

الفصل الثالث

" الدراسات السابقة "

~~~~~

#### المحور الاول :

- دراسات تناولت بناء الاختبار المرجعي المحك
- تعقيب على دراسات المحور الاول •

#### المحور الثاني :

- دراسات تناولت " الطرق المختلفة لتقدير الدرجة وأثرها على ثبات الاختبارات مرجعية المعيار "
- تعقيب على دراسات المحور الثاني •
- فروع البحث •

## مقدمة :

تهدف الدراسة الحالية الى دراسة اثربعض طرق تقدير الدرجات على ثبات الاختبارات مرجعية المحك ذات الاختيار من متعدد ، وثمة مجالات أخرى وثيقة الصلة بذلك لعل أهمها الدراسات المتعلقة ببناء الاختبار المرجعي المحك والدراسات التي تناولت الطرق المختلفة لتقدير الدرجات وأثرها على ثبات الاختبارات مرجعة المعيار .

حيث أن هذه الدراسات التي توافرت لدى الباحث ولم توجد أية دراسات تتعلق بأثر طرق تقدير الدرجات على ثبات الاختبارات مرجعية المحك في حدود علم الباحث ، لذلك سوف يتناول هذا الفصل تلك البحوث والدراسات بهدف الوقوف على طريقة المعالجة البحثية التي استخدمت في هذه الدراسات ، حيث يركز الباحث في عرضه لكل دراسة على النقاط الهامة المتعلقة بمجال البحث الحالي فقط ، وفيما يلي تفصيل ذلك :

أولا : دراسات تناولت بناء الاختبار المرجعي المحك ، ثم تعقيب

ثانيا : الدراسات التي تناولت الطرق المختلفة لتقدير الدرجات وأثرها على ثبات الاختبارات مرجعة المعيار ، ثم تعقيب .

## أولاً : المحور الأول

دراسات تناولت بناء الاختبار

المرجعي المحك •

( ١ ) دراسة جانوبول Ganopole ( ١٩٧٨ )

وتهدف الى بناء اختبار مرجعى المحك فى المهارات الاساسية للقراءة ،  
وقد مرت هذه الدراسة بثلاث مراحل هى :-

— تحديد المهارات الاساسية للقراءة ، وتم ذلك عن طريق جمع بيانات من  
المؤسسات العلمية والبحثية المختصة عن هذه المهارات .

— قد تم بناء مفردات الاختبار فى خطوتين هما :

( أ ) اعداد مواصفات الاختبار ، وهذه تعنى الوصول الى مجموعة الاهداف التى

تبنى على أساسها مفردات الاختبار المرجعى المحك .

( ب ) صياغة المفردات طبقا لهذه المواصفات .

— تقدير صدق وثبات الاختبار :

— تم ايجاد الصدق الوصفى للاختبار بعدة طرق .

— تم تقدير الثبات بالنسبة للاختبار ولكن اختبار فرعى بطريقتين خاصتين :

( أ ) طريقة الارتباط التقليدى المستخدمة فى المقاييس المرجعية المستوى

وتشمل ما يلى : اعادة الاختبار ، الصور المتكافئة .

( ب ) اتساق إقرار — وهذه أفضل الطرق للاستخدام فى المقاييس المرجعية

المحك .

( ٢٩ : ٢١٦ )

( ٢ ) دراسة سميث Smith ( ١٩٨٠ )

وتهدف الى بناء اختبار مرجعى المحك فى مادة القراءة وتميز المفردات

للطلاب الجدد بالجامعة ، ومرت تلك الدراسة بالخطوات الآتية :-

— تحديد الاهداف التحصيلية .

— تحليل هذه الاهداف الى مجموعات مثل (الهجاء — النطق — الفهم )

— تحديد ووصف نطاق كل اختبار فرعى .

— كتابة قوائم المفردات لكل اختبار فرعى .

— عمل دراسة استطلاعية لتحديد صدق المحتوى ومن ثم تنقيح الاداة بناء على البيانات الناتجة .

— اعتبار ٨٠ ٪ هو مستوى الاتفاق في الاختبارات الفرعية ، ١٠ ٪ مستوى ثقة ، مما جعل أدنى مستوى محك مقبول هو ٧٠ ٪ ، فالذين يصل أداءهم ٧٠ ٪ يعتبرون متقنين .

وتكونت عينة الدراسة من ٤٠ طالبا بالسنة الاولى بجامعة ولاية نيو مكسيكو New Mexico تم تقسيمها عشوائيا الى مجموعتين تضم كل منها ٢٠ طالبا .

وقد تم فحص كل جزء من اجزاء الاختبار بمعزل عن الاجزاء الاخرى ثم اجراء مقارنات بين أداء المجموعتين على كل اختبار فرعى ، وباستخدام تحليل التباين وأسلوب التحليل الوصفي ، تم تحليل مفردات كل اختبار فرعى من حيث مستوى الدقة — مدى تمثيلها للمحتوى — الكفاءة والفاعلية .

ومن حيث قوة تمييز المفردات ، فالمفردات التي ميزت بين المجموعتين " متقنين وغير متقنين " .

— قبلت على أنها جيدة وتمثل المحتوى وتتسم بالفاعلية .  
( ١٠٠٩ — ١٠٠٨ : ٤٩ )

### ( ٣ ) دراسة صلاح علام ( ١٩٨٠ )

تهدف الى بناء وتقدير صدق محتوى اختبار تشخيصي هدفى المرجع وذلك لقياس الحد الادنى لمهارات المعلمين ، اللازمة لبناء الاختبارات المدرسية .

وقد تم بناء صورتين متكافئتين من الاختبار ، كل صورة تتكون من ٩١ مفردة وضعت على أساس ٦٩ هدفا اجرائيا ، وتكون كل اختبار من خمسة اختبارات فرعية يقيس كل منها احدى المهارات التالية :

- مهارة صياغة الاهداف التعليمية التي يقيسها الاختبار اجرائيا .
- مهارة تصميم خطة الاختبار .

- مهارة بناء وتصحيح مفردات الاختبار .
- مهارة تحليل البيانات الناتجة عن تطبيق الاختبار .
- مهارة تقرير نتائج الاختبار بحيث يمكن التقرير مدى تحقيق الاهداف .

وقد تم بناء الاختبار طبقا للمراحل الاتية :-

( أ ) تحديد المجال السلوكي الذي يقيسه الاختبار .

( ب ) بناء مفردات الاختبار وتشمل ما يلي :-

( ١ ) اعداد مواصفات الاختبار

( ٢ ) صياغة مفردات الاختبار

( ج ) تقدير صدق محتوى الاختبار .

( د ) وضع مستويات الاداء .

وقد تحقق صلاح علام من صدق محتوى الاختبار ، وذلك عن طريق تقدير صدق المفردات بالاستعانة بمجموعة من خبراء القياس والتقويم ، وقد قام كل منهم على حدة بمضاهاة كل مفردة من مفردات الاختبار بكل من الهدف الذي تقيسه ، بقية الاهداف الاخرى ، ومجموعة مواصفات الاختبار ، ثم قام الباحث بتحليل البيانات التقديرية للخبراء ، ومن ثم بينت نتائج هذا التحليل أن كل مفردة تعتبر مقياسا صادقا من حيث محتواها ، للهدف الذي تقيسه ، وكذلك امتدت تلك النتائج الى تقدير صدق محتوى الاختبار ككل ، وذلك بالنظر الى جميع مفردات كل اختبار فرعى ، والتي تقيس الاهداف المختلفة في الاختبار .

