الفصل الثالث

الدراسات السابقة

- دراسات تناولت مقارنة بعض طرق حساب معامل ثبات الاختبارات محكية المرجع
- دراسات تناولت حساب معامل ثبات الاختبارات محكية المرجع وعلاقته ببعض المتغيرات
- تعقيب عام على الدراسات السابقة
- فروض الدراسة

الفصل الثالث

الدراسات السابقة

يمكن تقسيم الدراسات السابقة التي توصلت إليها الباحثة في محورين رئيسيين :

- دراسات تناولت مقارنة بعض طرق حساب معامل ثبات الاختبارات محكية المرجع .
- دراسات تناولت حساب معامل ثبات الاختبارات محكية المرجع وعلاقته ببعض المتغيرات .

أولاً : دراسات تناولت مقارنة بعض طرق حساب معامل ثبات الاختبارات محكية المرجع :

١- دراسة صبكوفياك (Subkoviack , 1976) :

الهدف من هذه الدراسة حساب ثبات الاختبار محكي المرجع باستخدام الطرق التالية طريقة ليفنجستون ، طريقة هارس ، طريقة سوامينثان وهامبلتون وألجينا ، طريقة مارشال وهيرتل .

وقد توصلت هذه الدراسة إلى ما يلي :

- معامل ليفنجستون حساس للاختلاف بين درجة القطع والمتوسط .
- معامل هارس حساس للنسب في مجموعات الإتقان وعدم الإتقان أكثر من حساسيته للاتساق الذي يعين الطلاب على أساسه في حالات الإتقان .
- مؤشرات سوامينثان وهامبلتون وألجينا ، ومارشال وهيرتل أكثر حساسية لاتساق التعيين حيث يتطلب الأول تطبيق الاختبار مرتين .

٢- دراسة صبكوفياك (Subkoviack , 1980) :

الهدف من هذه الدراسة مقارنة طرق حساب معامل ثبات الاختبارات محكية المرجع التالية : طريقة سوامينثان وهامبلتون وألجينا ، طريقة هينا ، طريقة صبكوفياك ، وطريقة مارشال وهيرتل .

استخدمت عينة مكونة من (١٥٨٦) طالباً طبق على كل منهم أشكال متكافئة لاختبارات مكونة من (١٠ ، ٣٠ ، ٥٠) مفردة ، وذلك باستخدام درجات قطع (٥٠% ، ٦٠% ، ٧٠% ، ٨٠%) من المفردات الصحيحة ، ثم حسبت نسبة الطلاب المصنفين باتساق

كمتقن / متقن أو غير متقن / غير متقن في كلا الشكلين وذلك لكل درجة قطع ولكل طول اختبار .

سحبت من العينة الأصلية (٥٠) عينة بشكل عشوائي مع إرجاع بحيث تكون كل منها من (٣٠) طالباً ، ثم أعيد نفس الإجراء حيث سحبت (٥٠) عينة أخرى تكون كل منها من (٣٠٠) طالب .

وقد بينت نتائج هذه الدراسة ما يلي :

- طريقة سوامينثان هي الطريقة الأبسط في الحساب كما أنها تعطي تقديرات غير متحيزة ، ومن مساوئها أنها تتطلب مرتي تطبيق وتميل أخطاء تقديراتها لأن تكون كبيرة نسبياً لعينات في حجم الفصل الدراسي . وقد ينظر إلى السيئة الأولى كميزة إذا كان هناك اهتمام بدراسة تأثيرات أشكال غير متكافئة من الاختبار .

- تتشابه مميزات ومساوئ طرق هينا ، صبكوفياك ، ومارشال وهيرتل عموماً ، فلها جميعاً ميزة أنها تحتاج التطبيق مرة واحدة ولتقديراتها أخطاء معيارية صغيرة نسبياً لعينات بحجم الفصل الدراسي ، ومن مساوئها جميعاً أنها مملّة حسابياً وتعطي تقديرات متحيزة للاختبارات القصيرة .

- فيما يتصل بالمشكلة الأخيرة ــ التقديرات المتحيزة للاختبارات القصيرة ــ لكل طريقة نوع مختلف من التحيز ويظهر ذلك بوضوح في اختبار مكون من (١٠) مفردات .

- يبدو أن طريقة صبكوفياك تعطي أقل تقديرات مُهملة لدرجات قطع مثل (٥٠%) قرب مركز توزيع الدرجات وأعلى تقديرات لدرجات قطع مثل (٨٠%) عند الأطراف وذلك للاختبارات القصيرة .

- على العكس مما سبق ، تعطي طريقة مارشال وهيرتل أعلى تقديرات لدرجات قطع في مركز التوزيع وأقل التقديرات لدرجات قطع عند الأطراف .

- إذا أخذت جميع الأمور بعين الاعتبار ، يوصى باستخدام طريقة هينا حيث أنها تتطلب التطبيق مرة واحدة وتميل لإعطاء تقديرات مهملة ، ومن ناحية أخرى ، يبدو أن طريقتي صبكوفياك ومارشال وهيرتل تعطيان أيضاً تقديرات دقة معقولة في العادة ، وطبعاً تبدو طريقة سوامينثان الطريقة الوحيدة غير المتحيزة بشكل كامل .

٣- دراسة سشفير (Schafeer , 1983) :

الهدف من هذه الدراسة مقارنة طرق حساب معامل ثبات الاختبارات محكية المرجع التي تستخدم تطبيق الاختبار مرة واحدة أو مرتين ، هذه الطرق هي : معامل كابا ، تقدير هينا لكابا ، ومعامل الاتفاق لصبكوفياك .

استخدمت هذه الدراسة بيانات (٣٢٥) طالباً اختبروا مرتين بمفردات متماثلة وذلك لمعرفة أثر خمس خصائص للاختبار والطالب وهي (طول الاختبار ، درجة القطع ، قدرة الطالب ، حجم العينة ، ومحتوى الاختبار متغاير الخواص) على معامل الثبات .

قورنت معاملات ثبات الاختبارات محكية المرجع هذه مع معاملات ثبات الاختبارات معيارية المرجع (كيودر ريتشاردسون ، إعادة التطبيق) .

وقد توصلت هذه الدراسة إلى ما يلي :

- قيم معامل الاتفاق أكثر اتساقاً من قيم معامل كابا وربما يرجع ذلك لكون معامل الاتفاق لم يصحح من عامل الصدفة .
- لم تلاحظ أية علاقات اتساق بين معاملات ثبات الاختبارات محكية المرجع ومعاملات ثبات الاختبارات معيارية المرجع .
- أشارت البيانات عند تقدير ثبات الاختبارات محكية المرجع بأن تقدير هينا على عكس معامل الاتفاق ينفع كتقدير معقول للثبات عندما حدد بالمعيار (K) .

