

البحث الأول

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

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

قرار التصنيف الناتج عن طريقتي أنجوف و ندلسكاي في تحديد درجة القطع لاختبار محكي المرجح " دراسة تجريبية "

إعداد

دكتور/حجاج غانم أحمد علي*

(مدرس علم النفس التربوي - كلية التربية بقنا)

مقدمة البحث:

يعتبر مجال القياس النفسي و التربوي من المجالات المهمة في علم النفس، لما 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 كانت الظواهر الطبيعية مثل الطول و الوزن و درجة الحرارة و الضغط يمكن قياسها بأدوات مقننة و ثابتة و بوحدات متدرجة على طول المقياس، فان الظواهر السلوكية أيضا يمكن قياسها بأدوات تعد لذلك يطلق عليها الاختبارات و المقاييس النفسية و التربوية.

و يعتمد تفسير الدرجات المستمدة من أدوات القياس النفسي و التربوي على أنظمة مرجعية مختلفة، و يقصد بالنظام المرجعي *Referenced System* الأساس الذي يبني 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 هناك نظامان مرجعيان أساسيان يعتمد عليهما تفسير الدرجات المستمدة من الاختبارات النفسية و التربوية ، و يستخدمان في صناعة قرارات تربوية ، النوع الأول يسمى الاختبار المعياري المرجح (*NRT*)

*E-mail(hagaggg@yahoo.com)

Norm-Referenced Test و هذا النوع من الاختبارات يمدنا بمعلومات عن الوضع النسبي للفرد بالنسبة لأقرانه الآخرين ، و ذلك بمقارنة أداء الفرد (درجته) بمعيار مستمد من جماعة معيارية مشابهة، و تتحدد أفضلية الفرد في ضوء ترتيبه النسبي بين أقرانه بغض النظر عن كفاءته، أما النوع الثاني من الاختبارات يسمى الاختبار المحكي المرجع *Criterion -Referenced 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 أداء الفرد (درجته) بمستوى معين يطلق عليه محك *Criterion* (Kubiszyn&Borich,2000,26).

ويختلف الاختبار المحكي المرجع عن الاختبار المعياري المرجع ، في أن هذا النوع من الاختبارات يعد اختبار تشخيصي و يدور حول مجموعة محددة تحديداً جيداً من الأهداف التي يتم اختيارها من جانب المختصين لكي تنطبق في موقف محدد، و المفردات في هذا الاختبار يتم اختيارها لكي تغطي أهدافاً تعليمية محددة ، غالباً (٣-٥) مفردات لكل هدف و لهذا السبب يطلق عليها أيضاً *Objective-Referenced Test (ORT)*، كما تختلف أيضاً عن الاختبارات المعيارية في أن قياس أداء الطالب ينسب إلى أهداف تعليمية محددة بدلاً من الاختبار ككل، و بالتالي يمكن التعرف على أي الأهداف التي حققها الطالب، و من ثم يمكن الاستفادة من نتائج الاختبار في توجيه التعليم المستقبلي، أو الأنشطة العلاجية (Slavin,1997,535).

و يهدف 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 ، و على 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

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*) .
و تعرف درجة 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 الاختبار، و يطلق عليها عدة مصطلحات مثل مستويات التمكن، درجات النجاح أو الاجتياز ، الحد الأدنى للكفاية ، مستويات المحك (صلاح الدين علام، ١٩٩١، ٧٧).

و هناك طرق عديدة لتحديد درجة القطع، و هي طرق إحصائية و لكنها تعتمد بدرجات متفاوتة على أحكام الخبراء *Judgments* في المجال الذي يقيسه الاختبار، و كذلك على الدرجات الفعلية للمفحوصين و من هذه الطرق طريقة ندلسكاي *Nedlesky* في عام ١٩٥٤، و طريقة أنجوف *Angoff* في عام ١٩٧١، و طريقة ايبيل *Ebel* في عام ١٩٦٢، و طريقة جيجر *Jaeger* في عام ١٩٨٢، و طريقة أنجوف المعدلة بواسطة بيرك *Berk* في عام ١٩٨٤، و طريقة المجموعات المحكمة، و طريقة المجموعات المتناقضة، طريقة المجموعات الحدية (*Bontempo et al. , 1998*; صلاح الدين علام، ١٩٩٥، ٢٣٣).

و سيتم إلقاء الضوء على طريقتي ندلسكاي و أنجوف لخضوعهما للدراسة في البحث الحالي.

صممت طريقة ندلسكاي لكي تطبق على الاختبارات ذات الاختيار من متعدد لتحديد درجة قطع الاختبار محكي المرجع، و تعتمد هذه الطريقة على الاسترشاد بجمع من الخبراء و المتخصصين *Judges* في المجال الذي يقيسه الاختبار لكي يقوموا بثلاث خطوات هي:

١- تحديد عدد المشتتات فى كل مفردة من مفردات الاختبار و التى سيستبدها 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 يساوى ٠,٣٣).

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 القطع لكل المحكمين لكى نحصل على درجة القطع للاختبار

(صلاح علام، ١٩٩٥، ٢٣٩؛ Wang,2003,231).

و من ثم فان هناك طرق عديدة لتحديد درجة القطع ، هذه الدرجة التي في ضوءها يتم تصنيف *Classification* الأفراد إلى فئتين متبتين *Masters* و هم من يحصلون على درجة القطع أو ما يزيد عنها، و غير متبتين *Nonmasters* و هم من يحصلون على درجة أقل من درجة القطع، و بالتالي فانه كلما كان هناك صدق في تحديد درجة القطع ، كلما كان هناك دقة *Accuracy* في قرار التصنيف.

و لكن يتأثر قرار التصنيف بنوعين من الخطأ يؤثران بشكل ملحوظ على دقته، النوع الاول يسمى بالخطأ الموجب *False Positive Error*، النوع الثاني يسمى بالخطأ السالب *False Negative Error* .(Berk,1995,100-101).

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

النوع الاول: التصنيف المحك و فيه يصنف الفرد على أساس البرنامج التعليمي الذي مر به، بحيث إذا مر بالبرنامج التعليمي يحصل على الدرجة (١)، و إذا لم يمر بهذا البرنامج يحصل على الدرجة (٠).
النوع الثاني: التصنيف المنبأ و فيه يصنف الفرد على أساس درجة القطع التي تم تحديدها، بحيث إذا حصل الفرد على درجة تساوى أو أعلى من درجة القطع يحصل على الدرجة (١)، و إذا حصل على درجة أقل من درجة القطع يحصل على الدرجة (٠).

فالخطأ الموجب هو خطأ في القرار المتعلق بالتصنيف *Misclassification* ، و يحدث عندما يحصل الطالب في المجموعة غير المتعلمة [أى يأخذ القيمة (٠) في التصنيف المحك] ، على درجة القطع أو ما يزيد عنه [أى يأخذ القيمة (١) في التصنيف المنبئ] ، و يصنف خطأ على أنه متقن، أما الخطأ السالب هو خطأ آخر في القرار المتعلق بالتصنيف *misclassification* ، و يحدث عندما يحصل الطالب في

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 يؤثر هذان النوعان من الخطأ فى صحة قرار التصنيف *Decision Accuracy* ، و كذلك معامل صدق القرار *Validity Coefficient* ، و دوال المنفعة *Utility Function* و الضرر *Disutility Function* و المنفعة العظمى *Maximal Utility Function* الناتجة عن قرار التصنيف ، و يعرف صحة القرار بأنه هو نسبة القرارات الصحيحة *Probability Correct Decisions P(CD)* التي تم اتخاذها فى التصنيف المنبئ(المبنى على درجة القطع)، أما معامل الصدق فهو عبارة عن معامل الارتباط الرباعى (معامل ارتباط فاى) بين المتغير الثنائى (التصنيف المحك) (١ ، ٠) و المتغير الثنائى (التصنيف المنبأ) (١ ، ٠) ، و فى ضوء القرارات الصحيحة يتم حساب دالة المنفعة ، كما أنه فى ضوء الأخطاء الموجبة و السالبة يتم حساب دالة الضرر، كما أن دالة المنفعة العظمى تأتى من مزج التقديرات الناتجة عن دالتى الضرر و المنفعة خلال كل الطلاب الخاضعين لقرار التصنيف

. (Berk,1976,5-8)

و بذلك يتضح مما سبق أن قرار تصنيف الفرد على أنه متقن أو غير متقن له مؤشرات و هى (الخطأ الموجب و الخطأ السالب و صحة قرار التصنيف و معامل الصدق و دوال الضرر و المنفعة و المنفعة العظمى).

و الدراسة الحالية هى محاولة لمقارنة هذه المؤشرات بين طريقتى أنجوف و ندلسكاي 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 فإن التحديد الدقيق لدرجة القطع يحظى بأهمية كبيرة نظراً لتأثيره فى اتخاذ القرار.

فتحديد مستويات الأداء (درجات القطع) يحظى بأهمية كبيرة لما يتم فى ضوءه 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 فإن 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 فإن 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) التى 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 المنحنيين درجة القطع، و توصلت الدراسة إلى أن درجة القطع المثالية و التى

تقابل أفضل مؤشرات للتصنيف، كانت درجة القطع الناتجة عن الطريقة التجريبية، و التي كانت (٤,٥) و تم تقريبها إلى (٥) لاختبار درجته الكلية (٨) ، حيث كانت مؤشرات التصنيف هي (القرار الصحيح للتصنيف "٠,٧٤" و الخطأ الموجب "٠,١٢" و الخطأ السالب "٠,١٤" و معامل الصدق "٠,٤٨" و دالة المنفعة "١,١" و دالة الضرر "٠,٣٨" و دالة المنفعة العظمى "١٤٤")، و هي أفضل من مؤشرات التصنيف التي تقابل درجات القطع و التي تم وضعها بطريقة تصفية ، و دراسة (صلاح الدين علام، ١٩٩١) التي توصلت إلى اختلاف درجة القطع ، باختلاف الطريقة المستخدمة في تحديد هذه الدرجة ، فقد كانت درجة القطع لاختبار درجته الكلية (٦٠) ، باستخدام طريقة ندلسكاي (٤٠) ، و طريقة المجموعة الحديدية (٤٤) ، و طريقة أنجوف (٤٣) ، و طريقة المجموعات المتناقضة (٤١) ، و دراسة (نادية عبد السلام ، ١٩٩١) التي استخدمت نفس طريقة Berk التجريبية في تحديد درجة القطع ، و توصلت الدراسة إلى عدة نتائج منها أن مؤشرات التصنيف (الخطأ الموجب، الخطأ السالب ، القرار الصحيح للتصنيف، معامل صدق القرار، دالة المنفعة ، دالة الضرر) الناتجة عن الطريقة التجريبية، أفضل من مؤشرات التصنيف الناتجة عن وضع درجات قطع بصورة تصفية و ذلك بالنسبة للبعدين الاول و الثالث (مستويات القياس ، و مقاييس النزعة المركزية) من أبعاد الاختبار المحكي موضوع البحث، أما باقي الأبعاد فتم التوصل إلى عكس النتيجة ، و هذا إن دل على شيء فإتما يدل على القصور الذي يحدو الطريقة التجريبية المقترحة في تحديدها لدرجة القطع ، و دراسة (Chang, 1996) التي توصلت إلى اختلاف درجة القطع باختلاف الطريقة المطبقة فكانت ٧١% ، باستخدام طريقة أنجوف ، و ٥٧% باستخدام طريقة ندلسكاي ، كما توصلت الدراسة إلى أن طريقة ندلسكاي أعلى ثبات داخلي بين المحكمين مقارنة بطريقة أنجوف ، و

