أثر بعض الطـــرق الوزنية لتقدير الدرجـــات على صدق الإختبارات مرجعية المحك ذات الاختيار من متعدد

د/ صلاح شريف عبدالوهاب
كلية التربية النوعية - جامعة الزقازيق

أثر بعض الطـــرق الوزنية لتقدير الدرجـــات على صدق الإختبارات مرجعية المحك ذات الاختيار من متعدد

د/ صلاح شريف عبدالوهاب

كلية التربية النوعية – جامعة الزقازيق

مقدمـــة :

يعد القياس والتقويم التربوى أحد المكونات الرئيسية للمنظومة التربوية وأكثرها تأثيراً فى تقدمها ورفع كفاءتها ، ويسهم تطويرها إسهاما إيجابيا فى إصلاح وتطوير بقية تلك المكونات .

لذلك عكف علماء القياس والتقويم التربوى من أمثال بلوك Block ١٩٧٠ ، بابام Popham ١٩٧٨ وبلوم Bloom ١٩٨١ وغيرهم فى النصف الأخير من القرن العشرين على دراسة أساليب جديدة لتطوير نظم قياس وتقويم التحصيل للمتعلمين محاولين الإستفادة من تلك التطورات التقنية التى حدثت فى المجالين التربوى والنفسى ، واستطاع هؤلاء العلمـــاء بفضل جهودهم الدائبة التوصل إلى استراتيجيات جديدة فى بناء الاختبارات التحصيلية Achievement Tests اعتمادا على الأهداف السلوكية Behavioral Obgectives

وتعد تلك الإختبارات أحد الوسائل الرئسية لعمليات التقويم التى تعتمد عليها المؤسسات التعليمية والتربوية للتعرف على مقدار التقدم فى تحصيل المتعلمين ومدى تحقيقهم للأهداف التعليمية ، لذلك يمكن إعتبارها إسهاما مهما فى تشخيص صعوبات التعلم لدى المتعلمين ، وتحديد مدى كفاءتهم ، وكذلك صنع القرارات التربوية لكل من تطبق عليهم تلك الإختبارات ، وبالتالى فإن دقة وسلامة تلك القرارت تتوقف على مدى جودة وكفاءة الأختبارات المستخدمة فى عملية التقويم . (٥ : ٢١)

ومن ثم كان هناك أكثر من نوع للإختبارات التحصيلية وأهمها فى المجال التعليمى إختبارات مرجعية المعيار Norm-Referenced Tests والإختبارات مرجعية المحك Criterion-Referenced Tests حيث يتميز النوع الأول من الإختبارات NRT بفلسفة تربوية تركز على تصنيف المتعلمين بحسب مركزهم النسبى بين أقرانهم فى القدرات المختلفة ، والتوازن بين أدائهم فى هذه القدرات دون تحديد لمكتسباتهم العقلية وكفاياتهم الفردية لذا يتبين محدودية هذا النوع من الإختبارات للتشخيص الدقيق لهذه المكتسبات والكفايات ، وكذلك قلة فاعليتها فى

إثراء المنظومة التعليمية وتطوير البرامج التعليمية والتدريبية . (٢٩: ٥)

وعلى الجانب الآخر يتميز النوع الثانى من الإختبارات التحصيلية مرجعية المحك (CRTS) بأنه يستخدم فى تقدير آداء الفرد بالنسبة إلى نطاق سلوكى معين Be-havioral Domain أى مجموعة من المعارف والمهارات المعرفة بطريقة إجرائية (سلوكية) بغض النظر عن علاقة أدائة بأداء غيره من أقرانه الذين يطبق عليهم نفس الإختبار . (٦ : ١٤)

لذلك يعد هذا النوع من الإختبارات أكثر إرتباطا بالأهداف التعليمية التى يسعى المعلم إلى تحقيقها لدى المتعلمين مما يجعلها أكثر فاعلية وملاءمة لمتطلبات التحصيل الدراسى للمتعلمين .

ونظرا للاهتمام المستمر بتطوير أساليب القياس والتقويم يجب أن يكون هناك إعادة النظر بصورة مستمرة فى أدوات القياس وتطويرها مراعاة الشروط والأسس الفنية الواجب توافرها عند بنائها لكى يعتمد على نتائجها بدرجة أكبر من الثقة .

ويعد الصدق والثبات من أهم شروط الإختبار الجيد ، لذلك فان الثقة فى إختبار ما سواء كان جماعى المرجع أو محكى المرجع تتوقف على مقدار صدقه وثباته . حيث يقصد بالصدق فى الإختبارات مرجعية المحك Validity مدى ملاءمة الدرجات المستمدة من الإختبارت للإستخدامات المعينة المناسبة للغرض الذى بنى من أجله (١٢: ٢٥١)

ويقصد بمفهوم الثبات Reliability فى الإختبارات مرجعية المحك أنه يشير إلى مدى إتساق ما بين بقية الإختبار ، ويزداد ثبات الإختبار بزيادة درجة هذا الإتساق . (٦ : ٢٥٢)

وقد اهتمت كثير من الدراسات والبحوث السيكومترية بالتوصل إلى طرق وأساليب يمكن باستخدامها تحسين وزيادة صدق الإختبارات وثبات درجاتها ، حيث قام بعض المشتغلين بالقياس النفسى والتربوى بدراسة صدق وثبات درجات مفردات الإختبارات ذات الإختيار من متعدد فى حالة استخدام طرق لتصحيح إجابات هذه الإختبارت تختلف عن طريقة التصحيح التقليدية التى تقوم على أساس اعتبار الدرجة الكلية للإختبار هى عدد المفردات التى يجيب عنها الفرد إجابة صحيحة Number-Right Score.

ومن بين هذه الطرق الوزنية لجتمان Guttman Weighting, Method (1941)، وطريقة الدرجات التجريبية لكومبس Coombs وومر Woomer

(1956) ، وطريقة الدرجة الوزنية للإختبار Choice Weight Scoring التى اقترحها ديفز (1959) Davis، وطريقة الإختيار الإستنتاجى لجلاسى، ووايلى (1959) Wiley وطريقة الإحتمال المقترح للإجابة Probapaility Assigned Correct Response لشطرد Shutord ومسنيجل (1968) Massengili ، وطريقة الإقليدية فى تقدير الدرجات لمفردات الإختبار Euclidean Method التى اقترحها ربى (1968) Rippey ، وطريقة درجات الثقة Confidence Scores لمشيل ,(1968) Michael ، وطريقة التصحيح المتعدد الرتب Multiple Category Scoring لبوك (1972) ;Boc، وطريقة التصحيح من أثر التخمين The Correction Guessing Method لدياموند Diamond وايفانس (1973) Evans، وطريقة دافيد (1975) David التى يتم بواسطتها تقدير درجات الإختبار ذات الإختيار من متعدد على أساس عدد المفردات التى يتكون منها الإختيار ، وكذلك عدد الإختبارات لكل مفردة ، وطريقة تصحيح أثر التخمين Correction For Guessing Method لتوماس Thomas (1981) (٢١٩:١٥) ، (١٨: ٢١٩-٢٢٠)

وقد بينت الدراسات التى أجريت لمقارنة هذه الطرق بعضها بالبعض الآخر ومقارنتها بالنظام التقليدى للتصحيح أن هذه الطرق قد أدت إلى زيادة ملحوظة فى صدق درجات الإختبارات مرجعية المحك أكثر مما يعطيه النظام التقليدى فى بعضها والبعض الآخر لم يعط أى فروق بين هذه الطرق ، ومن بين هذه الدراسات دراسة كل من جيمس وديوند (1975James Diamond) ، ويثزن 1976 Thissen وبرانبوم وتاستوك 1983 Biremboun & Totsua;al وسلوجوف 1987 Slogoff وجاردوت 1988 Gardot ، روس 1994 Ross وفيشر 1996 Fisher واميرال 1997 Ammeraal وريوز 1998 Ruiz .

مشكلة البحث :

من خلال العرض السابق يتبين أهمية الصدق كأحد شروط الإختبار الجيد ، وأيضا لتباين نتائج الدراسات فى هذا المجال حول مقارنة الطريقة التقليدية بالطرق الوزنية وأثرها على صدق الإختبارات مرجعية المحك وعلى هذا يمكن صياغة مشكلة البحث الحالى فى التساؤل التالي :

س١ هل يختلف صدق الإختبار مرجعى المحك فى حالة استخدام الطريقة التقليدية

فى تقدير الدرجات عنه فى حالة استخدام الطرق الوزنية المستخدمة فى البحث الحالى مثل طريقة الإحتمال المقترح أو طريقة الإختبار الإستنتاجى أو الطريقة التربيعية ؟

حـــدود البحث :

١- عينة من تلاميذ الصف الثالث الإعدادى بنين بمحافظة الشرقية .

٢- بناء اختبار تحصيلى مرجعى المحك ذات الإختيارمن متعدد فى وحدة التكاثر فى الكائنات الحية المقررة على الصف الثالث الإعدادى فى محتوى مادة العلوم .

٣- الإقتصار على أربع طرق وزنية فى تقـــدير الدرجات هى طريقة الإحتمال وطريقـة الدرجـــات التجريبية والطريقـة التربيعية وطريقة الإختيار الإستنتاجى إضافة إلى الطريقة التقليدية .

أهميـة البحث :

تتمثل أهمية البحث فيما يلى :

١- قد تساعد دراسة ألطرق الوزنية كطرق مقترحة لتقدير الدرجات ومدى فاعلية هذه الطرق مقارنة بالطريقة التقليدية على جودة الإختبارات مرجعية المحك من خلال إعطاء قيم أعلى للصدق والذى يعد أحد شروط الإختبار الجيد .

٢- الإرتقاء بعملية القياس والتقويم النفسى والتربوى من خلال جودة أدواته كى تواكب التطوير المستمر فى نظم التعليم .

٣- قد يفيد القائمين على تصميم أدوات القياس إلى اختيار إحدى الطرق الوزنية للتصحيح التى تعطى أعلى صدق وتعالج مشكلة التخمين فى الإختبارات مرجعية المحك ذات الإختيار من متعدد .

مصطلحات البحث :

١- **الإختبارات التحصيلية مرجعية المحك** Criterion - Refrenced Tests هى "تلك الإختبارات التى تستخدم لتقدير أداء الفرد بالنسبة إلى نطاق سلوكى

(مجموعة من المعارف والمهارات المعرفة بطريقة إجرائية) بغض النظر عن علاقة أدائه بأداء غيره من الأفراد الذين طبق عليهم نفس الإختبار " (٦ : ٤١)

(٢) المحك : Criterion – هو "مستوى أداء معين ذى مواصفات محددة يمكن استخدامه فى مقارنة أداء الفرد بهدف تقويم ذلك الأداء" (٢٦: ١٥).

(٣) الطرق الوزنية لتقدير الدرجات : هى "نظم وطرق للتصحيح معينة لإعطاء الدرجات وتقديرها حيث يختلف فيها درجة كل مفردة علي حسب صعوبتها بحيث لا يشترط أن تكون جميع درجات مفردات الإختبار متساوية " (١٧ : ٦٢٧ – ٦٣٥)

(٤) صدق الإختبارات مرجعية المحك : « يقصد بالصدق « Validity » مدى ملاءمة الدرجات المستمدة من الإختبار للإستخدامات المعينة المناسبة للغرض الذى بنى من أجله» (١٢:٢٥١)

مفـاهيم البحث :

أولا : الاختبارات مرجعية المحك Critrion - Referenced Tests

وهذا النوع من الاختبارات يحدد أداء الفرد أو مستواه بالرجوع إلى محك محدد أو مستوى أداء معين ، وسوف يتناول الباحث الحالى هذا النوع بشئ من الإسهاب من حيث تعريفاته وخصائصه وأهميته فى العملية التعليمية ، وقبل استعراض تلك التعريفات يود الباحث أن يوضح فيما يلى ماذا يقصد بكلمة "**محك**" "Critrion" ، فهذا المصطلح له أكثر من معنى يستخدم فيه ، ومن أكثر هذه المعانى التى يتضمنها هذا المصطلح أن علماء النفس والباحثين فى مجال التربية يميلون إلى تأكيد هذ المعنى والذى يستخدم عادة فى تجارب التعلم ، أى مستوى الأداء الذى يصل إليه المتعلم ، أو الدرجة التى يحصل عليها فى أداء ما ، أما أخصائيو القياس فهم يميلون إلى استخدام كلمة "**محك**" للإشارة إلى السلوك أو الأداء فى الحياة الواقعية ، ذلك فى مقابل الأداء الصادر عن الفرد . (٢٦ : ٢٤)

ويرى كل من جليزر ونيتكو Glazer & Nitko ١٩٧١ فى (٢٣ : ٢٣) أن مفهوم "**المحك**" يعنى نطاقاً شاملاً من المعارف والمهارات المحددة تحديداً جيداً بحيث يمكن نتيجة لموازنة أداء الفرد فى الاختبار بهذا النطاق أن يتضح ما يستطيع أن يؤديه هذا الفرد أو ما لايستطيع أن يؤديه .

أما بابام Popham ١٩٨٨ (١٦ : ١٤) فيعرف "المحك" على أنه مجموعة معقدة من أنماط السلوك التى تقاس من خلال بعض الإجراءات أو العمليات التى يمارسها الفرد بصورة غيرَ مباشرة .

لذلك يعرف نيتكو Nitko ١٩٨٠ (٢٠ : ٦٠) الاختبارات مرجعية "المحك" (CRTs) على أنها "تلك الاختبارات التى صممت عن قصد لكى تؤدى إلى قياسات تفسر مباشرة بالنسبة إلى معيار أداء محدد" .

أما بابام Popham ١٩٨٨ (١٦ : ١٥ – ١٦) يعرفها على أنها "تلك الاختبارات التى تستخدم فى تقدير أداء الفرد فى نطاق سلوكىBehavioral Domain محدد تحديداً دقيقاً" .

ويعرفها "صلاح أبو ناهية" ١٩٩٤ (٤ : ٦٧) بأنها "تلك الاختبارات التى تحدد أداء الفرد أو مستواه بالرجوع إلى محك محدد أو مستوى أداء معين".

كما يعرفها "صلاح علام" ١٩٩٥ (٦ : ٢٨) بأنها " الاختبارات التى تهتم بقياس مدى تحقق مجموعة متكاملة محددة من المهام ، وتزود المتعلم بأدلة عن درجة مدى تمكنه منها ، كما توضح تقدمه أثناء مطلب تعليمى محدد وما يعترضه من صعوبات تعوق هذا التقدم ، وبالتالى يمكنه أن يجرى تعديلاً فى بيئة تعلمه ونوعية تعليمه" .

أما "أنور الشرقاوى" ١٩٩٦ (٢ : ٢٧ – ٢٨) يعرفها على أنها "تلك الاختبارات التى تذكر شئاً عن الفرد دون الرجوع إلى أداء فرد آخر ، ففى المجال المعرفى يوضح الاختبار ماذا يعرف الفرد أو ماذا يستطيع أن يفعل".

وقد أدى تعدد تعريفات الاختبارات مرجعية المحك CRT إلى تطور مفهوم هذا النوع من الاختبارات ليحمل معنى مزدوجاً كما يوضحه كل من ويرسم وجيرس Wiersm & Jurs ١٩٩٠ (٣٢ : ٧٨) على النحو الآتى .. المعنى الأول : يعنى القياس مرجعى المحك إرجاع أداء الفرد إلى معيار ما ، وهو مستوى أداء محدد ، ويتم تفسير درجات الفرد على الاختبار بصورة مطلقة وليس بصورة تقريبية . والمعنى الثانى : فإن القياس مرجعى المحك يتضمن فكرة عن مجال سلوكى محدد ، أى مجموعة محددة من سلوكيات المتعلم ويتم إرجاع أداء الفرد فى الاختبار إلى مجموعة من السلوكيات المعرفة بشكل محدد ، والمحك فى هذه الحالة هو السلوك المطلوب .

