

المملكة العربية السعودية
وزارة التعليم العالي
جامعة أمّ القرى بمكة المكرمة
كلية التربية - قسم علم النفس

أثر كل من نوع المحكم وطول الاختبار على تحديد درجة
القطع لاختبار محكي المرجع يقيس الكفايات
الرياضية في العمليات الحسابية على
الأعداد بالصف السادس
الابتدائي بمدينة جدة

إعداد الطالب
خالد بن حسن شيبان التميمي
إشراف الدكتور
عبدالله بن عبدالغني صيرفي

بحث مكمل لمتطلبات نيل درجة الماجستير في علم النفس "تخصص اختبارات ومقاييس"

١٤٢٠هـ / ١٩٩٩م

ملخص الدراسة

أثر كل من نوع المحكم وطول الاختبار على تحديد درجة القطع لاختبار محكي المرجع يقيس الكفايات الرياضية في العمليات الحسابية على الأعداد بالصف السادس الابتدائي بمدينة جدة.

خالد بن حسن شيبان التميمي

هدفت الدراسة إلى تحديد أثر كل من نوع المحكم وطول الاختبار على تحديد درجة قطع اختبار محكي المرجع يقيس الكفايات الرياضية في العمليات الحسابية على الأعداد بالصف السادس الابتدائي بمدينة جدة. ولتحقيق الدراسة استخدم الباحث أربعة طرق مختلفة من طرق تحديد درجة القطع وهي: طريقة أنجوف، ونيدلسكي، والمجموعات المتضادة، والمجموعات المحكية. وأجريت الدراسة على عدة عينات هي: عينة من طلاب الصف السادس الابتدائي وبلغ عددهم (٦٥٩) طالباً، وعينة من معلمي مادة الرياضيات بلغ عددهم (٢٠) معلماً، وعينة من مشرفي مادة الرياضيات بلغ عددهم (١٠) مشرفاً، وعينة من أعضاء هيئة التدريس بلغ عددهم (١٠) أعضاء. وكان التساؤل الرئيس للدراسة هل تختلف درجة القطع في الاختبار محكي المرجع عند تطبيق طرق تحديد درجة القطع الأربعة التالية: نيدلسكاي، وأنجوف، والمجموعات المتضادة، والمجموعات المحكية) على اختبار موحد وعينة دراسية واحدة باختلاف نوع المحكم وطول الاختبار؟ وأسفرت نتائج الدراسة عن مايلي: عدم وجود فروق دالة إحصائية بين درجات قطع الاختبار الناتجة باختلاف نوع المحكم وفق طريقة أنجوف. وارتفاع تقديرات المحكمين لدرجة قطع الاختبار وفق طريقة أنجوف مقارنة مع الطرق المختلفة، وتباين درجات القطع باختلاف طول الاختبار وفق طرق تحديد درجة القطع الأربعة. تكرار نسبة درجة القطع (٥٧,١) بصورة كبيرة بين النسب المحددة لدرجة قطع الاختبارات الثلاثة (طويل، متوسط، قصير) وفق الطرق المختلفة باستثناء طريقة أنجوف. وعدم وجود علاقة بين طول الاختبار ودرجة القطع. وحققت طريقتي المجموعات المحكية ونيدلسكي أدنى نسبة لدرجة قطع الاختبار وذلك في نتائج الاختبارات الثلاث. كما تقدم الدراسة المقترحات التالية: إجراء دراسة مقارنة بين طرق تحديد درجات القطع التجريبية التي تعتمد على نظرية الاستجابة للمفردات الاختبارية، وإجراء دراسة لمعرفة أثر عقد الجلسات الدورية للمحكمين على تحديد درجة القطع وفق الطرق التحكيمية. إجراء دراسة تعتمد على النموذج ذي الحدين لتحديد درجة القطع.

الإهداء

يَ إلى من أدعو الله أن يطيل في عمريهما ويمدهما بالصحة والعافية ويختم لهما
بالصالحات أعمالهما.

"أبي وأمي"

يَ إلى التي وقفت بجانبتي وقاسمتني عناء البحث طيلة مشواري العلمي والعملية.

"زوجتي العزيزة"

يَ إلى الذين تطلعوا إلى رؤية هذا الجهد المتواضع.

"أخواني وأخواتي"

يَ إلى اللاتي شغلني عنهن إنجاز هذه الدراسة.

"بناتي سفانة وسلافة وسارة"

يَ إلى كل من أفاد في خروج هذا الجهد المتواضع إلى حيز النور.

الباحث

شكر وتقدير

الحمد لله الذي أيدني بعونه سبحانه وتعالى، وأصلي واسلم على أفضل الأنبياء والمرسلين سيدنا محمد بن عبدالله عليه أفضل صلاة وأزكى تسليم، وبعد.

لا يسعني في هذا المقام إلا أن أتوجه بالشكر الجزيل وخالص التقدير وعظيم الامتنان إلى كل من أعانني على إخراج هذه الدراسة المتواضعة إلى النور، وأخص بالشكر أستاذي ومشرفي سعادة الدكتور الفاضل عبدالله بن عبدالغني صيرفي، وفاء لما قام به من متابعة مستمرة وجهد متواصل وتوجيهات حكيمة كان لها أكبر الأثر في إنجاز الدراسة، فله مني صادق الدعاء بموفور الصحة والعافية والعطاء المستمر.

كما أتقدم بالشكر والتقدير أيضاً لجميع أعضاء هيئة التدريس بقسم علم النفس علة ما بذلوه من جهد في تنقيف وتعليم الدراسين فجزاهم الله عنا خير الجزاء.

كما أخص بالشكر والتقدير أيضاً سعادة الدكتور علي بن سعيد عسيري، والأستاذ الدكتور زايد بن عجير الحارثي اللذين كان لإسهاماتهما في مناقشة خطة الدراسة الأثر الكبير على بلورتها. كما لايفوتني أن أسجل شكري وامتناني لصاحبي السعادة عضوي لجنة المناقشة الدكتور عبداللطيف بن حميد الرائقي، والدكتور عبدالحفيظ بن سعيد مقدم على توجيهاتهما الطيبة وآرائهما السديدة وملاحظتهما الصائبة في تحسين صورة الدراسة النهائية، فلساني يعجز عن شكرهما فجزاهما الله عني خير الجزاء.

كما أتوجه بالشكر والامتنان لأستاذي القدير يوسف محمود عليوه المشرف على معمل قسم علم النفس على توجيهاته ونصحه وتشجيعه فله مني عظيم الشكر والتقدير.

كما أشكر الزميلين الفاضلين صالح بن عبدالعزيز الزهراني وعبدالرحمن بن معتوق زمزمي اللذين كان لتعاونهما ومشورتهما الأثر العظيم في نفسي فلهما مني جزيل الشكر والتقدير.

وأخص بالشكر والتقدير العميق جميع أفراد لجنة التحكيم من معلمين ومشرفين وأعضاء هيئة تدريس الذين ساهموا بصورة فاعلة في تحكيم بعض من أدوات الدراسة. كما أشكر جميع الأخوان العاملين في عينة المدارس بمدينة جدة على تعاونهم مع الباحث في تطبيق أدوات الدراسة، فلهم نمي جزيل الشكر وعظيم الامتنان وجزاهم الله عني خير الجزاء.

الباحث

قائمة المحتويات

أ	ملخص الدراسة.....
ب	الإهداء
ج	شكر وتقدير.....
د	قائمة المحتويات.....
ز	قائمة الجداول.....
١٠-١	الفصل الأول:
٢	١ - مقدمة الدراسة.
٥	٢ - مشكلة الدراسة.
٧	٣ - أهمية الدراسة.
٨	٤ - أهداف الدراسة.
٨	٥ - مصطلحات الدراسة.
١٠	٦ - حدود الدراسة.
٥٧-١١	الفصل الثاني: الإطار النظري والدراسات السابقة.
١١	الإطار النظري.
١٢	مفهوم الاختبار التحصيلي.
١٢	استخدامات الاختبارات التحصيلية.
١٢	تصنيفات اختبارات التحصيل.
١٣	أوجه الاختلاف بين الاختبارات معيارية المرجع ومحكية المرجع.
١٤	تعريف الاختبار محكي المرجع.
١٥	صدق الاختبارات محكية المرجع.
١٧	ثبات الاختبارات محكية المرجع.
٢٢	تحديد طول الاختبار محكي المرجع.
٢٣	مسميات وتعريف درجة القطع.
٥٥-٢٤	طرق تحديد درجة القطع.
٢٤	أولاً: الطرق التحكيمية.

٣٠	ثانياً: الطرق التي تعتمد جزئياً على التحكيم وتسترشد ببيانات تجريبية.
٣٥	ثالثاً: الطرق التي تعتمد على بيانات تجريبية وتسترشد بالتحكيم.
٤١	أسس تقويم طرق تحديد درجة القطع.
٤١	١ - الكفاية الفنية.
٤٢	٢ - الجانب التطبيقي.
٥٥-٤٣		الدراسات السابقة.
٤٣	١ - الدراسات العربية.
٤٧	٢ - الدراسات الأجنبية.
٥٥	٣ - ملخص الدراسات السابقة وعلاقتها بالدراسة الحالية.
٧٥-٥٨		الفصل الثالث: إجراءات الدراسة.
٥٩	أولاً: فرضيات الدراسة.
٥٩	ثانياً: منهج الدراسة.
٥٩	ثالثاً: مجتمع وعينة الدراسة.
٦٢	رابعاً: أدوات الدراسة.
٦٢	صدق أداة الدراسة.
٦٢	ثبات أداة الدراسة.
٦٣	١ - تحديد مفردات اختبارات الدراسة الحالية.
٦٨	٢ - تحديد درجات قطع الاختبار وفق طريقة نيدلسكي.
٦٨	٣ - تحديد درجات قطع الاختبار وفق طريقة أنجوف.
٦٩	٤ - تحديد درجات قطع الاختبار وفق طريقة المجموعات المتضادة.
٧٠	٥ - تحديد درجات قطع الاختبار وفق طريقة المجموعات المحكية.
٧٢	خامساً: التحليلات الإحصائية المستخدمة في الدراسة.
٨٩-٧٦		الفصل الرابع: عرض وتفسير نتائج الدراسة.
٨١-٧٧	أولاً: عرض نتائج أثر نوع المحكم على تحديد درجة القطع.
٧٧	١ - تحديد درجة القطع وفق طريقة نيدلسكي باختلاف نوع المحكم.

٧٩	٢- تحديد درجة القطع وفق طريقة أنجوف باختلاف نوع المحكم.
٨٩-٨١	ثانياً: عرض ومناقشة أثر متغير طول الاختبار.
٨٤-٨١	١- عرض ومناقشة أثر متغير طول الاختبار وفق كل طريقة على حدة.
٨١	أ- طريقة نيدلسكي.
٨٢	ب- طريقة أنجوف.
٨٣	ج- طريقة المجموعات المتضادة.
٨٤	د- طريقة المجموعات المحكية.
٨٩-٨٥	٢- عرض ومناقشة أثر متغير طول الاختبار على طرق الدراسة مجتمعة.
٨٥	أ- الاختبار الطويل.
٨٦	ب- الاختبار المتوسط.
٨٧	ج- الاختبار القصير.
٩٣-٩٠	الفصل الخامس: ملخص وتوصيات ومقترحات الدراسة.
٩١	أولاً: ملخص نتائج الدراسة.
٩٢	ثانياً: توصيات ومقترحات الدراسة.
٩٤	مراجع الدراسة.
١١٢-٩٧	ملاحق الدراسة.

قائمة الجداول

- جدول رقم ١: تكرارات استخدام طرق تحديد درجة القطع خلال دراسات سابقة. ٥٦
- جدول رقم ٢: أسماء مدارس عينة الطلاب ومواقعها الجغرافي وعدد طلاب العينة. ٦٠
- جدول رقم ٣: عدد المحكمين ووظائفهم والمؤسسات التربوية التي ينتمون إليها. ٦١
- جدول رقم ٤: القيم الناتجة من طرق تحديد معاملات الثبات للاختبارات الفرعية الثلاثة. ٦٢
- جدول رقم ٥: نتائج التحليلات الإحصائية المبدئية من تطبيق الاختبار على أفراد العينة. ٦٤
- جدول رقم ٦: تصنيفات مستويات الصعوبة وعدد المفردات ومعامل الثبات (a). ٦٥
- جدول رقم ٧: مستويات صعوبة مفردات الاختبار الطويل (٣٥) مفردة. ٦٦
- جدول رقم ٨: مستويات صعوبة مفردات الاختبار المتوسط. ٦٦
- جدول رقم ٩: مستويات صعوبة مفردات الاختبار القصير. ٦٧
- جدول رقم ١٠: قيم معامل ثبات (α) للاختبارات الثلاثة. ٦٧
- جدول رقم ١١: تكرارات ونسب درجة قطع الاختبار الطويل (٣٥) مفردة الناتجة وفق طريقة نيدلسكي باختلاف نوع المحكم وطول الاختبار. ٦٨
- جدول رقم ١٢: تكرارات ونسب درجة القطع الناتجة وفق طريقة المجموعات المتضادة باختلاف طول الاختبار. ٦٩
- جدول رقم ١٣: درجات القطع الناتجة وفق طريقة المجموعات المحكية باختلاف طول الاختبار. ٧٢
- جدول رقم ١٤: مميزات وعيوب الطرق المستخدمة في الدراسة. ٧٢
- جدول رقم ١٥: تكرارات ونسب درجة قطع الاختبار الطويل وفق طريقة نيدلسكي. ٧٧
- جدول رقم ١٦: تحليل التباين أحادي الاتجاه لمتوسط تقديرات فئات المحكمين الثلاثة لدرجة قطع الاختبار الطويل (٣٥) مفردة وفق طريقة نيدلسكي. ٧٨
- جدول رقم ١٧: نتائج تطبيق اختبار (توكي) الناتجة عن تطبيق طريقة نيدلسكي. ٧٨
- جدول رقم ١٨: تكرارات ونسب درجة قطع الاختبار الطويل الناتجة وفق طريقة أنجوف باختلاف نوع المحكم. ٧٩
- جدول رقم ١٩: نتيجة تحليل التباين أحادي الاتجاه لمتوسط تقديرات المحكمين الثلاثة لدرجة قطع الاختبار الطويل وفق طريقة أنجوف. ٨٠
- جدول رقم ٢٠: درجات القطع ونسبة الطلاب المتقنين ونسبة الطلاب غير المتقنين للاختبار الطويل (٣٥) مفردة الناتجة عن الطريقتين التحكيميتين باختلاف نوع المحكم. ٨٠
- جدول رقم ٢١: نسب درجات القطع الناتجة وفق طريقة نيدلسكي باختلاف طول الاختبار. ٨٢

- ٨٢ جدول رقم ٢٢: تحليل التباين أحادي الاتجاه لقيمة درجة القطع وفق طريقة نيدلسكي باختلاف طول الاختبار.
- ٨٣ جدول رقم ٢٣: درجات القطع الناتجة وفق طريقة أنجوف باختلاف طول الاختبار.
- ٨٣ جدول رقم ٢٤: تحليل التباين أحادي الاتجاه لقيمة درجة القطع وفق طريقة أنجوف باختلاف طول الاختبار.
- ٨٤ جدول رقم ٢٥: درجات القطع الناتجة وفق طريقة المجموعات المتضادة باختلاف طول الاختبار.
- ٨٤ جدول رقم ٢٦: درجات القطع الناتجة وفق طريقة المجموعات المحكية باختلاف طول الاختبار.
- ٨٥ جدول رقم ٢٧: تكرارات ونسب الطلاب المتقنين وغير المتقنين باختلاف الطرق وفق الاختبار الطويل.
- ٨٥ جدول رقم ٢٨: تكرار ونسب الخطأين الأول والثاني والطلاب المتقنين وغير المتقنين فعلاً للاختبار الطويل وفق الطرق المختلفة.
- ٨٦ جدول رقم ٢٩: نسب وتكرارات الطلاب المتقنين وغير المتقنين للاختبار المتوسط وفق الطرق المختلفة.
- ٨٦ جدول رقم ٣٠: نسب وتكرارات الخطأين الأول والثاني والطلاب المتقنين وغير المتقنين فعلاً وفق الطرق المختلفة للاختبار المتوسط.
- ٨٧ جدول رقم ٣١: تكرار ونسب الطلاب المتقنين وغير المتقنين للاختبار القصير وفق الطرق المختلفة.
- ٨٧ جدول رقم ٣٢: نسب وتكرار الخطأين الأول والثاني والطلاب المتقنين وغير المتقنين فعلاً للاختبار القصير وفق الطرق المختلفة.
- ٨٨ جدول رقم ٣٣: تكرارات ونسب درجات القطع الناتجة باختلاف طول الاختبارات وفق الطرق المختلفة.

الفصل الأول

- مقدمة الدراسة.
- مشكلة الدراسة.
- أهمية الدراسة.
- أهداف الدراسة.
- مصطلحات الدراسة.
- حدود الدراسة.

مقدمة الدراسة

تعتمد عمليتي القياس والتقويم التربوي على أدوات ومقاييس تقدم بيانات كمية تتيح للتربويين فهم الظاهرة التربوية بشكل دقيق ومحدد، وتعتبر الاختبارات والمقاييس التربوية جزءاً هاماً من أجزاء العملية التربوية فمن خلالها يستطيع المعلم الحكم على مدى تحقق أهداف البرامج التعليمية والتدريبية التي يقوم بتدريسها، كما أنها تفيد الطالب أيضاً في التعرف على مستوى تحصيله الدراسي من خلال أدائه في الاختبار. كما تتعدد استخدامات الاختبارات في المؤسسات التربوية فمنها ما يستخدم للقياس والتقويم أو التشخيص... الخ.

بنظرة سريعة على حركة بناء المقاييس التربوية المعاصرة يلاحظ أن هناك ثلاثة اتجاهات رئيسية يختلف كل اتجاه منها عن الآخر في الافتراضات التي يقوم عليها، وفي الإطار المرجعي الذي يعتمد عليه تفسير نتائج الطلاب في الاختبار، وفي إعداد أدوات القياس وهي القياس معياري المرجع Norm Referenced Measurement، والقياس محكي المرجع Criterion Referenced Measurement، ونظرية السمات الكامنة Latent Traits Theory. وتجدر الإشارة إلى أن كثيراً من علماء القياس والتقويم التربوي تناولوا أوجه الشبه والاختلاف بين هذه الاتجاهات المختلفة في القياس المعاصر وذلك لأهمية هذه الاتجاهات في قياس وتقويم مخرجات العملية التعليمية، وفي هذا الصدد أشار كل من جيلزر وبابام وهيوزيك (Popham & Husek, 1969; Glaser, 1994) بأن الاتجاه الأول يعتمد على تفسير درجاته في ضوء معايير محددة (Norms). أي تُرد درجات الطالب في الاختبار إلى جداول معيارية ومن ثم تحديد مستوى أداء الطالب النسبي بناءً على موقع درجة الطالب مقارنةً بمتوسط تحصيل المجموعة المعيارية في الاختبار، ويركز هذا الاتجاه على الفروق الفردية بين الطلاب بحيث يوضح مدى الفارق بين مستوى طالب وآخر، أي أنه يقيس مقدرة طالب معين مقارنةً بمقدرة طلبة آخرين.

وتعرض الاتجاه معياري المرجع لبعض أوجه النقد أشير إليها في أدبيات القياس التربوي ومنها ما ذكرها بابام وهيوزيك (Popham & Husek (1969 حيث تبيّن أن الاختبارات معيارية المرجع تعتمد على مقارنة أداء الطالب بأداء مجموعته الصفية، وبالتالي فإنه من المحتمل أن يختلف موقع الطالب باختلاف خصائص مجموعته المعيارية، بالإضافة إلى أنه ليس بالضرورة أن يمثل متوسط المجموعة المعيارية الأداء المطلوب للنجاح، وتهتم هذه الاختبارات أيضاً بإبراز الفروق الفردية بين الطلاب دون النظر إلى مدى إتقان الطالب للمهارات والمعلومات المراد قياسها وذلك لتأهيله لبرامج تدريبية وتعليمية جديدة، وبالتالي فهي لا تساعد على تقويم وتشخيص كل من كفاية الطالب وبعض البرامج التدريبية (برامج التعليم الفردي - البرامج القائمة على

الكفايات) بالقدر المطلوب. أي أنها لا تزود المعلم بمعلومات دقيقة تساعد على اتخاذ القرارات التربوية المناسبة حول مستوى كفاية الطالب والمنهج الدراسي.

كما أكد بابام (1978) Popham بأن هذه الانتقادات لعبت دوراً في ظهور تطورات جديدة في منهجية الاختبارات والمقاييس التربوية أدت إلى ظهور اتجاه آخر أطلق عليه القياس محكي المرجع Criterion-Referenced-Measurement، حيث تُسجّل أول مقالة لجليزر في عام 1963م تناقش هذا الاتجاه. كما أشار كل من هامبلتون وآخرون (1976) Hambleton, et al إلى أن هذا الاتجاه لا يعتمد على تحديد مستوى الطالب النسبي داخل مجموعته الصفية ولكنه يعتمد على تحديد ما يستطيع وما لا يستطيع الطالب أن ينجزه. أي ينسب أداء الطالب إلى محتوى الاختبار ذاته، ويعتمد هذا المقياس في بناءه على تحديد نطاق الأهداف السلوكية تحديداً دقيقاً، ومن ثم قياس مدى إتقان الطالب لهذه الأهداف من خلال مقارنة أداء الطالب بمستوى معين من الأداء (درجة القطع). وأضاف إيبيل وهامبلتون وإيجنور (1979) Ebel, (1979; Hambleton & Eignor) إلى أن القياس محكي المرجع يهدف إلى تصنيف الطلاب إلى مجموعتين الأولى متقنة Mastery والثانية غير متقنة Non-Mastery للمهارات والمعارف وذلك اعتماداً على (مستوى أداء - معيار - محك - درجة قطع - درجة نجاح) محددة تحديداً مسبقاً، ويتمثل هذا المحك (Criterion) في نسبة عدد الأسئلة التي يجب أن يقوم الطالب المتقن بالإجابة عليها بصورة صحيحة، وبالتالي إذا كانت درجة الطالب في الاختبار تزيد أو تساوي درجة القطع (Cut off Score) اعتبر الطالب متقناً للمهارات وإذا لم تكن كذلك اعتبر الطالب غير متقن. يتضح من ذلك أن عملية تصنيف الطلاب (Classification) إلى فئتين تعتمد على قيم درجات قطع الاختبارات، لذا يجب تحديد هذه الدرجة بعناية فائقة جداً لأنه يترتب على تحديدها اتخاذ قرارات هامة منها تقويم مستوى تحصيل الطالب في الاختبار وتقويم البرامج والمناهج التعليمية والتدريبية ومن ثم تشخيصها بصورة مناسبة.

ويعتبر تحديد درجة القطع (النجاح) أمراً بالغ في الصعوبة والأهمية لذلك كثرت المداخلات بين علماء القياس التربوي حول كيفية تحديد درجة القطع المثلى التي يمكن على ضوءها اتخاذ القرارات المناسبة، نتج عن تلك المداخلات طرائق متعددة ومختلفة لتحديد درجة القطع في الاختبارات محكية المرجع، والتي ذكر بيرك (1986) Berk) أنها بلغت ثمان وثلاثين طريقة ونقسم طرق تحديد درجة القطع إلى ثلاثة أقسام، يحتوي القسم الأول على الطرق التحكيمية Judgmental - Methods حيث تعتمد هذه الطرق على آراء (المحكمين) المختصين في المجال أو المحتوى المراد قياسه لتحديد درجة قطع الاختبار، ويحتوي القسم الثاني على طرق تعتمد جزئياً على التحكيم وتسترشد ببيانات تجريبية Judgmental - Empirical - Methods وهي تعتمد على آراء المحكمين مع الاعتماد على بيانات ترتبط بالأداء الفعلي للطلاب. ويحتوي القسم الثالث

على طرق تعتمد بشكل رئيس على البيانات التجريبية وجزئياً على التحكيم Empirical Methods Judgmental حيث تعتمد هذه الطرق على أداء الطلاب الفعلي في الاختبارات والتحليلات الإحصائية، وينحصر دور عملية التحكيم في انتقاء الطلاب المتقنين وغير المتقنين من الطلاب. كما أشار علام (١٩٩٥م) إلى أن كثرة الانتقادات الموجهة إلى القياس معياري المرجع كانت سبباً في ظهور نظرية السمات الكامنة (LTT)، ومن تلك الانتقادات مايلي: اعتماد بناء جميع خصائص الاختبارات السيكومترية على خصائص عينة الدراسة وعلى صعوبة عينة مفردات الاختبار، يفترض خطية ميزان درجات الطالب في الاختبار غير أنه عادة ما يكون الميزان منحنيًا لأن الفرق الثابت بين درجتين من درجات الاختبار يختلف تفسيره بناء على موقع الدرجات على متصل السمة أو القدرة المراد قياسها، كما يفترض أن العلاقة بين زيادة درجة الطالب في الاختبار وبين مقدار السمة أو القدرة علاقة خطية مطّردة وقد لا يكون هذا الافتراض صائباً إذا لم يحصل بعض الطلاب من ذوي القدرات المتميزة على درجات مرتفعة في الاختبار وفي المقابل يمكن أن يحقق بعض الطلاب من ذوي القدرات المتدنية على درجات مرتفعة في الاختبار، كما تتأثر مفردات الاختبار بالظروف المحيطة به حيث يمكن أن تتغير درجات الطلاب بحذف أو تغيير أي مفردة من مفردات الاختبار الأمر الذي يؤثر في تمثيل المفردات للنطاق السلوكي المراد قياسه. كما أشارت كاظم (١٩٨٦م) بأن هذه النظرية وفّرت نماذج تقدر السمات الكامنة (القدرة) في الفرد بصورة مستقلة عن أفراد العينة الدراسية، وتوجد ثلاثة أبعاد توفر أسس الموضوعية في نماذج السمات وهي: بُعد السمة المقاسة الذي يعرف بواسطة مجموعة من البنود وهي إما أن تكون أحادية (صعوبة المفردة) أو ثنائية (صعوبة المفردة ومعامل التمييز) أو ثلاثية المعلم (صعوبة المفردة ومعامل التمييز وعامل تخمين الإجابة)، وبُعد استقلالية القياس حيث لا يعتمد تقدير صعوبة البند على صعوبة البنود الأخرى، كما لا يعتمد أيضاً تقدير قدرة الأفراد على قدرة أي مجموعة أخرى، وأما بُعد توازي المنحنيات المميزة للبنود الذي يوفر الحد الذي تميّز فيه البنود بين الأفراد ذوي المستويات المختلفة في قدرة ما. كما أشار علام (١٩٩٥) إلى نماذج النظرية الثلاثة وهي نموذج (راش) ويُعد نموذجاً أحادي المعلم، ونموذج (لورد) ويُعد ثنائي المعلم، ونموذج (بيرنبوم) ويُعد ثلاثي المعلم.

مدخل إلى الدراسة

مشكلة الدراسة وتساؤلاتها:

تعتبر الاختبارات محكية المرجع أكثر أنواع الاختبارات التحصيلية ملاءمة لقياس وتقويم تحصيل الطلاب، وذلك لأنها تقوم على تحديد المهارات والكفايات المطلوب إتقانها بدقة فائقة لكي يتمكن المعلم من قياسها وملاحظتها بشكل مباشر ومن ثم تقدير مدى ما حققه الطالب من تلك الأهداف، بناءً على مستوى أداء محدد، وهذا مما يساعد المعلم على عملية التشخيص، حيث يتم تصنيف الطلاب إلى فئتين متقنة وغير متقنة للمهارات والكفايات المحددة، يلي ذلك تحديد نقاط الضعف والقوة في مستوى تحصيل الطالب ومعرفة المهارات التي أتقنها والتي لم يتقنها، وبذلك يستطيع المعلم وضع البرامج العلاجية المناسبة لذلك.

وتعتبر خطوة تحديد درجة قطع الاختبار محور ارتكاز بناء الاختبارات محكية المرجع، فمن خلالها تتحقق الافتراضات الأساسية التي يقوم عليها القياس محكي المرجع، فهي تزود المعلم بتقديرات كمية لتحديد مدى تمكن الطالب من المهارات والحكم على أدائه ومن ثم اتخاذ القرارات المناسبة.

لذلك فإن عملية تحديد درجة قطع الاختبار تؤثر تأثيراً مباشراً على تلك القرارات التربوية، فأبي خطأ ينتج عن تحديد هذه الدرجة يؤدي إلى الحصول على نتائج سلبية.

وتعتبر عملية تحديد درجة القطع في الاختبارات محكية المرجع من قضايا القياس والتقويم الهامة التي اختلفت فيها رؤى علماء القياس والتقويم التربوي والنفسي، حيث كثر الجدل والنقاش بينهم حول تصميم الطريقة المثلى لتحديد درجة القطع، لذلك حاولوا تقديم الأسس التي بموجبها يتم تحديد درجة القطع المناسبة التي يمكن الوثوق بها في تصنيف الطلاب، وقد حرصوا في هذه المحاولات على تلافي خطأ التداخل بين الفئتين المتقنة وغير المتقنة للمهارات التي يقيسها الاختبار والتي عادة ما تنتج عن أخطاء القياس ومن تلك الأخطاء ما ذكرها زاكي Zieky (عبدالسلام، ١٩٩٢م) وهي عامل الصدفة وعامل الغش وعامل التخمين، وعادة ما تؤدي هذه العوامل إلى ظهور خطأي التصنيف الأول (α) (تصنيف الطالب المتقن ضمن فئة غير المتقنين) والثاني (β) (تصنيف الطالب غير المتقن ضمن فئة المتقنين) والتي يترتب عليهما نتائج وأضرار سلبية على الطالب، وهي إعادة الطالب المتقن الوحدة الدراسية مرة أخرى بالرغم من تمكنه من المهارات المطلوبة، وفي المقابل تمكين الطالب غير المتقن بالوحدات الدراسية اللاحقة.

وقد تعددت طرق تحديد درجة القطع باختلاف رؤى علماء القياس التربوي والنفسي، الأمر الذي أدى إلى اختلاف قيم درجات القطع الناتجة عند تطبيق الطرق المختلفة على اختبار موحد

وعلى عينة دراسية موحدة، وذلك ما أكدته دراسات هالبن، وآخرون؛ وسكاكون وكلينج؛ وأندرود وجيمس (Haplin et al,1983; Skakun&Kling,1980; Andrew&James,1976).

وبالرغم من أن معظم الدراسات السابقة التي قارنت طرق تحديد درجات القطع المختلفة أكدت على وجود فروق جوهرية، إلا أن ما هو متوفر في أدبيات البحث العربي لا يكفي لتوفير القاعدة العلمية التي يمكن أن يعتمد عليها مصممو الاختبارات في اختيار الطريقة التي تتلاءم مع طبيعة الاختبار وطبيعة القرارات التعليمية المرتبطة به، فمن الملاحظ مثلاً أن تحديد درجة القطع (النجاح) في مراحل التعليم العام في المملكة العربية السعودية ثابتة تتراوح بين نسبتي (٤٠% و ٥٠%) تمّ تحديدهما بصورة ذاتية لا تفرّق بين طبيعة المادة الدراسية وأداء الطلاب وصعوبة الاختبار.