دراسة (Donnoe&Amato,1997) التي توصلت إلى اختلاف درجة القطع باختلاف الطريقة المطبقة فكانت ٥٠% استخدام طريقة أنجوف ، ٣٣% باستخدام طريقة ندلسكاي ، و ٦٠% باستخدام طريقة ايبيل، كما توصلت الدراسة إلى وجود فروق ذات دلالة إحصائية بين الطرق الثلاثة ، حيث كانت قيمة ف تساوي ٢٧ ، عند درجات حرية (٢،٢٥) (عدد المحكمين = ٢٨)، كما لا توجد ارتباطات بين طريقة ندلسكاي و كل من طريقتي أنجوف و ايبيل، بينما يوجد ارتباط بين طريقتي أنجوف و ايبيل ، و دراسة (Bontempo et al.,1998) التي توصلت باستخدام التحليل البعدي إلى أن طريقتي المجموعة الحديدية و جيجر تؤديان إلى درجات قطع أعلى من طريقة أنجوف ،بينما طريقة ندلسكاي تؤدى إلى درجة قطع أقل،كما أن عدد المحكمين يؤثر في تباين درجات القطع. و دراسة (Shen,2001) التي توصلت إلى أن نسبة الأخطاء الناجمة عن استخدام طريقة أنجوف في تحديد درجة القطع كانت ٢١%، و هي أكثر من نسبة الأخطاء لطريقة خريطة المفردة لراش و التي كانت ١٢,٥%، كما توصلت الدراسة إلى أن درجة القطع المحددة بواسطة طريقة أنجوف أكثر حدة من درجة القطع 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) التي توصلت إلى وجود اختلاف بسيط بين درجات القطع المحددة باستخدام طريقتي أنجوف و طريقة الترتيب التصاعدي لصعوبة المفردات (Bookmark) ، فكانت درجة القطع لطريقة أنجوف خلال محاولتين ٥١%، ٤٨%، و درجة القطع لطريقة الترتيب التصاعدي لصعوبة المفردات (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، و دراسة (Wang,2003)

و التى توصلت إلى 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 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 ضرر و المنفعة و المنفعة العظمى) كجوانب فى عملية المقارنة، و لعل هذا يمثل بعداً مهماً لمشكلة البحث ، أما البعد الآخر لمشكلة البحث فيعتمد على الأهمية التربوية التى تحظى بها طرق تحديد درجة القطع ، لأنه باختلاف الطريقة المستخدمة تختلف درجة 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 ما سوف يترتب على ذلك من نتائج، فعندما يصنف المتقن خطأ على أنه غير متقن (الخطأ السالب) سيترتب على ذلك تكليفه بمهارات جديدة كان فى غنى عنها فضلاً عن الأثر النفسى الذى يلحق به، فقد اعتبر غير متقن مع أنه فى الأصل متقن ، و فى حالة العكس أى تصنيف الفرد غير

المتقن على أنه متقن (الخطأ الموجب) ،سيترتب على ذلك أن يحتل هذا الفرد مكاناً أو مركزاً لا يستحقه ،فضلاً عن أنه سوف يغش في الأغلب و الأعم في المهارات المتدرجة التالية لأنها تفترض أنه أتقن المهارات التعليمية السابقة مع أنه لم يتقنها في واقع الأمر (نادية عبد السلام، ١٩٩٢ ، ٩٢).

يتضح من العرض السابق أن نوع الطريقة التي سيتم اختيارها لتحديد درجة القطع ،له أبعاد تربوية مهمة جداً نظراً لارتباطها بمدى صحة عملية التصنيف، وبالطبع فإن 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 فإن المقارنة بين طرق تحديد الدرجة الفاصلة تتطلب أخذ مؤشرات قرار التصنيف كجوانب مهمة في عملية المقارنة.

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

مسلمة البحث:

يستند البحث على مسلمة رئيسية و هي أن أي طالب يمر ببرنامج تعليمي فاعل (تابع للمجموعة المتعلمة) من المفترض أن يكون متقن لهذا المحتوى ، و العكس صحيح فان أي 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 في المجموعة المتعلمة المعرفة التي يعكسها الهدف التعليمي للبرنامج ، و ذلك بعكس المجموعة غير المتعلمة التي تكتسب معرفة أقل ، و لذلك يمكن وصف اكتساب أو عدم اكتساب المعرفة التي يعكسها الهدف التعليمي بمتغير ثنائي حقيقي *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 طريقتي ندلسكاي و أنجوف في درجة القطع المحددة؟.

٢- ما مدى الاختلاف بين طريقتي ندلسكاي و أنجوف في 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 الحالي إلى التعرف على :

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

٢- مدى الاختلاف في 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 (ندلسكاي - أنجوف) ؟ .

أهمية البحث:

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

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

٣- يتم فى هذا البحث المقارنة بين طريقتى ندلسكاي و أنجوف، و التعرف على الطريقة الأفضل يودى إلى دقة عملية التصنيف، مما يودى 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 كفاءته، 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 التى ينتمى إليها هذا الفرد أو بالنسبة إلى جماعته المعيارية (أنور الشرفاوى، 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 فإن 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 فإن 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 للإجازات 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 إتقان الطالب لمجموعة من المهارات الأساسية المتعلقة بمادة دراسية أو محتوى دراسى معين

(صلاح الدين علام، ٢٠٠١، ٣).

و يطلق على الاختبارات محكية المرجع اختبارات التفوق أو الإتقان *Mastery* أو اختبارات المقدره أو الكفافية 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 الاشطة العلاجية (Slavin,1997,535).

و يفيد نتائج الاختبار محكى المرجع المعلم فى قياس الإِتقان 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 لينغ 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 علام، ١٩٩٥، ٢٤)

و يهدف الاختبار المحكى المرجع إلى تحديد الفرد الذى اجتاز محك معين يتطلبه الاختبار من أجل اتخاذ قرار خاص بهذا الفرد سواء تسكينه فى مرحلة تعليمية أعلى ، أو التحاقه بمهنة معينة ، و على 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 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*):

تفيد الاختبارات محكية المرجع فى اتخاذ قرارات تعليمية تتعلق بدرجة إتقان الفرد لنطاق سلوكى من المعارف و المهارات المحددة و مستوى تمكنه من محتوى هذا النطاق، كما تستخدم فى تصنيف الأفراد بحسب مستويات تمكنهم أو إتقانهم 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 بين المتمكنين و غير المتمكنين ، أو بين مستويات مختلفة من التمكن على متصل درجات النطاق السلوكى الشامل للاختبار (صلاح الدين علام، ١٩٩٥، ٢٢٦).

و يطلق على درجة القطع مسميات عديدة منها مستوى أو مستويات الأداء *Performance Standard*، درجات التمكن *Mastery Scores*، درجات النجاح أو الاجتياز *Passing Scores*، الحد الأدنى للكفاية *Minimum Competency Level*، درجات المحك *Criterion Scores* (صلاح الدين علام، ١٩٩١، ٧٧)، و تعرف درجة القطع بأنها الدرجة التى تشير إلى الحد الأدنى اللازم لإتقان *Mastery* مادة معينة أو موضوع

معين (Slavin,1997,535)، كما تعرف بأنها الدرجة التي 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 ١٩٩٤ إلى أن درجة القطع يطلق عليها درجة الاجتياز *Passing Score* و هى نقطة على تدرج 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 ضونها 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 درجات القطع تصنع قرارات مهمة جداً في التربية لأنها تصنف الأفراد إلى ناجح أو سب، متمكن أو غير متمكن، مؤهل أو غير مؤهل، كما يمكن أن تصنفهم إلى مستويات متعددة، فعلى سبيل المثال يصنف التلاميذ بناءً على درجات القطع إلى مستويات مختلفة أ ب، ج، د، هـ، و في ضوء درجة القطع ينتقل التلميذ إلى مرحلة تعليمية تالية، أو يتم انتقاله في الكلية، أو يتم تسكينه في فصل دراسي معين، كما أن درجة القطع تصنع قرارات مهمة في سوق العمل، حيث في ضوء درجة القطع يتم تقرير ما إذا كان الطالب كفي لدرجة تؤهله للعمل كمدرس، أو مدير مدرسة، أو طبيب، أو أي نوع آخر من المهن (Bontempo et al., 1998).

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

هناك العديد من الطرق التي تستخدم في تحديد درجة القطع، و كل طريقة تم مهاجمتها و الدفاع عنها، من الوجهة النظرية و الإمبريقية، و لقد أثبتت الدراسات السابقة أن الطرق المختلفة لتحديد درجة القطع تؤدي إلى درجات قطع مختلفة، و من هذه الطرق طريقة ندلسكاي *Nedlesky* في عام ١٩٥٤، و طريقة أنجوف *Angoff* في عام ١٩٧١، و طريقة إيبيل *Ebel* في عام ١٩٦٢، و طريقة جيجر *Jaeger* في عام ١٩٨٢، و طريقة أنجوف المعدلة بواسطة بيرك *Berk* في عام ١٩٨٤، و طريقة المجموعات المحكّة، و طريقة المجموعات المتناقضة، و طريقة المجموعات الحدية (Bontempo et al., 1998).

و سيتم إلقاء الضوء على طريقتي ندلسكاي و أنجوف المستخدمتين في الدراسة الحالية:

أولاً: طريقة ندلسكاي *Nedlesky Method*:

سميت طريقة ندلسكاي بهذا الاسم لمؤسسها ندلسكاي و تم تأسيسها عام ١٩٥٤ و هي صممت لكي تطبق على الاختبارات ذات الاختيار من متعدد للحكم على درجة قطع الاختبار محكي المرجع، و تعتمد

هذه الطريقة على الاسترشاد بجمع من الخبراء و المتخصصين *Judges*

في المجال الذي يقيسه الاختبار لكي يقوموا بثلاث خطوات هي:

(١)- تحديد البدائل في كل مفردة من مفردات الاختبار و التي سيستبدها

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 يساوي ٠,٣٣

٣-تجميع مقلوبات مفردات الاختبار لكي 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 في البحث الحالي بأنه الطالب الذي 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) التي تؤهله لتطبيق

مفهوم التعلم عملياً داخل معمل علم النفس.

و تفترض هذه الطريقة أن الأفراد المختبرين ذوي الحد الأدنى من

الكفاية المقبولة يمتلكون مستويات مقبولة من المعرفة و المهارات و

القدرات و يستطيعون استبعاد البدائل التي تتضح أنها غير صحيحة أثناء

التفكير في اختيار البديل الصحيح،و مهمة المحكمين هي تحديد هذه

(Donnoe & Amato,1997).

البدائل

و أشار ندلسكاي إلى أنه لتلافى مشكلة خطأ القياس، يمكن الحصول على الانحراف المعياري لدرجات القطع الناتجة ، ثم يتم ضرب قيمة الانحراف المعياري في مقدار ثابت يتفق عليه المحكمون و هذا المقدار الثابت يعتمد على التوزيع الطبيعي لدرجات القطع للاختبار، ثم تضاف القيمة الأخيرة إلى درجة القطع التي تم تحديدها لنصل إلى درجة القطع بالطريقة المعدلة، و قيم الثابت التي يمكن أن يختار منها المحكمون هي (٢ أو ١) و هي تقابل النسبتين المئويتين (٨٤،٩٨) على الترتيب ، و هذه النسب تعنى نسبة فشل المجموعة الحدية (ذوى الحد الأدنى من الكفاية) فى الوصول لدرجة القطع

(صلاح الدين علام، ١٩٩٥، ٢٣٤)

غير أن الطريقة المعدلة تعثرها الصعوبة من ناحية فهم المحكمين لها بالإضافة إلى أن معظم الدراسات تستخدم الطريقة الأصلية لسهولة فهمها من جانب المحكمين .