ويتضح من التعريفات السابقة المتعددة للاختبارات مرجعية المحك أنها تصف أداء المتعلم بدلالة أنماط محددة من المهارات أو المهام التعليمية التى يتضمنها الاختبار

ويتضح أيضا أن التركيز فى هذه الاختبارات يكون حول وصول ذلك المتعلم إلى مستوى معين من الأداء على مهارة أو مهمة معينة تغطيها فقرات الاختبار .

ويعتمد تحديد النطاق السلوكى على عدد الأهداف أو المهارات المطلوب قياسها وصياغتها فى عبارات سلوكية بحيث يمكن ملاحظة سلوك المتعلم وقياسه ، وتختلف درجة تحديد النطاق السلوكى بحسب نوعه ومحتواه ومدى نقص الأهداف التى يشتمل عليها ، ولذلك اختلفت الأسماء التى تشير إلى الاختبارات مرجعية المحك بحسب درجة تحديد نطاقها السلوكى على النحو الآتى :

١- الاختبارات مرجعية الهدف (Objective Referenced Tests) (ORT) حيث عرفها كل من ميلمان Millman ١٩٧٤ ، وبابام Popham ١٩٧٨، وهامبلتون Hambleton ١٩٧٨ ، علام ١٩٨١ ، ١٩٨٢ بأنها "تلك الاختبارات التى تقيس مجموعة من الأهداف التعليمية السلوكية ، وهو يشتمل عادة على عدد قليل نسبياً من المفردات أو الاسئلة التى يفترض أنها تمثل جميع المفردات التى تقيس مجموعة الأهداف".

٢- الاختبارات مرجعية النطاق Domain - Referenced Tests (DRT) "تبنى هذه الاختبارات بتحديد نطاق سلوكى شامل من المهام أو المهارات أو المتطلبات تحديداً واضحاً ودقيقاً ، وتنتقى المفردات التى يشتمل عليها الاختبار من هذا النطاق انتقاء عشوائياً " . (٦ : ٢٥)

٣- اختبارات التمكن أو الإتقان Mastery Tests "تصمم هذه الاختبارات من أجل تحديد ما إذا كان فرد معين اكتسب السلوك الذى يهدف برنامج تعليمى أو تدريبى إلى تنميته لديه ، واختبار التمكن أو الإتقان يساعد فى اتخاذ قرارات تتعلق بإتقان الفرد لهدف تعليمى أو مهارة معينة أو نطاق من المهارات المحددة " (٦ : ٢٥)

ويتضح من ذلك أن كل تلك المسميات السابقة تندرج تحت الاختبارات التحصيلية مرجعية المحك ، ولكن يختلف المسمى وفقاً لدرجة تحديد واكتمال النطاق السلوكى للأداء الذى يقيسه الاختبار .

وسوف يتبنى الباحث مسمى الاختبارات التحصيلية مرجعية المحك ذات النطاق السلوكى المحدد تحديداً دقيقاً وهو ما يسمى بالنطاق السلوكى المغلق .

٤- خصائص الاختبارات التحصيلية مرجعية المحك وأهميتها فى العملية التعليمية :

تهتم الاختبارات مرجعية المحك (CRT) بقياس مدى تحقق مجموعة متكاملة محددة من المهام وتزود المتعلم بأدلة عن درجة تمكنه منها ،. كما توضح تقدمه أثناء دراسة مطلب تعليمى محدد وما يعترضه من صعوبات تعوق هذا التقدم ، وبالتالى يمكن أن يجرى تعديلاً فى بيئة تعلمه ونوعية تعليمه مستفيداً فى ذلك من مقترحات معلمه . وحيث إن هذا النوع من الاختبارات يركز على مجموعة محددة من سلوكيات المتعلم ، فإنه يساعد المعلم على أن يحدد ما إذا كان طلابه قد أتقنوا أم لم يتقنوا المهارات المحددة . (١٦ : ١٣٦ – ١٣٧)

ويشير "صلاح علام" ١٩٩٥ (٦ : ٢٢) إلى أهمية الاختبارات مرجعية المحك CRT بأنها تحول الاهتمام من الموازنة بين المتعلمين على أساس نتائج الاختبارات مرجعية المعيار إلى التحقق من كفاءة كل منهم واكتسابه للمعارف والمهارات والاتجاهات المرجوة ، وتشخيص الصعوبات التى تعوقهم فى تحقيق ذلك ، ومن ثم تصميم برامج علاجية مناسبة للتغلب على هذه الصعوبات .

ويتضح من قول نورمان Norman ١٩٨٥ (٢١ : ١٦) أن الاختبارات مرجعية المحك (CRT) تفيد فى التقويم البنائى حيث إن هذا النوع من التقويم يركز على قوة وضعف المتعلمين فى تعلم مجموعة محددة من الأعمال ، ولذلك فإن هذا النوع من الاختبارات المعدة إعداداً جيداً من الممكن استخدامه لتوضيح ماذا تعلم كل تلميذ فى وحدة دراسية محددة . ولأهمية هذا النوع من الاختبارات ولكونه محور الاهتمام فى الدراسة الحالية ، سوف يقوم الباحث الحالى فيما يلى باستعراض خطوات تصميم الاختبارات التحصيلية مرجعية المحك (CRT) .

٥- خطوات تصميم الاختبارات التحصيلية مرجعية المحك :

تُعد عملية بناء وتصميم الاختبارات التحصيلية مرجعية المحك أكثر صعوبة ودقة منها فى حالة الأنواع الأخرى من الاختبارات التحصيلية ، لما تتميز به من خصائص ومهام ، حيث تهدف الاختبارات التحصيلية مرجعية المحك إلى التحقق من اكتساب المتعلم كفايات أو مهارات أساسية تعبر عن نواتج تعليمية محدودة ، وتشخيص الصعوبات التى تصادفه أثناء تعلمه أو تدريبه ، وبذلك تساعد المعلم فى تصميم أساليب تعليمية علاجية مناسبة تيسر على المتعلم تصحيح هذه الأخطاء ومتابعة التعلم من أجل تحقيق الكفايات أو الأهداف المرجو تحقيقها . ولذلك يمر الاختبار التحصيلى مرجعى المحك بخطوات متعددة عند بنائه وتصميمه ويختلف بعضها باختلاف نوع

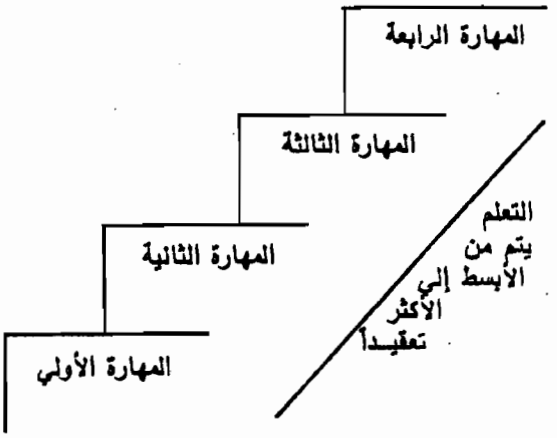
الاختبار وفقاً لطبيعة المحتوى التعليمى أو التدريبى والنطاق السلوكى للمعـارف والمهارات التى يتضمنها ذلك المحتوى .

وفيما يلى عرض لتلك الخطوات ، حيث اتفق كل من بابام ١٩٨٨ Popham (١٦ : ١٢٩ – ١٤٩) ، صلاح علام ١٩٩٥ (٦: ١٠٢ – ١٤١) ، نادية بعيبع ١٩٩٧ (١٠ : ١٧٢ – ١٧٥) عليها :

الخطوة الأولى : تحديد نطاق المحتوى السلوكى الذى يقيسه الاختبار وتحليله

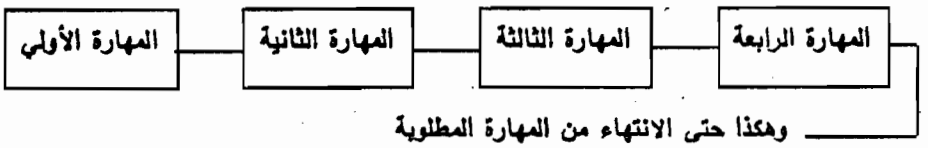
يعتمد تحديد نطاق المحتوى التعليمى أو التدريبى المراد قياسه على طبيعة ومدى اتساع هذا المحتوى ، فإذا كان المحتوى محدوداً مثل وحدة دراسية معينة ، أو موضوع تعليمى واضح المعالم ، فإنه يمكن الاكتفاء بمعرفة مكونات هذه الوحدة أو هذا الموضوع ، أما إذا كان المحتوى التعليمى المراد قياسه متسعاً فإنه يحسن تقسيمه إلى موضوعات فرعية مرتبطة ببعضها بحيث يمكن قياسها كوحدة واحدة . ولكن هذا كله يعتمد على طبيعة المحتوى الدراسى أو التدريبى وما ينطوى عليه من معارف ومهارات ، وهناك أساليب تستخدم لتحليل المحتوى الدراسى منها طريقة تحليل الاجراءاتProcedural Approach ، طريقة التحليل البنائى الهرمى Hierarchical Approach وطريقة تجمع بين تحليل الإجراءات والتحليل الهرميCombination Approach

ويمكن لمصممى الاختبارات التحصيلية مرجعية المحك (CRT) اختيار طريقة التحليل المناسبة للكفايات أو المهارات الرئيسية التى يتضمنها المحتوى الذى يقيسه الاختبار ، عن طريق تصور مكونات المهارة الرئيسية أو كيفية تأديتها أو تحققها تصوراً عقلياً إجمالياً ، وعلى سبيل المثال إذا تبين له أنه من غير الممكن تعلم عناصر أو مكونات المهارة الرئيسية دون تعلم مهارات أخرى فرعية تسبقها ، فإن هذا يعنى أن المهارة تتميز ببنية هرمية كما يتضح فى الشكل رقم (١) ، أما إذا تبين أن تأدية المهارة الرئيسية يتطلب سلسلة خطية متتابعة من الخطوات أو الإجراءات المستقلة ولكنها متكاملة ، فإن تحليل الاجراءات يكون مناسباً فى هذه الحالة كما يتضح من شكل (٢) .



شكل (١)

شكل تخطيطي يوضح مثال لمهارة رئيسية تتميز ببنية هرمية



شكل (٢)

شكل تخطيطي يوضح مثال لمهارة رئيسية تتطلب سلسلة خطية متتابعة من الخطوات والإجراءات

الخطوة الثانية : تديد الأهداف العامة وتحليلها إلى اهداف سلوكية

بعد الانتهاء من عملية تحليل الكفايات أو المهارات الرئيسية للمحتوى التعليمى إلى نواتج من معارف ومهارات فرعية ، ينبغى تحديد الأهداف العامة لذلك المحتوى ، فلكل محتوى تعليمى أو تدريبى أهداف مرتبطة به يسعى المعلم لقياس مدى تحققها لذى المتعلمين .

وتلك الأهداف تعبر عن نواتج متوقعة لعملية التعلم ، ولكنها فى نفس الوقت

تحتاج إلى صياغة سلوكية إجرائية يمكن أن تتخذ دليلاً على أن المتعلم قد حقق كل هدف من الأهداف العامة السابق تحديدها . لذلك يعرف الهدف السلوكى بأنه "الهدف الذى يصف التغير المطلوب إحداثه فى سلوك المتعلم نتيجة للخبرة التعليمية بحيث يمكن ملاحظة هذا السلوك وقياسه" .

ويجب مراعاة أن يكون المستوى المعرفى للأهداف السلوكية الإجرائية متفقاً مع المستوى المعرفى للهدف العام ، لأن ذلك يساعد على اختيار مفردات الاختبار المناسبة . (٧ : ١٠٨ - ١٠٩)

ويشير "صلاح علام" ١٩٩٥ (٦ : ٦١ - ٦٢) إلى أن الهدف السلوكى يتكون من أربعة عناصر هى : (فعل سلوكى - محتوى مرجعى - شرط الأداء - مستوى الأداء)" .

الخطوة الثالثة : إعداد المواصفات التفصيلية للاختبار

يرى بابام Popham ١٩٨٨(١٦١ : ١٤٠ - ١٤٢) ، صلاح علام ١٩٩٥ (٦ : ١٠٥ - ١٠٦) ، نادية بعيبع ١٩٩٧ (١٠ : ١٧٢ - ١٧٥) أن الاختبار مرجعى المحك (CRT) لايكون ذا فائدة فى تحديد ما يستطيع المتعلم أداءه وما لا يستطيع إلا إذا كانت هناك مواصفات تفصيلية واضحة تحدد طبيعة المثيرات التى يمكن أن تشتمل عليها المفردات ، وما تضمنه من موضوعات متعلقة بمحتوى تعليمى معين . كذلك ينبغى أن تتضمن هذه المواصفات كيفية الإجابة الصحيحة وحدودها ، ومحكات الحكم عليها إذا كانت غير مقيدة كما فى أسئلة المقال ، وخصائص ومكونات المشتتات أو الإجابات غير الصحيحة أو الأقل قبولا إذا كانت الإجابات مقيدة كما فى مفردات الاختبار من متعدد والصواب أو الخطأ .

وينبغى أن توضح هذه المواصفات أنواع المفردات التى يمكن استخدامها وتعليمات الإجابة عن كل منها ، وهذه التفاصيل الغرض منها تحديد النطاق السلوكى للاختبار بحيث يمكن الاسترشاد بها فى بناء مفردات اختبارية متجانسة تقيس جميعها هدفاً سلوكياً محدداً ، وهذه المواصفات تشتمل على أربعة عناصر هى :

١- وصف عام موجز لما يقيسه الاختبار
٢- مثال لمفردة اختبارية
٣- خصائص المثيرات
٤- خصائص الاستجابات

الخطوة الرابعة : كتابة المفردات الاختبارية استرشاداً بالمواصفات :

تتطلب هذه الخطوة دراية تامة من جانب مصمم الاختبار بكيفية انتقاء أنسب أنواع المفردات التى تقيس الأهداف السلوكية المحددة قياسا مباشراً ، ومهارة كبيرة فى فنيات بناء هذه المفردات ، كما تتطلب التمكن من المحتوى التعليمى وفهم خصائص المتعلمين . وذلك لأن هذه المفردات تستخدم فى التمييز بين الذين استطاعوا تحقيق الأهداف المحددة والذين واجهتهم صعوبات ، وتشخيص أخطاء وفجوات التعلم .

لذلك قام كل من تـــوم Tom ١٩٨١ (٣١ : ٣٩ – ٤٩) ، بيـتـر Peter ١٩٩٤ (٢٢ : ١٧٤ – ١٨٥) ، صـــلاح علام ١٩٩٥ (٦ : ٦٤ – ٦٧) ، ودونالد ١٩٩٨ (١٣ : ٩١٢ – ٩١٣) بتحديد نوعين أساسيين لمفردات الاختبار هما :

١- مفردات الاستجابة المنتجة أو المصاغة Constructed - Response Items

ويقصد بهذا النوع من المفردات أنه ذو الاستجابة الحرة أو المفتوحة وفيه يكتب المتعلم بأسلوبه الخاص ما يراه مناسباً للمطلوب الذى تحدده المفردة ويتباين طول الإجابة بتباين المطلوب ومن هذا النوع :

أ – مفردات الإكمال والإجابة القصيرة .

ب- أسئلة المقـــــــــــــال .

٢- مفردات الاستجابة المنتقاة أو المختارة Selected - Response Items

ويقصد بهذا النوع أنه ذو الاستجابة المقيدة التى تعتمد على التعرف ، وفيه يختار المتعلم إجابته من بين بدائل إجابات معطاة ، ولا تتدخل العوامل الذاتية للقائم بالتصحيح فى تصحيح الإجابات ، ومن هذا النوع :

أ – مفردات الاختيار من متعدد .