٤- دراسة وانغ (Wang , 1983) :

الهدف من هذه الدراسة التحقق من دقة أربع طرق اتساق قرار تحديد معامل ثبات الاختبارات محكية المرجع ، هذه الطرق هي :

طريقة سوامينثان ، طريقة هينا ، طريقة صبكوفياك ، وطريقة مارشال وهيرتل .

استخدمت محاكاة الكمبيوتر للتحقق من تأثيرات عدد المفردات (٥ ، ١٠ ، ٢٠) ، عدد المختبرين (١٥ ، ٣٠) ، ودرجات القطع على دقة التقديرات المستمدة من الطرق الأربع هذه وذلك باستخدام ثلاث مجموعات من الافتراضات (تغير الدرجة الحقيقية منخفضة ، متوسطة ، مرتفعة) .

وقد أشارت النتائج إلى ما يلي :

- لا يتغير متوسط تقديرات مؤشر الاتفاق عندما ينقص عدد المختبرين من (٣٠) إلى (١٥) .

- يتغير تأثير درجة القطع على متوسط التقديرات تابعاً لتغيرات الدرجة الحقيقية والطرق .

- يرتفع متوسط التقديرات مع زيادة طول الاختبار وتغير الدرجة الحقيقية في معظم الظروف المدروسة .

- تتشابه طريقتا هينا وصبكوفياك تماماً عندما تؤخذ كافة الظروف بعين الاعتبار حيث تميل هاتان الطريقتان لإعطاء أقل تقديرات كما تزودان بأقل أخطاء معيارية للتقدير .

- طريقة مارشال وهيرتل هي الأقل حساسية لتغيرات درجة القطع .

هذا وقد خلصت هذه الدراسة إلى أن مقدر سوامينثان معادل تقريباً لمؤشر الاتفاق الحقيقي مقارنة مع باقي الطرق كما أنه يميل لإعطاء أقل تقدير .

وأخيراً فقد أشارت إلى أن التغير في التقدير باستخدام طريقة سوامينثان أكبر دائماً من التغير في التقدير لباقي الطرق .

٥- دراسة هاجين (Hagen , 1983) :

قامت هذه الدراسة على مقارنة اجراءات تقدير هينا (١٩٧٦) وصبكوفياك (١٩٧٦) لمؤشرات اتساق التصنيف وكابا التي تستخدم عندما يطبق الاختبار مرة واحدة ، حيث يشير(P_o) إلى درجة الاتساق فيما يشير كابا الى درجة الاتساق بعد استبعاد عوامل الصدفة ، وعالجت كلتا الدراستين أطوال الاختبارات (٢٥ ، ٥٠ ، ٧٥) مفردة ، شكل التوزيع (طبيعي وانحرافين يساراً) ودرجات القطع (٧٠% ، ٨٠% ، ٩٠%) من الدرجات الصحيحة ، هذا وقد استخدمت الدراسة الأولى بيانات محاكاة بينما استخدمت الدراسة الثانية بيانات من تطبيقات فعلية للاختبار.

جاءت نتائج هذه الدراسة لتبين ما يلي :

- قدمت كلتا اجراءات هينا وصبكوفياك تقديرات (P_o) بأهمية متشابهة ومقبولية عالية كمعاملات ثبات .

- تقديرات صبكوفياك لكابا أكثر اتساقا من تقديرات هينا لكابا ولم تر اختلافات في سلوك المعاملات بين الدراستين الأولى والثانية .

- أثر طول الاختبار في كافة التقديرات زيادة في الأهمية مع زيادة في طول الاختبار

- لوحظ نقصان في أهمية معاملات (P_o) مع زيادة الانحراف وكان كابا في أقل قيمه مع توزيعات أقل انحرافاً .

- شكل التوزيع كان تابعاً للقرب بين درجة القطع وشكل التوزيع فقد كان (P_o) في أقل قيمه وكابا في أعلى قيمه عند اقتراب درجة القطع من منتصف التوزيع .

٦- دراسة داتشك (Dutschke , 1988) :

الهدف من هذه الدراسة تعيين خصائص التوزيع لمؤشري ثبات يستخدمان تابع مربع خطأ الاتفاق وهما : ليفنجستون ، برينان وكان ، كما حاولت تعيين خصائص توزيع خمسة مؤشرات تستخدم تابع الاتفاق وهي صبكوفياك ، هينا لـ (P_o) و (K) ، سوامينثان لـ (P_o) و (K) ، وقد تمت مقارنة الطرق السبع لحساب الثبات هذه في ظروف تغيير حجم العينة ، طول الاختبار ، ودرجة القطع .

زود توليد البيانات بالكمبيوتر باختبارات عشوائية متكافئة لـ (٢٠٠٠) مجموعة وحسبت كل من مؤشرات الثبات السبعة ، كما جمعت إحصاءات وصفية لكل مجموعة واختبرت خصائص التوزيع .

أشارت نتائج هذه الدراسة إلى ما يلي :

- اقتربت خصائص توزيع مؤشرات الثبات السبعة من المميزات الطبيعية مع زيادة حجم العينة .

- تقدير هينا لـ (P_o) هو التقدير الأكثر دقة ، مع أقل درجة من التحيز السالب .

- تقدير سوامينثان لـ (P_o) هو التقدير التالي الأفضل ولكن من مساوئه أنه يتطلب التطبيق مرتين في حين يتطلب مؤشر تقدير هينا لـ (P_o) التطبيق لمرة واحدة .

٧- دراسة آمال محروس (١٩٨٨) :

هدفت هذه الدراسة إلى تحديد أفضل طريقة لحساب معاملات ثبات الاختبارات محكية المرجع عند إجراء مقارنات بين أكثر من طريقتين حيث قامت على مقارنة الطرق التالية : طريقة ليفنجستون ، طريقة هارس ، طريقة هامبلتون ونوفيك ، وطريقة صبكوفياك .

استخدمت ثلاثة اختبارات محكية المرجع في محتوى المواد التالية :

١- التقويم والقياس النفسي ، ٢- علم النفس العام ، ٣- علم نفس التعلم المقررة على طالبات الصف الثاني بكليات البنات المتوسطة بالسعودية ، حيث تكون كل اختبار من صورتين متكافئتين ، وبلغ حجم العينة (٣٢٩) طالبة .

اختيرت درجة القطع (٦٠%) كمستوى أداء مقبول وهي الدرجة التي حددتها الرئاسة العامة لتعليم البنات كمستوى يفصل بين الناجحات والراسبات ثم حسب الثبات بالطرق المذكورة سابقاً ، كما تم حساب صدق مفردات الاختبارات بالاعتماد على مجموعة من المحكمين .

