الفصل الثالث

الدراسات السابقة

مقـــــــــدمة.

المحور الأول: الدراسات والبحوث التي تناولت معالجة أثر عدد بدائل المفردة في أسئلة الاختيار من متعدد على قيمة معامل الثبات.

- تعقيب الباحثة على البحوث والدراسات المرتبطة بالمحور الأول.

المحور الثاني: البحوث والدراسات التي اهتمت بمعالجة أثر موقع وترتيب مفردات الاختبار على قيمة معامل الثبات.

- تعقيب الباحثة على البحوث والدراسات المرتبطة بالمحور الثاني.

المحور الثالث: البحوث والدراسات التي تناولت أثر متغير طول الاختبار (عدد المفردات) على قيمة معامل الثبات.

- تعقيب الباحثة على البحوث والدراسات المرتبطة بالمحور الثالث.

المحور الرابع: البحوث والدراسات التي تناولت معالجة أثر حجم العينة على قيمة معامل الثبات.

- تعقيب الباحثة على البحوث والدراسات المرتبطة بالمحور الرابع.
- تعليق عام على الدراسات السابقة.

فروض الدراسة.

الفصل الثالث

الدراسات السابقة

مقدمة:

تعرض الباحثة في هذا الفصل للدراسات والبحوث ذات العلاقة بمتغيرات الدراسة الحالية في إطار تصنيفي وفقاً للمحاور الآتية:

المحور الأول: الدراسات والبحوث التي تناولت معالجة أثر عدد بدائل المفردة في أسئلة الاختيار من متعدد على قيمة معامل الثبات:

دراسة ساكس Sax, 1980:

بعنوان "المقارنة بين معاملات صدق وثبات أسئلة الاختيار من متعدد ذات البديلين وأسئلة الصواب والخطأ". وقد تمثل هدف الدراسة فى المقارنة بين معاملات الثبات والصدق الخاصة بأسئلة الاختيار من متعدد وأسئلة الصواب والخطأ عندما يكونان متطابقين، فالفرق بينهما فقط أن أسئلة الاختيار من متعدد تقدم للطلاب في صورة جذر متبوع بعدد من البدائل أحدهما البديل الصواب والبدائل الأخرى خطأ، بينما أسئلة الصواب والخطأ، أحد هذه البدائل موضوع في رأس السؤال (صواب ــ خطأ).

وقد قام الباحث باختيار مجموعة من أسئلة الاختيار من متعدد تم تجريبها سابقاً باستخدام معامل الارتباط الثنائي الأصيل، واختبار قوة تشتيت بدائلها، ثم قام بتحويلها إلى أسئلة صواب وخطأ طبقت على عينة من الطلاب بجامعة واشنطن.

وباستخدام أسلوب تحليل التباين الأحادي أظهرت النتائج ما يلي:

١- وجود فروق دالة إحصائياً بين أسئلة الاختيار من متعدد وأسئلة الصواب والخطأ عند مستوى دلالة (٠,٠١) لصالح أسئلة الاختيار من متعدد ذات البديلين فيما يتعلق بمتوسطات الأداء الاختباري للطلاب.

٢- لا توجد فروق دالة بين كلا النوعين من الأسئلة في معاملات الثبات باستخدام معادلة كيودر وريتشاردسون (٢٠).

٣- لا توجد فروق دالة إحصائياً في معاملات الصدق بين كلا النوعين من الأسئلة، باستخدام معامل الارتباط.

دراسة ستراتون وكاتس Straton & Catts, 1980:

بعنوان "المقارنة بين استخدام أعداد مختلفة من بدائل مفردات الاختبار (٤/٣/٢) بدائل للمفردة بمعلومية عدد ثابت من البدائل" وقد تمثلت أهداف الدراسة فى الآتى:

١- المقارنة بين الخصائص السيكومترية لمفردات اختبارات الاختيار من متعدد التي تصمم وتطبق تحت شرط Tversky (1964)، وهو وجود عدد كلي ثابت من البدائل في جميع صور الاختبارات التي نقارن فيما بينها، بحيث يكون الاختلاف بينها فقط في عدد بدائل الفقرة في كل صورة من هذه الصور الاختبارية.

٢- الاستفادة من شرط Tversky (1964)، كأساس للمقارنة بين الاختبارات ذات الأعداد المختلفة من البدائل لكل مفردة وذلك فيما يتعلق بمتوسط الوقت المستغرق في الإجابة على المفردات ذات الأعداد المختلفة من البدائل.

تألفت عينة الدراسة من (٢٦٠) طالب من الطلاب المقيدين لشهادة المدرسة الثانوية، وهؤلاء الطلاب كانوا يعدون للامتحان الأساسي، والذي تم استخدامه كأساس للانتقاء لدخول الكليات والجامعات في ولاية ساوس ويلز (New South Wales).

استخدم الباحثان اختبار اختيار من متعدد يتألف من (٦٠) مفردة بأربعة بدائل في مادة الاقتصاد (المستوى الثاني)، حيث تم الاعتماد على هذه المفردات في إعداد أربعة صور اختبارية، يشتمل كلاً منها على عدد كلي ثابت من البدائل (١٢٠) بديل وبيانها كالتالي:

١- الصورة (A) تتألف من (٦٠) مفردة ببديلين لكل مفردة.

٢- الصورة (B) تتألف من (٤٠) مفردة بثلاثة بدائل لكل مفردة.

٣- الصورة (C) تتألف من (٤٠) مفردة بثلاثة بدائل لكل مفردة.

٤- الصورة (D) تتألف من (٣٠) مفردة بأربعة بدائل.

وقد أوضح الباحثان أن الاختلاف بين الصورة (B)، (C) تمثل في أسلوب حذف أحد المشتتات من الاختبار الأصلي المتضمن أربعة بدائل، فعند إعداد الصورة (B) بثلاثة بدائل تم تجاهل أحد المشتتات من جميع المفردات بأسلوب عشوائي، بينما في الصورة (C) تم تجاهل أحد المشتتات وفقاً لآراء مجموعة من المعلمين والتي اعتبرت أقل جاذبية من وجهة نظرهم.

وقد تم تعيين صور الاختبارات الأربعة على الطلاب بأسلوب عشوائي وتم إرشادهم لضرورة تسجيل وقت بدايتهم وانتهائهم من الإجابة على الاختبار في أوراق إجاباتهم.

وقد توصلت الدراسة إلى النتائج التالية:

١- فيما يتعلق بأثر عدد البدائل لكل مفردة على معاملات التمييز أوضحت النتائج أن صورة الاختبار بأربعة بدائل حققت أعلى معاملات للتمييز (الصورة D)، وذلك بالمقارنة بالصورة (A) والتي تضمنت بديلين لكل مفردة، حيث حققت أقل معاملات تمييز.

٢- فيما يتعلق بمعاملات الثبات، وجد أعلى تقدير لمعاملات الثبات في الصورة (B) والتي تضمنت ثلاثة بدائل، وقيمته (٠,٧٣) وذلك بالمقارنة بالصورة (C, D) حيث حققت قيم أقل وهي (٠,٦٨)، (٠,٦٦) على التوالي، وباختبار الفروق بينهما إحصائياً وجد أنها غير دالة، ولكن بمقارنة معامل الثبات الخاص بالصورة (A) ببديلين لكل مفردة وقيمته (٠,٤٧)، وجد أن الفروق الإحصائية بينه وبين المعاملات الخاصة بالصور الثلاثة الأخرى دالة إحصائياً عند مستوى (٠,٠٥).

٣- الأخطاء المعيارية للقياس المحسوبة من الصورة (D) بأربعة بدائل لكل مفردة كانت أقل بالمقارنة بصور الاختبار الأخرى وخاصة الصورة (A) ببديلين لكل مفردة حيث حققت أعلى قيمة لأخطاء القياس المعيارية.

٤- متوسط وقت إتمام الاختبار لا يتناسب مباشرة مع عدد البدائل حيث ازداد بزيادة عدد المفردات، بينما متوسط الوقت لكل مفردة انخفض بانخفاض عدد البدائل.

٥- أوضحت النتائج أن أسلوب الاستعانة بالمعلمين الخبراء في المادة عند تجاهل أحد المشتتات من الاختبار الأصلي المتضمن (٤) بدائل، وذلك لإعداد صورة أخرى بثلاثة بدائل، حقق تقديرات للثبات أقل، بالمقارنة بالأسلوب العشوائي في تجاهل أحد المشتتات، وعلى الرغم من ذلك كانت الأخطاء المعيارية الخاصة بالأسلوب الأول أقل.

ويؤكد الباحثان في نهاية الدراسة على فاعلية استخدام المعلمين في المدارس والجامعات لصور اختبارات الاختيار من متعدد بثلاثة بدائل وذلك بالمقارنة بأعداد البدائل الأخرى (٢، ٤، ٥).

دراسة جرين وآخرون Green & et al., 1982:

بعنوان "صدق وثبات الاختبارات بأعداد مختلفة من بدائل المفردة للطلاب ذوى المستويات المتباينة من القدرة". وقد هدفت الدراسة إلى مقارنة صدق وثبات اختبارات الاختيار من متعدد ذات البدائل (٣، ٤، ٥) بالنسبة للطلاب ذوي المستويات المتباينة من القدرة (منخفضة ــ متوسطة ــ مرتفعة).

تألفت عينة الدراسة من (١٩٣) طالب من جامعة واشنطن.

وقد تمثلت إجراءات الدراسة فيما يلي:

١- تطبيق اختبار اختيار من متعدد يتألف من (٤٠) مفردة، بعد حذف المشتتات غير الجيدة بزمن قدره (٤٠) دقيقة.

٢- تم تصنيف الطلاب إلى ثلاثة مجموعات من حيث القدرة (منخفضة ــ متوسطة ــ مرتفعة) بالاعتماد على درجات القطع، ولم يكن لعامل السرعة أي تأثير على تطبيق الاختبار.

٣- استخدمت درجات المقدر كمحك خارجي لحساب معامل الصدق للاختبار.

٤- تم حساب المتوسطات والانحرافات المعيارية ومعاملات الارتباط للمفردات لحساب الصدق، وكذلك حسبت معاملات الثبات بتطبيق معادلة كيودر وريتشاردسون (٢٠) ومعادلة براون للتجزئة النصفية ومعادلة فيلدت باستخدام الحاسب الآلي، وتم استخدام أسلوب تحليل التباين الثنائي لمعرفة الفروق في مجموعات القدرة.

وقد أوضحت نتائج الدراسة ما يلي:

١- وجود فروق دالة عند مستوى (٠,٠٥) في الثبات بالنسبة للمجموعة منخفضة القدرة، حيث انخفض معامل الثبات مع الزيادة في عدد البدائل، وكانت الفروق لصالح البدائل الخمسة.

٢- وجود فروق دالة إحصائياً في مجموعة القدرة المرتفعة في اختبار الاختيار من متعدد (٣/ ٤/ ٥) بدائل لصالح البدائل (٥).

٣- لا توجد فروق دالة في معاملات الثبات بالنسبة للمجموعة متوسطة القدرة عند مستوى (٠,٠٥).

٤- وجدت فروق دالة إحصائياً بين معاملات الثبات الخاصة بمجموعة القدرة المشتركة (المجموعات الثلاثة ككل) لأسئلة الاختيار من متعدد بالبدائل (٣/ ٤/ ٥) لصالح البدائل (٤).

٥- لم تختلف النتائج الخاصة بمعاملات الصدق عن نتائج معاملات الثبات.

دراسة بوديسكو ونيفو Budescu & Nevo, 1985:

بعنوان "العدد الأمثل للبدائل: اختبار فرض التناسب". وقد تمثل الهدف الرئيسى للدراسة فى مجموعة من الأهداف الفرعية وهى:

١- اختبار العدد الأمثل لبدائل الفقرة لمفردات اختبار الاختيار من متعدد.

٢- اختبار صدق فرض التناسب (تناسب الزمن الكلي للاختبار مع عدد المفردات وعدد البدائل لكل مفردة).

٣- اختبار صحة الفرض القائل، أن كل المشتتات تتساوى في جاذبيتها، وأنه يمكن الحصول على الاختبارات المناسبة بأعداد مختلفة من البدائل عن طريق حذف المشتتات بشكل تحكمي (غير منطقي) أو بأسلوب المعاينة العشوائية.

تألفت عينة الدراسة من (١٠١٨) مرشح للقبول في العام الأكاديمي (١٩٨٣-١٩٨٤) بجامعة هايفا، طبق عليهم اختبار دخول دورة القياس النفسي.

اعتمد الباحثان على ثلاثة اختبارات من بطارية اختبارات للقدرات تم تطبيقها في العام السابق، صنفت كالآتي:

(i) الاختبار الأول تضمن (٦٠) مفردة تقيس القدرات اللغوية.

(ii) الاختبار الثاني تضمن (٢٥) مفردة تقيس القدرة على التفكير الرياضي.

(ج) الاختبار الثالث تضمن (٢٤) مفردة تقيس القدرة على فهم وقراءة القطعة في اللغة.

وكان من المفترض أن هذه الاختبارات الثلاثة تتطلب استراتيجيات استجابة مختلفة، وأن الفروق في استراتيجية الاستجابة تتأثر بالدرجة التي يكون عندها فرض التناسب متحقق (تناسب زمن الإجابة).

بالنسبة للصورة الأصلية للاختبار، كانت تطبق باستخدام (٥) بدائل، ولتحقيق أهداف الدراسة الحالية تم إعداد (٦) صور إضافية لكل اختبار فرعي بأعداد مختلفة من البدائل (٢/ ٣/ ٤).

وكان المجموع الكلي للاختبارات التي طبقت على أفراد العينة (٢١) اختبار، تم تعيينها بشكل عشوائي عليهم.

قام الباحثان بتوجيه المفحوصين إلى ضرورة تسجيل عدد المفردات التي انتهوا من الإجابة عليها بعد العشرة دقائق الأولى، وكذلك بعد الخمسة عشر دقيقة الأولى، وكذلك الوقت الكلي الذي استغرقه كلاً منهم في إنهاء الاختبار.

وقد توصلت الدراسة إلى النتائج التالية:

١- ارتفاع قيم معاملات الصعوبة، والثبات، والتمييز وقيمة التباين مع الزيادة في عدد البدائل بالنسبة لبعض الاختبارات.

٢- أوضحت نتائج التحليل رفض فرض التناسب عند مستوى معنوية (٠,٠٥)، بالإضافة إلى أن الوقت المناسب لكل بديل كان دالة في عدد البدائل، فقد لوحظ في اختبار قراءة القطعة أن الوقت المحسوب للبدائل (٤) أطول من الوقت المحسوب للبدائل (٢، ٣).