ب- مفردات الصواب أو الخطـــأ .

جـ- مفردات المزاوجـــــــــــة .

ومع ذلك فإن تحديد المعلم أو مصمم الاختبار لشكل المفردات الاختبارية يبقى مرتبطاً إلى حد كبير بالهدف الذى يسعى لقياسه ، فهناك أشكال مختلفة للمفردات بعضها يصلح استخدامه بكفاءة فى قياس أهداف سلوكية معينة ، فى حين أن البعض الآخر لايصلح تماماً لقياس مدى تحقق مثل هذه الأهداف . ويجب أن ترتبط صعوبة كل مفردة اختبارية بمستوى صعوبة الهدف الذى تقيسه ومستواه المعرفى .

الخطوة الخامسة : تحديد صدق محتوى الاختبار مرجعى المحك وثبات درجاته

بعد الانتهاء من كتابة مفردات الاختبار ومعرفة خصائص كل مفردة على حده باعتبار أن كل منها تقيس هدفاً تعليمياً معيناً يحتويه النطاق السلوكى للاختبار ، غير أن هذه المفردات ينبغى أن تعمل معاً كوحدة وظيفية متكاملة ومتسقة تشكل فى مجملها الاختبار باعتباره أداة قياس ، ولكى يفى الاختبار بالغرض الذى أعد من أجله ويستند إلى نتائجه فى اتخاذ قرارات صائبة ينبغى أن يتميز الاختيار بخصائص أخرى وأهمها الصدق والثبات .

أ- صدق الاختبارات مرجعية المحك :

يقصد بالصدق Validity "مدى ملاءمة الدرجات المستمدة من الاختبار للاستخدامات المعينة المناسبة للغرض الذى بنى من أجله" (١٢ : ٢٥١) .

طرق تقدير صدق الاختبارات مرجعية المحك :

توجد ثلاث طرق رئيسية لتقدير صدق الاختبارات مرجعية المحك هى :

١- الصدق الوصفى Descriptive Validity

٢- الصدق الوظيفى Functional Validity

٣- صدق انتقاء النطاق السلوكى : Domain Selection Validity

وسوف يتناول الباحث الصدق بشئ من التفصيل فيما بعد ، باعتباره أحد مفاهيم البحث الحالى .

ب- ثبات الاختبارات مرجعية المحك :

ويقصد بمفهوم الثبات Reliability أنه يشير إلى مدى اتساق ما يقيسه الاختبار ، ويزداد ثبات الاختبار بزيادة درجة هذا الاتساق ، فالاختبار مرجعى المحك يجب أن يتميز بالثبات شأنه شأن الاختبارات مرجعية المعيار ، غير أن الثبات هنا يعنى اتساق قياس نطاق سلوكى معين . (٦ : ٢٥٢)

ونظراً لأن توزيع درجات الطلاب فى الاختبارات مرجعية المحك (CRT) يفترض أن يكون متجانساً إذا كان التعليم ذا فاعلية ، فإن هذا يؤدى إلى انخفاض قيم معامل

الثبات إذا تم تقدير هذه القيم بالطرق التقليدية التى تتبع فى تقدير ثبات الاختبارات مرجعية المعيار ، لذلك فقد ظهرت الحاجة إلى مفاهيم أخرى تتعلق بثبات درجات الاختبارات مرجعية المحك (٦: ٢٦٠ – ٢٦١)

ومن بين طرق تقدير ثبات درجات الاختيارات مرجعية المحك (CRT) التى أشارت إليها أدبيات القياس التربوى والنفسى ، طرق تتطلب تطبيق الاختبار مرة واحدة ، بينما يتطلب البعض الآخر تطبيق الاختبار مرتين بفاصل زمنى على نفس المجموعة ، أو يمكن تطبيق اختبارين متوازيين على مجموعة واحدة .

وقد وضح كل من رونالد بيرك Ronald A. Berk ١٩٨٠ (٢٤ : ٣٢٣ – ٣٤٦) بابام Popham ١٩٨٨ (١٦) رونالد هامبلتون Ronald K. Hambleton ١٩٩٠ (٢٥ : ٢٢ – ٢٦) ، وصلاح علام ١٩٩٥ (٦ : ٢٦٣ – ٢٩٢) العديد من هذه الطرق منها :-

١- معامل كابا لسواميناثان ، هاميلتون ، وألجاينا ١٩٧٨ (٢٦ : ١ – ٤٧)

Swaminathan, Hambleton and Algina; Kappa Coefficiet

٢- معامل الاتفاق لسابكوفاك

Subkoviak Coeffcient of Agreement (Pc)

٢- معامل ليفينجستون ***Livingston Index*** **١٩٧٢** (٢٥ : ٣٢٣ – ٣٤٦)

ثانيا : صدق الإختبارات مرجعية المحك :

يقصد بالصدق Validity مدى ملاءمة الدرجات المستمدة من الإختيار للإستخدامات المعينة المناسبة للغرض الذى بنى من أجله (١٢ : ٢٥١) ، ولذلك يعتبر الصدق من أهم الشروط التى يجب توافرها فى الإختبار الجيد ، لذا يتعلق صدق الإختبارات مرجعية المحك بما يقيسه الإختبار وإلى أى حد ينجح فى قياسه ويتوقف ذلك على مدى الوصول إلى تنبوء دقيق أو استنتاج صحيح من الدرجة التى يحصل عليها المفحوص فى الإختبار (١٢: ١٣٣)

ويعد مفهوم الصدق فى الإختبارات مرجعية المحك واحد من أكثر المفاهيم الأساسية أهمية فى مجال القياس النفسى ، ورغم هذا فهو أكثر إثارة للجدل ، وتنوعا فى التعريف ، وهى إختلافات يدور بعضها حول صلاحية المفهوم نفسه ، بينما يدور

البعض الآخر حول فلسفة المفهوم فى صلته بأغراض القياس ، ويدور البعض الآخر حول إجرائية التعريفات وتطبيقاتها . (٣: ٢٧٥).

وقد عرف ثورندايك Thorndike وجانيه Gagen الصدق على أنه تقدير لمعرفة إذا ما كان الإختبار يقيس ما نريد قياسه ، وكل ما نريد قياس ولا شيء ما يراد قياسه أم لا فى (٣ : ٣٠٦). الثانى

طرق تقدير صدق الإختبارات مرجعية المحك:

تختلف طرق وأساليب التحقق من صدق الإختبارات مرجعية المحك إختلافا ملحوظا عن أساليب تقدير صدق درجات الإختبارات مرجعية الجماعة أو المعيار ، وذلك على الرغم من عدم اختلاف المعنى لمفهوم الصدق فى كل من نوعى الإختبارات ، فمفهوم الصدق يتعلق بمنطقية تفسير الدرجات وما يترتب عليها من قرارات صائبة . لذلك توجد ثلاث طرق رئيسية هى :

١- الصدق الوصفى *Descriptive Validity*

يعد الاختبار مرجعى المحك صادقاً إذا استطاع واضعو الاختبار استخدامه لوصف أداء الفرد بالنسبة لنطاق سلوكى يقيسه الاختبار ويكون محدداً تحديداً دقيقاً ، وهذا النوع من الصدق يعد أساساً لأنواع الصدق الأخرى ، فالاختبار الذى يفتقر إلى الصدق الوصفى لايصلح للغرض المشار إليه وبالتالى لاتكون هناك ثقة فى نتائجه . ويتفق كل من سميث Smith ١٩٧٤ ، وسواميناثان وهامبلتون والجانيا Swamimathan . Hambleton and Algina ١٩٧٤ ، وجرونلاند Gronlund ١٩٧٦ على أن أفضل أنواع الصدق وأكثرها أهمية بالنسبة للاختبارات مرجعية المحك هو الصدق الوصفى الذى يناظر صدق المحتوى فى الاختبارات العارية .

٢- الصدق الوظيفى *Functional Validity*

إذا اقتصر استخدام الدرجات المستمدة من الاختبارات مرجعية المحك على وصف أداء الفرد ، فإن هذا لا يمكننا من التنبوء بنجاحه فى عمل معين ، وهذا يتطلب التحقق من نوع آخر من الصدق يسمى الصدق الوظيفى ، والذى يقصد به أن يؤدى الاختبار مرجعى المحك الوظيفة التى صمم من أجلها ، ولا يقتصر فقط على وصف أداء الفرد فى النطاق السلوكى الذى يقيسه الاختبار . ويقابل مفهوم الصدق الوظيفى مفهوم الصدق المرتبط بمحك فى الأنواع الأخرى من الاختبارات .

٣- صدق انتقاء النطاق السلوكى : *Domain Selection Validity*

فى ضوء ما تم عرضه سابقاً يتضح أن الخطوة الأساسية فى بناء الاختبارات مرجعية المحك هى تحديد النطاق السلوكى لمفردات الاختبار وصياغة مواصفات تفصيلية أو قواعد منظمة لهذا النطاق ، ولابد أن يتميز هذا النطاق السلوكى بالصدق ، ويرتبط صدق النطاق بمفهوم إمكانية التعميم الذى نادى به كرونباك Cronback ١٩٧٣ فى نظريته التى تناولت هذا المفهوم Generalizability Theory ، ويؤيد هذا الاتجاه رونالد هامبلتون Ronald Hambleton ١٩٩٠ (٢٤ : ٢٦ - ٣٠) ، فالنطاق السلوكى الذى نختاره من بين النطاقات السلوكية الممكنة يجب أن يسمح بالتعميم على النطاق الشامل للبعد الذى يقيسه الاختبار .

ولتوضيح ذلك نفترض أنه عند قياس قدرة المتعلم على تحليل الأفكار التى ينطوى عليها موضوع ما تحليلاً سليماً فعند تصنيف السلوك الذى تتضمنه هذه القدرة إلى أقسام متعددة يجب أن ننتقى من بين هذه الأقسام النطاق السلوكى الذى لو أتقنه المتعلم يمكن تعميمه علىَ النطاقات الثلاثة الأخري وبذلك يمكن تحديد النطاق السلوكى الذى يتميز بإمكانية تعميم أداء المتعلمين فيه . وهذا النوع من الصدق يشبه إلى حد ما الصدق التكوين الفرضى Construct Validity فى الأنواع الأخرى من الاختبارات التحصيلية .

ثالثا:الطرق الوزنية لتقدير الدرجات للإختبارات مرجعية المحك ذات الإختيار من متعدد :

تعد طرق تقدير الدرجات نظم معينة لإعطاء الدرجات أثناء التصحيح يختلف فيها درجة كل مفردة على حسب صعوبتها بحيث لا يشترط أن تكون جميع درجات الأسئلة متساوية . وفيه يطلب من المفحوص أحيانا أن يضع نسبة وزنية وتقيم درجته على اساس هذه النسبة مباشرة أو بتطبيق معادلات خاصة عليها ، واحيانا يطلب من المفحوص أن يضع رتبة على كل إجابة تعبر عن درجة صحتها وعامل فى التصحيح على اساس هذه الرتب .

ولذلك وجه إلى الإختبارات ذات الإختيار من متعدد كثير من النقد . ويرجع ذلك إلى عامل التخمين الذى يحدث من جانب المفحوصين وذلك فى حالة عدم تأكدهم من الإجابة الصحيحة ، وتسمى هذه الطريقة بالطريقة التقليدية فى الإجابة ، حيث يجيب الممتحن عن كل مفردة (أى محاولة واحدة لإجابة كل مفردة) ، وعلى أساس ذلك إما أن يحصل المفحوص على الدرجة (واحد صحيح) عن المفردة إذا كانت إجابته صحيحة ، أو يحصل على الدرجة (صفر) إذا كانت إجابته خطأ ، وهذا الإسلوب فى تقدير

الدرجات لا يحقق العدالة فى التصحيح ، لأن الممتحن الذى حصل على الدرجة (صفر) عن مفردة ما يكون فى كثير من الأحيان لديه معلومات جزئية عن الإجابة الصحيحة لهذه المفردة Pariol Knowledge ويستحق عليها تقديرا جزئيا من الدرجات Par-tiol Credit .. ، واتباع هذا الأسلوب التقليدى لا يتيح للمصحح الفرصة لتقدير تلك المعلومات الجزئية والتغلب على عامل التخمين . (١٨: ٢١٩-٢٢٠) .

وفيما يلى عرضا لبعض الطرق الوزنية التى تستخدم فى تقدير درجات الإختبارات ذات الإختيار من متعدد وهذه الطرق هى :

١- طريقة الدرجات التجريبية : Expetemmtel S corig Method

فى هذه الطريقة يطلب من الممتحن أن يضع أمام كل إختبار يختاره رتبة معينة تدل على مدى ثقته فى الإجابة فإذا كان يعتقد أن إجابة ما هى الصحيحة يضع أمامها (١) فإذا كانت صحيحة تكون درجته (٣) . وإذا كان يعتقد أن إجابة أخرى ربما تكون صحيحة يضع أمامها رقم (٢) فإذا كانت إحداهما صحيحة تكون درجته (٢) . أما إذا كان يعتقد أن إجابة ثالثة ربما تكون صحيحة يضع أمامها رقم (٣) ، فإذا كانت إحداهما صحيحة تكون درجته (١) . ولا يسمح للممتحن بوضع رتب أمام الأربعة إختيارات ، وأن حدث تكون درجته صفر . (١٤: ١٢٩-١٣٣) (٢)

٢- الطريقة الإقليدية Euclideam Method

وفيها يطلب من الممتحن أن يعطى توقعات إحتمالية أو درجات ثقة على كل الإختبارات فى صورة نسبة مئوية ، بحيث يكون مجموع النسب ١٠٠٪ ثم يتم حساب متوسط التوقعات الإحتمالية ويرمز له بالرمز م (س) ثم تستخدم الدالة الأتية :

$$\text{الدرجة الإقليدية } (ا) = ١ - \frac{\left[\text{مجـ} \frac{ن}{س-١} \text{ ح} \times \text{س} - \text{م}(\text{س})^{٢}\right]^{\frac{١}{٢}}}{\sqrt{٢}}$$

وهكذا بالنسبة لبقية المفردات ، ثم تجمع الدرجات جمعا عاديا للحصول على الدرجة الكلية . (٢٣: ١٦٥-١٧٠) .

(٣) طريقة الإختيار الإستنتاجى Imferred Choice Method

فى هذه الطريقة يطلب من الممتحن أن يعطى توقعات إحتمالية أو درجات ثقة على كل الإختبارات فى صورة نسبة مئوية بحيث يكون مجموع النسب التى يضعها الممتحن ١٠٠٪ ، فإذا كانت النسبة الموضوعة للإختيار نسب أكبر نسبة فى الاختيار الصحيح تكون درجة المِمتحن (١) ، أما إذا كان هناك نسب أكبر منها فى الاختيار الخطأ فإن درجة الممتحن (صفر) (٢٣: ١٦٩- ١٧٠)

(٤) الطريقة التربيعية Squaral Method

فى هذه الطريقة يطلب من الممتحن أن يعطى توقعات إحتمالية أو درجات ثقة علي كل الإختبارات فى صورة نسبة مئوية بحيث يكون مجموع النسب التى يضعها الممتحن ١٠٠٪ ثم تحسب درجة المفردة من المعادلة التالية : وهكذا بالنسبة لبقية المفردات ثم تجمع جمعا عاديا .