أشارت نتائج هذه الدراسة إلى ما يلي :

- أظهرت معاملات ثبات اثنين من الاختبارات المستخدمة تفوق طريقة ليفنجستون على طريقة هارس .

- تبين عند مقارنة معامل الثبات المحسوب باستخدام طريقتي ليفنجستون وصبكوفياك أن كلتاهما تعطيان معاملات ثبات متقاربة .

- ما ينطبق على طريقة ليفنجستون وهارس ينطبق على طريقة صبكوفياك وهارس من حيث تفوق طريقة صبكوفياك على طريقة هارس .

- إذا تمت المقارنة بين طريقة هامبلتون ونوفيك والطرق الثلاث الأخرى (ليفنجستون ، هارس ، صبكوفياك) نجد تفوق طريقة هامبلتون ونوفيك وعلى الرغم من تفوقها لا يشيع استخدامها لأن هذه الطريقة تعطي نتائج مرتبطة بمدى تطابق صورتي الاختبار ، فإذا كانت الصورتان أكثر تطابقاً أعطت ثباتاً عالياً وإذا كانت الصورتان أقل تطابقاً أعطت ثباتاً منخفضاً .

- إذا تمت المقارنة بين الثلاث طرق التي تعتمد على صورة واحدة للاختبار في ايجاد معامل الثبات وهي (ليفنجستون ، هارس ، صبكوفياك) نجد أن طريقة ليفنجستون تعطي ثباتاً أعلى من ثبات الطريقتين المتبقيتين .

وخلصت هذه الدراسة إلى أن طريقة ليفنجستون تعطي أعلى معدل ثبات وأن طريقة هامبلتون ونوفيك تعتبر أبسط وأسهل الطرق استخداماً .

٨- دراسة محمود إبراهيم (١٩٩٠) :

الهدف من هذه الدراسة محاولة التوصل إلى أنسب طريقة لتحديد معامل الثبات في الاختبارات محكية المرجع عند نقاط قطع مختلفة ، وعند أطوال مختلفة للاختبار ، وعند أحجام مختلفة من العينات .

وقد قامت هذه الدراسة على مقارنة الطرق التالية : طريقة هينا ، طريقة صبكوفياك ، طريقة ليفنجستون التي تعتمد على تطبيق الاختبار مرة واحدة كما حاولت التوصل للعلاقة

بين منحنى (معامل الثبات – الحد الأدنى ، التوزيع التكراري) لدرجات الاختبار عند استخدام الطرق السابقة الذكر .

استخدم اختبار محكي المرجع في مادة الرياضيات " باب المجموعات " للصف السابع تم إعداده في ضوء طريقة بابام Popham وهو من نوع الاختيار من متعدد وتكون من (١٩٠) مفردة ، أعد مفتاح التصحيح وقدر صدق المحتوى باستخدام طريقة المحكمين وحسب صدق المفردات باستخدام معامل تطابق المفردة بالهدف ، حيث تكونت عينة الدراسة من (٨١٨) طالباً وطالبة من طلاب الصف السابع بمحافظة القاهرة .

وقد جاءت نتائج الدراسة مصنفة حسب المتغيرات التي تم دراستها كما يلي :

١- بالنسبة لطول الاختبار :

– تستخدم طريقة هينا مع الاختبارات الطويلة ولأحجام صغيرة أو متوسطة من العينات ، كما تستخدم طريقة ليفنجستون للأحجام الكبيرة من العينات .

– بالنسبة للاختبارات متوسطة الطول تستخدم طريقة ليفنجستون وذلك لأحجام عينات صغيرة ومتوسطة ، ويفضل استخدام طريقة هينا مع حجم العينات الكبيرة .

– بالنسبة للاختبارات القصيرة يفضل استخدام طريقة صبكوفياك مع عينة صغيرة وطريقة ليفنجستون مع عينة متوسطة وطريقة هينا مع عينة كبيرة ، ومع الحجم الكلي للعينة تستخدم طريقة صبكوفياك .

٢- بالنسبة لحجم العينة :

– تستخدم طريقة ليفنجستون لأحجام العينات الصغيرة والمتوسطة وذلك للاختبارات ذات الأطوال من (٣٢ – ٩٦) مفردة وتستخدم طريقة هينا للاختبارات ذات الأطوال (١١٢ – ١٥٨) مفردة ، لحجم العينة نفسها .

– لحجم العينات الكبيرة (٥٠ – ٦٠) طالباً تستخدم طريقة هينا مع اختبارات ذات أطوال (٣٢ – ٩٦) مفردة ، وتستخدم طريقة ليفنجستون للاختبارات ذات الأطوال (١١٢ – ١٥٨) مفردة ، ولحجم العينة الكلي تستخدم طريقة صبكوفياك للاختبارات ذات الأطوال (٣٢ – ٩٦) مفردة ، وتستخدم طريقة هينا لباقي الأطوال من الاختبارات .

٣- بالنسبة لدرجة القطع :

- لدرجات القطع الصغيرة (١٠% ، ٢٠%) تستخدم أياً من الطرق الثلاث لأحجام العينات المختلفة مع الاختبارات ذات الأطوال القصيرة .

- لدرجات القطع من (٣٠% إلى ٥٠%) تستخدم طريقة ليفنجستون لجميع أحجام العينات ولأي طول من الاختبارات ، ولاختبارات ذات طول (٣٢) مفردة مع أحجام العينات (٣٠ ، ٤٠ ، ٥٠) طالباً تستخدم طريقة هينا أو صبكوفياك .

- لدرجة القطع (٦٠%) يفضل استخدم طريقة هينا ، ويمكن استخدم طريقة صبكوفياك للأحجام الكبيرة جداً أو القصيرة مع أطوال الاختبارات القصيرة .

- لدرجات قطع من (٧٠% إلى ٩٠%) تستخدم طريقة ليفنجستون لأي حجم ولأي طول وكذلك طريقة هينا ، كما تستخدم طريقة صبكوفياك عند طول (١١٢ - ١٥٨) مفردة لدرجة قطع (٩٠%) .

ثانياً : دراسات تناولت حساب معامل ثبات الاختبارات محكية المرجع وعلاقته ببعض المتغيرات :

١- دراسة ايجنور ورونالد وهامبلتون ورونالد (Eignor , Ronald-R , Hambelton , Ronald-K , 1978) :

الهدف من هذه الدراسة التوصل إلى علاقات بين طول الاختبار ، شكل توزيع درجات النطاق ، درجات التقدم ، ومؤشرات صدق وثبات متعددة للاختبار محكي المرجع

وقد تمت هذه الدراسة باستخدام طرق المحاكاة بالكمبيوتر حيث وضعت قيم متغيرات الدراسة لتكون بشكل نموذجي كالمستخدمة في العادة غالباً .