٣- أظهرت نتائج التحليل أن صور الاختبار الأكثر صعوبة تميزت بانخفاض قيمة الخصائص السيكومترية المرتبطة بها (معامل الثبات ــ فعالية وقوة المشتتات ــ قوة تمييز المفردات) وذلك بالمقارنة بصور الاختبارات الأقل تعقيداً.

٤- فيما يتعلق باختبار العدد الأمثل لبدائل الفقرة، يرى الباحثان أن كل الحلول المقترحة لمعالجة هذه المشكلة والتي تؤيد استخدام ثلاثة بدائل غير مقبولة، لأنه في حقيقة الأمر تعتمد معظم الدراسات المنشورة على البدائل (٤، ٥).

دراسة فتحي الزيات (١٩٨٩):

بعنوان "أثر قوة المشتتات فى أسئلة الاختيار من متعدد على صدق الاختبار وثباتـــه". حيث تمثل هدف الدراسة فى اختبار أثر قوة تشتيت البدائل في أسئلة الاختيار من متعدد علـــى الخصائص السيكومترية للاختبار (الصدق العاملي والصدق الفارقي ــ الانحراف المعيـــاري) وخصائص مفردات الاختبار (قيم معاملات الصعوبة ومعاملات التمييز).

تألفت عينة الدراسة من (٤١٤) طالب وطالبة بالصفوف الأولـــى والثـــاني والثـــالث بالمرحلة الثانوية بمدرستي الفلاح الثانوية بمكة المكرمة وجدة الثانوية، تتراوح أعمارهم بيـــن (١٥-٢٠) عاماً بمتوسط (١٧,٢) وانحراف معياري (١,٤).

استخدم الباحث اختبار معاني الكلمات لأحمد زكي صالح، وهـــو أحـــد الاختبـــارات الفرعية لاختبار القدرات العقلية الأولية، والذي عربه عن ثيرستون، حيـــث يقيـــس القـــدرة اللغوية، وقد تم استخدامه ثلاثة مرات، مرة قبل تعديل البدائل ومرتين بعد التعديل.

وقد تمثلت نتائج الدراسة فيما يلي:

١- وجود فروق دالة إحصائياً عند مستوى (٠,٠٠١) بين معاملات الصعوبة وقوة التميـــيز للمفردات قبل وبعد تعديل البدائل لصالح بعد التعديل.

٢- زيادة قيمة التباين والانحراف المعياري للدرجات بعد تعديل المشتتات.

٣- ارتفاع قيمة معامل ثبات الاختيار بعد تعديل قوة المشتتات باســـتخدام معادلـــة التجزئـــة النصفية.

٤- ارتفاع قيمة (ت) لدلالة الفروق بين متوسطات الإرباعي الأعلى والأدنى بعد تعديل قـــوة تشتيت البدائل، وهذا يعني أن قوة المشتتات تؤثر على الصدق الفارقي للاختبار.

٥- زيادة قيم معاملات الارتباط لاختبار معاني الكلمات بالمحكات المستخدمة بعد التعديـــل، حيث تراوحت بين (٠,٢٩١ ــ ٠,٢٨٧) وهي دالة عند مستوى (٠,٠٠١).

دراسة تريفيسون Trevison, 1990:

بعنوان "ثبات وصدق اختبارات الاختيار من متعدد كدالة فى عدد البدائل لكل مفـــردة وقدرة التلاميذ". وقد هدفت الدراسة إلى تحديد العدد الأمثل لبدائل الفقرة، الذي يمكنه أن يرفـع من ثبات وصدق الاختبار وفقاً لقدرات الممتحنين.

تألفت عينة الدراسة من (٤٣٣) طالب من طلاب السنة الحادية عشر بالمدرسة العليـا في (بورتلاند).

استخدم الباحث في هذه الدراسة ثلاثة اختبارات فرعية (لفظية) من بطارية الاســتعداد (اختبارات الاستعداد الدراسي)، كل صورة من صور الاختبارات تكونت من أعــداد مختلفــة لبدائل الفقرة (٣/ ٤/ ٥).

وقد تم حذف الفقرات الأقل تمييزاً من الاختبار وصولاً إلى الصورة النهائية.

قام الباحث بتقسيم أفراد العينة إلى ثلاثة مجموعات وفقاً لمستويات القدرة (مرتفعــة ــ متوسطة ــ منخفضة).

وقد توصل الباحث إلى النتائج الآتية:

١- وجود فروق دالة عند مستوى (٠,٠٥) في معاملات الثبات للطلاب منخفضـــي القــدرة لأسئلة الاختيار من متعدد ذات البدائل (٣/ ٤/ ٥) لصالح البدائل (٥).

٢- لا توجد فروق دالة إحصائياً في معاملات الثبـــات بالنســبة للمجموعتيــن المتوســطة والمرتفعة في القدرة عند اختلاف عدد بدائل الفقرة (٣/ ٤/ ٥) بدائل.

٣- العدد الأمثل للبدائل عبر مجموعات القدرة ثلاثة بدائل.

٤- لم تُظهر نتائج الدراسة معاملات صدق غير متوقعة.

دراسة مارشال Marshal, 1990:

بعنوان "أثر استخدام أعداد مختلفة من بدائل الإجابة على الثبات والصدق لمقياس ليكرت للأطفال فى مستويات تعليمية مختلفة" وقد هدفت الدراسة إلى اختبار أثر استخدام أعداد مختلفة من بدائل الاستجابة (٢/ ٣/ ٤/ ٥) بدائل، على صدق وثبات مقياس ليكرت للأطفال في مستويات تعليمية مختلفة.

تألفت عينة الدراسة من (٢٥٢) طفل من المراحل التعليمية الثانية وحتى السابعة.

تم تعيين هؤلاء الأطفال في مجموعات عشوائياً وفقاً للمرحلة التي ينتمون إليها، حيث شملت المجموعة الأولى الأطفال في المرحلتين الثانية والثالثة، والمجموعة الثانية شملت المرحلتين الرابعة والخامسة، والمجموعة الثالثة شملت المرحلتين السادسة والسابعة.

اعتمدت الدراسة على مقياس القلق الظاهر للأطفال وقائمة القلق (سمة ـ حالة) للأطفال أيضاً، وقد تم تطبيق المقاييس بأعداد مختلفة من البدائل (٢/ ٣/ ٤/ ٥).

تم حساب معاملات ألفا لقياس الاتساق الداخلي للمقاييس، واستخدم معامل الارتباط بين درجات العينة في كلا المقياسين للتحقق من توافر الصدق التلازمي للمقاييس.

وقد أكدت نتائج الدراسة أن استخدام أربعة بدائل للإجابة زاد من قيمة ثبات وصدق المقاييس بالنسبة للأطفال في المرحلتين الثانية والثالثة، بينما كانت هذه المعاملات مستقلة عن بعضها البعض بالنسبة للأطفال في المراحل من الرابعة وحتى السابعة.

ويؤكد الباحث في نهاية الدراسة على ضرورة تركيز البحوث على المتغيرات المرتبطة بالعينة التي يطبق عليها المقاييس، بدلاً من التركيز على المتغيرات المرتبطة بالمقاييس ذاتها.

دراسة مايكل وآخرون Michael & et al., 1991:

بعنوان "دراسة أثر عدد البدائل لكل مفردة وقدرات التلاميذ على صدق وثبات الاختبار". حيث هدفت الدراسة إلى التحقق من خاصيتي ثبات وصدق أسئلة الاختيار من متعدد، كدالة في عدد البدائل لكل مفردة، وقدرات التلاميذ التحصلية.

تألفت عينة الدراسة من (٤٣٥) طالب بالمدارس الثانوية في بورتلاند، طبقت عليهم بطارية اختبارات قبل دخولهم الجامعة بواشنطن.

وقد تمثلت إجراءات الدراسة فيما يلي:

١- تطبيق اختبار اختيار من متعدد يتألف من ثلاثة صور بأعداد مختلفة من البدائل (٣/ ٤/ ٥) بطريقة عشوائية بعد حذف المشتتات غير الجيدة.

٢- تصنيف الطلاب حسب مجموعات القدرة (مرتفعة ــ متوسطة ــ منخفضة) باستخدام درجات القطع.

٣- تم حساب المتوسطات، والانحرافات المعيارية ومعاملات الصدق، وحسبت معاملات الثبات بتطبيق معادلة سبيرمان براون وكيودر وريتشاردسون وذلك لكل صورة من صور الاختبار، ولكل مجموعة من مجموعات القدرة.

وقد توصلت الدراسة إلى النتائج التالية:

١- وجود فروق دالة عند مستوى (٠,٠٥) بين معاملات الثبات للطلاب ذوي القدرات المنخفضة لصالح البدائل الثلاثة وكذلك بالنسبة للصدق.

٢- لا توجد فروق دالة بين معاملات الثبات للطلاب ذوي القدرة العالية أو القدرة المتوسطة نتيجة لاختلاف عدد بدائل الفقرة.

٣- لا توجد فروق دالة بين معاملات الصدق للطلاب ذوي القدرة المرتفعة والمتوسطة نتيجة اختلاف عدد البدائل في صور الاختبار المختلفة (٣/ ٤/ ٥) بدائل.

٤- العدد الأمثل من البدائل عبر المجموعات متفاوتة القدرة ثلاثة بدائل.

٥- ارتفاع قيمة معاملات تمييز المفردات بانخفاض عدد البدائل إلى ثلاث بدائل وخاصة بالنسبة للمجموعة مرتفعة القدرة.

٦- انخفاض عدد المفردات ذات المشتتات غير الفعالة عند استخدام ثلاثة بدائل للفقرة.

٧- جميع النتائج تؤكد فاعلية صورة الاختبار بثلاثة بدائل بالنسبة لمجموعات القدرة ككل (مستقلة ــ متحدة).

دراسة هانكوك وآخرون Hancock & etal., 1993:

بعنوان "المقارنة بين معاملات الثبات لمفردات الاختيار من متعدد ذات البديلين ومفردات الصواب والخطأ". حيث هدفت الدراسة إلى المقارنة بين أسئلة الاختيار من متعدد ذات البديلين وأسئلة الصواب والخطأ فيما يتعلق بمعاملات الثبات الخاصة بكلا النوعين من المفردات.

تألفت عينة الدراسة من (١١١) طالب من طلاب مدرسة بدسترون الثانوية، حيث تم تصنيفهم إلى (٤٠) طالب في الصف العاشر والحادي عشر (٣٥) طالب والثاني عشر (٣٦) طالب.

تم إعداد ثلاثة اختبارات موضوعية، الأول من نوع الاختيار من متعدد ويتألف من (٤٠) مفردة، والثاني والثالث أسئلة صواب وخطأ، وكان النوع الأول من المفردات وهو الاختيار من متعدد ببديلين لكل مفردة.

وقد قام الباحثين بتطبيق الثلاثة صور الاختبارية على أفراد العينة بطريقة التعيين العشوائي.

وقد توصلت الدراسة إلى النتائج التالية:

١- وجود فروق ذات دلالة إحصائية عند مستوى معنوية (٠,٠٥) بين قيم معاملات الثبات المحسوبة لكلا النوعين من المفردات وكانت الفروق لصالح أسئلة الاختيار من متعدد ذات البديلين.

٢- إن أسئلة الاختيار من متعدد تقدم مقياساً أكثر ثقة يمكن الاعتماد عليه بصورة أفضل من أسئلة الصواب والخطأ. بالإضافة لذلك أوضحت النتائج أن أسئلة الصواب والخطأ لا تقيس نفس القدرات التي تقيسها أسئلة الاختيار من متعدد.

٣- كشف اختبار توكي لدلالة الفروق بين المتوسطات أن الفروق لصالح أسئلة الاختيار من متعدد ذات المستوى المتوسط الصعوبة عند مستوى دلالة (٠,٠٥).

٤- ارتفاع معامل الثبات لأسئلة الاختيار من متعدد يرجع إلى الصعوبة النسبية لأسئلة الاختيار من متعدد ذات البديلين بالمقارنة بأسئلة الصواب والخطأ.

دراسة جيمس برونو James Bruno, 1995:

بعنوان "تحديد العدد الأمثل للبدائل لمفردات الاختيار من متعدد من وجهة نظر نموذج المعلومات"، وقد تمثلت أهداف الدراسة فيما يلى:

١- الاختبار التحليلي الدقيق للتساؤل الخاص بالعدد الأمثل لبدائل فقرات اختبار الاختيار من متعدد، في إطار نظري وليس تجريبي من منظور "نموذج نظرية المعلومات".

٢- مقارنة النتائج المستخلصة من تطبيق الأسلوب الجديد المقترح مع غيرها من النتائج المستمدة من تطبيق الأساليب الإحصائية التقليدية، والتي توصلت إليها العديد من الدراسات والبحوث.

أوضحت الدراسة أن محتوى المعلومات الذي تتضمنه المفردة يعتبر مصدر هام من مصادر قوة فقرات الاختبار، وهو مدخل هام يجب أخذه في الاعتبار عند تحديد العدد الأمثل لبدائل فقرات اختبار الاختيار من متعدد، وأن قوة الفقرة تزيد بعدد معين من البدائل، ولكن التحسن ينخفض كلما ابتعدنا عن ثلاثة أو أربعة بدائل.

وقد أوضح الباحث في هذه الدراسة أن المدخل المقترح لتحديد العدد الأمثل لبدائل الفقرة (مدخل نظرية المعلومات)، يعتمد أساساً على اعتبار محتوى المعلومات المرتبطة بالمفردة دالة خطية تأخذ شكل منحنى يمثل إحداثه السيني (عدد البدائل) والإحداث الصادي (محتوى المعلومات).

والعلاقة بينهم طردية، فكلما زاد عدد البدائل زاد تبعاً له محتوى المعلومات المرتبطة بالمفردة، وهناك نقطة حد أقصى لهذه الزيادة وهي (متوسط محتوى المعلومات)، وأن مفردة الاختبار التي تقدم للطالب باستخدام بديل واحد سوف ينتج عنها صفر في محتوى المعلومات بالنسبة للطالب.

وقد اعتمدت الدراسة على أساليب التحليل الرياضي في إطار نموذج نظرية المعلومات لحل مشكلة العدد الأمثل للبدائل في فقرات الاختيار من متعدد، وقدمت مجموعة من المعادلات الرياضية التي تعتمد على التفاضلات الجزئية وقواعد اللوغاريتمات الطبيعية والدوال الرياضية.

وكان من أهم نتائج الدراسة أن استخدام ثلاثة بدائل للفقرة هو الأمثل فيما يتعلق بمتوسط محتوى المعلومات الخاصة بكل بديل.