الدرجة التربيعية = ٢ ح (ق) - مجـ [ح (س)]^{٢} (١٩: ٢٩٧ - ٣٠٣)

(٥) الطريقة اللوغاريتمية : Logarithmic Method

وفيها يطلب من الممتحن أن يعطى توقعات إحتمالية ، أو درجات ثقة على كل الإختيارات فى صورة نسبة مئوية ، بحيث يكون مجموع النسب التى يضعها الممتحن ١٠٠٪ ، وفى هذه الطريقة يتم تقدير درجة أى سؤال بواسطة المعادلة الأتية :

الدرجة اللوغاريتمية ل = $\frac{١}{٢}$ [٢ لو ١٠ (ق)]

(٢٣: ١٦٥ - ١٧٠)

(٦) طريقة الإحتمال المقترح للإجابة الصحيحة
Probility Assigned to Correct Respose

فى هذه الطريقة يطلب من الممتحن أن يعطى توقعات إحتمالية أو درجات ثقة على كل الإختيارات التى تتكون منها المفردة فى صورة نسبة مئوية ، بحيث يكون مجموع النسب ١٠٠٪ ثم يؤخذ التوقع الإحتمالى المعطى للإختيار الصحيح على أنه درجة المفردة . (٢٣: ١٦٥)

(٧) الطريقة الكروية : Spherical Method

فى هذه الطريقة يطلب من الممتحن أن يعطى توقعات إحتمالية أو درجات ثقة على كل الإختبارات فى صورة نسبة مئوية ، بحيث يكون مجموع النسب التى يضعها الممتحن ١٠٠٪

فإذا رمزنا للتوقع الإحتمالى المعطى للإجابة الصحيحة بالرمز (ق) ولبقية الإحتمالات بالرمز (س) ، فإن درجة أى سؤال نسب بواسطة المعادلة التالية :

$$\text{الدرجة الكروية ك} = \frac{\text{ح (ق)}}{\left[\text{مجـ}\frac{\text{ن}}{\text{س}-١}\ (\text{ح (س)})^{٢}\right]^{\frac{١}{٢}}}$$

(٢٣: ١٦٥ – ١٧٠)

الدراسات والبحوث السابقة :

تعددت الدراسات التى تناولت أثر الطرق الوزنية لتقدير الدرجات على صدق الإختبارات سواء مرجعية المعيار أو مرجعية المحك وفيما يلى عرض لتلك الدراسات وذلك فى تسلسل تاريخى على النحو التالى :

حيث هدفت دراسة كل من بوجيت وكوهلر ١٩٧١ Poget & Koehler إلى معرفة أثر استخدام طريقتين من الطرق الوزنية لتقدير الدرجات على ثبات وصدق اختبار تحصيلى مقسم إلى ثلاثة إختبارات فرعية تقيس مفرداتها اللغة – الدراسات الإجتماعية والعلوم ، حيث يتكون كل إختبار فرعى من ٢٠ مفردة من نوع الإختيار من متعدد

وقد بلغ حجم العينة الكلية ٥٣٥ طالبا من مدرستين مدينة نيويورك بالصف الحادى عشر ، حيث قسمت العينة إلى ثلاث مجموعات على النحو التالى :

- المجموعة الأولى ١٨٠ طالبا وطلبوا منهم الإجابة بالطريقة التقليدية ،
- أما المجموعة الثانية ١٥٦ طالبا واجابوا عن طريق وضع نسب إحتمالية بحيث تكون النسب تساوى واحد صحيح .
- أما المجموعة الثالثة فقد بلغت ١٩٩ طالبا حيث طلب منهم وضع درجة لكل إختبارإحتمالى تعبر عن مقدار صوابه بحيث يتم توزيع الدرجات على كل الإحتمالات الإختيارية للمفردة الواحدة .

وتم معالجة إجابات الطلاب فى المجموعات الثلاث على النحو التالى :

المجموعة الأولى بالطريقة التقليدية والمجموعة الثانية بالطريقة التربيعية المجموعة الثالثة بالطريقة اللوغاريتمية وتم حساب معاملات الثبات بواسطة معامل الفا واتضح وجود فروق طفيفة لصالح الطرق الوزنية مقارنة بالطريقة التقليدية .

أما معاملات الصدق فقد تم حساب صدق التكوين بالفرض حيث أوضحت النتائج وجود فروق طفيفة فى الصدق لصالح الطريقة التقليدية مقارنة بالطريقتين الوزنيتين (١٩:٢٩٧-٣٠٣)

أما دراسة جيمس ١٩٧٥ Gams G. Diamend فقد هدفت إلى الكشف عن أثر طريقتين من طرق تقدير الدرجات على صدق وثبات أربعة إختبارات تحصيلية فى مادة القياس التربوى ، حيث بلغت حجم العينة ٨٤ طالبا من طلاب الجامعة ، حيث تم بناء أربعة اختبارات تحصيلية تختلف فى عدد مفرداتها حيث كانت على النحو التالى الإختبار الأول ٢٠ مفردة ، وكان الهدف من هذا الإختبار إجراء قياس قبلى لأفراد العينة وتدريب الطلاب على الطرق المستخدمة فى تقدير الدرجات وهما طريقتين إحداهما الطريقة التقليدية ، والثانية تتوافر فيها معلومات على شكل توجيهات تعطى للممتحنين ، وهذه الدرجات تسمى تجريبية . أما الإختبار الثانى فقد بلغ ٣٢ مفردة ، والثالث ٣٢ مفردة ، والرابع ٣٦ مفردة ، وهذه الإختبارات الثلاث تعطى جوانب من المحتوى الكلى للمنهج .

وتم تطبيق تلك الإختبارات على أفراد العينة وطلب من بعضهم الإجابة بالطريقة التقليدية والبعض الأخر بالإجابة بالطريقة التجريبية ، وقد اسفرت النتائج عن :

أن معاملات الصدق التى تم حسابها عن طريق اختيار آخر كمحك أنها غير واضحة الفروق ، حيث لا تدل أى من الطريقتين افضل من الأخرى فى إعطاء معاملات صدق كبيرة . أما فى الثبات فقد أعطت الطريقة التجريبية ثباتًا أكبر من الطريقة التقليدية فى (١ : ٢٩٤ - ٢٩٧)

بينما هدفت دراسة كل من برانبيم وتاتسوك Birenbbbaum & Tat-suoka ١٩٨٣ إلى الكشف عن فاعلية بعض الطرق الوزنية فى تقدير الدرجات لإختبارا تحصيليا مكون من ٦٤ مفردة مقسمة إلى أربع صور متكافئة

تحتوى كل صورة على ١٦ مفردة وتم تطبيقها على عينة قوامها ١٢٥ تلميذاً فى الصف الثامن ، وبعد التصحيح بالطريقة التقليدية والطريقة التجريبية المقترحة تم حساب صدق الإختبار باستخدام محك آخر ، وجد أن هناك فروقا دالة إحصائيا لصالح طريقة تقدير الدرجات التجريبية .(١١: ١٧ - ٢٦) .

أما دراسة جرادات Garadat ١٩٨٨ هدفت إلى المقارنة بين عدد من طرق تقدير درجات الإختبارات مرجعية المحك ذات الإختيار من متعدد فى تأثيرها على صدق وثبات تلك الإختبارات وكذا للتغلب على عامل التخمين عند الإجابة على مثل هذا النوع من المفردات ، حيث بلغت حجم العينة ٥٤ طالبا جامعيا وبلغت مفردات الإختيار ٢٤ مفردة من نوع الإختيار من متعدد ، وبعد تطبيق الإختبار على أفراد العينة المقسمة إلى مجموعتين إحداهم ضابطة والأخرى تجريبية وباستخدام الطريقة التقليدية والطريقة الإقليدية فى تقدير الدرجات تم حساب معاملات الصدق والثبات لمفردات الإختبار المستخدم حيث تبين أن الفروق فى معامل الثبات غير دالة بين الطريقتين بينما عند حساب معامل الصدق وجد أن هناك فروقا واضحة وارتفاع ملحوظ فى معامل الصدق عند استخدام الطريقة التجريبية المذكورة . (١٧: ٦٢٧-٦٣٥).

كما هدفت دراسة روس-ستيفن١٩٩٤ Ross-Steven إلى الكشف عن أثر إختلاف طرق تقدير الدرجات على كل من الصدق والثبات قبل وبعد تقديم المحتوى للإختبارات التحصيلية مرجعية المحك ذات الإختيار من متعدد فى اللغة ، واستخدام تلك النتائج فى تطوير دراسة اللغة لدى المتعلمين ، حيث بلغت العينة ١٢٧ طالبا بالمدارس الثانوية مما يدرسون اللغة وتم تطبيق الإختبار قبل تقديم المحتوى التعليمى وايضا بعد تقديم المحتوى وباستخدام طريقة تقليدية وأخرى تجريبية لتقدير الدرجات ، حيث أتضح من النتائج تباين فى قيم الثبات والصدق بالنسبة لدرجات الإختبارات التى تم تطبيقها قبل وبعد تقديم المحتوى وذلك لصالح الإختبارات التى طبقت بعد تقديم المحتوى أما بالنسبة لقيم الثبات والصدق مع اختلاف طرق تقدير الدرجات أتضح أن هناك فروق دالة فى قيم الثبات لصالح الطريقة التجريبية ، أما بالنسبة لقيم الصدق كان هناك تباين فى الفروق أيضا لصالح الطريقة التجريبية مقارنة بالطريقة التقليدية .(٢٧)

أماريوز - بريمو ١٩٩٨ Ruiz-primo فقد قام بدراسة المقارنة صدق وثبات الإختبارات مرجعية المحك فى مادة الكيمياء وطريقتين من طرق تقدير الدرجات وكانت أحدى الطرق تمثل تغذية راجعة كلية عند الإجابة على مفردات الإختبار والطريقة الأخرى تستخدم فيها تغذية راجعة جزئية ، وتم تقسيم العينة التى بلغت ١٥٢ من طلبة

المدارس الثانوية فى سبعة فصـــول إلى مجموعتين تجريبيتين ، وطلب من كل مجموعة أن تجيب على الإختبار بالطريقة المحددة بناء على التعليمات ، وتقدير الدرجات بالطريقتين السابقتين نتبين أن معاملات الثبات تختلف كثيرعند أستخدام الطريقتين أما معاملات الصدق كان هناك تمـــايز أيضا ولكن فروق غير دالة لصالح أحـــــد الطريقتين .(٢٨)

وتوحى هذه الدراسة بإجراء العديد من الدراسات لإستخدام طرق وزنية أخرى لتقدير الدرجات وذلك لتحسين كل من معاملات الصدق والثبات للإختبارات مرجعية المحك .

التعقيب علي الدراسات السابقة ومدي استفادة الباحث

١- معظم الدراسات السابقة إهتمت بالصدق والثبات الكلى للإختبار دون الإهتمام بصدق وثبات مفردات الإختبار .

٢- يتضح أن معظم الدراسات ركزت على أخذ عينات من طلاب الجامعات والمدارس الثانوية وكان هناك إهمال واضح بالنسبة لباقى المراحل .

٣- اعتمدت معظم الدراسات على تقسيم العينة إلى عدد من المجموعات منه الضابطة والتجريبية ومنها ما قام بتقسيم العينة إلى مجموعتين تجريبيتين .

٤- تباينت الدراسات فى نوع الإختبار المستخدم ، فمنها ما استخدم إختبارات مرجعية المعيار والبعض الآخـر إستخدمت إختبارات مرجعية المحك .

٥- تراوح حجم العينات من ٥٤ إلى ٥٣٥ فى معظم الدراسات .

٦- تنوعت طرق وأساليب حساب معاملات الصدق ولكن معظمها إستخدم صدق المحك فى الإختبارات مرجعية المعيار أو ما يسمى بالصدق الوظيفى فى الإختبارات مرجعية المحك

٧- تعددت الأساليب الإحصائية لإظهار الفروق بين الطرق الوزنية فى تقدير الدرجات فى تأثيرها على صدق الإختبارات ، فمنها ما استخدم بعض مقياس الإحصاء الوصفى مثل المتوسط والوسيط والإنحراف المعيارى ومعامل الإلتواء ، ومنها استخدم معاملات الإرتباط ومنها استخدم معادلة (ذ) للفروق بين النسب .

٨- اختلفت معظم الدراسات فى نتائجها ، فمنها ما أظهر فروقاً دالة لصالح الطرق الوزنية فى الثبات ومنها ما تبين أن الفروق لصالح الطرق التقليدية ، والبعض

الأخر لم يظهر فروق واضحة وغير دالة لأى من الطرق سواء التقليدية أو التجريبية .

وتتحدد استفادة الباحث من خلال هذا التعقيب على الدراسات السابقة فى تحديد حجم العينة وتحديد نوع الإختبار التحصيلى – التصميم التجريبى للبحث – اختبار الطرق الوزنية التجريبية – إختبار الأساليب الإحصائية المناسبة لمعالجة البيانات الخاصة بالدراسة ، كما يمكن صياغة الفرض الثانى كإجابة محتملة على السؤال الذى أثير سابقا فى مشكلة البحث الحالى :

الفرض الرئيسي للبحث :

"لاتوجد فروق ذات دلالة إحصائية بين مجموعة الطريقة التقليدية فى تقدير الدرجات وبعض الطرق الوزنية (طريقة الدرجات التجريبية – طريقة الاحتمال المقترح – الطريقة التربيعية – طريق الاختيار الاستنتاجى) فى صدق درجات الاختبار التحصيلى مرجعى المحك .

ويشتق من الفرض الرئيسى الفروض الآتية على النحو التالى :

فـــروض البحـــث :

١- لا توجد فروق ذات دلاله إحصائية بين مجموعتى الطريقة التقليدية وطريقة الدرجات التجريبية فى صدق درجات الإختبار مرجعى المحك .

٢- لاتوجد فروق ذات دلالة إحصائية بين مجموعتى الطريقة التقليدية وطريقة الإحتمال المقترح فى صدق درجات الإختبار مرجعى المحك .

٣- لاتوجد فروق ذات دلالة إحصائية بين مجموعتى الطريقة التقليدية والطريقة التربيعية فى صدق درجات الإختبار مرجعى المحك

٤- لاتوجد فروق ذات دلالة إحصائية بين مجموعتى الطريقة التقليدية وطريقة الإختيارالإستنتاجى فى صدق درجات الإختبار مرجعى المحك

٥- لاتوجد فروق ذات دلالة إحصائية بين مجموعتى طريقة الدرجات التجريبية وطريقة الإحتمال المقترح فى صدق درجات الإختبار مرجعى المحك .

٦- لاتوجد فروق ذات دلالة إحصائية بين مجموعتى طريقة الدرجات التجريبية والطريقة التربيعية فى صدق درجات الإختبار مرجعى المحك .

٧- لاتوجد فروق ذات دلالة إحصائية بين مجموعتى طريقة الدرجات التجريبية وطريقة الإختيار الإستنتاجى فى صدق درجات الإختبار مرجعى المحك .

