريقة أنجوف-طريقة ندلسكاى '' ؟. للتحقق من صحة هذا الفرض تم استخدام المعادلة $\upsilon_K = P(TM)\ (U_1) + P(TN)\ (U_2)$، و لقد تم شرح هذه المعادلة ، و التى يتم فيها وضع $U_2 = 1$ ، و $U_1 = 2$ ، لأن المنفعة الناتجة عن التحديد الدقيق للمتقنين $P(TM)$ أعلى من المنفعة الناتجة عن التحديد الدقيق لغير المتقنين $P(TN)$، و بذلك تصبح المعادلة $\upsilon_K = +2\ P(TM) + P(TN)$ ، و فيما يلى الجدول الذى يبين نتيجة الفرض الفرعى السادس

جدول(١٧)

نتيجة الفرض الفرعى السادس

البيان / الطريقة	نسبة المتقنين الحقيقيين $P(TM)$	نسبة غير المتقنين الحقيقيين $P(TN)$	دالة المنفعة υ_K
ندلسكاى	٠,٤٤	٠,٤٨	١,٣٦
أنجوف	٠,٠٦	٠,٤٨	٠,٦٠

يلاحظ من الجدول السابق أن دالة المنفعة لقرار التصنيف الناتج عن طريقة ندلسكاى [υ_K= 1.36]، أكبر من دالة المنفعة لقرار التصنيف الناتج عن طريقة أنجوف [υ_K= 0.60] و ذلك راجع إلى انخفاض درجة القطع ، و الذى قابله انخفاض فى معدلات حدوث الخطأ السالب و ارتفاع نسبة القرار الصحيح ، و ارتفاع معامل الصدق ،كل هذه المؤشرات الايجابية و الخاصة بطريقة ندلسكاى ، تمثل منفعة ناتجة من قرار التصنيف المطبق، و العكس نجده فى طريقة أنجوف ، حيث أن الارتفاع الملحوظ فى درجة القطع ، و الذى قابله ارتفاع الخطأ السالب للقرار ، و انخفاض نسبة القرار الصحيح ، و انخفاض معامل الصدق ، كلها مؤشرات سلبية لقرار التصنيف ينتج عن انخفاض فى معدلات المنفعة المتوقعة فى ضوء طريقة أنجوف .

نتيجة الفرض الفرعى السابع:

تختلف دالة المنفعة العظمى الناتجة عن قرار التصنيف باختلاف الطريقة المستخدمة فى تحديد درجة القطع''طريقة أنجوف–طريقة ندلسكاى '' ؟.

للتحقق من صحة هذا الفرض تم استخدام المعادلة $\tau = \sum^{m+n}_{k=1} (\upsilon_K + \delta_K$ ، حيث υ_K تمثل دالة المنفعة ، δ_K تمثل دالة الضرر ، و فيما يلى الجدول الذى يبين نتيجة الفرض الفرعى السابع

جدول(١٨)

نتيجة الفرض الفرعى السابع من الفرض الثانى

البيان / الطريقة	دالة المنفعة υ_K	دالة الضرر δ_K	دالة المنفعة العظمى τ
ندلسكاى	١,٣٦	-٠,٠٨	٢٢٧,٨٤
أنجوف	٠,٦٠	-٠,٤٧	٢٣,١٤

يلاحظ من الجدول السابق أن دالة المنفعة العظمى لقرار التصنيف الناتج عن طريقة ندلسكاى [τ = 227,84]، أكبر من دالة المنفعة العظمى لقرار التصنيف الناتج عن طريقة أنجوف [τ = 23,14]، و يفسر الباحث هذه النتيجة بأنها منطقية و أنها ترجمة للنتائج المعروضة سابقاً و التى تم التوصل إليها فى نتيجة الفرض الأول ، و النتائج الفرعية السابقة الخاصة بالفرض الثانى ، لأن قرار التصنيف المبنى على درجة القطع المحددة بواسطة ندلسكاى و الذى يحظى بأقل قدر من نسبة الأخطاء السالبة ، وبالتبعية أقل قدر من الأضرار ، و أكبر قدر من نسبة القرارات الصحيحة ، و بالتبعية ارتفاع معامل صدق القرار ، و ارتفاع المنفعة الناتجة عن القرار ، من الطبيعى هذه المؤشرات الايجابية ينتج عنها ، أقصى قدر من المنفعة (المنفعة العظمى)، و العكس يحدث فى طريقة أنجوف ، فقرار التصنيف المبنى على درجة القطع المحددة بواسطة أنجوف و الذى يحظى بأكبر قدر من نسبة الأخطاء السالبة ، وبالتبعية أكبر قدر من الأضرار ، و أقل قدر من نسبة القرارات