و هناك العديد من الدراسات التي أجريت على طريقة ندلسكاي و معظم هذه الدراسات توصلت إلى انخفاض درجة القطع الناتجة عن هذه الطريقة مقارنة بالطرق الأخرى و من هذه الدراسات

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

ثانياً: طريقة أنجوف *Angoff Method*

سميت طريقة أنجوف بهذا الاسم لمؤسسها أنجوف *Angoff* فى عام ١٩٧١ و هى تعد من أشهر الطرق *Widely Used* التي تستخدم فى تحديد درجات القطع، و صممت لكى تطبق على الاختبارات ذات الاختيار من متعدد 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 لكي نحصل على درجة القطع من وجهة نظر المحكم ، بعد ذلك يتم أخذ المتوسط لدرجات القطع لكل المحكمين لكي نحصل على درجة القطع للاختبار

(Fehrmann et al.,1991,858;Plake & Impara,1996; Hurtz& Hertz, 1999 , 885; Plake et al.,1999;

Wang,2003,231; Ferdous & Plake,2003)

و هناك عدة صور معدلة لطريقة أنجوف ، ففي عام ١٩٧٦ أجرى أول تعديل لطريقة أنجوف ، و فيها لا يطلب من المحكمين تقدير النسبة المئوية على كل مفردة من تلقاء أنفسهم، و لكن باختيار نسبة من (٧) اختيارات (٥%، ٢٠%، ٤٠%، ٦٠%، ٧٥%، ٩٠%، ٩٥%)، و لكن يعاب على هذا التعديل أنها تقيد المحكمين بالنسب المعطاة، و لا تعطى لهم الحرية لاختيار نسب أخرى، و هناك صورة أخرى لطريقة أنجوف ظهرت عام ١٩٧٨ و هى أيضاً تعتمد على اختيار المحكمين لنسبة مئوية ، و لكن يكون الاختيار من اختياريين و ليس سبعة كما فى الطريقة السابقة، حيث يتم سؤال المحكمين عما إذا كان الفرد ذا الحد الأدنى من الكفاية قادر على إجابة المفردة بصورة صحيحة ، و بالتالى يختار المحكم (نعم ، أو لا)، و يتم تحديد درجة القطع فى ضوء نسبة الموافقة على (نعم)، ثم يتم تحرير هذه الدرجة من أخطاء القياس، و لكن يعاب على هذا التعديل أيضاً أنه يقيد المحكمين بنسبتين فقط (١٠٠% "نعم" ، صفره% "لا")، مما يزيد من أخطاء القياس إلى أقصى حد . (Berk,1986,148)

و هناك تعديل آخر لطريقة أنجوف عام ١٩٨٦ حيث يسترشد الخبراء فى عملية التحكيم ببعض البيانات المستمدة من التجريب الميدانى لمفردات الاختبار، و فى هذه الطريقة يطلب من كل محكم تقدير احتمال إجابة الأفراد من ذوى الحد الأدنى من الكفاية المطلوبة عن مفردات الاختبار ككل إجابة صحيحة أى أن تقديراتهم تتعلق بالاختبار ككل و ليس

كل مفردة كما سبق، و بعد الحصول على هذه التقديرات يقدم لكل منهم بيانات تجريبية عن مفردات الاختبار تتضمن متوسط صعوبة كل مفردة، ثم يطلب منهم تعديل تقديراتهم الاحتمالية في ضوء ذلك، و تكون درجة القطع هي متوسط تقديراتهم المعدلة

(صلاح الدين علام، ١٩٩٥، ٢٤٣) .

ولقد أوضح Ried في عام ١٩٨٤ أن هناك تعديل آخر لطريقة أنجوف ، و هو مرتبط أيضا بطريقة ندلسكاي ، و يعتمد على أخذ المتوسط *average* لدرجتى القطع الناتجتين عن طريقة أنجوف و ندلسكاي ، و تتميز هذه الطريقة بسهولة تنفيذها و فهمها (berk,1986,152) .

و جدير بالذكر أن طريقة أنجوف و صورها المعدلة *modified* بالرغم من النقد الموجه لها، إلا أنها حظيت باهتمام و تفضيل علماء القياس النفسى و الباحثين لسهولة فهم المحكمين للإجراءات التى تتطلبها هذه الطريقة (Goodwin,1996,252).

و لقد أجريت العديد من الدراسات على طريقة أنجوف سواء للتعرف على مدى دقتها فى تحديد درجة القطع ، أو مقارنتها بالطرق الأخرى و أغلب الدراسات أثبتت ارتفاع درجة القطع الناتجة عن طريقة أنجوف و ذلك مقارنة بالطرق الأخرى، كما أثبتت انخفاض معدل الثبات الداخلى بين المحكمين و من هذه الدراسات (Fehrmann et Shen,2001; Wang,2003 Chang,1996) al.,1991; ، و من الدراسات العربية دراسة (صلاح الدين علام، ١٩٩١) .

و لقد توصلت دراسة (Fehrmann et al.,1991) إلى أن إمداد المحكمين 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 لندلسكاي، و الطريقة الأصلية لأجوف، للمقارنة فى البحث الحالى.

٤- مؤشرات قرار التصنيف : *Classification Decision*

بعد تحديد درجة القطع للاختبار محكى 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 في 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 ينبغي معرفة أن حساب هذه المعاملات يعتمد في الأصل على اختيار مجموعتين محكيتين هما المجموعة المتعلمة (*Instructional Group(I)* و هي تتكون من الطلاب الذين تلقوا تعليم فاعل *Effective Instruction* على الأهداف المطلوب تحقيقها، و يمثلون الطلاب في هذه المجموعة الطلاب المتقنون (*M*) ، و المجموعة غير المتعلمة (*Uninstructional Group(U)* و هي تتكون من الطلاب الذين لم يتلقوا تعليم على الأهداف المطلوب تحقيقها، و يمثلون الطلاب في هذه المجموعة الطلاب غير المتقنين (*N*) ، و ينبغي أن تتساوى العينتان في الحجم ، و كل مجموعة ينبغي أن تتكون من (١٠٠) طالب ، لكي نحصل على احتمالات مستقرة و ثابتة، و لو كانت هناك صعوبة في 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، و ينحصر توزيع درجات الطلاب بين (صفر- ن) ، حيث (ن) عدد مفردات الاختبار و التي 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 نوعان من التصنيف هما:التصنيف المحك *Criterion Classification* ، فيه يصنف الفرد على أساس البرنامج التعليمي الذي مر به ، حيث يحصل كل طالب فى المجموعة المتعلمة على القيمة (١) ، و يحصل كل طالب فى المجموعة غير المتعلمة على القيمة(٠) ، و التصنيف المنبئ *Predicted Classification* ، و فيه يصنف الفرد على أساس درجة القطع التى تم تحديدها، بحيث إذا حصل الفرد على درجة تساوى أو أعلى من درجة القطع يحصل على الدرجة (١)، و إذا حصل على درجة أقل من درجة القطع يحصل على الدرجة(٠) ، و تنتج التصنيفات المحكية و المتنبأ بها للطلاب بناءً على درجاتهم فى الاختبار أربعة نواتج ممكنة:-

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

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

٣-غير المتقنين الحقيقيين(*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 السالب لسوء 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 غير المتعلمة (U)

المجموعة المتعلمة (I)

<p>متقنون حقيقيون (TM)</p> <p>متقنون غير حقيقيين (FM)</p> <p>(الخطأ الموجب)</p>	<p>متقنون حقيقيون (TM)</p>
<p>غير متقنون غير حقيقيين (TN)</p>	<p>غير متقنون غير حقيقيين (FN)</p> <p>الخطأ الأول (الخطأ السالب)</p>

المتقنون المتنبأ بهم
PM=TM+FM

غير المتقنون المتنبأ بهم
PN=FN+TN

التصنيف المنبئ
(درجة القطع)

غير المتقنين

$$N = FM + TN$$

المتقنون

$$M = TM + FN$$

(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 رأى يأخذ

القيمة (٠) في التصنيف المحك/، على درجة القطع أو ما يزيد عنه/أى يأخذ القيمة (١) في التصنيف المنبئ/، و يصنف خطأ على أنه متقن، و يمكن حساب نسبة احتمال الخطأ الموجب من القانون $P(FM) = FM/M+N$ ، حيث FM هي 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 تؤدي 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 يؤدي إلى تأثير سلبي 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 هي عدد القرارات الخاطئة من النوع السالب (Berk, 1976, 5).

و يمكن أن يرجع سبب الخطأ السالب إلى بعض العوامل التي يتعرض لها الطالب (المتعلم) ، و التي تؤثر على استجابته بالسلب مما يجعله يحصل على درجة أقل من درجة القطع ، و من هذه العوامل الإرهاق أو شرود الذهن أو التعب أو التعرض لمؤثرات سلبية قبل الامتحان.

و لقد أوضحت نادية عبد السلام أن لكل من الخطأين أثره على عملية اتخاذ القرار التعليمي، و لكن ضرر الخطأ السالب حيث يدرس محتوى الأهداف مرة أخرى للطالب في المجموعة المتعلمة بالرغم من سابق دراسته لها ، أقل من الخطأ الموجب و الذي فيه ينتقل الطالب إلى المرحلة التعليمية التالية بالرغم من عدم إتقانه للأهداف في الوحدة السابقة(نادية عبد السلام ، ١٩٩١ ، ١٠٨) .

فبالرغم من المرغوب فيه تقليل احتمالات كلا النوعين من الخطأ عند اختيار درجة القطع ، فان هناك بعض المواقف التي يكون فيها اقتراف نوع معين من الخطأ أكثر حدة من النتائج المترتبة عن الخطأ الآخر، فعلى سبيل المثال النتائج المترتبة على تكليف تلميذ بمواد تعليمية جديدة و هذا التلميذ في الأصل غير قادر على دراسة هذه المواد و التعامل معها(الخطأ الموجب)، هذا الخطأ أكثر حدة من النتائج المترتبة على إعادة تكليف مواد تعليمية و أنشطة علاجية لتلميذ درس هذه المواد و الأنشطة من قبل (الخطأ السالب) (Berk , 1986 , 139) .

مما سبق يتضح أن أساس حدوث كلا النوعين من الخطأ هو اختلاف القيمة التي يحصل عليها الطالب في التصنيف المحك عن القيمة التي يحصل عليها في التصنيف المنبئ((١ ، ٠) ؛ (٠ ، ١))، أما اتفاق القيمة في كلا التصنيفين المحك و المنبئ أي ((١ ، ١) ؛ (٠ ، ٠))، فهذا يعني أن قرار التصنيف صحيح ، و يمكن توضيح القرارات الناتجة عن تفاعل

المتغيرين الثنائيين التصنيف المحك (١، ٠)، و التصنيف المنبئ (١، ٠) ،
 ، في الجدول التالي:

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: $P(CD)$ Correct Decisions Probability

هو نسبة القرارات الصحيحة $P(CD)$ Correct Decisions Probability التي تم اتخاذها في التصنيف المنبئ (المبنى على درجة القطع) ، و يحسب من القانون $P(CD) = P(TM) + P(TN)$ حيث $P(TM) = TM/M+N$ ، و هي نسبة المتقين الحقيقيين (TM) في الحجم الكلي للعينة $(M+N)$ ، $P(TN) = TN/M+N$ و هي نسبة غير المتقين الحقيقيين (TN) في الحجم الكلي للعينة $(M+N)$ (Berk, 1986, 5-6).