( ١٢ : ٣ — ٢٣ )

(٤) دراسة اسحاق حنا Ishaq Hanna (١٩٨١)

وتهدف الى تقويم التحصيل الدراسي في الفيزياء في مستوى المدارس الثانوية :  
دراسة عن أداء طلبة شهادة الثانوية العامة GSEF طبقا لمحتك معينة معرفة  
اجرائيا .

يعالج ههنا البحث : القياسات المرجعية المحك بطريقة حديثة  
في مجال التقويم التربوي . وتتلخص المشكلة هي :-

معالجة تقويم التحصيل في مادة الفيزياء لطلاب شهادة G.S.E. ( احدى

شهادات الدراسة الثانوية في المملكة المتحدة ) طبقا لمحكات معينة ولأنماط أداء معرفة  
اجرائيا .

وهناك ثلاث أهداف رئيسية لهذا البحث :

— تقويم تحصيل طلبة شهادة C.S.E الحاصلين على مختلف التقديرات Grades  
في مادة الفيزياء على أساس المحكات Criteria المعينة والمعرفة اجرائيا  
في هذا البحث .

— التمييز بين طلاب C.S.E الحاصلين على مختلف التقديرات باستخدام أنماط  
الاداء والمجالات المعرفية التي تحتويها المحكات .

— الوصول من نتائج هذا البحث الى ما يمكن اعتباره مفيدا وقابلا للتطبيق على  
النظم التربوية والخاصة بالامتحانات في جمهورية مصر العربية ، وفي المملكة  
المتحدة . وهذه الاهداف الثلاثة مترجمة اجرائيا الى " ثلاثة عشر " هدف  
اجرائي . وقد تمت دراسة هذه الاهداف في ثلاثة دراسات يحتويها هذا  
البحث .

وتحتوي هذه الرسالة على معالجة الترميم المرجعي المستوى والتقويم المرجعي  
المحك ، وبصفة خاصة الانجازات الحديثة للتقويم المرجعي المحك ومنها :

مفهوم التمكن (Concept of Mastery) وحالات التمكن Mastery States  
ودرجات القطع Cutt-off scors ، وكذلك ثبات ومدى القياسات المرجعية  
المحك وبعض الانجازات الاخرى .

ولقد تطلب اجراء هذا البحث بأن يقوم الباحث بوضع محكات للاداء في علم  
الفيزياء بأسلوب اجرائي وقد طبقت هذه المحكات بطريقتين هما :

( أ ) بأثر رجعي على امتحان الفيزياء لشهادة C.S.E عام ١٩٧٧ واشتملت العينة  
على ٢٠٠٠ طالب وطالبة .

( ب ) بوضع اختبارين من الاختبارات مرجعية المحك في الفيزياء . الاول طبق  
على طلبة C.S.E عام ١٩٧٨ واشتملت العينة على ٨٠٠ طالب وطالبة . ( ٥ )

والثاني طبق على طلبة CSE عام ١٩٧٩ واشتملت العينة على ٢٣٠ طالب وطالبة . وقام الباحث بوضع هذين الاختبارين ، وكذلك بفحص الثبات والصدق لهم .

وبوجه عام : عند تحليل نتائج هذا البحث روى أن :

تطبيق الوسائل الإحصائية البارامترية Non-parametric يوافق ( جدا ) هذا البحث نظرا لان الدرجات التي يحصل عليها في اختبار مرجعي المحك لا يشترط أن تكون موزعة توزيعا اعتداليا ، ولا أن تكون ذات تباين Variance كبير ، وقد استخدمت هذه الدراسة برامج بالكمبيوتر كوسيلة ناجحة للتمييز بين مجموعات الطلاب الحاصلين على مختلف التقديرات ومن بين النتائج الرئيسية التي حصل عليها في هذا البحث :

- وضع أوصاف إجرائية للاداء في علم الفيزياء لطلاب شهادة C.S.E ووضع تعاريف إجرائية لمجموعات طلاب شهادة C.S.E الحاصلين على مختلف التقديرات في علم الفيزياء وذلك باستخدام المحكلات المقترحة في هذا البحث .
- التمييز بين مجموعات طلاب C.S.E الحاصلين على مختلف التقديرات في علم الفيزياء .
- اقتراح نموذج لوضع التقديرات لمجموعات الطلبة على أساس مطلق .
- الحصول على ترتيب بالنسبة للصعوبة لأنماط أداء معينة ولمجالات معرفية معينة في مادة الفيزياء .
- الحصول على بعض الفروقات في التحصيل الدراسي طبقا لجنس الطالب " ذكر وأنثى " وطبقا لنوع المدارس من حيث احتوائها لاحد الجنسين أو كليهما . ( ٥ )



(٥) دراسة عماد عبدالمسيح : ( ١٩٨٢ )

وتهدف الى بناء مقياس مرجعي المحك في مادة الطبيعة بالصف الاول الثانوى وقد تم بناء المقياس حسب الخطوات الآتية :-

أولا : تحديد محتوى المادة الدراسية المراد قياسها :

قد حدد الباحث وحدة خواص المادة " فى مادة الطبيعة المقررة على طلاب الصف الاول الثانوى ليصمم على أساسها مقررات المقياس بعد أن حدد عناصر الموضوعات لمحتوى هذه الوحدة . ثم عرضت قائمة عناصر موضوعات المحتوى على مجموعة من المحكمين فى صورة استطلاع رأى لمراجعتها وتقرير مدى صحتها .

ثانيا : تحديد الاهداف التعليمية لمجال المحتوى المراد قياسه ، اجرائيا :

فقد شملت هدفين نهائيين ( عامين ) تم تحليلها الى ثمانية عشر هدفا اجرائيا ، ثم عرضت هذه القائمة على مجموعة من المحكمين فى صورة استطلاع رأى لمراجعتها من الناحية الاجرائية ، ومن حيث مدى انتمائية كل مفردة للهدف الاجرائي الخاص بها . وقد تم استبعاد ثلاثة أهداف من القائمة ، وقد أبقى على خمسة عشر هدفا موزعة على المستويات الثلاث لبلوم Bloom وهى ( المعرفة - الفهم - التطبيق )

ثالثا : اعداد مواصفات الاختبار :

اعتمدت هذه الدراسة طريقة بوبهام Popham ( ١٩٧٥ ، ١٩٧٨ ، ١٩٨١ ) فى اعداد مواصفات الاختبار لكل هدف من الاهداف الخمسة عشر ، وتتكون هذه المواصفات من خمسة عناصر أساسية هى : الوصف العام ( عبارة الهدف ) ، عينة المفردة ، خصائص المثير ، خصائص الاستجابة ، ملحق المواصفات . وقد تم عرضها على مجموعة من المحكمين فى صورة استطلاع رأى لتحديد مدى اختصارها ووضوحها فى تحديد نوع السلوك المراد قياسه .

#### رابعاً : بناء مفردات المقياس :

تم صياغة مفردات الاختبار من بين الاختيار من متعدد ، على أساس المستويات المعرفية الثلاثة الاولى لبلوم Bloom ، وهي : " المعرفة — الفهم — التطبيق " وبلغت عدد المفردات ٨٢ مفردة ، ثم أصبحت ٧٠ مفردة بعد عرضها على المحكمين . ثم عرضت قائمة المفردات على مجموعة من المحكمين وعرضت الاهداف الاجرائية المعدلة على لجنة اخرى من المحكمين لتقدير صدق محتوى المقياس ، عن طريق مزاجعة مفردات المقياس مع الاهداف . وتم حساب صدق مفردات المقياس من خلال البيانات التقديرية للمحكمين . وذلك باستخدام معامل تطابق المفردة بالهدف ، والذي اقترحه روفينيلي وها ملتسون Rovinelli and Hambleton ١٩٧٧ ووجد أن قيمة هذا المعامل لكل مفردة كانت واحداً صحيحاً . وذلك دليل على صدق جميع مفردات المقياس ومن ثم صدق محتوى المقياس .