بناءً على ما تقدم تحاول الدراسة الحالية معرفة أثر كل من نوع المحكّم وطول الاختبار على تحديد درجة قطع الاختبار الناتجة عن تطبيق أربعة من طرق تحديد درجة القطع اثنتين تحكيميّتان وهما: طريقتي أنجوف ونيديلسكي واثنتين تعتمدان جزئياً على التحكيم وتسترشدان ببيانات تجريبية وهما طريقتي المجموعات المتضادة والمجموعات المحكّية.

في ضوء ما سبق تحدد مشكلة الدراسة في التساؤل الرئيس التالي:

هل تختلف درجة قطع الاختبار الناتجة عن طرق تحديد درجة القطع المختلفة نيدلسكاي - أنجوف - المجموعات المتضادة - المجموعات المحكّية عند تطبيقها على اختبار موحد وعينة دراسية واحدة باختلاف نوع المحكّم وطول الاختبار؟

ويتفرع من التساؤل الرئيس ما يلي:

١- هل تختلف درجة القطع في الاختبار محكي المرجع عند تطبيق طريقة نيدلسكاي التحكيمية على اختبار موحد وعينة دراسية واحدة باختلاف نوع المحكّم؟

٢- هل تختلف درجة القطع في الاختبار محكي المرجع عند تطبيق طريقة أنجوف التحكيمية على اختبار موحد وعينة دراسية واحدة باختلاف نوع المحكّم؟

٣- هل تختلف درجة القطع في الاختبار محكي المرجع عند تطبيق طريقة نيدلسكي التحكيمية على اختبار موحد وعينة دراسية واحدة باختلاف طول الاختبار؟

٤- هل تختلف درجة القطع في الاختبار محكي المرجع عند تطبيق طريقة أنجوف التحكيمية على اختبار موحد وعينة دراسية واحدة باختلاف طول الاختبار؟

٥- هل تختلف درجة القطع في الاختبار محكي المرجع عند تطبيق طريقة (المجموعات المتضادة) على اختبار موحد وعينة دراسية واحدة باختلاف طول الاختبار؟

٦- هل تختلف درجة القطع في الاختبار محكي المرجع عند تطبيق طريقة (المجموعات المحكية) على اختبار موحد وعينة دراسية واحدة باختلاف طول الاختبار؟

فرضيات الدراسة:

- ١- لا توجد فروق دالة إحصائية عند مستوى دلالة ($\alpha = 0,05$) بين درجات القطع الناتجة عن تطبيق طريقة نيدلسكاي على اختبار موحد وعينة دراسية واحدة باختلاف نوع المحكم.
- ٢- لا توجد فروق دالة إحصائية عند مستوى دلالة ($\alpha = 0,05$) بين درجات القطع الناتجة عن تطبيق طريقة أنجوف على اختبار موحد وعينة دراسية واحدة باختلاف نوع المحكم.
- ٣- لا توجد فروق دالة إحصائية عند مستوى دلالة ($\alpha = 0,05$) بين درجات القطع الناتجة عن تطبيق طريقة نيدلسكي على اختبار موحد وعينة دراسية واحدة باختلاف طول الاختبار.
- ٤- لا توجد فروق دالة إحصائية عند مستوى دلالة ($\alpha = 0,05$) بين درجات القطع الناتجة عن تطبيق طريقة أنجوف على اختبار موحد وعينة دراسية واحدة باختلاف طول الاختبار.

أهمية الدراسة:

تظهر أهمية الدراسة الحالية فيما يلي:

- ١- تتناول هذه الدراسة بالبحث أحد اتجاهات القياس والتقويم التربوي المعاصرة التي كثر فيها النقاش والجدل بين علماء القياس والتقويم التربوي والنفسي، والتي تحتاج إلى مزيد من العناية والاهتمام والبحث، وذلك نظراً لحدائثة هذا النوع من القياس مقارنة مع الأنواع الأخرى من المقاييس التربوية والنفسية.
- ٢- تأتي أهمية الدراسة من الحاجة الماسة لتبصير المعلم بأهمية تحديد درجة القطع في الاختبارات محكية المرجع، لأنها يعول عليها الحكم على ما يستطيع وما لا يستطيع أن يتقنه الطالب، ومن ثم اتخاذ القرارات التربوية المناسبة حول تمكين الطالب المتقن من الوحدات الدراسية اللاحقة، وإعادة تأهيل الطالب غير المتقن للمهارات المراد قياسها في الاختبار.
- ٣- تنبع أهمية الدراسة من أهمية تحديد درجة القطع في تصنيف الطلاب إلى فئتين متقنة وغير متقنة للمهارات المطلوبة، يليه تشخيص تلك المهارات وتحديد نقاط الضعف والقوة، ومن ثم وضع الخطط العلاجية المناسبة.
- ٤- تسهم هذه الدراسة في توضيح طرق تحديد درجة القطع في الاختبارات محكية المرجع، بهدف مساعدة المعلم على اختيار الطريقة المناسبة لأداته، وذلك لتعدد هذه الطرق واختلاف افتراضاتها وأهدافها ونتائجها، وحيث أن صدق القرارات التربوية يعتمد على الاختيار

والتطبيق المناسبين لطريقة تحديد درجة القطع، لذا فإنه يمكن القول بأن دراسة طرق تحديد درجة القطع أمر في غاية الأهمية وذلك لأنها تسهم في إيضاح العلاقة بين اختيار الطريقة والهدف من تطبيق الاختبار، بالإضافة إلى أن التعرف على هذه الطرق سيعطي المعلم مؤشرات حول نقاط الضعف والقوة لهذه الطرق، الأمر الذي يساعد على إزالة الحيرة عند اختيار طريقة تحديد درجة القطع.

٥- قد تكون أول دراسة تطبق على مدارس المرحلة الابتدائية بالمملكة العربية السعودية - على حد علم الباحث - والتي تهتم بالتعرف على أثر كل من نوع المحكم وطول الاختبار على طرق تحديد درجة القطع في الاختبارات محكية المرجع.

أهداف الدراسة:

معرفة أثر متغيري الدراسة (نوع المحكم - طول الاختبار) على عملية تحديد درجة القطع وفق الطرق المستخدمة في الدراسة وهي: نيدلسكاي - أنجوف - المجموعات المتضادة - المجموعات المحكية عند تطبيقها على اختبار موحد وعينة دراسية واحدة.

مصطلحات الدراسة:

١- الاختبار محكي المرجع Criterion referenced Test

عرّف بابام (Popham, 1978 P. 93) الاختبار محكي المرجع بأنه ذلك "الاختبار الذي يستخدم لتقدير أداء الفرد في نطاق سلوكي محدد تحديداً دقيقاً". كما عرفه كل من جلاسر ونيكو Glaser&Nitko بأنه ذلك "الاختبار الذي يُعد بهدف إحداث قياسات يمكن تفسيرها بشكل مباشر بناءً على معايير (مستويات) أداء محددة" (Gray,1978,P.226).

أما التعريف الإجرائي للاختبار المحكي المرجع فهو اختبار الكفايات الرياضية في العمليات الحسابية على الأعداد بالمرحلة الابتدائية بمدارس مكة المكرمة الذي أعده بشناق (١٤١٢هـ) حيث اتبع في بناءه الأساليب والطرق المتعارف عليها في بناء هذا النوع من الاختبارات، ويتكون الاختبار في مجمله من ٩٥ مفردة، ويحتوي على ثلاثة اختبارات فرعية يقيس الاختبار الأول كفاية جمع الأعداد و طرحها ويتكون من ٣٣ مفردة، و يقيس الاختبار الثاني كفاية ضرب الأعداد وقسمتها ويتكون من ٤٠ مفردة، و يقيس الاختبار الثالث كفاية خصائص العمليات الحسابية ويتكون من ٢٢ مفردة.

٢ - درجة القطع Cut off Score

عرّف هامبلتون (Hambleton, 1978, P.279) درجة القطع بأنها "نقطة على متصل درجات الاختبار تستخدم لتصنيف الطلاب إلى فئتين تعكس مستويات الأداء المختلفة بالنسبة لهدف معين (أو الأهداف) المراد قياسها في الاختبار".

كما عرّف بيرك Berk درجة القطع المثلّي بأنها "تلك الدرجة التي تزيد من نسبة التصنيفات الصحيحة أو في المقابل تقلل من نسبة التصنيفات الخاطئة" (Hambleton et al , 1978 , P. 31) •
كما عرّفها بابام Popham "بأنها مقياس لمدى ملائمة أداء الطالب لهدف محدد" (Halpin et al, 1983, P. 185) •

أما التعريف الإجرائي لدرجة القطع في هذه الدراسة فهو عبارة عن الدرجات الناتجة من تطبيق الطرق المختلفة التي استخدمت في الدراسة وهي:

أ - طريقة نيدلسكاي Nedlesky's - Method

أشار نيدلسكي (Nedelsky 1954) إلى أن هذه الطريقة تستخدم مع الاختبار من نوع الاختيار من متعدد، وتتطلب من كل محكم فحص كل مفردة من مفردات الاختبار ثم يحدد من البدائل الخاطئة لكل مفردة تلك التي من الممكن أن يتجنب اختيارها الطالب ذو الحد الأدنى من الكفاية، ويكون الحد الأدنى لاحتمال الإجابة الصحيحة عن المفردة هو مقلوب عدد البدائل الباقية.

ب - طريقة أنجوف Angoff's Method (1976)

ذكر أنجوف (Angof 1971) بأن هذه الطريقة تستخدم أيضاً مع الاختبار من نوع الاختيار من متعدد وتتطلب هذه الطريقة من كل محكم التنبؤ بمستوى صعوبة كل مفردة في الاختبار وذلك عن طريق فحص كل محكم لمفردات الاختبار، ثم يطلب منهم تصور مجموعة من الطلاب من ذوي الحد الأدنى من الكفاية، ثم يقدر احتمال أن يعرف الطالب ذو الحد الأدنى من الكفاية المناسبة إجابة المفردة بصورة صحيحة دون أن يلجأ إلى التخمين، ثم يضع دائرة حول الرقم الذي يمثل مقدار الاحتمال المبين في ميزان فتري، ثم تجمع القيم الاحتمالية المقدرّة لكل مفردة، ويمثّل متوسط مجموع هذه الاحتمالات درجة القطع.

ج - طريقة المجموعات المتضادة Contrasting Groups

أشار هامبلتون (Hambleton 1982) بأنه يتعيّن على المحكمين في هذه الطريقة تحديد مجموعتين من الطلاب الأولى متقنة والثانية غير متقنة للمهارات التي يقيسها الاختبار، وبعد تطبيق الاختبار يتم رسم منحنيين يمثلان التوزيع التكراري لدرجات الطلاب، وتمثّل نقطة تقاطع

المنحنيين درجة قطع الاختبار، ويمكن تعديل الدرجة للتقليل من خطأ التصنيف الأول (α) وخطأ التصنيف الثاني (β).

د - طريقة المجموعات المحكية Criterion Groups

ذكر هامبلتون (1982) Hambleton بأن هدف هذه الطريقة هو تجنب أخطاء التصنيفات الخاطئة وزيادة التصنيفات الصحيحة، وتشابه هذه الطريقة إلى حد ما طريقة المجموعات المتضادة إلا أنها تتطلب إيجاد التصنيفات الثنائية للاحتتمالات التالية (طلاب متقنين فعلاً - طلاب غير متقنين فعلاً - طلاب متقنين غير حقيقيين - طلاب غير متقنين غير حقيقيين) يلي ذلك تحديد معامل صدق درجات القطع المحتملة، ثم حساب المنفعة والضرر النسبيين لأخطاء التصنيف.

٣ - طول الاختبار Test Length

يقصد بطول الاختبار عدد المفردات المكونة للاختبارات الفرعية التي قام الباحث بتقسيمها في هذه الدراسة إلى ثلاثة أنواع كالتالي:
أ- قصير (يتكون من ١٤ فقرة).
ب- متوسط (يتكون من ٢١ فقرة).
ج- طويل (يتكون من ٣٥ فقرة).

٤ - نوع المحكّم

يقصد بنوع المحكّم في هذه الدراسة نوع العمل الذي يمارسه المحكّم المشارك في تحديد درجة قطع الاختبار (معلم - مشرف تربوي - عضو هيئة تدريس).

حدود الدراسة:

الحدود الزمانية

يقتصر البحث على طلاب الصف السادس الابتدائي في بعض مدارس مدينة جدة الحكومية خلال الفصل الدراسي الثاني للعام الدراسي ١٤١٦-١٤١٧هـ.

الحدود المكاتبة

يقتصر البحث على طلاب الصف السادس الابتدائي في بعض مدارس مدينة جدة الحكومية وعلى أعضاء هيئة التدريس والمشرفين التربويين والمعلمين العاملين في مدينتي جدة ومكة المكرمة.

الفصل الثاني

الإطار النظري.
الدراسات السابقة.

الإطار النظري

مفهوم الاختبار التحصيلي Achievement Test

عرّف أبو حطب (١٩٨٧م) "الاختبار التحصيلي بأنه الأداة التي تستخدم في قياس المعرفة والفهم والمهارة في مادة دراسية أو تدريبية معينة أو مجموعة من المواد" ص٣٨٧. وأشار حفني، وآخرون (١٩٨٧م) "بأن الاختبارات التحصيلية هي تلك الاختبارات التي تهدف إلى قياس التعلم الماضي للفرد أو الخبرة السابقة" ص١٩. وذكر الطائي (١٤٠٣هـ) بأن الاختبار التحصيلي "عبارة عن قياس لأثر الخبرة الماضية والتنبؤ بالإنجاز المقبل" ص٥٩.

استخدامات الاختبارات التحصيلية:

ذكر كل من (حفني، وآخرون، ١٩٨٧م ؛ أمريزيق، ١٩٩٢م) بأن الاختبارات التحصيلية تهدف إلى مايلي:

- ١ – التأكد من توافر الحد الأدنى اللازم من مستويات الأداء بالنسبة لعمل أو مهنة معينة.
 - ٢ – التصنيف المهني والتعليمي والعسكري
 - ٣ – مقارنة أداء الطالب بأداء أقرانه.
 - ٤ – تحفيز دافعية التلاميذ على التحصيل.
- ولكي تحقق الاختبارات التحصيلية الأهداف المنشودة يجب أن يكون الاختبار معداً بطريقة جيدة بحيث يغطي عينة ممثلة من النواتج التعليمية والمادة الدراسية وأن يصاغ بطريقة تتناسب مع الغايات التي ستستخدم النتائج من أجلها وأن يتمتع بقدر كبير من الصدق والثبات وأن يتوخى الحذر في تفسير نتائجه وذلك بسبب أخطاء القياس وأن يستخدم لتحسين مستوى تعلم الطلاب وأن يحتوي على أسئلة ذات نوعية ملائمة لقياس الناتج التعليمي المطلوب.

تصنيفات اختبارات التحصيل:

يوجد العديد من تصنيفات الاختبارات التحصيلية وذلك حسب نوع الاستجابة وطريقة التطبيق وصياغة الأسئلة وتفسير الدرجات الناتجة، ولعل أهم تصنيف للاختبارات التحصيلية من حيث طريقة تفسير الدرجات الناتجة من تطبيق الاختبار كالتالي:

١: الاختبارات معيارية المرجع NRT

نكر بابام وهيوسك (1969) Pohpam & Husek بأن هذه الاختبارات تعتمد عند تفسيرها للدرجات المستحصلة على خصائص المجموعة التي طبقت عليها حيث يجب أن تكون هذه الخصائص مماثلة مع خصائص الطلاب من حيث السن والنوع والمنطقة الجغرافية والمرحلة الدراسية، وتهتم هذه الاختبارات بالكشف عن الفروق الفردية بين الطلاب وذلك من خلال مقارنة أداء الطالب بأداء أقرانه كجماعة معيارية بمعنى أن هذه الاختبارات تركز على وضع الطالب النسبي في علاقته بالآخرين.

٢ – الاختبارات محكمة المرجع CRT

أشار بابام وهيوسك (1969) Pohpam & Husek بأن هذه الاختبارات لا تعتمد عند تفسيرها لدرجة الطالب على خصائص مجموعته التي طبق عليها الاختبار، ولكنها تعتمد على مقارنة أداء الطالب بمحك أو مستويات أداء متوقعة ومحددة مسبقاً ويتم تحديد هذه المستويات في ضوء الأهداف السلوكية (الإجرائية) المراد قياسها وتقيس هذه الاختبارات مدى تحقيق الطالب لهذه الأهداف السلوكية وبذلك يمكن التحقق من كفاءة كل طالب من اكتسابه للمهارات والمعارف المرجوة ومن ثم تشخيص نواحي الضعف والقوة.

أوجه الاختلاف بين الاختبارات معيارية المرجع ومحكية المرجع

أشار كثير من علماء القياس التربوي مثل جليزر، وإيبيل، وبابام وهيوسك، وهامبلتون، وآخرون (Hambleton et al, Papham&Husek, 1969; Ebel, 1979; Glaser, 1994) إلى وجود عدد من الاختلافات بين الاختبارات معيارية المرجع ومحكية المرجع وذلك في عدة نواحي منها مفهومها وبناءها وأهدافها وخصائصها والتي يمكن أن تقدم على النحو التالي:

١ – الهدف الرئيس للاختبار:

يهدف الاختبار محكي المرجع إلى مقارنة أداء الطالب بمحك محدد مسبقاً وذلك لمعرفة الأهداف التي حققها والتي لم يحققها الطالب، أما الاختبار معياري المرجع فيهدف إلى مقارنة أداء الطالب بأداء مجموعته المعيارية من خلال تحديد عدد الأسئلة التي أجاب عليها الطالب بشكل صحيح.

٢- استخدام الاختبار:

يستخدم الاختبار محكي المرجع للتأكد من تحقيق الطالب للأهداف السلوكية المحددة، أما الاختبار معياري المرجع فيعتمد على تزويدنا بمؤشرات عن درجة نجاح الطالب، وبالتالي إظهار الفروق الفردية بين الطلاب.

٣- خصائص أسئلة الاختبار:

تتجمع الأسئلة في الاختبار محكي المرجع حول عدد محدد من الأهداف، بينما تنتشر الأسئلة في الاختبار معياري المرجع بشكل واسع حول نطاق الأهداف.

٤- تفسير الأداء:

تفسر الدرجة في الاختبار محكي المرجع بناء على درجة القطع، بينما تفسر درجة الطالب في الاختبار معياري المرجع بناء على درجات معيارية تمثل موقع الطالب بين مجموعته الصفية والرتب المئينية.

٥ - بناء الفقرات:

يعتمد بناء الفقرات في الاختبار محكي المرجع على مدى تحقيق المفردة للهدف السلوكي التعليمي، بينما يعتمد بناءها في الاختبار معياري المرجع على تباين الدرجات ومعامل التمييز، ويتجنب اختيار المفردات الصعبة جداً والسهلة جداً.

٦ - التقويم:

يسعى الاختبار محكي المرجع إلى تحديد جوانب الضعف والقوة في كل من أداء الطلاب والبرامج التعليمية، بينما يسعى الاختبار معياري المرجع إلى تقويم أداء الطالب ومقارنته مع أداء الطلاب الآخرين.

تعريف الاختبارات محكية المرجع Definition of a CRT

أشار عدد من علماء القياس التربوي مثل بيرك، وجايجر، وهامبلتون ونوفيك (Berk, 1982; Hambleton & Novick, 1973; Jaeger, 1989) إلى أن تعريف الاختبارات محكية المرجع يتعدد وفقاً لاختلاف آراء وخبرات علماء القياس التربوي حيث يوجد في أدبيات القياس التربوي أكثر من خمسين تعريفاً تصف الاختبارات محكية المرجع، وتتمحور تلك التعريفات حول النقاط التالية:

- ١ - تعريف النطاق السلوكي Content - domain الذي يقيسه الاختبار تعريفاً دقيقاً وذلك مما يساعد على اختيار مفردات تمثل هذا النطاق.
- ٢ - استخدام نتائج الاختبارات محكية المرجع لاتخاذ القرارات الخاصة بتحديد مستوى كفاءة الطالب مما يساعد على تشخيص نواحي ضعفه وقوته
- ٣ - مقارنة أداء الطالب بمستويات أداء محددة مسبقاً (أهداف سلوكية محددة) يتخذ في ضوءها قرارات تربوية تتعلق بمدى إتقان الطالب أو عدم إتقانه للمهارات المراد قياسها.

تصنيفات الاختبارات محكية المرجع Categories of CRT

تصنف الاختبارات محكية المرجع على حسب نطاقها السلوكي إلى قسمين وذلك ما أشار به نيتكو (1980) Nitko هما:

١: اختبارات محكية المرجع تعتمد على نطاقات مرتبة Ordered Domains

- يعتمد هذا النوع على محكات تساعد على ترتيب النطاقات السلوكية وهذه المحكات عبارة عن:
- أ- تحكيم المعايير الاجتماعية أو نوعية الأداء (Judged Social or esthetic quality of Performance)
 - ب- مستوى صعوبة أو تعقد المقرر الدراسي (Complexity or difficulty level of subject matter)
 - ج- درجة الكفاءة المطلوبة للمهارات المركبة (Degree of proficiency with which complex skills are performed)
 - د- متطلبات اكتساب المهارات العقلية والحركية (Prerequisite sequence for acquiring intellectual and psychomotor skills)
 - هـ - موقع السمات الكامنة المعرفة (Location on empirically defined latent trait)

٢-: اختبارات محكية المرجع تعتمد على نطاقات غير مرتبة:

- كما أشار نيتكو (1980) Nitko إلى أن هذا النوع يتطلب تعريف النطاقات السلوكية ويحتوي على أربعة أنواع من النطاقات يمكن إيجازها كالتالي:
- أ - خصائص مثيرات النطاق وخطة اختيار مفردات الاختبار (Stimulus Properties of Domain and Sampling Plan of Test)
 - ب- نطاقات تهتم بالخصائص اللفظية للمثيرات والاستجابات (Verbal Statements of Stimuli and responses in domain)

ج- نطاقات تعتمد على تشخيص تصنيفات الأداء (Diagnostic Categories of Performance)

د- نطاقات تعتمد على التجريد والسمات أو التكوينات الفرضية , Traits , (Abstractions , or Constructs).

طرق تقدير صدق الاختبارات محكية المرجع:

١ - الصدق الوصفي Descriptive - Validity

ذكر بابام (1978) Popham بأن الصدق الوصفي عبارة عن محاولة لتأكيد مدى ما يقيسه فعلاً الاختبار محكي المرجع، ويعتبر تحديد الصدق الوصفي شرطاً أساسياً لأنواع الصدق الأخرى وذلك لأنه أفضل وسيلة لتفسير أداء الطالب في الاختبار. ويمكن تحديد الصدق الوصفي باستخدام أساليب مختلفة، ومن ذلك الاستعانة بمجموعة من المتخصصين لتحديد مدى مطابقتها (Congruent) مفردات الاختبار مع المجال السلوكي المراد قياسه.

٢ - الصدق الوظيفي Functional - Validity

أشار بابام (1978) Popham بأن المقصود بالصدق الوظيفي هو دقة الاختبار محكي المرجع في تحقيق الغرض (الوظيفة) الذي صمم من أجله الاختبار، أي أنه لا يقتصر على وصف أداء الطالب في الاختبار، ويستخدم الصدق الوظيفي عندما يكون الهدف من تطبيق الاختبار وصف وتحديد الوظيفة المراد تحقيقها من الاختبار، لذا فإن الصدق الوصفي يعتبر شرطاً لازماً لحساب الصدق الوظيفي. ويقابل الصدق الوظيفي في القياس المعياري المرجع صدق المحك الخارجي (Criterion Related Validity) والذي يدل على قدرة الاختبار على التنبؤ بأداء طالب ما في موقف تالي من أداء الطالب على اختبار سبق التأكد من صدقه، وأطلق على هذا النوع من الصدق (بالصدق الوظيفي) بدلاً من صدق المحك الخارجي لأن بعض الوظائف المراد قياسها في الاختبار محكي المرجع تتطلب محكاً خارجياً والبعض الآخر لا يتطلب ذلك.

وأضاف هامبلتون (Hambleton) بأنه يمكن تحديد صدق القرار (Decision Validity) بمقارنة أداء مجموعتين محكيتين أو أكثر بدرجة قطع محددة. كما يمكن تحديد صدق القرارات التعليمية (Instructional Decision) المبنية على درجات الاختبارات محكية المرجع بمقارنة أداء مجموعتين (الأولى تلقت التعليم والثانية لم تتلق التعليم) بدرجة قطع الاختبار. ويمكن تقييم صدق القرار لكل هدف تعليمي بجمع النسب المئوية للطلاب المتقنين الذين اجتازوا درجة القطع والنسب المئوية للطلاب غير المتقنين الذين لم يجتازوا درجة القطع، وذلك بهدف تحديد الطلاب

المتقنين فعلاً والطلاب غير المتقنين فعلاً. ويمكن تحديد صدق القرار أيضاً بحساب معامل الارتباط بين متغيرين ثنائيين الأول متغير الانتماء إلى المجموعة (Group Membership) والثاني متغير قرار الإتقان (Mastery Decision) (Berk, 1982).

٣ - صدق انتقاء النطاق السلوكي Domain Selection Validity

أشار بابام (Popham 1978) بأن صدق انتقاء النطاق يتعلق بمدى دقة باني الاختبار في اختيار نطاق سلوكي محدد يمكن أن يستخدم كمؤشر (Indicator) لإيضاح مستوى الطالب بالنسبة إلى (نطاق عام) يعد أكثر عمومية. ويمكن تحديد صدق انتقاء النطاق السلوكي بأساليب متعددة منها مايلي:

أ- تدريس مجموعات صغيرة من الأفراد المهارات التي يشتمل عليها كل نطاق من النطاقات السلوكية حتى يصلوا إلى درجة الإتقان، يلي ذلك قياس إمكانية تعميم إتقان الطلاب لنطاق واحد تجاه النطاقات الأخرى.

ب- مقارنة أداء مجموعتين من الأفراد إحداهما تلقت التعليم (النطاقات المراد قياسها) والأخرى لم تتلق التعليم، وذلك بتطبيق عدد من الاختبارات، ثم يحدد النطاق السلوكي الذي يميز (Discriminates) بصورة أفضل بين المجموعتين (Popham, 1987).

طرق تقدير ثبات الاختبارات محكية المرجع:

تختلف وتتعدد طرق تقدير ثبات الاختبارات محكية المرجع باختلاف استخداماتها، وصنّف هامبلتون، وآخرون (Hambleton, et al (1978) تلك الطرق إلى ثلاث مجموعات رئيسية، وستطرق الدراسة الحالية لمجموعتين هما: تقدير ثبات درجات الطلاب في النطاق السلوكي وتقدير ثبات تصنيف الطلاب في مجموعات على حسب درجة تمكنهم من النطاق السلوكي.

أولاً - تقدير ثبات درجات الطلاب في النطاق السلوكي:

طرق هذه المجموعة تهتم بتقليل خطأ التباين الناتج عن الفروق بين درجات الطلاب في النطاق السلوكي بالإضافة إلى درجات النطاق، ومن تلك الطرق مايلي:-

معامل ليفنجستون Livingston-Index

ذكر كل من كروكر، والجائنا (1986) Crocker & Algina بأن هذا المعامل يهتم بإيجاد انحراف درجات الأفراد عن درجة القطع، ويهتم هذا المعامل أيضاً بمفهوم متوسط مجموع مربعات انحرافات الدرجات الملاحظة المتوقعة وذلك بحساب قيمة انحراف درجات كل طالب عن درجة القطع في الاختبار، ويتم ذلك وفق الصيغة الرياضية التالية:

$$K^2(X, T) = \frac{\sigma^2 T + (M_T - n_i C)^2}{\sigma^2 X + (M_x - n_i C)^2}$$

حيث إن:

$K^2(X, T)$ ترمز إلى معامل ليفنجستون

$\sigma^2 T$ ترمز إلى متوسط مجموع مربعات انحرافات الدرجات الحقيقية عن درجة القطع (C).

$\sigma^2 X$ ترمز إلى متوسط مربعات انحرافات الدرجات الحقيقية عن درجة القطع.

M ترمز إلى متوسط درجات الطلاب.

N_i ترمز إلى عدد الأسئلة.

C ترمز إلى درجة القطع.

ولتقدير معامل ليفنجستون $K^2(X, T)$ في حالة استخدام اختبارين لكل طالب تستخدم الصيغة الرياضية التالية:

$$\hat{K}^2(x, T) = \frac{\hat{P}_{xx'} \hat{\sigma}_x \hat{\sigma}_{x'} + (\hat{M}_x - n_i C)(\hat{M}_{x'} - n_i C)}{\sqrt{[\hat{\sigma}_X^2 + (\hat{M}_x - n_i C)^2] [(\hat{\sigma}_{X'}^2 + (\hat{M}_{x'} - n_i C)^2]}}$$

حيث إن:

$\hat{P}_{xx'}$ ترمز إلى معامل الارتباط بين درجات كلا الاختبارين.

$\sigma^2 X$ ترمز إلى متوسط مربعات انحرافات الدرجات الحقيقية عن درجة القطع.

\hat{M} ترمز إلى متوسط الدرجات الخام.

N_i ترمز إلى عدد الأسئلة.

C ترمز إلى درجة القطع.

أما في حالة استخدام اختبار واحد تستخدم الصيغة التالية:

$$\hat{K}^2(x, T) = \frac{\sigma^2 x(KR20) + (M_x - n_i C)^2}{\sigma^2 x + (M - n_i C)^2}$$

حيث ترمز (KR20) إلى معامل ارتباط كودرريتشارسون (٢٠).

ثانياً - تقدير ثبات تصنيف الطلاب في مجموعات على حسب درجة تمكنهم من النطاق السلوكي:

تهتم طرق هذه المجموعة بتحديد الأخطاء الناتجة عن عدم اتساق التصنيف عند تطبيق اختبارين متوازيين أو في حالة إعادة تطبيق الاختبار على أفراد العينة، ومن تلك الطرق مايلي:-

١: طرق تتطلب تطبيق الاختبار مرة واحدة:

أ- طريقة هاريس Harris Method

أشار هاريس (1974) Harris بأن هذه الطريقة لا تعتمد في تحديدها على متغير طول الاختبار (k) وإنما تقوم على الارتباط بين متغير يمثل مجموع درجة الطالب في الاختبار ومتغير ثنائي آخر يمثل تصنيف الطالب إلى متقن أو غير متقن على ضوء درجة القطع، لذا فإنه يتم تصنيف الطلاب إلى أربع مجموعات في (جدول رباعي) وهي طلاب متقنين فعلاً وطلاب غير متقنين فعلاً ومجموعة طلاب الخطأ الأول (α) ومجموعة طلاب الخطأ الثاني (β) ويتم تحديدهم بناءً على درجة قطع الاختبار.

ويطلق على هذه الطريقة بمعامل (هاريس) ويرمز له بالرمز (Mc) ويتم تقديره بالصيغة التالية:-
حيث إن:

$$MC = \frac{ssb}{ssb + ssW}$$

MC = معامل ثبات هاريس.

ssb، ssW = عبارة عن مجموع المربعات (داخل، بين) المجموعات.
وتمتد قيمة المعامل ما بين (صفر وواحد).

ب- معامل (كابا) لهائنا Huynh Kappa Coefficient

يمكن تقدير معامل (هاينا) في حال تطبيق اختبار واحد أو اختبارين متوازيين، ونظراً لصعوبة تقدير هذا المعامل في حال تطبيق اختبارين متوازيين، لذا سيتم التطرق لكيفية تقدير معامل الثبات

(هاينا) لاختبار واحد فقط، وفي هذا الصدد أشارت سابكوفياك Subkoviak بأن تقدير هذا المعامل يستند إلى عدة فروض هي: -

\bar{y} أن يتمثل توزيع درجات الطلاب الحقيقية في الاختبار شكل توزيع بيتا (Beta Distribution).