كذلك أكدت الدراسة على أن ما توصلت إليه من نتائج يكاد يتفق بشكل كبير مع ما توصلت إليه البحوث الإحصائية والسيكومترية والملاحظات التجريبية التي اهتمت بمعالجة مشكلة عدد البدائل الأمثل لمفردات اختبارات الاختيار من متعدد.

وقد أشار الباحث في نهاية الدراسة إلى أهمية مراعاة أبعاد الوقت والتكلفة أثناء التحليل، بمعنى كم عدد من البدائل يحقق لنا أكبر قدر من المعلومات الخاصة بالمفردة ويكون تأثيره إيجابي على خصائص الاختبار السيكومترية (الصدق ــ الثبات ....) وذلك بتكلفة مناسبة وفي حدود وقت مناسب.

دراسة نعمان الموسوي ١٩٩٧(*):

بعنوان "تحديد العدد الأمثل لخيارات الفقرة فى اختبار الاختيار من متعدد فى ضوء بعض خصائصه السيكومترية". وقد هدفت الدراسة إلى تحديد العدد الأمثل لبدائل الفقرة في اختبار الاختيار من متعدد، والذي يحقق أقصى قيمة ممكنة لبعض الخصائص السيكومترية للاختبار (الصدق ــ الثبات ــ القدرة التمييزية) عند تطبيقه على مفحوصين ذوي مستويات تحصيلية متباينة.

تألفت عينة الدراسة من (١٤٤) طالب وطالبة بجامعة البحرين وقد قام الباحث بإعداد اختبار تحصيلي يتألف من (٤٠) مفردة بأربعة بدائل في مقرر القياس والتقويم، وقام بتطبيقه على أفراد العينة، ثم أعاد تطبيقه عليها بعد أسبوعين، بعد تخفيض عدد بدائل الفقرة إلى ثلاثة، وبعد تصحيح الاختبار قسم الطلاب إلى ثلاث فئات وفقاً لدرجة الطالب النهائية في المقرر والتي اعتبرها الباحث محكاً لاحتساب معامل الصدق التنبؤي للاختبار.

وقد توصل الباحث إلى النتائج التالية:

١- وجود فروق معنوية عند مستوى (٠,٠٥) بين معاملي الصدق التنبؤي لصورتي الاختبار بثلاثة بدائل وبأربعة بدائل للفئة الوسطى لصالح البدائل الثلاثة.

(*) نقلاً عن حسن مصطفى، ٢٠٠٠.

٢- لم تظهر النتائج وجود فروق دالة إحصائياً عند مستوى (٠,٠٥) بين معـــاملي الثبـــات والتمييز لصورتي الاختبار بثلاثة بدائل وأربعة بدائل لكل فئة ولجميع الفئات (منخفضـــة ـــ متوسطة ـــ مرتفعة) القدرة.

دراسة روجرز وهارلي Rogers & Harley, 1999:

بعنوان "المقارنة الامبيريقية بين استخدام ثلاثة وأربعة بدائل لمفردات اختبار الاختيار من متعدد وأثرها على ثبات الاتساق الداخلي للاختبار". حيث هدفت الدراسة إلـــى المقارنـــة الامبيريقية بين استخدام ثلاثة بدائل وأربعة بدائل لمفردات اختبارات الاختيـــار مـــن متعـــدد واختبار أثر ذلك على الخصائص السيكومترية للمقياس وثبات الاتساق الداخلي.

تألفت عينة الدراسة من ٨٠% من طلاب ثلاثة مدارس ثانوية عليا في إقليـــم ألبرتـــا بكندا، والذين يتم إعدادهم لاختبار نهاية العام الدراسي.

اعتمد الباحثان على اختبار للرياضيات تم تطبيقه من قبل على تلاميذ الصف الثـــاني عشر في نهاية العام الدراسي (١٩٩١-١٩٩٢)، في إقليم ألبرتا بكندا، حيـــث تضمـــن هـــذا الاختبار قسمين:

(أ) القسم الأول خاص بأسئلة الاختيار من متعدد.

(ب) القسم الثاني خاص بالأسئلة ذات الاستجابات البنائية.

وقد تم التركيز على القسم الأول، حيث تضمن (٤٠) مفردة بأربعـــة بدائـــل، وقـــام الباحثان بإعداد صورة أخرى لنفس الاختبار بثلاثة بدائل للفقرة وذلك عن طريق حذف أحـــد المشتتات.

تم تطبيق الاختبار على التلاميذ بأسلوب التعيين العشوائي لصورتي الاختبار بثلاثـــة بدائل وأربعة بدائل، وكان زمن تطبيق الاختبار ساعتين.

ولم تكشف نتائج التطبيق عن وجود فروق معنوية في متوسطات الأداء بين طـــلاب المدارس المختلفة، وتبعاً لذلك كانت عينات المدارس متحدة معاً وكان العدد الكلي لأفراد العينة (١٥٨) تلميذ.

وقد توصلت الدراسة إلى النتائج التالية:

١- وجود فروق معنوية بين متوسطات الأداء على صورتي الاختبار بثلاثة بدائل وأربعة بدائل لصالح البدائل الثلاثة، كذلك كانت الفروق في التباينات بين كلا الصورتين (F= 1, 12)، والفروق في معاملات ثبات الاتساق الداخلي (W= 1, 25) لم تكن دالة إحصائياً.

٢- فيما يتعلق بأثر عدد بدائل الفقرة على مستوى الصعوبة ومستوى تمييز الفقرات، أظهرت النتائج وجود فروق دالة إحصائياً بين متوسطات صعوبة الفقرات لكلاً من صورتي الاختبار بثلاثة بدائل وأربعة بدائل، حيث بلغت قيمتها للصورة الأولى (0.659)، والثانية (0.574) على التوالي، وهذا يعني أن مستوى صعوبة الفقرة ارتبط ارتباط عكسي بعدد البدائل وبالنسبة للفروق في متوسط تمييز الفقرة بين كلا الصورتين للاختبار (٣/ ٤) بدائل لم تكن معنوية.

٣- فيما يتعلق بأثر عدد بدائل الفقرة على متوسطات الوقت المطلوب لإنجاز الاختبار، أوضحت النتائج عدم وجود فروق معنوية بين صورتي الاختبار بثلاثة بدائل، أربعة بدائل فيما يتعلق بمتوسط الوقت المطلوب للإجابة على كلاً منها، ويرى الباحثان أن هذه النتيجة تتناقض مع ما أسفرت عنه نتائج البحوث والدراسات السابقة التي تؤكد أن اختبار الاختيار من متعدد بثلاثة بدائل للفقرة يتم الإجابة عليه بصورة أسرع من الاختبار بأربعة بدائل.

تعقيب على البحوث والدراسات المرتبطة بالمحور الأول (أثر عدد بدائل مفردات الاختيار من متعدد على قيمة معامل ثبات القياس:

باستعراض الباحثة لنتائج الدراسات التي اهتمت بمعالجة أثر عدد بدائل مفردات الاختيار من متعدد على قيمة معامل الثبات، لاحظت اختلاف وجهات نظر الباحثين حول العدد الأمثل لبدائل المفردة، فقد اتفقت نتائج دراسات كل من Straton & Catts, 80، Trevison, 90، Trevison & etal., 91، James Bruno, 95، نعمان الموسوي ١٩٩٧، Rogers & Harley, 99، على فاعلية استخدام ثلاثة بدائل وأثرها الإيجابي على متوسط الأداء الاختباري للطلاب وبالتالي على خصائص الاختبار السيكومترية وأهمها الصدق والثبات وقوة تمييز المفردات.

ومن ناحية أخرى تناقضت هذه النتائج مع نتائج دراسات كل من Green & etal., 82، Marshal, 90، Budescu & Nevo, 85، حيث أيدت الاقتراح الخاص بفاعلية زيادة عدد بدائل المفردة عن ثلاثة بدائل، واستخدام أربعة بدائل للمفردة لأن ذلك له تأثيره الإيجابي على خصائص الاختبار السيكومترية.

كذلك وجدت بعض الدراسات التي حاولت اختبار العدد الأمثل لبدائل المفردة في علاقته بقدرات الطلاب التحصيلية (مرتفعة ــ متوسطة ــ منخفضة)، ومنها دراسة Green & etal., 82، Trevison, 90، Trevison & etal., 91، نعمان الموسوي ٩٧، وقد كان التعارض والاتفاق فيما بينها حول عدد البدائل الأمثل عبر المجموعات متباينة القدرة كالآتي:

أولاً: اتفقت نتائج دراسة كل من Green & etal., 82، ودراسة Trevison, 90، ودراسة Trevison & etal., 91، على وجود فروق دالة في ثبات أسئلة الاختيار من متعدد ذات البدائل (٣/ ٤/ ٥) بالنسبة للطلاب ذوي القدرة المنخفضة عند استخدام خمسة بدائل للمفردة.

ثانياً: توصلت دراسة Green & etal., 82 إلى عدم وجود فروق في معاملات الثبات لأسئلة الاختيار من متعدد ذات البدائل (٣/ ٤/ ٥) بالنسبة للمجموعة متوسطة القدرة واتفقت معها دراسة Trevison, 90، ودراسة Trevison & etal., 91.

ثالثاً: توصلت دراسة Green & etal., 82 إلى وجود فروق في معاملات ثبات مفردات الاختيار من متعدد ذات البدائل (٣/ ٤/ ٥) بالنسبة لمجموعة القدرة المرتفعة لصالح الخمسة بدائل وتناقضت مع دراسة Trevison, 90، ودراسة Trevison & etal., 91.

رابعاً: بالنسبة لعدد البدائل الأمثل عبر مجموعات القدرة ككل، فقد توصلت دراسة Green & etal., 82 إلى فاعلية أربع بدائل على قيمة معامل الثبات وتناقضت معها دراسة Trevison, 90، ودراسة Trevison & etal., 91 حيث توصلتا إلى أن العدد المناسب لبدائل الفقرة هو ثلاثة بدائل، واتفقت معهم في ذلك أيضاً دراسة نعمان الموسوي ٩٧.

وترى الباحثة أن من أهم الأسباب التي تؤكد أو تبرز أهمية الدراسة الحالية بالإضافة إلى تعارض وجهات نظر الباحثين السابقين حول عدد البدائل الأمثل، هو ندرة الدراسات العربية حول هذا الموضوع بالرغم من انتشار استخدام الاختبارات الموضوعية وبصفة خاصة أسئلة الاختيار من متعدد في عمليات القياس والتقويم.

المحور الثاني: البحوث والدراسات التي اهتمت بمعالجة أثر موقع أو ترتيب مفردات الاختبار على قيمة معامل ثبات القياس:

دراسة برينير Brenner, 1965:

بعنوان "صعوبة الاختبار وثباته، وتمييزه كدوال فى ترتيب صعوبة المفردات". وقد تمثل الهدف الرئيسي للدراسة في اختبار أثر اختلاف أساليب ترتيب مفردات الاختبار على الخصائص الإحصائية للمفردة (قيم معاملات الصعوبة ومعاملات التمييز)، وعلى قيمة معامل ثبات الاختبار.

قام الباحث بإجراء التطبيق الفعلي للدراسة على الطلاب المقيدين لدورة علم النفس التربوي رقم (٤٠٧) في جامعة أوهايو (Ohio).

وقد قام الباحث بإجراء أربعة تجارب، وداخل كل تجربة، قام بتطبيق أكثر من صورة للاختبار وهو من نوع الاختيار من متعدد، يتألف كل اختبار من (٤٠) مفردة بأربعة بدائل، تم عرضها بأساليب مختلفة لترتيب المفردات، وذلك لدراسة أثر التنوع في موقع المفردات أو اختلاف ترتيبها داخل الاختبار على متغيرات الدراسة (الثبات ــ التمييز ــ الصعوبة) وبيانها كالتالي:

التجربة الأولى:

وفيها تم تطبيق الاختبار بأشكال مختلفة لترتيب المفردات وهي:

١- الترتيب العشوائي للمفردات من حيث درجة الصعوبة.

٢- الترتيب من السهل إلى الصعب (متزايد الصعوبة تدريجياً).

٣- عكس الترتيب السابق من الصعب إلى السهل.

التجربة الثانية:

وفيها تم انتقاء (٤٠) مفردة عشوائياً من اختبار يتألف من (٨٠) مفردة، وتمت عملية الترتيب بالشكل الآتي:

١- تم انتقاء أسهل (١٠) مفردات وأصعب (١٠) مفردات، من الاختبار المكون من (٤٠) مفردة.

٢- قام الباحث بإعداد صورتين للاختبار بأسلوبين من أساليب الترتيب للمفردات، في الصورة الأولى تم وضع أسهل عشرات مفردات في بداية الاختبار يليها الثلاثون مفردة الباقية مرتبة عشوائياً، وفي الصورة الثانية تم العكس، حيث قام الباحث بوضع أصعب عشرة مفردات في بداية الاختبار يليها الثلاثون مفردة الباقية مرتبة أيضاً بأسلوب عشوائي.

التجربة الثالثة:

وفيها قام الباحث بتطبيق الاختبار بصورتين مختلفتين لترتيب المفردات وهي:

١- الصورة الأولى تم ترتيب وعرض مفرداتها بأسلوب متدرج الصعوبة (من السهل إلى الصعب).

٢- الصورة الثانية وفيها تم ترتيب المفردات بأسلوب عكسي متناقص تدريجياً في الصعوبة (من الصعب إلى السهل).

التجربة الرابعة:

وفيها قام الباحث بتطبيق نفس الإجراء السابق الذي استخدمه في التجربة الثانية عند انتقاء المفردات، ثم قام بتطبيق الاختبار بأسلوبين لترتيب وعرض المفردات، إحداهما من السهل إلى الصعب والآخر من الصعب إلى السهل.

وقد قام الباحث بحساب المقاييس الإحصائية الخاصة بكل اختبار وبكل صورة من صور ترتيب المفردات، ثم قام بتطبيق اختبارات الدلالة الإحصائية لكل التجارب.

وقد أسفرت نتائج الدراسة عن الآتي:

١- فيما يتعلق بأثر موقع وترتيب المفردات على قيم معاملات الصعوبة ومعاملات الثبات لم تسفر نتائج التحليل عن وجود أي فروق دالة إحصائية بين أشكال الاختبارات المختلفة الترتيب في قيم هذه المعاملات.