استخدمت مؤشرات الصدق والثبات (اتساق القرار ، كابا ، دقة القرار ، الصدق التنبؤي ، والفعالية) من بين أكثر المؤشرات فائدة لمطوري ومقيمي الاختبار محكي المرجع .

اقترحت موجهات عملية لاستخدام نتائج هذه الدراسة وذلك لتحديد أطوال اختبارات محكية المرجع مناسبة في حالات التقييم الدقيقة .

٢- دراسة ميلز وكريج وسيمون وروبرت (Mills , Craig , Simon , Robert , 1981) :

الهدف من هذه الدراسة تقديم نظام كمبيوتر سمي TESTLEN يسمح لمستخدمي الاختبار بتحديد أطوال الاختبار محكي المرجع المثلى بواسطة المحاكاة وفقاً لظروف المستخدم الخاصة مثل توزيع القدرة ، الخصائص الإحصائية للمفردة ، درجة القطع ، درجة التقدم ، نموذج الاختبار ، عدد المختبرين .

تقدم نتائج هذا النظام قيم اتساق القرار ، المعامل كابا ، ودقة القرار بالإضافة إلى المتوسط والانحراف المعياري لكل منها وتسجل نسبة مجموعة المختبرين المعينة في كل تصنيف إتقان .

٣- دراسة سشفير وماري وجروس وسوسان (Schafer , Mary , Gross , & Susan , 1983) :

الهدف من هذه الدراسة تحديد تأثر معاملات ثبات الاختبارات محكية المرجع التالية معامل كابا ، تقدير كابا ، معامل الاتفاق التي تعتمد تطبيق الاختبار مرة واحدة أو مرتين بكل من طول الاختبار ، طول الاختبار ودرجة القطع معاً ، قدرة الطالب ودرجة القطع معاً ، حجم العينة ، وتغيير خواص محتوى الاختبار .

كما اختبرت في هذه الدراسة العلاقات بين ثبات وصدق الاختبارات محكية المرجع وبين ثبات وصدق الاختبارات معيارية المرجع لنفس البيانات وذلك باستخدام نظام تعليمي في مادة الرياضيات .

طبقت ثلاثة اختبارات محكية المرجع على عينة مكونة من (٣٢٥) طالباً قبل وبعد أسبوعين من عطلة نصف العام .

وقد توصلت هذه الدراسة إلى ما يلي :

- معامل كابا وهو الوحيد الذي يمكن استخدامه عند تطبيق الاختبار مرتين له عموماً أقل القيم وأكبر الأخطاء المعيارية .
- أظهر تقدير معامل كابا متوسط قيم أكبر بقليل وأخطاء معيارية أقل .
- أعطى معامل الاتفاق أعلى معامل ثبات .

٤- دراسة نورمان (Norman , 1984) :

عرفت هذه الدراسة العلاقة بين طول الاختبار واتساق قرار تصنيف الاختبارات محكية المرجع بالاعتماد على مجموعة من العوامل كتجانس الاختبار ، توزيع مجال الدرجات ، ومستويات صعوبة المفردة .

حسبت معاملات الثبات باستخدام مؤشرات نسبة الاتفاق (P_o) ومعاملات كابا (K) .

وقد أشارت نتائج الدراسة إلى ما يلي :

- على عكس التوقعات المفترضة لم تزود الأهداف المبنية على (٢٠) مفردة بتحسينات عملية ومعنوية في اتساق القرار أكثر من الأهداف المبنية على (١٠) أهداف .

- كما تم افتراضه أظهرت فئة الأهداف المبنية على مفردة من نوع الاختيار من متعدد (٤ بدائل) تحسينات معتدلة ومعنوية في اتساق القرار أكثر من الأهداف المبنية على مفردة محددة .

- لم تظهر الأهداف المبنية على (١٠) مفردات تحسينات اتساق قرار معنوية أكثر من فئة الأهداف المبنية على مفردة من نوع الاختيار من متعدد (٤ بدائل) .

٥- دراسة سليمان محمد سليمان (١٩٨٦) :

الهدف من هذه الدراسة معرفة اختلاف ثبات الاختبارات محكية المرجع ذات الاختيار من متعدد باختلاف طرق تقدير الدرجات للمفردات وكذلك باختلاف طرق حساب هذا الثبات وفيما إذا كان هناك تأثير مشترك لكل من طرق تقدير الدرجات وطرق حساب الثبات على ثبات هذه الاختبارات .

تم في هذه الدراسة بناء صورتين متكافئين من اختبار تحصيلي محكي المرجع في مادة الجبر للصف الثالث الإعدادي في وحدتي " الأعداد الحقيقية والتحليل " .

طبقت الصورة (أ) من الاختبار على عينة استطلاعية تكونت من (١٨٠) طالباً وطالبة وبعد شهر ونصف طبقت الصورة (ب) من الاختبار وذلك بهدف تقنين الاختبار وحساب الزمن المناسب له ، وقد تكونت العينة النهائية من (٥٨٠) طالباً وطالبة موزعة على (١٣) مدرسة بمحافظة الشرقية . وقد تم حساب ثبات الاختبارات محكية المرجع باستخدام الطرق التالية : طريقة ليفنجستون ، طريقة هارس ، طريقة نسبة الاتفاق ، كما تم تقدير درجات الاختبار باستخدام الطرق التالية :

التقليدية ، أسلوب الدرجات الذاتية ، وزن الثقة (احتمال مقترح) ، وزن الثقة (لوغارتمية) ، توزيع الثقة الكروية .

وتوصلت هذه الدراسة إلى ما يلي :

– تؤثر الطرق المختلفة لتقدير الدرجات على ثبات الاختبار محكي المرجع المحسوب باستخدام الطرق المختلفة لحساب الثبات .

– تؤثر الطرق المختلفة لحساب ثبات الاختبار محكي المرجع على ثبات الاختبار حيث أعطت طريقة ليفنجستون أعلى معامل ثبات وذلك في طرق تقدير الدرجات المقترحة يليها طريقة هارس ثم طريقة نسبة الاتفاق .

– وجد تأثير مشترك لكل من طرق تقدير الدرجات وطرق حساب الثبات على ثبات الاختبارات محكية المرجع .

٦– دراسة هامبلتون (Hambelton , 1987) :

الهدف من هذه الدراسة تحديد الطول الأمثل للاختبار محكي المرجع في ضوء زمن محدد وذلك بأخذ أهمية صدق وثبات الدرجات التي تقيس الأهداف بعين الاعتبار ، حيث تكمن المشكلة في أنه عندما يثبت طول الاختبار فعلى مطور الاختبار أن يقرر فيما إذا كان من الأكثر فائدة قياس عدد قليل نسبياً من الأهداف بدقة أم قياس عدد أكبر منها ولكن بدقة أقل ؟

وقد تم في هذه الدراسة وضع الافتراضات التالية :

– توفر قائمة كبيرة صادقة من مفردات الاختبار لقياس كل هدف .