\bar{y} إذا طبق عدد من الاختبارات على طالب ما، فإنه يفترض أن يتخذ توزيع درجاته في الاختبار شكل التوزيع ذي الحدين، وفي هذه الحالة يجب أن يكون تقدير درجات أسئلة الاختبار إما صفر أو واحد، وأن تكون الإجابة على الأسئلة مستقلة إحصائياً بحيث لا تؤثر الإجابة على إحدى المفردات على إجابة المفردة الأخرى، وأن تتساوى تقريباً معاملات صعوبة جميع أسئلة الاختبار.

ويتم تقدير معامل ثبات (هاينا) وفق إحدى الصيغتين التاليتين:

$$\hat{K} = \frac{\hat{P}_{zz} - \hat{P}_z^2}{\hat{P}_z - \hat{P}_z^2}$$

$$\hat{P}_0 = 1 - 2(\hat{P}_z - \hat{P}_{zz})$$

ويتطلب تحقيق الفروض السابقة إلى إجراء الخطوات التالية: -

\bar{y} تقدير متوسط درجات الطلاب (\hat{M}) والانحراف المعياري ($\hat{\sigma}^2$) ومعامل ثبات

كودريتشاردسون ٢١ ($\alpha 21$) وتحديد درجة القطع (C) وبارمتر توزيع ألفا (α).

\bar{y} تحويل القيم السابقة إلى بيانات معيارية (**normalize data**) وفق الصيغ الآتية: -
حيث إن:

$$\hat{M}' = \text{Sin}^{-1} \sqrt{\hat{M}/n} \quad \hat{M} \text{ ترمز إلى المتوسط.}$$

n ترمز إلى عدد المفردات.

$$\hat{\sigma}' = \sqrt{(\hat{\alpha}21+1)/(\hat{\alpha}+n)} \quad \hat{\sigma}' \text{ ترمز إلى الانحراف المعياري.}$$

$$\hat{P}' = \hat{\alpha}21 \sqrt{(n-1)/(n+\hat{\alpha}21)} \quad \hat{P}' \text{ ترمز إلى معامل الثبات.}$$

$$\hat{C}' = \text{Sin}^{-1} \sqrt{(C-5)n} \quad \hat{C}' \text{ ترمز إلى درجة القطع.}$$

\bar{y} تقدير الانحراف المعياري (normal deviate) بناءً على درجة القطع (Z).

$$Z = (C' - \hat{M}) / \hat{\sigma}'$$

\bar{y} تقدير قيمة الاحتمال ($\hat{P}z$) من الجداول الخاصة بذلك بحيث تكون هذه القيمة أقل من قيمة (Z).

\bar{y} ثم تحديد قيمة الاحتمال ($\hat{P}z$) من جداول دالة التوزيع الثنائي الاعتدالي بحيث تكون قيمة الاحتمال أيضاً أقل من قيمة (Z). (Berk, 1982).

٢: طرق تتطلب تطبيق الاختبار مرتين:

أ- معامل كارفر Carver Method

أشارت سابكوفياك Subkoviak إلى أن معامل كارفر يعد من الطرق الأولى المقترحة لتقدير ثبات الاختبارات محكية المرجع التي تتعلق باتساق قرار التصنيف، وتقوم هذه الطريقة على تطبيق اختبارين متوازيين على مجموعة واحدة من الأفراد، يليه مقارنة نسبة عدد الطلاب الذين تم تصنيفهم ضمن المتقنين في كلا الاختبارين فإذا كانت النسبتين متماثلتين أو قريبتين من بعضهما فبذلك يعتبر الاختبارين ثابتين، ويعاب على هذه الطريقة بأنها غير حساسة لاتساق قرارات كل طالب في الاختبارين.

ويمكن إيجاد قيمة معامل (كارفر) وفق الجدول التالي:-

الاختبار (ب)

	متقن	غير متقن	
متقن	أ	ب	
غير متقن	د	ج	

(١)

ويحسب معامل كارفر باستخدام الصيغة التالية:-

$$\frac{أ + ج}{ن} =$$

حيث إن:

أ = مجموع الطلاب المتقنين بالفعل. ج = مجموع الطلاب غير المتقنين بالفعل

ن = أ + ب + ج + د (Berk, 1982).

ب- معامل (كابا) لسواميناثان وهامبلتون وألجاينا - Swaminathan-

Hambleton-Algina

ذكر كل من هامبلتون، وآخرون (Hambleton, et al (1978) بأن هذه الطريقة تعتبر امتداداً لطريقة (كارفر) السابقة حيث تعتمد على فرق النسبة بين الطلاب المتمكنين وغير المتمكنين في صورتى الاختبار المتوازيتين أو في مرتى تطبيق الاختبار، وتمتد قيمة معامل الثبات بين (+، 1) - (1-).

ويمكن تقدير معامل الثبات وفق الصيغة التالية: -

$$\hat{P}_o = \sum_{k=1}^m PKK,$$

حيث إن:

\hat{P}_o ترمز إلى معامل الثبات أي إتساق التصنيف.

PKK ترمز إلى نسبة الطلاب الذين تم تصنيفهم ضمن المتقنين في المستوى (K) في مرتى تطبيق الاختبار.

M ترمز إلى عدد مستويات الإتقان.

ونظراً لأن هذه الطريقة لا تأخذ بعين الاعتبار إتساق التصنيف الناتج عن عامل الصدفة، لذا أقترح كل من سواميناثان وهامبلتون وألجاينا استخدام معامل كابا (**Kappa Coefficient**) الذي ينسب إلى كوهن (Cohen, 1960).

$$\hat{K} = \frac{\hat{P}_o - \hat{P}_c}{1 - \hat{P}_c}$$

حيث إن:

K = معامل كابا.

\hat{P}_o = نسبة الإتفاق الملاحظ في التصنيفات وتحسب بالصيغة السابقة.

$$\hat{P}_o = \sum_{k=1}^m PKK,$$

\hat{P}_c = ترمز إلى نسبة الإتفاق المتوقع في التصنيفات.

وتحسب بالصيغة التالية: -

$$\hat{P}_o = \sum_{k=1}^m Pk.P.K.$$

حيث إن:

$(PK.) (P.K.) =$ نسب الطلاب المصنفين ضمن فئة المتقنين في المجموعة (K) في مرتي تطبيق الاختبار على التوالي.

تحديد طول الاختبار محكي المرجع:

ذكر كل من هامبلتون وإيقنور (1980) Hambleton & Eignor بأن هناك عاملين ينبغي وضعهما في الاعتبار عند تحديد أسئلة الاختبار وهما:

١ - العلاقة بين عدد أسئلة الاختبار والأهمية الخاصة بكفاية معينة:

تتفاوت أهمية الكفايات المراد قياسها في الاختبار، لذا ينبغي أن يختار مصمم الاختبار عند قياسه كفايات متعددة أسئلة تتناسب مع أهمية النطاق المراد قياسه.

٢ - العلاقة بين عدد أسئلة الاختبار ومستوى الحد الأدنى المقبول لدرجة ثبات الاختبار:

يستدعى ذلك تحديد قيمة الدقة المطلوبة (Degree of Precision) في تقدير درجة الطالب في النطاق، ويتم ذلك وفق الصيغة التالية: -

$$0,25$$

طول الاختبار = _____

(درجة الدقة المطلوبة)^٢

حيث (0,25) مقدار ثابت.

وأشار كل من هامبلتون، وآخرون (1978) Hambleton et al بأن الهدف من تطبيق الاختبار محكي المرجع هو تصنيف الطلاب إلى فئتين متقنة وغير متقنة، لذا فإن تحديد طول الاختبار المناسب الذي يقلل من أخطاء التصنيف (**Number of Classification Errors**) يتطلب تطبيق اختبار طويل يقلل من أخطاء القياس.

وأضاف كل من هالادين ورويد **Haladyn & Roid** بأن قيمة أخطاء التصنيف تقل بشكل عام عند زيادة طول الاختبار لكن النقص في قيمة الخطأ يكون ملحوظاً عند الانتقال من اختبار طوله (١٠) فقرات إلى اختبار طوله (٢٠) فقرة، وأما النقص في قيمة الخطأ عند الانتقال من اختبار طوله (٢٠) فقرة إلى اختبار طوله (٣٠-٤٠) فقرة يكون صغيراً نسبياً لدرجة قد لا تبرر استخدام فقرات تزيد عن (٢٠) فقرة (الأحمد، ١٩٩٢م).

مسميات وتعريف درجة القطع :Definitions of cut off score

مسميات درجة القطع:

تتعدد مسميات درجة قطع الاختبارات بتعدد آراء وخبرات علماء القياس والتقويم التربوي في مجال الاختبارات محكية المرجع، ومن تلك المسميات ما ذكرها هالبن، وآخرون Halpin, et al (1983) وهي كما يلي:

- ١- درجة القطع Cut off Score.
- ٢- الحد الأدنى من الكفاية Minimum Competency.
- ٣- درجة المحك Criterion Score.
- ٤- درجة الاجتياز Passing Score.
- ٥- مستوى الحد الأدنى من الكفاية Level of Minimal Competency.

تعريف درجة القطع:

نظراً لأهمية مفهوم درجة القطع في القياس محكي المرجع، الأمر الذي أدى إلى تعدد تعريفات علماء القياس التربوي والنفسي لمفهوم درجة القطع، ومن تلك التعريفات ما أشار إليه هامبلتون (Hambleton, 1978) حيث عرّف درجة القطع بأنها "نقطة على متصل درجات الاختبار تستخدم لتصنيف الطلاب إلى فئتين تعكس مستويات الأداء المختلفة بالنسبة لهدف معين (أو الأهداف) المراد قياسها في الاختبار" (P. 279).

كما عرّف بيرك Berk درجة القطع بأنها "تلك الدرجة التي تزيد من نسبة التصنيفات الصحيحة أو في المقابل تقلل من نسبة التصنيفات الخاطئة" (Hambleton et al, 1978, p. 31).

كما عرّفها بابام Popham بأنها "مقياس لمدى ملائمة (أداء الطالب) لهدف محدد" (Halpin et al, 1983, P.185)

طرق تحديد درجة القطع:

تبنى خبراء القياس والتقويم التربوي العديد من الطرق والمحاولات لتحديد درجة القطع، واختلفت تلك الطرق والأساليب باختلاف خبرات ودرجات تأهيل وتخصصات هؤلاء الخبراء، فهناك عدة طرق متعددة لتحديدها والتي ذكر بيرك (1986) Berk أنها بلغت ثمان وثلاثين طريقة مختلفة من حيث أسلوب بناءها وتطبيقها وتفسير نتائجها، ويمكن تصنيفها إلى ثلاثة تصنيفات كالتالي:

أولاً: الطرق التحكيمية Judgmental - methods

تستند الطرق التحكيمية على آراء المتخصصين في مجال القياس أو في محتوى الاختبار المراد قياسه، وذلك لتحديد درجة القطع المناسبة، حيث يسترشد بأرائهم في تحديد مستوى الطالب ذو الحد الأدنى من الكفاية (تحديد عدد الأسئلة التي يجب أن يجيب عليها الطالب بصورة صحيحة) واضعين في اعتبارهم المتغيرات التي تؤثر في تحديد هذه الكفاية.

مميزات وعيوب الطرق التحكيمية:

لكل طريقة من الطرق التحكيمية مميزات وعيوب خاصة بها وهناك عيوب ومميزات تكاد تكون مشتركة بينها وهي كالتالي:

مميزات الطرق التحكيمية:

- كما استرسل بيرك (Berk) بأن من مميزات الطرق التحكيمية مايلي:
- 1- أنها سهلة الاستخدام مقارنة بالطرق الأخرى.
 - 2- أنها سهلة التطبيق والفهم.
 - 3- أنها سهلة التفسير.
 - 4- لا تحتاج إلى جهد ووقت كبيرين.
 - 5- أنها سهلة الحساب ولا تعتمد على أساليب إحصائية معقدة.

عيوب الطرق التحكيمية:

أشار كل من بيرك، وهامبلتون، وايجنور (Hambleton & Eignor, 1980; Berk, 1986) إلى عيوب استخدام الطرق التحكيمية وهي:

- 1- اختيارية (Arbitrary) وسميت كذلك لأن هذه الطرق تعتمد كلياً على آراء المحكمين وتلعب كفاية هؤلاء المحكمون دوراً بارزاً في الوصول إلى المعيار المناسب لتصنيف الطلاب وتحديد بنود الاختبار التي يمكن لأقل الطلاب كفاءة الإجابة عليها بشكل صحيح ويعتقد بأن كون هذه

الطرق اختيارية لا يعد سبباً كافياً لرفض هذه الطرق لأن هذه العملية الاختيارية ليست مطلقة ولكنها تعتمد على مجموعة من الأسس في تحديد مستويات الأداء.

٢- صعوبة تحديد الطلاب الأقل كفاءة.

٣- لا تعتمد على بيانات تجريبية فعلية لكي يسترشد بها المحكمون في تقدير أحكامهم.

٤- الحصول على درجات قطع مختلفة عند تطبيقها على اختبار موحد وعينة دراسية واحدة.

٥- ذاتية المحكمين في اختيار الحد الأدنى من الكفاية وذلك بسبب اختلاف تصوراتهم حول عملية الإيقان.

٦- تفاوت عملية الاتساق الداخلي في عملية التحكيم.

٧- صعوبة التنبؤ بأداء الطلاب في الاختبار.

على الرغم من وجود بعض العيوب في استخدام الطرق التحكيمية إلا أنه توجد بعض المعايير المطروحة في أدبيات القياس التربوي التي تحد من تأثير هذه العيوب حيث ذكر بيرك (1986) Berk بعض المعايير نذكر منها ما يلي:

أ - استخدام طرق لتوحيد بيانات الأداء الفعلي.

ب - يجب أن تكون عملية تحكيم مفردات الاختبار متزامنة مع تطبيق الاختبار.

ج - الاهتمام بتدريب المحكمين.

د - تنظيم عملية تحكيم بنود الاختبار.

هـ - إيجاد الحلول المناسبة لتسوية الاختلافات في أحكام المحكمين.

و- تصميم استمارات تنظم عملية التحكيم.

ز - تزويد المحكمين ببيانات مسبقة حول أداء الاختبار.

وتتاول كثير من الباحثين الطرق التحكيمية وأسهبوا في ذكر تفصيلاتها وتطبيقاتها ومميزاتها وعيوبها ومن ضمن هؤلاء الباحثين علام، وبيرك، وجايفر (Jaeger, 1989; Berk, 1986)؛ علام، (١٩٩٥) ومن هذه الطرق ما يلي:

١ - طريقة إيبيل Ebel's method:

اقترح إيبيل Ebel طريقة لتحديد درجة القطع تعتمد على تعليقات المحكمين على مفردات الاختبار وذلك من خلال اعتبار أن مواصفات المفردات لها بعدين هما بعد الصلة Relevance وبعد الصعوبة difficulty ويشتمل بعد الصعوبة على ثلاثة مستويات (سهلة، متوسطة، صعبة) ويشتمل بعد الصلة على أربعة مستويات (أساسية، مهمة، مقبولة، غير مهمة) وبذلك يستخدم المحكم أعداداً مختلفة من تصنيفات الصعوبة والصلة، ثم يطلب من كل محكم مايلي:

أ - تحديد الخلية المناسبة لكل مفردة من المفردات بناءً على توافق محتواها مع تصنيفات البعدين.

ب - تحديد النسبة المئوية لعدد الطلاب ممن لديهم الحد الأدنى من الكفاية الذين يجب أن يجيبوا عن كل مفردة بصورة صحيحة وذلك لكل خلية من الخلايا، ثم يضرب عدد مفردات كل خلية في النسبة المئوية، ثم يقسم مجموع نواتج جميع الخلايا على عدد المفردات، وبالتالي نحصل على درجة القطع (Jeager, 1989).

عيوبها: -

تناول بيرك (1986) Berk عيوب هذه الطريقة حيث ذكر بأن أيبل لم يلتزم بالموصفات التي تستخدم في كل من البعدين والتمييز بينهما، لكي يسترشد بها المحكمين في عملية التحكيم، لذلك ربما يعتمد كل محكم على مواصفات يحددها لنفسه مما يؤثر بلا شك في دقة التصنيف واتساقه، كما أن هذه الطريقة اعتمدت على تحكيم المفردة ككل ولم تأخذ بعين الاعتبار بدائل الإجابات التي تشتمل عليها مفردة الاختيار من متعدد.

بالإضافة إلى أن طريقة إيبل تحتاج تحكيم خاص يتمثل بنسب مئوية للمفردات وذلك لكل خلية عبر بعدي الصلة والصعوبة المستمرين، فإذا لم تكن هنالك معايير خارجية يبني عليها هذا الحكم فإنه يبدو أنه سيكون حكماً اعتباطياً.

للحد من هذه العيوب قام كل من سكاكون وكليبق Skakun & Kling بإجراء تعديلات على طريقة إيبل بحيث تعتمد هذه الطريقة على بعدين هما معامل الصعوبة ويشتمل على (سهل - متوسط الصعوبة - صعب جداً) ومعامل التصنيف ويشتمل على (واقعي - شامل - حل المشكلات) وتستخدم بيانات الأداء في البنود لتقدير تصنيفات الصعوبة ويحدد كلاً من باني الأسئلة ولجنة الاختبار الأسئلة طبقاً لمستويات التصنيف، بناءً على ذلك تضرب الأسئلة في الخلايا المكونة من المصفوفة 3×3 ثم يقوم المحكمين بالإشارة إلى نسبة المفردات في كل خلية والتي يجب أن يجيب عليها الطالب المتقن إجابة صحيحة (Berk, 1986).

المميزات:

لها نفس مميزات طريقة إيبل ويحدد تصنيف معامل صعوبة البند بالاستعانة ببيانات الأداء، وتضرب في كلا البعدين مما يسهل عمل المحكم.

العيوب:

إن عيوب هذه الطريقة تكمن في عدم وضع عاملي الصلة والأهمية في الاعتبار ومن المحتمل أيضاً أن يصعب على بعد التصنيف تحديد مناطق المحتوى المحددة (Berk, 1986).

للحد من هذه العيوب قام كل من سكاكون وكليبق Skakun & Kling بإجراء تعديلات على طريقة إيبل بحيث أصبحت هذه الطريقة تعتمد على بعدين أيضاً وهما معاملي الصلة والتصنيف وأصبح

معامل الصلة ذو أبعاد ثلاثة (مهمة جداً - مهمة - مقبولة) وهذه الأبعاد مقترنة ببعدها معامل التصنيف.

المميزات:

لها نفس مميزات طريقة إيبيل وتعتبر عملية ضرب كلا البعدين ميسرة لعمل المحكمين (Berk , 1986).

العيوب:

هذه الطريقة لا تضع في اعتبارها مستويات صعوبة البنود الواقعية، ومن المحتمل أن يصعب في بعد التصنيف تحديد مناطق المحتوى المحددة (Berk , 1986).

٢ - طريقة أنجوف Angoffs method (١٩٧١م)

يطلب من كل محكم في هذه الطريقة فحص كل فقرة من فقرات الاختبار، ثم يطلب منهم تصور مجموعة من الأفراد الذين حققوا الحد الأدنى للكفاية التي يقيسها الاختبار، ثم يقدر نسبة عدد الأفراد الذين يحتمل أن يجيبوا إجابة صحيحة عن كل مفردة من مفرداته، ويمثل متوسط هذه النسب الحد الأدنى لمستوى الاجتياز في الاختبار (Angoff, 1971).

مميزاتها

أ - سهولة التنفيذ. ب - سهولة الفهم. ج - أنها تعتمد على تصور عدد من الأفراد الذي حققوا الحد الأدنى للكفاية المطلوبة بدلاً من تصور فرد واحد.

عيوبها:

يصعب على كثير من المحكمين تعريف الطلاب من ذوي الحد الأدنى للكفاية، ويصعب أيضاً تقدير القيم الاحتمالية لبعض المفردات المصاغة في عبارات منفية ((Negatively أو المتضمنة عمليات حسابية (Computational) (Berk, 1986).

للحد من هذه العيوب أشار كل من جايفر وكروس وآخرون (Jaeger, 1989; Cross et al., 1984) بأن أنجوف Angoff أجرى تعديلات على طريقته وذلك لتسهيل عملية تقديرات قيم المحكمين الاحتمالية لأسئلة الاختبار حيث يطلب من كل محكم تقدير احتمال أن يجيب الطلاب ذوي الحد الأدنى للكفاية كل سؤال دون أن يلجئوا إلى تخمين (Guessing) الإجابة، وتسجل تقديراتهم الاحتمالية في استمارة خاصة تحتوي على مقياس عشري الفترات (Equal Interval Scale) (١ = صفر - ٠,١٠) (٢ = ١١ - ٠,٢٠) (٣ = ٢١ - ٠,٣٠) (٤ = ٣١ - ٠,٤٠) (٥ = ٤١ - ٠,٥٠) (٦ = ٥١ - ٠,٦٠) (٧ = ٦١ - ٠,٧٠) (٨ = ٧١ - ٠,٨٠) (٩ = ٨١ - ٠,٩٠) (١٠ = ٩١ - ١,٠) ثم يطلب من كل محكم اختيار القيمة الاحتمالية التي تمثل تقديره الخاص بكل سؤال

من أسئلة الاختبار، ثم تجمع القيمة الاحتمالية التي قدرها كل محكم لكل مفردة، ثم نوجد متوسطها لتمثل درجة القطع.

٣ - طريقة ندلسكاي Nedlesky's Method:

أفترح نيدلسكي (1954) Nedelsky هذه الطريقة لتحديد درجة القطع للاختبارات التي تشتمل على أسئلة ذات اختيار من متعدد حيث تقوم مجموعة من المحكمين المختصين بتحديد درجة القطع المناسبة للاختبار، وفي هذه الطريقة يطلب من كل محكم فحص كل مفردة من مفردات الاختبار، ثم يحدد من بين الخيارات الخاصة لكل مفردة تلك التي يمكن أن يتجنب الطلاب ذوي الحد الأدنى للكفاية اختيارها لأنها لا تمثل في نظرهم الإجابة الصحيحة للمفردة، بعد ذلك يضع المحكم درجة للسؤال بناءً على ما تبقى من الخيارات، وذلك بأخذ مقلوب الخيارات المتبقية. فمثلاً إذا كانت المفردة تشتمل على خمسة خيارات ورأى المحكم أن الطالب ذو الحد الأدنى للكفاية يمكن أن يتجنب اختيار ثلاثة منها عندها تكون درجة السؤال عبارة عن:

$$0,5 = \frac{1}{2} = \frac{1}{3-5} =$$

بعد ذلك يقوم المحكم بجمع درجات الأسئلة، ويكون المجموع الكلي هو درجة القطع بالنسبة لذلك المحكم، وتجمع في النهاية الدرجات التي يتحصل عليها المحكمون، ثم يحسب المتوسط وبالتالي نحصل على درجة القطع للاختبار ككل، غير أن ندلسكاي وجد أنه من الأفضل إيجاد الانحراف المعياري (σ) لدرجات القطع ثم ضربها في مقدار ثابت (K) يتفق عليه المحكمون وتعتمد قيمة هذا المقدار على النسبة المئوية التي يرى المحكمون أنها مطلوبة لنجاح الطلاب من ذوي الحد الأدنى للكفاية.

ويتم تحديد درجة القطع في هذه الطريقة وفق المعادلة التالية:

$$MPL = \overline{MFD} + KsFD$$

$$\overline{MFD} = \frac{\sum FD}{N}$$

$$sFD = \sqrt{\frac{\sum (MFD - \overline{MFD})^2}{N}}$$

حيث إن:

MPL تمثل درجة القطع.

MFD تمثل متوسط الحد الأدنى المحدد من المحكمين.

K تمثل المقدار الثابت المحدد من المحكمين.

σ FD تمثل الانحراف المعياري لدرجات القطع

N تمثل عدد المحكمين.

وأشار بيرك (1986) Berk, إلى بعض عيوب هذه الطريقة وهي كالتالي:

لا يوجد تبرير منطقي يؤيد افتراض (ندلسكاي) أن الطالب الذي لا يعرف الإجابة الصحيحة لسؤال معين يلجأ إلى التخمين العشوائي لأن مشتتات الأسئلة تصمم عادةً لجذب غير المتمكنين، كما أن هذه الطريقة تسمح فقط بقيم احتمالية منفصلة تعتمد على عدد خيارات الإجابات، لذلك يضطر المحكمون لتعيين القيمة (0,5) لمعظم المفردات، هذا بالإضافة إلى أن هذه الطريقة تحتاج لقدرة كبير من التدريب لكي يستخدمها المحكمون استخداماً مناسباً.

٤ - طريق أنجوف-ندلسكاي المختلطة . (Angoff Nedelsky combination, Reid, 1984)

في هذه الطريقة يتم حساب المعيار على أساس حساب متوسط المعايير الناتجة عن استخدام طريقتي أنجوف وندلسكاي.

المميزات:

لها نفس مميزات طريقة أنجوف.

العيوب:

لها نفس عيوب طريقتي أنجوف وندلسكاي بالإضافة إلى احتمالية زيادة مصادر الخطأ المختلفة وذلك عندما نحسب متوسط المعيارين (Berk, 1986).

٥ - طريقة تقدير أهمية الصعوبة difficulty importance estimate (cangelosi,1984)

في هذه الطريقة يقوم المحكمون بفحص كل هدف، وذلك لتقدير الحد الأدنى للنجاح ويجب أن تتفق عينة المحكمين على الحد الأدنى وعلى كل هدف أيضاً، وذلك لأن للحد الأدنى وزن خاص طبقاً لأهمية صلته بالهدف المنشود، ثم تضرب النسبة المئوية لدرجة النجاح في نسبة أهمية الدرجة لكل هدف، والنتيجة يجمع من خلال جميع الأهداف، ويضرب في عدد بنود الاختبار وذلك للوصول للمعيار، ويجب أن يضع الحكام في الاعتبار الخطأ المعياري وذلك عند اتخاذهم القرارات الخاصة بدرجة القطع.

المميزات:

أ - سهولة الفهم.

ب - سهولة الاستخدام وتعمل على ربط درجة القطع بقائمة خصائص النطاق السلوكي.

العيوب:

استخدمت تقديرات المحكمين للأهداف لتقدير معيار بنود الاختبار، بينما نجد أنه من المفروض أن تكون العلاقة بين الأسئلة والأهداف مباشرة، وليس بالضرورة افتراض أن تقديرات الأسئلة تساوي تقديرات الأهداف، ومن المحتمل أن يكون نوع الخطأ المعياري المستخدم غير ملائم لتفسير درجة القطع. (Berk, 1986).

٦ - طريقة مواصفات المفردة (ميلز و بار ١٩٨٣م):

في هذه الطريقة يقوم المحكمون بقراءة مواصفات كل مفردة والتي تحتوي على المهارة المراد قياسها ووصف محتوى البند المقبول وخواص المثير والقوانين التي تحكم إحداث خيارات الاستجابة وعينة المفردات، ومن ثم فإن على المحكمين إدراك وتصور الإطار الكلي للمفردات والمتمثل في المواصفات والتنبؤ بنسبة البنود في الإطار الكلي، والتي يمكن أن يجيب عليها الممتحن ذو الحد الأدنى من الكفاية إجابة صحيحة.

المميزات:

سهولة الفهم وسهولة الاستخدام وتقوم بربط درجة القطع بمواصفات النطاق السلوكي وربطهما بكافة البنود التي تقيس النطاق السلوكي وتستغرق زمناً أقل مما لو عمدنا إلى تقدير كل مفردة بشكل منفصل، وتكون درجة القطع فيها مرتبطة بالنطاق السلوكي وليس بصيغة الاختيار المحدد.

العيوب:

تستخدم تقديرات المحكمين لمواصفات البنود لتقدير درجة القطع لمفردات الاختبار، بينما العلاقة بين المفردات ومواصفات البنود مباشرة، لذا ليس من الضروري أن تكون تقديرات درجة القطع مماثلة لتقديرات مواصفات البنود (Berk, 1986).

ثانياً: الطرق التي تعتمد جزئياً على التحكيم وتسترشد ببيانات

تجريبية:

أشار (علام، ١٩٩٥م، ٢٤١-٢٤٢) بأن الباحثين أدركوا قصوراً ملحوظاً في الطرق التحكيمية ومن ذلك القصور اعتمادها على تحكيم مفردات الاختبار دون النظر إلى الأداء الفعلي للمختبرين مما يؤدي ذلك إلى الحصول على معايير غير متميزة بدرجة عالية من الصدق.

وأشار كل من سكون وجاليون وفيرارا بأن درجة القطع المحددة باستخدام بعض من الطرق التحكيمية تؤدي إلى رسوب أكثر من نصف المتقدمين للحصول على ترخيص مهنة الطب في أمريكا، أما في حالة استخدام الطرق التجريبية فإنه يتبين خطأ تلك الدرجات حيث يجتاز معظم الراسبين الحد الأدنى لمستوى النجاح (علام، ١٩٩٥م، ٢٤٢).

لذا اقترح علماء القياس العديد من الطرق المختلفة لتحديد مستويات الأداء التي تتضمن طرقاً تجريبية محاولين في ذلك تجاوز عيوب الطرق التحكيمية والحصول على مستويات أداء دقيقة يمكن الاعتماد عليها ومن هذه الطرق مايلي:

١ - الطريقة التحكيمية المعززة بالمعلومات Informed Judgement method (popham,1981)

يتضح من مسمى هذه الطريقة أنها تعتمد على بيانات مسبقة عن أداء الطلاب حيث تقيّد هذه البيانات في إعطاء دلائل للمحكّمين عن المستوى العام للطلاب، وبالتالي توضع هذه البيانات في اعتبار المحكّمين لكي يسترشدوا بها أثناء تحديدهم للمعايير وفي هذا الشأن ذكر بيرك (1986) Berk بأن هذه الطريقة تعتمد على بيانات تتعلق بالأداء الفعلي للأفراد، وتتضمن هذه البيانات إحصاءات خاصة بالمفردات مثل معاملات الصعوبة والتمييز ومتوسط درجات الاختبار بعد تطبيقه على العينة المناسبة من الأفراد، وبعض تعليقات وأحكام الخبراء حول مفردات الاختبار، واقتراحات المختصين مثل المدرسين والإداريين وأساتذة الكليات وأعضاء مجالس المدارس، وتقدم جميع هذه المعلومات للمحكّمين لفحصها وتبادل المقترحات حولها ومناقشتها وإعادة النظر فيها وذلك لكي يتم التوصل إلى اتفاق حول درجة القطع المناسبة.

المميزات:

سهولة التطبيق و سهولة الفهم و يسهل حسابها، كما أنها تجمع وتوحد بيانات الأداء والبيانات المتميزة من مدى واسع بواسطة المجموعات المختصة بتلك البيانات (Berk, 1986).

العيوب:

من عيوب هذه الطريقة أنها تتطلب جمع كثير من البيانات التجريبية والمعلومات المتعلقة بالاختبار ومحتواها من مصادر متعددة، وقد يستغرق ذلك وقتاً وجهداً كبيرين، ويبدو أن تحليلات المحكّمين للبيانات غير منظمة، كما تبدو الحاجة لوجود خطوط تفصيلية تسهل استخدام أنواع البيانات المختلفة والمناقشة الجماعية للأداء الفردي، والتي لا تبدو مفيدة في هذه الطريقة بسبب تأثيراتها المعيارية (Berk, 1986).