كما أشار *Hambelton* و زملاؤه في عام ١٩٧٦ أن صحة القرار *Decision Accuracy* الخاص بالتصنيف يمكن حسابه من المعادلة:

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

حيث " *E* " هي خطأ التصنيف و يأخذ إحدى ثلاث قيم (٠) ، (١-) ، (٠) ، (١) ، و تتحدد قيمة *E* على حسب نوع الخطأ ، فإذا صنف الطالب تصنيفاً صحيحاً ، فإن $E = (٠)$ ، أما في حالة الخطأ الموجب ، فإن $E = (١)$ ، و في حالة الخطأ السالب فإن $E = (١-)$ ، أما *N* فهي العدد الكلي لأفراد المجموعتين المحكيتين

(في:نادية عبد السلام، ١٩٩٢ ، ٩٨ -٩٩).

مما سبق يتضح أن هناك طريقتين للتعرف على القرار الصحيح للتصنيف أحدهم بحساب نسبة القرارات الصحيحة مباشرة و هي طريقة *Berk* ،أما الطريقة الثانية فتأتى عن طريق طرح نسبة القرارات الخاطئة(الخطأ الموجب و الخطأ السالب) من الواحد الصحيح ، في ضوء المعادلة التي ذكرها *Hambelton* و زملاؤه.

(٤)-معامل صدق قرار التصنيف *Validity Coefficient of*

Classification Decision

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

$\phi = P(TM) - BR(SR) / [BR (1-BR) SR (1-SR)]^{1/2}$ ، حيث *BR* ترمز إلى *Base Rate* و تعنى معدل الأصل و هي تشير إلى نسبة عدد أفراد

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

المجموعة المتعلمة (M) إلى العدد الكلي لأفراد المجموعتين المحكيتين ($M+N$) ، حيث تأتي من القانون: $BR=P(FN)+P(TM) = M/M+N$ ، أما SR ترمز إلى *Selection Ratio* وتعني نسبة الانتقاء و هي تشير إلى نسبة عدد الأفراد المتقنين المتنبأ بهم (PM) إلى العدد الكلي لأفراد المجموعتين المحكيتين ($M+N$) ، حيث تأتي من القانون: $SR=P(TM)+P(FM) = PM/M+N$ ، و تقيد كل من معدل الأصل و نسبة الانتقاء مدى القيم لمعامل الصدق، فعندما $BR=SR$ تكون أقصى قيمة لمعامل الصدق و تصل إلى (1) ، و كلما بعدت النسبتين عن بعضهما، يتناقص معامل الصدق (Berk, 1986, 7) .

(5)- دالة الضرر: *Disutility Function*

إن خطأي التصنيف (الخطأ الموجب و الخطأ السالب) يتضمنان بعض الخسارة و الفاقد، هذه الخسارة تعد دالة للخطورة النسبية الناتجة عن نوعي الخطأ، و عندما يتم استخدام نسبة القرارات الصحيحة و معامل الصدق في اختيار درجة القطع ، يفترض أن هذين النوعين من الخطأ لهما نفس الخطورة، و لكن هذين النوعين من الخطأ لهما درجات متفاوتة من الخطورة، و من ثم فإن احتمالات الخطأ تحتاج أن يتم تعديلها لكي تعكس الخسارة الحقيقية المرتبطة بقرار التصنيف المتخذ، فالخسارة المقدرة في إعادة الوقت التعليمي و إحباط الطالب المتقن غير الحقيقي، و الذي يحاول تحقيق أهداف تعليمية أصعب من الأهداف السابقة (و التي لم يتقنها أصلاً) الخطأ الموجب) ، هذه الخسارة تكافئ مرتين الخسارة الناتجة عن تكرار الطالب غير المتقن غير الحقيقي لدراسة المادة التعليمية لأنه حصل على درجة دون درجة القطع بالرغم من دراسته للبرنامج التعليمي (الخطأ السالب)، و على ذلك فان تقدير الخسارة (الضرر) المتوقع و المرتبط بدرجة قطع معينة يعد أصلاً الضرر المتوقع لقرار التصنيف المتخذ، و تحسب دالة الضرر (5) لقرار

وحيد(خاص بدرجة قطع) من مجموع نواتج احتمالات الخطأ(الخطأ الموجب ، و الخطأ السالب) كما في المعادلة التالية: $\delta_x = P(FN) (D_1) + P(FM) (D_2)$ ، حيث D_1 ، و D_2 تمثل القيمتان الوزنيتان لتعديل حجم الخسارة الناتجة عن أخطاء قرارات التصنيف، و في ضوء ما سبق تأخذ D_1 القيمة

(-1) ، و تأخذ D_2 القيمة (-2) ، نظراً لأن $P(FN)$ ترمز إلى نسبة تكرار الأخطاء السالبة و التي فيها يحصل الطالب المتقن *Master* (مر ببرنامج تعليمي) على درجة دون درجة القطع ، و يصنف خطأ على أنه غير متقن و من ثم يعيد *Retain* دراسة المادة التعليمية ، و $P(FM)$ ترمز إلى نسبة تكرار الأخطاء الموجبة و التي يحصل فيها الطالب غير المتقن *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 في الضرر الناجم عنه، لذلك أخذ الخطأ الموجب القيمة الوزنية $(D_2 = -2)$ ، و أخذ الخطأ السالب القيمة الوزنية $(D_1 = -1)$ (Berk,1986,7) .

فالضرر الناتج عن سوء التصنيف هو في الواقع دالة لأخطاء التصنيف الموجبة و السالبة ، و هذه الدالة يتم تحديدها بصورة ذاتية (في ضوء اختيار القيم الوزنية)، و هي تعتمد على سلوك المخاطرة لصانع القرار و اتجاهه نحو نسبة العوامل المسببة للضرر. و أشار Vos إلى أن العوامل المسببة للضرر يمكن أن تنحصر في العواقب الاجتماعية و النفسية و الاقتصادية التي تصاحب القرار الخاطئ المتخذ (Vos,2000,197).

(٦)- دالة المنفعة : *Utility Function*

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

، و لكن الفارق هنا أن القيم الوزنية تأخذ قيمة موجبة لأنها دالة منفعة، و يرمز لدالة المنفعة لقرار تصنيفي بالرمز (v_k) ، أما القيمتين الوزنيتين في دالة المنفعة فيرمز لهما U_1 ، و U_2 ، و قيمتهما تأتي من أن الفائدة أو المكسب في الكفاءة و الفاعلية التعليمية الناتجة عن التصنيف الصحيح لفرد متقن (TM) ، يكافئ مرتين الفائدة الناتجة عن التصنيف الصحيح لفرد غير متقن (TN) ، و بالتالي فإن القيمة الوزنية المقابلة لنسبة الإتقان الحقيقي (TM) تساوي $2 + (U_1 = +2)$ ، كما أن القيمة الوزنية المقابلة لنسبة عدم الإتقان الحقيقي (TN) تساوي $1 + (U_2 = +1)$ ، و تكون الفائدة المتوقعة *Expected Utility* لقرار التصنيف K المبني على درجة قطع معينة يساوي مجموع نواتج القرارات الصحيحة معدلة بواسطة القيمتين الوزنيتين $(U_1 = +2 ; U_2 = +1)$ و دالة الفائدة موضحة في المعادلة التالية:

$$v_k = P(TM) (U_1) + P(TN) (U_2) \quad (Berk, 1986, 7)$$

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

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

الدراسات المرتبطة و فروض البحث الحالي :

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

محكى المرجع فى حساب رؤوس الأموال *Capitalization*، يتكون من (٨) مفردات ، و لا توجد عينة للمحكمين نظراً لأن الطريقة المستخدمة فى تحديد درجة القطع لا تعتمد على المحكمين و لكن تعتمد على الرسم البياتى، أما عينة المفحوصين فتكونت من مجموعتين : مجموعة المتعلمين و تكونت من (١٠٠) طالب، و مجموعة غير المتعلمين و تكونت أيضاً من (١٠٠) طالب، و تعتبر هذه الدراسة من الدراسات النادرة التى 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 تصفية ، و توصلت الدراسة إلى أن درجة القطع المثالية و التى تقابل أفضل مؤشرات للتصنيف، كانت درجة القطع المثالية و التى تقابل أفضل مؤشرات التجريبية، و التى كانت (٤,٥) و تم تقريبها إلى (٥) ، حيث كانت مؤشرات التصنيف هى (القرار الصحيح للتصنيف "٠,٧٤" و الخطأ الموجب "٠,١٢" و الخطأ السالب "٠,١٤" و معامل الصدق "٠,٤٨" و دالة المنفعة "١,١" و دالة الضرر "٠,٣٨" و دالة المنفعة العظمى "١٤٤")، و هى أفضل من مؤشرات التصنيف التى تقابل درجات القطع و التى تم وضعها بطريقة تصفية .

٢- دراسة (Brennan & Lockwood, 1980) التى هدفت إلى المقارنة بين طريقتى ندلسكاي و أنجوف ، و تركزت المقارنة على قيمة درجة القطع ، و كان الاختبار المستخدم يتكون من (١٢٦) مفردة من نوع الاختيار من متعدد (رباعية البدائل) فى مجال الصحة العامة، و تكونت عينة المحكمين من (٥) محكمين فى المجال الذى يقيسه الاختبار ، و توصلت الدراسة إلى عدة نتائج منها اختلاف درجة القطع باختلاف الطريقة المطبقة فكانت (٨٤) باستخدام طريقة أنجوف باتحراف معيارى قدره

(٤,٧) ، و (٧٠) باستخدام طريقة ندلسكاي بانحراف معياري قدره (٩,٠٣).

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

٤-دراسة (DeMauro & Powers, 1990) هدفت الدراسة إلى معرفة الاتساق المنطقي لطريقة أنجوف ، و تكونت عينة المحكّمين 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 المحكّمين و الأداء الفعلي للطلاب قدره (٠,٧١) ، كما يوجد ارتباط بين تقديرات المحكّمين على المفردات و الصعوبات الحقيقية

* تم أخذ متوسط درجات القطع المحددة في الدراسة على الأربعة مدارس في كل طريقة (انجوف - ندلسكاي) ، و كل اختبار (الرياضيات - القراءة).

للمفردات تراوح بين (٠,٢٥) إلى (٠,٥٦) بوسيط قدره (٠,٤٢)، و أوصى 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 إستراتيجيتين مقدمتين للمحكمين على ثبات و دقة القرارات الناتجة عن طريقة أنجوف ، و كانت عينة المحكمين عبارة عن (٢١) محكم من الخبراء الذين لهم دراية تامة بعلم النفس العام، و لقد تم تقسيم المحكمين إلى ثلاث مجموعات بالتساوى بحيث تقوم المجموعة الأولى بتقدير المفردات بدون استراتيجية مقدمة، بينما تقدم للمجموعتين الأخرين الاستراتيجيتان موضوع البحث، و كان الاختبار المستخدم، هو اختبار الكفاءة في 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 الأخطاء الموجبة(٠,٠٥) و السالبة(٠,٤١) لقرار التصنيف الناتج عنها ، و هذا ما حدا بالباحث إلى تقديم استراتيجيتين تعتمدان على إمداد المحكمين في طريقة أنجوف سواء بصورة فردية(الاستراتيجية الأولى) أو بصورة جماعية(الاستراتيجية الثانية) بأطر مرجعية *Frame - Referenced* عن الفرد ذى الحد الأدنى من الكفاية و تعريف دقيق له و أداء حقيقى عن أدائه و الذى أدى إلى ارتفاع معدلات القرارات الصحيحة إلى (٠,٧٤ ؛ ٠,٦٩) ، و انخفاض نسبة الأخطاء السالبة إلى (٠,١٤ ؛ ٠,٢٣) للاستراتيجيتين على الترتيب ، و لكن فى 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 هي طريقة أنجوف و طريقة ندلسكاي 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 ندلسكاي (٤٠)، و طريقة المجموعة الحدية (٤٤)، و طريقة أنجوف (٤٣)، و طريقة المجموعات المتناقضة (٤١).