الصحيحة ، و بالتبعية انخفاض معامل صدق القرار ، و انخفاض المنفعة الناتجة عن القرار ، من الطبيعى هذه المؤشرات السلبية ينتج عنها ، أقل قدر من المنفعة (المنفعة العظمى) ،و فيما جدول يبين ملخص لنتيجة الفرض الثانى بفروضه الفرعية السبع :-

جدول(١٩)

نتائج الفرض الثانى

البيان / الطريقة	القرار الصحيح	الخطأ الموجب	الخطأ السالب	معامل صدق القرار	دالة المنفعة v_k	دالة الضرر δ_k	دالة المنفعة العظمى τ
ندلسكاى	٠,٩٢	صفر	٠,٠٨	٠,٨٥	١,٣٦	- ٠,٠٨	٢٢٧,٨٤
أنجوف	٠,٥٤	صفر	٠,٤٧	٠,٢٤	٠,٦٠	- ٠,٤٧	٢٣,١٤

تشير نتيجة الفرض الثانى إلى أن خصائص قرار التصنيف الناتج عن طريقة ندلسكاى/(الخطأ السالب(٠,٠٨) ، القرار الصحيح(٠,٩٢) ، معامل صدق القرار(٠,٨٥)، دوال الضرر(- ٠,٠٨) و المنفعة(١,٣٦) و المنفعة العظمى(٢٢٧,٨٤)/ تتسم بالايجابية مقارنة بخصائص قرار التصنيف الناتج عن طريقة أنجوف /(الخطأ السالب(٠,٤٧) ، القرار الصحيح(٠,٥٤) ، معامل صدق القرار(٠,٢٤)، دوال الضرر(- ٠,٤٧) و المنفعة(٠,٦٠) و المنفعة العظمى(٢٣,١٤)/، حيث نجد أن قرار تصنيف الطالب فى ضوء درجة القطع المحددة بطريقة أنجوف يتسم بارتفاع الخطأ السالب و انخفاض نسبة القرارات الصحيحة ، و انخفاض معامل صدق القرار و ارتفاع معدلات الضرر الناتجة عن القرار و بالتبعية انخفاض معدلات المنفعة ، و المنفعة العظمى ، و ذلك مقارنة

بطـريقة ندلـسكاى ، و التـى نـتج عنها قرار تصنيفى يتسم بإيجابية مؤشراته.

و بالتالـى فـإن طريقة أنجوف ، بالرغم من انتشارها الواسع، إلا أنها يعتريها بعض العيوب التى تجعلنا لا نثق فى قرار التصنيف الناتج عنها، فيـشير العديد من الباحثين و المصادر البحثية أن هناك نقد كبير موجه لطـريقة أنجوف و هو أن تقديرات المحكمين للقيم الاحتمالية للمفردات مـن المحتمل جداً أن تكون غير دقيقة و أحياناً غير متسقة و متناقضة *Contradictory (Wang ,2003 ,231-232)*. ،كما أشارت الأكاديمية التربوية فـى التربية *(NAE)* إلى أن هناك خطأ جوهرى فى طريقة أنجوف و هو أن المحكمـين مطلـوب مـنهم إنجاز مهمة معرفية تكاد تكون مستحيلة *Nearly Impossible* ، هذه المهمة هى ''تقدير احتمالية إجابة الطالب ذى الحـد الأدنـى مـن الكفايـة و الذى هو فى الأصل شخصية افتراضية *Hypothesis* إجابة صحيحة على كل مفردة من مفردات الاختبار *(In : Berk ,1996 ,216)* ،حـيث أشـار *poggio* فى عام ١٩٨٤ إلى أن إحدى عيوب طـريقة أنجـوف هى أن معظم المحكمين يواجهون صعوبة فى تعريف الطالـب ذى الحد الأدنى من الكفاية *(In : Berk , 1986 , 148)* ، و فى هذا الـصدد أشـار *Sizmur* إلـى أن طـريقة أنجوف تعتمد على مدى فهم المحكمين معنى الحد الأدنى من الكفاية ، و كذلك فهمهم لكيفية استجابة الأفراد للمهام التى يتطلبها الاختبار ، و عدم فهم المحكمين لهذه الأمور يـؤدى إلـى انخفـاض مـصداقية الطريقة فى التوصل إلى درجة قطع موضوعية يمكن الوثوق بها*(sizmur , 1997 ,3)* .

