البحث السادس :

" **أثر إعادة ترتيب بدائل الاستجابة في صعوبة فقرة الاختيار من متعدد** ".

إعـــداد :

د / باسل خميس سالم أبو فودة
استاذ مساعد القياس والتقويم التربوي بقسم علم النفس
كلية التربية جامعة الأميرة نورة بنت عبد الرحمن

" أثر إعادة ترتيب بدائل الاستجابة في صعوبة فقرة الاختيار من متعدد "

د / باسل خميس سالم أبو فودة

• **مستخلص الدراسة** :

هدفت الدراسة إلى بيان أثر إعادة ترتيب بدائل الاستجابة في صعوبة فقرة الاختيار من متعدد ، ولتحقيق ذلك تمّ بناء اختبار في مبحث الرياضيات لطلبة الصف التاسع الأساسي ، يتكون من متون فقط عددها (٢٠) فقرة ، ثم جرى إعداد كراستين للاستجابة تتكون كل منها من (٣٣) بديلاً ، وتختلف الكراستان في ترتيب بدائل الاستجابة الصحيحة لفقرات نموذج الاختبار ، ثمّ جرى تطبيق نموذج الاختبار على عينة مكونة من (٦٠٠) طالبٍ وطالبة في تربية محافظة جرش ، بواقع (٣٠٠) طالبٍ وطالبة لكل كراسة استجابة ؛ وذلك خلال عام ٢٠١٢ / ٢٠١٣م.

وبعد إجراءات التطبيق حسبت قيم معاملات صعوبة الفقرات في نموذج الاختبار وفق كراستي الاستجابة ، وجرى إيجاد الفرق بين قيم معاملي الصعوبة في الفقرات المتناظرة لنموذج الاختبار حسب كراستي الاستجابة ، وكذلك جرى إيجاد الفرق المطلق بين مواقع الاستجابة الصحيحة في كراستي الاستجابة.

أظهرت النتائج أن الفرق لم يكن دالاً إحصائيًا بين معاملات صعوبة الفقرات ، تعزى إلى إعادة ترتيب بدائل الاستجابة في كراستي الاستجابة ، على الرغم من أنّ توزيع قيم معاملات صعوبة فقرات نموذج الاختبار جرّاء الاستجابة في كراستي الاستجابة مختلفة ، وأشارت النتائج إلى أنه لم يكن هناك نمط ثابت فيما يتعلق بإعادة ترتيب بدائل الاستجابة وأثر ذلك في صعوبة فقرات الاختبار. وأظهرت هذه الدراسة أنّ إعادة ترتيب بدائل الاستجابة قد تُعدّ ممارسة خطيرة خصوصًا في الاختبارات المتكافئة ، وتوصي الدراسة بإجراء المزيد من الأبحاث المستقبلية في هذا المجال.

الكلمات المفتاحية: إعادة ترتيب بدائل الاستجابة ، صعوبة فقرة الاختيار من متعدد.

the effect of reordering response options on the difficulty level of multiple choice test items

Abstract

This study aimed at investigating the effect of reordering response options on the difficulty level of multiple choice test items. To achieve the aim of the study, an achievement test (20 items) was designed for the ninth grade students during the academic year 2012/ 2013. Two response booklet forms, differing only in the order of the options of the test items, were prepared and used. The test was administered to a sample consisting of (600) students: (300) examinees responding for each booklet. Coeffieicents of the difficulty level for each item was calculated for each set of questions. The difference between the coefficients of difficulty for the matching items were calculated. The results of the study showed that reordering response options often had no significant effects on item difficulty although the coefficients of the difficulty levels for the response options were different in the two booklets. The results a here was no stable pattern

concerning reordering test options which affected the difficulty of the level of the test response options. In addition, the study revealed that reordering response options may be considered a dangerous practice especially in the equivellent pictures. The study recommended that test designers should be cautioned about such practice and that this area should be addressed further in future research papers.

Keywords: Reordering Response Options, Item difficulty.