٢- فيما يتعلق بقوة الاختبار التمييزية لم تظهر النتائج وجود فروق دالة بين قيم معاملات التمييز المحسوبة في كل التجارب ولكل صور الاختبارات التي اختلفت في موقع مفرداتها، باستثناء التجربة الثانية فقط، حيث أوضحت النتائج وجو فرق معنوي عند مستوى (٠,٠٥) بين قيم معاملات التمييز المحسوبة للاختبارات المستخدمة.

دراسة كروماك وساكس Cromack & Sax, 1966:

بعنوان "أثر الأشكال المختلفة لترتيب المفردات على الأداء الاختبارى". حيث هدفت الدراسة إلى اختبار أثر ترتيب وتقديم مفردات الاختبار إلى التلاميذ في أربعة صور تجريبية مختلفة الترتيب على الأداء الاختبارى للطلاب.

تألفت عينة الدراسة من (٢٧٦) طالباً من الطلاب المقيدين للدورة التمهيدية في التربية في جامعة هاواى.

اعتمد الباحثان على اختبار للقدرة العددية يتألف من (٧٠) مفردة تم ترتيبها وفقاً لمستويات الصعوبة، في أربع صور اختبارية تجريبية وهي كالتالي:

١- الصورة (E) وفيها تم عرض وتقديم المفردات بترتيب متدرج من الصعوبة أو متزايد تدريجياً في الصعوبة (من السهل إلى الصعب).

٢- الصورة (H) وفيها تم عرض وترتيب المفردات بأسلوب منخفض تدريجياً في مستوى الصعوبة (من الصعب إلى السهل).

٣- الصورة (M) وفيها تم ترتيب المفردات بأسلوب يزيد من دافعية الطلاب (Motivation) حيث تضمن وجود مفردة سهلة يتم وضعها بين كل ستة مفردات لها مستوى صعوبة متوسطة (٠,٥) تقريباً.

٤- الصورة (R) وفيها تم ترتيب مفردات الاختبار بأسلوب عشوائي.

وقد قام الباحثان بتعيين صور الاختبار المختلفة بأسلوب عشوائي على أفراد العينة، وتم تحديد زمن (٣٠) دقيقة لإنهاء الاختبار.

وفي نفس التوقيت قام الباحثان بتطبيق تجربة مماثلة للصور الاختبارية الأربعة، على عينة تتألف من (١٩١) طالباً من المقيدين للدورة التمهيدية في علم النفس، بعد تخصيص زمن للإجابة قدرة (٤٨) دقيقة، وكان الهدف من هذه التجربة دراسة تأثير عامل الوقت (إطالة الوقت) على الأداء الاختباري للطلاب.

وقد أسفرت نتائج الدراسة عن الآتي:

١- وجود فروق دالة إحصائياً في قيمة (F) عند مستوى معنوية (٠,٠٠١) بين الصور الاختبارية مختلفة الترتيب والتي استغرق تطبيقها (٣٠) دقيقة.

٢- ارتفاع في متوسط درجات الصورة (E) والتي تم ترتيب مفرداتها من السهل إلى الصعب بالمقارنة بالصورة (H) والتي رتبت مفرداتها بأسلوب عكسي من الصعب إلى السهل، بينما لم توجد فروق دالة في متوسطات صور الاختبار (M, R).

٣- النسبة (F) الخاصة بالمجموعة الثانية والتي طبق عليها الاختبار بصوره المختلفة الترتيب، في زمن قدره (٤٨) دقيقة لم تكن دالة إحصائياً، ويرى الباحث أن ترتيب المفردات في أشكال متنوعة تكون تأثيراتها على الأداء الاختباري واضحة، ولكن في حدود معينة من الوقت الاختباري.

دراسة سموس وماتز Smouse & Munz, 1968:

بعنوان "تأثيرات التفاعل بين تسلسل صعوبة المفردة وقلق التحصيل على الأداء الأكاديمي". حيث هدفت الدراسة إلى اختبار أثر ثلاثة أساليب لعرض وترتيب مفردات الاختبار، وكذلك دراسة أثر قلق الاختبار على الأداء الاختباري للطلاب.

تألفت عينة الدراسة من (١٨١) طالباً من الطلاب المسجلين لأداء الامتحان النهائي في الدورة التمهيدية لعلم النفس.

وقد تضمن الاختبار (١٠٠) مفردة، قام الباحثان بعرضها وترتيبها في ثلاثة صور اختباريه وهي:

١- الصورة الأولى تم ترتيب مفرداتها من السهل إلى الصعب.

٢- الصورة الثانية تم ترتيب مفرداتها من الصعب إلى السهل.

٣- الصورة الثالثة وفيها عرضت المفردات مرتبة عشوائياً.

وبالإضافة لذلك قاموا بتعريض مجموعة واحدة فقط من التلاميذ لخبرات ومعلومات مثيرة للقلق.

وفي تجربة ثانية، قام الباحثان بتطبيق نفس الاختبار على نفس المفحوصين، ولكن بعد تصنيفهم وفقاً لنوع قلق الاختبار لديهم (مرتفع ــ متوسط ــ منخفض).

أوضحت نتائج الدراسة أن اختلاف موقع وترتيب المفردات من الاختبار لم يؤثر على الأداء الاختباري للطلاب، وكان هناك صعوبة في تفسير التفاعل بين أنواع قلق الاختبار وأسلوب عرض وترتيب المفردات.

ويرى الباحثان أن القائمين بتطبيق الاختبارات أو الباحثين يجب عليهم أن يتحكموا في متغير قلق الاختبار عند دراستهم لأثر ترتيب المفردة على الأداء الاختباري.

دراسة فلاجير وآخرون Fleugher & et al., 1968:

بعنوان "إعادة ترتيب المفردات تحت شروط الاختبار النموذجية". حيث هدفت الدراسة إلى اختبار أثر تسلسل وإعادة ترتيب مفردات الاختبار على الأداء الاختباري، والإحصاءات المرتبطة بالمفردة (الصعوبة ــ الارتباط الجزئي ــ عدم الاستجابة).

اعتمد الباحثون على استخدام اختبار الاستعداد الدراسي (SAT) للطلاب المتقدمين لدخول الكلية لأول مرة في مارس (١٩٦٥)، مع التركيز على القسم اللفظي المرتبط بالمفردات اللفظية، والقسم الرياضي المرتبط بالمفردات الرياضية، بالنسبة للقسم اللفظي تألف من (٤٠) مفردة تم توزيعها على (٨) مجموعات من المفردات، اشتملت كل مجموعة على خمسة مفردات، تم ترتيبها داخلياً من السهل إلى الصعب، وبالنسبة للقسم الرياضي تألف من (٢٥) مفردة تم توزيعها على مجموعات يتضمن كلاً منها خمسة مفردات متماثلة ولكنها مختلفة الترتيب.

تألفت عينة الدراسة من (٥٠٠٠) طالب من المرشحين لدخول الكلية لأول مرة وذلك للتحليلات الخاصة بأثر إعادة الترتيب على الدرجة الكلية، (٤٩٥) طالب تم انتقاؤهم كعينات عشوائية مستقلة للتحليلات الخاصة بإحصاءات المفردة.

وقد أسفرت نتائج الدراسة عن ما يلي:

١- وجود فروق معنوية جديرة بالملاحظة في مستويات الصعوبة الخاصة بالأجزاء اللفظية نتيجة إعادة تسلسل وترتيب المفردات.

٢- لم توجد فروق معنوية دالة إحصائياً بين معاملات الارتباط المحسوبة بين اختبار الاستعداد الدراسي واختبارات تحصيلية أخرى، نتيجة إعادة ترتيب وتسلسل المفردات في صورة مختلفة.

٣- ارتفاع تقديرات معاملات الصعوبة بالنسبة للترتيب المعياري (من السهل إلى الصعب)، خاصة بالنسبة للمفردات التي ظهرت في نهاية الاختبار.

٤- فيما يتعلق بأثر إعادة ترتيب المفردات داخل الاختبار على الارتباط الجزئي للمفردات بالأداء الكلي على الاختبار، لم توجد فروق معنوية بين قيم معاملات الارتباط الجزئي المحسوبة.

٥- وجود عدد كبير من الطلاب الذين لم يستجيبوا استجابة كلية على بعض المفردات وخاصة التي جاء موقعها في نهاية الاختبار ويرى الباحث أن هذه النتيجة قد ترجع إلى عدم قدرة الطالب على الوصول إلى هذه المفردات، ولا يجب أن نعتبر ذلك مؤشراً لارتفاع مستوى الصعوبة، وأن الافتراض المسبق بالتكافؤ بين طرق الترتيب المختلفة، من المحتمل أن يكون افتراضاً خاطئاً.

دراسة مارسو Marso, 1970:

بعنوان "ترتيب مفردات الاختبار، زمن الاختبار والأداء الاختباري". حيث هدفت الدراسة إلى اختبار أثر تسلسل وترتيب المفردات وقلق الاختبار على الأداء الأكاديمي للطلاب وعلى الوقت المطلوب لإنهاء الاختبار.

وقد قام الباحث بإجراء تجربتين لاختبار فروض الدراسة وهي كالتالي:

١- التجربة الأولى:

وفيها تألفت عينة الدراسة من (١٢٢) طالباً من الطلاب المسجلين لدورة علم النفس التربوي في جامعة مدينة جرين (Green)، وتم استخدام اختبار اختيار من متعدد يتألف من (١٣٩) مفردة يتراوح مدى صعوبتها من (صفر -١٠٠%) تقيس القدرة اللغوية.

وقد قام الباحث بإعداد ثلاثة صور للاختبار باستخدام ثلاثة أساليب لعرض وترتيب المفردات وهي:

(i) الترتيب من السهل إلى الصعب (متزايد الصعوبة تدريجياً).

(ii) الترتيب من الصعب إلى السهل (منخفض الصعوبة تدريجياً).

(iii) الترتيب العشوائي.

وتم تعيين صور الاختبار على أفراد العينة عشوائياً، وكانت الاختبارات تطبق على أنها اختبارات قوة، وبرغم ذلك كان يطلب من كل تلميذ تسجيل الوقت الذي أنهى فيه الاختبار على ورقة الإجابة.

وقد أوضحت نتائج التجربة الأولى:

١- باستخدام درجات التحصيل كمحك فإن العامل الخاص بقلق الاختبار كان له تأثير دال إحصائياً على الأداء الاختباري للطلاب عند مستوى معنوية (٠,٠١)، وأن متغير موقع المفردة أو اختلاف ترتيب المفردات لم يؤثر على الأداء الاختباري للطلاب.

٢- لا يوجد تأثير دال إحصائياً للتفاعل بين قلق الاختبار وموقع المفردة أو تسلسل المفردات على أداء الطلاب، وأن التلاميذ الذين حصلوا على درجات مرتفعة في مقاييس قلق الاختبار كان أداؤهم الاختباري أقل جودة على الثلاثة صور الاختبارية.

٣- لم يكن هناك أي تأثير معنوي لمتغير موقع المفردات وكذلك قلق الاختبار على مقدار الوقت الذي يستغرقه الطالب في الإجابة عند مستوى معنوية (٠,٠١).

التجربة الثانية:

وقد هدف الباحث من إجراء التجربة إلى دراسة أثر استخدام أسلوب التماثل في محتوى المادة التعليمية كإجراء لترتيب المفردات وكذلك قلق الاختبار على الدرجات التحصيلية للطلاب، وعلى مقدار الوقت المستغرق في إجابة الاختبار.

وقد اعتمد الباحث في هذه التجربة على عينة تتألف من (١٥٦) طالب من الطلاب المسجلين لنفس الدورة، وتم إعداد اختبار لقياس القدرة اللغوية، يتألف من (١٠٣) مفردة بأربعة بدائل، من ثلاثة صور تم تعيينها عشوائياً على الطلاب.

وقبل تطبيق الوسيلة التجريبية قام الباحث بتطبيق مقياس للقلق المرتبط بالاختبار، ثم قام بتصنيف العينة إلى ثلاث مجموعات (مرتفعة القلق ــ متوسطة القلق ــ منخفضة القلــق) وفقاً لدرجاتهم في المقياس.

وقد أسفرت نتائج التجربة الثانية عن الآتي:

١- لا توجد تأثيرات دالة إحصائياً لمتغير موقع وترتيب المفردات على آداء الطــلاب فــي الاختبار عند مستوى (٠,٠١).

٢- وجود تأثير دال إحصائياً لمتغير قلق الاختبار على أداء الطلاب الاختباري عند مستوى (٠,٠١).

٣- لا يوجد تأثير دال لمتغير ترتيب المفردات على مقدار الوقت الذي يستغرقه الطلاب في إجابة الاختبار.

وكان من أهم توصيات الدراسة ضرورة إجراء بحوث إضافية لاختبار أثــر موقــع وترتيب المفردات على صدق وثبات الاختبار.

دراسة ويليام ومونك William & Monk, 1970:

بعنوان "أثر ترتيب المفردات على درجات الاختبار". حيث هدفت الدراسة إلى اختبار إمكانية إعداد صور اختباريه متكافئة من حيث متوســطات الــدرجات الكليــة، ومعــاملات الصعوبة ومعاملات الثبات باستخدام أساليب مختلفة لعرض وترتيب المفردات.

اعتمد الباحثان في هذه الدراسة على استخدام (١١) زوجاً من أزواج الاختبارات التي تقيس المعلومات الجغرافية، والتي كانت متاحة في الفترة من (١٩٦٥- ١٩٦٨) في جامعــة إلينوى (Illinois).

وكانت غالبية هذه المفردات تركز على قياس استدعاء وتذكر الحقائق وحوالي (٢٠-٣٠%) من المفردات تركز على تفسير الإحصاءات والرسومات البيانية والخرائط.

وقد قام الباحثان بإعادة ترتيب المفردات بالاعتماد على نوع المفردة (صواب وخطأ ــ اختيار من متعدد ...إلخ)، وكانت عملية الترتيب واخــتلاف موقــع المفــردات تــتم داخــل مجموعات الأسئلة وبين المجموعات بعضها البعض.

وتم استخدام اختبار (t. test) لدلالة الفروق بين المتوسطات وذلك بهدف المقارنة بين متوسطات كل زوج من أزواج الاختبارات وفي بعض الاختبارات كانت العينات غير متساوية والتباينات غير متجانسة لذلك تم استخدام الاختبار المناسب في هاتين الحالتين، كذلك حسب معاملات الثبات باستخدام معادلة كيودر وريتشاردسون (٢٠).

وقد أسفرت نتائج الدراسة عن الآتي:

١- لا توجد فروق دالة إحصائياً بين متوسطات درجات تسعة أزواج من (١١) زوج للاختبارات المتكافئة عند مستوى معنوية (٠,٠٥).