٢ - طريقة أنجوف التجريبية المعدلة Modified Angoff empirical (mclean & Halpin,1984)

ذكر بيرك (1986) Berk أن هذه الطريقة عبارة عن تعديل لطريقة أنجوف التحكيمية، وتتطلب هذه الطريقة المعدلة تزويد المحكّمين ببعض البيانات المستمدة من التجريب الميداني لمفردات الاختبار، وذلك لكي يسترشد بها المحكّمين في عملية التحكيم، ويطلب من كل محكم تقدير احتمال إجابة الأفراد ذوي الحد الأدنى من الكفاية المطلوبة لمفردات الاختبار ككل وليس

من كل مفردة على حده كما في الطريقة السابقة (التحكيمية)، وبعد الحصول على هذه التقديرات يتم تزويد المحكمين ببيانات تجريبية عن مفردات الاختبار تشتمل على متوسط صعوبة كل مفردة، ثم يطلب منهم تعديل تقديراتهم الاحتمالية في ضوء تلك البيانات التجريبية، ويكون الحد الأدنى لمستوى الاجتياز في الاختبار هو متوسط مجموع تقديراتهم المعدلة.
المميزات:

سهولة التطبيق، سهولة الفهم ويسهل حسابها وتعمل على ربط المعيار بالكفايات المقاسة، وتدمج معامل صعوبة البند الحقيقي خلال عمليات التحكيم.
العيوب:

إن عملية استخدام تقديرات المحكمين للكفايات يجب أن تكون على درجة عالية من التوافق والانسجام، وذلك بين المفردات والكفاية ولا يمكن افتراض أن تكون التقديرات الواقعية متطابقة (Berk , 1986).

وأضاف جايجر Jaeger بأنه أجريت تعديلات في عام ١٩٧٨م على طريقة أنجوف الأساسية وذلك بهدف تبسيط مهمة المحكمين بحيث يطلب من كل محكم تقييم كل مفردة من مفردات الاختبار بالإجابة (بنعم) أو (لا)، ومن ثم يتم تحديد الحد الأدنى للكفاية وبعد الحصول على هذه التقديرات يستعرض عدد مختلف من المحكمين ثلاث أنواع من البيانات التجريبية، وذلك خلال ثلاث لقاءات منفصلة وتشتمل هذه البيانات التجريبية على التوزيع التكراري لدرجات الطلاب في الاختبار والأساليب الإحصائية الوصفية ودرجات القطع المحددة من قبل المحكمين، ثم يطلب من عينات المحكمين المختلفة تعديل تقديراتهم بناءً على تلك البيانات التجريبية، وتصبح درجة القطع عبارة عن متوسط تلك التقديرات التي حددتها عينات المحكمين في اللقاءات الثلاثة المنفصلة (Linn, 1989).

المميزات:

تتيح للمحكمين فرصة لتحسين تقديراتهم بناءً على ثلاث أنواع من البيانات المختلفة، ويحدد درجة القطع عينات مختلفة من المحكمين وتميل هذه الطريقة إلى زيادة ثبات درجة القطع (Berk, 1986).

العيوب:

تحديد الإجابة بنعم أو لا، بالإضافة إلى أن هذه الطريقة تفرض على المحكمين أن يلتقوا ثلاثاً لقاءات منفصلة، كما أنها تستغرق كثيراً من الوقت، وتعتبر أكثر تعقيداً من الطرق السابقة، و يعتبر تفسير المعيار النهائي صعب جداً بالنسبة للمحكمين.

لتلافي بعض من هذه العيوب قام كل من سنديرس ومابوس بإجراء تعديلات على الطريقة فأصبحت تحتوى على عينة واحدة من المحكمين، وأصبح تحديد المعيار النهائي يتم خلال لقاء واحد يجمع بين المحكمين (Berk, 1986).

كما قام شيررد أيضاً بإجراء تعديلات على طريقة أنجوف الأساسية بحيث أصبحت هذه الطريقة تتطلب أربعة أنواع من المعطيات التي يجب أن توضع بعين الاعتبار وهي:

أ- أحكام مطلقة مبنية على فحص المفردات (مثل طريقة أنجوف).
ب- بيانات تجريبية على عينة محكمة صنف المتقنين وغير المتقنين (مثل طريقة المجموعات المتضادة).

ج- تقديرات صانعي القرار (المحكمين) لمستوى النجاح المقبول.

د- التوفيق بين تقديرات المحكمين المختلفة.

فإذا كانت درجات القطع الممكنة والمختلفة قد نشأت من هذه المعطيات فيجب على محدد المعايير أن يوفق بين هذه الاختلافات (Berk, 1986).
المميزات:

من مميزات أنها تضم أربعة أنواع من البيانات التجريبية والتحكيمية وتضع في حسابها معظم المصادر الرئيسية للقيام بعملية تحديد المعايير، وتحفظ بمميزات طريقتي أنجوف والمجموعات المتضادة (Berk, 1986).

العيوب:

لها نفس عيوب طريقتي أنجوف والمجموعات المتضادة بالإضافة إلى صعوبة التوفيق بين الاختلافات بين المعايير وتعديل المعايير المبنية على معدلات النجاح المقبولة وعدم الثبات، بينما نجد أن إستراتيجية (التثليث) لها مميزات معينة وخطوط إرشادية محددة، والتي يمكن أن تكون مقبولة بشكل كبير حتى تصل بشكل منظم إلى المعيار النهائي (Berk, 1986).

٣ - طريقة توفيق بين الطرق المطلقة والطرق النسبية Absolute Relative Compromise (I)(II)

أشار بيرك (1986) Brek إلى أن هذه الطريقة تعمل على التوفيق بين أحكام المحكمين المتعلقة بأداء الأفراد في الامتحان والأداء الفعلي المستمد من البيانات التجريبية المبنية من تطبيق الاختبار على العينة، وفي هذه الطريقة يطلب من المحكمين تحديد قيمتين هما:

أ- الحد الأدنى للنسبة المئوية لعدد المفردات التي يجب أن يجيب عليها الفرد إجابة صحيحة لكي يعتبر ناجحاً في الاختبار، ويرمز لهذه النسبة بالرمز (k).

ب- النسبة المئوية لعدد الأفراد المتوقع اجتيازهم للاختبار ويرمز لهذه النسبة بالرمز (V)، ويمكن التوصل إلى معادلة تربط بين النسبتين (K)، (V)، والتي يمكن استخدامها لتعديل تلك القيم التي لا تتفق مع التوزيع الفعلي لدرجات الاختبار، إذ يمكن رسم منحنى للمتوسطات والانحرافات المعيارية لقيم (K)، (V) التي حددها المحكمون، وكذلك النسبة المئوية لعدد الأفراد المتوقع نجاحهم ويرمز لها بالرمز (Y) كدالة تناقصية لدرجات الاختبار ويرمز لها بالرمز (X)، ثم يرسم منحنى لتحديد النقطة ويرمز لها بالرمز (I) والتي تمثل الحد الأدنى لمستوى النجاح، بحيث توفق بشكل أمثل بين المستويات المطلقة (التحكيم) والمستويات النسبية (الأداء الفعلي للأفراد في الاختبار).

المميزات:

من مميزات هذه الطريقة أنها توفق بين الأحكام المتعلقة بمفردات الاختبار وبين الأداء الفعلي للمختبرين أي بين المستويات النسبية.

العيوب:

من عيوبها أنها تتطلب من المحكمين أن يوفروا تقديرين وهذا يجعل تنفيذ هذه الطريقة أكثر صعوبة من غيرها سواء في عملية التحكيم أو من الناحية الإحصائية، وكذلك تتطلب جهداً كبيراً من جانب المحكمين، بالإضافة إلى أن تفسير المعيار من قبل المتخصصين قد يكون صعب جداً. لذلك أجرى هوفستي تعديلات لهذه الطريقة حيث يتطلب من المحكمين تحديد أربع قيم هي:

أ- أعلى نسبة مئوية لعدد المفردات التي يمكن أن يجيب عليها الفرد إجابة صحيحة ويرمز لها بالرمز (Kmax) وهي عبارة عن درجة القطع المرضية حتى إذا حصل عليها جميع الطلاب.
ب- أدنى نسبة مئوية من عدد المفردات التي يمكن أن يجيب عليها الفرد إجابة صحيحة ويرمز لها بالرمز (Kmin) وهي عبارة عن درجة القطع الدنيا التي لا يمكن لأي الطالب أن يحرز أقل منها.

ج- أعلى نسبة مئوية مقبولة من الراسبين ويرمز لها بالرمز (Fmax).

د- أدنى نسبة مئوية مقبولة من الراسبين ويرمز لها بالرمز (Fmin).

وباستخدام توزيع درجات الاختبار المتجمع (التراكمي) فإننا نستطيع تحديد العلاقة التجريبية بين (K) و (F)، ويكون المعيار هو نقطة التقاطع بين النموذج (المطلق) والمنحنى (النسبي).

المميزات:

من مميزات هذه الطريقة أنها توفق بين الأحكام المتعلقة بمفردات الاختبار وبين الأداء الفعلي للمختبرين، أي بين المستويات المطلقة والمستويات النسبية كما تضع في اعتبارها الحد الأعلى للنسبة المئوية للمفردات التي من الممكن أن يجيب عليها الفرد إجابة صحيحة (Kmax)، كما

تقوم بتبديل النسبة المئوية من الأفراد المتوقع اجتيازهم للاختبار (V) بأعلى نسبة مئوية مقبولة من الراسبين (Fmax) وبأدنى نسبة مئوية مقبولة من الراسبين (Fmin) (Brek, 1986).
العيوب:

من عيوب هذه الطريقة أنها تتطلب جهداً كبيراً من المحكمين، وتتطلب توفير أربعة تقديرات وهذا يجعل تنفيذ هذه الطريقة أكثر صعوبة من غيرها سواء في عملية التحكم أو الإحصاء (Berk, 1986).

ثالثاً: طرق تعتمد على البيانات التجريبية وتسترشد بالتحكيم

:Empirical Judgmental

١ - طريقة المجموعات المحكية (BERK, 1984) CRITERION GROUPS

أشار كل من بيرك وهامبلتون وآخرون (BERK, 1986; Hambleton et al, 1978) إلى أن هذه الطريقة تتطلب انتقاء مجموعتين من الطلاب الأولى متقنة (تلقت التعليم) والثانية غير متقنة (لم تتلق التعليم)، ويكون مدى توزيع درجات الطلاب من الدرجة (صفر) إلى الدرجة (ن) حيث (ن) تمثل عدد الأسئلة التي نقيس كل هدف على حده، وتتطلب الطريقة تقسيم الطلاب بناءً على درجة القطع (س) إلى قسمين (متقنين - غير متقنين)، وبافتراض أن الطلاب في المجموعة المتقنة (التي تلقت التعليم) عبارة عن طلاب متقنين فإنه يمكن تقسيمهم بناءً على درجة القطع إلى فئتين الأولى متقنين بالفعل (TM) والثانية غير متقنين غير حقيقيين (FN)، وبالمثل يتم تصنيف طلاب المجموعة الثانية غير المتقنة (التي لم تتلق التعليم) إلى فئتين الأولى غير متقنين فعلاً (TN) والثانية متقنين غير حقيقيين (FM).

وتعتبر توزيعات درجات الطلاب في المجموعتين المتقنة وغير المتقنة المحددات الأولية (Primary Determinants) لمدى دقة تصنيف الاختبار للطلاب كمتقنين وغير متقنين فعلاً للهدف المراد قياسه.

وتحدد قيمة درجة القطع (س) من خلال فحص خلية الاحتمالات لكل درجة قطع محتملة، وتكون درجة القطع المثلى تلك التي تزيد نسبة التصنيفات الصحيحة (TM - TN) وفي المقابل تحجم نسبة التصنيفات الخاطئة (FM - FN).

ويتم تحديد درجة الإتقان بشكل أولي من خلال توزيعات تكرارات درجات الطلاب في المجموعتين المتقنة (التي تلقت التعليم) وغير المتقنة (التي لم تتلق التعليم) حيث يصنف الطلاب بناءً على نقطة تقاطع التوزيعين إلى فئتين متقنين بالفعل وغير متقنين بالفعل للهدف المقاس.

ثم استخدم بيرك (1976) Berk طريقة بديلة تهدف لإيجاد معامل الصدق لكل درجة قطع
محتملة (س)، لذا فإن متغيرات التصنيف تعطي قيم ثنائية هما: -

أ- متغير التصنيف المنبئ The Predictor Classification

حيث يقدر للطالب القيمة (1) عندما تكون درجته أعلى من درجة القطع، ويقدر للطالب القيمة
(صفر) عندما تكون درجته أدنى من درجة القطع.

ب- متغير التصنيف المحك (Criterion Classification)

يقدر لكل طالب ضمن مجموعة المتقنين القيمة (1) بينما يقدر لكل طالب ضمن مجموعة غير
المتقنين (صفر)، ويحسب معامل الصدق من قيمة معامل (فاي) بين المتغيرين الثنائيين.

يلي ذلك حساب المنفعة والضرر النسبيين لأخطاء التصنيف، ويمكن تمثيل قيمة الضرر
للخطأ (FN) بالقيمة (-1) وللخطأ (FM) بالقيمة (-2)، في المقابل فإنه يتم حساب الفائدة
النسبية لتحديد الإتقان بالفعل (TM) بالقيمة (2) ولتحديد عدم الإتقان بالفعل (TN) بالقيمة (1).
مميزاتها:

أنها نموذج للصدق التكويني وتعتبر هذه الطريقة ذات حساسية خاصة للخصائص الإحصائية
للمفردة وكذلك الصدق البنائي.

عيوبها:

من عيوبها أنها لا تحدد المعيار الحقيقي وفي أفضل الحالات تقوم بتحديد موقع الأداء على متصل
الدرجات حيث يتعذر التمييز بين المجموعتين المتقنة وغير المتقنة.

٢ - طريقة المجموعات المتضادة (LIVINGSTON & ZIEKY, 1982) **CONTRASTING GROUPS**

ذكر كل من بيرك وجايفر بأن هذه الطريقة ماثلة لحد ما لطريقة المجموعات المحكية
وطريقة المجموعة الحدية، ولكنها تختلف في كيفية تحديد درجة القطع في الاختبار، ففي هذه
الطريقة يقوم المحكمون المتخصصون بتحديد مجموعتين من الطلاب بحيث تكون الأولى متقنة
والأخرى غير متقنة بناءً على ما يلي: -

المعلومات المتوفرة حول مستويات مهارات الطلاب.

استناداً على درجات الطلاب وأدائهم داخل الفصل حيث يتم تصنيف الطلاب إلى فئتين الأولى
الحاصلون على تقدير (جيد جداً فما فوق) والثانية الطلاب الحاصلون على تقدير (مقبول فما
دون).

ثم يطبق الاختبار على كلتا المجموعتين ويرسم منحنيان يمثلان التوزيعين التكراريين
لدرجات الطلاب في الاختبار، وتمثل نقطة تقاطع المنحنيين درجة قطع الاختبار، هذا الاقتراح

يفترض تداخل التوزيعين التكراريين وعدم تطابقهما، ويمكن تعديل هذه الدرجة آخذةً في اعتبارها خطأً التصنيفين، أي يمكن اختيار الدرجة التي تقلل من خطأ التصنيف الموجب (FALSE POSITIVE) أو خطأ التصنيف السالب (FALSE NEGATIVE).

عيوبها:

لها نفس عيوب طريقة المجموعات المحكية، بالإضافة إلى أنه من الصعب تجاوز الخطأ الناجم من المحك (Linn, 1989).

٣ - طريقة المجموعة الحدية (LIVINGSTON&ZIEKY,1982) **BORDERLINE GROUP**

ذكر كل من بيرك وجايفر (BERK, 1986; JAEGER, 1989) بأن هذه الطريقة تتطلب من المحكمين انتقاء عينة من الطلاب الذين يرون بأن كفايتهم المراد قياسها تقع عند الحد الفاصل بين التمكن وعدم التمكن، ومن المحتمل أن تكون نسبة الطلاب قليلة بالنسبة للمجموعة الكلية، ثم يطبق الاختبار على أفراد العينة وبعد ذلك يتم إيجاد وسيط درجات هذه المجموعة ليكون بمثابة مستوى الأداء المطلوب.

مميزاتها:

تمتاز هذه الطريقة بأنها سهلة الفهم والتطبيق والإحصاء غير أنها تحتاج إلى عدد كاف من الطلاب وذلك لكي نحصل على تقديرات مستقرة نسبياً.

عيوبها:

من عيوبها أنه يصعب على المحكمين أو المدرسين انتقاء المجموعة الحدية، إذ أن تعريف أو تحديد الطالب الذي يقع عند الحد الفاصل بين الإتقان وعدم الإتقان أصعب بكثير من تحديد الفرد المتقن أو غير المتقن، كما تتدخل بعض المتغيرات في عملية انتقاء أفراد هذه المجموعة وقد يصعب تحديد الحجم المناسب للعينة.

٤ - طريقة المحك معياري المرجع (NORM REFENCEED CRITERION):

ذكر بيرك (BERK 1986) بأنه في هذه الطريقة يتم اختيار المعيار بناءً على العلاقة بين درجات القطع في الاختبار محكي المرجع وبين درجات الأداء في الاختبار معياري المرجع (NRT) ويتم اختيار درجة القطع أولاً في الاختبار معياري المرجع بطريقة اختيارية (على سبيل المثال: مناظرة الدرجة الخام بالدرجة المئوية ٥٠ أو ٧٥) ثم تختار درجة القطع في الاختبار محكي المرجع على أساس أنها تزيد مايلي:

أ- نسبة ثبات القرارات الناتجة عن تطبيق الاختبار معياري المرجع ويرمز لها بالرمز (P).

ب- نسبة ثبات القرارات المصححة من عامل الصدفة ويرمز لها بالرمز (K).

ج- تعيين دالة الفقد (Loss Function) ويرمز له بالرمز (L).

المميزات:

من مميزات أنها سهلة التطبيق والفهم والإحصاء ويمكن توفير بيانات محك الاختبار معياري المرجح بسهولة ويسر، بالإضافة إلى أن هذه الطريقة تضع في اعتبارها دالة الفقد (LOSS FUNCTION) بالإضافة إلى العلاقة المباشرة بين الاختبارين.

العيوب:

إن استخدام درجات القطع الاختيارية في الاختبار معياري المرجح كمحك يعد خطأ فني وغير منطقي.

٥ - طريقة النتائج التربوية (BLOCK, 1972) EDUCATIONAL CONSEQUENCES:

ذكر بيرك (1986) BERK بأن هذه الطريقة تعتمد أساساً على اختيار المعيار من خلال العلاقة بين درجات الاختبار ومحك التعلم المستقبلي (نتائج انفعالية أو معرفية) حيث يتوقع أن يزداد المنحنى طردياً ليربط بين المتغيرين، وتحدد درجة القطع وفقاً على أثرها في زيادة أداء المحك.

المميزات:

سهولة التطبيق والفهم والحساب.

العيوب:

إن معيار (التعلم المستقبلي) يحتاج إلى تحديد واضح و لا تبدو الافتراضات في هذا النموذج متطابقة مع البيانات الواقعية، بالإضافة إلى أن العلاقة بين المتغيرات المقترحة لا توفر "دالة درجية" (STEP FUNCTION) ، أو خاصية فاعلة في المحك لتحديد درجة القطع.

٦ - طريقة إمريك لتقييم اختبار التمكن Emrick's mastery Testing Evaluation

ذكر إمريك (1971) EMRICK بأن هذه الطريقة تناسب الاختبارات الفرعية القصيرة التي تتكون من خمس فقرات أو أقل، وتعتمد هذه الطريقة على نظرية القرار، إذ يتم من خلالها تحليل ثلاثة من أخطاء القرار وهي الخطأ الأول (α) والخطأ الثاني (β) ونسبة الفاقد لأخطاء القرارين الأول والثاني (RR) لذلك يجب تحديد الخطأ الأول والخطأ الثاني ونسبة الفاقد وطول الاختبار، ويتم حساب نسبة الفاقد من خلال تقييم الأنواع المختلفة لأخطاء القرار، ومن ثم تسجيل مجموع المخاطر بتحديد درجة القطع باستخدام المعادلة التالية:

$$K = \frac{\log \frac{\beta}{1-\alpha} + 1/n(\log RR)}{\log \frac{\alpha\beta}{(1-\alpha)(1-\beta)}}$$

حيث إن:

K ترمز إلى درجة القطع. α ترمز إلى الخطأ الأول.

β ترمز إلى الخطأ الثاني. RR ترمز إلى نسبة الفاقد للخطأين الأول والثاني.

n ترمز إلى عدد الأسئلة.

٧ - طريقة نوفيك ومعاونيه THE WORK OF NOVICK & COLLABORATORS

ذكر (1976) MESKAUSKAS بأن هذه الطريقة عبارة عن مزيج من الطرق المختلفة، وتهدف إلى تحديد كل من درجة القطع والتوزيع الاحتمالي للأداء و الفوائد الناتجة عن الخطأ الأول (α) والخطأ الثاني (β) ويفيد تحديد هذه المتغيرات في زيادة دقة القرارات المتخذة، حيث يتم وضعها في شكل توزيع احتمالي، ويضاف إلى هذا التوزيع درجات الاختبار بحيث ينتج عنه توزيع احتمالي أكثر حساسية، ويتم تحديد القيمة المقدرة للمستوى الحقيقي للطالب المتقن عن طريق تحديد عينات من المفردات، ثم يضرب الاحتمال الذي تزيد فيه الدرجة الملاحظة على التحصيل المعطى، ثم تقارن النتيجة بالدالة المشابهة التي تشير إلى الاحتمال الذي لا تزيد فيه الدرجة الملاحظة.

٨ - طريقة رودابش للدرجة الحقيقية الثنائية ROUDABUSH'S DICOTOMOUS

TRUE SCORE

ذكر (1976) MESKAUSKAS بأن طريقة رودابش تستخدم مع الاختبارات ذات الفقرة الواحدة، وتشتمل هذه الطريقة على نموذجين من النماذج القطعية وكلا النموذجين يشتملان على مقياس ثنائي للدرجة الحقيقية، وتقوم هذه الطريقة على تحديد أربعة من الأخطاء المحتملة وهي α_1, α_2 (احتمال أن يكون الفرد المتقن ضمن غير المتقنين في المحك) و β_1, β_2 (احتمال أن يكون الفرد المتقن ضمن غير المتقنين في المحك) ويتم تحديد الأخطاء الأربعة وفق العلاقات التالية:

$$\alpha_2 = P(X \geq X^c \mid T = 0) \quad \alpha_1 = P(X \geq X^c \mid T = 1)$$

$$\beta_2 = P(X \geq X^c \mid T = 0) \quad \beta_1 = P(X \geq X^c \mid T = 1)$$

حيث إن:

الدرجة الحقيقية = T الدرجة الناتجة = X درجة القطع = XC

وتستخدم العلاقات التالية لتحديد درجة القطع وهي كالتالي :

$$f_{00} = N_0(1 - \alpha_1)(1 - \alpha_2) + N_1\beta_1\beta_2$$

$$f_{01} = N_0(1 - \alpha_1)\alpha_2 + N_1\beta_1(1 - \beta_2)$$

$$f_{10} = N_0\alpha_1(1 - \alpha_2) + N_1(1 - \beta_1)\beta_2$$

$$f_{11} = N_0\alpha_1\alpha_2 + N_1(1 - \beta_1)(1 - \beta_2)$$

٩- طريقة ملمان ذات القرار المبني على الحدين MILLMAN'S BINOMIAL BASED

DECISION

ذكر كل من هامبلتون، وآخرون وميسكاوسكاس (Hambleton et al, 1978; Meskauskas, 1976) بأن ملمان طور مجموعة من الجداول التي تعتمد على التوزيع ذو الحدين، والذي يشير إلى كل من تقديرات الخطأ المتوقعة لمجموعة متعددة من الدرجات الحقيقية وطول الاختبار ودرجة القطع، وهذه الطريقة تقوم على افتراض أن الاختبار يتكون من مجموعة عشوائية من الدرجات تتراوح ما بين (١ - صفر)، ويستخدم التوزيع ذو الحدين لإيجاد البيانات حول قيم محددة مفترضة للدرجة الحقيقية في كل من أطوال الاختبار والاختبارات الكبيرة والتوزيع الهندسي الفوقي أو ذي الحدين، فيتم التنبؤ بالتكرارات المتصلة، ثم يرسم التكرار المتصل للنسب المئوية المختلفة من الأسئلة التي تمت الإجابة عليها بشكل صحيح.

١٠- طريقة ديفز ودايموند (١٩٧٤م) THE DAVIS DIAMOND BAYESIAN

METHOD

أشار ميسكاوسكاس (1976) MESKAUSKAS إلى أن هذه الطريقة تعتمد على نموذج بيبز، وقد حدد كل من ديفز ودايموند جدولاً يضع تقديرات تعتمد على الدرجة الملاحظة واحتمال أن يكون المستوى الحقيقي لكفاءة الطالب بالقرب أو أعلى من المستويات المختارة. بالنسبة للاختبار (ذو خمس فقرات) فإن احتمال أن يكون الطالب قد أحرز خمس درجات وله مستوى كفاءة حقيقي يساوي (٠,٩٩) أو أعلى من ذلك يساوي (٠,١١٣٥) وهو رقم أدنى، ويعرض الجدول درجات القطع الضرورية في مستويين من التصنيفات الحقيقية هما (٠,٨٥) ودرجة القطع للاختبارات حتى عشرون فقرة. فإذا أراد صانع القرار أن يقتطع نسبة ٨٥% من التصنيفات الحقيقية للطلاب مع وجود مستويات كفاية عند نسبة ٩٠% أو أعلى فيجب أن تكون هنالك مجموعة متساوية لعدد المفردات في الاختبار، ولتحقيق هذا الهدف يجب استخدام اختبار تتراوح فقراته ما بين ١٢ إلى ٢٠ فقرة.

أسس تقويم طرق تحديد درجة القطع:

ذكر بيرك (Berk, 1986) بأنه يوجد معيارين لتقويم طرق تحديد درجة القطع وهما كما يلي:

١: الكفاية الفنية: Technical - Adequacy

أ- أن تحدث الطريقة تصنيفاً مناسباً للمعلومات: Appropriate Classification of Information

بمعنى أن تحدث الطريقة تصنيفات ثنائية مثل الإتيقان / عدم الإتيقان، الفشل / النجاح، ويعتبر التصنيف أساسياً لمعظم القرارات المتخذة.

ب- أن تكون حساسة لأداء المختبر Sensitive to Examinee Performance

يجب أن تكون الطريقة حساسة لمستويات صعوبة الأسئلة المختلفة ولأي خصائص أخرى مثل معامل التمييز.

ج- أن تكون الطريقة حساسة لكل من المتعلم أو لعملية التدريب Sensitive to Instruction or Training

يجب أن تراعي الطريقة عملية التعلم والتدريب الفعلي الذي تلقاه الطالب بحيث تقاس المهارات والكفايات التي تلقاها الطالب بالفعل.

د- أن تتضمن الطريقة طرقاً إحصائية مناسبة Statistically Sound

يجب استخدام الطرق الإحصائية المناسبة، ويجب أن تفسر بشكل صحيح.

هـ- أن تحدد الطريقة المعيار الصحيح Identify the true standard

بمعنى أنه يجب أن تراعي الطريقة أخطاء القياس.

و- أن تثبت الطريقة صدق القرار Decision validity evidence

بمعنى أنه يجب الحصول من الطريقة تقدير احتمالات قرارات التصنيفات الصحيحة والخاطئة.

٢- الجانب التطبيقي (العملي) Practicability

أ- أن تكون الطريقة سهلة التطبيق Easy to implement

يجب أن تكون الخطوات المستخدمة في تطبيق الطريقة واضحة وسهلة الفهم للمشاركين فيها، وأن تتم في قدر معقول من الوقت.

ب- أن تحسب الطريقة بسهولة Easy to Compute:

يجب أن تكون الطرق الإحصائية المستخدمة للحصول على المعيار النهائي سهلة الاستخدام سواءً باستخدام الآلة الحاسبة أو البرامج الإحصائية المتوفرة في الحاسب الآلي الشخصي أو المركزي.

ج- أن قابلة للتفسير بسهولة للعامة Easy to interpret to lay people

يجب أن تكون الطريقة المستخدمة في تحديد درجة القطع قابلة للتفسير وقابلة للفهم، ويجب أن تكون تفسيرات الطريقة واضحة وسهلة الإدراك والفهم للعامة والمهتمين على حد سواء.

د- أن تكون الطريقة جديرة بثقة العامة Credible to lay people

يجب أن تكون الطريقة مقنعة وجديرة بثقة العامة.

كما أضاف بيرك (Berk, 1982) بأن هناك عدة عوامل تساعد على اختيار الطريقة المناسبة

وهي:

هـ - مدى أهمية القرارات المبنية على تحديد درجة القطع.

و- الكمية المتاحة من الوقت لتحديد درجة القطع.

ح- مدى توفر المصادر (المادية - الطاقة البشرية) لإنجاز مهمة تحديد درجة القطع.

ط- كفاية المحكمين (بعض الطرق تتطلب دراية أكثر بالمحتوى المراد قياسه وبمستوى

الطلاب التحصيلي).

ص- مدى مناسبة طريقة تحديد درجة القطع لنوعية الاختبار المراد قياسه.

الدراسات السابقة

نظراً لما لتحديد درجة القطع والقرارات التي تبنى على نتائجها من أهمية كبرى، كان من الضروري إجراء دراسات متعددة ومختلفة تدور حول الطرق المناسبة لتحديد درجة القطع، وذلك باستخدام طرق وأساليب متعددة تهدف إلى مقارنة تلك الطرق المختلفة وأثر العديد من المتغيرات عليها، وذلك بهدف الوصول إلى درجة قطع يمكن الوثوق بمصداقيتها وموضوعيتها. تكمن أهمية الدراسات السابقة في كونها تلقي الضوء على الخطوات المتبعة في تحديد درجة القطع وأثر المتغيرات المختلفة على تحديدها، بالإضافة إلى أن هذه الدراسات سوف تكشف عن الأساليب الإحصائية المناسبة التي تساعد على تحليل نتائج الدراسة الحالية ويمكن تصنيف هذه الدراسات إلى مجموعتين هما:

أولاً- الدراسات العربية. ثانياً- الدراسات الأجنبية.

أولاً: الدراسات العربية:

١ - دراسة صلاح الدين محمود علام (١٩٨٥م):

عنوان الدراسة: استخدام النموذج ذي الحدين في تقدير درجة القطع لاختبار محكي المرجع (دراسة إحصائية وتجريبية).

هدف الدراسة:

تهدف الدراسة إلى ما يلي:-

أ- بناء توزيع نظري احتمالي على أساس النموذج ذي الحدين Binomial Model لدرجات الاختبار المستخدمة في الدراسة لبعض مستويات التمكن بعد تصحيح هذه الدرجات من أثر التخمين.

ب- استخدام هذا التوزيع الاحتمالي لبحث إمكانية تحديد درجة القطع للاختبار على أساس البيانات الواقعية المستخدمة من عينة الدراسة.

أداة الدراسة:

أعد الباحث اختباراً يتألف من صورتين متكافئتين (أ، ب) يشتمل كل منهما على أربعة اختبارات فرعية تقيس المهارة في صياغة الأهداف السلوكية.

عينة الدراسة:

كانت عينة الدراسة عبارة عن مجموعة من طلاب كلية التربية بجامعة الأزهر، وبلغ عدد أفرادها (١٥٤) طالباً.