- **المقدمة** :

أن التطور الحاصل في جميع مناحي الحياة ، وتسارع التطور في العقود الأخيرة ، يجعلنا ندرك أن هناك حاجة لتطوير المناهج التعليمية وتطوير أدوات التقويم من مقاييس واختبارات عقلية وأدائية ؛ لتستطيع قياس قدرات المفحوصين بدقة وموضوعية ؛ لذا فقد حظيت الاختبارات بالنصيب الأكبر من هذا التطوير لما لها من استخدامات واسعة في جميع المجالات التعليمية والمهنية.

وتلعب الاختبارات دورًا أساسيًا في حياة الطالب المدرسية بجميع مراحل التعليم ، فبناءً على نتائجها يتحدد مستقبل الطالب ، ويتخذ بحقه الكثير من القرارات ، وما زالت الاختبارات الوسيلة الأساسية في تقويم العملية التربوية ؛ وذلك لتعرف ما حققه الطالب من أهداف تعليمية. والاختبارات كأداة من أدوات التقويم التربوي ، تعدّ من أقدم أدوات التقويم المستخدمة وأكثرها شيوعًا وقبولاً عند أولياء الأمور ، ومع أنّ الاتجاهات في التقويم التربوي تشجع على استخدام أساليب أخرى غير اختبارات التحصيل مثل: ملفات الإنجاز ، فإنه لا يزال للاختبارات التحصيلية دور مهم في عملية التقويم التربوي.

ويفترض في الاختبارات المدرسية أن تكون أداة صادقة تنصف الطالب وتعطيه حقه ، وتقيس ما عنده من قدرات ومعارف ومهارات في المواد الدراسية ، ويفترض المدرس أنّ الفروق بين الطلبة في نتائجهم في الاختبارات تعزى إلى قدراتهم في مجال المادة الدراسية ، إلا أنه من السهل أن يكتشف المدرس بالتغذية الراجعة في الاختبارات ، أنّ حكمة الطالب في الإجابة أو حكمة الاختبار (Test wiseness) يمكن أن تكون عاملاً مساهمًا في الحيلولة دون تحقيق هذا الغرض.

وتتعدد أشكال فقرات اختبارات التحصيل من حيث: صياغتها ؛ فمنها ما يتطلب من المفحوص أن يعطي الإجابة بلغته الخاصة مثل: أسئلة المقال ، والأسئلة التي تتطلب الإجابة القصيرة ، وأسئلة التكميل ، ومنها ما يتطلب من المفحوص أن يختار الإجابة الصحيحة من بين مجموعة من البدائل كفقرات الاختيار من متعدد ، وفقرات الصواب والخطأ ، وفقرات المزاوجة.

ويتحكم في اختيار الشكل المناسب للفقرات مجموعة من العوامل منها: نوع المعلومات المطلوب تقييمها ، وطبيعة المحتوى العلمي للمقرر أو الموضوع نفسه ، والأهداف التعليمية المراد تحقيقها من محتوى الموضوع ، والوقت المحدد لأداء الاختبار(زمن تطبيق الاختبار) ، والمستوى العلمي للطلبة الذين سيوضع لهم الاختبار ، ومستوى الهدف المر[illegible]فحوص ، وغرض الاختبار.

واختبارات التحصيل الكثير من الإيجابيات التي ساعدت في انتشارها نذكر منها ما أورده دودين (٢٠٠٥ ، ص ١٠٢ – ١١٧) على النحو التالي:

- الاختبارات قادرة على توفير معلومات دقيقة ومفيدة ؛ إذ تزود الطالب وولي الأمر وأصحاب القرار بالتغذية الراجعة عن مستوى تحصيل الطلبة ؛ فيتم تعرف مواطن القوة والضعف لدى الطلبة.
- لا تتطلب الاختبارات الكثير من التجهيزات مقارنةً بأدوات التقويم الأخرى.
- إضافة إلى فائدة الاختبارات في التقويم ، فإنّ الاختبارات تعدّ وسيلة تدريس فعالة يمكن استخدامها لتعليم مواد تعليمية متنوعة بكفاءة عالية.
- يمكن استغلال الاختبارات لتوفير خبرات تعليم فريدة ، وبالتالي توظف هذه الخبرات لتعلم محتوى معين أو لتعلم مهارات واتجاهات مفيدة.
- تساعد الطالب على تنظيم وقته أثناء المذاكرة ، وملاحظة الترابط بين أجزائها والعلاقات بين مكوناتها ؛ أي أنها تساعد الطالب في فهم المادة الدراسية فهمًا شموليًا.
- إثارة دافعية الطلبة للتعلم.
- تقييم البرنامج التعليمي.
- تقييم طرائق التدريس.
- تقييم المناهج التعليمية ومدى ملاءمتها لحاجات الطلبة.

وبالرغم من هذه الإيجابيات ، فإنّ معظم الطلبة يكرهون الاختبارات ، ويحاولون تجنب تجربة الاختبار بأي شكل ، وقد يكون السبب من ذلك أنّ الاختبارات . مثل أي وسيلة تقويم . قد تستخدم لتحقيق أهداف لا علاقة لها بالتقويم ؛ كالعقاب مثلا ، أما السبب الآخر والمباشر وراء كره الطلبة للاختبارات فهو ارتباطها الدائم بالعلامات ، وما يترتب على ذلك من قراراتٍ ووقائع سلبية ؛ كالرسوب ؛ مما ينقل هذا الشعور إلى الاختبارات ذاتها ، إضافة إلى ذلك فإنّ ثقافة المجتمع لا تنظر للاختبار باعتباره أداة تقويم عادية ، وإنما يضخم الاختبار إلى حدث مصيري ، حتى يشعر الطالب بالقلق والتوتر ؛ أي أنّ الأمر تعدى النتيجة التربوية المتوقعة من تقديم الاختبار إلى شيء أكبر بكثير ؛ وبالتالي يصاب الطالب بالفزع عند التفكير بذلك ، وعندما يربط أولياء أمور الطلبة مدحهم لأبنائهم بنتائج الاختبارات ، فإنهم في الواقع يزيدون الأمر سوءًا ؛ إذ تتضاعف أهمية نتيجة الاختبار بالنسبة للطالب.

وفيما يتعلق بفقرات الاختيار من متعدد على وجه التحديد ، فهناك عدة صيغ لفقرات الاختبار من متعدد ، وأبسط هذه الصيغ عندما يكون للفقرة بديلين ؛ بحيث تكون إجابات الطلبة على الفقرة (بنعم أو لا) أو(صح أو خطأ) ، ويمكن استخدام هذا النوع من صيغ الفقرات لقياس قدرة الطلبة على تمييز التعريفات الصحيحة والحقائق البسيطة ، بدلا من أسئلة الاختيار من متعدد عندما تكون اختبارات الاختيار من متعدد لا تحتمل سوى بديلين ، واحد يمثل الإجابة الصحيحة ، والآخر يصرف الانتباه عن الإجابة الصحيحة ، وتفضل هذه الحالة ، لأنها تشغل حيزًا أقل من الورقة ، ولا تحتاج قراءتها إلى وقت طويل.

ومن الصيغ الأخرى لفقرات الاختيار من متعدد ، عندما يكون للفقرة ثلاثة بدائل أو أربعة بدائل أو خمسة [illegible] اج هذا النوع من الاختبارات إلى

وقت ومهارة أكثر من صيغ الفقرات التي يكون لها بديلان ، وكذلك يحتاج الاختبار الذي لفقراته خمسة بدائل إلى وقت ومهارة أكثر من الاختبار الذي لفقراته أربعة أو ثلاثة بدائل ؛ حيث يحتاج باني الاختبار إلى التفكير في عدة احتمالات ممكنة للإجابة ، كما يحتاج الطلبة إلى وقت أطول في الإجابة بما يتناسب مع عدد البدائل.