٧-دراسة (نادية عبد السلام ، ١٩٩١) التى 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 المنحنيين 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 البياتى، 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 خطنا 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 تصفية ، و توصلت الدراسة إلى عدة نتائج منها أن مؤشرات التصنيف الناتجة عن الطريقة التجريبية، أفضل من مؤشرات التصنيف الناتجة عن وضع درجات قطع بصورة تصفية و ذلك بالنسبة للبعدين الأول 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 في محتوى الإحصاء الوصفي)، أما باقي الأبعاد فتم التوصل إلى عكس النتيجة ، و هذا إن دل على شئ فإنما يدل على القصور الذي يحدو الطريقة التجريبية موضوع البحث في تحديدها لدرجة القطع .

٨-دراسة (Chang,1996) التي هدفت إلى المقارنة بين طريقتي ندلسكاي و أنجوف ، و تركزت المقارنة على الثبات الداخلي بين المحكمين *Intrajudge Consistency* في كل طريقة، و أيضاً قيمة درجة القطع ، و تكونت عينة المحكمين من (٢٢) محكم ممن حصلوا على برنامج في الدراسات العليا بكلية التربية، و كان الاختبار المستخدم يتكون من (٩) مفردات من نوع الاختيار من متعدد (رباعية البدائل) مأخوذ من الاختبار النهائي لهؤلاء المحكمين في 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 ندلسكاي ، كما توصلت الدراسة إلى أن طريقة ندلسكاي أعلى ثبات داخلي بين المحكمين مقارنة بطريقة أنجوف.

٩- دراسة (Donnoe&Amato,1997) التى هدفت إلى المقارنة بين طرق أنجوف و ندلسكاي و ايبيل ، و تركزت المقارنة على درجات القطع باستخدام 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 المهني ، و توصلت الدراسة إلى عدة نتائج منها اختلاف درجة القطع باختلاف الطريقة المطبقة فكانت ٥٠% استخدام طريقة أنجوف ، ٣٣% باستخدام طريقة ندلسكاي ، و ٦٠% باستخدام طريقة ايبيل، كما توصلت الدراسة إلى وجود فروق ذات دلالة احصائية بين الطرق الثلاثة ، حيث كانت قيمة ف تساوى ٢٧ ، عند درجات حرية (٢،٢٥) (عدد المحكمين=٢٨) ، كما لا توجد ارتباطات بين طريقة ندلسكاي و كل من طريقتى أنجوف و ايبيل، بينما يوجد ارتباط بين طريقتى أنجوف و ايبيل.

١٠- دراسة (Bontempo et al.,1998) أجرت هذه الدراسة تحليل بعدى *Meta-Analytic* لطرق تحديد درجة القطع ، باستخدام حجم التأثير، حيث تمت مقارنة طريقة أنجوف المعدلة بأربع طرق أخرى هي (طريقة المجموعة الحديدية، طريقة المجموعات المتناقضة، طريقة ايبيل، طريقة جيجر)، و توصلت الدراسة إلى عدة نتائج منها أن طريقتى المجموعة الحديدية و جيجر يؤديان إلى درجات قطع أعلى من طريقة أنجوف ، بينما طريقة ندلسكاي تؤدي 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 طريقتى أنجوف و المجموعة الحديدية ، 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 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 التصاعدي لصعوبة البنود " *Bookmark* " ، و كان وجه المقارنة بين الطريقتين عبارة عن متوسط درجات القطع عبر المحكمين، و الانحراف المعياري 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%، ودرجة 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 استخدمتا طريقة في تحديد درجة القطع لا تعتمد على المحكمين، و دراسة (Fehrmann et al.,1991) و التي استخدمت القرار الصحيح للتصنيف و الخطأ الموجب و الخطأ السالب ، كما أن هذه الدراسة استخدمت طريقة واحدة في تحديد درجة القطع ، و هي طريقة أنجوف و درست تأثير متغيرات تجريبية معينة على قرار التصنيف الناتج عن هذه الطريقة، و من ثم فهناك ندرة في الدراسات التي 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 في الدراسات السابقة أنها تغطي مجالات متعددة في النواحي الأكاديمية، و الانتقاء المهني، و الرخصة في مجال الطب، مما يعطى أهمية للطرق المستخدمة في تحديد درجة القطع ، كما يلاحظ أن معظم هذه المقاييس من نوع الاختيار من متعدد رباعى البدائل، و هذا ما اتبعه الباحث في إعداد المقياس المستخدم في البحث الحالي.

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

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

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

العظمى) باختلاف الطريقة المستخدمة فى تحديد درجة القطع (طريقة ندلسكاي، طريقة أنجوف).

إجراءات البحث

أولاً: عينة البحث:

هناك نوعان من الأفراد الذين يمثلون عينة البحث كالتالى:
النوع الأول: المحكمون *Judges* و هم الأفراد الذين يقومون بتقدير مفردات الاختبار موضوع البحث بطريقتى أنجوف و ندلسكاي للحصول على درجة القطع الناتجة عن كل طريقة ، و بلغ عددهم (١٢) محكماً، تم اختيارهم من بين المعيدىن و المدرسىن المساعدين و المدرسىن الذىن قاموا بتدريس محتوى التدريبات المعملية (سكاشن علم النفس) بمعامل كليات التربية (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

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

النوع الثانى: الأفراد المفحوصون *Examiners* و هم الأفراد الذىن يقومون بالاستجابة لمفردات الاختبار المحكى المرجع فى التدريبات المعملية لعلم النفس ، و حيث أن مؤشرات التصنيف تتطلب مجموعتين من الأفراد

إحداهما متعلمة و الأخرى غير متعلمة، لذلك فإن الأفراد ينقسمون إلى مجموعتين: المجموعة المتعلمة: وهم الطلاب الذين مروا ببرنامج تعليمي فى التدريبات المعملية لعلم النفس(سكاشن علم النفس) خلال الفصل الدراسى الأول من العام الجامعى ٢٠٠٤-٢٠٠٥، و بلغ قوامهم فى البداية (١٠٠) طالب من طلاب كلية التربية بقنا شعبة التعليم العام للقسم العلمى، ولكن تم استبعاد(٧) طلاب منهم لأسباب تتعلق بعدم الجدية فى الإجابة ، أو تطرف مستوى الذكاء(خارج الفئة ٩٠ - ١١٠ %) نسبة ذكاء، و بذلك وصل عددهم النهائى إلى (٩٣)طالب،

والمجموعة غير المتعلمة و هم الطلاب الذين لم يمروا ببرنامج تعليمي(سكاشن علم النفس)،خلال نفس الفصل الدراسى، وبلغ قوامهم فى البداية(٨٨)طالب تم اختيارهم من بين طلاب الفرقة الثالثة بكلية التربية شعبة التربية الرياضية خلال نفس الفصل الدراسى،و تم استبعاد(٣) طلاب منهم، لنفس الأسباب السابقة ، و بذلك وصل عددهم النهائى إلى (٨٥)طالب.

و بالتالى فإن المجموعة المتعلمة تدرس سكاشن علم النفس ، و المجموعة غير المتعلمة لا تدرس هذه السكاشن بالرغم من أن كلا المجموعتين تدرس مقررأ نظرياً فى علم النفس التعلم .
و توزيع عينة الأفراد المفحوصين مبين فى الجدول التالى:

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

جدول (٤)

توزيع عينة الأفراد المختبرين في البحث

المجموعة	المتعلمة (Masters)			غير المتعلمة (Nonmasters)
الشعبة	الرياضيات	الآحياء الكيمياء الطبيعية و	التربية الرياضية	
العدد	٥٩	٢٤	١٠	٨٥
المجموع	٩٣			٨٥
المجموع الكلّي	١٧٨			

ثانياً : أدوات البحث:

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

١- بناء الاختبار:

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

أ-تحديد الكفايات الرئيسية المكونة لمحتوى الاختبار.

ب-تحليل الكفايات الرئيسية إلى مكوناتها الفرعية.

ج-صياغة الأهداف السلوكية التي تعكس الكفايات الفرعية.

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 الوصفي ، و إجراء الثبات بطريقة معامل سابكوفياك، كما سيلي شرحه.

و فيما يلي عرض هذه الخطوات على الاختبار موضوع البحث:

الخطوة الأولى- تحديد الكفايات الرئيسية المكونة لمحتوى الاختبار(التدريبات العملية لعلم النفس):إن التدريبات العملية لعلم النفس من صميم اختصاص الباحث ، نظراً لقيامه بتدريس هذا المحتوى العملي للعديد من الطلاب عبر سنوات متعاقبة ، لذا فقد قام الباحث بتحديد خمس كفايات رئيسية تعبر عن أساسيات التدريبات العملية 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 في 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 تنظيمياً هرمياً بحيث أن إتقان المتعلم لمهارة أو كفاية معينة يعد متطلباً أساسياً لإتقانه المهارات التالية، و هذه ما ينطبق على المهارات المتطلبية في معمل علم النفس، و لقد قام الباحث بتحليل الكفايات الخمس الرئيسية إلى (٢٥) كفاية فرعية، فعلى سبيل المثال، الكفاية الأولى (المعارف و

المهارات اللازمة لممارسة تجربة التعلم)، هذه الكفاية الرئيسية تم تحليلها إلى خمس كفايات أو مهارات فرعية و هي: ١- معرفة شكل التجربة، ٢- تصنيف تجربة التعلم في ضوء النظرية التابع لها، ٣- تحديد متغيرات التجربة، ٤- تحديد منحنيات التعلم في التجربة، ٥- معرفة الخطوات الإجرائية للتجربة، وهذه المهارات الفرعية تسير في ترتيب منطقي، حيث أن معرفة شكل التجربة يؤدي بالمفحوص إلى تصنيف التجربة في ضوء النظرية التابع لها، مما يؤدي إلى تحديد متغيرات التجربة (المستقلة و التابعة)، و الذي يؤدي به تلقائيا إلى تحديد منحنيات التعلم في التجربة ، و كل هذه المعلومات تؤدي به إلى معرفة الخطوات الصحيحة لإجراء التجربة.

و هكذا بالنسبة لباقي الكفايات ، و الذي يوضحه الجدول التالي:

جدول (٥)

الكفايات الرئيسية و الكفايات المتفرعة منها

الكفاية الرئيسية	المعارف و المهارات اللازمة لممارسة تجربة التعلم	إعداد جدول البيانات	رسم منحنى التعلم	التطبيق على منحنى التعلم	معرفة التطبيقات التربوية لتجربة التعلم
الكفايات الفرعية	معرفة شكل التجربة- تصنيف تجربة التعلم في ضوء النظرية التابع لها- تحديد متغيرات التجربة- تحديد منحنيات التعلم في التجربة- معرفة الخطوات الإجرائية للتجربة.	تحديد عدد صفوف الجدول- تحديد عدد أعمدة الجدول- رسم جدول بيانات- تدوين بيانات التجربة في الجدول.	معرفة العلاقة بين متغيرات التجربة و المحاور الأفقية و الرأسية- رسم المحور الأفقي و المحور الرأسى- تحديد الرأسى- مقاييس الرسم المناسب- الدقة في وضع نقاط المنحنى- توصيل نقاط المنحنى- توضيح جزء المنحنى الذي يمثل حدوث التعلم.	التمييز بين المنحنى الإيجابي و المنحنى السلبي- تحديد مستوى التعلم من المنحنى- تحديد سرعة التعلم من المنحنى- تحديد نقاط الذبذبات و الهضاب على المنحنى- معرفة حدود أقصى التعلم- التمييز بين المنحنى المنتظم و غير المنتظم.	استخلاص أهداف تربوية لتجربة التعلم- تحديد العوامل العقلية و غير العقلية التي من الممكن أن تؤثر على التعلم أثناء ممارستها- التجربة- تقديم أمثلة تربوية على شكل التجربة- تقديم أمثلة تربوية على قوانين التعلم للنظرية التابعة لها- التجربة.