نتائج الدراسة:

كون الباحث توزيعاً نظرياً احتمالياً للنسب المئوية لأعداد الطلاب الذين يمكن اعتبارهم متمكنين وفقاً لدرجات قطع ومستويات تمكّن مختلفة، مستخدماً في ذلك النموذج ذي الحدين، ثم اختبر حسن مطابقة التوزيع الواقعي للنسب المئوية لأعداد طلاب العينة اعتماداً على درجاتهم الفعلية في الصورة (ب)، وكان التوزيع النظري الاحتمالي عند مستوى تمكّن (٧٠، ٠) بين المتغيرين الثنائيين. والذي أختاره الباحث للمهارة المطلوبة، ووجد أنه لا توجد فروق ذات دلالة إحصائية بين كل من التوزيعين، مما أدى إلى اعتبار مستوى التمكن (٧٠، ٠) مناسباً لمجموعة الطلاب المشاركين في البرنامج.

٢ - دراسة صلاح الدين علام (١٩٩١م):

عنوانها: مقارنة بعض طرق تحديد مستويات الأداء في اختبار مرجعي المحك

تهدف هذه الدراسة إلى ما يلي:

الموازنة بين أربع طرق لتحديد مستويات الأداء في الاختبار، الذي أعده الباحث اثنتان منها تعدان من الطرق المطلقة، أي التي تعتمد على تحكيم محتوى الاختبار ومفرداته، وهما طريقتا أنجوف وندلسكاي، والطريقتين الأخرين تعدان من الطرق النسبية، أي التي تعتمد على تحكيم أداء المختبرين وجمع بيانات تجريبية فعلية تتعلق بهذا الأداء وهما طريقتي المجموعة الحديثة والمجموعة المتناقضة.

عينة الدراسة:

تكونت عينة المحكمين من (٢٦) محكماً من ثلاث مجموعات متباينة في مستوى تمكّنها في مجال القياس والتقويم التربوي والنفسي، اشتملت المجموعة الأولى على (١٤) معلماً من معلمي المدارس الثانوية العامة الحاصلين على دبلوم خاص في التربية أو على دورة تدريبية في مجال القياس والتقويم، واشتملت المجموعة الثانية على ستة مدرسين مساعدين حاصلين على درجة الدكتوراه في علم النفس التعليمي ولديهم خبرة لا تقل عن ثلاثة أعوام في تدريس مادة الاختبارات والمقاييس، واشتملت المجموعة الثالثة على ستة من أساتذة علم النفس التعليمي، أما بالنسبة لعينة الطلاب الذين أجرى تحكيم أدائهم في الاختبار فكانت (٣٨٢) طالباً من طلاب كلية التربية ممن يدرسون مقرراً سيكولوجياً في الفروق الفردية، ويقوم بتدريسهم عينة المحكمين المكونة من الأساتذة والمدرسين المساعدين التي اختيرت في هذه الدراسة.

نتائج الدراسة:

أدى استخدام الطرق الأربع المختلفة المستخدمة في الدراسة إلى الحصول على درجات قطع مختلفة نسبياً فقد كانت متوسطاتها ٤٠، ٤٤، ٤٣، ٤١، لكل من طريقة ندلسكاي والمجموعة الحدية وأنجوف والمجموعات المتناقضة على التوالي.

٣ - دراسة نادية محمد عبد السلام (١٩٩١م)

عنوانها: تعدد معطيات تحديد الدرجات الفاصلة في القياس محكي المرجع

تهدف الدراسة إلى ما يلي: -

تحديد الدرجة الفاصلة لاختبار في الإحصاء الوصفي يتكون من (٥١) مفردة تقيس ستة مجالات متعددة (مستويات القياس - تصنيف البيانات - مقاييس النزعة المركزية - مقاييس التشتت - الارتباط - التحويلات - المعايير - الدرجات المعيارية) وذلك وفق طريقة المجموعات المحكية (Criterion Groups).

عينة الدراسة: تتكون عينة الدراسة من مجموعتين محكيتين كما يلي: -

أ - مجموعة المتعلمين:

وتتكون من الطلبة الذين تلقوا تعليم فعال على منهج الإحصاء الوصفي، وبلغ عدد أفراد هذه العينة (٧٥) طالب وطالبة من طلبة الدراسات العليا.

ب - مجموعة غير المتعلمين: -

وتتكون من الطلبة الذين لم يتلقوا تعليم فعال على مقياس الدراسة، وبلغ عدد أفراد العينة (٧٥) طالبة من طالبات السنة الأولى قسم علم النفس.

خطوات تحديد الدرجة الفاصلة: -

لتحديد الدرجة الفاصلة قامت الباحثة بإيجاد تصنيف ثنائي لكل توزيع تكراري ولكل درجة فاصلة ممكنة، وتم حساب نتائج الاحتمالات الأربعة التالية (طلاب متقنين فعلاً - طلاب متقنين غير حقيقيين - طلاب غير متقنين فعلاً - طلاب غير متقنين غير حقيقيين) وذلك بقسمة تكراراتهم على مجموع أفراد العينتين، ثم تحديد الدرجة الفاصلة بناءً على نقطة تقاطع منحنيات التوزيعات التكرارية لدرجات العينتين في الاختبارات الفرعية الستة، ثم أوجدت احتمالات القرارات الصحيحة وأخطاء التصنيفين الأول والثاني ومعامل صدق المفردات والفائدة والضرر النسبيين، وذلك لكل درجة فاصلة ممكنة ولكل اختبار فرعي على حده.

نتائج الدراسة:-

اعتمدت الباحثة على المعايير السابقة في اختيار الدرجة الفاصلة التي تحجم نسبة أخطاء التصنيفات الخاطئة، وذلك لكل اختبار فرعي.
وكانت الدرجات الفاصلة للاختبارات الفرعية (مستويات القياس - تصنيف البيانات - مقاييس النزعة المركزية - مقاييس التشتت - الارتباط - تحويلات المعايير و الدرجات المعيارية) (٤ - ١١ - ٦ - ٦ - ٨ - ٤) على التوالي.

٤ - دراسة نادية محمد عبدا لسلام (١٩٩٢م):

عنوانها: ثلاثة محكات رئيسة لتحديد مستوى الإتقان في الاختبارات محكية المرجع (دراسة إمبريقية).

تهدف هذه الدراسة إلى مايلي:

تحديد الدرجات الفاصلة لاختبار محكي المرجع، والحكم على مناسبتها لتقدير إتقان الطالب باستخدام ثلاث محكات هي:

أ- متوسط الفرق المطلق (AAD) Average Absolute Difference

ب- قرار الإتقان Decision Accuracy

ج- نموذج الاحتمالية.

أداة الدراسة:

أعدت الباحثة اختباراً محكي المرجع يقيس منهج الإحصاء الوصفي، يتكون من ستون مفردة تقيس ستة مجالات مختلفة.

عينة الدراسة:

تشتمل عينة الدراسة على تسع وأربعون طالباً وطالبة من طلبة الدراسات العليا (دبلوم خاص + تمهيدي ماجستير) في كلية البنات.

نتائج الدراسة:

الجدول التالي يوضح نتائج الدراسة الناتجة:

مجالات الاختبار	عدد المفردات	الحد الأدنى للنجاح ن	قرار الإلتقان	متوسط الفرق المطلق	احتمالية النجاح لمن أجاب على الحد الأدنى للنجاح أو أكثر
١- مستويات القياس	٦	٥	٨٢،	١١،	٠٠٥
٢- تصنيف البيانات	١٥	١٣	٧٢،	١٢،	٠٠٠١
٣- مقاييس النزعة المركزية	١١	٩	٥١،	١٧،	٠٠٠١
٤- مقاييس التشتت	٨	٧	٨٢،	٠٩،	٠٠٠٣
٥- معامل الارتباط	١٠	٨	٦٥،	٣١،	٠٠٠٤
٦- التحويلات، المعايير، الدرجات المعيارية	١٠	٩	٨٤،	٠٨،	٠٠٠٣

وترى الباحثة أن قرار الإلتقان للقسم الثالث والخامس منخفض نوعاً ما، وذلك لبعدها عن درجة الحد الأدنى للنجاح عن النهاية العظمى.

ثانياً: الدراسات الأجنبية:

١- دراسة كلاً من **Barbara J. Andrew & James - Hecht (1976)**:

عنوان الدراسة (دراسة تمهيدية لطريقتين من طرق تحديد درجات الاختبار).

أهداف الدراسة:

تحدد أهداف الدراسة في التساؤلات التالية:

أ- هل تختلف القيم التقديرية لمستويات الأداء في الاختبار باختلاف الطريقة المستخدمة في التوصل إلى هذه القيم؟

ب- هل تختلف القيم التقديرية لمستويات الأداء في الاختبار باختلاف مستوى كفاية المحكمين عند استخدامهم نفس الطريقة؟

ج- هل هناك تفاعل بين الطريقة المستخدمة في تقدير مستويات الأداء الاختبار وبين مجموعات المحكمين؟ أي هل تختلف القيم التقديرية لمستويات الأداء الناتجة عن استخدام الطرق المختلفة باختلاف مجموعة المحكمين؟

أداة الدراسة:

طبق الباحثان جزء من الاختبار التالي:

Nationally administered Certifying examination in the health) (profession

ويحتوي الاختبار على مائة وثمانين سؤالاً من متعدد الخيارات.

عينة الدراسة:

تتألف عينة المحكمين من ثمان محكمين متخصصين في بناء الاختبارات التأهيلية، ثم قسم المحكمون إلى مجموعتين كل مجموعة تحتوي على أربعة محكمين.

طرق تحديد المعايير:

استخدم الباحثان طريقتي إيبيل وندلسكاي لتحديد درجات القطع.

نتائج الدراسة:

- أ- أدى استخدام الطريقتين إلى الحصول على معايير مختلفة للاختبارات ككل.
- ب- أظهرت المجموعات المختلفة من المحكمين المشاركين في تحديد المعايير مستويات كفاية متشابهة للاختبار ككل.
- ج- أظهرت الدراسة اختلافاً في المعايير الناتجة من كلتا المجموعتين من الحكام.

٢ - دراسة (1980) Ernest N. Skakun and Samuel Kling:

عنوان الدراسة (دراسة مقارنة لطرق تحديد المعايير)

أهداف الدراسة:

تهدف الدراسة إلى مايلي:

- أ- تعديل طريقة إيبيل الخاصة بتحديد المعايير.
- ب- معرفة مدى اختلاف المعايير الناتجة باختلاف الطريقة المستخدمة في التوصل إلى هذه المعايير.
- ج- معرفة أثر درجات القطع المختلفة على جميع تقديرات النجاح.
- د - معرفة مدى ثبات متوسط تقديرات المحكمين لدرجات القطع، وذلك لكل طريقة من طرق تحديد درجة القطع.

أداة الدراسة:

طبق الباحثان (١٤٩) مفردة من مفردات اختبار الجراحة العامة.

طرق تحديد درجة القطع:

استخدم الباحثان كل من الطرق التالية إيبيل الأولى والثانية وندلسكاي.

عينة الدراسة:

بلغ عدد عينة الدراسة (١٦٨) طالباً من طلاب كلية الطب تخصص جراحة عامة، وقسم الباحثين العينة إلى قسمين كان القسم الأول يحتوي على (٤٨) طالباً يشكلون المجموعة المرجعية، والقسم الثاني يحتوي على (١٢٠) طالباً يشكلون المجموعة غير المرجعية، وبلغ عدد عينة المحكمين ثمانية وهم أساتذة ساهموا في إعداد اختبار الجراحة العامة، ومن بين هؤلاء الثمانية شارك سبعة منهم في تحديد درجة القطع وفق طريقة ندلسكاي، وشارك ستة منهم في تحديد درجة القطع وفق طريقتي إيبيل المعدلتين، وخمسة منهم شاركوا في تحديد الدرجة وفق الطرق الثلاث مجتمعة.

نتائج الدراسة:

يتضح من الجدول التالي نتائج درجات القطع والإحصاء الوصفي للاختبار.

إيبيل ٢	إيبيل ١	ندلسكاي	
٦	٦	٧	عدد الحكام
٠،٧٨٤ - ٠،٦٤٤	٠،٧٥٧ - ٠،٦١٣	٠،٧٨١ - ٠،٤٧٧	التقدير
٠،٧١٧	٠،٦٩٧	٠،٦٦٧	المتوسط
٠،٠٥٨	٠،٠٥٦	٠،٠٩٣	الانحراف المعياري لدرجات القطع
١٣٩	١٣٥	١٢٩	درجة القطع

وتدل هذه النتائج على أن استخدام الطرق المختلفة لتحديد درجة القطع يؤدي إلى درجات مختلفة، حيث كانت الدرجات كما يلي ندلسكاي ٦٦،٧% وطريقة إيبيل (١) ٦٩،٧% وطريقة إيبيل (٢) ٧١،٧%، ونتج أيضاً من هذه الدراسة درجة ثبات عالية بالنسبة لمتوسط تقدير المفردات وكذلك بالنسبة لمتوسط درجات القطع.

ويعزى الاختلاف في درجات القطع إلى عوامل عدة:

أ- اختيار الإجابات الخاطئة أيسر من اختيار الإجابات الصحيحة.

ب- تحصر طريقة ندلسكاي المحكمين في عدد محدود من الاحتمالات في الاختبار المتعدد الاختيارات.

ج- يواجه المحكمون في طريقة ندلسكاي مشكلات في فصل معامل الصعوبة الخاص بهم، فيما يتعلق بالوصول إلى إجابات صحيحة.

د- اختلاف مفهوم الطالب المتقن، لأن كل محكم لديه تعريف خاص يقوم على تجربة خاصة مع طلابه.

٣ - دراسة (1983) Gerald Halpin , Gary. Sigmon and Glennelle

Halpin

عنوان الدراسة: (تطبيقات لتحديد صدق درجات القطع باستخدام ثلاث مجموعات مختلفة من المحكمين وفق ثلاث طرق تحكيمية).

أهداف الدراسة:

تهدف هذه الدراسة إلى الإجابة على التساؤلات التالية:-

- ١- هل ينتج من استخدام كل من الطرق التالية إيبيل ونيدلسكاي وأنجوف درجات قطع مختلفة في حال استخدام اختبار مقنن وموحد؟
- ٢- هل تؤثر كفاية المحكمين المختلفة في حال استخدامهم طرق موحدة على معايير الاختبار. أي هل ينتج عن اختلاف كفاياتهم درجات قطع مختلفة وذلك لاختبار موحد؟
- ٣- هل هناك تفاعل بين مجموعة المحكمين أي هل تختلف درجات القطع المحددة بطرق مختلف باختلاف مجموعات المحكمين؟
- ٤- مامدى صدق التطبيقات لنتائج هذه الدراسة؟

أداة الدراسة:

طبق الباحثان اختبار جامعة ميسوري للغة الإنجليزية ويشتمل على (٩٠) سؤالاً من الاختيارات المتعددة، ويهدف إلى تقييم الكفاية في إملاء الكلمات والترقيم وتكوين الجمل وبناء القواعد وبناء الفقرات.

عينة الدراسة:

كانت عينة الدراسة من المحكمين عبارة عن خمسة عشر محكماً، وزعوا إلى ثلاث مجموعات كل مجموعة تحتوي على خمسة أعضاء، وتشتمل المجموعة الأولى على خمسة من طلاب درجة الدكتوراه في تعليم اللغة الإنجليزية، وتشتمل المجموعة الثانية على خمسة من معلمي اللغة الإنجليزية في مدارس المرحلة الثانوية أربعة منهم حاصلون على درجة الماجستير في تعليم اللغة الإنجليزية والخامس حاصل على درجة البكالوريوس، وتشتمل المجموعة الثالثة على خمسة أعضاء من الكلية لديهم خبرة وتدرّيبات في تعليم اللغة الإنجليزية، وبعضهم حاصل على درجة الدكتوراه في الفلسفة والبعض الآخر حاصل على درجة الدكتوراه في التربية.

وكانت عينة الدراسة من الطلاب (١٧٣) طالباً من طلاب جامعة ميسوري.

نتائج الدراسة:

أوضحت الدراسة مايلي:

أ- فروعاً في درجات القطع الناتجة بالطرق المختلفة، وكانت متوسطات درجات القطع الناتجة كما يلي:-

طريقة إيبيل(٦٢,٦١) طريقة نيدلسكي (٤٢,٦٣) طريق أنجوف (٥٦,٠٢).

ب- عندما تضم تلك الطرق المستخدمة فإن تحديد درجات القطع لا يختلف برغم أن مجموعة المحكمين يمثلون مستويات مختلفة من الكفاية.

ج- عند عمليات المقارنة بين المجموعات الأخرى من المحكمين في هذه الدراسة فإن عينة مدرسي الثانوية العامة حددت أدنى درجات باستخدام طريقة نيدلسكاي، وحددت أيضاً أعلى درجات عند استخدامها لطريقة أنجوف.

٤ - دراسة: (Behuanik, et al (1982)

عنوان الدراسة (تطبيقات لصدق تفسير درجة اختبار التمكن باستخدام طريقتي أنجوف ونيدلسكاي).

أهداف الدراسة:

تحدد أهداف الدراسة من خلال الإجابة على التساؤلات التالية وهي:

- Y هل تؤثر الخصائص الديموغرافية المختلفة للمحكمين على درجات القطع؟
- Y هل ينتج عن المجموعات المختلفة من المحكمين درجات قطع مختلفة عند تطبيق طريقة موحدة لتحديد المعايير؟
- Y هل تظهر الطريقتين المختلفتين (نيدلسكاي وأنجوف) درجات قطع مختلفة؟
- Y هل تعتبر الدرجات المحددة من قبل المحكمين الذين استخدموا طريقة أنجوف أكثر تغييراً من الدرجات المحددة بطريقة نيدلسكاي؟
- Y هل تعتبر تقديرات الأسئلة المبنية على التحكيم الفردي في المجموعة ذاتها المستخدمة الطريقة ذاتها، أكثر تشابهاً من تلك التقديرات المحددة من المحكمين في المجموعات المختلفة الذين استخدموا الطريقة ذاتها؟
- Y هل توجد فروق في مستوى إجماع المحكمين بوصفه يتألف من المجموعة ككل، عند استخدامهم طريقتي أنجوف نيدلسكاي لتحديد درجة القطع؟

أداة الدراسة:

عبارة عن اختبارين هديتي المرجع في مادتي القراءة والرياضيات، ويحتوي اختبار مادة الرياضيات على ثلاثة نماذج، كل نموذج يحتوي على (٩٠) سؤالاً لكل سؤال أربعة خيارات. أما بالنسبة لاختبار مادة القراءة فيحتوي على ثلاثة نماذج أيضاً كل نموذج يحتوي على (٨٠) سؤالاً من الأسئلة متعددة الاختيارات، تحتوي على إحدى عشر موضوعاً.

وكان معامل الثبات لاختبار الرياضيات (٩٣،) ولاختبار القراءة (٩١،).

عينة الدراسة:

اشتملت عينة الدراسة على (٤٦٠) طالباً، بالإضافة إلى (١٤) عضواً من أعضاء هيئة التدريس المتخصصين في مادة القراءة و(١٣) عضواً من أعضاء هيئة التدريس المتخصصين في مادة الرياضيات.

نتائج الدراسة:

- ȳ لا توجد علاقة بين خصائص المحكمين الديموغرافية وبين درجات القطع التي حددها.
- ȳ إن الطرق المختلفة لتحديد المعايير أظهرت درجات قطع مختلفة إحصائياً وعملياً.
- ȳ يجب أن لا نتوقع من مجموعات المحكمين المستقلة التي تستخدم طريقة تحديد المعايير ذاتها أن تحدد معايير متشابهة.
- ȳ لا يتوقع من طريقتي أنجوف وندلسكاي اتساقاً في تحديد أكبر أو أصغر درجة قطع متغيرة.
- ȳ يميل المحكمون في المجموعة الواحدة الذين يستخدمون طريقة تحديد معايير موحدة، إلى تحديد أحكام أكثر تشابهاً من المحكمين الذين ينتمون إلى مجموعات مختلفة والذين يستخدمون طريقة تحديد معايير موحدة.

٥ - دراسة (1981) Harasym, P.H

عنوان الدراسة (دراسة مقارنة لطرق تحديد المعايير في مخرجات التقويم)

أهداف الدراسة:

تهدف الدراسة إلى تحديد أي من طريقتي تحديد المعايير ستقوم تجريبياً (نتائج تقييمية متكافئة)، وإذا لم تقدم ذلك إلى أي مدى تكون نتائج تطبيق الطريقتين أفضل من الأخرى.

أداة الدراسة:

صممت ثلاثة اختبارات متوازنة لقياس مواضيع مقرر الغدد الصماء الأيضية (EM)، ويحتوي كل اختبار على نوعين من أنواع الأسئلة، فالنوع الأول عبارة عن اختيار الإجابة الأفضل، وأما النوع الثاني عبارة عن أسئلة من نوع إجابات الصح - الخطأ المتعددة ويحتوي كلا النوعين على أربعة أو خمسة بدائل.

طرق تحديد درجة القطع:

استخدم في هذه الدراسة طريقتي أنجوف المعدلة وندلسكاي، وقد استخدمت طريقة نيدلسكاي لتحديد معايير الأسئلة ذات النوع الأول، واستخدمت طريقة أنجوف المعدلة لتحديد معايير النوع الثاني من الأسئلة.

عينة الدراسة:

اشتملت عينة الدراسة على (٢١٢) طالباً من طلاب السنة الثانية من كلية الطب، وكانت العينة من طلاب السنوات التالية ١٩٧٩م و ١٩٨٠م و ١٩٨١م.

نتائج الدراسة:

- أ- أدى استخدام طريقة نيدلسكاي إلى الحصول على معايير أقل اتساقاً من تلك المعايير الناتجة بطريقة أنجوف المعدلة.
- ب- ظهرت المعايير المنخفضة الناتجة بطريقة نيدلسكاي في عدد كبير من الطلاب المصنفين ضمن الفئة المرضية في طريقة أنجوف المعدلة.
- ج- يعزى سبب الحصول على معايير منخفضة عند استخدام طريقة نيدلسكاي إلى النقص الأساسي الكبير في قيم مستوى الحد الأدنى للنجاح (على سبيل المثال من ٠,١ إلى ٠,٥).

٦ - دراسة: - (1991) Woehr Et al

عنوان الدراسة (مقارنة تجريبية لطرق تحديد درجة القطع المتصلة بالمحتوى والخاصة بتحديد صدق المحك المتصل).

هدف الدراسة:

تهدف هذه الدراسة إلى تقييم مقارنة للطرق المختلفة لتحديد درجات القطع المتصلة بالمحتوى والخاصة بتحديد صدق المحك المتصل، وقد قورنت طرق تحديد درجة القطع استناداً على كل من الخطأ المعياري المتنبأ به في القياس، وعدد تصنيفات القرارات الصحيحة والخطئة الناتجة من كل طريقة من طرق تحديد درجة القطع، والأداء المعياري الحقيقي ككل للأفراد الذين اجتازوا كل درجات القطع.

أداة الدراسة:

صمم اختبار لقياس الكفاية في مقرر علم النفس العام يحتوي على (٣٠) سؤالاً من الخيارات المتعددة.

عينة الدراسة:

كانت عينة الدراسة عبارة عن (٣٧٢) طالباً من طلاب قسم علم النفس، قسّمت تلك العينة إلى مجموعتين:

المجموعة الأولى:

عبارة عن (١٢١) طالباً استخدمت هذه المجموعة كعينة لصدق أو لنمو طرق تحديد درجة القطع المبنية على مايلي: الانحدار وأساس التقدير والمجموعات المتضادة.

المجموعة الثانية:

عبارة عن (٢٥١) طالباً استخدمت هذه المجموعة كعينة للمقارنة الفعلية.

طرق تحديد درجة القطع:

استخدمت في هذه الدراسة سبع طرق مختلفة من طرق تحديد درجة القطع وهي
١- طريقة الانحدار ٢- الطريقة المبنية على أساس التقدير ٣- الطريقة المبنية على انحراف
معياري واحد أقل من المتوسط ٤- الطريقة المبنية على المتوسط ٥- طريقة أنجوف ٦- طريقة
المجموعات المتضادة (أ) ٧- طريقة المجموعات المتضادة (ب).

نتائج الدراسة:

تظهر نتائج الدراسة اختلافات بين طرق تحديد درجات القطع وفق الطرق المختلفة،
وكان متوسط مستويات أداء المحك للأفراد الذين اجتازوا كل درجات القطع متشابهة جداً، لذلك
يجب أن يكون اختيار طريقة تحديد درجة القطع متسقاً مع استخدام الاختبار، ويكون تفسيره
معرفاً بوضوح، وعند اختيار الطريقة يجب أن تكون مصممة بشكل جيد.

٧- دراسة (1983) C.N.Mills:

عنوان الدراسة (مقارنة ثلاثة طرق لتحديد درجة القطع في الاختبارات محكية المرجع).

هدف الدراسة:

تهدف هذه الدراسة إلى مقارنة المعايير المحددة باستخدام الطرق التحكيمية والطرق التي
تتطلب بيانات عن أداء المفحوصين.

طرق تحديد درجة القطع:

استخدمت هذه الدراسة كل من الطرق التالية (أنجوف والمجموعات المتضادة والمجموعة
الحدية).

نتائج الدراسة:

أظهرت نتائج الدراسة أنه عند استخدام ثلاثة طرق أو أكثر فإنه من الممكن الحصول على
بعض التقارب بين طريقتين على الأقل، فعلى سبيل المثال فإن المعايير المحددة بطريقة أنجوف
وطريقة المجموعات المتضادة تكون أكثر اتساقاً مع بعضها البعض من المعايير الناتجة باستخدام
طريقة المجموعة الحدية، ولمزيد من الإيضاح فأنا نجد تبايناً موجوداً في المعايير الناتجة
بالطرق المختلفة.

٨- دراسة (1980) Stephen L. Koffler:

عنوان الدراسة: مقارنة لطرق تحديد معايير التمكن.

هدف الدراسة:

تهدف هذه الدراسة إلى مقارنة معايير التمكن الناتجة من استخدام طريقتي ندلسكاي والمجموعات المتضادة.

أداة الدراسة:

كانت أداة الدراسة عبارة عن مجموعة من اختبارات ولاية نيوجرسي لقياس الحد الأدنى للمهارات الأساسية في القراءة والرياضيات (MBS)، وتتكون أسئلة هذه الاختبارات من الأسئلة ذات الاختيارات المتعددة، وتطبق هذه الاختبارات سنوياً على طلاب المستوى الثالث والسادس والتاسع والحادي عشر.

عينة الدراسة:

اشتملت عينة الدراسة من المحكمين لطريقة ندلسكاي على ٧١ محكماً.

نتائج الدراسة:

لا يوجد اتجاه عام لاختلاف درجات القطع الناتجة من استخدام طريقتي نيدلسكاي والمجموعات المتضادة، وكانت المعايير متناقضة إلى حد بعيد على ثلاث من الاختبارات الثمان، وكان الاختلاف ملحوظاً في اختبارين آخرين.

ملخص الدراسات السابقة:

تعتمد الاختبارات محكية المرجع اعتماداً كلياً على درجة القطع، فمن خلالها يتمكن المعلم من تشخيص مدخلات ومخرجات العملية التربوية، ونظراً لحدثة هذا النوع من القياس، فقد تعددت واختلقت في الوقت ذاته طرق وأساليب تحديد درجة القطع، ومع تعدد هذه الطرق تعددت أيضاً القيم التقديرية لدرجات القطع وأصبح من الصعب اختيار الطريقة المناسبة التي تؤدي إلى قرارات صحيحة.

لذلك اهتمت كثير من الدراسات بمقارنة بعض من طرق تحديد درجة القطع وذلك بهدف تحديد مدى اتفاق أو اختلاف نتائج استخدام تلك الطرق، وتبين من الدراسات السابقة تعدد واختلاف متغيراتها وعدد مفردات الاختبارات المستخدمة وتساؤلاتها وأهدافها، وذلك بهدف تحديد مدى تأثير هذه المتغيرات على القيم التقديرية لدرجات القطع.

فقد أكدت جميع الدراسات السابقة أن القيم التقديرية لدرجات القطع تختلف باختلاف الطريقة المستخدمة عند تطبيقها على اختبار موحد وعينة دراسية موحدة.

كما أكدت دراسات كل من (Behuniak et al,1982:Woehr et al,1991:Koffler,1980) على يكون اختيار طريقة تحديد درجة القطع متسقاً مع هدف الاختبار وأن يكون الاختبار واضحاً ومعرفاً تعريفاً جيداً، وعند تطبيق الطريقة يجب أن تكون مصممة بشكل جيد، وذلك لأن صدق

القرار التربوي يعتمد على كل من الاختيار المناسب لطريقة درجة القطع وعلى كيفية التطبيق، فمن الملاحظ أن طريقتي نيدلسكي والمجموعات المتضادة مثلاً تهدفان إلى تحديد الحد الأدنى من الإلتقان، وإلى تجميع صدق بيانات المجموعات المتقنة، حيث تتطلب طريقة نيدلسكي تقديرات المحكمين حول مفردات الاختبار التي تمثل الحد الأدنى للإلتقان، في المقابل نجد أن طريقة المجموعات المتضادة تتطلب تحكيمات المحكمين حول مستويات الإلتقان بناءً على الأداء الفعلي للطلاب.

كما أوضحت دراستي (C. N. Mills, 1983; Koffler, 1980) أنه في حال استخدام ثلاث طرق فأكثر فإنه من الممكن الحصول على بعض التقارب النسبي بين طريقتين على الأقل على سبيل المثال (أنجوف - المجموعات المتضادة).

كما أشارت دراسة (Jager, 1989) أن كثير من الدراسات السابقة ركزت على الطرق التحكيمية دون غيرها من الطرق حيث لا تخلو دراسة تقريباً من تطبيق إحدى الطرق التحكيمية، وذلك نظراً لأهميتها وسهولة استخدامها وتحليل نتائجها ويتضح ذلك من الجدول التالي:
جدول رقم ١ * تكرارات استخدام طرق تحديد درجة القطع خلال دراسات سابقة:

الطريقة	ندلسكاي	أنجوف	إيبيل	المجموعات المتضادة	المجموعات الحدية	جايقر	إيبيل ذات معامل الصعوبة	أيبيل ذات معامل الصلة
التكرارات	٢٩	٢٠	١٢	٢٢	١١	٢	١	١

• قام البحث بإعداد الجدول أعلاه لتوضيح عدد تكرارات تطبيق الطرق المستخدمة في الفترة ما بين ١٩٧٦م - ١٩٨٤م وذلك لاثنتان وثلاثون دراسة تضمنتها دراسة جايقر (Jager, 1989).

كما بينت دراسات كل من هارسيم، وبهيوينيك، وهالبن (Harasym, 1981; Behuniak, 1982; Halpin et al, 1983) أن تقديرات المحكمين لدرجة القطع وفق طريقة أنجوف حققت أعلى نسبة درجة قطع بين الطرق الأخرى.

كما بينت دراسة كل من سكاكون وكلينق، وهارسيم، وهالبن وآخرون (HARASYM, 1981; SKAKUN&KLING,1980; HALPIN et al, 1983) أن طريقة ندلسكي تؤدي إلى درجات قطع منخفضة، وعزت هذه الدراسات انخفاض درجات القطع وفق هذه الطريقة للأسباب التالية:-
أ - يواجه المحكمون صعوبة في تحكيم المشتتات التي يجب على الطالب المتقن أن يتجنب اختيارها مما يؤدي ذلك إلى تشتت أفكارهم.