و إذا كانـت طـريقة ندلسكاى تتطلب أيضاً من المحكمين تصور و تخـيل أداء الطالـب ذى الحد الأدنى من الكفاية ، فان حجم الخطأ الناتج عن عدم قدرة المحكم فى طريقة أنجوف على تخيل هذا الطالب أكبر من حجـم الخطـأ الناتج فى طريقة ندلسكاى ، فهذه العملية التخيلية و التى

تحـــدث فى الطريقتين معرضة للخطأ و بالتالى ينجم عنها خطأ فى قرار التـــصنيف ، فـــإذا تطـــرقنا لطريقة أنجوف نجد هناك قرار وحيد يتخذه المحكـــم علـــى تقدير كل مفردة ، لأن المحكم يتعامل مع البديل الصحيح للإجابـــة، وهذا القرار معرض أن يكون خاطئ و خطأ القرار الحادث فى طريقة أنجوف لا يمكن تعديله ، و فى المقابل نجد فى طريقة ندلسكاى أن المحكـــم لا يـــتخذ قرار وحيد على تقدير كل مفردة و إنما عدة قرارات *Multi Decisions Making* ، حـــيث أن المحكم فى طريقة ندلسكاى يتعامل مـــع مشتتات المفردة ، و تمثل استجابته لكل مشتت قرار(و بالتالى فإن المحكـــم فـــى طريقة ندلسكاى لا يصدر قرار وحيد عن المفردة ،كما فى طـــريقة أنجـــوف ، و إنما عدد من القرارات يساوى عدد المشتتات)، و بـــذلك فان الخطأ الناتج عن التقدير النهائى للمفردة هو معدل (متوسط) القرارات المعرضة للخطأ ، وهذا التعديل يخفف من الأخطاء الناتجة عن القـــرار، فعلـــى سبيل المثال ، إذا أصدر المحكم تقدير خاطئ فى طريقة أنجـــوف بأن نسبة إجابة الطالب ذى الحد الأدنى من الكفاية على (مفردة مـــا رباعـــية البدائل ، و البديل الصحيح هو ' جـــ ') هى ١٠٠% ، و بالتالـــى فهـــناك خطأ فى القرار بنسبة ١٠٠% ،فإذا أصدر المحكم نفس الحكـــم علـــى المفردة و لكن بطريقة ندلسكاى ، نجد هناك فرصة كبيرة لتقلـــيل نسبة الخطأ لأن المحكم فى هذه الطريقة لا يصدر قرار وحيد كما فى طريقة أنجوف ، و لكن يصدر ثلاثة قرارات على المفردة و هى:

القرار الأول بأن الطالب ذا الحد الأدنى من الكفاية سيستبعد المشتت 'أ'.

القـــرار الثانـــى بـــأن الطالـــب ذا الحـــد الأدنـــى مـــن الكفاية سيستبعد المشتت 'ب'.

القـــرار الـــثالث بأن الطالب ذا الحد الأدنى من الكفاية سيستبعد المشتت 'د'.