– يمكن تحديد زمن الاختبار المثبت في شكل العدد الكلي لمفردات الاختبار التي يمكن تطبيقها .

– تتطلب مفردات الاختبار مقادير متساوية من زمن التطبيق ، وبالتالي يمكن تطبيق نتائج البحث في اختبارات تستخدم نفس شكل المفردات ولها مجال محدد من مستوى الصعوبة .

– يقيس الاختبار أهداف متعددة .

– تستخدم بيانات أداء المختبر لتقدير درجة المجال .

– يمكن استخدام الارتباط بين تقديرات درجة المجال الناتجة عن قياس كل هدف ومحك مختار ملائم (معامل صدق الاختبار) كقياس تقويمي لتأثير الاختبار .

ضمن هذه الافتراضات سيكون طول الاختبار هو العامل الرئيسي المؤثر في صدق وثبات درجة المجال وذلك في ضوء بعض الاعتبارات مثل :

- بعض الأهداف أكثر أهمية من غيرها .
- بعض الأهداف أكثر صعوبة في القياس من غيرها .
- تستخدم الاختبارات محكية المرجع لقياس أهداف مختلفة في مستوى أهميتها .
- زمن الاختبار يجب أن يبقى منخفضاً (في معظم الحالات) .

ولتحديد طول الاختبار يجب تحديد عدد الأهداف والأهمية النسبية لكل منها ويمكن قبول مستويات معتدلة من صدق وثبات المفردة للأهداف التعليمية الصفية الأقل أهمية ، في حين أن مستويات أعلى بكثير ستكون مطلوبة للأهداف التعليمية الأكثر أهمية .

وكتطبيق للخوارزمية المقترحة أعد جدول يوضح طول الاختبار المناسب عند تغيير العوامل الأخرى وقدمت في البداية اختبارات متساوية الطول :

ظهر في التطبيق الأول تأثير الأهمية النسبية للهدف على اختبارات متساوية الصدق والثبات ، حيث نتج الحل الأمثل من خلال زيادة طول الاختبار ذي الأهداف الأكثر أهمية بمقدار (٦) مفردات عن الاختبار ذي الأهداف الأقل أهمية .

وضح التطبيق الثاني دور ثبات الاختبار في الحلول المثلى حيث كانت عوامل أخرى متساوية (كطول الاختبار ، الصدق ، الأهمية النسبية) ، في الحل الأمثل المصمم لتعزيز ثبات الاختبار كانت الاختبارات ذات عوامل الثبات الأعلى أقصر بينما زاد طول الاختبارات ذات عوامل الثبات الأقل .

وضح التطبيق الثالث تأثير معاملات صدق مختلفة على الحل الأمثل حيث تساوت العوامل الأخرى ، وقد طولت الاختبارات الأكثر صدقاً في الحل الأمثل .

أما في التطبيق الرابع فقد تنوعت جميع العوامل (الأهمية النسبية ، الثبات ، الصدق) فقد كانت الاختبارات متساوية الطول في البداية (٢٠) مفردة ثم تراوحت أطوالها في الحل الأمثل من (١٣ – ٣١) مفردة .

وأخيراً فإن الحل المقترح في هذه الدراسة لتحديد زمن الاختبار سهل التطبيق وممكن عندما تكون بعض البيانات الأساسية كالصدق والثبات متاحة .

٧- دراسة لين (Lin , 1988) :

الهدف من هذه الدراسة مقارنة ثلاث طرق لاختيار أفضل مفردات اختبار مميزة ونتيجة ذلك على ثبات تصنيفات الإتقان وعدم الإتقان .

تمت مقارنة الطرق التالية : معامل الاتفاق ، معامل فاي ، وطريقة الاختيار العشوائي .

استخدمت استجابات اختبار علوم فيزيائية مكون من (٥٠) مفردة ، ثلاثة أحجام مختلفة من العينات (٧٥ ، ١٥٠ ، ٣٠٠) ، ثلاثة أطوال مختلفة من الاختبار (١٥ ، ٢٥ ، ٣٥) وذلك على (١٨٣٦) طالباً .

أشارت نتائج هذه الدراسة إلى ما يلي :

- طريقة الاتفاق مناسبة للاستخدام لاختيار مفردات اختبار محكي المرجع في مستوى الصف الدراسي .

- طريقة فاي مناسبة للاستخدام لاختيار مفردات اختبار محكي المرجع في مستويات الولاية .

- لم يكن لطريقة الاختيار العشوائي خصائص مشابهة في اختيار مفردات الاختبار وأعطت أقل معاملات ثبات عندما قورنت بطريقتي الاتفاق ومعامل فاي .

٨- دراسة بوزيل (Pozehl , 1990) :

الهدف الأول لهذه الدراسة هو التحقق من تأثيرات اختيار النموذج وخفض طول الاختبار من خلال الطرق النموذجية لاختيار المفردة على ثبات الاختبار محكي المرجع .

أما الهدف الثاني فهو التحقق من تأثيرات اختيار النموذج ، حجم العينة ، وخفض طول الاختبار على دقة التقدير عند تطبيق نظرية الاستجابة للمفردة .

استخدمت أحجام عينات مكونة من (٢٥٠ ، ٥٠٠) مختبر وأطوال اختبار (٥٠ ، ١٠٠) مفردة لدراسة التأثيرات على دقة التقدير مع خمسة نماذج مختلفة .

أشارت نتائج هذه الدراسة إلى ما يلي :

- أظهر نموذج أداء عينة المختبرين كاملة أن نماذج ثنائية وثلاثية البارامتر قد زودت بأفضل ملاءمة بيانات _ نموذج خلال أطوال الاختبار الثلاثة .

- أظهرت مقارنة دقة التقدير مع نماذج نظرية الاستجابة للمفردة في ظروف خفض حجم العينة وطول الاختبار بأن نماذج البارامتر الواحد والبارامتر الواحد المعدل قد زودا بتقديرات أكثر دقة من النموذجين ثنائي وثلاثي البارامتر .

- أعطت معظم النماذج أفضل النتائج مع العينات الأكبر عموماً ، بينما أعطى النموذجان أحادي البارامتر وأحادي البارامتر المعدل أفضل النتائج مع العينات الأصغر .

٩- دراسة كالون (Kalohn , 1992) :

الهدف من هذه الدراسة اختبار تأثيرات شكل توزيع درجات الاختبار ، حجم العينة ، معاملات الثبات التقليدية ، وموضع درجة القطع على تعيين خصائص كل من : نسبة الاتفاق ، معامل كابا ، كابا المصححة ، وفاي .