وبشكل عام تُعدّ فقرات الاختيار من متعدد من أفضل أنواع الفقرات الموضوعية ، وأكثرها شيوعًا وانتشارًا في اختبارات التحصيل (Aiken, 1987, pp. 44 – 57) ، ويتكون اختبار الاختيار من متعدد من مجموعة من الفقرات ، التي تشكل عينة عشوائية من مجموع الفقرات التي يمكن صياغتها لقياس البناء النفسي أو العقلي المراد قياسه ، ويفترض أن تقيس الفقرات بمجموعها سمة واحدة ، وهذا ما يسمى بأحادية البعد (Unidimensionality) (Hambleton & Swaminathan, 1985, p. 16 - 22) ، ولكن من الصعب الوصول إلى هذه الفقرات من الناحية العملية ، فبعض الاختبارات تقيس أكثر من سمة ، فامتلاك بعض الأفراد لسمة غير السمة المراد قياسها بواسطة الاختبار قد تساعدهم في الإجابة عن بعض الفقرات ولا تساعد غيرهم ، والفقرة التي تتصف بهذه الخاصية هي فقرة متحيزة.

وتتكون جميع فقرات الاختيار من متعدد من العناصر نفسها ، وأول هذه العناصر هو المتن أو الأرومة (Stem) ، الذي يمكن أن يكون على شكل سؤال ، أو جملة إخبارية تتضمن قضية معينة تحتاج إلى إجابة ، أما العناصر الأخرى في فقرة الاختيار من متعدد فهي الخيارات أو البدائل (Options or alternatives) ، وهي إجابات محتملة للسؤال الوارد في المتن ، وأحد هذه البدائل هو الإجابة الصحيحة ، أما بقية البدائل فتعرف بالموهات أو المشتتات (Distracters or Foils) (Roid & Haladyna, 1982, PP. 49 - 55). وتعرف الاختبارات الموضوعية بأنها تلك الاختبارات التي تخلو أسئلتها من الذاتية ؛ مما يعني أنها تتحرر من ذاتية المصحح عند تقدير درجات الطلاب (Gronlund & Linn, 1990,PP.166 - 189).

وعند تطوير اختبار من نوع الاختيار من متعدد يجب أخذ عدة أمور تؤثر في الخصائص السيكومترية للاختبار بعين الاعتبار ؛ مثل (الثبات والصدق) ، وفي الخصائص السيكومترية للفقرة (الصعوبة والتمييز) ، ومن هذه الأمور عدد البدائل ؛ حيث أوضحت بعض الدراسات ؛ مثل دراسة ستارتون وكاتس Straton) & Catts,1980, PP. 357 - 365) أنّ اختلاف عدد البدائل يؤثر في الخصائص السيكومترية للاختبار وفقراته ، وفي هذا الصدد يرى كل من أبيل وفريسبي (Eble & Frisbie, 1986) أن اختلاف عدد البدائل في اختبارات الاختيار من متعدد يؤدي إلى اختلاف عدد المقارنات للإجابة المعطاة ؛ حيث يؤثر ذلك على صعوبة الفقرة ومعامل تمييزها ، وباختلاف الدالة التمييزية لها يختلف ثبات أداء المفحوصين على الفقرة وعلى الاختبار ككل ؛ مما يؤثر على معامل ثبات الاختبار ، فكلما زادت الدالة التمييزية للفقرات زاد ثبات الاختبار ؛ لأنّ انخفاض الدالة التمييزية للفقرة يجعل النجاح على الفقرة مرتبطا بالصدفة ، وبذلك فهي لا [illegible] ة الحقيقية ، كما يؤثر مدى

صعوبة الفقرة على ثبات الاختبار ؛ وذلك لأنه عندما تكون الفقرات متوسطة الصعوبة يكون الثبات أعلى منه في حالة الفقرات السهلة جداً أو الصعبة جداً.