#### خامساً : تحديد مستوى الاداء المقبول ( الدرجة الفاصلة )

بالاستعانة بمجموعة من المحكمين ، حدد الباحث النسبة ٨٠ ٪ كمستوى أداء جيد ، ومقبول لكل جزء من اجزاء المقياس ( المعرفة — الفهم — التطبيق ) وأيضا المقياس ككل .

#### سادساً : تقدير صدق محتوى الاختبار وثباته :

تم تقدير صدق محتوى الاختبار عن طريق مزاجعة مفردات الاختبار بالاهداف وكذلك تم حساب معاملات الثبات لاجزاء المقياس الثلاثة . وذلك بتطبيق المقياس في تجربة استطلاعية على عينة تتكون من ٨٢٢ طالبا وطالبة . ثم اعادته تطبيقه على عينة تتكون من ٣١٤ طالبا وطالبة . رغم حساب معامل السهولة لاسئلة المقياس ، وانحصرت هذه المعاملات بين ١٩ ٪ — ٨٩ ٪ ، وتم حساب معامل تمييز وحدات القياس باستخدام طريقة كوكس وفارجاس Cox and Vargas

١٩٦٦ ووجدت معاملات التمييز لوحدة القياس محصورة بين ٠.٤ ر — ٠.٥٨ ر وبذلك أخذ المقياس صورته النهائية .

( ٦ ) دراسة أحمد الرفاعي غنيم ( ١٩٨٤ ) :

تهدف الى العلاقة بين اجتماعية المهنة للمعلم وتحصيل تلاميذ الصف السابع الاساسى فى الرياضيات .

فى هذه الدراسة :

تم اعداد اختبار تحصيلى مرجعى المحك فى مادة الرياضيات للصف السابع الاساسى ، حيث اتخذت الخطوات التالية فى عملية الاعداد :

— حلل محتوى مقرر الرياضيات للصف الاول الاعدادى وتم الخرج فى النهاية جدول مواصفات ، تم حساب ثبات التحليل عن طريق الاستعانة بمحلل آخر ، حيث تبين أنه لا توجد فروق دالة بين نسب تحليل الباحث والمحلل الآخر .

— حددت أهداف مقرر الرياضيات للصف السابع الاساسى ثم حلت الى أهداف سلوكية اجرائية اكر تخصيصا ، وتحديد ا ، ثم نظمت فى ترتيب هرمى وفقا لطريقة جانبيهه Gané

— وضعت مفردات الاختبار بحيث تقيس الاهداف السلوكية الاجرائية المحللة ويتناسب عددها مع جدول المواصفات المستخرج .

— تم الاستعانة بعشرين مدرسا للرياضيات كمحكمين للحكم على مدى قياس المفردات واستبعدت المفردات التى اتفق المحكمين على انها لا تقيس الهدف المندرجة تحته .

— تم تطبيق الاختبار على عينة استطلاعية مكونة من ٢٠٥ تلميذا وتلميذة ، حيث حسب الزمن المناسب للاختبار .

— حسب ثبات الاختبار المرجعى المحك باستخدام طريقة ميشيل ج • سيكوفالك  
( ١٩٧٦ ) مبلغ ٧٨

— حسب معامل الارتباط بين الاختبار ودرجات الطلاب ، فى مادة الرياضيات فى  
نهاية العام الدراسى ، فبلغ معامل الارتباط ٧٢  
( ٨ — ٧ : ٣ )

( ٧ ) دراسة محمد الدسوقي عبدالعزيز الشافعى ( ١٩٨٥ ) :

وتهدف الى بناء اختبار مرجع المحك فى مادة الهندسة الفراغية لطلاب  
الصف الثانى الثانى العلمى ، وقد تم بناء الاختبار طبقا للخطوات الاتية :-

أولا : تحديد محال ومحتوى المادة الدراسية :

اقتصر هذا البحث على منهج الهندسة الفراغية المقرر على طلاب الصف الثانى  
ثانوى العلمى فى العام الدراسى ١٩٨٣ / ١٩٨٤ وذلك ما عدا الباب الاخير  
وهو " المنشور والهرم " وذلك لتكرار دراسته فى بداية الصف الثالث الثانوى —  
وبذلك اشتمل البحث على الابواب الثلاثة الاولى وهى : المستقيمت والمستويات  
— الاسقاط العمودى — الزوايا الزوجية • وذلك ليصمم الباحث على أساسها  
مفردات الاختبار بعد أن حدد عناصر الموضوعات لمحتوى كل باب على حدة • ثم  
عرضت قائمة عناصر موضوعات المحتوى بصفة عامة على مجموعة من المحكمين فى  
صورة استطلاع رأى لمراجعتها وتقدير مدى صحتها •

ثانيا : تحديد الاهداف التعليمية لمجال المحتوى المراد قياسه ، اجرائيا :

وقد شملت ٨٢ هدفا اجرائيا ثم عرضت هذه القائمة على مجموعة من المحكمين  
فى صورة استطلاع رأى لمراجعتها من الناحية الاجرائية ومن حيث مدى انتبايسة  
كل مفردة للهدف الاجرائى الخاص بها • وهذه الاهداف موزعة على مستويات  
بلوم Bloom وهى : ( المعرفة — الفهم — التطبيق ، حل المشكلات  
وتشمل المستويات الاربعة العليا من تصنيف بلوم •

### ثالثا : اعداد مواصفات الاختبار :

اعتمدت هذه الدراسة على طريقة بوفام Popham ( ١٩٧٥ ، ١٩٧٨ ، ١٩٨١ ) فى اعداد مواصفات الاختبار لكل هدف من الاهداف الاجرائية وتتكون هذه المواصفات من خمسة عناصر أساسية هى :

الوصف العام " عبارة الهدف " ، عينة المفردة ، خصائص المثير ، خصائص الاستجابة ، ملحق المواصفات . وقد تم عرضها على مجموعة من المحكمين فى صورة استطلاع رأى لتحديد مدى اختبارها ووضوحها فى تحديد نوع السلوك المراد قياسه .

### رابعا : بناء مفردات الاختبار :

قام الباحث ببناء مفردتى اختبار متكافئتين لقياس كل هدف سلوكى وقد استخدم الباحث الثلاثة أنواع من المفردات :

( ١ ) مفردات اختيار من متعدد .

( ٢ ) مفردات اجابة قصيرة .

( ٣ ) مفردات مقال .

وبلغت عدد المفردات ١٦٤ مفردة ، ثم عرضت قائمة المفردات على مجموعة من المحكمين وعرضت الاهداف الاجرائية المعدلة على لجنة اخرى من المحكمين لتقدير صدق محتوى المقياس ، عن طريق مزاجعة مفردات الاختبار مع الاهداف وتم حساب صدق مفردات الاختبار من خلال البيانات التقديرية للمحكمين وذلك باستخدام معامل تطابق المفردة بالهدف ، ووجد أن قيمة هذا المعامل لكل مفردة كانت واحدا صحيحا وذلك دليل على صدق جميع مفردات الاختبار ومن ثم صدق محتوى الاختبار .

### خامسا : تحديد مستوى الاداء المقبول ( الدرجة الفاصلة )

بالاستعانة بمجموعة من المحكمين حدد الباحث النسبة ٨٠ ٪ كمستوى أداء جيد ، ومقبول لكل جزء من أجزاء الاختبار ( المعرفة — الفهم — التطبيق — حل المشكلات ) وأيضا الاختبار ككل .