ب - اختلاف مفهوم الطالب المتقن لدى المحكمين، لأن كل محكم لديه تعريف مختلف يقوم على تجربة خاصة مع طلابه.

ج - تحصر الطريقة المحكمين في عدد محدود من الاحتمالات غير المستمرة (Irregularly) (١,٠)، (٠,٥٠)، (٠,٣٣)، (٠,٢٥)، (٠,٢٠)، (٠,٠).

د - يجد المحكمون صعوبة في تحديد المشتتات، لأن اختيار المشتتات الخاطئة أيسر من تحديد المشتتات الصحيحة.

وأوضحت دراسة هالبن وآخرون (Halpin, et al (1983) بأن تقديرات فئة المعلمين أدت إلى الحصول على درجات قطع مرتفعة مقارنةً مع فئات المحكمين الأخرى وفق طريقة أنجوف، بينما أدت تقديراتهم إلى الحصول على درجات قطع منخفضة وفق طريقة نيدلسكي. كما أشارت دراسة بيهيونيك (Behuniak, et al (1981) إلى أن المحكمين في فئة واحدة يميلون إلى تحديد تقدير درجات قطع أكثر تشابهاً من المحكمين الذين ينتمون إلى فئات مختلفة عند تطبيق طريقة موحدة وعينة موحدة. وأوضحت دراستي أبو علام (١٩٩١م) وكروس وآخرون (Cross, et al (1984) أن تقديرات المحكمين أظهرت ثباتاً عالياً وفق طريقة أنجوف بينما أظهرت طريقة نيدلسكي ثباتاً أدنى لتقديرات المحكمين. كما أشارت دراستي بيهيونيك و هالبن وآخرون (Behuniak, et al, 1982؛ HALPIN, et al, 1983) أن القيم التقديرية لدرجة القطع تتسق نسبياً إذا استخدمت طريقة واحدة، وطبق الاختبار على عينة موحدة على الرغم من اختلاف مجموعة المحكمين وكفايتهم وخصائصهم الديموغرافية. وأكدت دراسة كل من أندرو وهيتشت و بيهيونيك (Andrew & Hecht, 1976; Behuniak et al, 1981) و(أبو علام، ١٩٩١م) عدم وجود فروق دالة إحصائية بين تقديرات فئات المحكمين المختلفة لدرجات القطع وفق الطرق التحكيمية عند تطبيقها على اختبار موحد وعينة موحدة.

في المقابل أشارت دراستي بيهيونيك وهالبن وآخرون (Behuniak et al, 1981; Halpin et al, 1983) إلى وجود فروق ذات دلالة إحصائية بين مجموعات المحكمين الذين طبقوا طرقاً تحكيمية مختلفة وبين المجموعات التي طبقت طريقة موحدة.

علاقة الدراسة الحالية بالدراسات السابقة:

لقد أفادت الدراسات السابقة في اشتقاق تساؤلات الدراسة الحالية وأهدافها وبعض متغيراتها والطرق الإحصائية المستخدمة، إلا أن هذه الدراسة تختلف عن كثير من الدراسات السابقة في تضمين متغير طول الاختبار، لذلك تهدف هذه الدراسة إلى مقارنة بعض من طرق تحديد درجة القطع (أنجوف - نيدلسكي - المجموعات المتضادة - المجموعات المحكية) وأثر كل من طول الاختبار ونوع المحكم على درجة القطع.

الفصل الثالث: إجراءات الدراسات

فرضيات الدراسة.

منهج الدراسة.

مجتمع وعينة الدراسة.

أدوات الدراسة.

التحليل الإحصائي.

أولاً: فرضيات الدراسة:

- لتحقيق أهداف الدراسة الحالية حاول الباحث اختبار فرضيات الدراسة وهي كالتالي:-
- ١- لا توجد فروق دالة إحصائية عند مستوى دلالة ($\alpha = 0,05$) بين درجات القطع الناتجة عن تطبيق طريقة نيدلسكي على اختبار موحد وعينة دراسية واحدة باختلاف نوع المحكم.
 - ٢- لا توجد فروق دالة إحصائية عند مستوى دلالة ($\alpha = 0,05$) بين درجات القطع الناتجة عن تطبيق طريقة أنجوف على اختبار موحد وعينة دراسية واحدة باختلاف نوع المحكم.
 - ٣- لا توجد فروق دالة إحصائية عند مستوى دلالة ($\alpha = 0,05$) بين درجات القطع الناتجة عن تطبيق طريقة نيدلسكي على اختبار موحد وعينة دراسية واحدة باختلاف طول الاختبار.
 - ٤- لا توجد فروق دالة إحصائية عند مستوى دلالة ($\alpha = 0,05$) بين درجات القطع الناتجة عن تطبيق طريقة أنجوف على اختبار موحد وعينة دراسية واحدة باختلاف طول الاختبار.

ثانياً: منهج الدراسة:

- استخدمت الدراسة الحالية المنهجين التاليين لتحقيق أهداف الدراسة الحالية:
- ١- استخدمت الدراسة منهج البحث الوصفي المقارن لأنها تسعى إلى وصف طرق تحديد درجة القطع المختلفة، كما استخدمت الدراسة منهج الدراسات المقارنة لأنها ركزت على مقارنة درجات قطع الاختبار الناتجة عن تطبيق طرق تحديد درجة القطع في الاختبارات محكية المرجع، وذلك في ضوء متغير طول الاختبار ومتغير نوع المحكم.
 - ٢- استخدمت الدراسة أيضاً منهج البحث شبه التجريبي للحصول على تفسيرات علمية على تساؤلات الدراسة وفرضياتها، وتحليل نتائجها واستخلاص دلالاتها العلمية.

ثالثاً: مجتمع وعينة الدراسة:

جرت الدراسة الحالية على مجتمعين أساسيين هما:-

١ - مجتمع وعينة الطلاب:

اختار الباحث عينة الطلاب بناءً على المعايير التي أشار إليها جايقر وهي كالتالي:

أ- التمثيل Representativeness

أن يشتمل مجتمع العينة على جميع المختصين في المجال المراد قياسه.

ب- العشوائية Randomness

أن يحظى كل فرد من أفراد المجتمع بفرصة ماثلة لاختياره ضمن أفراد العينة.

ج- حجم العينة Sample Size

أن يكون حجم العينة كافياً للقيام بعملية تحديد المعايير (Selina J. 1980).
ويتمثل هذا المجتمع في طلاب الصف السادس الابتدائي بمدارس محافظة جدة الحكومية في الفصل الدراسي الثاني للعام ١٤١٦ - ١٤١٧هـ.
بعد تحديد مجتمع الطلاب قام الباحث باختيار عينة الدراسة من الطلاب عشوائياً على النحو التالي:-

- تحديد أسماء وعناوين جميع المدارس الابتدائية الحكومية الواقعة داخل مدينة جدة.
- اختيار عينة من المدارس بطريقة عشوائية طبقية بناءً على التوزيع الجغرافي لمراكز الإشراف التربوي داخل مدينة جدة، ثم اختيرت عينة عشوائية من الفصول الدراسية من كل مدرسة.

والجدول التالي يوضح عينة المدارس وتوزيعها الجغرافي وعدد عينة الطلاب:

جدول رقم ٢ أسماء مدارس عينة الطلاب وموقعها الجغرافي وعدد طلاب العينة:

مسلسل	المدرسة	الموقع الجغرافي	عدد الطلاب
١	جاير بن حيان	وسط	٧٢
٢	النموذجية الثالثة	وسط	١٠١
٣	السيوطي	شمال *	٣٠
٤	سعد بن أبي وقاص	جنوب	٥٦
٥	الوزيرية	جنوب	٩٦
٦	الأمين	جنوب	٤٥
٧	الإمام النووي	شرق	٥٩
٨	عثمان بن عفان	شرق	١١٢
٩	حبيب بن زيد	شرق	٨٨
	المجموع		٦٥٩

* يقترب مستوى عينة الطلاب الاقتصادي والاجتماعي من مستوى أقرانهم في منطقة الوسط.

٢ - مجتمع وعينة المحكمين:

من أهداف الدراسة الحالية تحديد درجة قطع الاختبار باستخدام الطرق التحكيمية، الأمر الذي يتطلب الاستعانة بمجموعة من المحكمين المتخصصين في تدريس مادة الرياضيات بصورة

عامه، وبتدريس المادة بالصف السادس الابتدائي بصفة خاصة، وذلك للمشاركة في تحديد درجة قطع الاختبار وفق تعليمات الطريقتين التحكيميتين، وكان مجتمع المحكمين كالتالي:

أ- أعضاء هيئة تدريس مادتي طرق تدريس الرياضيات والعلوم بالمؤسسات التعليمية بمدينة جدة ومكة المكرمة.

ب- مدرسو مادة الرياضيات بالصف السادس الابتدائي بمدينة جدة.

ج- مشرفو مادة الرياضيات بمدينة جدة. وكانت عينة المحكمين كالتالي:

أ: أساتذة طرق تدريس مادتي الرياضيات والعلوم:

نظراً لقلّة عدد أفراد مجتمع أساتذة طرق تدريس مادتي الرياضيات والعلوم في المؤسسات التربوية بمدينة جدة ومكة المكرمة قام الباحث باختيار جميع أفراد المجتمع للقيام بمهمة التحكيم. كما هو موضح في الجدول رقم (٣).

ب: معلمو مادة الرياضيات:

اختار الباحث عينة عشوائية طبقية من معلمي مادة الرياضيات بالصف السادس بناءً على ما يلي:

- التوزيع الجغرافي لمراكز الإشراف التربوي داخل مدينة جدة.
- تقديرات الكفاية المرتفعة، وذلك للحصول على تقديرات دقيقة لدرجات قطع الاختبارات.

ج: مشرفو مادة الرياضيات:

قام الباحث باختيار جميع أفراد مجتمع مشرفو مادة الرياضيات بمدينة جدة. والجدول التالي يوضح أنواع المحكمين وعددهم والمؤسسات التربوية التي ينتمون إليها. جدول رقم ٣ عدد المحكمين ووظائفهم والمؤسسات التربوية التي ينتمون إليها:

م	المؤسسة التربوية	المدينة	الوظيفة	العدد
١	جامعة أم القرى	مكة المكرمة	عضو هيئة تدريس	٤
٢	جامعة الملك عبد العزيز	جدة	عضو هيئة تدريس	٢
٣	كلية المعلمين	جدة	عضو هيئة تدريس	٤
٤	الإدارة العامة للتعليم	جدة	مشرف تربوي	١٠
٥	الإدارة العامة للتعليم	جدة	معلم	٢٠
٦	المجموع الكلي			٤٠

رابعاً: أدوات الدراسة:

لتحقيق أهداف الدراسة استخدم الباحث الأدوات التالية: -

١- اختبار محكي المرجع يقيس الكفايات الرياضية في العمليات الحسابية على الأعداد بالمرحلة الابتدائية إعداد بشناق (١٤١٢هـ).

ويتصف هذا الاختبار بأنه استخدم في بناء الطرق والأساليب المتعارف عليها في بناء الاختبارات محكية المرجع ويتكون الاختبار من (٩٥) فقرة من نوع الاختيار المتعدد، ويتكون الاختبار من ثلاثة اختبارات فرعية ملحق رقم (١) على النحو التالي:

أ- الاختبار الفرعي الأول:

يتكون من (٣٣) فقرة تهدف إلى قياس كفاية جمع الأعداد وطرحها.

ب- الاختبار الفرعي الثاني:

يتكون من (٤٠) فقرة تهدف إلى قياس كفاية ضرب الأعداد وقسمتها.

ج- الاختبار الفرعي الثالث:

يتكون من (٢٢) فقرة تهدف إلى قياس كفاية خصائص العمليات الحسابية على الأعداد.

صدق الاختبار:

استخدم بشناق (١٤١٢هـ) الصدق الوصفي لتقدير صدق الاختبار، حيث قام بتحليل المفردات بهدف تعيين صدق مفردات الاختبار مستخدماً في ذلك طريقة التمايز اللغوي وتهدف هذه الطريق إلى تحديد مدى ملائمة وارتباط المفردات بالأهداف التي تقيسها، وذلك من خلال الاستعانة بمجموعة من المحكمين المتخصصين في القياس والتقويم والمحتوى الدراسي المراد قياسه.

ثبات الاختبار:

أعتمد بشناق (١٤١٢هـ) في تحديد ثبات الاختبارات الفرعية الثلاث على الطرق الثلاث الموضحة بالجدول التالي:-

جدول رقم ٤ القيم الناتجة من طرق تحديد معاملات الثبات للاختبارات الفرعية الثلاثة:

الاختبار	عدد المفردات	ألفا	ليفنجستون	هاريس
الأول	٣٣	٨٣	٨٤	٧٨
الثاني	٤٠	٦٨	٩٢	٧٤
الثالث	٢٢	٦٦	٩٣	٦٣

على الرغم من أن دراسة بشناق (١٤١٢هـ) لم تشر إلى قيم معامل ثبات الاختبار الكلي (٩٥) مفردة، إلا أنه يتضح من الجدول السابق أن جميع قيم معاملات الثبات للاختبارات الفرعية المحددة وفق الطرق المختلفة مناسبة، ويمكن الحكم من خلالها على ثبات الاختبارات الفرعية، ويلاحظ من الجدول أيضاً اختلاف قيم معاملات الثبات، حيث حقق معامل (ليفنجستون) أعلى قيمة للاختبارات الفرعية الثلاث، بينما تفاوتت قيم معامل الثبات المحددة وفق معاملي (ألفا وهاريس).

وللتأكد من ثبات الاختبار قام الباحث بحساب معامل ثبات (ألفا) كرنباخ، وذلك للاختبارات الفرعية الثلاث والاختبار الكلي، والتي كانت كالتالي:
الاختبار الفرعي الأول (٠,٨٠) والاختبار الفرعي الثاني (٠,٠٨٥) والاختبار الفرعي الثالث (٠,٠٧) والاختبار الكلي (٠,٩١)، يلاحظ من هذه القيم ارتفاع معامل ثبات الاختبار الكلي الأمر الذي يدل على اتسام الاختبار الكلي بالثبات، ويتضح أيضاً وجود علاقة طردية بين قيم معامل ثبات (ألفا) وطول الاختبارات الفرعية، حيث ترتفع قيمة معامل الثبات بزيادة عدد مفردات الاختبار، وهذا ناتج طبيعي.

تحديد مفردات اختبارات الدراسة الحالية:

حيث إن أهداف الدراسة الحالية تختلف عن أهداف دراسة بشناق (١٤١٢هـ)، لذا قام الباحث بتطبيق الاختبار الكلي (٩٥) مفردة على عينة الطلاب بلغت (٦٥٩) طالباً، وتمت عملية تطبيق الاختبار وفق التعليمات المحددة، وذلك في الفترة الأخيرة من الفصل الدراسي الثاني للعام الدراسي ١٤١٦-١٤١٧هـ، وبعد الانتهاء من عملية تصحيح الإجابات تمّ تفرغها في الاستمارات المعدة، ثم استعان الباحث بمركز المعلومات والحاسب الآلي بجامعة أم القرى باستخدام برنامج (SPSS) لإجراء التحليلات الإحصائية المناسبة، وكانت نتائج التحليلات الإحصائية للاختبار الكلي (٩٥) مفردة كما يلي:-

١ - المتوسطات والانحراف المعياري:

لتحليل إجابات عينة الطلاب على الاختبار قام الباحث بحساب متوسط درجات الطلاب وانحرافها المعياري وذلك للاختبار الكلي (٩٥) مفردة وللاختبارات الفرعية الثلاثة التي تقيس الكفايات الرياضية المختلفة، والجدول التالي يوضح نتائج هذه التحليلات الإحصائية.

جدول رقم ٥ نتائج التحليلات الإحصائية المبدئية من تطبيق الاختبار على أفراد العينة:

الاختبار	عدد المفردات	المتوسط	الانحراف المعياري	الكفايات	عدد الطلاب
الفرعي الأول	٣٣	٢٣,٩	٤,٧	الجمع والطرح	٦٥٩
الفرعي الثاني	٤٠	٢٠,٨	٦,٨	الضرب والقسمة	٦٥٩
الفرعي الثالث	٢٢	٩,٥	٣,٩	خصائص العمليات الحسابية	٦٥٩
الكلي	٩٥	٥٤,١٤	١٣,٣	جميع الكفايات	٦٥٩

يتضح من الجدول السابق ارتفاع متوسط درجات الطلاب في الاختبار الفرعي الأول مقارنةً بالاختبارين الفرعيين الآخرين.

٢: تحديد طول الاختبار:

يعتبر متغير طول الاختبار من متغيرات الدراسة الأمر الذي يتطلب تعيين اختبارات مختلفة في عدد مفرداتها لقياس أثر المتغير على طرق تحديد درجة القطع المستخدمة في الدراسة. ونظراً لطول مفردات الاختبار الكلي (٩٥) مفردة، وتقديراً لإجهاد المحكمين المشاركين في تحكيم مفردات الاختبار، قام الباحث بانتقاء عينة قسدية من مفردات الاختبار الذي سبق تطبيقه على عينة الدراسة وذلك بناءً على المعايير الثلاثة التالية:-

أ - مستويات معامل الصعوبة المختلفة (صعبة جداً - صعبة - صعبة لحد ما - متوسطة الصعوبة - سهلة لحد ما - سهلة جداً) ملحق رقم (٢).

ب - قيمة معامل الثبات (a) إذا استثنيت المفردة من مجموع المفردات الكلي (Alpha if item Deleted)

ج - معامل ارتباط المفردة المصحح مع مجموع الارتباط الكلي للمفردات (Corrected Item Total Correlation)

نتج عن ذلك حذف (٢٤) مفردة واختيار أفضل (٧١) مفردة من الاختبار الكلي تمثل الاختبار بصورة مبدئية ملحق رقم (٣).

والجدول التالي يوضح عدد المفردات وقيم معامل ثبات (a) لجميع مستويات الصعوبة المختلفة، وذلك للمفردات المختارة (٧١) مفردة التي تمثل الاختبار بصورته المبدئية.

جدول رقم ٦ تصنيفات مستويات الصعوبة وعدد المفردات ومعامل الثبات (a):

التسلسل	تصنيف المفردة	مستوى الصعوبة	a	عدد الأسئلة
١	صعبة جداً	٩٠ - ٨٠	٠,٤٣	٥
٢	صعبة	٧٩ - ٧٠	٠,٤٠	١١
٣	صعبة لحد ما	٦٩ - ٦١	٠,٥٠	١٢
٤	متوسطة الصعوبة	٦٠ - ٥٠	٠,٦١	١١
٥	سهلة لحد ما	٤٠ - ٣٠	٠,٧٥	١٢
٦	سهلة	٢٩ - ٢٠	٠,٨٠	١١
٧	سهلة جداً	١٩ - ١٠	٠,٦٥	٩
٨	المجموع الكلي	٩٠ - ١٠		٧١

ثم حدّد الباحث طول اختبارات الدراسة الحالية الثلاث وهي: اختبار طويل يتكون من (٣٥) مفردة، واختبار متوسط يتكون من (٢١) مفردة، واختبار قصير يتكون من (١٤) مفردة. ولتيسير عملية تحكيم مفردات الاختبار قام الباحث أيضاً باختيار أفضل (٣٥) مفردة من مفردات الاختبار بصورته المبدئية (٧١) مفردة بناءً على معاملات صعوبة المفردات، وبذلك تمكّن الباحث من تحديد مفردات الاختبار في صورته النهائية. ونظراً لإمكانات الباحث المحدودة، ولقلة عينة أساتذة طرق تدريس الرياضيات، ولصعوبة تطبيق ثلاثة اختبارات لكل طريقة تحكيمية على جميع أفراد عينة المحكمين (أي بواقع ستة اختبارات لكل محكم)، اكتفى الباحث بتطبيق الاختبار الطويل (٣٥) مفردة على جميع أفراد عينة المحكمين، ومن ثمّ اختار من مفرداته بناءً على مستويات صعوبة المفردة مفردات الاختبارين المتوسط (٢١) مفردة والقصير (١٤) مفردة، وبذلك تمكّن الباحث من تحديد مفردات اختبارات الدراسة الثلاثة بصورتها النهائية وهي:

١ - الاختبار الطويل

يتكون من (٣٥) مفردة تمثّل جميع مستويات الصعوبة بواقع (٥) مفردات لكل مستوى من المستويات السبعة، ويتمثّل الاختبار في أرقام المفردات التالية المنتقاة من الاختبار الكلي (٩٥) مفردة وهي: صعبة جداً (٢٣-٢٤-٢٤-٦٤-٦٨-٦١) صعبة (٦٦-٩٠-٧٢-٩١-٨٣) صعبة لحد ما (٧١-٩٢-٩٦-٤٨-٥٠) متوسطة الصعوبة (٢٢-٢١-٤٤-٧٠-٥٩) سهلة لحد ما (٥٧-٦٣-٤٥-٤٦-٥٢) سهلة (٢٠-١٥-١٦-١٨-١٧) سهلة جداً (١٩-٤٠-٣٧-٣٨-٣٦).

والجدول التالي يوضح مستويات صعوبة مفردات الاختبار الطويل (٣٥) مفردة بصورته النهائية.

جدول رقم ٧ مستويات صعوبة مفردات الاختبار الطويل (٣٥) مفردة:

التسلسل	تصنيف المفردة	مستوى الصعوبة	عدد الأسئلة
١	صعبة جداً	٩٠-٨٠	٥
٢	صعبة	٧٩-٧٠	٥
٣	صعبة لحد ما	٦٩-٦١	٥
٤	متوسطة الصعوبة	٦٠-٥٠	٥
٥	سهلة لحد ما	٤٠-٣٠	٥
٦	سهلة	٢٩-٢٠	٥
٧	سهلة جداً	١٩-١٠	٥
٨	الاختبار الفرعي الطويل	٩٠ - ١٠	٣٥

٢ - الاختبار المتوسط:

يتكون من (٢١) مفردة تمثل جميع مستويات الصعوبة بواقع (٣) مفردات لكل مستوى من المستويات السبعة، ويتمثل الاختبار في أرقام المفردات التالية المنتقاة من الاختبار الكلي (٩٥) مفردة وهي كما يلي: صعبة جداً (٢٣-٢٤-٦٤) صعبة (٦٦-٧٢-٨٣) صعبة لحد ما (٧١-٩٢-٤٨) متوسطة الصعوبة (٢٢-٢١-٥٩) سهلة لحد ما (٥٧-٤٥-٥٢) سهلة (١٦-١٨-١٧) سهلة جداً (١٩-٣٨-٣٦).

والجدول التالي مستويات صعوبة مفردات الاختبار المتوسط (٢١) مفردة بصورته النهائية.

جدول رقم ٨ مستويات صعوبة مفردات الاختبار المتوسط:

التسلسل	تصنيف المفردة	مستوى الصعوبة	عدد الأسئلة
١	صعبة جداً	٩٠-٨٠	٣
٢	صعبة	٧٩-٧٠	٣
٣	صعبة لحد ما	٦٩-٦١	٣
٤	متوسطة الصعوبة	٦٠-٥٠	٣
٥	سهلة لحد ما	٤٠-٣٠	٣
٦	سهلة	٢٩-٢٠	٣
٧	سهلة جداً	١٩-١٠	٣
٨	الاختبار الفرعي المتوسط	٩٠ - ١٠	٢١

٣ - الاختبار القصير:

يتكون من (١٤) مفردة تمثل جميع مستويات الصعوبة بواقع مفردتين لكل مستوى من المستويات السبعة، ويتمثل الاختبار في أرقام المفردات التالية المنتقاة من الاختبار الكلي (٩٥) مفردة وهي كما يلي: صعبة جداً (٢٣-٢٤) صعبة (٦٦-٧٢) صعبة لحد ما (٧١-٤٨) متوسطة الصعوبة (٢٢-٢١) سهلة لحد ما (٤٥-٥٢) سهلة (١٨-١٧) سهلة جداً (٣٨-٣٦).
والجدول التالي مستويات صعوبة مفردات الاختبار القصير (١٤) مفردة بصورته النهائية.

جدول رقم ٩ مستويات صعوبة مفردات الاختبار القصير:

التسلسل	تصنيف المفردة	مستوى الصعوبة	عدد الأسئلة
١	صعبة جداً	٩٠-٨٠	٢
٢	صعبة	٧٩-٧٠	٢
٣	صعبة لحد ما	٦٩-٦١	٢
٤	متوسطة الصعوبة	٦٠-٥٠	٢
٥	سهلة لحد ما	٤٠-٣٠	٢
٦	سهلة	٢٩-٢٠	٢
٧	سهلة جداً	١٩-١٠	٢
٨	الاختبار الفرعي القصير	٩٠-١٠	١٤

ثبات الاختبارات الثلاث:

للتأكد من ثبات الاختبارات قام الباحث بحساب معامل ثبات ألفا (α) (كرونباخ) وذلك للاختبارات الثلاثة، وكانت قيم معامل الثبات كما يلي:
جدول رقم ١٠ قيم معامل ثبات (a) للاختبارات الثلاثة:

الاختبار	عدد المفردات	قيم معامل ثبات (α)
الطويل	٣٥	٠,٨٠
المتوسط	٢١	٠,٧٣
القصير	١٤	٠,٦٧

يتبين من الجدول السابق أن جميع قيم معاملات الثبات للاختبارات الثلاثة مناسبة ويمكن الحكم من خلال هذه القيم على اتسام الاختبارات بالثبات، ويلاحظ من الجدول أيضاً اختلاف قيم معاملات الثبات باختلاف طول الاختبار، حيث ترتفع قيمة معامل الثبات بارتفاع عدد مفردات الاختبار وهذا ناتج طبيعي.

تحديد درجة قطع الاختبارات وفق الطريق المختلفة:

١ - تحديد درجة القطع وفق طريقة نيدلسكاي:

أسترشد الباحث بما أشير إليه في الإطار النظري حول تصميم هذه الاستمارة بهدف تحديد تقديرات المحكمين لمفردات المقياس، وذلك بعد تحديد كل من أطوال الاختبارات المختلفة ومعامل صعوبة مفردات الاختبار، وتم تحديد تقديرات المحكمين بعد فحص كل محكم مفردات الاختبار، ثم طلب منهم تصور مجموعة من طلاب الصف السادس الابتدائي لديهم الحد الأدنى المقبول من الكفاية الرياضية، ثم طلب منهم أيضاً تحديد البدائل الخاطئة التي من الممكن أن يستبعد اختيارها هؤلاء الطلاب لأنها لا تمثل في نظرهم الإجابة الصحيحة للمفردة، وبذلك يكون الحد الأدنى لمستوى اجتياز المفردة هو مقلوب عدد البدائل المتبقية للمفردة ملحق رقم (٤).

ثم تحقق الباحث من ثبات تقديرات المحكمين وفق هذه الطريقة باستخدام طريقة سبيرمان براون (Eble, 1951) حيث بلغت قيمته (٠,٤٥) وتعد هذه القيمة مقبولة، ولعل ذلك يعزى إلى صعوبة تحكيم البدائل التي يجب على الطالب ذو الحد الأدنى من الكفاية أن يتجنب اختيارها، بالإضافة إلى اختلاف مفهوم الطالب ذو الحد الأدنى من الكفاية لأن كل محكم لديه مفهوم خاص به. والجدول التالي يوضح قيم درجة القطع الناتجة عن اختلاف نوع المحكم وطول الاختبار.

جدول رقم ١١ تكرارات ونسب درجة قطع الاختبار الطويل (٣٥) مفردة الناتجة وفق طريقة نيدلسكاي باختلاف نوع المحكم وطول الاختبار:

نوع وعدد المحكمون	درجة القطع	النسبة	طول الاختبار	درجة القطع	النسبة
أعضاء هيئة التدريس (١٠)	٢٥	٧١,٤	طويل ٣٥ مفردة	٢٠	٥٧,١
المشرفون (٢٠)	١٨	٥١,٤	متوسط ٢١ مفردة	١٢	٥٧,١
المعلمون (١٠)	١٨	٥١,٤	قصير ١٤ مفردة	٨	٥٧,١

٢ - تحديد درجة القطع وفق طريقة أنجوف:

أسترشد الباحث أيضاً بما أشير إليه في الإطار النظري حول تصميم هذه الاستمارة حيث تشتمل على تعليمات وأرقام مفردات الاختبار، وكُتبت أمام كل مفردة حرف مناظر للإجابة الصحيحة للمفردة داخل المربع، وكذلك ميزان يشتمل على عشر فترات متساوية تقريباً (١ = صفر - ١٠)، (٢ = ١١ - ٢٠)، (٣ = ٢١ - ٣٠)، (٤ = ٣١ - ٤٠)، (٥ = ٤١ - ٥٠)، (٦ = ٥١ - ٦٠)، (٧ = ٦١ - ٧٠)، (٨ = ٧١ - ٨٠)، (٩ = ٨١ - ٩٠)، (١٠ = ٩١ - ١٠٠)، ثم طلب من كل محكم تقدير احتمال أن يعرف طالب الصف السادس الذي يمتلك الحد الأدنى من الكفاية المطلوبة إجابة المفردة دون أن يلجأ إلى تخمين الإجابة، ثم يضع دائرة حول الرقم الذي

يمثل هذه القيمة على الميزان المعطى، ثم تُجمع القيم الاحتمالية المقدره لكل مفردة، يلي ذلك إيجاد متوسط هذه القيم لجميع المحكمين ولجميع مفردات الاختبار ملحق رقم (٨).

٣ - طريقة المجموعات المتضادة:

لتحديد درجة القطع للاختبارات الثلاثة وفق هذه الطريقة استرشد الباحث بما أشير إليه في الفصل الثاني (ص ص ٣٩، ٣٨) حيث قام بتحديد مجموعتين من الطلاب الأولى متقنة والثانية غير متقنة بناءً على الأداء الفعلي (Actual data) لعينة الدراسة من الطلاب في مادة الرياضيات بالصف السادس الابتدائي، حيث استعان بالحاسب الآلي الشخصي باستخدام برنامج الرزمة الإحصائية للعلوم الاجتماعية (spss) لتحديد محك يصنف الطلاب إلى فئتين متقنة وغير متقنة للكفايات الرياضية، لذا قام بحساب معامل ارتباط (بيرسون) بين درجات جميع أفراد عينة الطلاب (٦٥٩) طالباً في مادة الرياضيات خلال الفصلين الدراسيين للعام الدراسي ١٤١٦ هـ / ١٤١٧ هـ وبين درجاتهم الكلية في الاختبار الطويل (٣٥) مفردة، حيث تبين أن معامل الارتباط بين مجموع اختباري الفصلين (٧٠) درجة وبين الاختبار الطويل (٣٥) مفردة الأعلى ارتباطاً (م = ٠,٦) وبذلك استبعد الباحث مجموع درجات أعمال الفصلين الأول والثاني لتدني معامل ارتباطهما، ثم قام بتحديد عينة قصدية تمثل طلاب المجموعة الأولى (التي أتقنت الكفايات الرياضية) بناءً على درجاتهم المرتفعة (< أو = ٦٤ درجة) أي الحاصلين على نسبة (٩١,٤٣%) فأعلى في اختبار الفصلين الدراسيين وبذلك أصبح عدد أفراد عينة المتقنين ٤٥ طالباً، ثم حدّد طلاب المجموعة الثانية (التي لم تتقن الكفايات الرياضية) بناءً على درجاتهم المتدنية (> أو = ٢٣ درجة) أي الحاصلين على نسبة (٣٢,٨٦%) فأدنى في اختبار الفصلين وبذلك أصبح عدد أفراد عينة غير المتقنين ٤٤ طالباً، ثم حصل على مجموع درجات المجموعتين المتقنة وغير المتقنة المحددتين وفق المحك السابق ذكره في الاختبارات الثلاثة، ثم قام بتمثيل توزيع مجموع تكرار الدرجات الكلية للمجموعتين بيانياً، وبذلك حصل على درجة قطع الاختبارات الثلاثة من تقاطع المنحنيات، كما هو موضح بالملاحق (١٢، ١٣).