و بالتالـــى تكـــون القيمة الاحتمالية للمفردة هى (١) أى إجابة صحيحة بنسبة ١٠٠% هذا التقدير النهائى على المفردة ليس قرار وحيد كما فى طـــريقة أنجـــوف ، و لكـــن عبارة عن ثلاثة قرارات،و هناك صعوبة أن تكـــون الـــثلاث قـــرارات خاطئة فى نفس الوقت ، بمعنى آخر أن مجرد إصدار قرار فقط صحيح من الثلاثة يعنى تخفيض نسبة الخطأ الناتج عن التقديـــر النهائـــى للمفـــردة ، و هذا الذى يميز درجة القطع الناتجة عن طريقة ندلسكاى(*Chang , 1996*) . و بذلك نجد أن احتمال تعرض طريقة ندلـــسكاى لأخطاء التقدير من جانب المحكم ، أقل من طريقة أنجوف ، و هـــذا الـــذى جعـــل طريقة ندلسكاى تتميز بقرار تصنيفى ذات مؤشرات إيجابـــية، فطريقة ندلسكاى تلقى استحسان عند عدد من الباحثين ، فعلى الأقـــل لـــم يتم توجيه لها نقد مثل الكم الكبير جداً من النقد الموجه إلى طـــريقة أنجـــوف ،فلقـــد توصـــلت دراسة(*Chang,1996*) إلى أن طريقة ندلسكاى تحقق أعلى ثبات داخلى بين المحكمين مقارنة بطريقة أنجوف ، و ذلك لأن المحكمين يستخدمون نفس المصدر من المعلومات فى تحكيم المفردة، حيث أنه فى هذه الطريقة يحدد المحكمون البدائل الخاطئة ، و التى يستطيع الأفراد ذوو الحد الأدنى من الكفاية استبعادها أثناء التفكير فى الإجابة ، و بالتالى فان العملية المعرفية للمحكمين تركز على دراسة مدى منطقية البدائل و ليس على أى شئ آخر.

كمـــا أظهـــر عـــدد من الدراسات النتائج السلبية التى تترتب عن استخدام طريقة أنجوف فى تحديد درجة القطع ، منها دراسة (*Fehrmann et al.,1991*) التـــى توصلت إلى أن طريقة أنجوف تتسم بانخفاض معدلات القـــرارات الـــصحيحة(٠,٥٤) ، و كذلك ارتفاع نسبة الأخطاء الموجبة(٠,٠٥) و الـــسالبة(٠,٤١) لقرار التصنيف الناتج عنها ، و هذا ما حدا بالباحث إلى تقديم استراتيجيتين تعتمدان على إمداد المحكمين فى طريقة أنجوف سواء بصورة فردية أو بصورة جماعية بأطر مرجعية – *Frame*

Referenced عــن الفرد ذى الحد الأدنى من الكفاية و تعريف دقيق له و أداء حقيقى عن أدائه و الذى أدى إلى ارتفاع معدلات القرارات الصحيحة ، و انخفــاض نسبة الأخطاء السالبة ،كما توصلت دراسة*(Shen , 2001)* إلى أن معدل الفشل أو الإخفاق *Failure Rate* الناتج عن استخدام طريقة أنجــوف فى تحديد درجة القطع يساوى ٢١% ، و هو أعلى من المعدل الموجود فى طريقة خريطة المفردة ، كما توصلت دراسة *(Wang , 2003)* إلى انخفاض معدلات الثبات الداخلى بين المحكمين فى طريقة أنجوف ، و ترتبط بهذه النتيجة دراسة *(Chang , 1996)* التى توصلت إلى أن طريقة ندلــسكاى تحظــى بأكبــر قدر من الثبات الداخلى بين المحكمين مقارنة بطريقة أنجوف ، مما يعطى مصداقية لقرار التصنيف الناتج عن طريقة ندلسكاى .

و لــذلك أشــار *(Shen , 2001)* إلــى أنه ينبغى إمداد المحكمين فى طريقة أنجوف ببيانات حقيقية عن المفحوصين حتى تكون تقديراتهم أكثر واقعية ، و هذا الإجراء يجعل درجة القطع المحددة بواسطة هذه الطريقة أقل حدة.

توصيات البحث:

فى ضوء ما تم عرضه و ما تم التوصل إليه من نتائج يمكن تقديم التوصيات التالية:

١- تقويم أداء الطالب بواسطة اختبارات محكية المرجع تعكس مدى من الأهداف التربوية يمكن قياسها ، و يتم قياس أداء الطالب فى ضوء مدى تحقيقه لهذه الأهداف.