إدسا : تقدير صدق محتوى الاختبار وثباته :

تم تقدير صدق محتوى الاختبار عن طريق مزاجية مفردات الاختبار بالاهداف وكذلك تم حساب معاملات الثبات لاجزاء الاختبار الثلاثة وذلك بتطبيق الاختبار في تجربة استطلاعية بصورتين أ ، ب على عينة من طلاب وطالبات المدارس الثانوية بمحافظة الدقهلية والغربية ، وقد بلغ عدد أفراد العينة الذين طبقت عليهم الصورة ( أ ) من الاختبار ٤٦٠ طالبا وطالبة ، بينما بلغ عدد أفراد الذين طبقت عليهم الصورة ( ب ) من الاختبار ٣٢٦ طالبا وطالبة وتم استبعاد أوراق اجابة ١٥٣ طالبا وطالبة وذلك لاسباب مخلة بقواعد الامتحانات -- وتم حساب معامل السهولة لاسئلة الاختبار ، وتم تقدير الثبات باستخدام الطرق غير الارتباطية التي اقترحها سوامتران وزملائه ( ١٩٧٤ ) ، والطرق الارتباطية التي تتمثل في معادلة كيودر ريتشاردسون ومعامل ألفا (  $\alpha$  ) وكان معامل الثبات الناتج تتراوح بين ٨٧ر - ٩٥ر

( ٢١ )

(( فيسب على دراسات المحور الأول ))

بعد عرض لبعض الدراسات السابقة التي تمكن الباحث من الوصول إليها في مجال بناء الاختبارات المرجعية المحك • فانه يمكن التوصل الى بعض الملاحظات التي تبين نقاط التشابه والاختلاف بين هذه الدراسات ومدى امكانية الاستفادة من هذه الدراسات بصفة عامة •

(١) تتشابه الدراسات السابقة في بعض الخطوات البحثية والادوات المستخدمة فيه • فقد اتبعت معظم الدراسات السابقة طريقة تحليل المقدرات الدراسية وعرض هذا التحليل على مجموعة من الخبراء والمحكمين لاقرارها أو لتعديلها ثم اعداد قائمة التحليل في صورتها النهائية مثل دراسة سميث ( ١٩٨٠ ) ودراسة صلاح علام ( ١٩٨٠ ) ، ودراسة اسحاق حنا ( ١٩٨١ ) ، ودراسة عماد عبدالمسيح ( ١٩٨٢ ) ، ودراسة أحمد الرفاعي غنيم ( ١٩٨٤ ) ، ودراسة محمد الدسوقي الشافعي ( ١٩٨٥ ) •

(٢) تتفق معظم الدراسات السابقة في اختيار نمط الاختيار من متعدد من الاختبارات لبناء المقياس التحصيلي المرجعي المحك الخاص بالدراسة مثل دراسة اسحاق حنا بطرس ( ١٩٨١ ) ، ودراسة عماد عبدالمسيح ( ١٩٨٢ ) ، ودراسة محمد الدسوقي الشافعي ( ١٩٨٥ )

(٣) قيام الاختبارات المرجعية المحك على الاهداف التعليمية الاجرائية عند بنائها مثل دراسة سميث ( ١٩٨٠ ) ، دراسة اسحاق حنا بطرس ( ١٩٨١ ) ، ودراسة عماد عبدالمسيح ( ١٩٨٢ ) ، ودراسة أحمد الرفاعي غنيم ( ١٩٨٤ ) •

(٤) اعتماد جميع الاختبارات المرجعية المحك على صدق المحتوى عند تحديد صدق الاختبار مثل دراسة عماد عبدالمسيح ، ( ١٩٨٢ ) ، ودراسة أحمد الرفاعي غنيم ( ١٩٨٤ ) ودراسة محمد الدسوقي الشافعي ( ١٩٨٥ ) •

(٥) اعتمدت معظم الدراسات السابقة على تحديد المستوى ٨٠% كحد أدنى لمستوى الاداء المقبول لتصنيف الطلاب الى متمكين وغير متمكين مثل دراسة عماد عبدالمسيح (١٩٨٢) ، دراسة محمد الدسوقي (١٩٨٥) ، دراسة أحمد الرفاعي غنيم (١٩٨٤) .

(٦) اعتمدت معظم الدراسات السابقة في تحليلها لمفردات الاختبارات المرجعية المحك على أسلوبين :

(أ) أسلوب استخدام نسب المحكمين .

(ب) أسلوب كو كس وفيرجاس .

مثل دراسة عماد عبدالمسيح (١٩٨٢) ، دراسة محمد الدسوقي (١٩٨٥) ، دراسة أحمد الرفاعي غنيم (١٩٨٤) .

(٧) اعتمدت معظم الدراسات السابقة في ايجادها الماملات ثبات الاختبارات المرجعية — المحك على تصنيف الطلاب الى فئتين متقنين وغير متقنين مثل دراسة اسحاق حنا (١٩٨١) ، ودراسة عماد عبدالمسيح (١٩٨٢) ، ودراسة محمد الدسوقي الشافعى (١٩٨٥) .

(٨) بعض الدراسات السابقة استخدمت عينات صغيرة نسبيا مثل دراسة سميث (١٩٨٠) حيث بلغ العدد الكلى لافراد العينة ٤٠ طالبا ، دراسة أحمد الرفاعي غنيم (١٩٨٤) حيث بلغ العدد الكلى لافراد العينة ٢٠٥ طالبا .

والبعض الاخر من الدراسات السابقة استخدمت عينات كبيرة مثل دراسة اسحاق حنا بطرس (١٩٨١) حيث بلغ العدد الكلى لافراد العينة ١٠٣٠ طالبا ، دراسة عماد عبدالمسيح (١٩٨٢) حيث بلغ العدد الكلى لافراد العينة ١١٣٦ طالبا ، دراسة محمد الدسوقي الشافعى (١٩٨٥) حيث بلغ العدد الكلى لافراد العينة ٢٨٦ طالبا .



## ثانياً : المحور الثاني

دراسات تنساولت

" الطرق المختلفة لتقدير الدرجات وأثرها  
على ثبات الاختبارات مرجعية المعيار "

=====

(١) دراسة روبرت م : / بي (١٩٧٠)

Robert, M. Rippey, The University of Chicago

تهدف الى مقارنة خمس طرق لتقدير الدرجات بعضهم ببعض لمعرفة أى من هذه الطرق تعطى ثباتا أكثر للاختبار المتعدد الاختيارات .

— الطرق الخمس هـى :

- ١- الاحتمال المقترح للاجابة الصحيحة .
- ٢- الطريقة اللوغارتمية .
- ٣- الطريقة الكروية .
- ٤- الطريقة الاقليدية .
- ٥- طريقة الاختيار الاستنتاجى .