استخدم أسلوب المحاكاة بالكمبيوتر للتحقق من تأثيرات كل من هذه العوامل ، حيث قيِّم مجموع ثمانية توزيعات ، أربع أحجام عينات (٣٠ ، ٦٠ ، ١٢٠ ، ٢٤٠) ، خمس درجات قطع ، ثلاثة معاملات ثبات (ر = ٠,٧٥ ، ٠,٨٥ ، ٠,٩٥) .

وقد توصلت هذه الدراسة إلى ما يلي :

- تأثر قيم البارامترات في جميع الإحصاءات بشكل التوزيع ، الثبات ، ودرجة القطع

- أعطى فاي تقديرات أقل تحيزاً من معامل كابا ، كما كان لتوضع درجة القطع وشكل التوزيع التأثير الأكبر في التحيز لتقدير كابا وفاي فقد زودت درجات القطع الموجودة قرب منتصف التوزيع بتقديرات غير متحيزة لكل من كابا وفاي وأنه كلما اقتربت درجة القطع من أطراف التوزيع وكانت كثافة الدرجات قليلة زاد التحيز .

- أظهرت الزيادة في حجم العينة نقص تحيز معنوي خصوصاً عندما وجدت درجة القطع عند أطراف التوزيع ، وربما انخفض تأثير حجم العينة عندما وجدت درجة القطع عند منتصف التوزيع .

- قدمت نسبة الاتفاق وكابا المصححة تقديرات غير متحيزة في قيم البارامترات في كافة الظروف .

١٠- دراسة إسماعيل الوليلي (١٩٩٦) :

الهدف من هذه الدراسة التعرف على أثر استخدام بعض طرق التغذية المرتدة الاختبارية على ثبات درجات اختبار هدفي المرجع وعلى استبقاء المعلومات ، حيث تمت دراسة التغذية المرتدة بمستوياتها الثلاثة : كلية - جزئية - ومؤجلة .

استخدم في هذه الدراسة اختبار تحصيلي هدفي المرجع في حساب المثلثات وحدة " الدوال المثلثية " تكون من صورتين متكافئتين عدد مفردات كل منهما (٣٢) مفردة .

ثم أجريت تجربة استطلاعية للاختبار لتحديد الزمن المناسب لإجابة مفردات كل من صورتي الاختبار ، وقد تم التحقق من صدق محتوى الاختبار باستخدام طريقة المحكمين وقدر ثبات الدرجات بطريقتين : إحداهما معامل كابا لسوامينثان وهامبلتون وألجينا عند درجة قطع (٧٨,١٣%) محددة بطريقة أنجوف حيث بلغت قيمة هذا المعامل (٠,٨١) ، والطريقة الثانية هي معامل التكافؤ ممثلاً بمعامل ارتباط بيرسون بين درجات الطلاب في مرتي التطبيق حيث بلغت (٠,٩٣) وهي دالة إحصائياً عند مستوى (٠,٠١) ، كما حسب معامل صعوبة المفردات ومعامل تمييزها .

واختيرت عينة عشوائية مكونة من (٢٨٥) طالباً موزعة على ثمانية فصول من طلاب الصف الأول الثانوي بمحافظة الشرقية .

وقد توصلت هذه الدراسة إلى ما يلي :

- أدت التغذية المرتدة الكلية والمؤجلة في الاختبار إلى زيادة ثبات درجاته عن حالة التغذية الجزئية أو عدم وجود تغذية مرتدة .

- قيمة معامل الثبات في حالة التغذية المرتدة الكلية أعلى منه في حالة التغذية المرتدة المؤجلة .

١١- دراسة لي (Lee , 1996) :

الهدف من هذه الدراسة معرفة مدى حساسية ثبات الاختبار محكي المرجع بالنسبة لموضع درجة القطع ، فبالاعتماد على ما هو معروف بأن الثبات يتأثر بموضع درجة القطع تم افتراض أنه كلما اقتربت درجة القطع من منتصف التوزيع سيقل ثبات الاختبار بصرف النظر عن نوع التوزيع .

تم تغيير موضع درجة القطع إلى عشرة مواضع مختلفة اعتباراً من منتصف التوزيع وباتجاه الأطراف بهدف معرفة أثر ذلك على قيم معاملات الثبات وذلك لكل اختبار ومع كل نوع توزيع ، ثم استخدم معامل كابا لحساب قيم معاملات الثبات .

وقد توصلت هذه الدراسة إلى تأثر ثبات كل اختبار بموضع درجة القطع على الرغم من نشوء علاقة عكسية لما تم افتراضه ، فقد ازداد المعامل كابا كلما اقتربت درجة القطع من منتصف التوزيع لذلك تم التوصية بإجراء بحث تجريبي في هذا النطاق للتحقق من عوامل أخرى تؤثر في الثبات .

١٢- دراسة شحته عبد المولى (١٩٩٩) :

الهدف من هذه الدراسة الإجابة على الأسئلة المتعلقة بالثبات ودقة القياس ومدى الخطأ المعياري للقياس وذلك لكل من الاختبارات محكية المرجع ومعيارية المرجع ، وإلى التعرف على أهم المتغيرات التي تؤثر على ثبات ودقة الاختبارات والموازنة بين اتجاهين من اتجاهات القياس (التقليدي والمعاصر) ، بالإضافة إلى التحقق من إمكانية تطبيق طرق نظرية الاستجابة للمفردة في تحليل وبناء اختبارات على درجة عالية من الدقة والموضوعية في مجال القياس محكي / معياري المرجع وكذلك المقارنة بين أربع طرق لتحليل الاختبارات محكية ومعيارية المرجع .

كما حاولت هذه الدراسة معرفة إلى أي حد تختلف قيم معاملات ثبات الاختبارات محكية / معيارية المرجع (الطريقة التقليدية ، النموذج أحادي البارامتر ، ثنائي البارامتر ، ثلاثي البارامتر) .

ولتوليد البيانات المتطلبة لإجراءات هذه الدراسة استخدام أسلوب المحاكاة عن طريق الكمبيوتر مثل أسلوب مونت كارلو ، وبذلك تم الحصول على (٢٠٠) مجموعة اشتملت كل منها على عينة من الأفراد عددها (١٠٠٠) فرد وعدد من المفردات يصل إلى (١٠٠) مفردة وجميع هذه العينات تتميز بالصفة الثنائية (صفر ، واحد) .

حسبت معاملات ثبات الاختبارات معيارية المرجع باستخدام معامل كيودر ريتشاردسون (٢٠) ، في حين حسبت معاملات ثبات الاختبارات محكية المرجع باستخدام طريقة ليفنجستون عند درجة قطع (٨٠%) .