٢- تم التوصل فى البحث الحالى إلى أفضلية طريقة ندلسكاى فى تحديد درجة القطع ، لذلك يوصى الباحث باتباع هذه الطريقة ، و لكن مع الحرص على تقديم مزيد من التدريب للمحكمين .

٣-يوصى الباحث باستخدام درجات قطع صادقة تحظى بأكبر قدر من

المؤشرات الإيجابية لقرار التصنيف عند اختيار الطلاب فى الجامعات المختلفة ، و خاصة فى الكليات التى تحتاج إلى قدرات خاصة مثل كلية التربية النوعية ، و كلية التربية الرياضية .

٤- يوصى الباحث باستخدام درجات قطع صادقة تحظى بأكبر قدر من المؤشرات الايجابية لقرار التصنيف عند قبول الأفراد للالتحاق بالمهن التى لها أهمية خاصة مثل الطب و التدريس.

٥- على المسئولين و الجهات المعنية توجيه مزيد من الاهتمام إلى موضوع درجة القطع ، و التحقق من المنافع و الأضرار التى تقابل هذه الدرجة و التى ينجم عنها اتخاذ قرارات تربوية خاصة بالانتقاء أو التسكين أو الانتقال لمرحلة تعليمية تالية .

مراجع البحث

١-أنور محمد الشرقاوى (١٩٩٦) :''الاختبارات المرجعة إلى محك ، و سائل جديدة فى القياس النفسى و التربوى'' ، فى : أنـــور محمـــد الشرقاوى ، سليمان الخضرى الشيخ ، أمينة محمد كاظم ، نادية محمد عبد السلام(محررون): اتجاهات معاصرة فى القياس و التقويم النفسى و التربوى ، القاهرة ، مكتبة الأنجلو المصرية ، ص ص ١٩ – ٣٩

٢-حامد عبد السلام زهران(١٩٧٧) : اختبار ذكاء الشباب اللفظى ،القاهرة ، عالم الكتب.

٣- رجـــاء محمـــود أبـــو علام (٢٠٠٤):مناهج البحث فى العلوم النفـــسية و التـــربوية، القاهـــرة ، ط ٤ ، دار النـــشر للجامعات.

٤-صلاح الدين محمود علام (١٩٩٠) : ''استخدام نموذج راش فى بـــناء مقياس هدفى المرجع للمعارف الأساسية فى إعداد خطـــة الـــبحوث النفسية و التربوية ، دراسة سيكومترية تجريبية'' ، مجلة كلية التربية ، جامعة الأزهر ، ع ١٧ ، ص ص ١٢٩ – ١٨٥ .

٥-صلاح الدين محمود علام(١٩٩١) : ''دراسة مقارنة لبعض طرق تحديد مستويات الأداء فى اختبار مرجعى المحك ، المجلة المصرية للدراسات النفسية ، ع١ ، ص ص ٧٧-٩٦.

٦-صلاح الدين محمود علام(١٩٩٥) : الاختبارات التشخيصية مرجعية المحك فى المجالات التربوية و النفسية و التدريبية،القاهرة ، دار الفكر العربى .

٧- صـــلاح الـــدين محمـــود علام(٢٠٠١) :دليل الاختبار التشخيصى مرجعـــى المحـــك للمهارات الأساسية فى بناء الاختبارات التحصيلية ، القاهرة ، مكتبة الأنجلو المصرية .

٨-صـــلاح الدين محمود علام(٢٠٠٤) :التقويم التربوى البديل، أسسه النظرية و المنهجية و تطبيقاته الميدانية ، القاهرة ، دار الفكر العربى .

٩-صـــلاح الـــدين محمـــود عـــلام (٢٠٠٥): الأســـاليب الإحصائية الاستدلالية فى تحليل بيانات البحوث النفسية و التربوية و الاجتماعية البارامترية و اللابارامترية ،القاهرة ، دار الفكر العربى .

١٠-محمد محمد شوكت (١٩٨٧): اختبار تونى للذكاء الصورة(ب) ، القاهرة ، مكتبة الأنجلو المصرية .