وقد استخدم الباحث اختبار ( step ) للكتابة ، وهذا الاختبار له عدة نويات ، وقد التقى الباحث بالمستوى الاول وفى هذا المستوى توجد صورتان أ ، ب وعلى صورة جزأان وقد طبق هذا الاختبار على ١٤ مجموعة من طلاب المدرسة العليا وكانت هذه المجموعات مختلفة فى نوعية الدراسة كما كانت مختلفة فى العدد حيث تراوح عدد كل مجموعة من تلك المجموعات ما بين ١٩ ، ٥٢ طالبا وكان يطلب منهم أن يعطوا الاجابات المتعددة لكل سؤال درجات احتمالية تعبر عن مدى صحة هذه الاجابات ، بحيث يكون مجموع تلك الدرجات الاحتمالية يساوى واحدا صحيحا وتصحيح اجابات الطلاب المختلفة على الاختبار بالخمس طرق السالفة الذكر ، واستخراج معاملات الثبات فى كل طريقة ولكل مجموعة من المجموعات الاربعة عشر ، وحساب متوسط معاملات الثبات لكل الطريقة من الطرفين الخمس ، تم الحصول على النتائج الآتية :

- متوسط معاملات ثبات المجموعات فى طريقة الاحتمال المقترح للاجابة الصحيحة ٦٩ر
- متوسط معاملات ثبات المجموعات فى الطريقة اللوغارتمية ٥٠ر
- متوسط معاملات ثبات المجموعات فى الطريقة الكروية ٤٩ر

- متوسط معاملات ثبات المجموعات في الطريقة الاقلدية ٨ م
- متوسط معاملات ثبات المجموعات في طريقة الاختيار الاستنتاجي ٤٧ ر

ومن هذه النتائج يتضح أن طريقة الاحتمال المقترح للإجابة الصحيحة تعطى أفضل ثبات ممكن يليها الطريقة الاقلدية يليها الطريقة الكووية ، أما طريقة الاختيار الاستنتاجي فهي التي تعطى أقل ثباتا في الطرق الخمس المذكورة .  
( ٤٧ : ١٦٥ — ١٧٠ ) ، ( ٢ : ٢٧٣ — ٢٧٦ )

(٢) دراسة روجر أ. كوهلر Roger, A. & Koehler (١٩٧١)  
Ohio University.

تهدف الى معرفة أثر استخدام طريقتين من طرق الثقة على ثبات وصدق اختبار تحصيلي تقيس مفرداته اللغة والمواد الاجتماعية والعلوم . وهذا الاختبار يتكون من ٦٠ مفردة ، ومكون من ثلاثة أجزاء كل جزء يتكون من ٢٠ سؤالا ، منها ١٠ أسئلة ذات اختياريين فقط ، والبقية الثانية ذات خمسة اختيارات .

وقد طبق الاختبار على ٥٣٥ طالبا باعصف الحادي عشر ثم اختياريهم من مدرستين تقعان غرب مدينة نيويورك . ونسبت العينة الى ثلاثة مجموعات :

المجموعة الأولى : تتكون من ١٨٠ طالبا ويجيبون على الاختيار بالطريقة التقليدية المعروفة .

المجموعة الثانية : تتكون من ١٥٦ طالبا ، طلب منهم وضع نسبة احتمالية تعبر عن مقدار صوت كل اختيار احتمالي موجود في السؤال ، حيث كان مجموع النسب الاحتمالية المحتملة ( الموزعة ) في أي سؤال تساوي واحدا صحيحا .

المجموعة الثالثة : تتكون من ١٩٩ طالبا ، طلب منهم وضع درجة لكل اختيار احتمالي تعبر عن مقدار صوابه بحسب - يتم توزيع ١٠ درجات على كل الاحتمالات الاختيارية للسؤال الواحد .

وبعد الحصول على اجابات الطلاب في المجموعات الثلاث تم تصحيح اجابات طلاب المجموعة الاولى بالطريقة العادية ( صفر ، ١ ) أما اجابات طلاب المجموعتين الثانية والثالثة فقد عولجت في التصحيح بواسطة دالتين هما الدالة التريعبسية ، والدالة اللوغاريتمية ، وبعد تصحيح اجابات المجموعتين الثانية والثالثة بالدالتين التريعبية ، اللوغاريتمية واستخراج معاملات الثبات للاختبار من تلك المجموعتين بالاضافة الى المجموعة التقليدية — بواسطة معامل ألفا تم الحصول على النتائج الآتية :

- معامل الثبات في المجموعة الاولى " التقليدية " ٥٠ ر
- معامل الثبات في المجموعة الثانية " الدالة التريعبية " ٥٢ ر
- معامل الثبات في المجموعة الثالثة " الدالة التريعبية " ٥٠ ر
- معامل الثبات في المجموعة الثانية " الدالة اللوغاريتمية " ٥١ ر
- معامل الثبات في المجموعة الثالثة " الدالة اللوغاريتمية " ٥٣ ر

ومن الواضح أن هناك زيادة طفيفة بين المجموعة الثانية والاولى ، كما أن هناك زيادة بين معامل الثبات قدرها ٠٣ ر بين المجموعة الاولى من المجموعة الثالثة المعالجة بالدالة اللوغاريتمية ، وصفة عامة فإن تلك الفروق بسيطة .

وبالنسبة للصدق فقد حسب صدق الذكويين الفرضي ولقد أثبتت النتائج وجود فروق طفيفة في الصدق لصالح الطريقة التقليدية عن الطريقتين الاخرتين .

( ٢٩٧ : ٤١ — ٣٠٣ ) ، ( ٢ : ٢٧٧ — ٢٧٩ )

(٣) دراسة ريتشارد م. ايفانس ، كاثي مسفيلدت ( ١٩٧٤ )

Richard, M.E. Vans and Kathy Misfeldt

تهدف الى تحديد أثر اجراءات الدرجات الذاتية ( المتمثلة فى اعطاء المفحوصين تغذية رجعية كلية ) على ثبات الاختبار فى هذه الدراسة . كانت العينة مكونة من ٥٢ طالبا تحت التخرج من قسم علم النفس التعليمى قسموا الى مجموعتين متساويتين ( بواقع ٢٦ طالبا فى كل مجموعة ) .

طبق على احدى المجموعتين امتحان نصف العام " وكان مكونا من خمسين مفردة ) بالطريقة التقليدية ، بينما طبق على المجموعة الثانية نفس الامتحان ولكن فى وجود التغذية الرجعية الكلية ، وقد تم تصحيح أوراق المجموعة التقليدية على أساس واحد أو صفر لكل سؤال ثم جمع الدرجات جمعا ناديا .

أما مجموعة التغذية الرجعية فقد تم تصحيح أوراقها بطريقتين :

الطريقة الاولى : كان ينظر الى الاستجابة الاولى فقط ( دون اعتبار لبقية استجابات نفس السؤال ) فإذا كانت هذه الاستجابة صحيحة حصل المفحوص فى هذا السؤال على الدرجة " واحدا " وإذا كانت خاطئة حصل الدرجة " صفر " وهكذا .

الطريقة الثانية : وفى هذه الطريقة كان التصحيح يتم على أساس النظر الى الاستجابات الخاصة بالسؤال ف يأخذ المفحوص الدرجة ٣ اذا وصل الى الاجابة الصحيحة على السؤال من أول محاولة ، ويحصل على الدرجة ٢ أو ١ أو صفر ، اذا وصل الى الاجابة الصحيحة من المحاولة الثانية أو الثالثة أو الرابعة على الترتيب وتسمى هذه الطريقة بأسلوب الدرجات الذاتية .

وقد أعيد تطبيق نفس الاختبار على نفس المجموعتين بعد أسبوع من الاجراء الاول ولان طريقة التطبيق ماثلة تماما للطريقة التى اتبعت فى الاجراء الاول .

وفى نهاية العام الدراسى حصل أفراد المجموعتين ( وعددهم ٥٢ طالبا ) على الامتحان النهائى فى وجود التغذية الرجعية الكلية ، وتم تصحيح الاوراق بنفس الطريقتين السابقتين .

وقد حسب الثبات فى كل حالة بواسطة التجزئة النصفية وباستخدام معادلة سبيرمان براون .

وقد أظهرت النتائج عموما أن أسلوب الدرجات الذاتية تزيد الثبات عن الأسلوب التقليدى وعن أسلوب الرقم الصحيح الاستنتاجى ( INR ) .