٢- أوضحت نتائج مقارنة معاملات الثبات المحسوبة لأزواج الاختبارات المختلفة أن إعادة ترتيب مفردات الاختبار بصور مختلفة أدى إلى اختلاف معاملات الثبات بدرجة طفيفة.

٣- فيما يتعلق بأثر إعادة ترتيب المفردات على قيم معاملات الصعوبة أظهرت النتائج وجود فروق ضئيلة بين القيم المحسوبة لمعاملات الصعوبة عند اختلاف طرق الترتيب، كما أن معامل ضرب العزوم لهذه القيم بالنسبة للأشكال الاختبارية المختلفة في مواقع مفرداتها بلغ (٠,٩٥).

وقد أكد الباحثان في نهاية الدراسة على إمكانية إعداد صور اختبارات متكافئة باستخدام أساليب مختلفة لترتيب المفردات وذلك في حالة ما إذا كان ترتيب الصعوبة غير نظامي (non-systematic)، وكانت الاختبارات من نوع اختبارات القوة، كما أن معاملات الصعوبة ومعاملات الثبات لا تتأثر بشكل ملحوظ بإعادة تسلسل وترتيب المفردات بطرق متنوعة.

دراسة هك وباورز: Huck & Bowers, 1972

بعنوان "مستوى صعوبة المفردة وأثر التسلسل فى اختبارات الاختيار من متعدد التحصيلية". حيث هدفت الدراسة إلى اختبار أثر التحيز في مستوى صعوبة المفردة نتيجة لتأثرها بمستوى صعوبة المفردة السابقة لها مباشرة في اختبار الاختيار من متعدد.

وقد قام الباحثان بإجراء تجربتين متشابهتين كان الاختلاف الوحيد بينهما في اعتبارات التصميم، وبيانهما كالتالي:

التجربة الأولى:

تألفت العينة من (١٢٠) طالباً من المقيدين للامتحان النهائي لدورة علم النفس في جامعة نورث ويسترن (North Western) في العام الدراسي (١٩٦٩).

استخدم الباحثان اختباراً موضوعياً (اختيار من متعدد)، يتألف من (٦٠) مفردة لقياس المهارات الأساسية المرتبطة بالدورة، وقد قام الباحثان بإعداد (١٠) صور اختباريه متماثلة في المحتوى والاختلاف الوحيد بينهما في تسلسل وترتيب المفردات.

وقد أوضحت نتائج تحليل التباين عدم وجود تأثير دال أو معنوي لأثر ترتيب المفردات أو اختلاف موقع المفردة على التحيز في مستوى صعوبة المفردة عند مستوى معنوية (٠,٠٥).

التجربة الثانية:

وقد قام الباحثان بإجرائها بهدف إثبات وتأكيد صحة نتائج التجربة الأولى، وذلك بإعداد اختبار يتألف من (٥٠) مفردة اختيار من متعدد في امتحان نصف العام المرتبط بالدورة.

تضمنت عينة الدراسة (١٦٢) طالباً من نفس مجموعة الطلاب المقيدين لهذه الدورة، في نفس الجامعة، وبدلاً من إعداد عشر صور اختباريه، قام الباحثان بإعداد ست صور فقط متماثلة في المحتوى ولكنها تختلف من حيث ترتيب وتسلسل المفردات الخاصة بكلاً منها، قام الباحثان بتعيين صور الاختبارات عشوائياً على أفراد العينة في الوقت المحدد للامتحان.

وقد اتفقت نتائج التجربة الثانية مع نتائج التجربة الأولى في عدم وجود تأثير دال إحصائياً لإعادة ترتيب المفردات أو اختلاف مواقعها من الاختبار على التحيز في مستوى صعوبة المفردة، وذلك عند مستوى معنوية (٠,٠٥).

ويرى الباحثان أن هذه النتيجة لا يجب أن تعمم على كل التلاميذ وكل المواقف الاختباريه، وهذا بالطبع يتطلب مزيداً من الدراسات لاختبار أثر اختلاف موقع وترتيب المفردات على الأداء الاختباري وبالتالي على خصائص الاختبار السيكومترية.

دراسة هامبلتون وتروب Hambleton & Traub, 1974:

بعنوان "أثر ترتيب المفردات على الأداء الاختبارى والجهد"، وقد تمثلت أهداف الدراسة فيما يلى:

١- اختبار أثر ترتيب وتغير موقع المفردات على الأداء الاختباري.

٢- اختبار أثر ترتيب المفردات على مقدار الجهد المبذول أثناء الاختبار.

٣- اختبار أثر ترتيب المفردات على ارتفاع أو انخفاض قلق الاختبار.

تألفت عينة الدراسة من (١٠٦) طالباً في الصف الحادي عشر من طلاب المدرسة الصيفية، تم انتقاؤهم عشوائياً من اثنتين من المدارس الثانوية في (تورنتو)، وتم ترتيبهم وفقاً لمستوياتهم التحصيلية في الرياضيات.

ولتحقيق أغراض الدراسة قام الباحثان بتطبيق المقاييس التالية:

١- مقياس التحصيل: وقد اشتمل على (٣٠) مفردة في مادة الرياضيات فرع الجبر، وبعد تطبيق الاختبار حسبت معاملات السهولة والصعوبة للمفردات، وبناء على ذلك تم تقسيم أو تصنيف الاختبار إلى صورتين إحداهما تم ترتيب وعرض مفرداته من السهل إلى الصعب، والصورة الأخرى تم ترتيب وتسلسل مفرداتها من الصعب إلى السهل.

٢- مقياس الجهد: حيث قام الباحثان باستخدام مقياس النبض لقياس معدل ضربات القلب، كأفضل مؤشر فسيولوجي للجهد المبذول أثناء الاختبار.

٣- مقياس قلق الاختبار: وهو يتألف من اثنين من المقاييس المستقلة إحداها يعتبر القلق دافع ييسر من أداء الاختبار والآخر يعتبر قلق الاختبار عائق أمام الأداء الجيد.

وقد تمثلت نتائج الدراسة فيما يلي:

١- فيما يتعلق بأثر موقع وترتيب المفردات على درجات الطلاب التحصيلية أسفرت النتائج، عن ارتفاع درجات مجموعة الطلاب على صورة الاختبار الذي تم ترتيب وعرض مفرداته من السهل إلى الصعب بالمقارنة بصورة الاختبار الأخرى والذي رتبت مفرداته

من الصعب إلى السهل، حيث كانت درجات المجموعة الأولى في المتوسط أعلى دلالة ومعنوية.

٢- فيما يتعلق بأثر موقع وترتيب المفردات على الجهد المبذول أثناء الاختبار بلغ متوسط درجات الجهد بالنسبة لمجموعة الطلاب التي طبق عليها صورة الاختبار مرتبة من السهل إلى الصعب (٧٥,٢) بينما بلغ متوسط درجات الجهد بالنسبة للمجموعة الثانية والتي طبق عليها الاختبار بترتيب عكسي للمفردات، من الصعب إلى السهل (٧٦,٤٨)، وهذا يعني أن النوع الثاني لترتيب المفردات جعل الموقف الاختباري أكثر إجهاداً بالنسبة للتلاميذ بالمقارنة بالنوع الأول من الترتيب.

٣- ارتفاع مستوى قلق عينة الطلاب التي طبق عليها الاختبار بعد ترتيب مفرداته من الصعب إلى السهل بالمقارنة بالترتيب العكسي (من السهل إلى الصعب).

دراسة هيك Huck, 1978:

بعنوان "الأداء الاختبارى تحت شرط معرفة صعوبة المفردة" حيث هدفت الدراسة إلى اختبار أثر معرفة الطلاب لمستوى صعوبة المفردات وترتيبها وفقاً لمستوى صعوبتها على التحصيل وثبات الاختبار.

تم تطبيق ثلاثة اختبارات فرعية شملت مقرر الإحصاء على عينة تتألف من (٧٢) طالباً من طلاب السنة الأولى بكلية التربية جامعة تينيس (Tenness).

وقد قام الباحث بإجراء تعيين عشوائي لأفراد العينة إلى مجموعتين متساويتين، المجموعة الأولى (ن= ٣٦) توافرت لديهم معلومات عن صعوبة المفردات وترتيبها حسب مستوى صعوبتها، بينما المجموعة الأخرى (ن= ٣٦) لم يكن لديهم أي معلومات ترتبط بصعوبة المفردات أو بترتيب المفردات داخل الاختبار.

وقد توصلت نتائج الدراسة إلى أن توافر معلومات لدى الطلاب عن صعوبة المفردات وترتيبها من حيث مستوى صعوبتها، ساعد في الحصول على درجات مرتفعة في الاختبار، وبالرغم من ذلك فإن هذه المتغيرات لم تؤثر على قيمة معامل ثبات درجات الاختبار.

دراسة شريشم Chrieshem, 1980:

بعنوان "أسلوب عرض المفردة وأثره على صدق الاستبيان، تجربة ميدانية". وقد هدفت الدراسة إلى اختبار أثر أسلوب تنظيم وترتيب مفردات الاستبيان (التصنيف في مجموعات ــ التصنيف العشوائي) على الصدق التقاربي والصدق التمييزي للأداة.

تألفت عينة الدراسة من (٦٠) موظفاً، طلب منهم الإجابة على مفردات الاستبيان لقياس السلوك القيادى للمديرين.

وقد قام الباحثان بإعداد الاستبيان في صورتين وفقاً لنوع المفردات وتم تقسيم العينة إلى مجموعتين متساويتين، وزعت عليهم الاستبيانات بطريقة عشوائية، وذلك بعد ترتيب المفردات وعرضها بأسلوبين أحدهما يعتمد على تصنيف المفردات في مجموعات منظمة وفقاً لأبعاد محددة والثاني يعتمد على عرض وترتيب المفردات عشوائياً.

وقد أوضحت نتائج الدراسة فاعلية أسلوب الترتيب العشوائي لمفردات الاستبيان على الخصائص الإحصائية للأداة، حيث كانت النتائج دالة إحصائياً عند مستوى (٠,٠١) وذلك بالمقارنة بالأسلوب التصنيفي في مجموعات للمفردات داخل الاستبيان.

دراسة بلاك Plack, 1982:

بعنوان " ترتيب المفردات ومعرفة أسلوب الترتيب على درجات الاختبار". حيث هدفت الدراسة إلى فحص أثر ترتيب مفردات الاختبار بثلاثة طرق (من السهل إلى الصعب، الدائري، العشوائي) وكذلك معرفة الطلاب لمستوى صعوبة المفردة على أدائهم الاختباري.

أعد الباحث اختبار يتألف من (٩٦) مفردة اختيار من متعدد بمستويات صعوبة تتراوح بين (٠,٢٠ ــ ٠,٩٦) بمتوسط (٠,٦٥) في مقرر التمريض والطب النفسي، تم تطبيقه على عينة تتألف من (١٠٤) طالب وطالبة تم تقسيمها إلى ثلاث مجموعات تجريبية متساوية العدد تقريباً وبطريقة عشوائية، وكان معامل ثبات الاختبار بطريقة كيودر وريتشاردسون (٢٠)، (٠,٨٥).

وقد استخدم الباحث تحليل التباين ذو التصميم العاملي (٢× ٣).

وقد أسفرت النتائج على الآتي:

١- لا يوجد تأثير دال إحصائياً لترتيب مستويات الصعوبة على الأداء الاختباري (من السهل إلى الصعب ــ الدائري ــ العشوائي).

٢- لا يوجد تأثير دال إحصائياً لمعرفة الطلاب لمستوى صعوبة الأسئلة على أدائهم الاختباري.

٣- لا يوجد تأثير دال للتفاعل بين العاملين السابقين، ترتيب الأسئلة حسب مستوى صعوبتها ومعرفة الطلاب لهذا الترتيب على الأداء الاختباري للطلاب.

دراسة شير وهينركسن Schurr & Henriksen, 1983:

بعنوان "أثر تسلسل المفردات وتنظيمها فى مجموعات على الإحصاءات الخاصة بمفردات الاستبيان". وقد هدفت الدراسة إلى اختبار أثر تنظيم وترتيب المفردات في مجموعات (Item Grouping) على المتوسطات والارتباطات والتغايرات الخاصة بهذه المفردات.

تألفت عينة الدراسة من مجموعة المديرين والمشرفين التربويين في بعض المدارس الحكومية، تم تقسيمهم إلى ثلاث عينات مستقلة.

وقد اعتمد الباحثان على استبيان يتألف من (٦١) مفردة لقياس مجموعة من المهارات والأنشطة التي يجب توافرها في المعلمين بصفة عامة.

وقد قام الباحثان بتقديم وعرض مفردات الاستبيان بثلاث صور مختلفة لترتيب المفردات وهي كالتالي:

١- الصورة الأصلية (Original): وفيها تم تصنيف وتوزيع مفردات الاستبيان في سبعة مجموعات مستقلة ومنظمة.

٢- الصورة الثانية: وفيها تم ترتيب وعرض مفردات الاستبيان ولكن بدون تصنيفها في مجموعات منظمة أو مستقلة وكان يشار إليها بالرمز (UG – Ungrouped)، وبالرغم من عدم تصنيفها إلا أنها كانت مرتبة أو معروضة وفقاً لترتيب معين.

٣- الصورة الثالثة: وفيها تم عرض وتقديم المفردات لأفراد العينة بنفس الأسلوب السابق، غير مصنفة في مجموعات ولكن ترتيبها عشوائي وكان يرمز إليها بالرمز (RD-Random).

وقد أسفرت نتائج الدراسة عن الآتي:

١- وجود فروق إحصائية دالة عند مستوى (٠,٠٠١) بين صورة الاستبيان الأصلي، والذي تم تصنيف مفرداته في شكل مجموعات مستقلة وبين الصورة الأخرى (UG) والتي عرضت مفرداتها غير مصنفة في مجموعات لصالح الصورة الأولى.

٢- وجود فروق معنوية ذات دلالة إحصائية بين متوسطات مفردات الاستبيان الذي عرضت مفرداته وفقاً لترتيب معين والصورة الأخرى التي عرضت مفرداتها مرتبة بأسلوب عشوائي لصالح الصورة الأولى.

٣- لاحظ الباحثان أن الاستجابات الأكثر إيجابية ظهرت بصورة واضحة بالنسبة للمفردات في المواقع المتطرفة من الاستبيان (البداية ــ النهاية)، وذلك بالمقارنة بالمفردات في مواقع الوسط من الأداة.

٤- بمعاينة قيم الارتباطات بين المفردات اتضح أن قيم الارتباطات بين المفردات المصنفة في مجموعات لم تكن متزايدة بشكل منتظم بالمقارنة بالارتباطات الخاصة بالمفردات غير المصنفة في مجموعات.