٨- لاتوجد فروق ذات دلالة إحصائية بين مجموعتى طريقة الإحتمال المقترح والطريقة التربيعية فى صدق درجات الإختبار مرجعى المحك

٩- لاتوجد فروق ذات دلالة إحصائية بين مجموعتى طريقة الإحتمال المقترح وطريقة الإختيار الإستنتاجى فى صدق درجات الإختبار مرجعى المحك

١٠- لاتوجد فروق ذات دلالة إحصائية بين مجموعتى الطريقة التربيعية وطريقة الإختيار الإستنتاجى فى صدق درجات الإختبار مرجعى المحك

إجراءات البحث :

فيما يلى يتناول الباحث إجراءات البحث الحالى وتتضمن التصميم التجريبى للبحث ، وكذلك خصائص أفراد العينة ، والأداة المستخدمة من حيث خطوات الإعداد ، وقد اعتمد فى ذلك على نتائج الدراسات والبحوث السابقة المتعلقة بمجال البحث الحالى والتى تم عرضها فيما سبق وذلك للتحقق من صحة فروض البحث الحالى :

أولا : التصميم التجريبي للبحث EXPERIMENTAL DESIGN

يستخدم البحث الحالى المنهج التجريبى طبقا للمتغيرات التالية :

١- المتغير المستقل : هو الطرق الوزنية لتقدير الدرجات وهى :

١- الطريقة التجريبية ٢- الطريقة التربيعية

٣- طريقة الإحتمال المقترح ٤- طريقة الإختبار الإستنتاجى

١- المتغير التابع :

هو الصدق فى الإختبارات مرجعية المحك ذات الإختيارات من متعدد ويتطلب التصميم التجريبى للبحث الحالى ثلاث مجموعات من تلاميذ الصف الثالث الإعدادى . حيث يكون هناك مجموعة ضابطة ومجموعتان تجريبيتان ويوضح ذلك الجدول التالى :

جدول رقم (١)

التصميم التجريبى للبحث الحالى

مجموعات البحث	طرق تدريس الدرجات
المجموعة الضابطة (ض)	الطريقة التقليدية
المجموعة التجريبية الأولى	طريقة الدرجات التجريبية
جـ١	١- طريقة الاحتمال المقترح
المجموعة التجريبية الثانية	٢- الطريقة التربوية
جـ٢	٣- طريقة الاختيار الاستنتاجى

ثانيا : المشاركون في البحث

تم اختار المشاركون فى البحث الحالى من تلاميذ الصف الثالث الإعدادى ممن درسوا وحدة التكاثر فى الكائنات فى محافظة الشرقية بإدارة ديرب نجم التعليمية من ثلاث مدارس إعدادية حيث تم اختيار المدارس عشوائيا باستخدام جدول الإعداد العشوائية (٩ : ٣٧٦- ٣٧٧) ، وتم اختيار ثلاثة فصول من كل مدرسة بنفس الطريقة السابقة ، وقد بلغ العدد المبدئى لتلاميذ الفصول التسعة ٣١٧ تلميذاً . وبعد استبعاد الطالبات فى المدرسة المشتركة والطلبة الباقين للإعادة وأيضا الذين تخلفوا عن إجراء أحد الإختبارات بلغ عدد العينة الكلية ٢٨٩ تلميذا . حيث تراوح العمر الزمنى لهؤلاء التلاميذ بين ١٣ ، ١٤ عاما بمتوسط قدره ١٤,٥٦ وإنحراف معيارى قدره ٠,٥٨ . وتم تحديد كل ثلاث فصول فى مدرسة لمجموعة طبقا للتصميم التجريبى وتم ذلك بطريقة عشوائية أيضاً .

جدول رقم (٢)

اسماء المدارس التى طبقت فيها الاختبارات وعدد الفصول التى اختيرت بكل منها وعدد التلاميذ

اسم المدرسة	عدد الفصول	عدد التلاميذ	اسم المجموعة
مدرسة التبين الإعدادية	٣	٩٨	المجموعة الضابطة
مدرسة عمر بن الخطاب ع	٣	٩٦	المجموعة التحريبية (١)
مدرسة المشتركة الإعدادية	٣	٩٥	المجموعة التجريبية (٢)

ثالثا : الأداة المستخدمة في البحث

بناء إختبار تحصيلى مرجعى المحك : C - R - T

تتطلب الدراسة الحالية بنا ء إختبار تحصيلى مرجعى المحك من صورتين متكافئتين لقياس تحصيل الطلاب فى وحدة التكاثر فى الكائنات الحية المقررة ضمن منهج العلوم للصف الثالث الإعدادى .

وفيما يلى خطوات بنا ء الإختبار التحصيلى مرجعى المحك فى وحدة التكاثر :

الخطوة الأولي :

تحديد النطاق السلوكى الذى يقيسه الإختبار وتحليله : ويندرج تحت هذه الخطوة عدد من العناصر ، هى :

١ – تحديد محتوى الموضوعات فى الوحدة الدراسية :

قام الباحث بتحديد العناصر الرئيسية فى وحدة التكاثر فى الكائنات الحية ثم قام بتحليل كل عنصر رئيسى إلى عدد من العناصر الفرعية ، وكذلك تحديد الأهمية النسبية لكل موضوع رئيسى فى ضوء الموضوعات الفرعية التى يشملها داخل المحتوى التعليمى ، وقد استخدمت المعادلة التالية لتحديد الأهمية النسبية ·

$$\text{الأهمية النسبية للعنصر الرئيسى} = \frac{\text{عدد الموضوعات الفرعية التى تندرج تحت هذا العنصر}}{\text{إجمالى عدد الموضوعات الفرعية فى المحتوى التعليمي}} \times ١٠٠$$

وهذه الموضوعات وما يندرج تحتها من موضوعات فرعية موضحة بالجدول رقم (٣)

جدول رقم (٣)

الموضوعات الرئيسية لوحدة التكاثر فى الكائنات الحية وما يندرج تحتها من موضوعات فرعية والأهمية النسبية لكل منها

الموضوع الرئيسي	الموضوعات الفرعية	عدد الموضوعات الفرعية	النسبة المئوية
١- صورة التكاثر فى الكائنات الحية	أ - مفهوم عملية التكاثر ب- أهمية عملية التكاثر جـ- التكاثر اللاجنسى (اللاتزاوجى) ١- كيفية حدوثه ٢- أمثلة لبعض الكائنات التى تتكاثر لاتزاوجيا (لاجنسى) د- التكاثر الجنسى (التزاوجى) ١- كيفية حدوثه ٢- أمثلة لبعض الكائنات التى تتكاثر تزاوجيا (جنسى)	٤	٢٥٪
٢- التكاثر فى النباتات الزهرية	ا- مكونات الزهرة ب تعريف عضو التكاثر فى النباتات الزهرية جـ- اعضاء التكاثر فى الزهرة . ء- مفهوم عملية التلقيح . هـ- مفهوم عملية الإخصاب . و - التكنولوجيا والإنتاج الغذائى .	٦	٣٧,٥٪
٣- التكاثر فى الإنسان	ا- أهمية التكاثر فى حياة الإنسان ب- تركيب الجهاز التناسلى فى ذكر الإنسان جـ- تركيب الجهاز التناسلى فى أنثى الإنسان ء- هرمونات التكاثر ومظاهر البلوغ . هـ - الصحة الإنجابية . و- بعض الأمراض التى لها علاقة بالجهاز التناسلى وطرق الوقاية منها. ١- الزهرى . ٢- السيلان . ٣- الأيدز .	٦	٣٧,٥٪
المجموع الكلي والنسبة المئوية للموضوعات الفرعية		١٦	١٠٠٪

يتضح من الجدول السابق رقم (٣) أن الموضوعات الفرعية التى تندرج تحت كل موضوع رئيسى متقارب إلى حد ما حيث تراوحت النسبة المئوية للموضوعات الفرعية ما بين ٢٥٪ ٣٧,٥ ٪.

ثم قام الباحث بعرض تحليل محتوى الوحدة الدراسية على مجموعة من المحكمين تتكون من سبعة محكمين متخصصون فى مناهج وطرق تدريس العلوم وأيضا فى مجال علم النفس التعليمى . ثم قام الباحث لحساب نسب الإتفاق بين نتائج التحليل الذى قام به ونتائج التحكيم من قبل المحكمين حيث اتضح أن هناك إتفاق بين المحكمين مع الباحث حول عناصر تحليل محتوى الوحدة الدراسية ، وقام الباحث بإجراء بعض التعديلات التى اقترحها بعض المحكمين .

(ب)- تحديد المهارات الرئيسية وتحليلها إلى مكوناتها من معارف ومهارات فرعية :

قام الباحث الحالى بتحديد المهارات الأساسية (الرئيسية) المرجو تحقيقها لدى الطلاب بعد دراستهم لوحدة التكاثر فى الكائنات الحية ، مسترشدا فى ذلك بنتائج تحليل محتوى الوحدة الدراسية وآراء مجموعة من المحكمين المشار إليهم سابقا . لذلك أيضا خبرة الباحث الشخصية فى تدريس مادة العلوم للصف الثالث الإعدادى والتى تزيد على تسع سنوات حيث توصل الباحث إلى ثلاث مهارات رئيسية ينبغى أن يتمكن منها كل طالب بعد دراسة هذه الوحدة وتلك المهارات هى :

١- تحديد صور التكاثر فى الكائنات الحية .

٢- تفسير التكاثر فى النبـــات الزهرية .

٣- تفسير التكاثـــر فى الإنسان .

وتعد هذه المهارات الرئيسية لهذا الترتيب تنظيما بنائيا ، حيث أن إتقان المتعلم لإحداها يعد متطلبا أساسيا لإتقانه المهارات الرئيسية التالية :

نظرا لأن هذه المهارات الرئيسية ليست مجرد مهارة بسيطة وإنما تتضمن بعض المعارف والمهارات الفرعية التى تسهم فى تحقيقها، فقد قام الباحث بإجراد تحليل سلوكى لهذه المهارات الرئيسية ، باستخدام كلا من أسلوب تحليل البنية التعليمية Structural Analysis والذى اقترحه ساندیورا Suandiora فى (٦ : ١٤٥) ، وأسلوب تحليل المهام التعليمية Tasks analysis والذى اقترحه جانيه Gagné (١٩٩٢) فى (٦ : ١٤٥) . بعد ذلك قام الباحث بتنظيم المعارف والمهارات التى تندرج

تحتها فى شكل هرمى بحسب أولية إسهام كل منها فى تحقيق المهارة الرئيسية المتعلقة بها :

ثم قام الباحث بعرض هذه المهارات الرئيسية وكذلك نواتج تحليلها إلى معارف ومهارات فرعية على مجموعة من المحكمين لإبداء الرأى حول مدى أهمية هذه المهارات الرئيسية للتلاميذ الصف الثالث الإعدادى ، ومدى كفايتها فى تحقيق الهدف من تدريس الوحدة الدراسية المتعلقة بها ، وكذلك أيضا لابد أن الرأى حول مدى كفاية نواتج تحليل كل مهارة منها وترتيبها المنطقى واقتراح ما يرونه مناسبا من تعديلات سواء بالحذف أو بالإضافة .

جدول رقم (٤)

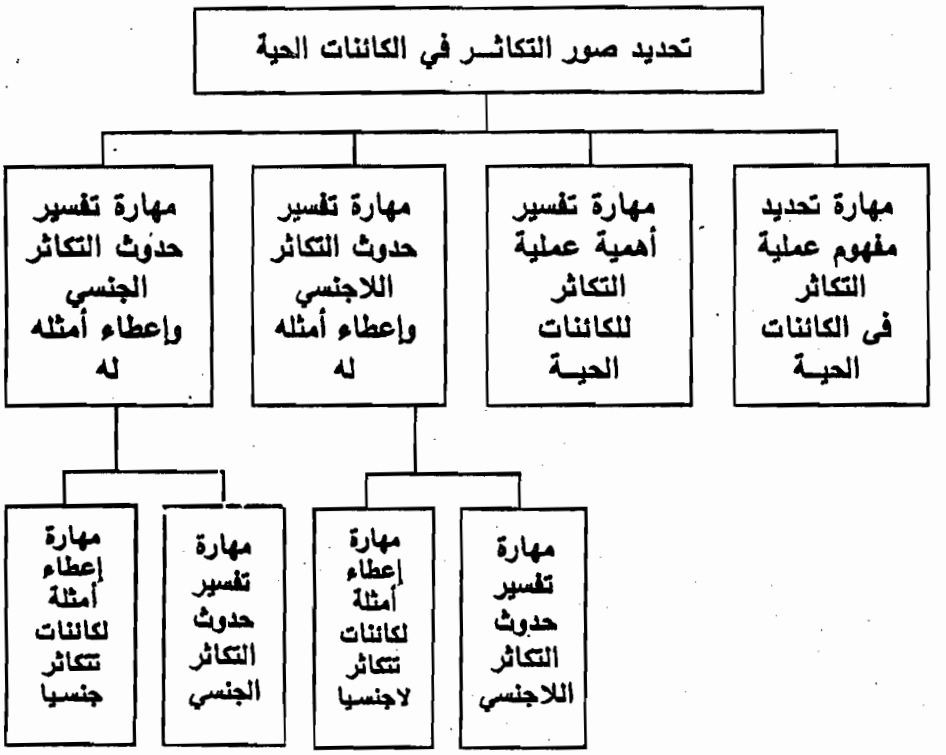
نسب اتفاق مجموعة المحكمين حول أهمية كل مهارة رئيسية وايضا فى تقييم نواتج التحليل السلوكى المتعلقة بالمهارات الثلاث

المهارات / عناصر التحكيم	نسب الاتفاق لكل مهارة ٪		
	الأولى	الثانية	الثالثة
س١ : هل كل مهارة رئيسية لها أهمية فى وحدة التكاثر فى الكائنات الحية لتلاميذ الصف الثالث الإعدادى من حيث تحقيق الهدف من تدريس الوحدة؟	٪١٠٠	٪١٠٠	١٠٠
س٢ : هل نواتج التحليل المتعلق بكل مهارة رئيسة مرتبة ترتيبا منطقيا مترابطا .	٪٨٥,٧	١٠٠	١٠٠
س٣ : هل نواتج التحليل المتعلق بكل مهارة رئيسية كافية لتحقيقها	٪٨٥,٧	١٠٠	١٠٠

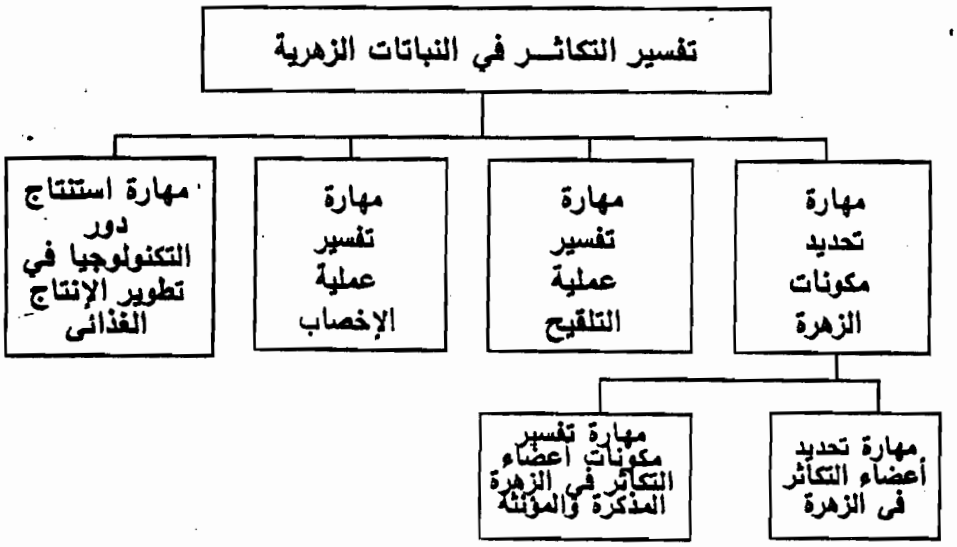
وقام الباحث بإجراء جميع التعديلات التى اقترحها المحكمون وحساب نسب اتفاق المحكمين فى تحديد المهارات الرئيسية لوحدة التكاثر فى الكائنات الحية ، وكذلك حول الصياغة السلوكية لنواتج تحليلها لكل مهارة من المهارات .

وفيما يلى عرضا لتلك المهارات الثلاث الرئيسية فى شكل مخططات توضح نواتج تحليلها فى وحدة التكاثر فى الكائنات الحية والتى تسهم فى تحقيقها كما فى الشكل (٣) ، (٤) ، (٥) .

شكل رقم (٣) المهارة الرئيسية الأولى



شكل (٤) المهارة الرئيسية الثانية



شكل (٥) المهارة الرئيسية الثالثة

الخطوة الثانية :

تحديد الأهداف العامة للمحتوى وتحليلها إلى أهداف سلوكية فى ضوء نواتج التحليل السلوكى السابق للكفايات ويندرج تحت هذه الخطوة عنصرين هما .