أشارت نتائج هذه الدراسة إلى ما يلي :

- أهمية نظرية الاستجابة للمفردة وما تنطوي عليه من نماذج رياضية في بناء الاختبارات معيارية المرجع والاختبارات محكية المرجع وذلك لما قدمته من دفعة في قيم معاملات الثبات ودقة تقدير القدرة وانخفاض الخطأ المعياري لتقدير القدرة وزيادة القدرة من خلال معلومات الاختبار بدرجة أفضل من طرق تحليل المفردة التقليدية .

١٣- دراسة سعاد حسانين (٢٠٠٠) :

الهدف من هذه الدراسة التعرف على مدى فعالية نوع المفردة وبعض طرق اختيارها على ثبات الاختبارات محكية المرجع ، حيث حاولت الإجابة على الأسئلة التالية:

١- هل يختلف معامل ثبات الاختبار محكي المرجع باختلاف طرق اختيار المفردة؟

٢- هل يختلف معامل ثبات الاختبار محكي المرجع باختلاف نوع المفردة ؟

٣- هل يختلف معامل ثبات الاختبار محكي المرجع باختلاف تغيير المفردات ، باختلاف نوع المفردة ؟

٤- هل يختلف معامل ثبات الاختبار محكي المرجع باختلاف طرق حساب الثبات ؟

استخدمت أنواع المفردات : الاختيار من متعدد ، أسئلة الصواب والخطأ ، وأسئلة المقال .

وفي طرق اختيار المفردة استخدمت : قوة تمييز المفردة ، معامل صعوبة المفردة ، معامل الفروق القبلية - البعدية ، ومعامل برينان .

بينما حسبت معاملات الثبات باستخدام الطرق التالية : طريقة هارس ، طريقة هامبلتون ، وطريقة ليفنجستون .

تم بعد ذلك بناء ثلاثة اختبارات تحصيلية في باب " العمليات المعرفية " من مادة علم النفس للصف الثاني الثانوي العام وذلك في أنواع المفردات الثلاثة المختارة ، اعتمدت درجة القطع (٨٠%) ، وحسب صدق مفردات الاختبار باستخدام طريقة المحكمين .

تكونت عينة الدراسة من (٢٨٠) طالباً وطالبة من طلاب الصف الثاني الثانوي .

وقد أشارت نتائج هذه الدراسة إلى ما يلي :

- حقق اختبار الصواب / الخطأ أعلى معامل ثبات يليه اختبار المقال ثم اختبار الاختيار من متعدد عند حساب الثبات باستخدام طريقة هارس وليفنجستون .

- لم يختلف معامل ثبات الاختبار محكي المرجع باختلاف نوع المفردة عند حساب الثبات باستخدام طريقة هامبلتون وألجينا .

- وجدت اختلافات في حال تغيير ترتيب المفردات واختلاف نوع المفردة وذلك عند حساب الثبات باستخدام طريقة هارس وليفنجستون .

- تؤثر الطرق المختلفة لحساب ثبات الاختبار محكي المرجع على ثبات الاختبار حيث :

* تتفوق طريقة ليفنجستون على طريقة هارس ، هامبلتون وألجينا .

* تتفوق طريقة هامبلتون وألجينا على طريقة هارس .

- وأخيراً بيَّنت نتائج هذه الدراسة إمكانية استخدام الطرق التقليدية لتحليل المفردة معيارية المرجع في تحليل مفردات الاختبار محكي المرجع حيث أعطت أعلى معامل ثبات الاختبارات المكونة من مفردات مختارة بطرق أخرى في اختبار الصواب والخطأ ، الاختيار من متعدد عند حساب الثبات باستخدام طريقة هامبلتون وألجينا ، وفي اختبار الصواب والخطأ عند حساب الثبات باستخدام طريقة هارس .

تعقيب عام على الدراسات السابقة :

يلاحظ من العرض السابق للدراسات التي تمت في هذا المجال :

١- استخدم عدد من هذه الدراسات طريقة مواصفات الاختبار المقترحة من قبل بابام Popham في بناء أداة الدراسة كدراسة (آمال محروس ، ١٩٨٨ ؛ محمود إبراهيم ، ١٩٩٠ ؛ إسماعيل الوليلي ، ١٩٩٦ ؛ سعاد حسانين ، ٢٠٠٠) .

٢- أجمعت الدراسات التي اعتمدت طريقة مواصفات الاختبار في بناء اختبارات محكية المرجع على استخدام طريقة الصدق الوصفي حيث استخدمت جميعها تقديرات المحكمين في تقدير صدق مفردات الاختبار وصدق المحتوى .

٣- اعتمدت معظم الدراسات على مفردات من نوع الاختيار من متعدد .

٤- قامت بعض الدراسات التي تناولت مقارنة بعض طرق حساب معامل ثبات الاختبارات محكية المرجع على مقارنة بعض الطرق التي تتطلب تطبيق الاختبار مرة واحدة كدراسة (محمود إبراهيم ، ١٩٩٠ ؛ Hagen , 1983) .

في حين قارنت دراسات أخرى بعض الطرق التي تتطلب تطبيق الاختبار مرة واحدة مع واحدة أو اثنتين من الطرق التي تتطلب تطبيق الاختبار مرتين كدراسة (آمال محروس ، ١٩٨٨ ؛ Subkoviack , 1976 ; Subkoviack , 1980 ; Schafer et al. , 1983 ; Wang , 1983) .

٥- قامت بعض الدراسات التي تناولت حساب معامل ثبات الاختبارات محكية المرجع وعلاقته ببعض المتغيرات بدراسة إحدى هذه الطرق في علاقتها ببعض المتغيرات كدراسة (شحته عبد المولى ، ١٩٩٩ ؛ Lin , 1988 ; Lee , 1996) .

في حين اهتمت دراسات أخرى بدراسة أكثر من طريقة في علاقتها ببعض المتغيرات كدراسة (سليمان سليمان ، ١٩٨٦ ؛ إسماعيل الوليلي ، ١٩٩٦ ؛ سعاد حسانين ، ٢٠٠٠ ؛ Schafer et al . , 1983 ; Norman , 1984 ; Kalohn , 1992) .

٦- قارنت بعض هذه الدراسات معاملات ثبات الاختبارات محكية المرجع مع معاملات ثبات الاختبارات معيارية المرجع كدراسة (شحته عبد المولى ، ١٩٩٩ ؛ Schafer et al. , 1983 ; Schafeer , 1983) .

٧- من أكثر المتغيرات التي تمت دراسة علاقتها بمعاملات ثبات الاختبارات محكية المرجع طول الاختبار ، درجة القطع ، حجم العينة .