١١-نادية محمد عبد السلام (١٩٩١) : ''تعدد معطيات تحديد الدرجات الفاصلة فى القياس محكى المرجع، دراسة تجريبية '' ، حولية كلية البنات ، جامعة عين شمس ، ع١٦ ، الجزء الأول ، ص ص ١٠٢ - ١٢١.

١٢-نادية محمد عبد السلام(١٩٩٢) : ''ثلاثة محكات رئيسية لتحديد مستوى الإتقان فى الاختبارات محكية المرجع دراسة امبريقية ''، المجلة المصرية للدراسات النفسية ، ع١٣ ، ص ص ٨٨ - ١٠٥.

١٣-نادية محمد عبد السلام(١٩٩٦:أ):'' حول معايير التقويم ، نظرة جديدة ''، فى : أنور محمد الشرقاوى ، سليمان الخضرى الشيخ ، أمينة محمد كاظم ، نادية محمد عبد السلام(محررون): اتجاهات معاصرة فى القياس و التقويم انفسى و التربوى ، القاهرة ، مكتبة الأنجلو المصرية ، ص ص ٤١ - ٦٢.

١٤-نادية محمد عبد السلام(١٩٩٦:ب): '' بناء اختبار هدفى المرجع و جماعى المرجع ، دراسة مقارنة بين النموذجين''، فى : أنور محمد الشرقاوى ، سليمان الخضرى الشيخ ، أمينة محمد كاظم ، نادية محمد عبد السلام(محررون): اتجاهات معاصرة فى القياس و التقويم النفسى و التربوى ، القاهرة ، مكتبة الأنجلو المصرية ، ص ص ٨٦-١٢٠.

15-Abbott, M.(2003) : "Standard Setting For Complex Performance Assessments: A Critical Examination Of The Analytic Judgment Method", Paper Presented At The Annual Congress of The Canadian Society For The Study Of Education, Halifax, Nova Scotia , May.

16- Berk , R. (1995) : "Something Old , Something New , Something Borrowed , A Lot To Do " , Applied Measurement In Education , Vol.8 , No.1, Pp.99- 109.

17- Berk , R.(1996) : "Standard Setting : The Next Generation (Where Few Psychometricians Have Gone Before !) ", Applied Measurement In Education , Vol. 9 , No. 3 , Pp. 215-235.

18- Berk ,R.(1976):" Determination of Optimal Cutting Scores in Criterion - Referenced Measurement " , Journal Of Experimental Education,Vol.45 , No.1,Pp. 4-9.

19- Berk, R. (1986): " A Consumer's Guide to Setting Performance Standards on Criterion-Referenced Tests", Review Of Education Research, Vol.56, No.1 ,Pp. 137-172 .

20- Bontempo ,B. ;Marks ,C & Karabatsos ,G.(1998):" A Meta-Analytic Assessment Of Empirical Differences In Standard Setting Procedures ", Paper Presented at The Annul Meeting Of The American Educational Research Association, San Diego,CA, Aprlil

21-Brennan, R.& Lockwood, R(1980):" A Comparison of The Nedelsky and Angoff Cutting Score Procedures Using Generalizability Theory", Applied Psychological Measurement, Vol. 4 , No. 1,Pp. 219-240 .

22-Buckendahl , C. ; Smith , R. & Impara , J.(2002):"A Comparison of

Angoff and Bookmark Standard Setting Methods ", Journal of Educational Measurement, Vo.39, No.3, Pp. 253-263.

23- Chang, L. (1996) : "A Comparison Between The Nedelsky And Angoff Standard Setting Methods ", Paper Presented at The Annual Meeting of The National Council on Measurement in Education, New York, NY, 9-11 April .

24- Davidson, F. & Lynch, B. (1998) :" A Criterion Referenced Viewpoint on Standards Cut scores In Language Testing ", Paper Presented at The Annual Meeting of The American Association For Applied Linguistics, Seattle, WA, 14-17 March .

25-Davis, G.(1983): "Educational Psychology : Theory And Practice " ,California, Addison-Wesley Publishing Com. .