ا- تحديد الأهداف العامة للمحتوى الذى يقيسه الإختبار وفقا لمحتويات معرفية متعددة قام الباحث بتحديد تلك الأهداف العامة لوحدة التكاثر فى الكائنات الحية . وقد توصل إلى ثلاثة أهداف عامة بمعدل هدف لكل مستوى معرفى من مستويات بلوم المعرفية الثلاث الأولى وتلك الأهداف هى :

١- الهدف العام الأول فى مستوى (التذكر)

"يعرف التلاميذ المفاهيم والمصطلحات التى يتضمنها محتوى وحدة التكاثر فى الكائنات الحية المقررة على الصف الثالث الإعدادى" .

٢- الهدف العام الثانى فى مستوى (الفهم) .

"يفهم كيفية حدوث عملية التكاثر فى الكائنات الحية" .

٣- الهدف العام الثالث فى مستوى التطبيق.

" تستخدم الأساليب الحديثة والتكنولوجيا فى الإنتاج النباتى والصحة الإنجابية ومعالجة بعض الأمراض التى تصيب الجهاز التناسلى فى الإنسان " .

ثم قام الباحث بصياغة نواتج التحليل السلوكى لكل مهارة رئيسية صياغة سلوكية فى صورة أهداف ، ثم وضعها فى ترتيب منطقى حسب أولوية كل هدف سلوكى منها للإسهام فى تحقيق المهارة الرئيسية المتعلقة به .

تقييم الأهداف السلوكية :

جدول رقم (٥)

نتائج تقييم الأهداف العامة وتحليلها إلى أهداف سلوكية

عناصر التحكيم	نسب الإتفاق
س١ : هل الأهداف العامة والسلوكية واضحة وصياغتها سليمة ؟	٨٥,٧ ٪
س٢ : هل الأهداف السلوكية كافية لتحقيق كل من المهارات الرئيسية الثلاث ؟	١٠٠ ٪

للتحقق من وضوح الأهداف السلوكية وصحة صياغتها تم عرضها على مجموعة من المحكمين حيث كان هناك اتفاق واضح حول صياغة تلك الأهداف وصحة وصياغتها وعلى الرغم من ذلك كان هناك بعض المقترحات ، وقد قام الباحث بإجرائها وفيما يلى عرضا لتلك الأهداف والتى توضح كيفية صياغة الأهداف السلوكية من خلال تحليل الأهداف العامة .

وسوف يستعين الباحث بتلك الأهداف السلوكية كإحدى خطوات بناء الإختبار التحصيلى مرجعى المحك الخاص بالبحث الحالى .

١- الهدف العام الأول فى مستوى التذكر (المعرفى)

« يعرف التلاميذ المفاهيم والمصطلحات التى يتضمنها محتوى وحدة التكاثر فى الكائنات الحية المقررة على الصف الثالث الإعدادى »

ويتم تحليله إلى أهداف سلوكية فى نفس المستوى .

١- يذكر التعريف الصحيح للتكاثر اللاجنسى

٢- يتعرف على التعريف الصحيح للتكاثر الجنسى

٣- يحدد مكونات الزهرة كعضو للتكاثر فى النباتات الزهرية

٤- يحدد مكونات الجهاز التناسلى الذكرى فى الإنسان .

٥- يحدد مكونات الجهاز التناسلى الأنثوى فى الإنسان .

٦- يذكر بعض الكائنات الحية التى تتكاثر لا جنسيا .

٧- يتعرف على بعض الأمراض التى تصيب الجهاز التناسلى فى الإنسان .

٨- يذكر خطوات تكوين الجنـين .

٩- يتعرف على أعراض بعض الأمراض التى تصيب الجهاز التناسلى فى الإنسان

١٠- يتعرف على مظاهر البلوغ من كل من الذكر والأنثى فى الإنسان .

٢- الهدف العام الثانى فى مستوى (الفهم) .

" يفهم كيفية حدوث عملية التكاثر فى الكائنات الحية" ، ويتم تحليله إلى أهداف سلوكية فى نفس المستوى .

١- يفسر أهمية التكاثـر .

٢- يميز بين التكاثر الجنسى والتكاثر اللاجنسى .

٣ - يميز بين الحيوان المنوى والبويضة .

٤- يوضح الفرق بين التلقيح والإخصاب فى النبات الزهرية .

٥- يوضح وظائف مكونات الزهـــــرة فى عملية التكاثر .

٦- تميز بين اللقـاح والبويضة .

٧- يفسر دور الهرمونات ف عملية التكاثر ومظاهر البلوغ .

٨- يفسر سبب حدوث بعض الأمراض التى تصيب الجهاز التناسلى فى الإنسان .

٩- يوضح بعض صور التكاثر اللاجنسى .

٣- الهدف العام الثالث فى مستوى (التطبيق) .

"يستخدم الأساليب الحديثة والتكنولوجيا فى الإنتاج النباتى والصحة الإنجابية ومعالجة بعض الأمراض التى تصيب الجهاز التناسلى فى الإنسان" ، ويتم تحليله إلى

أهداف سلوكية فى نفس المستوى .

١- يعطى أمثلة لبعض الكائنات التى تتكاثر جنسيا ولاجنسيًا.

٢- يفحص قطاع عرض فى زهرة بواسطة المجهر .

٣- يستخدم التكنولوجيا الحديثة فى تحسين أنواع البذور والثمار .

٤- يعطى الطرق والأساليب الصحية التى يجب أن يتبعها الإنسان قبل الزواج .

٥- يعطى بعض الطرق والأساليب الحديثة لعلاج بعض الأمراض التى تصيب الجهاز التناسلى فى الإنسان ..

الخطوة الثالثة : إعداد المواصفات التفصيلية لاختبار

أعد الباحث مواصفات الإختبار مسترشدا بطريقة كل من بابام Papham١٩٧٨ ، وبـــرك ١٩٨٨ Berk وميـلـمـان ١٩٨٩ Melman فى تصميم الإختبارات التشخيصية مرجعية المحك ، والهدف من هذه المواصفات هو تحديد النطاق السلوكى الذى يقيسه الاختبار تحديدا أكثر تفصيلا مما تحدده عبارات الأهداف السلوكية التى تمت صياغتها ، ومن العناصر الأساسية لمواصفات الإختبار :

١- وصف عام موجز لما يقيسه الإختبار .

٢- مثال لمفردة إختبارية .

٣- خصائص المثيرات .

٤- خصائص الإستجابات .

الخطـــوة الرابـــعة : بناء مفردات الإختبار

قام الباحث الحالى ببناء مفردات الإختبار من نوع الإختيار من متعدد ، لقياس كل هدف من الأهداف السلوكية التى تمت صياغتها ومراجعتها . وقد ارتبط مستوى كل مفردة بالمستوى المعرفى الخاص بكل هدف الذى يقيسه (تذكر – فهم – تطبيق) .

حيث كان يتم صياغة مفردتين لكل هدف سلوكى ، وكان الهدف من ذلك بناء صورتين متكافئتين للإختبار التحصيلى مرجعى المحك فى وحدة التكاثر فى الكائنات الحية حيث بلغت كل صورة ٢٥ مفردة إختيارية .

جدول رقم (٦)

نسب اتفاق مجموعة المحكمين فى تقييم المفردات الاختبارية

عناصر التحكيـــم	نسب الاتفـــاق
س١ : هل صياغة كل مفردة اختبارية صحيحة مناسبة لمستوى تلاميذ الصف الثالث الإعدادى ؟	٨٥,٧ ٪
س٢ : هل التعليمات لكل مفردة واضحة وتحقق الشروط الخاصة لنوع المفردات ؟	٨٥,٧٪
س٣ : هل بدائل الإجابات لكل مفردة مناسبة من بينها إجابة واحدة هى الصحيحة .	١٠٠ ٪

بعد بناء المفردات الإختبارية تم عرضها على مجموعة من المحكمين فى مجال القياس والتقويم وتقييم الإختبارات من حيث صياغة كل مفردة وبدائل الإجابات لكل مفردة ، واتضح أن هناك نسب إتفاق بين المحكمين مع وجود بعض الإقتراحات التى قام الباحث بإجرادها لصورتى الإختبار . الصورة النهائية للإختبار ملحق (١) .

تعلـيمات الإخـتبار :

بعد الإنتهاء من بناء المفردات وعرضها على المحكمين فى صورتى الإختبار وإجراء التعديلات المقترحة وصل عدد المفردات لكل صورة من صورتى الاختبار ٢٥ مفردة إختبارية فى الصورة النهائية للإختبار ، وقام الباحث بكتابة تعليمات الإجابة عن الإختبار بصورتيه وطريقة تسجيل الإجابات ، وكذلك زمن الإختبار ، وبعض البيانات المطلوب تدوينها من الممتحنين مثل : الجنس ، السن ، المدرسة الإسم إختياريا .

الدراسـة الإستـطلاعية للإخـتبار

قام الباحث باختيار عينة من تلاميذ الصف الثالث الإعدادى قوامها ٧٠ تلميذا ممن درسوا وحدة التكاثر فى الكائنات الحية بإحدى المدارس بمركز ديرب نجم شرقية ، ثم

بعد ذلك تم تطبيق الإختبار بصورتيه على أفراد العينة الإستطلاعية وبعد مرور ١٥ يوم تم إعادة تطبيقه مرة أخرى . والهدف من تطبيق الإختبار على هذه العينة تحديد زمن الإجابة على مفردات الإختبار وأيضا تقدير ثباته وقياس مدى تكافؤ صورتى الإختبار ، وباستخدام المعادلة المناسبة لتحديد زمن الإختبار ثم تحديد الزمن ٣٠ دقيقة لكل صورة من صورتى الإختبار ، وأيضاً لتحديد درجة القطع باستخدام طريقة انجوف ، حيث تم تحديد درجة الإتقان ٨٠٪ .

ومن حيث تكافؤ صورتى الإختبار اتضح من حساب المتوسط والإنحراف المعيارى بصورتى الإختبار فى أحد التطبيقات تبين أنهما تقريبا متكافئتان حيث بلغ المتوسط بصورة ا (١٧,٤٢) والمتوسط للصورة (ب) (١٧,٦٢) بينما الإنحراف المعيارى بلغ (٣,٩٣) للصورة ا والصورة (ب) بلغ (٣,٨٦).

الخطـوة الخامسة : تقدير الخصائص السيكومترية للإختبار :

(ا) التحقق من صدق محتوى الإختبار :

١- إستخدام الباحث الحالى طريقة الصدق الوصفى Descriptive Validity

ويعد الإختبار مرجعى المحك صادقا إذا ما أمكن باستخدامه وصف أداء الفرد بالنسبة لنطاق سلوكى يقيسه الإختبار ويكون محدد تحديدا دقيقا .

لذا قام الباحث بإعداد إستمارة خاصة بذلك ثم قدم نسخة منها لكل محكم من مجموعة المحكمين مرفقا بها الأهداف السلوكية للمحتوى الخاص بالوحدة ويلى كل هدف منها المفردات التى تقيسه ، وطلب من كل محكم أن يجيب عن التساؤلات التالية الواردة فى الإستمارة إما نعم أو لا . لصورة الإختبار .

س١ هل كل مفردة إختبارية تعد قياسا مناسبا للهدف الذى تقيسه ؟

س٢ هل مفردات الإختبارككل تعد بمثابة عينة ممثلة للأهداف فى إطار محتوى الوحدة الخاصة بالدراسة ؟

س٣ هل مفردات الإختبار ككل متسقة مع الهدف الأساسى منه ؟

هذا وقد قام الباحث بحساب نسب الإتفاق بين المحكمين حول الإجابة على التساؤلات السابقة فكانت كما يوضحها جدول (٧) الآتى :

جدول رقم (٧)

نسب اتفاق المحكمين على صلاحية مفردات الإختبار

عناصر التحكيم	نسبة الاتفاق
س١ هل كل مفردة إختبارية تعد قياسا مناسبا للهدف الذى تقيسه ؟	٪١٠٠
س٢ هل مفردات الإختبار ككل تعد بمثابة عينة ممثلة للأهداف فى إطار محتوى الوحدة الخاصة بالدراسة ؟	٪ ٩٤,٦
س٣ هل مفردات الإختبار ككل متسقة مع الهدف الأساسى منه؟	٪ ١٠٠

ويلاحظ من الجدول (٧) أن نسب اتفاق المحكمين كانت مرتفعة ، وقد أفاد الباحث من ملاحظات ومقترحات المحكمين فى إجراء التديلات اللازمة .

وفى إطار هذه النسب يعد الإختبار صادقا وأن الإختبار التحصيلى مرجعى المحك صالحا لقياس ما وضع لقياسه ، بصورتيه (ا، ب) .

١- ثم استخدم الباحث الحالى أيضا طريقة الصدق الوصفى باستخدام محك آخر :

وقد بلغ معامل الإرتباط ٨٨ والذى يعتد به كعامل صدق مما يدل على إرتفاع صدق الإختبار بصورتيه (ا ، ب)

(ب) حساب ثبات درجات الإختبار :

استخدم الباحث طريقتين لحساب ثبات الإختبار التحصيلى مرجعى المحك بصورتيه (ا ، ب) طريقتين هما :

١ . معامل الإتفاق لسايكوفاك Subkoviak Coefficient for C.R.T

حيث بلغت قيمة معامل ثبات الإختبار باستخدام معدلة معامل الإتفاق لسابكوفاك ٧١، للصورة (ا) ، ٧٠، للصورة (ب) وهاتين القيمتي مقبولتين وتعبر عن ثبات الإختبار بصورتيه بدرجة معقولة .

٢- معامل ليفنجستون Livingeston Coefficient for C-R-T

وقد بلغت قيمة معامل ثبات الإختبار بصورتيه باستخدام تلك المعادلة ٠,٩٥ ، ٠,٩٦ وتعد هذه القيمة مرتفعة وأن الاختبار بصورتيه يتميز بدرجة عالية من الثبات .

ويمكن تفسير إختلاف معاملات ثبات الإختبار من طريقة إلى أخرى أن معامل الإتفاق لسايكوفاك يتأثر بعدد أفراد العينة وتجانس المحتوى الذى يقيسه الإختبار بصورتيه لذلك تكون قيمة منخفضة حيث بلغت ٧١، ، ٧٠، ، مما يؤكد أن هناك إتساقا داخليا بين مفردات الإختبار فى صورتيه ، ولذلك اختلفت قيم معاملات الثبات من طريقة إلى أخرى ، ويبين ملحق الدراسة صورتى الاختبار والتعليمات وورقة الإجابة .

رابعا: الإجراءات التجريبية للبحث الحالى

١. تم تقسيم العينة إلى ثلاث مجموعات كما هو مبين بالجدول (١)

٢- طبق الباحث صورتى الإختبار التحصيلى على المجموعات الثلاث بطرق مختلفة عن بعضها البعض كما يلى :

(ا) المجموعة الضابطة :

تم تطبيق الاختبار بصورتيه بأن طلب منهم أن يضعوا علاقة (√) على أحد البدائل الأربع التى يعتقدون أنها صحيحة ، فإذا كانت صحيحة تكون درجته (١) وإذا كانت خاطئة تكون الدرجة (صفر) .

(ب) المجموعة التجريبية الأولى :

تم تطبيق الإختبار التحصيلى بصورتيه بطريقة توضح تعليماتها للممتحنين بأن يضعوا رتبة تدل على صحة الإختبار ، فإذا اعتقد أن إجابة ما هى الصحيحة يضع عليها رقم (١) فإذا كانت صحيحة تكون درجته ٣ وإذا أعتقد أن إجابة أخرى هى الصحيحة يضع رقم (٢) عليها فإذا كانت إحداهما صحيحة تكون الدرجة ٢ ، وإذا اعتقد أن هناك إجابة ثالثة ربما تكون صحيحة يضع عليها رقم (٣) فإذا كانت إحداهما صحيحة تكون درجته ١ ولايسمح له بوضع رتب أمام الإختبارات الأربع وإذا حدث كانت درجته صفر .