٥- أثر تسلسل المفردات وتنظيمها في مجموعات على التغايرات الخاصة بالمفردة كان أكثر وضوحاً بالنسبة للأعداد المنخفضة من المفردات بالمقارنة بالأعداد الكبيرة للمفردات.

دراسة شيسوم Chissom, 1985:

بعنوان "اختبار العلاقة بين ترتيب المفردات والأداء الاختباري". حيث هدفت الدراسة إلى اختبار وفحص أثر اختلاف ترتيب مفردات الاختبار على مستوى آداء الطلاب وقيمة معامل ثبات الاختبار.

تألفت عينة الدراسة من (٩٠٠) طالب من الخريجين المقيدين لدورة البحث التربوي في جامعة آلاباما.

قام الباحث بإجراء التجربة أربع مرات باستخدام أربعة اختبارات مختلفة، وتم الاعتماد على ثلاثة أساليب متنوعة لترتيب مفردات الاختبار الأصلي والذي تألف من (٥٠) مفردة اختيار من متعدد كالآتي:

(أ) ترتيب المفردات بأسلوب متزايد تدريجياً في الصعوبة.

(ب) ترتيب المفردات بأسلوب منخفض تدريجياً في الصعوبة.

(جـ) الترتيب العشوائي للمفردات.

وقد تراوحت معاملات صعوبة المفردات بين (٠,٢٥ ــ ١).

وتراوحت معاملات ألفا للثبات بين (٠,٦٣ ــ ٠,٨٢).

ولم تستخدم معاملات التصحيح من أثر التخمين، وكذلك لم يكن هناك وقت محـدد أو زمن معين لإجابة الاختبار، ومفردات الاختبار كانت تمثل جميـع الموضـوعات المقـررة للدراسة.

وقد أسفرت الدراسة عن النتائج التالية:

١- أوضحت نتائج المعالجة الإحصائية الوصفية عدم وجود فروق أساسية بين متوسـطات الدرجات عبر مرات تكرار التجربة.

٢- أوضحت نتائج تحليل التباين أن متوسطات المجموعات لم توجـد بينهـا فـروق دالـة إحصائيا.

٣- اختلاف أساليب ترتيب مفردات الاختبار بالاعتماد على مستوى الصعوبة لا ترتبط بقوة الأداء على الاختبار، سواء بالمعنى الإيجابي أو السلبي.

ويرى الباحث أنه بالرغم من عدم وجود تأثير فعال لاختلاف مواقع ترتيب المفردات على الأداء الاختباري، إلا أنها يمكن أن تكون أسلوب إيجابي في تجنب الغـش مـن بعـض الطلاب أثناء أداء الاختبار.

دراسة نيومان وآخرون Newman & et al., 1988:

بعنوان "أثر تنوع ترتيب المفردات من حيث مستوى الصعوبة (الإحصائية والمعرفية) على درجات اختبار الاختيار من متعدد". وقد هدفت الدراسة إلى اختبار التأثير المزدوج لكلاً من الصعوبة الإحصائية والصعوبة المعرفية على الأداء الاختباري للطلاب.

اعتمدت الدراسة على عينة تتألف من (١٢٠) طالب وطالبة يدرسون مقرر علم النفس التربوي (٢٦% ذكور، ٧٤% إناث).

وقد قام الباحثين بانتقاء (٤٠) مفردة من بنك أسئلة اختيار من متعدد في مـــادة علـــم النفس التربوي تتراوح صعوبتها بين (٠,١٩ ــ ٠,٩٥). بمتوسط صعوبة (٠,٦٣)، ثم قـــاموا بتصنيفها إلى ثلاثة مستويات معرفية وفقاً لتصنيف بلوم (١٥ مفردة مســـتوى المعرفـــة، ١١ مفردة في مستوى الفهم ١٤ مفردة في مستوى التطبيق).

تم تصنيف هذه المفردات في أربعة صور اختبارية مختلفة من حيث تسلسل وتنظيـــم المفردات وبيانها كالتالي:

١- الصورة الأولى تكون مفرداتها متزايدة في صعوبتها الإحصائيـــة (مـــن الســـهل إلـــى الصعب)، وداخل كل مستوى للصعوبة تصنف معرفياً (معرفة ــ فهم ــ تطبيق).

٢- الصورة الثانية تكون مفرداتها متزايدة تدريجياً في الصعوبة المعرفية (معرفة ــ فـــهم ــ تطبيق) وداخل كل مستوى معرفي يتم تصنيفها إحصائياً من السهل إلى الصعب.

٣- الصورة الثالثة مفرداتها تتناقص تدريجياً من حيث الصعوبة المعرفية (تطبيق ــ فـــهم ــ معرفة)، وداخل كل مستوى معرفي تصنف وفقاً لصعوبتها الإحصائية (سهل ــ متوســـط ــ صعب).

٤- الصورة الرابعة، تكون مفرداتها متزايدة تدريجياً من حيث الصعوبة الإحصائية، وفـــي داخل كل مستوى تتناقص تدريجياً في صعوبتها المعرفية (تطبيق ــ فهم ــ معرفة).

وقد توصلت الدراسة إلى النتائج التالية:

١- عدم وجود تفاعل دال بين الصعوبة الإحصائية والصعوبة المعرفيـــة بالنســـبة لجميـــع المفردات على الصورة الأصلية للاختبار.

٢- بالنسبة للاختبارات الفرعية:

أ- وجود فروق دالة إحصائياً في الصعوبـــة الإحصائيـــة عنـــد مســـتوى (٠,٠٥) وباستخدام المقارنات المتعددة كانت الفروق لصالح المستوى المتوسط للصعوبة.

ب- وجود فروق دالة إحصائياً في الصعوبة المعرفية عند مســـتوى (٠,٠٥) لصـــالح مستوى التطبيق.

ج- وجود تفاعل دال بين الصعوبة الإحصائية والصعوبة المعرفيـــة عنـــد مســـتوى (٠,٠٥).

د -حصل الطلاب الذين تلقوا مفردات ذات ترتيب معرفي متزايد (تذكر ــ فهم ــ تطبيق) على درجات أعلى بغض النظر عن الصعوبة الإحصائية للمفردات (سهل ــ متوسط ــ صعب) عند مستوى (٠,٠٥).

هـ - حصل الطلاب الذين تلقوا مفردات ذات ترتيب إحصائي متزايد تدريجياً في صعوبته (سهل ــ متوسط ــ صعب) على درجات أفضل وخاصة بالنسبة للمستوى المتوسط الصعوبة بغض النظر عن الصعوبة المعرفية عند مستوى (٠,٠٥).

دراسة ستيفن Steven, 1993:

بعنوان "أثر ترتيب المفردات في مجموعات على الثبات ودرجات المقاييس". حيث هدفت الدراسة إلى اختبار أثر وضع أو موقع المفردة من المقياس (التصنيف في مجموعات ــ التصنيف العشوائي) على قيمة معامل ثبات الاتساق الداخلي لألفا، ومتوسطات المقياس وأثر ذلك على فاعلية عملية القياس.

تألفت عينة الدراسة من (٩٩٢) مدير مدرسة حكومية، تم انتقاؤهم من بين (٢٦٢٨) مدير مدرسة بولاية بنسلفانيا بمساعدة الإدارة التعليمية للتربية.

وقد اعتمد الباحث على استخدام مقياس تقييم الاتجاهات الحكومية نحو عدم الرضا الوظيفي عن المعلمين، يتألف من (٤١) مفردة تم توزيعها على ستة مقاييس فرعية وهي:

١- القيادة التعليمية
٢- الوسائل التعليمية المساعدة.
٣- الشهادات والمؤهلات الدراسية.
٤- تنظيم وقت الحصة.
٥- العوامل النفسية.
٦- فعالية الإدراك والملاحظة.

تم تطبيق المقياس على أفراد العينة عن طريق تقسيمها إلى مجموعتين، في المجموعة الأولى طبقت الأداة بعد تصنيف المفردات وتنظيمها في مجموعة متجانسة، وفي المجموعة الثانية تم تطبيق المقياس ولكن بأسلوب عشوائي في ترتيب مفرداته أي في شكل مجموعات غير متجانسة من المفردات.

وقد قام الباحث بحساب معاملات الثبات، لكل المقاييس الفرعية باستخدام معادلة ألفا لقياس التجانس الداخلي.

كذلك قام بتطبيق اختبار (t. test) لتحديد مدى دلالة الفروق بين متوسطات درجات كلا المجموعتين في كل مقياس فرعي وبالنسبة لكل أسلوب من أساليب ترتيب مفردات المقياس (مصنفة ــ عشوائية).

وقد توصلت الدراسة إلى النتائج التالية:

١- وجود فروق معنوية عند مستوى (٠,٠١) بين المتوسطات الخاصة بمقياس القيادة التعليمية ومقياس العوامل النفسية وذلك بالنسبة لكلا الأسلوبين في ترتيب المفردات وكانت الفروق لصالح الأسلوب التنظيمي أو التصنيفي في عرض مفردات المقياس بالمقارنة بالأسلوب العشوائي وذلك بالنسبة لمقياس القيادة التعليمية، بينما كانت الفروق دالة لصالح الأسلوب العشوائي في عرض المفردات في مقياس العوامل النفسية.

٢- بالنسبة للفروق بين المتوسطات الخاصة ببقية المقاييس الفرعية الأخرى، لم تكن دالة إحصائياً، ولكن كان هناك نزعة أو اتجاه ملحوظ في ارتفاع متوسطات درجات المقاييس التي عرضت مفرداتها بأسلوب عشوائي.

٣- فيما يتعلق بأثر وضع أو موقع المفردات داخل المقياس على قيمة معامل ثبات الاتساق الداخلي، أوضحت النتائج أن تقديرات الثبات كانت لصالح المقاييس التي تم ترتيب مفرداتها بشكل تصنيفي داخل مجموعات متجانسة.

تعقيب الباحثة على البحوث والدراسات المرتبطة بالمحور الثاني (أثر موقع أو ترتيب المفردات داخل الاختبار على قيمة معامل الثبات).

بملاحظة نتائج الدراسات والبحوث السابقة والمتعلقة بدراسة أثر اختلاف موقع المفردة من الاختبار أو ترتيب المفردات داخل الاختبار على خصائص الاختبار السيكومترية ومن بينها معامل الثبات، اتضح اختلاف وجهات نظر الباحثين حول الأسلوب الأمثل عند تقديم المفردات وترتيبها داخل الاختبار، والذي بواسطته يمكن الحصول على نتائج أفضل فيما يتعلق بأداء الطلاب الاختباري وأيضاً خصائص الاختبارات السيكومترية (الصعوبة ــ الثبات ــ قوة التمييز)، وسوف تتناول الباحثة عرض نقاط الاتفاق والاختلاف فيما بينها وفقاً للعناصر التالية:

أولاً: اتفقت نتائج دراسات كل من Cromack & Sax, 66، ودراسة Flaugher & etal., 68، ودراسة Hambleton & Ross, 74، ودراسة Newman & etal., 88، على وجود فروق دالة في متوسطات درجات الطلاب التحصيلية، وكذلك في قيم معاملات الثبات ومعاملات الصعوبة والتمييز، كنتيجة لاختلاف طرق ترتيب وتنظيم المفردات

داخل الاختبار، وكانت هذه الفروق لصالح الترتيب المعياري لعرض المفردات والمتزايد تدريجياً في صعوبتها (من السهل إلى الصعب)، بالمقارنة بأساليب الترتيب الأخرى ومنها الأسلوب العشوائي وكذلك أسلوب عرض المفردات بانخفاض تدريجي في صعوبتها (من الصعب إلى السهل).

ثانياً: توصلت نتائج دراسات كل من Brenner, 64، ودراسة Somuse & Munz, 68، ودراسة Marso, 70، ودراسة William & Monk, 70، ودراسة Norman & Huck, 72، ودراسة Husck, 78، ودراسة Plack, 82، إلى عدم وجود تأثير دال إحصائياً على متوسطات الدرجات، وكذلك على قيم معامل الثبات نتيجة اختلاف أساليب ترتيب مفردات الاختبار (من السهل إلى الصعب، الترتيب العشوائي، والترتيب من الصعب إلى السهل).

ثالثاً: تناولت بعض الدراسات معالجة أثر بعض طرق ترتيب مفردات المقياس على خصائصه السيكومترية ومنها طريقة تصنيف المفردات في مجموعات متجانسة وطريقة التصنيف العشوائي، فقد اتفقت نتائج دراسات كلاً من Chester, 80، ودراسة Steven, 93 على فاعلية أسلوب تصنيف مفردات المقياس في مجموعات متجانسة على الصدق التقاربي والصدق التمييزي وكذلك على ثبات الاتساق الداخلي لدرجات المقياس، وتناقضت معهم دراسة Terry & Henriksen, 83، حيث توصلت إلى وجود تأثير دال على المتوسطات والارتباطات الخاصة بالمفردات وكانت الفروق لصالح الترتيب العشوائي.

رابعاً: حاولت بعض الدراسات اختبار أثر متغير موقع أو ترتيب المفردات في علاقته بمتغير قلق الاختبار على الأداء الاختباري والخصائص السيكومترية للأداة، فقد اتفقت نتائج دراسة كل من Marso, 70، ودراسة Smouse & Munz, 68، على أن اختلاف طرق ترتيب وعرض المفردات داخل الاختبار ليس له تأثير دال إحصائياً على الأداء الاختباري، وأنه لا يوجد تفاعل دال بين متغير موقع المفردة أو ترتيب المفردات وأنواع قلق الاختبار، وتناقضت معهم دراسة Hambleton & Ross, 74، فقد توصلت إلى ارتفاع متوسط درجات الطلاب على الاختبار عندما تم ترتيب مفرداته من السهل إلى الصعب، بالمقارنة بالأسلوب العكسي في عرض وترتيب المفردات (من الصعب إلى السهل)، بالإضافة إلى أن الأسلوب الأخير في عملية الترتيب للمفردات أدى إلى ارتفاع مستوى قلق الاختبار لدى عينة الطلاب.