( ٢٨ : ١٢٤٨ ) ، ( ٢ : ٣٠٤ — ٣٠٧ )

( ٤ ) دراسة جيمس ج . ديموند ( ١٩٧٥ )

James J. Diamond; University of Pennsylvania

تهدف الى تقدير الصدق والثبات من خلال نوعين من الدرجات :

— درجات يتم التوصل اليها من اجراءات تتوافر فيها معلومات على شكل توجيهات تعطى للمفحوصين ، وهذه الدرجات تسمى درجات تجريبية

Experimental Scoring

— درجات يتم التوصل اليها بالطريقة التقليدية وتسمى درجات الصواب .

Number-Right Scoring

وقد استخدمت هذه التوجيهات فى أربعة اختبارات تحصيلية طبقت على

٨٤ طالبا من طلاب الجامعة ، حيث كانوا يدرسون مقررا فى القياس التربوى .

الاختبار الاول : يتكون من ٢٠ مفردة ، ولبس فى أول يوم من الدراسة بصورة مبدئية

وهذا الاختبار يساعد فى ادلاء معلومات أساسية عن الخلفية

السابقة للطلاب ، كما ساعد فى تعريف الطلاب على اجراءات اعطاء

الدرجات .

أما الاختبارات الثلاثة الأخرى فكانت تتكون من مفردات عددها ٣٢ ، ٣٢ ، ٣٦ على الترتيب .

وهذه الاختبارات تعطى جوانب من المحتوى الكلى للمقرر ، ولقد صممت الاختبارات لقياس التمكن فيما يختص بالقدرة أكثر من التمايز وهذا قد تم عن طريق تكوين مفردات سهلة ولها تمييز منخفض ، وبعد انتهاء الطلاب من الإجابة ثم التصحيح حيث استخرجت مجموعتين من الدرجات .

وقد استخدم اختبار مكون من ٣٢ مفردة كمعيار خارجي للاختبارات الأربع السابقة ، وهذا الاختبار كان صادقا في محتواه ومؤسس على المادة التي درست للطلاب وهي القياس التربوي .

وقد أسفرت نتيجة هذا البحث على أن الدرجات التجريبية تعطى ثباتا أكثر من الثبات التي تعطيه الدرجات التقليدية .

أما بالنسبة لمعاملات الصدق فهي متذبذبة ، حيث لا تدل أى من الطريقتين أنها أفضل من الأخرى في إعطاء معاملات صدق كبيرة .

( ٢٧ : ١٢٩ — ١٣٣ ) ، ( ٢ : ٢٩٤ — ٢٩٧ )

(٥) دراسة وانلوب كانسوب ( ١٩٧٥ )

Wanlop Kansup, A. Ralph Hakstion University of Alberta  
University of British Columbia.

تهدف الى حساب الثبات والصدق من معالجة الدرجات لعدة طرق واجراء مقارنة بينها ، كما تهدف أيضا الى اجراء مقارنة بين معاملات الثبات والصدق المستخرج من درجات مجموعتين من التلاميذ اجابوا على عدة اختبارات ذات اختيار من متعدد بطريقتين مختلفتين .

في هذه الدراسة تم اختيار مجموعتين من تلاميذ الصف التاسع " ثلاثة اعدادى " وقسمت الى مجموعتين متجانستين ، احدهما ضابطة وتشمل ٣٤٦ تلميذا وتلميذة ، والمجموعة الاخرى تجريبية وتتكون من ٣٤٨ تلميذا وتلميذة .

وقد اختيرت هذه العينة من ثمانى مدارس بمنطقة البرتا وايدنتون ، وذلك من المستويات الاجتماعية المنخفضة والمتوسطة .

وقد تم اختيار أربعة اختبارات ذات اختيار من متعدد لتطبيقها على العينة السابقة والاختبارات الاربع هى :

- ١- اختيار لقياس القدرة اللفظية ، صورة أولى ( VA )
- ٢- اختيار لقياس القدرة اللفظية ، صورة ثانية ( VB )
- ٣- اختيار لقياس القدرة الرياضية ، صورة أولى ( RA )
- ٤- اختيار لقياس القدرة الرياضية ، صورة ثانية ( RB )

كما تم الاستعانة بأربعة اختبارات مدرسية كمعايير لصدق الاختبارات السابقة .

وقد طلب من أفراد المجموعة الضابطة الاجابة على الاختيارين ( VA ) ،  
( RA ) بالطريقة التقليدية .

أما المجموعة التجريبية فقد طلب من أفرادها أن يحددوا درجة تأكدهم من صحة الاجابة على كل سؤال من أسئلة الاختبارين ، وذلك بوضع درجة من عشرة لكل اجابة يختارونها ويعتقدون أنها صحيحة .



وقد أعيد تطبيق هذين الاختبارين على نفس المجموعتين ونفس تعليمات المرة الأولى وذلك بعد أربعة أسابيع من التطبيق الأول ، وقد تم تطبيق الاختبارين ( $R_B$ ) ( $V_B$ ) على طلاب المجموعتين بالطريقة التقليدية المتبعة للإجابة على هذين الاختبارين .

وقد سارت عملية استخراج الدرجات ومعالجتها كالآتي :-

أولاً : بالنسبة للمجموعة الضابطة :

- (١) الطريقة التقليدية في تصحيح الاختبارين .
- (٢) طريقة الدرجات الوزنية ، وفيها يحدد وزن ( أهمية ) كل سؤال بالنسبة لبقية الاسئلة .

ثانياً : المجموعة التجريبية :

استخرجت وعولجت درجات طلاب هذه المجموعة بخمس طرق :

- (١) اعتبار درجة الثقة في الإجابة الصحيحة التي وضعها الطالب على كل سؤال ( هذه الدرجة من ١٠ ) هي درجة السؤال إذا كانت الإجابة صحيحة ، وإذا كانت الإجابة غير صحيحة طرحت هذه الدرجة من ١٠ واعتبر ناتج الطرح هو درجة ذلك السؤال .
- (٢) اعتبار درجة كل سؤال ( لو ١ ن ) أي يؤخذ لوغاريتم الدرجة المحددة بالطريقة الأولى للأساس ١٠
- (٣) اعتبار درجة كل سؤال ( ١ - لو ( ١٠ - ق ) حيث ( ق ) هي الدرجة المحددة بالطريقة الأولى .
- (٤) اعتبار درجة كل سؤال (  $\frac{1}{10^Q}$  ) حيث ( ق ) هي الدرجة المحددة بالطريقة الأولى .
- (٥) تحول الدرجة ( ق ) المحددة بالطريقة الأولى لكل سؤال إلى درجة معيارية .