والجدول التالي يوضح درجات القطع الناتجة للاختبارات الثلاثة المختلفة.

جدول رقم ١٢ تكرارات ونسب درجة القطع الناتجة وفق طريقة المجموعات المتضادة باختلاف طول الاختبار:

طول الاختبار	عدد المفردات	درجة القطع	النسبة
طويل	٣٥	٢٣	٦٥,٧
متوسط	٢١	١٢	٥٧,١
قصير	١٤	٨,٥	٦٠,٧

٤ - طريقة المجموعات المحكية:

استرشد الباحث بما أشير إليه في الفصل الثاني لتحديد درجة القطع (ص ص ٣٧، ٣٨) ثم قام بتحديد فئتين من الطلاب الأولى متقنة والثانية غير متقنة معتمداً في ذلك على الخطوات السابقة التي قام بها في المجموعات المتضادة مع اختلاف في درجة المحك ($<$ أو $= ٥٦$ درجة) أي الطلاب الحاصلين على نسبة (٨٠%) فأعلى، ($>$ أو $= ٢٨$ درجة) أي الطلاب الحاصلين على نسبة (٤٠%) فأدنى، وذلك بهدف زيادة عدد أفراد العينة القصدية (لتحقيق أحد شروط تطبيق الطريقة حيث يجب أن يزيد عدد أفراد العينة الواحدة عن مائة طالب)، وبذلك أصبح عدد المجموعة المتقنة ١٠٩ طالباً والمجموعة غير المتقنة ١١٠ طالباً، ولتحديد درجة القطع للاختبارات المختلفة قام الباحث بإيجاد مجموع درجات المجموعتين المتقنة وغير المتقنة لجميع اختبارات الدراسة الثلاثة (طويل - متوسط - قصير) بناءً على درجات المحك، ثم استعان بالحاسب الآلي الشخصي باستخدام برنامج الحزمة الإحصائية للعلوم الاجتماعية للتمثيل توزيع مجموع تكرار درجات المجموعتين الكلية بيانياً، وبذلك حصل على درجات القطع لجميع الاختبارات الثلاثة بصورتها الأولية من نقاط تقاطع منحنيات التوزيعات التكرارية، كما هو موضح في الملاحق رقم (١٤، ١٦).

ثم قام الباحث بحساب الاحتمالات والمعادلات التالية لجميع الاختبارات الثلاثة وهي كما يلي:

أ: حساب نتائج الاحتمالات الأربعة التالية:

- طلاب متقنون فعلاً True Master

- طلاب متقنون غير حقيقيين False Master

- طلاب غير متقنين فعلاً True Non Master

- طلاب غير متقنين غير حقيقيين False Non Master

وتمّ ذلك بتقسيم مجموع عدد أفراد العينة (٢١٩) طالباً على تكرار كل احتمال من الاحتمالات الأربعة.

ب: تم حساب نسبة احتمالات القرارات الصحيحة وفق الصيغة التالية:-

القرارات الصحيحة = احتمال (طلاب متقنين حقيقيين) + احتمال (طلاب غير متقنين حقيقيين)

P (TN)

P (TM)

ج: تم حساب نسبة احتمالات القرارات الخاطئة وفق الصيغة التالية:-

القرارات الخاطئة = احتمال (طلاب غير متقنين غير حقيقيين) + احتمال (طلاب متقنين غير حقيقيين)

P (FM)

P (FN)

د: تمّ حساب معامل صدق درجات القطع المحتملة لقياس مدى تقدير التصنيف المتنبأ به (بناءً على درجة القطع المحتملة) ويتم ذلك وفق الصيغة التالية:

$$\phi = \frac{P (TM) - BR (SR)}{BR (1- BR) SR (1 - SR)}$$

حيث إن:-

• BR (Base Rate) = احتمال الإلتقان في المجموعة

• 1 - BR احتمال عدم الإلتقان في المجموعة

• SR = احتمال المتقنين المتنبأ بهم في المجموعة

• 1 - SR = احتمال غير المتقنين المتنبأ بهم

وتمّ حساب الاحتمالين BR - SR وفق الصيغة التالية:-

$$BR = P (FN) + P (TM)$$

$$SR = P (TN) + P (FM)$$

وتعتبر القيمة (+) أعلى درجة لمعامل الصدق

خامساً:- تم حساب المنفعة والضرر النسبيين لأخطاء التصنيف.

ويحسب الضرر المتوقع (&) للقرار الواحد (K) وفق الصيغة التالية:-

$$K = P (FN) (D1) + P (FM) (D2)$$

حيث إن:

$$D1 = 1 - u_1, D2 = u_2 - 1$$

ثم قام الباحث بحساب الفائدة المتوقعة (u) للقرار الواحد (K) وفق الصيغة التالية:

$$K = P(TM)(u_1) + P(TN)(u_2)$$

$$حيث إن: u_1 = 1, u_2 = u$$

وتمّ حساب الضرر والفائدة المتوقعتين لكل درجة قطع محتملة، وبذلك حدد الباحث جميع

درجات القطع للاختبارات المختلفة، كما هو موضح بالملاحق (١٥، ١٧، ١٨).

والجدول التالي يوضح درجات القطع الناتجة للاختبارات الثلاثة.

جدول رقم ١٣ درجات القطع الناتجة وفق طريقة المجموعات المحكية باختلاف طول الاختبار:

طول الاختبار	عدد المفردات	درجة القطع	النسبة
طويل	٣٥	٢٠	٥٧,١
متوسط	٢١	١٢	٥٧,١
قصير	١٤	٨	٥٧,١

والجدول التالي يبين مميزات وعيوب الطرق المستخدمة في الدراسة.

جدول رقم ١٤ مميزات وعيوب الطرق المستخدمة في الدراسة:

الطريقة	المميزات	العيوب
نييلسكي	سهلة الفهم والتطبيق ولا تتطلب استخدام أساليب إحصائية معقدة.	لا يوجد مبرر يؤيد افتراض أن الطالب الذي لا يستطيع الإجابة بصورة صحيحة عن سؤال معين سيتمكن من الإجابة عن طريق التخمين العشوائي، لأن بدائل الأسئلة تصمم عادة لجذب الطلاب غير المتقنين، كما أن هذه الطريقة تحتاج إلى جهد كبير لتدريب المحكمين كي يستخدمونها بشكل مناسب، بالإضافة إلى أنها تسمح فقط بقيم احتمالية منفصلة تعتمد على بدائل الإجابات.
أنجوف	سهلة للتنفيذ والفهم وتعتمد على تصور مجموعة من الأفراد الذين حققوا الحد الأدنى للكفاية المطلوبة بدلاً من تصور فرداً واحداً.	يصعب على كثير من المحكمين تعريف الطلاب ذوي الحد الأدنى للكفاية، كما يصعب تقدير القيم الاحتمالية لبعض المفردات المصاغة في عبارات منفية أو المتضمنة على عبارات رياضية.
المجموعات المتضادة	تعتبر نموذج للصدق التكويني والبنائي، وهي ذات حساسية خاصة لخصائص المفردة الإحصائية.	لا تحدد المعيار الحقيقي وفي أفضل الحالات تحدد موقع الأداء على متصل الدرجات حيث يتعذر التمييز بين المجموعتين المتقنة وغير المتقنة، كما يصعب التجاوز عن الخطأ الناجم من المحك.
المجموعات المحكية	تعتبر نموذج للصدق التكويني والبنائي، وهي ذات حساسية خاصة لخصائص المفردة الإحصائية.	لا تحدد المعيار الحقيقي وفي أفضل الحالات تحدد موقع الأداء على متصل الدرجات حيث يتعذر التمييز بين المجموعتين المتقنة وغير المتقنة.

خامساً: التحليلات الإحصائية:-

لمعالجة بيانات الدراسة استخدم الباحث تحليلات الإحصائية التالية:-

١ - التكرارات والنسب المئوية:

للمقارنة بين درجات القطع الناتجة وفق الطرق المختلفة باختلاف متغيري الدراسة (طول الاختبار ونوع المحكم).

٢ - المتوسط الحسابي والانحراف المعياري:

لوصف نتائج أداء عينة الطلاب في الاختبار الرئيس (٩٥) مفردة.

٣ - معامل ثبات (a) كرونباخ:

استخدم معامل ثبات (α) كرونباخ لتحديد قيمة معامل ثبات الاختبارات المختلفة المطبقة على عينة الدراسة من الطلاب.

٤ - معامل ارتباط المفردة المصحح مع مجموع الارتباط الكلي للمفردات .Corrected item total Correlation

للمساهمة في اختيار مفردات الاختبار الطويل المكون من (٣٥) مفردة من الاختبار الرئيس (٩٥) مفردة.

٥ - معامل صعوبة المفردات Item Difficulty:

للمساهمة أيضاً في اختيار مفردات الاختبارات الثلاثة (الطويل - المتوسط - القصير) من الاختبار الرئيس (٩٥) مفردة، تمّ حساب معامل صعوبة المفردات وفق الصيغة التي ذكرها (Berk, 1982, P.59) وهي كما يلي:

$$\text{Diff} = \frac{C}{N} \times 100$$

حيث إن:

$$C = \text{عدد الإجابات الخاطئة.} \quad N = \text{العدد الكلي للطلاب.}$$

٦ - تحديد درجة القطع وفق طريقة نيدلسكي:

لتحديد درجة قطع الاختبار وفق هذه الطريقة استخدم الباحث الصيغة التالية التي أشار إليها (Nedelsky, 1954, PP.5-7):

$$MPL = \bar{MFD} + KsFD$$

$$\bar{MFD} = \frac{\sum FD}{N}$$

$$sFD = \sqrt{\frac{\sum (MFD - \bar{MFD})^2}{N}}$$

حيث إن:

MPL تمثل درجة القطع، N تمثل عدد المحكمين.

MFD تمثل متوسط الحد الأدنى المحدد من المحكمين.

K تمثل المقدار الثابت المحدد من المحكمين.

σ FD تمثل الانحراف المعياري لدرجات القطع.

٧ - تحديد درجة القطع وفق طريقة المجموعات المحكية:

لتحديد درجة قطع الاختبار وفق هذه الطريقة استخدم الباحث الصيغ (نظراً لكثرة المعادلات المستخدمة لتحديد درجة القطع وفق هذه الطريقة) التي أشار إليها بالتفصيل (ص ٧٢-٧٣).

٨ - تحديد ثبات تقديرات المحكمين (Estimation of The Reliability of

Ratings)

قام الباحث بتحديد ثبات تقديرات المحكمين لطريقتي نيدلسكي وأنجوف وفق الصيغة التي أشار إليها إيبيل (Ebel (1951 وهي كما يلي:

$$r_k = \frac{M\bar{X} - M}{M\bar{X}}$$

حيث أن:

rk = معامل ثبات تقديرات المحكمين.

Mx = متوسط مربع المفردات (يقدر من تباين متوسطات تقديرات المحكمين للمفردات).

M = متوسط مربع الخطأ (يقدر من تباين خطأ تقديرات المحكمين).

٩ - أسلوب تحليل التباين الأحادي (One way Anova (Unequal Sample

Size)

طبق الباحث أسلوب تحليل التباين أحادي الاتجاه للعينات غير المتساوية لاختبار فرضيات الدراسة وفق الصيغة التي أشار إليها واينر (Winer (1971 وهي كالتالي:

$$F = \frac{MS_{treat}}{MS_{Error}}$$

حيث إن:

F = اختبار ف.

Ms treat = التباين بين المجموعات. Ms Error = التباين داخل المجموعات.

١٠ - حساب قوة العلاقة بين المتغيرين:

تم حساب قوة العلاقة بين المتغيرين المستقل والتابع باستخدام معامل مربع ايتا (η) لحساب نسبة الترابط وفق الصيغة التي ذكرها الشربيني (١٩٩٥م) وهي كالتالي:

$$\frac{\text{مجموع المربعات بين المجموعات}}{\text{مجموع المربعات الكلي}} = \eta$$

١١ - الكشف عن تجانس التباين (Homogeneity of Variance).

لقياس دلالات الفروق بين متوسطات المحكمين في تقديرهم لدرجة القطع وفق طريقة نيدلسكي طبق الباحث طريقة Tukey الذي أشار إليها (Winer 1971) وهي كما يلي:

$$\sqrt{M_{\text{Serror}} \div \tilde{n} q}$$

حيث إن:

M_{Serror} = التباين داخل المجموعات. \tilde{N} = المتوسط التوافقي.

q = القيمة الحرجة من جدول (توكي) بدرجات حرية التباين داخل المجموعات وعدد المجموعات.

الفصل الرابع

عرض نتائج الدراسة.

تفسير نتائج الدراسة.

عرض وتفسير نتائج الدراسة

هدفت الدراسة إلى معرفة أثر كل من طول الاختبار ونوع المحكم على أربع من طرق تحديد درجة قطع الاختبارات محكية المرجع (أنجوف - نيدلسكاي - المجموعات المتضادة - المجموعات المحكية)، ولتحقيق أهداف الدراسة قام الباحث بالإجابة على تساؤلات وفرضيات الدراسة، والتي يمكن أن تقدم على النحو التالي:-

أولاً: عرض نتائج أثر نوع المحكم على تحديد درجة القطع:

سيتم عرض أثر نوع المحكم على تحديد درجة القطع، وفق الطريقتين التحكيمييتين كما يلي:

١ - تحديد درجة القطع وفق طريقة نيدلسكي باختلاف نوع المحكم

للإجابة عن تساؤل الدراسة الأول (هل تختلف درجة القطع الناتجة عن تطبيق طريقة نيدلسكي على اختيار موحد وعينة دراسية موحدة باختلاف نوع المحكم) تم استخدام التكرارات والنسب المئوية لإيضاح أثر نوع متغير نوع المحكم على تحديد درجة القطع وفق هذه الطريقة. والجدول التالي يوضح درجات القطع المحددة للاختبار الطويل (٣٥) مفردة باختلاف نوع المحكم.

جدول رقم ١٥ تكرارات ونسب درجة قطع الاختبار الطويل وفق طريقة نيدلسكي:

نوع المحكم	عدد المحكمون	درجة القطع	النسبة
أعضاء هيئة التدريس	١٠	٢٥	٧١,٤٣
المشرفون	١٠	١٨	٥١,٤٣
المعلمون	٢٠	١٨	٥١,٤٣

بمقارنة نسب درجة قطع الاختبار المبينة في الجدول أعلاه تبين اختلاف في نسب تقديرات المحكمين لدرجة قطع الاختبار، حيث تساوت تقديرات فئتي المعلمين والمشرفين لنسبة درجة قطع الاختبار (٥١,٤٣)، بينما اختلفت نسبة درجة القطع (٧١,٤٣) المحددة من قبل فئة أعضاء هيئة التدريس بصورة كبيرة. **لإضافة مزيداً من التوضيح** تم اختبار فرض الدراسة الأول لقياس الفروق بين متوسط تقديرات فئات المحكمين الثلاثة لدرجة قطع الاختبار الطويل (٣٥) مفردة، مع اعتبار أن درجات القطع متغير تابع ونوع المحكم متغير مستقل.

لذا قام الباحث بتطبيق أسلوب تحليل التباين أحادي الاتجاه (One Way Anova) للعينات غير المتساوية (Winer, 1971) Unequal Sample Size، لتحديد الفروق بين متوسطات تقديرات فئات

المحكمين الثلاثة لدرجة قطع الاختبار الطويل (٣٥) مفردة وفق طريقة نيدلسكي، والجدول التالي يوضح متوسط المربعات والنسبة الفائية ومستوى الدلالة.

جدول رقم ١٦ نتيجة تحليل التباين أحادي الاتجاه لمتوسط تقديرات فئات المحكمين الثلاثة لدرجة قطع الاختبار الطويل (٣٥) مفردة وفق طريقة نيدلسكي:

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة ف	مستوى الدلالة
بين مجموعات المحكمين	١٦٨,٠٢	٢	٨٤,٠١	٥,٣٥ ***	٠,٠٥
داخل المجموعات	٥٨٠,٨٤	٣٧	١٥,٧		
المجموع الكلي	٧٤٨,٩	٣٩			

دال عند مستوى دلالة ($\alpha = 0,05$)

من الجدول السابق تبين وجود فروق دالة إحصائية عند مستوى دلالة ($\alpha = 0,05$)، وهذا ما يؤدي إلى نفي الفرضية الصفرية، وإثبات وجود فروق بين متوسطات تقديرات فئات المحكمين الثلاثة تعزى لاختلاف مستويات كفايات المحكمين ووجهات نظرهم حول مستوى تحصيل الطلاب، وذلك ما أكدته دراسة بيهيونيك (Behuniak, et al (1981) ولمعرفة مدى تأثير المتغير المستقل على المتغير التابع، قام الباحث بحساب الدلالة العملية (الشرييني، ١٩٩٥م، ١٨١) حيث تبين أن حجم التأثير ($F = 22,4$) وهذه دلالة على وجود أثر أقل من المتوسط للمتغير المستقل (نوع المحكم) على درجات قطع الاختبار، ولاستخراج دلالات الفروق بين متوسطات المحكمين استخدم الباحث اختبار (Winer, 1971) Tukey كما هو موضح في الجدول التالي:

جدول رقم ١٧ نتائج تطبيق اختبار (توكي) الناتجة عن تطبيق طريقة نيدلسكي:

متوسط المجموعة	أعضاء هيئة التدريس ٢٥	المشرفون ١٨	المعلمون ١٨
أعضاء هيئة التدريس ٢٥	_____	٧ *	٧ *
المشرفون ١٨	_____	_____	صفر
المعلمون ١٨	_____	_____	_____

عند مستوى دلالة ($\alpha = 0,01$)

بمقارنة القيمتين المحددتين في الجدول أعلاه بالقيمة الحرجة في جدول توكي للفروق الدال الصادق ($5,5$) يتبين وجود فروق ذات دلالة إحصائية بين فئة المدرسين وفئة أعضاء هيئة التدريس وبين فئة المشرفين وفئة أعضاء هيئة التدريس.

مما تقدم يتضح أيضاً وجود اتفاق بين تقديرات فئة المدرسين وفئة المشرفين التربويين لدرجات القطع وفق طريقة نيدلسكي، وهذه دلالة على تقارب وجهتي نظرهما حول مستوى إتقان طلاب الصف السادس الابتدائي في مدارس مدينة جدة الحكومية، ويرى الباحث أن سبب ذلك يعود إلى ممارستهم الميدانية المشتركة في المجال التربوي.

٢ - تحديد درجة القطع وفق طريقة أنجوف باختلاف نوع المحكم:

للإجابة عن تساؤل الدراسة الثاني (هل تختلف درجة القطع الناتجة عن تطبيق طريقة أنجوف على اختيار موحد وعينة دراسية موحدة باختلاف نوع المحكم) تم استخدام التكرارات والنسب المئوية لإيضاح أثر نوع متغير نوع المحكم على تحديد درجة القطع وفق هذه الطريقة. والجدول التالي يوضح درجات القطع المحددة للاختبار الطويل (٣٥) مفردة باختلاف نوع المحكم.

جدول رقم ١٨ تكرارات ونسب درجة قطع الاختبار الطويل (٣٥) مفردة الناتجة وفق طريقة أنجوف باختلاف نوع المحكم:

نوع المحكم	عدد المحكمون	درجة القطع	النسبة
أعضاء هيئة التدريس	١٠	٢٧,٤	٧٨,٣
المشرفون	١٠	٢٨,٣	٨٠,٩
المعلمون	٢٠	٢٨,٩	٨٢,٦

وبمقارنة نسب درجة قطع الاختبار المبينة في الجدول أعلاه يتبين اختلاف ضئيل بين نسب تقديرات المحكمين لدرجة قطع الاختبار، حيث قدر المعلمون أعلى نسبة لدرجة القطع (٨٢,٦) اقتربت منها نسبة تقديرات المشرفين التربويين (٨٠,٩) بينما حققت تقديرات أعضاء هيئة التدريس أدنى نسبة لدرجة القطع (٧٨,٣)، ولعل سبب التقارب بين النسبتين المحدتين من قبل المعلمين والمشرفين يعود إلى تقارب خبرتهما الميدانية.

وللتحقق من فرض الدراسة الثاني قام الباحث بتطبيق أسلوب تحليل التباين أحادي الاتجاه للعينات غير المتساوية لتحديد الفروق بين متوسطات تقديرات فئات المحكمين الثلاثة لدرجة قطع الاختبار الطويل (٣٥) مفردة وفق طريقة أنجوف، والجدول التالي يوضح متوسط المربعات والنسب الفئوية ومستوى الدلالة.

جدول رقم ١٩ نتيجة تحليل التباين أحادي الاتجاه لمتوسط تقديرات المحكمين الثلاثة لدرجة قطع الاختبار الطويل (٣٥) مفردة وفق طريقة أنجوف:

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف
بين مجموعات المحكمين	١٤,٧٢	٢	٧,٣٦	٠,٤
داخل المجموعات	٦٨٤,٠٢	٣٧	١٨,٤٨	
المجموع	٦٩٨,٨	٣٩		

غير دال عند مستوى دلالة (٠,٠٥)

تشير النتائج في الجدول أعلاه إلى عدم وجود فروق ذات دلالة إحصائية عند مستوى دلالة α (٠,٠٥) بين متوسطات تقديرات فئات المحكمين الثلاثة لدرجة قطع الاختبار الطويل (٣٥) مفردة الناتجة وفق طريقة أنجوف، ربما يعود ذلك إلى ثبات تقديرات المحكمين (٠,٩٥) لدرجة قطع الاختبار وفق هذه الطريقة، وبالتالي فإن اختلاف نوع المحكم لا يؤثر تأثيراً جوهرياً على تقديرات درجات القطع، وذلك ما أكدته دراسة (أبو علام، ١٩٩١م) حيث لم تظهر الدراسة فروقاً دالة إحصائية لتقديرات درجة القطع باختلاف نوع المحكم.

وللمقارنة بين درجات القطع الناتجة عن تطبيق الطريقتين باختلاف نوع المحكم، تم تحديد التكرارات والنسب المئوية لدرجات القطع ولعدد الطلاب المتقنين وغير المتقنين للكفايات التي يقيسها الاختبار الطويل (٣٥) مفردة، كما هو موضح بالجدول التالي:

جدول رقم ٢٠ درجات القطع ونسبة الطلاب المتقنين ونسبة الطلاب غير المتقنين للاختبار الطويل (٣٥) مفردة الناتجة عن الطريقتين التحكيميتين باختلاف نوع المحكم:

الطريقة المستخدمة	نوع المحكم	درجة القطع الدرجة - النسبة	المتقنين العدد - النسبة	غير المتقنين العدد - النسبة	أفراد العينة
أنجوف	المعلمون	٢٩,٠	٢٢	٦٣٧	٦٥٩
أنجوف	المشرفون	٢٨,٠	٣١	٦٢٨	٦٥٩
أنجوف	أعضاء هيئة التدريس	٢٧,٠	٤٣	٦١٦	٦٥٩
نيدلسكي	المعلمون	١٨,٠	٣٣٠	٣٢٩	٦٥٩
نيدلسكي	المشرفون	١٨,٠	٣٣٠	٣٢٩	٦٥٩
نيدلسكي	أعضاء هيئة التدريس	٢٥,٠	٧٢	٥٨٧	٦٥٩

بالنظر في الجدول أعلاه نجد أن نسب تكرار درجات القطع المحددة وفق طريقة أنجوف تختلف باختلاف نوع المحكم، وتبين ذلك من خلال مقارنة النسب فيما بينها، حيث حقق المعلمون أعلى نسبة لدرجة القطع (٨٢,٩) بين فئات المحكمين، وذلك يعزز ما أكدته دراسة هالين وآخرون

Halpin, et al (1983)، بينما بلغت نسبة درجة القطع التي حددها المشرفون (٨٠,٠)، وكانت النسبة التي حددها أعضاء هيئة التدريس الأدنى (٧٧,١).

ويتضح أثر الفروق بين درجات القطع من خلال نسب تكرار الطلاب المتقنين (٣,٣، ٤,٧، ٦,٥) وغير المتقنين (٩٦,٧، ٩٥,٣، ٩٣,٥) باختلاف نوع المحكمين وفق طريقة أنجوف، حيث أنه كلما ارتفعت درجة القطع كلما قل عدد الطلاب المتقنين وزاد عدد الطلاب غير المتقنين. كما يتضح من الجدول أيضاً اختلاف نسب تكرارات درجات القطع المحددة وفق طريقة نيدلسكي باختلاف نوع المحكم، حيث بلغت نسبة درجة القطع التي حددها أعضاء هيئة التدريس (٧١,٤)، حيث كانت الأعلى بين فئات المحكمين الثلاثة، كما تساوت نسبة درجة القطع (٥١,٤) % التي حددها المعلمون والمشرفون (قد يعود ذلك إلى تقارب وجهة نظرهما لمعرفتهما الوطيدة بمستوى تحصيل الطلاب الفعلي مما ساعد على تطابق تقديراتهما في تحديد درجات قطع الاختبار) حيث كانت الأدنى مقارنةً مع عينة أعضاء هيئة التدريس، وذلك ما أكدته دراسة هالبن وآخرون (Halpin, et al, 1983) التي أشارت إلى أن تقديرات المعلمين وفق هذه طريقة نيدلسكي تؤدي إلى الحصول على درجات قطع متدنية مقارنة مع مجموعات المحكمين المختلفة.

ثانياً: عرض ومناقشة نتائج أثر متغير طول الاختبار:

ستتم مناقشة نتائج أثر متغير طول الاختبار على طرق تحديد درجة القطع المستخدمة في الدراسة على مستويين هما: مستوى كل طريقة من طرق تحديد درجة القطع على حدة ومستوى طرق تحديد درجة القطع مجتمعة.

١: عرض ومناقشة نتائج أثر متغير طول الاختبار على مستوى كل طريقة على

حده:

أ - طريقة نيدلسكي:

لتحديد درجة القطع وفق هذه الطريقة للاختبارات المختلفة (الطويل - متوسط - قصير) كرر الباحث الخطوات التي قام بها مسبقاً عند تحديد درجة القطع وفق طريقة نيدلسكي، كما هو موضح في الملاحق رقم (٥، ٦، ٧) والجدول التالي يوضح متوسط تقديرات جميع المحكمين للاختبارات الثلاثة.

جدول رقم ٢١ نسب درجات القطع الناتجة وفق طريقة نيدلسكي باختلاف طول الاختبار:

طول الاختبار	عدد المفردات	درجة القطع النسبية
طويل	٣٥	٥٧,١
متوسط	٢١	٥٧,١
قصير	١٤	٥٧,١

بمقارنة النسب المبينة في الجدول السابق يتضح تساوي نسب درجات قطع الاختبارات الثلاثة المحددة وفق طريقة نيدلسكي (٥٧,١) مما يدل على عدم وجود أثر لمتغير طول الاختبار على تحديد درجة القطع وفق هذه الطريقة.

ولإضافة مزيداً من التوضيح قام الباحث بالتحقق من اختبار الفرض الثالث بتطبيق أسلوب تحليل التباين أحادي الاتجاه للعينات غير المتساوية لتحديد الفروق بين قيمة درجة القطع وفق طريقة نيدلسكي باختلاف طول الاختبار، والجدول التالي يوضح متوسط المربعات والنسبة الفائية ومستوى الدلالة.

جدول رقم ٢٢ نتيجة تحليل التباين أحادي الاتجاه لقيمة درجة القطع وفق طريقة نيدلسكي باختلاف طول الاختبار:

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة ف
بين الاختبارات	١٠,٤١٦	٢	٥,٢٠٨	٠,٦٦
داخل الاختبارات	٠,٥٢٦	٦٧	٧,٨٥٣	
المجموع الكلي	١٠,٤٦٩	٦٩		

غير دال عند مستوى دلالة (٠,٠٥)

تشير النتائج في الجدول أعلاه إلى عدم وجود فروق دالة إحصائية عند مستوى دلالة $(\alpha) = 0,05$ بين متوسطات قيم درجات قطع الاختبارات الثلاثة وفق طريقة نيدلسكي الأمر الذي يدل على عدم علاقة بين طول الاختبار وبين تحديد درجة القطع.

ب - طريقة أنجوف:

لتحديد درجة القطع وفق هذه الطريقة للاختبارات المختلفة (الطويل - متوسط - قصير) كرر الباحث الخطوات التي قام بها مسبقاً عند تحديد درجة قطع وفق طريقة أنجوف، كما هو موضح في الملاحق رقم (٩، ١٠، ١١).

والجدول التالي يوضح متوسط تقديرات جميع المحكمين للاختبارات الثلاثة المختلفة.
جدول رقم ٢٣ يبين درجات القطع الناتجة وفق طريقة أنجوف باختلاف طول الاختبار:

درجة القطع النسبية		عدد المفردات	طول الاختبار
٨٠,٠	٢٨	٣٥	طويل
٨٠,٩٥	١٧	٢١	متوسط
٨٢,١	١١,٥	١٤	قصير

بمقارنة النسب المبيّنة في الجدول السابق يتضح فروق صغيرة في نسب درجات قطع الاختبارات الثلاثة المحددة وفق طريقة أنجوف، ويتبين أيضاً وجود علاقة عكسية بين طول الاختبار وقيمة درجة القطع حيث تقل قيمة درجة القطع كلما زاد عدد مفردات الاختبار.
ولإضافة مزيداً من التوضيح قام الباحث بالتحقق من اختبار الفرض الرابع بتطبيق أسلوب تحليل التباين أحادي الاتجاه للعينات غير المتساوية لتحديد الفروق بين قيمة درجة القطع وفق طريقة أنجوف باختلاف طول الاختبار، والجدول التالي يوضح متوسط المربعات والنسبة الفائية ومستوى الدلالة.

جدول رقم ٢٤ تحليل التباين أحادي الاتجاه لقيمة درجة القطع وفق طريقة أنجوف باختلاف طول الاختبار:

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة ف
بين الاختبارات	١٣,٥٤	٢	٦,٧٧	١,٠٨
داخل الاختبارات	٠,٤١٩	٦٧	٦,٢٥	٣
	١٣,٩٣	٦٩		

غير دال عند مستوى دلالة (٠,٠٥)

تشير النتائج في الجدول أعلاه إلى عدم وجود فروق دالة إحصائية عند مستوى دلالة $(\alpha) = 0,05$ بين متوسطات قيم درجات قطع الاختبارات الثلاثة وفق طريقة أنجوف مما يؤكد النتيجة السابقة النتيجة السابقة الناتجة عن تطبيق طريقة نيدلسكي.

ج - طريقة المجموعات المتضادة:

للإجابة عن تساؤل الدراسة الخامس (هل تختلف درجة القطع الناتجة عن تطبيق طريقة المجموعات المتضادة على اختبار موحد وعينة دراسية موحدة باختلاف طول الاختبار) تم استخدام التكرارات والنسب المئوية لإيضاح أثر نوع متغير طول الاختبار على تحديد درجة

القطع وفق هذه الطريقة. والجدول التالي يوضح درجات القطع الناتجة للاختبارات الثلاثة المختلفة.