26-Demauro, G. & Powers, D.(1990): " Logical Consistency of The Angoff Method of Standard Setting", Paper Presented at The Annual Meeting of The National Council on Measurement in Education Boston, MA ,17-19 April.

27-Donnoe, W. & Amato, R. (1997) : "Supportive Data & Guidelines For Using The Angoff, Ebel And Nedelsky Cutoff Score Methods ", Paper Presented at The International Personnel Management Assessment Council (IPMAAC) Conference, Newport Beach, June .

28- Fehrmann ,M. ; Woehr,D. & Arthur,W.(1991) :" The Angoff Cutoff Score Method : The Impact of Frame of Referenced Rater Training" , Educational And Psychological Measurement, Vol. 51, No.1, Pp. 857-872 .

29- Ferdous, A. & Plake, B.(2003) : "The Use of Subsets of Test Questions in an Angoff Standard Setting Method ", Paper Presented at The Annual Meeting of The National Council on Measurement in Education, Chicago, IL ,22-24 April.

30- Goodwin, L.(1996) : "Focus On Quantitative Methods Determining Cut-Off Scores ", Research In Nursing & Health, Vol. 19 ,No.1, Pp. 249-256 .

31- Hurtz, G. & Hertz,N.(1999) : "How Many Raters Should Be Used For Establishing Cutoff Scores With The Angoff Method ? A Generalizability Theory Study ", Educational And Psychological Measurement, Vol. 59, No. 6, Pp. 885-897.

32-Impara , J. & Plake , B.(2000): "A Comparison of Cut Scores Using Multiple Standard Setting Methods ", Paper Presented At The Large Scale Assessment Conference. Snowbird, UT , June.

33- Kubiszyn,T. & Borich , G.(2000) : Educational Testing and Measurement , 6 Th Edition , New York , John Wiley & Sons.

34- Livingston, S. & Zieky, M. (1989): " A Comparative Study of Standard Setting Methods", Applied Measurement In Education, Vol.2,No.2 ,Pp. 121-141.

35- Plake , B. & Impara, J.(1996) : "Intrajudge Consistency Using The Angoff Standard Setting Methods ", Paper Presented at The Annual Meeting of The National Council on Measurement in Education , New York , NY , 9-11 April.

36- Plake,B. ; Impara ,J. & Irwin , P.(1999) : "Validation Of Angoff-Based Predictions of Item Performance" , Paper Presented at The Annual Meeting of The American Educational Research Association, Montreal, Canada

37- Reckase , M. (2000):" The ACT/NAGB Standard Setting Process : How Modified Does It Have to Be Before It Is No Longer A Modified Angoff Process", Paper Presented at The Annual Meeting of The American? Educational Research Association, New Orleans , LA , 24- 28 April.

38- Shen,L. (2001) : "A Comparison of Angoff and Rasch Model Based Item Map Methods in Standard Setting " , Paper Presented at The Annual Meeting of The American Educational Research Association, Seattle , WA , 10-14 April .

39- Sizmur , S. (1997) : "Look Back In Angoff : A Cautionary Tale " , Journal of The British Educational Research Association , Vol. 23 , No. 1 , Pp. 3-13.

40- Slavin , R.(1997): Educational Psychology : Theory and Practice , 5 Th Edition , Boston , Allyn & Bacon.

41- Vos, H.(2000): "A Bayesian Procedure in The Context of Sequential Mastery Testing" ,Psicológica , Vol.21 , No. 1 ,Pp. 191-211.

42- Wang , N.(2003) : "Use of The Rasch IRT Model In Standard Setting : An Item-Mapping Method" ,Journal of Educational Measurement, Vol.40 , No.3 , Pp231-253.

43- Woolfolk , A. (1995) : "Educational Psychology , 6th Edition , Boston , Allyn& Bacon.