## نتائج الدراسة :

(١) بالنسبة للثبات فى المجموعة الضابطة :

توجد فروق فى معاملات الثبات بين الدرجات الوزنية ( رقم ٢ ) والدرجات التقليدية ( رقم ١ ) لصالح الدرجات الوزنية • وذلك عند حساب الثبات بمعامل ألفا ، أما فى حالة اعادة الاختبار فان تلك الفروق لم تظهر بصورة دالة احصائيا •

(٢) ثبت أنه لا توجد فروق فى معاملات الصدق المحسوبة لكلا النوعين من الدرجات وذلك بالنسبة للاختبار اللفظى (  $V_A$  ) • سواء عند استخدام الاختبار اللفظى (  $V_B$  ) كمعيار أو عند استخدام اختبار مدرسى مناظر كمعيار •

أما بالنسبة للاختبار الرياضى (  $R_A$  ) فان الصدق المحسوب له من الدرجات الوزنية يقل عن الصدق المحسوب من الدرجات التقليدية وهذا النقص له دلالة الاحصائية •

بالنسبة للثبات فى المجموعة التربوية :

توجد بعض الفروق فى معاملات الثبات بين الاجراءات الخمسة التى أتبع فى معالجة الدرجات وذلك عند استخدام معامل ألفا وهذه الفروق وجدت بصفة خاصة بين الطريقة اللوغاريتمية ( الثالثة ) والطريقة التربيعية ( الرابعة ) لصالح الطريقة التربيعية (

أما بالنسبة لاعادة الاختبار فقد أظهرت النتائج أنه لا توجد فروق بين معاملات الثبات المحسوبة بالاساليب الخمسة المذكورة •

بالنسبة للصدق فى المجموعة التجريبية :

أظهرت النتائج أنه توجد فروق فى معاملات صدق الاختبار اللفظى (  $V_A$  ) بين الطريقة الثالثة ( اللوغاريتمية ) والرابعة التربيعية ) لصالح الطريقة الثالثة ( اللوغاريتمية ) وذلك عند أخذ اختبار مدرسى مناظر كمعيار خارجى • أما عند أخذ الاختبار (  $V_B$  ) كمعيار داخلى فان هذه الفروق لا تكون دالة احصائيا •

وبالنسبة للاختبار (  $R_A$  ) للقدرة الرياضية فقد أظهرت النتائج بعض الفروق بين الطريقة الخامسة والثانية وذلك عند استخدام اختبار تحصيلي مناظر كمعيار داخلي ، بينما لا توجد فروق في صدق الاختبار الرياضي (  $R_A$  ) بين هذه الطرق عند استخدام الاختبار (  $R_B$  ) كمعيار داخلي .

( ٤٠ : ٢٢٩ — ٢٣٩ ) ، ( ٢ : ٢٨٧ — ٢٩٢ )

( ٦ ) دراسة جيرالدس ، وكارل أ. لونج ( ١٩٧٩ )

Geralds, Hanna and Carl, A. Long, J.R.

تهدف الى بحث أثر التغذية الرجعية الكلية على ثبات الاختبارات المتعددة الاختيارات ، فقد اختيرت العينة من طلاب الصف السابع والثامن باحدى المدارس العليا في منطقة ذات مستوى اجتماعي متوسط .

وقد قسمت العينة المختارة عشوائيا الى مجموعتين ، كان عدد الطلاب في احدهما ١٣١ طالبا وفي الاخرى ١٥٦ طالبا ، ولقد طبق على المجموعة الاولى ( وعددها ١٣١ طالبا ) اختبار تحصيلي في كيفية استخدام مواد مراجع المكتبة وكان هذا الاختبار مكون من ١٨ مفردة ، كل مفردة بها أربع اختيارات . ولكن اجراء هذا الاختبار على المجموعة يتم في وجود التغذية الرجعية الكلية حيث كانت هذه التغذية تعطى للمفحوصين عن طريق الصورة الخفية بهذه الطريقة توفر التغذية الرجعية الكلية بأسلوب أحب حتى تصل الى الصواب . وما يتبعها من اجراءات التصحيح الخاصة بها . وأما المجموعة الثانية ( عددها ٥٦ طالبا ) فقد طبق عليهم نفس الاختبار المذكور ولكن بالطريقة التقليدية ( أي بدون تغذية رجعية على الاطلاق ) وعند حساب الثبات الناتج من المجموعتين باستخدام معامل ألفا تبين أن قيمة  $r_{٦٤}$  للمجموعة الاولى التي تلقت تغذية رجعية كلية ، وأن قيمة هذا المعامل للمجموعة الثانية ( التي لم يترك أفرادها أي تغذية ، يساوي ٣ والفرق بينهما غير دال احصائيا .

( ١ : ٤٦٤ — ٤٦٦ )

تهدف الى دراسة أثر التغذية الرجعية الفورية على صدق وثبات الاختبارات العقلية ذات الاختيار من متعدد . وللتحقق من هذا الهدف أخذ الباحث عينة عددها ٧٢٠٠ طالب وطالبة منهم ١٨٠٠ طالب كعينة استطلاعية ، ٥٤٠٠ طالب عينة نهائية من تلاميذ المدارس الابتدائية والاعدادية والثانوية والجامعة ، وقسم هذه العينة الى ثلاث مجموعات كما يلي :—

- (أ) مجموعة ضابطة : تجيب على الاختبارات بالطريقة التقليدية .  
(ب) المجموعة التجريبية الاولى : تجيب على الاختبارات مع وجود التغذية المرتدة الجزئية .  
(ج) المجموعة التجريبية الثانية : تجيب على الاختبارات مع وجود التغذية المرتدة الكلية .

وكانت أدوات الباحث عبارة عن اختبارات عقلية وهى :

- |                              |                         |
|------------------------------|-------------------------|
| (١) اختبار الذكاء المصور     | (اعداد أحمد زكى صالح)   |
| (٢) اختبار الذكاء غير اللفظى | (اعداد عطية محمود هنا)  |
| (٣) اختبار الذكاء العالى     | (اعداد السيد محمد خيرى) |
| (٤) اختبار الذكاء الاعدادى   | (اعداد السيد محمد خيرى) |

وحاول الباحث أن يجعل المجموعات الثلاثة متكافئة باستخدام اختبارى الذكاء الاعدادى والعالى واستخدام مستوى التحصيل السابق .

وبعد تطبيق الاختبارات قام الباحث بتصحيح فى المجموعات الثلاث كما يلي :

المجموعة الضابطة : تم تصحيح اجاباتهم بالطريقة التقليدية .  
المجموعة التجريبية الاولى : تم تصحيح الاجابات بنفس طريقة المجموعة الضابطة والاختلاف فقط فى طريقة الاجراء .

المجموعة التجريبية الثانية : أخذت الدرجات طبقا للمعادلة الآتية :

درجة السؤال = عدد الاحتمالات الاختيارية — عدد المحاولات للوصول للإجابة

الصحيحة .

وبعد الحصول على الدرجات قام الباحث بحساب الثبات للمجموعات الثلاث  
بعدة معادلات هي : معامل ألفا ، معادلة جتمان ، معادلة سبيرمان وبراون  
، ومعادلة فلت .

توصل الباحث الى أن معامل الثبات باستخدام التغذية المرتدة الجزئية  
أعلى منه باستخدام الطريقة التقليدية ، كما توصل الى أن معامل الثبات باستخدام  
التغذية المرتدة الكلية أعلى منه باستخدام التغذية المرتدة الجزئية .  
وما يؤخذ على هذه الدراسة استخدامها نوعا واحدا من الاختبارات  
العقلية .

#### ( ٨ ) دراسة عادل العدل ( ١٩٨٦ )

تهدف الى دراسة أثر بعض طرق تقدير الدرجات للفردات على ثبات الاختبارات  
ذات الاختيار من متعدد .

في هذه الدراسة استخدمت أربع طرق من طرق تقدير الدرجات :

العينة : أولا — عينة الادوات :

١ — اختبار تحصيلي موضوعي في الفيزياء لتلاميذ الصف الاول الثانوي .

٢ — اختبار القدرة على التفكير الاستنباطي .

ثانيا — عينة الممتحنين :

— العينة الاستطلاعية وبلغت ١٩٨ طالبا من الصف الاول الثانوي .

— العينة النهائية وبلغت ٦٣٠ طالبا من الصف الاول الثانوي .

وبالتالي بلغت العينة الكلية ٨٢٨ طالبا من الصف الاول من المرحلة الثانوية :

الاجراءات :

١ — قيام الباحث باعداد اختبار تحصيلي موضوعي في مادة الفيزياء للصف الاول من المرحلة  
الثانوية .