جدول رقم ٢٥ يبين درجات القطع الناتجة وفق طريقة المجموعات المتضادة باختلاف طول الاختبار:

طول الاختبار	عدد المفردات	درجة القطع	النسبة
طويل	٣٥	٢٣	٦٥,٧
متوسط	٢١	١٢	٥٧,١
قصير	١٤	٨,٥	٦٠,٧

بمقارنة النسب المبينة في الجدول السابق يتضح اختلاف نسب درجات قطع الاختبارات الثلاثة المحددة وفق طريقة المجموعات المتضادة مما يدل على وجود أثر لمتغير طول الاختبار على تحديد درجة القطع وفق هذه الطريقة، حيث حقق الاختبار الطويل أعلى نسبة (٦٥,٧) لدرجة القطع مقارنةً مع الاختبارين المتوسط والقصير، وفي المقابل حقق الاختبار المتوسط أدنى نسبة لدرجة القطع بلغت (٥٧,١).

د - طريقة المجموعات المحكية:

للإجابة عن تساؤل الدراسة السادس (هل تختلف درجة القطع الناتجة عن تطبيق طريقة المجموعات المحكية على اختبار موحد وعينة دراسية موحدة باختلاف طول الاختبار) تم استخدام التكرارات والنسب المئوية لإيضاح أثر نوع متغير طول الاختبار على تحديد درجة القطع وفق هذه الطريقة. والجدول التالي يوضح درجات القطع الناتجة للاختبارات الثلاثة المختلفة.

جدول رقم ٢٦ درجات القطع الناتجة وفق طريقة المجموعات المحكية باختلاف طول الاختبار:

طول الاختبار	عدد المفردات	درجة القطع	النسبة
طويل	٣٥	٢٠	٥٧,١
متوسط	٢١	١٢	٥٧,١
قصير	١٤	٨	٥٧,١

بمقارنة النسب المبينة في الجدول السابق يتضح تساوي نسب درجات قطع الاختبارات الثلاثة المحددة وفق طريقة المجموعات المحكية (٥٧,١) مما يدل على عدم وجود أثر لمتغير طول الاختبار على تحديد درجة القطع وفق هذه الطريقة. الأمر الذي يؤكد النتائج السابقة الناتجة عن الطريقتين التحكيميتين.

٢: عرض ومناقشة نتائج الدراسة لمتغير طول الاختبار وفق طرق الدراسة

مجتمعة:

ستتم مناقشة أثر متغير طول الاختبار بناءً على نتائج طرق تحديد درجة القطع المستخدمة في الدراسة مجتمعة، الأمر الذي يساعد على مقارنة نتائج جميع طرق تحديد درجة القطع لكل اختبار من الاختبارات الثلاثة على حده.

لذا قام الباحث بإيجاد نسب جميع درجات القطع الناتجة عن تطبيق الطرق المختلفة باختلاف طول الاختبار، وخطأى التصنيف الأول (α) والثاني (β) والطلاب المتقنين فعلاً (True Master) والطلاب غير المتقنين فعلاً (True Non Master)، وذلك للاختبارات الثلاثة (طويل - متوسط - قصير) وهي كالتالي:-

أ - الاختبار الطويل ٣٥ مفردة:

الجدولين التاليين يوضحان عدد ونسب الطلاب المتقنين وغير المتقنين ونسب الخطأين الأول والثاني والطلاب المتقنين وغير المتقنين فعلاً الناتجة وفق الطرق المختلفة للاختبار الطويل.

جدول رقم ٢٧ تكرارات ونسب الطلاب المتقنين وغير المتقنين باختلاف الطرق وفق الاختبار الطويل:

عدد أفراد العينة	غير متقنين		متقنين		درجة القطع		الطريقة المستخدمة
	العدد	النسبة	العدد	النسبة	الدرجة - النسبة	الدرجة - النسبة	
٦٥٩	٦٢٨	٩٥,٣	٣١	٤,٧	٢٨	٨٠,٠	أنجوف
٦٥٩	٤٣٣	٦٥,٧	٢٢٦	٣٤,٣	٢٠	٥٧,١	نيدلسكي
٦٥٩	٥٣٠	٨٠,٤	١٢٩	١٩,٦	٢٣	٦٥,٧	المتضادة
٦٥٩	٤٣٣	٦٥,٧	٢٢٦	٣٤,٣	٢٠	٥٧,١	المحكية

جدول رقم ٢٨ تكرار ونسب الخطأين الأول والثاني والطلاب المتقنين وغير المتقنين فعلاً للاختبار الطويل وفق الطرق المختلفة:

عدد أفراد العينة القصدية	غير متقن فعلاً		متقن فعلاً		الخطأ الثاني		الخطأ الأول		درجة القطع	الطريقة المستخدمة
	العدد	النسبة	العدد	النسبة	العدد	النسبة	العدد	النسبة		
٨٩	٤٤	٤٩,٤	١٤	١٥,٧	٣١	٣٤,٨	—	—	٢٨	أنجوف
٨٩	٣٨	٤٢,٧	٤٢	٤٧,٢	٣	٣,٤	٦,٧	٦	٢٠	نيدلسكي
٨٩	٤٣	٤٨,٣	٣٥	٣٩,٣	١٠	١١,٢	١٠,١	١	٢٣	المتضادة
٢١٩	٩١	٤١,٥	٩١	٤١,٥	١٩	٨,٧	٨,٢	١٨	٢٠	المحكية

بمقارنة النسب الموضحة بالجدولين السابقين تبين أن نسبة درجة القطع المحددة وفق طريقة أنجوف كانت الأعلى (٨٠,٠) حيث نتج عنها أدنى نسبة للطلاب المتقنين (٤,٧) وأعلى نسبة للخطأ الثاني (٣٤,٨) وأدنى نسبة للطلاب المتقنين فعلاً (١٥,٧)، يليها نسبة طريقة المجموعات المتضادة (٦٥,٧) التي نتج عنها ارتفاع ملحوظ في نسبة المتقنين (١٩,٦) ونسبة طفيفة للخطأ الأول (١,١) ونسبة مرتفعة للطلاب المتقنين فعلاً (٣٩,٣) وارتفاع كبير في نسبة الطلاب غير المتقنين بالفعل (٤٩,٤).

يتضح من الجدولين أيضاً تساوي نسبي طريقتي نيدلسكاي والمجموعات المحكية (٥٧,١) على الرغم من الاختلاف الكبير في إجراءات تحديدهما، وترتب على هذه النسبة ارتفاع نسبة الطلاب المتقنين (٣٤,٣) وانخفاض نسبة غير المتقنين (٦٥,٧) مقارنةً مع الطريقتين الآخرين، بمقارنة نسبة الخطأين الأول (٦,٧ - ٨,٢) والثاني (٣,٤ - ٨,٧) لطريقتي نيدلسكي والمجموعات المحكية يلاحظ أنهما مختلفين بشكل واضح، أما فيما يخص نسب الطلاب المتقنين (٤٧,٢ - ٤١,٥) وغير المتقنين فعلاً (٤٢,٧ - ٤١,٥) فيلاحظ تساوي نسبي المتقنين وغير المتقنين وفق طريقة المجموعات المحكية (٤١,٥) واختلافهما وفق طريقة نيدلسكاي.

ب- الاختبار المتوسط ٢١ مفردة:

الجدولين التاليين يوضحان عدد ونسب الطلاب المتقنين وغير المتقنين ونسب الخطأين الأول والثاني والطلاب المتقنين وغير المتقنين فعلاً الناتجة وفق الطرق المختلفة للاختبار المتوسط.

جدول رقم ٢٩ نسب وتكرارات الطلاب المتقنين وغير المتقنين للاختبار المتوسط وفق الطرق المختلفة:

الطريقة المستخدمة	درجة القطع - النسبة	متقنين - العدد - النسبة	غير متقنين - العدد - النسبة	العينة القصدية
أنجوف	١٧ - ٨٠,٩٥	٣٧ - ٥,٦	٦٢٢ - ٩٤,٤	٦٥٩
نيدلسكي	١٢ - ٥٧,١	٢٢٤ - ٣٤,٠	٤٣٥ - ٦٦,٠	٦٥٩
المتضادة	١٢ - ٥٧,١	٢٢٤ - ٣٤,٠	٤٣٥ - ٦٦,٠	٦٥٩
المحكية	١٢ - ٥٧,١	٢٢٤ - ٣٤,٠	٤٣٥ - ٦٦,٠	٦٥٩

جدول رقم ٣٠ نسب وتكرارات الخطأين الأول والثاني والطلاب المتقنين وغير المتقنين فعلاً وفق

الطرق المختلفة للاختبار المتوسط:

الطريقة المستخدمة	درجة القطع	الخطأ الأول - العدد - النسبة	الخطأ الثاني - العدد - النسبة	متقن فعلاً - العدد - النسبة	غير متقن فعلاً - العدد - النسبة	العينة القصدية
أنجوف	١٧	—	٢٨	٣١,٥	١٧	٨٩
نيدلسكي	١٢	٤ - ٤,٥	٥	٥,٦	٤٠	٨٩
المتضادة	١٢	٤ - ٤,٥	٥	٥,٦	٤٠	٨٩
المحكية	١٢	١٣ - ٥,٩	٢٠	٩,٠	٩٠	٢١٩

يتضح من الجدولين السابقين تساوي نسبة درجة القطع (٥٧,١) للطرق الثلاث المختلفة (نيدلسكاي - المجموعات المتضادة - المجموعات المحكية) رغم الاختلاف الكبير في إجراءات تحديدها، ترتب على ذلك تساوي نسب الخطأ الأول (٤,٥) والخطأ الثاني (٥,٦) والطلاب المتقنين فعلاً (٤٤,٩) والطلاب غير المتقنين فعلاً (٤٤,٩) وذلك في نتائج الطريقتين نيدلسكاي والمجموعات المتضادة، أما طريقة أنجوف فقد حققت أعلى نسبة (٨٠,٩٥) درجة قطع، وكانت نسبة الخطأ الثاني (إعادة دراسة الطالب البرنامج مرة ثانية رغم إتقانه) (٣١,٥) الأعلى بين الطرق المختلفة، وبمقارنة نسب الطلاب المتقنين فعلاً بين الطرق المختلفة يتضح أن طريقة أنجوف حققت أدنى نسبة بين الطرق (١٩,١) على الرغم من تساوي عدد أفراد العينة القصدية بين معظم الطرق (٨٩) طالباً .

ج- الاختبار القصير ١٤ مفردة

الجدولين التاليين يوضحان عدد ونسب الطلاب المتقنين وغير المتقنين ونسب الخطأين الأول والثاني والطلاب المتقنين وغير المتقنين فعلاً الناتجة وفق الطرق المختلفة للاختبار القصير .
جدول رقم ٣١ تكرار ونسب الطلاب المتقنين وغير المتقنين للاختبار القصير وفق الطرق المختلفة:

الطريقة المستخدمة	درجة القطع - النسبة	المتقنين - العدد	غير المتقنين - النسبة	عدد أفراد العينة
أنجوف	١١,٥	٢٣	٦٣٦	٦٥٩
نيدلسكي	٨	٢٤٨	٤١١	٦٥٩
المتضادة	٨,٥	١٦٠	٤٩٩	٦٥٩
المحكية	٨	٢٤٨	٤١١	٦٥٩

جدول رقم ٣٢ نسب وتكرار الخطأين الأول والثاني والطلاب المتقنين وغير المتقنين فعلاً للاختبار القصير وفق الطرق المختلفة:

الطريقة المستخدمة	درجة القطع	الخطأ الأول - العدد - النسبة	الخطأ الثاني - العدد - النسبة	متقن فعلاً - العدد - النسبة	غير متقن فعلاً - العدد - النسبة	العينة القصدية
أنجوف	١١,٥	—	٣٣	١٢	٤٤	٨٩
نيدلسكي	٨	٧,٩	٤	٤١	٣٧	٨٩
المتضادة	٨,٥	٤,٥	١٠	٣٥	٤٠	٨٩
المحكية	٨	٦,٨	٢١	٨٩	٩٤	٢١٩

بمقارنة النسب الموضحة بالجدولين السابقين تبين أن طريقة أنجوف حققت أعلى نسبة (٨٢,١) ترتب عليها تدني نسبة الطلاب المتقنين (٣,٥) وارتفاع نسبتي الطلاب غير المتقنين (٦٦,٥)

والخطأ الثاني (٣٧,١)، تلت هذه الطريقة بفارق واضح طريقة المجموعات المتضادة (٦٠,٧) ترتب عليها ارتفاع في نسبة الطلاب المتقنين (٢٤,٣) وتدني نسبي الطلاب غير المتقنين (٧٥,٧) والخطأ الثاني (١١,٢) وذلك مقارنةً مع طريقة أنجوف، أما بالنسبة لطريقتي نيدلسكي والمجموعات المحكية فتساوت فيهما نسبي درجة القطع (٥٧,١) حيث حققنا أعلى نسبة للطلاب المتقنين (٣٧,٦)، أما بالنسبة للخطأ الأول (٧,٩ - ٦,٨) والخطأ الثاني (٩,٦ - ٤,٥) على التوالي فيلاحظ تقاربهما إلى حد ما بين الطريقتين في نسبة الخطأ الأول وتباينهما في الخطأ الثاني، بمقارنة نسبي الطلاب المتقنين فعلاً (٤٦,١، ٤٠,٦) والطلاب غير المتقنين فعلاً (٤١,٦، ٤٢,٩) على التوالي نجد أنهما يختلفان في المعيار الأول ويتقاربان إلى حد ما في الثاني.

والجدول التالي يوضح نسب درجات القطع الناتجة باختلاف طول الاختبارات وفق الطرق المختلفة.

جدول رقم ٣٣ تكرارات ونسب درجات القطع الناتجة باختلاف طول الاختبارات وفق الطرق المختلفة:

الطريقة المستخدمة	طويل		متوسط		قصير	
	الدرجة - النسبة	الدرجة - النسبة	الدرجة - النسبة	الدرجة - النسبة	الدرجة - النسبة	الدرجة - النسبة
أنجوف	٢٨	٨٠,٠	١٧	٨٠,٩٥	١١,٥	٨٢,١
نيدلسكي	٢٠	٥٧,١	١٢	٥٧,١	٨	٥٧,١
المتضادة	٢٣	٦٥,٧	١٢	٥٧,١	٨,٥	٦٠,٧
المحكية	٢٠	٥٧,١	١٢	٥٧,١	٨	٥٧,١

يتبين من مقارنة نسب درجات قطع الاختبارات الناتجة عن تطبيق الطرق المختلفة والمبينة في الجدول أعلاه تكرار نسبة درجة القطع (٥٧,١) في نتائج تقديرات درجة القطع وفق طريقة نيدلسكي والمجموعات المتضادة والمجموعات المحكية، وعلى الرغم من اختلاف الإجراءات المستخدمة في تحديد درجة القطع وفق الطرق المختلفة.

من خلال المناقشة السابقة أمكن التوصل للنتائج التالية:-

١. يقل مجموع خطأ التصنيف الأول والثاني عند زيادة طول الاختبار (القصير - الطويل) وفق طرق تحديد درجة القطع المختلفة، ويتفق ذلك مع دراسة هالادين ورويد Haladyn & Roid, (الأحمد، ١٩٩٢م).
٢. تساوي درجات القطع الناتجة عن تطبيق طريقتي نيدلسكي والمجموعات المحكية باختلاف طول الاختبار، وذلك يتفق مع دراسة ميلز (Mills, 1983).
٣. ارتفاع درجات القطع الناتجة عن تطبيق طريقة أنجوف وذلك عند تطبيقها على الاختبارات المختلفة (طويل - متوسط - قصير)، نتج عنها عدم ظهور أي نسبة للخطأ

- الأول، وارتفاع في نسبة الخطأ الثاني (٣٤,٨، ٣١,٥، ٣٧,١) على التوالي، وذلك يؤكد ما أشارت إليه الدراسات السابقة ومنها دراسة بيهيونيك، وهالبن وآخرون، وهاراسيم (Behuniak,1982; Halpin et al, 1983; Harasym,1981).
٤. تقارب نسبي الطلاب المتقنين (True Master) إلى حد ما (٣٤,٣، ٣٤,٦، ٣٧,٦) وغير المتقنين (True Non Master) (٦٥,٧، ٦٦,٤، ٦٢,٤) عند تطبيق طريقة المجموعات المحكية على الاختبارات المختلفة مقارنة بالطرق الأخرى .
٥. حققت طريقتي نيدلسكاي والمجموعات المحكية أدنى درجات قطع عند تطبيقهما على الاختبارات المختلفة (طويل - متوسط - قصير) وذلك يعزز ما أشارت إليه الدراسة السابقة التالية: دراسة كوفلر، وكروس وآخرون، وهاراسيم، وهالبن وآخرون، وسكاكون وكيلنق. (Cross et al, 1984; Harasym, 1981; Halpin et al, 1983; Skakun&Kling, 1980; Koffler, 1980).

الفصل الخامس

ملخص وتوصيات.

مقترحات الدراسة.

ملخص الدراسة:

هدفت الدراسة إلى التعرف على أثر كل من نوع المحكم وطول الاختبار على أربع من طرق تحديد درجة القطع وهي طريقة نيدلسكي وأنجوف والمجموعات المتضادة والمجموعات المحكية. ويقصد بنوع المحكم الوظيفة التي يمارسها المحكم المشارك في تحديد درجة القطع وهو المعلم أو المشرف التربوي أو عضو هيئة التدريس. ويقصد بطول الاختبار عدد مفردات الاختبار، وهي اختبار طويل يتكون من (٣٥) مفردة، واختبار متوسط يتكون من (٢١) مفردة، واختبار قصير يتكون من (١٤) مفردة. ولتحقق من أهداف الدراسة انتقى الباحث بناء على ثلاثة معايير إحصائية مفردات الاختبارات الثلاثة المستخدمة في الدراسة من اختبار محكي المرجح يتكون من (٩٥) مفردة تقيس الكفايات الرياضية في العمليات الحسابية على الأعداد بالمرحلة الابتدائية. كما استخدم الباحث استمارتين لتحديد درجة القطع وفق طريقتي نيدلسكي وأنجوف التحكيميتين. وبلغ عدد أفراد عينة الدراسة (٦٥٩) طالباً، كما بلغ عدد المحكمين (٤٠) محكماً شاركوا في تحديد درجة القطع وفق الطريقتين التحكيميتين. وللإجابة عن أسئلة الدراسة استخدم الباحث التحليلات الإحصائية المتمثلة في التكرارات والنسب المئوية وأسلوب تحليل التباين أحادي الاتجاه للعينات غير المتساوية وذلك للتعرف على دلالة ما قد يوجد من فروق بين المتوسطات تقديرات درجة قطع الاختبارات وفق الطرق المستخدمة باختلاف نوع المحكم وطول الاختبار.

أسفرت الدراسة عن عدد من النتائج أهمها:-

١. توجد فروق ذات دلالة إحصائية عند مستوى دلالة (٠,٠٥) بين تقديرات المحكمين لدرجات قطع الاختبار وفق طريقة نيدلسكي باختلاف نوع المحكم
٢. لا توجد فروق ذات دلالة إحصائية بين تقديرات المحكمين لدرجات قطع الاختبار وفق طريقة أنجوف باختلاف نوع المحكم.
٣. حققت تقديرات مجموعتي المعلمين والمشرفين أعلى نسبة لدرجة قطع الاختبار وفق طريقة أنجوف، بينما حققتا أدنى نسبة وفق طريقة نيدلسكي.
٤. تساوي تقديرات مجموعتي المعلمين والمشرفين لدرجات قطع الاختبار وفق طريقة نيدلسكي.
٥. تكررت نسبة درجة القطع (٥٧,١) بصورة كبيرة بين النسب المحددة لدرجة قطع الاختبارات الثلاثة الناتجة عن تطبيق الطرق المختلفة (باستثناء طريقة أنجوف).
٦. لا توجد علاقة بين طول الاختبار ودرجات قطع الاختبار. أي لا يترتب على زيادة عدد المفردات تدني في نسبة درجة القطع (باستثناء طريقة أنجوف).

٧. يقل مجموع خطأي التصنيف الأول والثاني عند زيادة طول الاختبار (القصير - الطويل) وفق طرق تحديد درجة القطع المختلفة.
٨. تساوي نسبة درجة قطع الاختبارات الثلاث (٥٧,١) الناتجة عن تطبيق طريقتي نيدلسكي والمجموعات المحكية.
٩. تقارب نسبي الطلاب المتقنين وغير المتقنين في الاختبارات الثلاث الناتجة عن تطبيق طريقة المجموعات المحكية.
١٠. حققت طريقتي المجموعات المحكية ونيدلسكي أدنى نسبة لدرجة قطع الاختبار وذلك في نتائج الاختبارات الثلاث.

توصيات ومقترحات الدراسة

في ضوء نتائج الدراسة يوصي الباحث بما يلي :-

١. تبين للباحث تشوش المحكمين عند تحكيم استمارات تحديد درجة القطع، لذا يوصى الباحث بتدريب المحكمين بصورة جيدة وذلك مما يساعد على الحصول على تقديرات دقيقة لدرجات قطع الاختبار.
٢. أظهرت نتائج الدراسة عدم وجود فروق إحصائية بين درجات قطع الاختبار باختلاف نوع المحكم وفق طريقة أنجوف، لذا يرى الباحث بأنه ليس بالضرورة تعدد فئات مجموعات المحكمين عند تطبيق طريقة أنجوف ويمكن الاكتفاء بعينة متخصصة ذات علاقة بكل من النطاق السلوكي المراد قياسه ومستوى عينة الدراسة التحصيلي.
٣. بينت الدراسة ارتفاع تقديرات المحكمين لدرجة قطع الاختبار وفق طريقة أنجوف، لذا توصي الدراسة باستخدام هذه الطريقة عندما يكون الهدف من الاختبار انتقاء طلاب ذوي كفاءة عالية في البرنامج المراد قياسه.
٤. من الصعوبات التي تواجه المحكمين عند تطبيق طريقة المجموعات المتضادة تصنيف الطلاب إلى فئتين متقنين وغير متقنين، لذا توصي الدراسة الحالية بالاعتماد على آليات أخرى مثل (درجات الطلاب في اختبارات تحصيلية سابقة) لتحديد فئتي الطلاب المتقنين وغير المتقنين بصورة موضوعية.
٥. تحتاج طريقة المجموعات المحكية إلى عمليات رياضية وإحصائية معقدة يصعب على بعض المعلمين تطبيقها، لذا توصي الدراسة بتعريب أو بتصميم برامج خاصة تسهل عملية تحديد درجة القطع بالحاسب الآلي.

مما تقدم يمكن للباحث تقديم المقترحات التالية:

١. إجراء دراسة مقارنة بين طرق تحديد درجات القطع التجريبية التي تعتمد على نظرية استجابة المفردة (IRT).
٢. إجراء دراسة لمعرفة أثر تدريب المحكمين على تحديد درجة القطع وفق الطرق التحكيمية.
٣. إجراء دراسة مماثلة على أداة الدراسة الحالية تعتمد على متغيرات وعينة دراسية مختلفة.
٤. إجراء دراسة لمعرفة أثر التداخل بين نوع المحكم ومستوى صعوبة المفردة.
٥. إجراء دراسة لمعرفة أثر استخدام الطرق التي تعتمد على بيانات واقعية (**Actual Data**) حول مستوى تحصيل الطلاب على تحديد درجة القطع.

قائمة المراجع

- ١- أبو حطب، فؤاد (١٩٨٧م). التقويم النفسى. القاهرة: الأنجلو المصرية.
- ٢- الأحمّد، أحمد (١٤١٣هـ). تأثير طريقة اختيار فقرات الاختبار المحكى المرجع على خصائصه السيكومترية. رسالة دكتوراه غير منشورة. عمان. الجامعة الأردنية.
- ٣- أمريزيق، خالد (١٩٩٢م). المقارنة بين اختبار الاختيار من متعدد واختبار الصح - خطأ واختبار الصح - خطأ المتعدد من حيث الخصائص السيكومترية وشكل منحى التوزيع للأداء على الاختبار. رسالة ماجستير غير منشورة. عمان. الجامعة الأردنية.
- ٤- بشناق، عدنان (١٤١٢هـ). اختبار محكى المرجع لقياس الكفايات الرياضية فى العمليات الحسابية على الأعداد بالمرحلة الابتدائية بمدارس مكة المكرمة. رسالة ماجستير غير منشورة. مكة المكرمة. جامعة أم القرى.
- ٥- الشربيني، زكريا (١٩٩٥م). الإحصاء وتصميم التجارب فى البحوث النفسية والتربوية والاجتماعية. القاهرة: مكتبة الأنجلو المصرية.
- ٦- الطائي، نزار (١٤٠٣هـ). القياس وعلاقته بالتقويم - محاضرات فى التقويم التربوي. الكويت: مكتب التربية العربي لدول الخليج.
- ٧- عبد السلام، نادية (١٩٩١م). تعدد معطيات تحديد الدرجات الفاصلة فى القياس محكى المرجع (دراسة تجريبية). حولية كلية البنات، جامعة عين شمس. المجلد ١، العدد ١٦.
- ٨- _____ (١٩٩٢م). ثلاث محكات رئيسية لتحديد مستوى الإتقان فى الاختبارات محكية المرجع. المجلة المصرية للدراسات النفسية، العدد ٣.
- ٩- علام، صلاح الدين (١٩٨٥م). استخدام النموذج ذى الحدين فى تقدير درجة القطع للاختبار محكى المرجع (دراسة إحصائية). المجلة العربية للعلوم الإنسانية، جامعة الكويت.
- ١٠- _____ (١٩٨٦م). تطورات معاصرة فى القياس النفسى والتربوي. الكويت: مطابع القبس التجارية.
- ١١- _____ (١٩٩١م). دراسة مقارنة لبعض طرق تحديد مستويات الأداء فى اختبار محكى المرجع، المجلة المصرية للدراسات النفسية، العدد ١: ٧٧-٩٦.
- ١٢- _____ (١٩٩٥م). الاختبارات التشخيصية مرجعية المحك فى المجالات التربوية والنفسية والتدريبية. القاهرة: دار الفكر العربي.
- ١٣- قدرى، حنفي وآخرون (١٩٨٧م). أصول القياس والبحث العلمى. القاهرة: دار آتون للنشر.
- ١٤- كاظم، أمينة (١٩٨٨م). دراسة نظرية نقدية حول القياس الموضوعى للسلوك - نموذج راش - الكويت: مؤسسة الكويت للتقدم العلمى.

- 15- Andrew, B.J. and Hecht, J.T. (1976). A preliminary investigation of two procedures for setting examination standard. Educational and Psychological Measurement. 36 : 45-50.
- 16- Angoff, W.H. (1971). Scales, Norms and Equivalent Scores. In R.L. Thorndik (Ed.), Educational Measurement. Washington, D.C.: American Council on Education.
- 17- Behuniak, P., Archambault, F.X., and Gable, R.K. (1982). Angoff and Nedelsky Standard Setting Procedures: Implications For The Validity of Proficiency Test Score Interpretation. Educational and Psychological Measurement. 42: 247-255.
- 18- Berk, R.A (1980). Criterion Referenced Measurement. The State of Art.
- 19- ----- (1986). Consumers Guide To Setting Performance Standard on Criterion Referenced Test. Review of Educational Research . 56: 137-172.
- 20- Crocker, L., & Algina, J. (1986). Introduction to Classical & Modern Test Theory. New York Holt . Rinehart & Winston.
- 21- Cross, I.H., Impara, I.C., Frary, R.B., and Jaeger, R.M. (1984). A Comparison of Three Methods for Establishing Minimum Examination. Journal of Educational Measurement . 21: 113-129.
- 22- Ebel, R.L. (1979). Essentials of Educational Measurement. Englewood Cliffs . NJ: Prentice Hall.
- 23- Emrick, J.A (1971). An evaluation models for mastery testing. Journal of Educational Measurement . 8: 321-326.
- 24- Fehrmann, M., Woehr, D.J. & Wimfred, A.Jr. (1991). The Angoff Cut Off Score Methods: The Impact of Frame of Reference Rater Training. Educational & Psychological Measurement . 51: 857-872.
- 25- Glaser, R (1994). Instructional Technology & the Measurement of Learning Outcomes: Some Questions. Educational Measurement: Issues & Practice . 13: 6-8.
- 26- Gray, W.M. (1978). A comparison of Piagetian theory & criterion referenced measurement. Review of Educational Research. 48: 223-249.
- 27- Halpin, G., Sigmon, & Halpin, G. (1983). Minimum Competency Standard Set by Three Divergent Groups of Raters Using Three Judgmental Procedures: Implications for Validity. Educational & Psychological Measurement. 43: 185-196.
- 28- Harasym, P.H. (1981). A comparison of the Nedelsky & Modified Angoff Standard Setting Procedure on Evaluation Outcomes. Educational & Psychological Measurement . 41: 725-734
- 29- Hambleton, R.K., & Novick, M.R. (1973). Toward An Integration of Theory and Method for Criterion Referenced Tests. JEM. 10: 159-170.
- 30- ----- (1978). On the Use of Cut Off Score with Criterion Referenced Tests in Instructional Setting, JEM. 15: 277-290

- 31- -----, Swaminathan, H., & Algina. (1976). Some contributions to the theory and practice of criterion referenced testing. In D.N.M. de gruijter, & van der kamp (Eds.), Advances in psychological and educational measurement. New York: Wiley.
- 32- ----- & swaminathan , H, Algina. J & couison. D. B. (1978). criterion referenced testing & measurement: A review of technical issues & development. Review of educational research . 48: 1-47.
- 33- ----- & Eignor, D.R. (1980). Competency Test development, Validation, and standard setting. IN R. M. jaeger & C.K. Tittle. (Eds.) , Minimum competency a chievement testing Motives, models, measures, and consequences. Berkeley, CA: Mccutchan
- 34- Jaeger, R. M (1989). certification of student competence. IN linn, R.L.(Eds.), educational measurement. new york: macmillan.
- 35- Koffler, S. L. (1980) A comparison of Approaches for setting proficiency standards. jem . 17: 167-178.
- 36- Meskauskas, J. A. (1976). evaluation models for criterion referenced testing: views regarding mastery & standard setting. Review of educational research . 46:133-158.
- 37- Mills, C.N. (1983). A comparison of three methods of establishing cut off scores on criterion referenced tests. jem . 20: 283 - 292.
- 38- Nedelsky, L. (1954). Absolute grading standards for objective tests. Educational and psychological measurement. 14: 3-19.
- 39- Nitko , A. J (1980). distinguishing the many varieties of criterion referenced tests, review of educational research. 50: 461 - 485.
- 40- Popham, W. J. (1978). criterion referenced measurement. englewood cliffs , NJ. prentic hall
- 41- ----- & Husek, T.R. (1969). Implications of criterion referenced measurement. JEM. 6: 19.
- 42- Skakun, E. N & Kling, S (1980). comparability of methods for setting standard.. jem. 17 , 229 - 235.
- 43- Subkoviak, M.J. (1980). decision consistency approaches. In R. A, BERK (Eds.), criterion referenced measurement. the state of art. PP. 129 - 183.
- 44- Winer, B.J, (1971). Statistical Principles In Experimental Design. New York: McGraw – Hill (2nd ed.).
- 45- Woehr, D. J, Arthur, W. JR & Fehrmann, M. L (1991). An empirical comparison of cut off score methods for content related & criterion related validity setting, educational & psychological measurement. 15: 1029 - 1039.