ويشير الصراف (٢٠٠٢) إلى أن ثبات الاختبار يتأثر بعدد فقراتهِ وعدد البدائل لهذه الفقرات ، فالاختبار القصير ؛ أي الذي يحوي عدداً قليلاً من الفقرات ، يميز بشكل بسيط بين المستويات المختلفة لمعرفة الطلبة ، كما أن زيادة عدد بدائل فقرات الاختبار قد تعمل على تشويش الطلبة وتزيد من أخطائهم ؛ مما يستدعي تفضيل الاختبار الذي يحوي على عدد أكبر من الفقرات ، وبعدد بدائل أقل لتلك الفقرات. ولتجنب تأثير عدد البدائل على معامل ثبات الاختبار فلابدّ أن تكون البدائل متجانسة ومبنية وفق استراتيجية جيدة لنحصل على اختبار قادر على قياس قدرات المفحوصين بدقة وموضوعية ، ويُزود واضعيه بمعلومات موثقة لعملية اتخاذ القرارات ، ويكون له دور إيجابي في عملية التعليم والتعلم.

والأمر الثاني في هذا المجال هو شكل بدائل فقرات الاختيار من متعدد ، التي تتلخص في وجود إجابة واحدة صحيحة وعدد من المموهات ، سواء كانت بدائل تقليدية أو بدائل مركبة ؛ كأن تحتوي على بدائل في شكل "لا شي مما ذكر" ، أو "جميع ما ذكر صحيح" ، أو الجمع بين البدائل مثل "أ + ب" ، وهذا ما نلاحظه في كثير من اختبارات التحصيل التي يعدها المعلمون ، أو قد يتم استخدام البدائل المركبة لفقرات الاختبار ، إما لتوفير الوقت والجهد في بناء مموهات جيدة ، أو عدم توفر الخبرة لدى مطوري الاختبارات بالمحتوى الدراسي أو بالمفحوصين ، أو عدم وجود استراتيجية معينة يتم من خلالها بناء بدائل جذابة ومناسبة لفقرات اختبارات الاختيار من متعدد.

لقد أشارت كريهان وهالادينا (Crehan & Haladyna,1991, PP.183 – 192) إلى أنّ الفقرات التي تحتوي على بدائل مركبة يؤدي إلى زيادة الصعوبة والتمييز لتلك الفقرات ، ومن ناحية أخرى يرى جرونلاند (Gronlund & Linn, 1990, PP. 166 - 189) أنّ استخدام البديل "جميع ما ذكر صحيح" يزيد من احتمالية تخمين الإجابة من بين البدائل ، بحيث إذا عرف المفحوص أن بديلين من بدائل الفقرة صحيحان ، فإنه سيختار البديل "جميع ما ذكر صحيح" بغض النظر عن عدد البدائل المتبقية للفقرة ، وإذا عرف المفحوص أنّ أحد البدائل . على الأقل . خاطئ ، فإنه سيستبعد البديل "جميع ما ذكر صحيح".

ويضيف عودة (٢٠٠٥) أنّ بعض المعلمين قد يلجأون إلى وضع البديل "لا شي مما ذكر" وخاصة في الأسئلة الحسابية ؛ وذلك تحسباً من عدم وجود إجابة صحيحة بين البدائل دون أن يقصد المعلم ذلك ، وفي هذه الحالة يبقى المعلم في حل من أي اعتراض حول عدم وجود إجابة صحيحة ، وبذلك يصبح البديل "لا شي مما ذكر" بمثابة البديل الهروبي.