(جـ) المجموعة التجريبية الثانية :

طبق الإختبار التحصيلى بصورتيه بطريقة تختلف عن الطريقتين السابقتين حيث القيت تعليمات للممتحنين وطلب منهم أن يضعوا نسبة مئوية على كل اختبار من الإختبارات الأربعة تدل على مدى ثقته من صحة الإجابة بحيث يكون مجموع النسب

١٠٠٪ أى طريقة الإجابة هنا طريقة وزن الثقة ، ثم بعد ذلك تم معالجة إجابات هذه المجموعة بثلاث طرق منها :

الطريقة الأولي : طريقة الإختيار الإستنتاجي

وفيها تكون درجة الممتحن (١) عند أعلى نسبة مئوية إذا كانت أمام الاختيار الصحيح أما إذا كانت هناك نسبة أكبر منها والاختيار خطأ فإن الممتحن يحصل على (صفر) .

الطريقة الثانية : طريقة الإحتمال المقترح للإجابة الصحيحة :

وفيها تكون درجة الممتحنين عبارة عن النسبة المئوية التى وضعها أمام الإجابة الصحيحة لكل مفردة وتجمع هذه النسب لكل المفردات وبذلك تكون الدرجة الكلية للممتحنين فى الإختبار .

الطريقة الثالثة : الطريقة التربيعية :

وفيها تحسب درجة المفردة من المعادلة :

$$\text{الدرجة اللوغاريتمية} = ١ + \frac{١}{٢} \text{ لو } (\text{ح ق})$$

لكل مفردة ثم تجمع القيم لجميع المفردات لكل ممتحن حتى يحصل على الدرجة الكلية

وسوف يستخدم الباحث الحالى هذه الدرجات فى التحليل الإحصائى للبيانات بعد استبعاد درجات أى فرد تخلف عن إحدى صورتى الإختبار .

٣- الأساليب الإحصائية المستخدمة :

ا - بعض أساليب الإحصاء الوصفى للتأكد من اعتدالية المنحنى .

ب - حساب معاملات الإرتباط

جـ- طريقة Fldt لحساب الفروق بين معاملات الصدق .

نتائج البحث وتفسيرها :

يعرض الباحث الحالى النتائج التى اسفر عنها البحث كما يقوم بتفسير تلك النتائج ومناقشتها فى ضوء فروض البحث والإطار النظرى والدراسات السابقة .

١- حساب معاملات الصدق للإختبار المرجعى المحك بصورتيه ويسمى الصدق الوظيفى :

كما يتضح ذلك فى الجدول التالى :

جدول (٨)

يبين معاملات الصدق فى العينة النهائية وذلك لكل طريقة من الطرق الوزنية لتقدير الدرجات على حده

المجموعات	طرق تقدير الدرجات للإختبار	معاملات الصدق
المجموعة الضابطة	الطريقة التقليدية	٠,٨٣٨
المجموعة التجريبية الأولى	طريقة الدرجات التجريبية	٠,٩١٧
المجموعة التجريبية الثانية (١)	طريقة الإحتمال المقترح	٠,٨٩٥
المجموعة التجريبية الثانية (٢)	الطريقة التربيعية	٠,٣٣١
المجموعة التجريبية الثانية (٣)	طريقة الإختبار الإستنتاجى	٠,٩٦٩

يتضح من الجدول (٨) السابق أن هناك تباين بين قيم معاملات صدق الإختبار مرجعى المحك بناء على اختلاف الطرق الوزنية لتقدير الدرجات .

٢- الطريقة الإحصائية المستخدمة لإختبار صحة الفروض :

لإختبار صحة الفرض استخدم الباحث الحالى طريقة Fldt للفروق بين معاملات الصدق باستخدام المعادلة الآتيـــة :

$$ف = \frac{١ - ر_{١}}{١ - ر_{٢}} \qquad (١ : ٤٢٣ - ٤٢٤)$$

حيث ر معامل الصدق الأول , ر٢ معامل الصدق الثانى

وقد ثبت أن $ف = \frac{١ - ر_{١}}{١ - ر_{٢}}$ تأخذ نفس توزيع (ف) ولذلك يمكن استخدام جداول النسب الفائية للكشف عن دلالة الفروق بين معاملات الصدق .

والجدول (٩) يبين مجموعات المقارنة وقيم معاملات صدقها والنسبة الفائية ومستوى دلالتها

جدول (٩)

يوضح المقارنة بين المجموعات وقيم معاملات صدقها والنسبة الفائية ومستوى دلالتها

مجموعات المقارنة	قيم معاملات الصدق	النسبة الفائية ومستوى الدلالة (الترتيل)
الطريقة التقليدية مع طريقة الدرجات التجريبية	٠,٨٣٨ ١,٩٥	١,٩٥ **
الطريقة التقليدية طريقة الإحتمال المقترح	٠,٨٣٨ ٠,٨٩٥	١,٥٤ *
الطريقة التقليدية الطريقة التربيعية	٠,٨٣٨ ٠,٣٣١	٤,١٢ **
الطريقة التقليدية طريقة الإختيار الإستنتاجى	٠,٨٣٨ ٠,٩٦٩	٥,٢٢ **
طريقة الدرجات التجريبية طريقة الإحمال المقترح	٠,٩١٧ ٠,٨٩٥	١,٢٦
طريقة الدرجات التجريبية الطريقة التربيعية	٠,٩١٧ ٠,٣٣١	٨,٠٦ **
طريقة الدرجات التجريبية طريقة الإختيار الإستنتاجى	٠,٩١٧ ٠,٩٦٩	٢,٦٧ *
طريقة الإحتمال المقترح الطريقة التربيعية	٠,٨٩٥ ٠,٣٣١	٦,٣٧ *
طريقة الإحتمال المقترح الطريقة الإختيار الإستنتاجي	٠,٨٩٥ ٠,٩٦٩	٣,٣٨ **
الطريقة التربيعية طريقة الإختيار الإستنتاجى	٠,٣٣١ ٠,٩٦٩	٢١,٥٨ **

* مستوى دلالة عند ٠,٠٥

** مستوى دلالة عند ٠,٠١

يتضح من الجدول (٩) السابق وجود فروق دالة إحصائيا عند مستوى دلالة ٠,٠٥ ، ٠,٠١ بين قيم معاملات صدق الإختبار مرجعى المحك باختلاف الطرق الوزنية لتقدير الدرجات وسوف يتضح ذلك عند التأكد من تحقق صحة فروض البحث الحالى كل على حده .

وللتحقق من صحة **الفرض الأول** والذى ينص على أنه« لا توجد فروق ذات دلاله إحصائية بين مجموعتى الطريقة التقليدية وطريقة الدرجات التجريبية فى صدق درجات الإختبار مرجعى المحك »

وبالنظر إلى جدول رقم (٩) فقد اتضح وجود فروق ذات دلالة إحصائية عند مستوى ٠,٠١ وأيضا ٠,٠٥ بين مجموعتى الطريقة التقليدية وطريقة الدرجات التجريبية ، وكانت تلك الفروق لصالح مجموعة طريقة الدرجات التجريبية لكونها ذات قيمة أكبر فى معامل الصدق حيث بلغت ٩١٧, مقارنة بالقيمة الآخرى والتى بلغت ٨٣٨,٠ . وبذلك لم يتحقق هذا الفرض لوجود تلك الفروق الدالة إحصائيا .

وتتفق هذه النتيجة مع دراسة كل من برانبوم وتاتسوك ١٩٨٣ Birenbbaum &Tatsuoka وروس واستيفين ١٩٩٤ Ross & Stevn بينما تختلف مع دراسة كل من كوهلر ١٩٧١ Koehler وريوز – بريمو Ruiz - Primo وربما يرجع السبب فى هذا التباين إلى اختلاف حجم العينة أو المحتوى الدراسى لذلك تفوقت الدرجات التجريبية على الطريقة التقليدية فى تأثيرها على صدق الإختبار مرجعى المحك ، ويمكن تفسير ذلك لزيادة عدد المحاولات التى تتاح للممتحن أثناء الإجابة على مفردات الإختبار لاعطاء قيمة لأى اختيار يتوقع أنه الإجابة الصحيحة أو قريب منها ، وهذا بخلاف الطريقة التقليدية التى فيها يكتفى الممتحن بمحاولة واحدة فقط لكل مفردة من مفردات الإختبار ، وزيادة عدد المحاولات يؤدى إلى زيادة طول الإختبار ومداه وبالتالى يؤدى إلى زيادة صدقه وثباته .

أما الفرض الثانى والذى ينص على أنه « لاتوجد فروق ذات دلالة إحصائية بين مجموعتى الطريقة التقليدية وطريقة الإحتمال المقترح فى صدق درجات الإختبار مرجعى المحك»

بالنظر إلى جدول رقم (٩) فقد وجد أن هناك فروقاً ذات دلالة إحصائية عند

مستوى ٠٥. وبين مجموعتى الطريقة التقليدية وطريقة الإحتمال المقترح ، وكانت تلك الفروق لصالح مجموعة الإحتمال المقترح لكونها ذات قيمة أكبر فى معامل الصدق حيث بلغت ٨٩٥ , مقارنة بالقيمة الأخرى والتى بلغت ٨٣٨ ,٠ وبذلك لم يتحقق هذا الفرض وذلك لوجود تلك الفروق الدالة إحصائيا .

وتتفق هذه النتيجة مع دراسة كل من برانبوم وتاتسوك ١٩٨٣ Birenb-baum &Tatsuoka وروس واستيفين ١٩٩٤ Ross & Stevn بينما تختلف مع دراسة كل من كوهلر ١٩٧١ Koehler وريوز - بريمو Ruiz - Primo وربما يرجع السبب فى هذا التباين إلى اختلاف حجم العينة أو المختوى الدراسى، ويمكن تفسيرتفوق طريقة الإحتمال المقترح على الطريقة التقليدية فى معامل صدق الإختبار مرجعى المحك ، لزيادة عدد المحاولات التى تتاح للممتحن باعطاء نسبة مئوية للإجابة التى يفترض أنها صحيحة حتى لو تم إعطاء نسبة مئوية لجميع الإختبارات بشرط أن يكون مجموع النسبة ١٠٠٪ ومن هنا يمكن أن يقيس الإختبار لما وضع لقياسه وبالتالى يزداد صدقه بخلاف الطريقة التقليدية .

أما الفرض الثالث والذى ينص على أنه «لاتوجد فروق ذات دلالة إحصائية بين مجموعتى الطريقة التقليدية والطريقة التربيعية فى صدق درجات الإختبار مرجعى المحك »

وبالنظر إلى جدول رقم (٩) يتضح وجود فروق ذات دلالة إحصائية عند مستوى ٠,٠٥ ، ٠,٠١ بين مجموعتى الطريقة التقليدية والطريقة التربيعية ، وكانت تلك الفروق لصالح مجموعة الطريقة التقليدية لكونها ذات قيمة أكبر فى معامل الصدق حيث بلغت ٨٣٨ , مقارنة بالقيمة الآخرى والتى بلغت ٣٣١, . وبذلك لم يتحقق هذا الفرض لوجود تلك الفروق الدالة إحصائيا.

وتتفق هذه النتيجة مع دراسة كوهلر ١٩٧١ Korhler بينما اختلفت مع دراسة كل من برابنوم وتاتسوك ١٩٨٣ وروس وستيفين ١٩٩٤ وجرادات ١٩٨٨ .

ويمكن تفسير تفوق الطريقة التقليدية على الطريقة التربيعية فى صدق الإختيارات مرجعية المحك لأن درجة الممتحن فى الطريقة التربيعية لم تعبر عن استجابة الممتحن بطريقة مباشرة لأنها جاءت بعد التطبيق بناء على معادلة وبالتالى فهى درجة

مشتقة لذلك اعطت معامل صدق منخفض

أما الفرض الرابع والذى ينص على أنه « لاتوجد فروق ذات دلالة إحصائية بين مجموعتى الطريقة التقليدية وطريقة الإختيارالإستنتاجى فى صدق درجات الإختبار مرجعى المحك »

بالنظر إلى جدول رقم (٩) يتضح وجود فروق ذات دلالة إحصائية عند مستوى ٠,٠٥ ، ٠,٠١ بين مجموعتى الطريقة التقليدية وطريقة الإختيار الإستنتاجى وكانت تلك الفروق لصالح مجموعة طريقة الإختيار الإستنتاجى لكونها ذات قيمة أكبر فى معامل الصدق حيث بلغت ٩٦٩, مقارنة بالقيمة الأخرى والتى بلغت ٨٣٨, وبذلك لم يتحقق هذا الفرض ، لوجود تلك الفروق الدالة إحصائيا

. وتتفق هذه النتيجة مع دراسة كل من برانبوم وتاتسوك ١٩٨٣ Birenb-baum &Tatsuoka وروس واستيفين ١٩٩٤ Ross & Stevn بينما تختلف مع دراسة كل من كوهلر ١٩٧١ Koehler وريوز - بريمو Ruiz - Primo وربما يرجع السبب فى هذا التباين إلى اختلاف حجم العينة أو المحتوى الدراسى ، وتفسير هذه النتيجة على أن طريقة الإختيار الإستنتاجى تعطى الممتحن الفرص فى اعطاء أكثر من احتمال للإجابة الصحيحة بوضع نسبة مئوية أمام الإختيار وذلك يؤدى إلى زيادة الثقة لدى الممتحن مما زاد من معامل الصدق عند استخدام هذه الطريقة .

أما الفرض الخامس والذى ينص على أنه « لاتوجد فروق ذات دلالة إحصائية بين مجموعتى طريقة الدرجات التجريبية وطريقة الإحتمال المقترح فى صدق درجات الإختبار مرجعى المحك » .

بالنظر إلى جدول (٩) يتضح عدم وجود فروق ذات دلالة إحصائية بين مجموعتى طريقة الدرجات التجريبية وطريقة الإحتمال المقترح وبذلك يتحقق الفرض لعدم وجود فروق ذات دلالة إحصائية .

وتتفق هذه النتيجة مع دراسة كل من برانبوم وتاتسوك ١٩٨٣ Birenb-baum &Tatsuoka وروس واستيفين ١٩٩٤ Ross & Stevn بينما تختلف مع دراسة كل من جيمس ١٩٧٥ Gams وريوز - بريمو Ruiz - Primo وربما يرجع السبب فى هذا التباين إلى اختلاف حجم العينة أو المحتوى الدراسى، وتفسير هذه

النتيجة التى توضح تفوق الطريقة التجريبية على طريقة الإحتمال المقترح لأن الطريقة التجريبية تتيح للممتحن اختيار ثلاث إجابات عندما يعجز عن اختيار إجابة واحدة أو اثنين ومهما تعددت تلك الإختيارات فانه يحصل على درجة ، مما يزيد تباين درجات الإختبار مما يؤدى إلى زيادة معامل الصدق عند استخدام الطريقتين لتقاربهما فى إتاحة الفرصة للممتحنين .

أما الفرض السادس والذى ينص على أنه « لاتوجد فروق ذات دلالة إحصائية بين مجموعتى طريقة الدرجات التجريبية والطريقة التربيعية فى صدق درجات الإختبار مرجعى المحك »

بالنظر إلى جدول (٩) يتضح وجود فروق ذات دلالة إحصائية عند مستوى ٠,٠٥ ، ٠,٠١ بين مجموعتى طريقة الدرجات التجريبية والطريقة التربيعية ، وكانت تلك الفروق لصالح مجموعة طريقة الدرجات التجريبية لكونها ذات قيمة أكبر فى معامل الصدق حيث بلغت ٠,٩١٧ مقارنة بالقيمة الآخرى والتى بلغت ٠,٣٣١ وبذلك لم يتحقق هذا الفرض لوجود تلك الفروق الدالة إحصائيا .