وترى الباحثة أن من أهم الأسباب التي تدعوا إلى ضرورة الاهتمام بدراسة أثر متغير موقع المفردات وترتيبها داخل الاختبار على الأداء الاختباري للطلاب وبالتالي على قيمة معامل الثبات، هو ندرة الدراسات العربية حول هذا الموضوع، ففي حدود علم الباحثة لا توجد دراسة عربية واحدة اهتمت بمعالجة أثر هذا المتغير، على الرغم من التوسع في استخدام الاختبارات التحصيلية الموضوعية ومن أهمها اختبارات الاختيار من متعدد في عمليات القياس، بالإضافة إلى لجوء عدد كبير من القائمين بتطبيق هذا النوع من الاختبارات إلى محاولة تغيير وتبديل مواقع المفردات داخل الاختبار تجنباً لمحاولات الغش من جانب بعض الطلاب دون الأخذ في الاعتبار أن اختلاف ترتيب هذه المفردات من الممكن أن يكون له تأثير دال إحصائياً على خصائص الاختبار السيكومترية ومنها معامل الثبات، وهذا ما تهدف الدراسة إلى التحقق منه.

المحور الثالث: البحوث والدراسات التي تناولت أثر متغير طول الاختبار (عدد المفردات) على قيمة معامل ثبات القياس:

دراسة إيبل Ebel, 1969:

بعنوان "الثبات المتوقع كدالة فى طول الاختبار وعدد الاختيارات لكل مفردة". حيث هدفت الدراسة إلى تقدير معامل الثبات المتوقع كدالة في طول الاختبار (عدد المفردات)، وعدد البدائل أو الاختيارات لكل مفردة.

قام الباحث بإعداد اختبار موضوعي من نوع أسئلة الاختيار من متعدد يتألف من (١٠٠) مفردة، وكذلك استخدام أعداد مختلفة من بدائل الإجابة (٢/ ٣/ ٤/ ٦) بدائل، كذلك كانت الفروق بين أعداد المفردات التي استخدمها الباحث مرتفعة في البداية، انخفضت بعد ذلك بالتدريج.

أوضح الباحث أنه عندما كانت أعداد بدائل الاختيار لكل مفردة ثابتة، كانت أعداد المفردات المستخدمة متغيرة والعكس صحيح، وقد استخدمت معادلة كيودر وريتشاردسون (٢١) في حساب معاملات الثبات.

وقد توصلت الدراسة إلى النتائج التالية:

١- وجود فروق ذات دلالة إحصائية في قيم معاملات الثبات لصالح الاختبارات الأكثر طولاً أو التي تتميز بارتفاع عدد مفرداتها حيث ارتفع معامل الثبات من (٠,٤٨) إلى (٠,٨٦) عندما زاد طول الاختبار من (٥٠) إلى (٢٠٠) مفردة.

٢- وجود فروق دال إحصائياً بين قيم معاملات الثبات عندما زاد عدد بدائل الفقرة من اثنين إلى ثلاثة بدائل بينما كانت الفروق منخفضة عندما تم استخدام أربعة بدائل فأكثر.

٣- أوضحت النتائج أن اختبار الصواب والخطأ الذي يتألف من (١٠٠) مفردة حقق معامل ثبات قدره (٠,٧٤) عندما تم الاعتماد على بديلين للإجابة، في حين زادت هذه القيمة إلى (٠,٨٦) مع زيادة عدد البدائل إلى أربعة.

٤- أن التلميذ يمكنه الإجابة على مفردتين صواب وخطأ ببديلين للإجابة في وقت أقل من الوقت المحدد للإجابة على مفردة اختيار من متعدد بأربعة بدائل.

دراسة واجنر Wagner, 1990:

بعنوان "الاختبار التجريبى لاستقرار الارتفاع فى قيمة معامل الارتباط كتقدير لثبات الاتساق الداخلى للاختبارات بأعداد منخفضة من المفردات". حيث هدفت الدراسة إلى الاختبار التجريبي لاستقرار الارتفاع في قيمة الارتباط كتقدير لمعامل ثبات الاتساق الداخلي للاختبارات ذات الأحجام الصغيرة من المفردات (الاختبارات القصيرة).

تألفت عينة الدراسة من (١٠٠) عينة اشتمل كلاً منها على (٥٠) حالة سحبت عشوائياً من المجتمع بإرجاع.

اعتمد الباحث على استخدام اختبار الشخصية الاسقاطي (بقع الحبر لرورشاخ) كمتغير تجريبي لدراسة الاستقرار في ارتفاع معاملات الارتباط بين المفردات كمقدر للثبات.

وقد قام الباحث بحساب معاملات ألفا للثبات ومعامل الارتباط الأقصى (r-max) ومعامل التجزئية النصفية كتقديرات للثبات بالنسبة لكل عينة وتمت المقارنة بين المقاييس الثلاثة المحسوبة، كذلك قام الباحث بالمقارنة بين توزيع قيم ألفا المحسوبة في هذه الدراسة والتوزيع المتوقع وكانت النتائج متشابهة إلى حد كبير.

وقد أوضحت نتائج الدراسة أن أسلوب الارتباط الأقصى (r-max) كإجراء معياري لتقدير الاتساق الداخلي للاختبارات ذات الأعداد المنخفضة من المفردات أكثر ثباتاً واستقراراً (٠,٨٠٥)، يليه في الترتيب معاملات ألفا (٠,٦٩٦) لأنها أقل احتمالاً في إعطاء تقديرات منخفضة للثبات، بالإضافة إلى أن خصائص توزيعه معروفة، وقد جاء أسلوب التجزئة النصفية في الترتيب الأخير (٠,٦٠٨) لأنه يعطى تقديرات منخفضة للثبات، وخاصة بالنسبة لاختبارات الشخصية شديدة التأثر بخاصية عدم التجانس بين مفرداتها، بالإضافة إلى أن هذا الأسلوب في تقدير الثبات يتأثر بدرجة كبيرة بطول أداة القياس (عدد المفردات).

دراسة ميريل Merrill, 1993:

بعنوان "العلاقة بين الصدق والثبات وطول الاختبار فى مقاييس الشخصية شائعة الاستخدام". حيث هدفت الدراسة إلى اختبار العلاقة بين خصائص المقياس السيكومترية (الصدق والثبات) وطول المقياس بالنسبة لمقاييس الشخصية شائعة الاستخدام، والتحقق مما إذا كانت هذه العلاقة يمكن أن تصل إلى الحد الأدنى أو تكون سالبة في ارتباطها بطول المقياس، وكذلك اختبار مدى تأثر هذه العلاقة بطريقة بناء الاختبار.

تألفت عينة الدراسة من (١٧٩) طالباً جامعياً، تم تقسيمهم عشوائياً إلى مجموعتين، تألفت المجموعة الأولى من (٩٠) طالباً والثانية من (٨٩) طالباً، وكانت المجموعة الثانية تستخدم للتحقق من صدق النتائج المستمدة من المجموعة الأولى.

اعتمد الباحث على استخدام خمسة مقاييس فرعية للشخصية والتي تم انتقاؤها لتعكس ثلاث استراتيجيات أولية لبناء الاختبار، وتم حساب معاملات الصدق والثبات للمقاييس الفرعية على التوالي وبتتابع أطول المقاييس والتي امتدت في الطول من مفردة واحدة وحتى الطول الكامل للمقياس الأصلي.

وقد أظهرت نتائج الدراسة أنه بالنسبة للمقاييس الفرعية الخمسة ولكل طريقة من طرق تنظيم وترتيب المفردات (الخبراء، معاملات الانحدار، معاملات الارتباط، والترتيب العشوائي) فإن الفروق بين معاملات الصدق للمقاييس ككل لم تكن معنوية بالمقارنة بالطريقة الأصلية في انتقاء المفردة، وقد تحققت نفس النتائج مع المجموعة الثانية.

كذلك أوضحت نتائج الدراسة أن معاملات الثبات بصفة عامة ارتبطت بطول المقياس، وبالنسبة للمقاييس الفرعية حققت معامل ثبات مرتفع بالمقارنة بالمقياس الأصلي الكامل.

دراسة ليون وآخرون Leon & et al., (1995):

بعنوان "مقاييس تحصيل أكثر ثباتاً يمكن أن تخفض من متطلبات حجم العينة". وقد تمثلت أهداف الدراسة في الآتي:

١- اختبار أثر عدد مفردات المقياس (طول المقياس) على مقدار التشتت داخل المجموعات (تباين الخطأ) وحجم الأثر بين المجموعات وكذلك ثبات أداة القياس.

٢- توضيح العلاقة بين الثبات والقوة الإحصائية تجريبياً.

٣- اختبار أثر جودة الخصائص السيكومترية للمقياس (الصدق والثبات) على تخفيض شروط ومتطلبات حجم العينة.

تألفت عينة الدراسة من (١٠٢) حالة، انقسمت إلى مجموعتين مجموعة العلاج العملي (٧٢) حالة، ومجموعة العلاج بالمهدئات (٣٠) حالة.

اعتمدت الدراسة على استخدام مقياس تقدير الانخفاض في مستوى النشاط الوظيفي والقوى الحيوية للجسم لهاملتون، يتألف من (٢٤) مفردة، ولتوضيح العلاقة بين الثبات والقوة الإحصائية تجريبياً تم اشتقاق مقاييس ذات أطوال مختلفة، تمثلت في (٢٠) مقياس فرعي مستقل امتدت في الطول من (٤) مفردات وحتى (٢٣) مفردة، وقد تم انتقاء مفردات المقاييس الفرعية عشوائياً باستخدام جداول الأعداد العشوائية.

وقد تم حساب المتوسطات والانحرافات المعيارية المشتركة والفروق بين متوسطات المجموعات وكذلك حجم الأثر بين المجموعات وذلك بالنسبة للمقاييس الفرعية وكذلك للمقياس الأصلي الذي يتألف من (٢٤) مفردة.

وقد توصلت الدراسة إلى النتائج التالية:

١- بزيادة عدد مفردات المقاييس الفرعية، اتجه التشتت داخل المجموعات (تباين الخطأ) إلى الانخفاض، وارتفعت قيم معامل الثبات والقوة الإحصائية للاختبار، فقد زادت قيمة معامل الثبات من (٠,٥) بالنسبة للمقياس الفرعي الذي يتألف من (٤) مفردات إلى (٠,٨٨) للمقياس الكلي الذي يتضمن (٢٤) مفردة.

٢- وجود علاقة إيجابية بين قيمة معامل الثبات وحجم الأثر، فعندما زاد ثبات القياس بالنسبة للمقاييس الفرعية، زاد بالتالي حجم الأثر بين المجموعات، حيث امتد من (٠,٥) إلى (٠,٦٦) ونتيجة لذلك ارتفعت القوة الإحصائية للمقاييس الفرعية.

٣- فيما يتعلق بحجم العينة المناسب لتحقيق القوة الإحصائية المرغوب فيها فقد اختلف نتيجة اختلاف عدد المفردات داخل المقاييس الفرعية، حيث امتد من (٧٢) مفحوص لكل مجموعة بالنسبة للمقياس الذي يتألف من (٤) مفردات إلى أقل من (٥٠) مفحوص بالنسبة لكل مقياس فرعي اشتمل على أكثر من (١٦) مفردة.

دراسة فيتزباتريك و ين Fitzpatrick, & Yen 2001:

بعنوان "أثر طول الاختبار وحجم العينة على الثبات ومعادلة الاختبارات التى تتألف من المفردات بنائية الاستجابة". حيث هدفت الدراسة إلى فحص واختبار أثر كل من طول الاختبار وحجم العينة على قيمة معامل الثبات وتعديل درجات الاختبارات التي تتألف من المفردات ذات الاستجابات البنائية Constructed Response Items.

قام الباحث بمساعدة الحاسب الآلي بتوليد اختبارات تقيس مهارات الرياضيات تتألف من أعداد مختلفة من المفردات (٢، ٤، ٨، ١٢، ٢٠) مفردة وكذلك بأعداد مختلفة من النقاط أو مستويات الدرجة لكل مفردة (٢ ــ ٤ ــ ٦) نقاط.

اعتمد الباحث على عينات من أحجام مختلفة من المختبرين (٢٠٠، ٥٠٠، ١٠٠٠) حالة لتحقيق أغراض الدراسة.

وقد قام الباحث ببناء ثلاثة صور اختبارية (A, B, C) وبالنسبة لصورتي الاختبار (A, B) يمثلان صورتين لاختبار واحد تم تطبيقه على نفس العينة، بينما صورة الاختبار (C) تم تطبيقها على عينة جديدة من المختبرين ولكنها متكافئة تماماً مع العينة التي طبق عليها الاختبار (A).

وكان هدف الباحث من ذلك هو تعديل الدرجات الملاحظة للمفحوصين على صورتي الاختبار (A, C) ثم تحويل درجات أفراد العينة على الصورة (C) إلى المقياس (A) ومقارنة الدرجات الحقيقية مع الدرجات المحولة أو المعدلة بهدف تقييم أثر طول الاختبار وحجم العينة على دقة تعديل درجات الاختبار.

وقد أسفرت نتائج الدراسة عن الآتي:

١- استخدام عينة تتألف من (٢٠٠) مفحوص نتج عنه ارتفاع ملحوظ في جذر فروق متوسط المربعات، وارتباطات أقل بين الصور المختلفة للاختبار وذلك بالمقارنة بأحجام العينات (٥٠٠، ١٠٠٠) مفحوص.

٢- زادت درجة الفروق بين نتائج القياس بالنسبة للعينات المختلفة عندما زادت أعداد المفردات على صور الاختبار المستخدمة وكانت هذه الفروق لصالح الأعداد الكبيرة من المفردات.

٣- أوضحت النتائج أن قيم معاملات الارتباط بين الدرجات الحقيقية للمختبرين على صور الاختبار (A, B) كمقدر للثبات تحسنت بشكل ملحوظ مع الزيادة في عدد المفردات وكذلك الزيادة في عدد نقاط أو مستويات الدرجة لكل مفردة.

٤- أوضحت النتائج أن الاختبارات التي بنيت باستخدام (١٢) مفردة، (٢٠) مفردة كانت نتائجها أفضل وأكثر دقة وثباتاً من الاختبارات التي تألفت من (٢ ــ ٤ــ ٨) مفردة.

٥- فيما يتعلق بأثر طول الاختبار وحجم العينة على دقة تعديل درجات الاختبارات التي تتألف من مفردات بنائية الاستجابة، أوضحت النتائج أن دقة تعديل الدرجات تتحقق عند استخدام (٨) مفردات بستة (٦) نقاط للدرجة لكل مفردة، أو على الأقل (١٢) مفردة بأربعة نقاط للدرجة وأن الزيادة في الدقة تتطلب زيادة عدد المفردات.

تعقيب الباحثة على نتائج الدراسات المرتبطة بالمحور الثالث (أثر طول الاختبار على قيمة معامل الثبات):

يتضح من مراجعة نتائج الدراسات السابقة والتي اهتمت بمعالجة أثر طول الاختبار (عدد المفردات) على قيمة معامل الثبات، اتفاق الباحثين بصفة عامة على مدى التأثير الإيجابي لزيادة عدد مفردات أداة القياس على دقة نتائج القياس وبالتالي تحسين قيمة معامل الثبات.