أما الأمر الثالث فهي الطريقة التي يتم من خلالها اختيار البدائل لفقرات اختبار الاختيار من متعدد ، فمن الإرشادات الواردة في أدبيات القياس أن تكون البدائل متجانسة مع بعضها [illegible] بد من قدرة المموهات على جذب

انتباه المفحوصين ؛ حيث يتوقع أن تزداد صعوبة فقرة اختبار الاختيار من متعدد بزيادة تجانس البدائل ، وإذا كانت مموهات الفقرة ضعيفة وغير جذابة ؛ فيتوقع أن يؤدي ذلك إلى تخفيض صعوبة الفقرة ؛ حيث سينعكس ذلك على قدرتها التمييزية ، فالفقرات التي معامل صعوبتها قريب من الصفر أو الواحد ، تكون قدرتها التمييزية ضعيفة. ولابدّ من أن يكون متن الفقرة مُصاغا بشكل جيد ، فعندما تكون الفقرة غامضة وضعيفة لغويًا ، فإنّ ذلك يؤدي إلى إرباك الطلبة ، ويزيد من احتمال استجابتهم للفقرة بشكل خاطئ، حتى وإن كانوا يمتلكون المعرفة اللازمة لإجابة الفقرة بشكل صحيح، وفي هذا الصدد يرى جرونلاند ولين (Gronland & Linn, 1990, PP. 166 - 189)، أن مثل هذا الغموض يؤدي إلى فهم مختلف من مفحوص لآخر لمضمون فقرة الاختبار.

وتحتل أسئلة فقرات الاختيار من متعدد مكاناً بارزاً في منظومة التقويم في الكثير من دول العالم، وذلك لتعدد إيجابياتها، ولعل من أهمها ما يأتي:

- سهولة وسرعة تصحيح الإجابات، مما يمكن المعلمين من تقديم التغذية الراجعة لتلاميذهم بصورة فورية.
- سهولة تغطية محتوى المقرر الدراسي، وبالتالي فهي تغطي معظم الأهداف التعليمية للمحتوى الدراسي ، وبذلك تعدّ أكثر شمولية ؛ مما يجعل الهدف منها واضحًا والإجابة عليها محددة ؛ فيقل غموض الأسئلة.
- لا تتأثر بقدرة الطالب اللفظية أو التعبيرية ، أو سوء خطه ، أو تنظيمه لورقة الإجابة ، أو أخطائه الإملائية والنحوية ، ونحو ذلك ؛ فالاستجابات تعتمد على تعرف الإجابة الصحيحة من بين الإجابات المعطاة.
- الموضوعية في تصحيح الإجابات وتقدير الدرجات وكذلك موضوعية النتيجة النهائية.
- عدم تأثير عامل الخداع في النتيجة النهائية للطالب ؛ فالإجابة عن كل سؤال محددة جيداً ومعروفه ، ولا تحتاج سوى اختيار حرف معين أو وضع علامة أو كتابة كلمة ونحو ذلك.

وكما أنّ لهذا النوع من الاختبارات مميزات كثيرة ، فله - أيضًا - نصيب من العيوب ينبغي التنويه عنها ؛ كي نضعها موضع الاعتبار عند تصميم تلك الاختبارات وإعداد أسئلتها، ومن بين تلك العيوب ما يأتي:

- إلغاء عملية التنظيم والقدرة التعبيرية وكذلك القدرة على الابتكار والإبداع والتفسير لدى الطلاب ؛ فالإجابات محددة ومعطاة عقب كل سؤال بورقة الإجابة.
- يحتاج معدوا ومصمموا هذه الأنواع من الاختبارات إلى التدريب الجيد على مهارات إعداد أسئلة تلك الأنواع بطريقة جيدة.
- يحتاج إعداد وبناء بدائل الإجابة الجيدة عن الأسئلة إلى وقت وجهد كبيرين وبخاصة إذا كانت هذه الأسئلة تقيس مستويات عقلية عليا، وعدم توافر فرصة تعبير الطالب عن نفسه أو عن رأيه عند الإجابة الصحيحة.
- ارتفاع أثر التخمين؛ فعند عجز الطالب عن معرفة الإجابة الصحيحة فإن أول ما يفكر فيه هو اختيار الإجابة عشوائيًا من بين بدائل الإجابات المعطاة عقب كل سؤل ، وبخاصة في أسئلة الصواب والخطأ. ففي أسئلة الاختيار من