- ٢- قام الباحث باعداد اختبار لقياس القدرة على التفكير الاستنباطى .
- ٣- اختار الباحث ثلاث مجموعات متكافئة فى مستوى التحصيل السابق من طلبة الصف الاول من المرحلة الثانوية وطلب عليهم الاختبارين على المجموعة الضابطة بالطريقة التقليدية ، وعلى المجموعات التجريبية بالطرق المقترحة ( احتمال مقترح ، لوغاريتمية ، التجريبية ) .
- ٤- قام الباحث بتصحيح الاختبارين فى المجموعة الضابطة على أساس ( ١ ، صفر ) وفى المجموعات التجريبية على أساس الطرق المقترحة .
- ٥- تم حساب معامل الثبات بطريقة التجزئة النصفية ( معادلة سبيرمان و براون ) ومعادلة قلت ، ومعادلة جتمان ، وطريقة معامل ألفا وطريقة اعادة الاختبار .
- ٦- تم تحليل النتائج بالطرق الاحصائية المناسبة

#### النتائج :

- من خلال الاجراءات السابقة تم اتوصل الى النتائج الاتية :
- أولا : عند مقارنة طرق تقدير الدرجات :
- يختلف الثبات المحسوب بالطرق المختلفة لحساب الثبات وذلك للطرق المختلفة لتقدير الدرجات للاختبارين العلى والتحصيلى .
- ثانيا : عند مقارنة طرق حساب الثبات :
- ( أ ) بالنسبة للاختبار التحصيلى :
- ( ١ ) لا توجد فروق ذات دلالة احصائية بين معاملات الثبات المحسوبة بمعادلة سبيرمان - براون ومعادلة كرونباخ وكذلك لا توجد فروق ذات دلالة احصائية بين معاملات الثبات المحسوبة بمعادلة كرونباخ وطريقة اعادة الاختبار وذلك لطرق تقدير الدرجات الاربع .
- ( ٢ ) توجد فروق ذات دلالة احصائية بين معاملات الثبات المحسوبة بمعادلة ، سبيرمان - براون وطريقة اعادة الاختبار وذلك لطريقتين من طرق تقدير الدرجات هما طريقة الاحتمال المقترح والطريقة اللوغاريتمية .

( ب ) بالنسبة للاختبار العقلى :

١- لا توجد فروق ذات دلالة احصائية بين معاملات الثبات المحسوبة بمعادلة سبيرمان - براون ومعادلة كرونباخ وذلك لطرق تقدير الدرجات الاربع .

٢- توجد فروق ذات دلالة احصائية بين معاملات الثبات المحسوبة بمعادلة سبيرمان - براون وطريقة اعادة الاختبار وكذلك بين معاملات الثبات المحسوبة بمعادلة كرونباخ وطريقة اعادة الاختبار وذلك لطرق تقدير الدرجات الاربع .

ثالثا : عند المقارنة بين نوعى الاختبار :

وجد الباحث أن النتائج فى الاختبار التحصيلى تختلف عنها فى حالة الاختبار العقلى .

---

(( تعقيب على دراسات المحور الثانى ))

أولا : ان طرق تقدير الدرجات التى استخدمت فى الدراسات السابقة تحتاج فى الغالب الى جهد اكبر ووقت أكثر وتكلفة أكثر من استخدام الطريقة التقليدية .

ثانيا : ثبت من الدراسات السابقة أن هناك عدة طرق لتقدير الدرجات تؤدى الى رفع معامل الثبات عن نظيره المستخرج من الدرجات التقليدية ومن هذه الدراسات دراسة كوهلر ( ١٩٧٠ ) ، والتى قد أثبتت أن الطريقة اللوغاريتمية والطريقة التربيعية لم تؤد الا الى زيادة طفيفة فى معامل الثبات بالمقارنة الى الطريقة التقليدية ، دراسة جيمس ج . ديوند ( ١٩٧٢ ) ، أثبتت أن الدرجات التجريبية تعطى ثباتا أكثر من التى تعطيه الدرجات التقليدية ، دراسة ريتشارد م . ايفانس ، كاثي مسفلدت ( ١٩٧٤ ) ، دراسة أحمد الرفاعى غنيم ( ١٩٨٣ ) ، أثبتا أن أسلوب الدرجات الذاتية يعطى ثباتا أفضل من الأسلوب التقليدى ، دراسة عادل الحدل ( ١٩٨٦ ) ، أثبتت أن طريقة الاحتمال المقترح أعطت أعلى ثبات من الطريقة التقليدية واللوغاريتمية والتجريبية .

ثالثا : اهتمت الدراسات السابقة بالصدق والثبات الكلى للاختبار دون الاهتمام بصدق وثبات مفردات الاختبار وهذه نقطة تستحق البحث حيث أن صدق وثبات الاختبار لكل يتوقف الى حد كبير على ثبات وصدق المفردات التى يتكون منها .

رابعا : ركزت معظم الدراسات السابقة على أخذ عينات من طلاب الجامعة والمرحلة الثانوية مثل دراسة روبرت م . ريس ( ١٩٧٠ ) ، اختيرت العينات من المدارس العليا ، دراسة وانلوب كسوب ( ١٩٧٥ ) اشتقت العينات من طلبة الصف التاسع والاساسى . ولم توجد الا دراسة أحمد الرفاعى غنيم التى استخدمت عينة من تلاميذ المدارس الابتدائية والاعدادية والثانوية والجامعة .



خامسا : هناك دراسات استخدمت مجموعة واحدة من الطلاب لاستخراج نوعين من الدرجات مثل دراسة دافيدالن جللمان وبولافرى ( ١٩٧٢ ) ، وقد ترتب على ذلك تأثير اجابات الطلاب بالتوجيهات المعطاه لهم ، وهذا قد يؤثر بدوره على الدرجات التقليدية المستخرجة من اجاباتهم .

سادسا : معظم الدراسات التى استخدمت أكثر من مجموعة مفحوصين لم تراعى تساوى هذه المجموعات فى العدد مثل دراسة روبرت م . رى ( ١٩٧٠ ) ، دراسة جيرالدس ، وكارل أ . لونج ( ١٩٧٩ ) ، دراسة روجر أ . كوهلر ( ١٩٧١ ) دراسة كنسوب ( ١٩٧٥ )

سابعا : معظم الدراسات السابقة لم تجرى تجارب استطلاعية لتحديد الزمن المناسب لافراد العينة ، حيث أن زمن الاختبار يؤثر على ثباته وصدقه ، دراسة روبرت م . رى ( ١٩٧٠ ) ، دراسة روجر أ . كوهلر ( ١٩٧١ ) ، دراسة جيمس ج . ديموند ( ١٩٧٥ ) ، دراسة وانلوت كنسوب ( ١٩٧٥ ) .

ثامنا : لم تستخدم أى دراسة من الدراسات السابقة الاختبارات المرجعية المحك .

### فروض البحث :

واستنادا الى ما أثير في الفصل الاول من تساؤلات ، وطبقا للدراسات السابقة ، فان الدراسة الحالية تحاول اختبار صحة الفروض التالية :

- ( ١ ) يختلف ثبات الاختبارات المرجعية المحك ذات الاختيار من متعدد باختلاف طرق تقدير الدرجات للمفردات .
- ( ٢ ) يختلف ثبات الاختبارات المرجعية المحك ذات الاختيار من متعدد باختلاف طرق حساب هذا الثبات .
- ( ٣ ) يوجد تأثير مشترك لكل من طرق تقدير الدرجات وطرق حساب الثبات على ثبات الاختبارات المرجعية المحك .