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

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

- ١- كل مفردة ينبغي أن تقيس ناتج مهم أو هدف للتعلم.
- ٢- ينبغي أن يقدم رأس السؤال فكرة واضحة في لغة بسيطة.
- ٣- تأكد من أن كل البدائل متفكة نحويًا مع رأس السؤال و ليس البديل الصحيح فقط.

- ٤- لا تستخدم العبارات المطلقة (عادة-كل-أبدأ-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 في كل المفردات.
- ٨- اجعل المفردات مستقلة بقدر الإمكان عن بعضها .

(Davis,1983,570-571)

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

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

البيان	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 التي تعتمد على مصدر تيار كهربائي و التي لا تعتمد على مصدر تيار كهربائي - يذكر اسم النظرية التي تنبئها التجربة المعطاة- يذكر نوع التعلم السائد لتجربة تعلم معطاة- يصنف مجموعة من تجارب تعلم معطاة في ضوء نوع التعلم السائد- يصنف مجموعة من تجارب تعلم معطاة في ضوء نظرية التعلم التابعة لها- يسمي المتغير المستقل الأساسي في تجارب التعلم- يذكر عدد المتغيرات التابعة لتجربة تعلم معطاة- يسمي المتغير التابع (المتغيرات التابعة) لتجربة تعلم معطاة- يحدد مجموعة تجارب التعلم التي تشترك في متغير تابع أو أكثر- يحدد المتغير التابع (المتغيرات التابعة) التي يشترك بين مجموعة من التجارب المعطاة- يذكر عدد منحنيات التعلم لتجربة معطاة- يتعرف على العلاقة بين عدد منحنيات التعلم و عدد المتغيرات التابعة- يسمي منحنى(منحنيات) التعلم لتجربة معطاة- يتعرف على الترتيب المنطقي لخطوات إجراء تجربة تعلم معطاة- يميز بين دور الفأحص و دور المفحوص أثناء أداء التجربة	١٧	٣٢-١	٣٢
إعداد جدول البيانات	يتعرف على عدد صفوف جدول البيانات لتجربة معطاة- يتعرف على عدد أعمدة جدول البيانات لتجربة معطاة- بدون البيانات في الجدول بوحدات مناسبة- بدون بيانات المتغيرات في الجدول بترتيب منطقي- يميز بين دور الفأحص و دور المفحوص أثناء تدوين بيانات التجربة- يكتشف رقم المحاولة التي عندها يتم إيقاف التجربة	٦	٤٢-٣٣	١٠
رسم منحنى التعلم	يتعرف على العلاقة بين المتغير المستقل و المحور الأفقي- يتعرف على العلاقة بين المتغير التابع و المحور الرأسي- يرسم المحور الأفقي في منحنى التعلم- يرسم المحور الرأسي في منحنى التعلم- يعرف الأساس الذي عليه يتم اختيار مقياس الرسم- يقترح مقياس رسم مناسب في ضوء بيانات معطاة- يفهم معنى بعد نقطة المنحنى عن المحور الأفقي- يفهم معنى بعد نقطة المنحنى عن المحور الرأسي- يعرف القيمتين التين على أساسهما يتم وضع نقطة المنحنى- يدقق في وضع النقطة على المنحنى- يكتشف الشكل الصحيح لمنحنى التعلم- يستخرج الجزء من المنحنى الذي يمثل حدوث التعلم- يميز بين حدود أقصى التعلم	١٣	٥٨-٤٣	١٦
التطبيق على منحنى التعلم	يعرف بداية المنحنى الإيجابي- يعرف بداية المنحنى السلبي- يميز بين المنحنى الإيجابي و المنحنى السلبي 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	٧٣-٥٩	١٥
معرفة التطبيقات النظرية التجريبية لتجربة التعلم	يستخرج هدف تعليمي من إجراء التجربة- يسرد العوامل العقلية التي من الممكن أن تؤثر على التعلم في تجربة معطاة- يسرد العوامل غير العقلية التي من الممكن أن تؤثر على التعلم في تجربة معطاة- يحدد فائزون أو أكثر لنظرية التعلم التابع لها التجربة- يستخلص تطبيقات تجريبية لتجربة التعلم التي تنبئها التجربة- يستخلص تطبيقات تربوية من شكل التجربة	٦	٨٧-٧٤	١٤
المجموع	٥٧	٨٧		

* انظر ملحق البحث للتعرف على مفردات الاختبار.

الخطوة الخامسة (ثبات و صدق الاختبار):

(١)-صدق الاختبار: تم استخدام أسلوب الصدق الوصفي *Descriptive Validity* للتحقق من صدق الاختبار، و يعنى الصدق الوصفي 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 يفترق إلى الصدق الوصفي 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 الوصفي فى ضوء ما أوضحه صلاح الدين علام (١٩٩٠، ١٤٣-١٤٥) كالتالى: تم عرض استمارة على المحكمين (الذين قاموا بتدريس التدريبات العملية 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 سلوكى للكفاية الرئيسية التابع لها.
- (٢)- ما مدى انطباق كل مفردة اختبارية على الهدف السلوكى التابعة له.

و يقوم المحكمون بالاستجابة للسؤالين السابقين بنعم (v) 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	العلو (تم)								
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 الثاني

انطباق كل مفردة اختبارية على الهدف السلوكي التابعة له

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

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 فى ضوء آراء المحكمين، و بذلك تم التحقق من الصدق الوصفى 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 $L = \frac{S}{n}$ ، حيث S عدد المفردات التى أجاب عليها المفحوص (ف) إجابة صحيحة، n 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) = $(س \leq ٠) / (٠ \leq ١) - ١$ ل (س \leq ٠) ، حيث أن ل (س \leq ٠) هي توزيع ذات الحدين.

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

ل س. = مج — ل س. (ف) ان ، و هي تعنى متوسط معاملات الثبات لكل فرد على حدة ، حيث أن ل س. تمثل معامل سابكوفياك للثبات ، ولقد 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 ندلسكاي ، و هي 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 اللفظي

(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)، و توصل الباحث إلى معامل صدق قدره (٠,٥٧٩) و هى قيمة دالة احصائياً عند مستوى ٠,٠١ ، 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 المحكى المرجع فى التدريبات المعملية لعلم النفس بعد التحقق من صدقه و ثباته على مجموعة المحكمين البالغ عددهم (١٢) محكم لتقدير مفرداته بطريقتى أنجوف و ندلسكاي ، و لقد تم فى هذه الخطوة شرح لطريقتى أنجوف و ندلسكاي للمحكمين مدعماً بمثال تطبيقي، 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 عشري) خلال جميع المفردات، أما درجة القطع النهائية فهي متوسط درجات القطع المحددة بواسطة المحكمين (طريقة أنجوف).

* تحديد عدد المشتتات (البدائل الخاطئة) التي سيستبدها (الطالب ذو الحد الأدنى من الكفاية) تلقائياً، عندما يفكر في اختيار بديل من البدائل الأربعة، ولقد تم تمييز البديل الصحيح عن المشتتات بخط مغاير، وبذلك تصبح درجة القطع المحددة بواسطة كل محكم هي مجموع مقبولات عدد البدائل المتبقية، أما درجة القطع النهائية فهي متوسط درجات القطع المحددة بواسطة المحكمين (طريقة ندلسكاي).

٣- تم تقدير درجات القطع على الاختبار وذلك بالنسبة لكل محكم، وكذلك تم حساب درجتى القطع النهائيتين للاختبار، بطريقتى أنجوف، و ندلسكاي .

٤- تم تطبيق اختبار ذكاء الشباب اللفظي على أفراد عينة المفحوصين (المجموعة المتعلمة "١٠٠" طالب-المجموعة غير المتعلمة "٨٨" طالب)، وذلك لضبط متغير الذكاء بين المجموعتين، وبعد تطبيق الاختبار وتصحيحه وجد الباحث أن طلاب المجموعتين يدخلون في فئة نسبة الذكاء (٩٠-١١٠)، باستثناء (٦) طلاب، منهم (٤) طلاب من المجموعة المتعلمة، و طالبين من المجموعة غير المتعلمة، وذلك لكي يكون هناك تقارب في مستوى الذكاء بين المجموعتين، كما تم استبعاد (٤) طلاب آخرين من المجموعتين لعدم جديتهم في الاستجابة لمفردات اختبار الذكاء، وبذلك تصبح المجموعتان متقاربتين إلى حد ما الذكاء، ومع تقارب المجموعتين في العمر بحكم أنهما من فرقة واحدة، وكذلك في 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 التي 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 المحكى المرجع فى 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 ندلسكاي).

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

نتائج البحث

نتيجة الفرض الأول :

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

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 ندلسكاي ، ثم تم تحليل استجابات المحكمين و التوصل إلى درجة قطع خاصة بكل محكم ، و كذلك متوسط

لدرجات القطع على كل المحكمين و هذا بالنسبة لكل طريقة على حدة ،
و هذه البيانات معرّضة في الجدول الآتي:

جدول (٩)

درجات القطع المحددة بواسطة المحكمين

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

و لقد تمت المقارنة بين درجة القطع الناتجة عن طريقة أنجوف و درجة القطع الناتجة عن طريقة ندلسكاي عن طريق الأسلوب الإحصائي 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	الدالة 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 ندلسكاي ، و ذلك لصالح طريقة أنجوف ، و هذا يعنى أن استخدام طريقة أنجوف يؤدي إلى تحديد درجات قطع أعلى من درجات القطع المحددة بواسطة طريقة ندلسكاي .

و يفسر الباحث هذه النتيجة على أساس أنه إذا كانت طريقة ندلسكاي تتطلب من المحكم تحديد عدد المشتتات التي سيستبدها الطالب نو الحد الأدنى من الكفاية تلقائياً أثناء التفكير في اختيار البديل الصحيح و بالتالى تتطلب من المحكم فهم لمفهوم الطالب ذى الحد الأدنى من الكفاية ، فان الفرق بين الطريقتين و الذى جعل طريقة أنجوف تودى إلى درجة قطع مرتفعة مقارنة بطريقة ندلسكاي ، هو أن طريقة أنجوف تتطلب من المحكمين تقدير المفردة دون التفحص الدقيق لبدائل هذا المفردة. أما طريقة ندلسكاي و خاصة فى الاختبار ذات المفردات رباعية البدائل فهناك أربعة احتمالات لتحديد عدد المشتتات لكل مفردة و يقابله أربع قيم احتمالية على المفردة بناءً على عدد المشتتات ، فعندما لا يحدد المحكم مشتتات تكون عدد البدائل المتبقية (٤) ، و مقلوبها (٠,٢٥) يمثل القيمة الاحتمالية على هذا المفردة ، أما عندما يحدد مشتت واحد تكون 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 يشير Shepard فى عام ١٩٨٤ إلى أن غالبية المحكمين يميلون إلى اختيار مشتتين و الذى يقابله قيمة احتمالية على المفردة قيمتها (٠,٥) ، و الذى يؤدى بدوره إلى انخفاض درجة القطع المحددة بواسطة طريقة ندلسكاي (In:Berk , 1986 , 152) ، فطريقة ندلسكاي تسمح فقط بقيم احتمالية منفصلة و هى (٠,٢٥ ، ٠,٣٣ ، ٠,٥ ، ١) بحسب عدد بدائل المفردة ، و لا تسمح بقيم تنحصر بين (٠,٥) و (١) ، و بالتبعية لا يميل معظم المحكمين فى كثير من الأحيان إلى تعيين القيمة الاحتمالية (١) (أى مستوى الإتيقان ١٠٠%) ، و لذلك يضطرون لتعيين القيمة الاحتمالية (٠,٥) لمعظم المفردات ، و حيث أن درجة القطع هى فى الواقع مجموع هذه القيم الاحتمالية على كل المفردات ، لذلك تكون