متعدد ذات أربعة البدائل وجد أنّ نسبة التخمين تكون (٢٥٪) ؛ وبالتالي فإن الطالب الذي لم يقرأ المادة التعليمية يمكنه الحصول على (٢٥٪) من الدرجات بالتخمين ، وقد ترتفع تلك النسبة إلى أكثر من ذلك ، فقد أوضح جونستون وآخرون (Johnstone et al., 1983, pp. 626 – 634) أنّ نسبة التخمين قد ترتفع إلى (٥٠٪) في أسئلة فقرات الاختيار من متعدد ذات أربعة البدائل ، نتيجة قيام الطلبة بحذف بديلين في البداية قبل القيام بعملية التخمين لأسباب منها: عدم تجانس البديلين مع باقي البدائل ، أو وجود كلمة مفتاحية تعمل على استبعاد هذين البديلين من قبل الطلبة.

وهناك – أيضًا – سيل من الانتقادات وجهت لهذا النوع من الاختبارات من أهمها في هذا البحث ما يعرف بالتحيز لموقع الاستجابة Bias Response Position، ويقصد بذلك اختيار المفحوص بديل ما اعتمادًا على موقع ذلك البديل في ترتيب البدائل الأخرى، بغض النظر عن محتوى السؤال، فقد يختار الطلبة البديل(أ) مثلا أكثر من البديل(ب) في حالة أنّ السؤال يتكون من أربعة بدائل، مع احتمال أن تكون الإجابة الصحيحة هي البديل(ب)، وهذا يرجع إلى أنّ بعض الطلبة قد لا يقرءون كل البدائل، وإنما يكتفون بقراءة البديلين الأول والثاني فقط، خاصة عندما يكتنف الغموض جذر السؤال.

لقد أشار بلونش(Blunch, 1984, pp. 216 - 220) إلى أنّ اختيار المفحوص للبديل الصحيح يعتمد على موقع البديل الصحيح بالنسبة للبدائل الأخرى، ومن ضمن بدائل المموهات أو المشتتات، بديل يظهر بصورة كثيرة في استجابات الطلبة، حتى أنّه أحيانا يختاره الطلبة أكثر من اختيارهم لبديل الإجابة الصحيحة، ويطلق عليه باسم المشتت القوي Srong Distractor، ويعدّ موقع هذا البديل بالنسبة للبديل الصحيح من ضمن العوامل التي تسبب مشكلات في هذا النوع من الأسئلة، فوجود هذا البديل بجانب البديل الصحيح، أو بعيدًا عنه، يؤثر في استجابات الطلبة، الأمر الذي ينعكس على احصائيات الفقرة والاختبار.

وهناك مجموعة من القواعد والإرشادات في كتابة فقرة الاختيار من متعدد، أوردها (عدس، 2002 ;Haladyna, Downing & Rodriguez, 2002)، منها: التأكد من أنّ الأرومة تطرح مشكلة محددة وواضحة، وجعل البدائل قصيرة ما أمكن؛ وذلك لتحقيق الاقتصاد في الوقت، ولضمان الوضوح في طرح المشكلة، وتَجنّب صيغ النفي في الأرومة، فعبارات النفي إذا ما اقترنت ببدائل الإجابة، تطرح مشكلات قرائية معقدة أمام الطلبة، والتأكد من أنّ بدائل الإجابة الخطأ تؤلف إجابات معقولة ظاهريا، وأن تكون جذابة للمفحوصين الذين تنقصهم المعرفة، وأن لا يتّضمن الاختبار فقرات تعتمد في إجابتها على فقرات أخرى، أي يجب أن تكون الفقرات مستقلة عن بعضها الآخر، وتَجنّب استخدام الخداع والغموض في أرومة الفقرة وبدائلها دون مبرر منطقي، وتجنب وجود المحدداتٍ اللفظية الخاصة، أو ما تسمى بمصطلحات التخصيص أو التعميم مثل: أبدا، دائما، كل، لا أحد نحو ذلك؛ لأنّ وجودها غالبا ما يدل على أنّ البديل خاطىءٍ، وجعل بدائل الفقرة متساوية في طولها؛ لكي لا يكون طول البديل المميز مؤشرا على صحة الإجابة، وعدم الإتساق القاعدي أو اللغوي بين أرومة الفقرة والبديل الصحيح؛ لأنه يُعدّ تلميحا .