وتتفق هذه النتيجة مع دراسة جرادات Gradat ١٩٨٨ بينما تختلف مع دراسة كل من جيمس Gams ١٩٧٥ وريوز – بريمو Ruiz - Primo وربما يرجع السبب فى هذا التباين إلى اختلاف حجم العينة أو المحتوى الدراسى ،وتفسير هذه النتيجة على أن تفوق الطريقة التجريبية كما ذكر سابقا أن الممتحن يحصل على درجة مهما تعددت الإختيارات بشرط عدم وصولها أربع اختيارات ، بينما انخفض معامل الصدق فى الطريقة التربيعية كما ذكر سابقا ربما يرجع إلى طريقة وتطبيق المعادلة .

أما الفرض السابع والذى ينص على أنه «لاتوجد فروق ذات دلالة إحصائية بين مجموعتى طريقة الدرجات التجريبية وطريقة الإختيار الإستنتاجى فى صدق درجات الإختبار مرجعى المحك »

بالنظر إلى جدول (٩) يتضح وجود فروق ذات دلالة إحصائية عند مستوى ٠,٠٥ ، ٠,٠١ بين مجموعتى طريقة الدرجات التجريبية وطريقة الإختيار الإستنتاجى ، وكانت تلك الفروق لصالح مجموعة الإختبار رالإستنتاجى لكونها ذات قيمة أكبر فى معامل الصدق حيث بلغت ٠,٩٦٩ مقارنة بالقيمة الآخرى والتى بلغت ٠,٩١٧ ، وبذلك لم يتحقق

هذا الفرض لوجود تلك الفروق الدالة إحصائيا .

بينما تختلف مع دراسة كل من جيمس ١٩٧٥ Gams وريوز – بريمو - Ruiz Primo وربما يرجع السبب فى هذا التباين إلى اختلاف حجم العينة أو المحتوى الدراسى، ويمكن تفسير تفوق طريقة الإختيار الإستنتاجى على الطريقة التجريبية فى معامل صدق الإختبار مرجعى المحك إلى أن طريقة الإختيار الإستنتاجى فيها يحصل الممتحن علي درجة واحدة على أعلى نسبة للإختيار الصحيح مما يزيد ثقة مفردات الإختيار ويزيد من معامل صدقه .

أما الفرض الثامن والذى ينص على أنه «لاتوجد فروق ذات دلالة إحصائية بين مجموعتى طريقة الإحتمال المقترح والطريقة التربيعية فى صدق درجات الإختبار مرجعى المحك»

بالنظر إلى جدول (٩) يتضح وجود فروق ذات دلالة إحصائية عند ٠,٠٥ ، ٠,٠١ بين مجموعتى طريقة الإحتمال المقترح والطريقة التربيعية وكانت تلك الفروق لصالح مجموعة طريقة الإحتمال المقترح لكونها ذات قيمة أكبر فى معامل الصدق حيث بلغت ٨٩٥, مقارنة بالقيمة الأخرى والتى بلغت ٣٣١, ، وبذلك لم يتحقق هذا الفرض لوجود تلك الفروق الدالة إحصائيا .

وتتفق هذه النتيجة مع دراسة كل من برانبوم وتاتسوك (١٩٨٣) Birenb-boum & Tatsum وجرادات (١٩٨٨) Garadat ، وروس (١٩٩٤) Ross ، بينما تختلف مع دراسة كل من جيمس (١٩٧٥) James ، ريوز (١٩٩٨) Ruiz ، وتفسر هذه النتيجة على أن طريقة الاحتمال المقترح تساعد الممتحن على وضع نسبة مئوية لأكثر من اختيار حتى ولو كانت جميع الاختيارات ، وأنه يحصل على مقدار النسبة المئوية حيث تزيد هذه الطريقة من تباين درجات الاختبار مما يؤدى إلى زيادة معامل الصدق ، حيث إن التباين القوى يزيد من قيمة معامل الصدق ، وذلك على عكس الطريقة التربيعية التى تستخرج نتيجة معادلة ، ويمكن من خلال تطبيق المعادلة التأثير على قيمة معامل الصدق .

أما الفرض التاسع والذى ينص على أنه «لاتوجد فروق ذات دلالة إحصائية بين مجموعتى طريقة الإحتمال المقترح وطريقة الإختيار الإستنتاجى فى صدق درجات

الإختبار مرجعى المحك »

بالنظر إلى جدول (٩) يتضح وجود فروق ذات دلالة إحصائية عند ٠,٠٥ ، ٠,٠١ بين مجموعتى طريقة الإحتمال المقترح وطريقة الإختيار الإستنتاجى وكانت تلك الفروق لصالح مجموعة طريقة الإختيار الإستنتاجى لكونها ذات قيمة أكبر فى معامل الصدق حيث بلغت ٩٦٩, مقارنة بالقيمة الآخرى والتى بلغت ٨٩٥ . وبذلك لم يتحقق هذا الفرض لوجود تلك الفروق الدالة إحصائيا .

وتتفق هذه النتيجة مع دراسة كل من برانبوم وتاتسوك (١٩٨٣) Birenb-boum & Tatsum وجرادات (١٩٨٨) Jaradat ، وروس (١٩٩٤) Ross ، بينما تختلف مع دراسة كل من جيمس (١٩٧٥) James ، ريوز (١٩٩٨) Ruiz ، وتفسر هذه النتيجة على أن طريقة الاختيار الاستنتاجى تعطى فرصة أكبرللممتحن فى عملية الاختيار وتعدد الاحتمالات وحصوله على درجة المفردة إذا كانت النسبة الموضوعة للاختيار أكبر نسبة للاختيار الصحيح مما يزيد من الثقة فى الاختبار وبالتالى زيادة صدقه .

أما الفرض العاشر والذى ينص على أنه «لاتوجد فروق ذات دلالة إحصائية بين مجموعتى الطريقة التربيعية وطريقة الإختيار الإستنتاجى فى صدق درجات الإختبار مرجعى المحك »

يتضح من الجدول (٩) وجود فروق ذات دلالة إحصائية عند ٠,٠٥ ، ٠,٠١ بين مجموعتى الطريقة التربيعية وطريقة الإختيار الإستنتاجى ، وكانت تلك الفروق لصالح مجموعة طريقة الإختيار الإستنتاجى لكونها ذات قيمة أكبر فى معامل الصدق حيث بلغت ٩٦٩, مقارنة بالقيمة الآخرى والتى بلغت ٣٣١, وبذلك لم يتحقق هذا الفرض لوجود تلك الفروق الدالة إحصائيا .

وتتفق هذه النتيجة مع دراسة كل من برانبوم وتاتسوك (١٩٨٣) Birenb-boum & Tatsum وجرادات (١٩٨٨) Jaradat ، وروس (١٩٩٤) Ross ، بينما تختلف مع دراسة كل من جيمس (١٩٧٥) James ، ريوز (١٩٩٨) Ruiz ، ويمكن تفسير تفوق طريقة الاختيار الاستنتاجى على الطريقة التربيعية فى معامل الصدق وذلك لتميز طريقة الاختيار الاستنتاجى – كما اتضح سابقا – والتى تعمل على زيادة الثقة فى الاختبار وأيضا الزيادة فى معامل الصدق .

التوصيات التربوية والبحوث المقترحة :

من خلال نتائج البحث الحالى يوصى الباحث باستخدام بعض الطرق الوزنية فى تقدير درجات الاختبارات مرجعية المحك ذات الاختيار من متعدد عند تصحيحها ، مما يزيد من الثقة فى الاختبار ومفرداته وأيضاً فى معامل صدقه مقارنة بالطريقة التقليدية . وأيضا تساعد هذه الطرق الوزنية فى تقدير الدرجات على التقليل من التأثر بعامل التخمين عند الإجابة فى مثل هذا النوع من الاختبارات مما يرفع من كفاءتها وزيادة صدقها وقياسها فيما وضعت لقياسه .

لذا يقترح الباحث الحالى إجراء بعض الدراسات فى هذا المجال ، وعلى سبيل المثال :

١- دراسة مقارنة لأثر بعض الطرق الوزنية لتقدير الدرجات لكل من الاختبارات مرجعية المعيار والاختبارات مرجعية المحك ذات الاختيار من متعدد على معامل صدق تلك الاختبارات .

٢- يمكن إجراء البحث الحالى على عينة من طلاب المدارس الثانوية ومحتوى دراسى آخر .

٣- يمكن إجراء البحث الحالى مع استخدام طرق وزنية لتقدير الدرجات تختلف عن الطرق المستخدمة فى البحث الحالى ، ودراسة تأثيرها على صدق الاختبارات مرجعية المحك ذات الاختيار من متعدد .

المراجـــــع

مجلة كلية التربية ببنها – عدد أكتوبر ٢٠٠١

أولاً: المراجع العربية

١- *أحمد الرفاعى غنيم (١٩٨٣)* : أثر التغذية الرجعية الفورية على صدق وثبات الاختبارات العقلية ذات الاختيار من متعدد ، رسالة دكتوراه غير منشورة ، كلية التربية ، جامعة الزقازيق .

٢- *أنور محمد الشرقاوى ، سليمان الخضرى الشيخ ، أمينة محمد كاظم ، نادية محمد عبدالسلام (١٩٩٦)* : "القياس والتقويم النفسى والتربوى" ، مكتبة الأنجلو المصرية ، القاهرة .

٣- *صفوت فرج (١٩٨٠)* : القياس النفسى ، دار الفكر العربى ، القاهرة .

٤- *صلاح الدين محمد أبو ناهية (١٩٩٤)* : "القياس التربوى" ، ط١ ، الأنجلو المصرية ، القاهـــرة .

٥- *صلاح الدين محمود علام (١٩٨٧)* : "دراسة موازنة ناقدة لنماذج السمات الكامنة والنماذج الكلاسيكية فى القياس النفسى والتربوى ، المجلة العربية للعولم الإنسانية ، العدد السابع والعشرين و المجلد السابع ، ص ص١٨ - ٤٣ ، جامعة الكويت ، الكويت .

٦- *صلاح الدين محمود علام (١٩٩٥)* : "الإختبارات التشخيصية مرجعية المحك فى المجالات التربوية والنفسية والتدريبية ، دار الفكر العربى ، الطبعة الأولى ، القاهرة .

٧- *فتحى الديب (١٩٩٣)* : التقويم وبناء الإختبارات فى التعليم الجامعى ، مجلة العلوم التربوية ، المجلة الأولى ، العدد الأول ، معهد الدراسات التربوية ، جامعة القاهرة .

٨- *فؤاد البهى السيد (١٩٥٨)* : الجداول الإحصائية لعلم النفس والعلوم الإنسانية الأخرى ، دار الفكر العربى القاهرة .

٩- *فؤاد البهى السيد (١٩٧٩)* : علم النفس الإحصائى وقياس العقل البشرى ، ط٣ ، دار الفكر العربى ، القاهرة .

١٠- *نادية بعيبع (١٩٩٧)* : أهمية الإختبارات المحكية فى التقويم التربوى ، بناء إختبار محكى المرجع لقياس التحصيل فى مادة الرياضيات للصف السادس من التعليم الأساسى ، مجلة علم النفس ، الهيئة المصرية العامة للكتاب ، العدد (٤٤) أكتوبر – نوفمبر – ديسمبر ، السنة الحادية عشر ، ص ص ١٧٢ : ١٧٥ .

ثانيا: المراجع الاجنبيـــة

11- *Birenbaum, M. and Tatsuaka, K. K. (1983) :* The affect of Ascoring system basedon algorthium underlying the students' Response Patterns of the demonationalty of achievement test data of the problem scoring type, Journal of Educational Measurement, 20, 1.

12- *Dalaba, B. A. (1979) :* The Validation of Objectives on Deveopment of Critterion - Referenced Test for Eighteen Units in Consumer Education, Dissertation Abstract International, A. Vol. 39 (12), pp. 7122 - 7123.

13- *Donald, G. Gardner (1998) :* Single - Item Versus Multiple-Item Measurement Scales: An Empirical Comparison, Educational and Psychological Measurement, Vol. 58, No. 6, pp. 898 - 915.

14- *Hanna, G. S. and Long, C. A. (1974) :* Effect of Answer Until Correct Testing on Reliability, Perceptual and Motor Skills, 1974.

15- *Hendrickson, G. F. (1971) :* An Interpretation of Livingston's Reliability Coeddicient for Criterion - Referenced Tests, Journal of Educational Measurement, Vol. 9, No. 1, pp. 27 - 29.

16- *James Popham (1988) :* Educational Evaluation. Second Edition. Library of Congress Cataloging in Publication Data. New York.

17- *Jaradat - Derar, Tollefson, Nona (1988)* : The Impact of Alternative Scoring Procedures for Multiple - Choice Items on Test Reliability, Validity, and Grading. Educational and Psychological Measurement; Vol. 48, No. 3, pp. 627 - 635.

18- *Kansup, W. & Hokstian, A. R. (1975) :* A Comparison of Several Methods of Assessing Partial Knowledge in Multiple Choice Tests : Scoring Procedures, Journal of Educational Measurement, Vol. 12, No. 4, pp. 219 - 229.

19- *Koehler4, R. A. (1971) :* A Comparison of the Validities of Conventional Choice Testing and Various Confidence Marking Procedures, Journal of Educational Measurement, Vol. 8, No. 4.

20- *Nitko, A. (1980) :* Distinguishing the Many Varieties of Criterion- Referenced Tests. Review of Educational Research, 73.

21- *Norman Gronlund (1985) :* Measurement & Evaluation in Teaching - 5 ed., Mcmillan Publishing Company, N.Y.

22- ***Peter, W. Airasian (1994) :*** Classroom Assessment, McGraw-Hill, Inc., U.S.A.

23- ***Rippy, R. M. (1970) :*** A Comparison of Five Different Scoring Function for Confidence Tests. Journal of Educational Measurement, Vol. 7, No. 3, 1970.

24- ***Ronald A. Berk (1980) :*** A Consumers Guide to Criterion Referenced Test Reliability. Journal of Educational Measurement. Vol. 17, No. 4, pp. 323 - 347.

25- ***Ronald K. Hambleton (1990) :*** Advance in Educational and Psychological Testing. Theory and Applications. Kluwer Academic Publishers. Whashington.

26- ***Ronald K. Hambleton; Swaminathau, H. ; Algina, J. & Coulson. D. B. (1978) :*** Criterion - Referenced Testing and Measurement. A Review of Technical Issues and Developments. Review of Educational Research.

27- ***Ross, Steven ; Te (1994) :*** An Approach to Gain Score Dependability and Validity for Criterion - Referenced Language Tests. Paper presented at the Annual Language Testing Research Colloquium 16th. In computer search. ED.

28- ***Ruiz - Primo, Maria - Aracoli and Others (1998) :*** Comparison of the Reliability and Validity of Scores from Two Concept- Mapping Techniques National Center for Research on Evaluation, Standards and Student Testing, Los Angeles, CA. In computer search. ED.

29- *Stanly, Julian C and D. Hopkins (972) :* Educational Psychological Measurement and Evaluation, 5th ed., Englewood Cliffs, N.J., Prentice-Hall.

30- *Thomas M. Haladyna (1994) :* Developing and Validating Multiple - Choice Test Items. Applied Psychological Measurement. Vol. 19, No. 2, pp. 205 - 207.

31- *Tom Haladyna (1981) :* The Role of Instructional Sensetivity in the Empirical Review of Criterion Referenced Test Items. Journal of Educational Measurement, Vol. 18. No. 1.

32- *Wiersma, W. & Jurs. S. (1990) :* Educational Measurement and Testing. (2nd. ed.) Boston : Allyn and Bacon.