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

و ربما يكون الفرق في درجة القطع بين طريقتي أنجوف و ندلسكاي دالة لمعرفة المحكمين بالمفردة و محتواها ، و قدرتهم في الإجابة عليها ، حيث يضع المحكمون قيماً احتمالية أعلى للمفردات الملمين بها و بمحتواها ، و في المقابل يضعون قيماً احتمالية أقل للمفردات التي لا يعرفون الإجابة عليها و تمثل مفردات صعبة بالنسبة لهم ، و التفحص الكامل للبدائل في طريقة ندلسكاي و ملاحظة المحكم مدى التشابه بين البدائل و منطقية كل بديل كإجابة محتملة تخلق تحدى لدى المحكم ، و الذى يؤدي بدوره إلى اقتناعه بعدم سهولة المفردة نظراً للمشتتات المتشابهة و التي تحير المفحوص ، و بالتبعية يحدد المحكم قيماً احتمالية أقل و الذى يؤدي إلى انخفاض درجة القطع المحددة بواسطة ندلسكاي ، و العكس يحدث في طريقة أنجوف ، حيث أن عدم تفحص المحكم لبدائل المفردة و ما بينها من فروق طفيفة و حضور الإجابة له أثناء تقدير المفردة ، تجعل المحكم يقدر تقدير مبالغ فيه *overestimate* لسهولة المفردة ، و بالتبعية يحدد المحكم قيماً احتمالية أعلى و الذى يؤدي إلى ارتفاع درجة القطع المحددة بواسطة أنجوف، و لذلك نجد أن نقص خبرة المحكم بالمفردة و ما تحويه من مفاهيم له تأثير أقل على قرار التصنيف المتخذ في ضوء طريقة أنجوف ، بعكس طريقة ندلسكاي و التي نجد فيها أن الصعوبة التي ربما يجدها المحكم في مفردة ما تؤثر في قراره و الذى يؤدي إلى انخفاض درجة القطع (Chang , 1996) .

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

و لقد اتفقت نتائج العديد من الدراسات مع نتيجة الفرض الحالي في اختلاف درجة القطع المحددة بواسطة طريقة أنجوف عن درجة القطع المحددة بواسطة طريقة ندلسكاي ، و هنا انقسمت النتائج إلى جانبين ، الأول و هو الغالب، و يشير إلى أن الفرق بين درجتى القطع المحدد بالطريقتين 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 باندلسكاي

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

و بالرغم من أن هذه الدراسات تشير إلى ارتفاع درجة القطع المحددة بطريقة أنجوف مقارنة بدرجة القطع المحددة بطريقة ندلسكاي ، إلا أن هناك جانب اخر من الدراسات توصل إلى انخفاض درجة القطع المحددة بطريقة أنجوف مقارنة بطريقة ندلسكاي ، منها دراسة (Livingston & Zikey, 1989) التي توصلت إلى انخفاض درجة القطع المحددة بطريقة أنجوف مقارنة بطريقة ندلسكاي و ذلك في اختبار القراءة فقط ، و لكن في اختبار الرياضيات توصلت إلى عكس النتيجة كما هو موضح في الجدول ، فبالنسبة لاختبار القراءة توصلت إلى درجة القطع باستخدام طريقة أنجوف قدرها (٣٤) ، و باستخدام طريقة ندلسكاي كانت درجة القطع (٤٢) . و دراسة smith & smith في عام ١٩٨٨ التي توصلت إلى أن درجة القطع المحددة بواسطة طريقة أنجوف قدرها (٦٩ %) ، و باستخدام طريقة ندلسكاي قدرها (٧١ %) (in: chang, 1996). و قد يرجع اختلاف نتيجة هاتين الدراستين مع نتيجة الدراسة الحالية إلى اختلاف محتوى الاختبار، و اختلاف طريقة القاء التعليمات على المحكمين . و مجمل ما سبق هو أن الفرض الذي تم وضعه و الذي يشير إلى اختلاف درجة القطع الناتجة عن طريقة أنجوف عن درجة القطع الناتجة عن طريقة ندلسكاي صحيح ، و جميع نتائج الدراسات التي تم عرضها تؤيد هذه الفرض، ومعظمها يشير إلى أن هذا الفرق لصالح طريقة أنجوف ، أما العدد القليل منها فيشير إلى أن هذا الفرق لصالح طريقة ندلسكاي .

نتيجة الفرض الثاني:

و الذي ينص على أنه: تختلف 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 في تحديد درجة القطع " طريقة أنجوف - طريقة ندلسكاي " .

أشارت نتيجة الفرض الأول إلى أن طريقة أنجوف تؤدي إلى درجة قطع قيمتها (٦٩) و هي أعلى من درجة القطع المحددة بطريقة ندلسكاي و التي قيمتها (٤٧) ، و نتيجة الفرض الأول ترتبط ارتباط كلى بنتيجة الفرض الثاني ، بمعنى أن ارتفاع درجة القطع الناتجة عن طريقة أنجوف ، و ما يقابلها من انخفاض في درجة القطع في طريقة ندلسكاي ، هو الذي سيجعل هناك اختلاف في مؤشرات قرار تصنيف الطلاب في الطريقتين ، و الذي يصب في صالح طريقة ندلسكاي ، و في الواقع فإن نتيجة الفرض الثاني تشير إلى اختلاف مؤشرات قرار تصنيف الطلاب الناتج عن الطريقتين ، باستثناء الخطأ الموجب للقرار ، و الذي كان معدوماً في الطريقتين و فيما يلي تحليل للفرض الثاني الرئيسي إلى سبعة فروض فرعية و التحقق من صحتها كالتالي :

الفرض الفرعي الأول:

يختلف الخطأ الموجب لقرار التصنيف باختلاف الطريقة المستخدمة في تحديد درجة القطع "طريقة أنجوف - طريقة ندلسكاي" ؟.

للتحقق من صحة هذا الفرض تم استخدام المعادلة $P(FM) = FM/M+N$ ، حيث FM هي عدد القرارات الخاطئة من النوع الموجب، أما المقام $(M+N)$ فيمثل العدد الكلي للمجموعتين (المتعلمة و غير المتعلمة)، حيث $M = 93$ (عدد أفراد المجموعة المتعلمة) ، $N = 85$ (عدد أفراد المجموعة غير المتعلمة) ، و بالتالي يكون العدد الكلي $(M + N) = 93 + 85 = 178$ ، و هذا العدد الكلي ثابت في الطريقتين المستخدمتين، و الجدول التالي يوضح نتيجة هذا الفرض:

جدول (١٢)

نتيجة الفرض الفرعى الأول

البيان	عدد الأخطاء الموجبة (FM)	نسبة احتمال الخطأ الموجب لقرار التصنيف $p(FM)$
الطريقة ندلسكاي	صفر	صفر
أنجوف	صفر	صفر

يتضح من الجدول السابق أنه لا توجد أخطاء موجبة فى الطريقتين المستخدمتين ، فكلتا 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 ندلسكاي .

الفرض الفرعى الثانى:

يختلف الخطأ السالب لقرار التصنيف باختلاف الطريقة المستخدمة فى تحديد درجة القطع "طريقة ندلسكاي - 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 $P(FN)$
ندلسكاي	١٤	٠,٠٨
أنجوف	٨٣	٠,٤٧

يتضح من الجدول السابق أن احتمال الخطأ السالب لقرار التصنيف الناتج عن طريقة ندلسكاي $[P(FN) = 0.08]$ ، 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 التظيمى (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 ندلسكاي (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 ندلسكاي 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 ندلسكاي" ؟.

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 $(M+N)$ ، 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	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)
ندلسكاي	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 ندلسكاي $[P(CD) = 0.92]$ ، أكبر من احتمال صحة قرار التصنيف الناتج عن طريقة أنجوف $[P(CD) = 0.54]$ ، وهذا يعنى أن درجة القطع الناتجة عن طريقة ندلسكاي 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 ندلسكاي (٤٧) ، مقارنة بطريقة أنجوف (٦٩) ، هذا الانخفاض الذى أدى إلى انخفاض حدوث الخطأ السالب فى طريقة ندلسكاي ، و الذى تم توضيحه فى نتيجة الفرض الفرعى الثانى يقابله منطقياً صحة فى القرار لطريقة ندلسكاي ، 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 المبني على درجة القطع المحددة بواسطة طريقة أنجوف .

الفرض الفرعى الرابع:

يختلف معامل صدق قرار التصنيف باختلاف الطريقة المستخدمة فى تحديد درجة القطع "طريقة أنجوف - طريقة ندلسكاي" ؟.

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

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

جدول (١٥)

نتيجة الفرض الفرعى الرابع

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

يلاحظ من الجدول السابق أن معامل صدق قرار التصنيف الناتج عن طريقة ندلسكاي $[\phi = 0.85]$ ، أعلى من معامل صدق قرار التصنيف الناتج عن طريقة أنجوف $[\phi = 0.24]$ ، و تعنى هذه النتيجة قدرة درجة القطع المحددة بطريقة ندلسكاي على التنبؤ بالمتقنين الحقيقيين (TM) ، و غير المتقنين الحقيقيين (TN) ، أى أن درجة القطع صادقة فى الفصل بين المتقنين و غير المتقنين ، مقارنة بأنجوف ، و هى نتيجة

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

الفرض الفرعي الخامس:

تختلف دالة الضرر الناتجة عن قرار التصنيف باختلاف الطريقة المستخدمة في تحديد درجة القطع 'طريقة أنجوف- طريقة ندلسكاي' .؟
 للتحقق من صحة هذا الفرض تم استخدام المعادلة $\delta_x = P(FN) (D_1) + P(FM) (D_2)$ ، و لقد تم شرح هذه المعادلة ، و التي 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 $\delta_x = -P(FN) - 2 P(FM)$ ، حيث $p(FN)$ هي نسبة احتمالية الأخطاء السالبة ، $p(FM)$ هي نسبة احتمال الأخطاء الموجبة ، و فيما يلي الجدول الذي يبين نتيجة الفرض الخامس

جدول (١٦)

نتيجة الفرض الفرعي الخامس

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

يلاحظ من الجدول السابق أن دالة الضرر لقرار التصنيف الناتج عن طريقة ندلسكاي $[\delta_k = -0.08]$ ، أقل من دالة الضرر لقرار التصنيف الناتج عن طريقة أنجوف $[\delta_k = -0.47]$.، ويفسر الباحث هذه النتيجة بأنها منطقية و تتفق مع النتائج الفرعية السابقة ، حيث أنه مع تثبيت الخطأ الموجب لقرار التصنيف و المعدوم في الطريقتين ، نجد أن انخفاض الخطأ السالب للتصنيف و الذى يميز طريقة ندلسكاي 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 ندلسكاي ، 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 المبني 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 ندلسكاي" ؟. للتحقق من صحة هذا الفرض تم استخدام المعادلة $U_1 = P(TM) + P(TN)$ ، 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 $U_1 = 2P(TM) + P(TN)$ ، و فيما يلى الجدول الذى

يبين نتيجة الفرض الفرعى السادس

جدول (١٧)

نتيجة الفرض الفرعى السادس

دالة المنفعة v_x	نسبة		البيان الطريقة
	نسبة غير المتقين الحقيقيين $P(TN)$	المتقين الحقيقيين $P(TM)$	
١,٣٦	٠,٤٨	٠,٤٤	ندلسكاي
٠,٦٠	٠,٤٨	٠,٠٦	أنجوف