وجدت بعض الدراسات التي حاولت اختبار أثر طول الاختبار في علاقته بمتغيرات أخرى على قيمة معامل الثبات، مثل دراسة Ebel, 69 والتي حاولت اختبار أثر عدد بدائل الفقرة وطول الاختبار على قيمة معامل الثبات، ودراسة كل من Leon & et al., 95، ودراسة Fitzpatrick, 2001 والتي حاولت اختبار أثر طول الاختبار وحجم العينة على قيمة معامل الثبات.

المحور الرابع: البحوث والدراسات التي تناولت معالجة أثر حجم العينة على قيمة معامل الثبات:

دراسة إنترابراسيرت Intraprasert, 1986:

بعنوان "اختبار ثبات خمسة طرق تستخدم فى اكتشاف تحيز مفردات الاختبار". حيث هدفت الدراسة إلى فحص واختبار ثبات خمسة طرق أو أساليب إحصائية تستخدم في اكتشاف تحيز مفردات الاختبار.

اعتمدت الدراسة على عينات ذات أحجام مختلفة (٢٠٠/ ٥٠٠/ ١٠٠٠) مفحوص وكذلك كانت هذه العينات من أجناس مختلفة وقد قام الباحث بتقسيمها إلى أربع مجموعات مختلفة لتحقيق أهداف الدراسة.

وقد استخدم الباحث اختباراً للمفردات اللغوية يتألف من (٥٠) مفردة.

وقد قام الباحث بحساب الصدق التقاربي بحساب معاملات ارتباط الرتب لسبيرمان، وكذلك تم حساب النسب المئوية للاتفاق بين الأربعة طرق الأكثر فاعلية في اكتشاف تحيز المفردات بين كل الأزواج الممكنة من الطرق، وقد تم حساب معاملات الثبات للطرق المختلفة.

وقد توصلت الدراسة إلى النتائج التالية:

١- ارتفاع في قيمة معامل ثبات الاتساق بين الطرق المستخدمة في تماثل المفردات في التحيز، والنسب المئوية للاتفاق كانت تزيد عن ٧٥% وذلك عندما اعتمدت الدراسة على عينة تتألف من (١٠٠٠) فرد.

٢- انخفاض الاتفاق بين هذه الطرق أو انخفاض قيمة معامل ثبات الاتساق عندما تم الاعتماد على عينات تتألف من (٢٠٠)، (٥٠٠) فرد وأنه باستخدام معامل الاتفاق لكندال زاد اتساق الترتيب مع زيادة حجم العينة.

٣- معاملات الارتباط كانت تزيد عن (٠,٦٥٤) بالنسبة لكل الطرق المستخدمة وكل أحجام العينات باستثناء طريقة واحدة.

دراسة كوبرايف وشاو Kopriva & Shaw, 1991:

بعنوان "تقديرات القوة: أثر ثبات المتغير التابع على قوة تحليل التباين أحادى العامل". حيث هدفت الدراسة إلى اختبار درجة تأثير ثبات المتغير التابع على قوة تحليل التباين

الأحادي باستخدام عينتين أو ثلاثة مع الأخذ في الاعتبار حجم العينة ومستوى الدلالة الإحصائية وأثره على الدرجة الحقيقية في الاختبار.

استخدم الباحثان (١٠) مستويات لمعامل الثبات وهي (٠,١، ٠,٠٢، .....١) وكذلك (٤) أربعة أحجام للدرجة الحقيقية وهي (٠,١ع، ٠,٢٥ع، ٠,٥ع، ١ع) حيث (ع) هي الانحراف المعياري للدرجة الحقيقية.

كذلك اعتمد الباحثان على (٨) ثمانية أحجام مختلفة للعينة تتراوح فيما بين (٥ إلى ٤٠٠) لكل مجموعة.

وقد تمت الإجراءات الإحصائية باستخدام اختبارات الذيل الواحد عند أربعة مستويات للدلالة الإحصائية وهي (٠,٠١، ٠,٠٢٥، ٠,٠٥، ٠,١٠).

وكان من أهم نتائج الدراسة: أن للثبات أثراً كبيراً على قوة الدلالة الإحصائية لتحليل التباين الأحادي للاختبارات التي يستخدمها الباحث، وخاصة عندما يكون حجم العينة أقل من (١٠٠) فرد لكل مجموعة، وكذلك يكون من المتوقع أن تكون الفروق بين حجم التأثير صغيرة أو متوسطة القيمة أو كلا المتغيرين معاً (حجم العينة وحجم الأثر).

دراسة فيلدت وأنكينمان Feldt & Ankenmann, 1998:

بعنوان "حجم العينة المناسب للمقارنة بين معاملات ألفا للثبات". حيث هدفت الدراسة إلى توجيه الباحثين في مواقف القياس المختلفة إلى كيفية تحديد حجم العينة المناسب والذي يجب استخدامه للمقارنة بين قيمتين من قيم معاملات ألفا للثبات.

وقد قام الباحثان في بداية الدراسة باستعراض مجموعة من المعادلات الرياضية المرتبطة بإحصائي الاختبار وتوزيع المعاينة الخاص به وأوضحوا أنه في معظم مواقف القياس يفضل أن تتجاوز قوة الاختبار الإحصائي من الجودة (٠,٩) ولكي يمكن تحقيق ذلك يجب زيادة أحجام العينات التي نقارن معاملات الثبات الخاصة بها.

وبهدف التخفيف من صعوبة المعادلات الرياضية التي يمكن استخدامها في الوصول إلى مستوى قوة مناسبة للاختبار باختيار قيمة تجريبية لحجم العينة، قام الباحثان بإعداد مجموعة من المنحنيات البيانية يطلق عليها منحنيات القوة "Power Curves" والتي تعكس العلاقة الوظيفية بين حجم العينة وحجم الأثر ومستوى قوة الاختبار عند مستوى معنوية محدد.

وهذه المنحنيات تم إعدادها لأحجام من العينات تتراوح فيما بين ٣٠ مفحوص وحتى ٥٠٠ مفحوص عند مستويات مختلفة من قوة الاختبار الإحصائي (٠,٥، ٠,٩، ٠,٧٥، ٠,٩٥)

وأوضحوا أنه بمجرد تحديد مستوى القوة المرغوب فيه يمكن للباحث أن يحدد حجم العينة المناسب لتحقيق هذا المستوى، بمعلومية قيمة حقيقية لحجم الأثر ومستوى معنوية محدد.

وقد أكد الباحثان في نهاية الدراسة على أنه من غير المرغوب فيه استخدام أحجام من العينات أقل من (٣٠) فرد لأن ذلك يضعف من إمكانية اكتشاف فروق دالة بين قيم معاملات ألفا للثبات، وأن الحد الأدنى لمستوى قوة الاختبار هو (٠,٥) وإذا رغب الباحث في زيادة هذا المستوى يجب عليه زيادة حجم العينة المستخدم لأن قوة الاختبار دالة في حجم العينة.

دراسة فان وبنج Fan & Ping, 2001:

بعنوان "خصائص العينة وثبات القياس". حيث هدفت الدراسة إلى فحص واختبار مدى تأثير خصائص عينة المختبرين على قيمة معامل الثبات وتباين درجات الخطأ.

تألفت عينة الدراسة من مجموعتين كبيرتين من الطلاب المختبرين تم توزيعهم كالآتي:

١- المجموعة الأولى تألفت من (٥٠٠٠) طالب تم اختبارهم في المهارات الأساسية الأكاديمية بأسلوب التقييم محكي المرجع.

٢- المجموعة الثانية تألفت من (١٠٠٠٠) طالب تم اختبارهم في المهارات الأكاديمية بأسلوب التقييم معياري المرجع.

وقد توصلت الدراسة إلى النتائج التالية:

١- تشتت كلا المجموعتين بالنسبة للسمة التي يقيسها الاختبار، وكذلك مستوى الأداء لكل مجموعة كان له تأثير دال على قيمة معامل الثبات.

٢- اتجهت أخطاء القياس إلى الانخفاض مع الزيادة في حجم العينة ومستوى قدراتها، وذلك بالمقارنة بالمجموعة منخفضة الحجم والقدرة.

وتتفق الباحثة مع هذه النتيجة، حيث أن الانخفاض في مستوى أداء العينة فيما يتعلق بالسمة التي يقيسها الاختبار يؤدي إلى انخفاض في قيمة تباين الدرجات الحقيقية للطلاب، ومن ثم ارتفاع في قيمة تباين درجات الخطأ (خطأ القياس) مما يؤثر سلبياً على قيمة معامل الثبات.

تعقيب الباحثة على نتائج البحوث والدراسات المرتبطة بالمحور الرابع (أثر حجم العينة على قيمة معامل ثبات القياس):

على الرغم من مدى تأثير حجم العينة على دقة نتائج القياس وخاصة تباين درجات الخطأ عند تطبيق الأداة، إلا أن هناك ندرة شديدة من جانب الدراسات والبحوث سواء كانت عربية أو أجنبية في معالجة مثل هذا التأثير واختباره، وتوعية الباحثين والقائمين بتطبيق الاختبارات بأهميته ومدى تأثيره على دقة نتائجهم من عملية القياس وكذلك على الخصائص السيكومترية للأدوات أو الاختبارات التي يقومون بتطبيقها ومنها معامل الثبات، مما يبرز أهمية الدراسة الحالية في اختبار ذلك والتحقق منه.

تعليق عام على الدراسات السابقة:

باستعراض الباحثة لنتائج البحوث والدراسات ذات الصلة بمتغيرا الدراسة الحالية، وجد بعض التناقض وبعض الاتفاق حول تأثير هذه المتغيرات على قيمة معامل الثبات، فبالنسبة للدراسات المرتبطة بالمحور الأول (أثر عدد بدائل المفردة على قيمة معامل الثبات)، اتفقت نتائج بعض الدراسات على فاعلية استخدام ثلاثة بدائل للمفردة فى تحسين قيمة معامل الثبات مثل دراسة Straton&Catts,80، Trevison,90، Trevison&etal,91، James Bruno,95، نعمان الموسوى ١٩٩٧، Rogers&Harley,99.

كذلك وجدت بعض الدراسات الأخرى التى أوصت بضرورة استخدام أربعة للمفردة لما لها من أثر إيجابى على زيادة قيمة معامل الثبات ومعاملات تمييز المفردات مثل دراسات كل من Green&etal,82، Marshal,90، Budescu&Nevo,85. وفيما يتعلق بالمحور الثانى (أثر موقع أو ترتيب مفردات الاختبار على قيمة معامل الثبات) لاحظت الباحثة اختلاف وجهات نظر الباحثين حول الأسلوب الأمثل فى عرض وترتيب مفردات الاختبار والذى يساعد فى تحسين الأداء الاختبارى للطلاب وبالتالى يزيد من ثبات درجات الاختبار، فقد اتفقت نتائج دراسات كل من Cromack&Sax,66، ودراسة Flaugher&etal.,68، ودراسة Hambleton&Ross,74، ودراسة Newman etal.,88، على فاعلية أسلوب الترتيب المعيارى للمفردات (من السهل إلى الصعب) فى تحسين الخصائص الإحصائية للمفردة وثبات درجات الاختبار، وذلك بالمقارنة بأساليب الترتيب الأخرى ومنها الأسلوب العشوائى، وعلى العكس من ذلك توصلت نتائج دراسات كل من Banner,65، ودراسة Somuse&Munz,68، ودراسة Marso,70، William&Monk,70، ودراسة

Norman&Huck,72، Huck,78، Plack,82 إلى عدم وجود تأثير دال إحصائياً لاختلاف أساليب ترتيب المفردات على قيمة معامل ثبات الاختبار.

وفيما يتعلق بنتائج دراسات المحور الثالث والرابع (أثر طول الاختبار وحجم عينة المختبرين على قيمة معامل الثبات) أكدت نتائج جميع الدراسات على التأثير الإيجابى لزيادة عدد المفردات وحجم عينة المختبرين على دقة واستقرار نتائج القياس.

وبمراجعة الباحثة وتحليلها للدراسات والبحوث السابقة والتي ارتبطت بمتغيرات الدراسة الحالية، وكذلك ملاحظة مدى الاتفاق والتناقض في آراء الباحثين حول مدى تأثير هذه المتغيرات على قيمة معامل الثبات، فقد استفادت من ذلك في توجيه تصنيف متغيرات الدراسة الحالية، لاختبار مدى تأثيرها على خصائص الاختبار السيكومترية ومنها معامل الثبات، وذلك على النحو التالي:

١- فيما يتعلق بعدد بدائل الفقرة، سوف تقوم الباحثة بتصنيف هذه البدائل في مستويين (٣/ ٤) بدائل للفقرة لاختبار مدى تأثيرها على قيمة معامل الثبات.

٢- فيما يتعلق بدراسة أثر طول الاختبار على قيمة معامل ثبات القياس سوف تقوم الباحثة بتطبيق الاختبار في صورتين (اختبار قصير أو عدد مفرداته أقل/ اختبار طويل أو يتميز بزيادة عدد مفرداته).

٣- لاختبار أثر موقع المفردة من الاختبار على قيمة معامل الثبات، سوف تقوم الباحثة بإعداد صورتين لنفس الاختبار ولكن الاختلاف فيما بينهما سوف ينحصر في أسلوب عرض وترتيب المفردات داخل الاختبار (الترتيب من السهل إلى الصعب ــ الترتيب من الصعب إلى السهل).

٤- لاختبار أثر حجم العينة على قيمة معامل ثبات الاختبار، سوف تعتمد الباحثة في تطبيق الاختبار على عينات ذات أحجام مختلفة من الطلاب، للتحقق من مدى هذا التأثير.

فروض البحث:

في ضوء مشكلة البحث والتساؤلات التي يطرحها، والدراسات والبحوث التي تم الإطلاع عليها صيغت فروض الدراسة الحالية على النحو التالي:

الفرض الأول: يختلف تقدير قيمة معامل ثبات الاختبار باختلاف طول الاختبار (عدد مفرداته).

الفرض الثاني: يختلف تقدير قيمة معامل الثبات باختلاف موقع المفردات من الاختبار (ترتيب المفردات داخل الاختبار).

الفرض الثالث: يختلف تقدير قيمة معامل الثبات باختلاف عدد بدائل فقرات الاختيار من متعدد.

الفرض الرابع: يختلف تقدير قيمة معامل الثبات نتيجة اختلاف حجم عينة المختبرين.