في حين اهتمت بعض الدراسات بدراسة متغيرات أخرى إلى جانب بعض من هذه المتغيرات كدراسة (Eignor et al. , 1978) التي درست شكل توزيع الدرجات ، دراسة (Schafer et al. , 1983) التي اهتمت بدراسة قدرة الطالب وخواص محتوى الاختبار ، دراسة (سليمان سليمان ، ١٩٨٨) التي اهتمت بدراسة طرق تقدير درجات المفردات ، دراسة (Lin , 1988) التي اهتمت بدراسة طرق اختيار المفردات ، دراسة (إسماعيل الوليلي ، ١٩٩٦) التي اهتمت بدراسة طرق التغذية ، ودراسة (سعاد حسانين ، ٢٠٠٠) التي اهتمت بدراسة نوع المفردات وطرق اختيارها وترتيبها .

٨- استخدمت بعض هذه الدراسات بيانات فعلية ناتجة عن تطبيق اختبارات محكية المرجع على عينة من الطلاب كدراسة (سليمان سليمان ، ١٩٨٨ ؛ آمال محروس ، ١٩٨٨؛ محمود إبراهيم ، ١٩٩٠ ؛ إسماعيل الوليلي ، ١٩٩٦ ؛ سعاد حسانين ، ٢٠٠٠ ؛ Subkoviack , 1980 ; Schafer et al. , 1983 ; Schafeer , 1983 ; Lin , 1988 ; Pozehl , 1990) .

في حين استخدمت دراسات أخرى بيانات عن طريق المحاكاة بالكمبيوتر كدراسة (شحته عبد المولى ، ١٩٩٩ ؛ Eignor et al. , 1978 ; Wang , 1983 ; Norman , 1984 ; Dutschke , 1988 ; Kalohn , 1992) .

وقد وجدت دراسة (Hagen , 1983) التي جمعت بين الطرفين حيث تضمنت دراستين استخدمت الأولى بيانات محاكاة بينما استخدمت الثانية بيانات ناتجة عن تطبيقات فعلية للاختبار .

٩- ظهر تناقض بين بعض نتائج هذه الدراسات فقد توصلت دراسة (Subkoviak , 1980) إلى أن تقديرات كابا غير متحيزة نهائياً ، في حين أشارت دراسة (Kalohon , 1992) إلى تحيز طريقة معامل كابا وأنه كلما اقتربت درجات القطع من منتصف التوزيع قل هذا التحيز . ومن هنا يتضح أنه لم يتم التوصل لرأي قاطع حول تحيز هذه الطريقة مما يترك الباب مفتوحاً لعمل مزيد من الدراسة .

هذا وقد حاولت الدراسة الحالية الاستفادة من الدراسات السابقة التي تم عرضها حيث قامت على ما يلي :

١- كون الدراسات السابقة لم تقارن طرق حساب معامل ثبات الاختبارات محكية المرجع التي تتطلب تطبيق الاختبار مرتين بشكل منفصل عن الطرق التي تتطلب تطبيق الاختبار مرة واحدة ، لذا تحاول الدراسة الحالية مقارنة بعضاً منها حيث تقوم على مقارنة الطرق التالية :

طريقة كارفر ، طريقة نسبة الاتفاق ، وطريقة معامل كابا .

٢- تهتم الدراسة الحالية بدراسة أكثر المتغيرات أثراً على قيم معاملات ثبات الاختبارات محكية المرجع وهذه المتغيرات هي :

طول الاختبار ، درجة القطع ، وحجم العينة .

٣- تقوم الدراسة الحالية باستخدام بيانات فعلية ناتجة عن تطبيق ثلاثة اختبارات تحصيلية محكية المرجع في وحدة " مجموعة الأعداد الطبيعية " لطلاب الصف الأول الإعدادي مبنية باستخدام طريقة مواصفات الاختبار المقترحة من قبل بابام ، وتستخدم مفردات من نوع الاختيار من متعدد .

٤- يتم حساب صدق مفردات الاختبارات باستخدام تقديرات المحكمين .

٥- تستفيد الدراسة الحالية في وضع فروض البحث من الدراسات السابقة التي أكدت ما يلي :

- تزيد قيم معاملات ثبات الاختبارات محكية المرجع بزيادة طول الاختبار.
- تزيد قيم معاملات ثبات الاختبارات محكية المرجع كلما اقتربت درجة القطع من أطراف التوزيع ، ما عدا في طريقة معامل كابا حيث أكدت عدة دراسات زيادة قيم معامل الثبات المحسوب باستخدام هذه الطريقة مع درجات قطع عند منتصف توزيع الدرجات .
- تزيد قيم معاملات ثبات الاختبارات محكية المرجع بزيادة حجم العينة .

بناءً على العرض السابق تحاول الدراسة الحالية اختبار الفروض التالية :

فروض الدراسة :

أ- بالنسبة لطول الاختبار :

١- تزيد قيمة معامل الثبات المحسوب باستخدام طريقة كارفر كلما زاد طول الاختبار

٢- تزيد قيمة معامل الثبات المحسوب باستخدام طريقة نسبة الاتفاق كلما زاد طول الاختبار

٣- تزيد قيمة معامل الثبات المحسوب باستخدام طريقة كابا كلما زاد طول الاختبار.

ب- بالنسبة لدرجة القطع :

١- تزيد قيمة معامل الثبات المحسوب باستخدام طريقة كارفر كلما اقتربت درجة القطع من أطراف التوزيع .

٢- تزيد قيمة معامل الثبات المحسوب باستخدام طريقة نسبة الاتفاق كلما اقتربت درجة القطع من أطراف التوزيع .

٣- تزيد قيمة معامل الثبات المحسوب باستخدام طريقة معامل كابا كلما اقتربت درجة القطع من منتصف التوزيع .

ج- بالنسبة لحجم العينة :

١- تزيد قيمة معامل الثبات المحسوب باستخدام طريقة كارفر كلما زاد حجم العينة.

٢- تزيد قيمة معامل الثبات المحسوب باستخدام طريقة نسبة الاتفاق كلما زاد حجم العينة .

٣- تزيد قيمة معامل الثبات المحسوب باستخدام طريقة معامل كابا كلما زاد حجم العينة .

د - بالنسبة للطريقة :

١- توجد فروق دالة إحصائياً بين قيمة معامل الثبات المحسوب باستخدام طريقة كارفر وقيمة معامل الثبات المحسوب باستخدام طريقة نسبة الاتفاق .

٢- توجد فروق دالة إحصائياً بين قيمة معامل الثبات المحسوب باستخدام طريقة نسبة الاتفاق وقيمة معامل الثبات المحسوب باستخدام طريقة معامل كابا .

٣- توجد فروق دالة إحصائياً بين قيمة معامل الثبات المحسوب باستخدام طريقة كارفر وقيمة معامل الثبات المحسوب باستخدام طريقة معامل كابا .