يلاحظ من الجدول السابق أن دالة المنفعة لقرار التصنيف للناتج عن طريقة ندلسكاي $[v_x = 1.36]$ ، أكبر من دالة المنفعة لقرار التصنيف الناتج عن طريقة أنجوف $[v_x = 0.60]$ وذلك 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 ندلسكاي، 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 ندلسكاي' .؟

للتحقق من صحة هذا الفرض تم استخدام المعادلة $\tau = \sum_{k=1}^{m+n} (v_k + \delta_k)$ حيث v_k تمثل دالة المنفعة ، δ_k تمثل دالة الضرر ، و فيما يلي الجدول الذى يبين نتيجة الفرض الفرعى السابع

جدول (١٨)

نتيجة الفرض الفرعى السابع من الفرض الثانى

البيان	دالة المنفعة v_k	دالة الضرر δ_k	دالة المنفعة العظمى τ
ندلسكاي	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 ندلسكاي $\tau = 227,84$ ، أكبر من دالة المنفعة العظمى لقرار التصنيف الناتج عن طريقة أنجوف $\tau = 23,14$ ، 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 المبني على درجة القطع المحددة بواسطة ندلسكاي 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 المبني 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 الطبيعي هذه المؤشرات السلبية ينتج عنها ، أقل قدر من المنفعة (المنفعة العظمى) ، و فيما جدول يبين ملخص لنتيجة الفرض الثاني بفروضه الفرعية السابع :-

جدول(١٩)

نتائج الفرض الثاني

البيان الطريقة	القرار الصحيح	الخطأ الموجب	الخطأ السالب	معامل صدق القرار	دالة المنفعة U_K	دالة الضرر δ_K	دالة المنفعة العظمى ϵ
ندلسكاي	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 الثاني إلى أن خصائص قرار التصنيف الناتج عن طريقة ندلسكاي ((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 بالاجابية 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 في 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

بطريقة ندلسكاي ، و التى نتج عنها قرار تصنيفى يتسم بإيجابية مؤشراتته.

و بالتالى فإن طريقة أنجوف ، بالرغم من انتشارها الواسع، إلا أنها يعترها بعض العيوب التى تجعلنا لا نثق فى قرار التصنيف الناتج عنها، فيشير العديد من الباحثين و المصادر البحثية أن هناك نقد كبير موجه لطريقة أنجوف و هو أن تقديرات المحكمين للقيم الاحتمالية للمفردات من المحتمل جداً أن تكون غير دقيقة و أحياناً غير متسقة و متناقضة *Contradictory* (Wang ,2003 ,231-232). كما أشارت الأكاديمية التربوية فى التربية (NAE) إلى أن هناك خطأ جوهري فى طريقة أنجوف و هو أن المحكمين مطلوب منهم إنجاز مهمة معرفية تكاد تكون مستحيلة *Nearly Impossible* ، هذه المهمة هى "تقدير احتمالية إجابة الطالب ذى الحد الأدنى من الكفاية و الذى هو فى الأصل شخصية افتراضية *Hypothesis* إجابة صحيحة على كل مفردة من مفردات الاختبار (In : Berk ,1996 ,216)، حيث أشار *poggio* فى عام ١٩٨٤ إلى أن إحدى عيوب طريقة أنجوف هى أن معظم المحكمين يواجهون صعوبة فى تعريف الطالب ذى الحد الأدنى من الكفاية (In : Berk , 1986 , 148) ، و فى هذا الصدد أشار *Sizmur* إلى أن طريقة 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 يؤدي إلى انخفاض مصداقية الطريقة فى التوصل إلى درجة قطع موضوعية يمكن الوثوق بها(3, 1997, sizmur) .

و إذا كانت طريقة ندلسكاي تتطلب أيضاً من المحكمين تصور و تخيل أداء الطالب ذى الحد الأدنى من الكفاية ، فإن حجم الخطأ الناتج عن عدم قدرة المحكم فى طريقة أنجوف على تخيل هذا الطالب أكبر من حجم الخطأ الناتج فى طريقة ندلسكاي ، فهذه العملية التخيلية و التى

تحدث في الطريقتين معرضة للخطأ و بالتالى ينجم عنها خطأ فى قرار التصنيف ، فإذا تطرقنا لطريقة أنجوف نجد هناك قرار وحيد يتخذه المحكم على تقدير كل مفردة ، لأن المحكم يتعامل مع البديل الصحيح للإجابة، وهذا القرار معرض أن يكون خاطئ و خطأ القرار الحادث فى طريقة أنجوف لا يمكن تعديله ، و فى المقابل نجد فى طريقة ندلسكاي أن المحكم لا يتخذ قرار وحيد على تقدير كل مفردة و إنما عدة قرارات *Multi Decisions Making* ، حيث أن المحكم فى طريقة ندلسكاي يتعامل مع مشتتات المفردة ، و تمثل استجابته لكل مشتت قرار (و بالتالى فإن المحكم فى طريقة ندلسكاي لا يصدر قرار وحيد عن المفردة ، كما فى طريقة أنجوف ، و إنما عدد من القرارات يساوى عدد المشتتات)، و بذلك فإن الخطأ الناتج عن التقدير النهائى للمفردة هو معدل (متوسط) القرارات المعرضة للخطأ ، وهذا التعديل يخفف من الأخطاء الناتجة عن القرار، فعلى سبيل المثال ، إذا أصدر المحكم تقدير خاطئ فى طريقة أنجوف بأن نسبة إجابة الطالب ذى الحد الأدنى من الكفاية على (مفردة ما رباعية البدائل ، و البديل الصحيح هو 'ج -') هى ١٠٠% ، و بالتالى فهناك خطأ فى القرار بنسبة ١٠٠% ، فإذا أصدر المحكم نفس الحكم على المفردة و لكن بطريقة ندلسكاي ، نجد هناك فرصة كبيرة لتقليل 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' .

القرار الثانى بأن الطالب ذا الحد الأدنى من الكفاية سيستبعد المشتت 'ب' .

القرار الثالث بأن الطالب ذا الحد الأدنى من الكفاية سيستبعد المشتت 'د' .

و بالتالى تكون القيمة الاحتمالية للمفردة هي (١) أى إجابة صحيحة بنسبة ١٠٠% هذا التقدير النهائى على المفردة ليس قرار وحيد كما فى طريقة أنجوف ، و لكن عبارة عن ثلاثة قرارات، و هناك صعوبة أن تكون الثلاث قرارات خاطئة فى نفس الوقت ، بمعنى آخر أن مجرد إصدار قرار فقط صحيح من الثلاثة يعنى تخفيض نسبة الخطأ الناتج عن التقدير النهائى للمفردة ، و هذا الذى يميز درجة القطع الناتجة عن طريقة ندلسكاي (Chang , 1996) . و بذلك نجد أن احتمال تعرض طريقة ندلسكاي لأخطاء التقدير من جانب المحكم ، أقل من طريقة أنجوف ، و هذا الذى جعل طريقة ندلسكاي تتميز بقرار تصنيفى ذات مؤشرات إيجابية، فطريقة ندلسكاي تلقى استحسان عند عدد من الباحثين ، فعلى الأقل لم يتم توجيه لها نقد مثل الكم الكبير جداً من النقد الموجه إلى طريقة أنجوف ، فلقد توصلت دراسة (Chang,1996) إلى أن طريقة ندلسكاي 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.

كما أظهر عدد من الدراسات النتائج السلبية التى تترتب عن استخدام طريقة أنجوف فى تحديد درجة القطع ، منها دراسة (Fehrmann et al.,1991) التى توصلت إلى أن طريقة أنجوف تتسم بانخفاض معدلات القرارات الصحيحة (٠,٥٤) ، و كذلك ارتفاع نسبة الأخطاء الموجبة (٠,٠٥) و السالبة (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 عن الفرد ذى الحد الأدنى من الكفالية و تعريف دقيق له و أداء حقيقى عن أدائه و الذى أدى إلى ارتفاع معدلات القرارات الصحيحة ، و انخفاض نسبة الأخطاء السالبة ، كما توصلت دراسة (Shen , 2001) إلى أن معدل الفشل أو الإخفاق *Failure Rate* الناتج عن استخدام طريقة أنجوف 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 ندلسكاي تحظى بأكبر قدر من الثبات الداخلى بين المحكمين مقارنة بطريقة أنجوف ، مما يعطى مصداقية لقرار التصنيف الناتج عن طريقة ندلسكاي .

و لذلك أشار (Shen , 2001) إلى 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 ندلسكاي 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 في الجامعات المختلفة ، و خاصة في الكليات التي تحتاج إلى قدرات خاصة مثل كلية التربية النوعية ، و كلية التربية الرياضية .

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

٥- على المسؤولين و الجهات المعنية توجيه مزيد من الاهتمام إلى موضوع درجة القطع ، و التحقق من المنافع و الأضرار التي تقابل هذه الدرجة و التي 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 : 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 ١٩ - ٣٩
- ٢- حامد عبد السلام زهران (١٩٧٧) : اختبار ذكاء الشباب اللفظى ، القاهرة ، عالم الكتب.
- ٣- رجاء محمود أبو علام (٢٠٠٤) : مناهج البحث فى العلوم النفسية و التربوية ، القاهرة ، ط ٤ ، دار النشر للجامعات.
- ٤- صلاح الدين محمود علام (١٩٩٠) : "استخدام نموذج راش فى بناء مقياس هدفى المرجع للمعارف الأساسية فى إعداد خطة البحوث 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 ، ع ١٤ ، 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- صلاح الدين محمود علام(٢٠٠١) :دليل الاختبار التشخيصى
مرجعى المحك للمهارات الأساسية فى بناء الاختبارات
التحصيلية ، القاهرة ، مكتبة الأنجلو المصرية .
- ٨- صلاح الدين محمود علام(٢٠٠٤) :التقويم التربوى البديل، أسسه
النظرية و المنهجية و تطبيقاته الميدانية ، القاهرة ، دار
الفكر العربى .
- ٩- صلاح الدين محمود علام (٢٠٠٥):الأساليب 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 “ ،
جولية كلية البنات ، جامعة عين شمس ، ١٦ع ، 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 ،
١٣ع ، 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 ، لراسة مقارنة بين النمنونجين" ، 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, April

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, 6th 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, 5th 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 ندلسكاي في تحديد درجة القطع لاختبار محكي المرجح " دراسة تجريبية "

يهدف البحث إلى المقارنة بين درجة القطع المحددة بطريقة ندلسكاي و درجة القطع المحددة بطريقة أنجوف ، كما يهدف أيضاً إلى مقارنة مؤشرات قرار تصنيف الطلاب (الخطأ الموجب و الخطأ السالب و قرار الإلتقان و معامل الصدق و دالة لضرر و دالة المنفعة و المنفعة العظمى) بين الطريقتين، و لقد قام (١٢) محكم باستخدام الطريقتين (ندلسكاي و أنجوف) لتحديد درجة القطع على اختبار محكي المرجح في التدريبات العملية لعلم النفس، كما تكونت عينة المفحوصين من مجموعتين ، المجموعة الأولى و هي 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 التجريبي 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 ندلسكاي ، و هذه النتيجة أدت إلى ارتفاع المؤشرات الايجابية لقرار التصنيف الخاص بطريقة ندلسكاي / (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 في ضوء

الدراسات المرتبطة ، و الإطار الخلفية النظرية ، و فى ضوء هذه النتائج تم تقديم مجموعة من التوصيات التربوية ، و التى تؤكد على مزيد من الاهتمام بموضوع درجة القطع و الدقة فى تحديد هذه الدرجة و إمكانية استخدامها لاتخاذ قرارات تربوية خاصة بالانتقاء و التسكين و الانتقال لمرحل 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 uninstructional 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 method 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).