ملخص البحث

قرار التصنيف الناتج عن طريقتي أنجوف و ندلسكاى في تحديد درجة القطع لاختبار محكي المرجع '' دراسة تجريبية''

يهدف البحث إلى المقارنة بين درجة القطع المحددة بطريقة ندلسكاى و درجة القطع المحددة بطريقة أنجوف ، كما يهدف أيضاً إلى مقارنة مؤشرات قرار تصنيف الطلاب (الخطأ الموجب و الخطأ السالب و قرار الإتقان و معامل الصدق و دالة لضرر و دالة المنفعة و المنفعة العظمى) بين الطريقتين، و لقد قام (١٢) محكم باستخدام الطريقتين (ندلسكاى و أنجوف) لتحديد درجة القطع على اختبار محكى المرجع فى التدريبات المعملية لعلم النفس ،كما تكونت عينة المفحوصين من مجموعتين ، المجموعة الأولى و هى المجموعة المتعلمة و تتكون من (٩٣) طالب من طلاب الفرقة الثالثة تربية عام الشعب العلمية بكلية التربية بقنا، أما المجموعة غير المتعلمة فتكونت من (٨٥) طالب من طلاب الفرقة الثالثة شعبة التربية لرياضية بنفس الكلية و لقد تم اتخاذ هذا الإجراء التجريبى للتعرف على مؤشرات قرار التصنيف المقابل لكل طريقة من الطريقتين ، و أشارت نتيجة التحليل اللابارامترى إلى أن درجة القطع المحددة بطريقة أنجوف أعلى (بصورة دالة إحصائياً عند مستوى ٠,٠١) من درجة القطع المحددة بطريقة ندلسكاى ، و هذه النتيجة أدت إلى ارتفاع المؤشرات الايجابية لقرار التصنيف الخاص بطريقة ندلسكاى [(الخطأ السالب(٠,٠٨) ، القرار الصحيح(٠,٩٢) ، معامل صدق القرار(٠,٨٥)، دوال الضرر(- ٠,٠٨) و المنفعة(١,٣٦) و المنفعة العظمى(٢٢٧,٨٤)] مقارنة بمؤشرات قرار التصنيف الناتج عن طريقة أنجوف [(الخطأ السالب(٠,٤٧) ، القرار الصحيح(٠,٥٤) ، معامل صدق القرار(٠,٢٤)، دوال الضرر(- ٠,٤٧) و المنفعة(٠,٦٠) و المنفعة العظمى(٢٣,١٤)]، و لقد تم تفسير هذه النتائج فى ضوء

الدراسات المرتبطة ، و الإطار الخلفية النظرية ، و فى ضوء هذه النتائج تـــم تقديم مجموعة من التوصيات التربوية ، و التى تؤكد على مزيد من الاهتمام بموضوع درجة القطع و الدقة فى تحديد هذه الدرجة و إمكانية اســـتخدامها لاتخاذ قرارات تربوية خاصة بالانتقاء و التسكين و الانتقال لمراحل تعليمية تالية(يتكون البحث من ١٩ جدول ، و ٤٦ مرجع).

Paper Abstract

Classification Decision Resulting From Angoff And Nedelsky Methods Which Are Used In Determining Criterion –Referenced Test' Cut-Off Score

Empirical Study

The aims of the study are to investigate the differences between the cut-off scores which are derived from two methods of standard setting (Angoff And Nedelsky) , and to investigate also the differences in characteristics of Classification decision(positive error , negative error, correct decisions , validity coefficient , disutility function, utility function,and maximal utility function) resulting from the two methods , twelve(12) judges use each method to set cut-off score on criterion-referenced test (87 four-choices items) relating psychological laboratory practices , after that ,two groups of students "the instructional group(93) students & the uninsructional group(85) students" , respond to the test to investigate the Classification decision resulting from the two methods(Nedelsky And Angoff), results indicated that , the cut-off score derived from Angoff method is significantly higher than the cut-off score derived from Nedelsky method, this result led to increasing the positive indicators of classification decision relating Nedelsky method (negative error"0.08", correct decisions"0.92" , validity coefficient"0.85" , disutility function"-0.08", utility function"1.36",and maximal utility function"228") comparing to Angoff method(negative error"0.47", correct decisions"0.54" , validity coefficient"0.24" , disutility function"-0.47", utility function"0.60",and maximal utility function"23") which mean that the cut-off score derived from Nedelsky methjod is more accurate except the positive error which is none(equal zero) in the two methods, results are discussed ,educational Recommendations are including the need for more attention given to determining cut-off scores for measures used in educational settings(contains 19 tables , and 43 references).