ولا تقتصر قواعد صياغة فقرات الاختيار من متعدد على هذه الإرشادات، بل هناك العديد من الإرشادات والقواعد التي قدمها الخبراء لكاتبي الأسئلة، وهي في مجملها مبنية على الخبرة، ولها ما يبررها. ويرى براون (Brown, 1983) أنّ اختيار شكل الفقرات المناسب يعتمد على خصائص الفقرة، ومهارة المعلم في كتابة الفقرات، وفلسفة التربية، فمثلًا إذا رأى المعلم أن طلابه يجب أن يكونوا قادرين على تذكر المادة التعليمية بدلًا من تعرف الاستجابات الصحيحة، فعليه أن يستخدم اختبارات التكميل والإجابة القصيرة، خاصة أنّ بناء فقرات التكميل أو الإجابة القصيرة أسهل من بناء فقرات الاختيار من متعدد جيدة المستوى، وفي كثير من الأحيان يمكن قياس الهدف بأكثر من نوع من الفقرات؛ إذ يمكن ترجمة الفقرة من نوع إلى آخر، ومع ذلك فغالبا ما يميل المعلمون إلى استخدام نوع من الفقرات دون الآخر.

ويتوقف مستوى أداء المفحوص في الاختبار، على خصائص الاختبار والمفحوص (Hambleton, 1983)، ومن خصائص فقرة الاختيار من متعدد المؤثرة في مستوى أداء المفحوص ما يعرف بصعوبة الفقرة، وقدرتها التمييزية، وطريقة ترتيب الفقرات، وطرق إعطاء التعليمات، إضافة إلى جاذبية الموهات، ومدى تأثرها بعامل التخمين (Plake, Thompson & Lowry, 1981, PP. 214 - 219)، ويمكن التعبير عن صعوبة الفقرة بطريقتين: الأولى إحصائية، والثانية معرفية، ويقصد بمستوى صعوبة الفقرة إحصائيا، نسبة الأفراد الذين أجابوا عن الفقرة إجابة صحيحة، وتتأثر هذه النسبة إضافة إلى القدرة اللازمة لإجابة الفقرة، بعوامل فنية تتعلق بدرجة الوضوح في صياغة المتن، ومدى تجانس بدائل الفقرة، ودرجة علاقتها مع متن السؤال، ودرجة وضوح ذلك المتن، ومدى تحرر الفقرة بشكل عام من خلل في صياغتها، أما صعوبة الفقرة المعرفية فيقصد بها نمط المعالجة العقلية التي تتطلبها المهمة، ويمكن تعريفها إجرائيا بالمستوى المعرفي الذي تقيسه تلك الفقرة (Lane & Bull,1987, PP. 865 - 879).

وفيما يتعلق بالقدرة التمييزية للفقرة، فيقصد بها قدرة الفقرة على التمييز بين الأفراد الذين يمتلكون قدرا مختلفا من السمة المقيسة، ولا تقتصر خصائص الاختبار المؤثرة في الأداء على صعوبة الفقرة وتمييزها ، بل إنّ طريقة ترتيب فقرات الاختبار تؤثر في الأداء، ومن هذه الطرق ترتيب الفقرات وفق مستوى صعوبتها الإحصائية، أو المعرفية، أو وفق المستويين معا؛ إذ يمكن ترتيب الفقرات من السهل إلى الصعب، أو من الصعب إلى السهل، أو ترتيبها حلزونيا، إذ ترتب فقرات كل موضوع من الموضوعات المتضمنة في الاختبار من السهل إلى الصعب (Plake et al., 1981).

أما خصائص المفحوص المؤثرة في مستوى الأداء على الفقرة، فتتمثل بمستوى قدرة المفحوص على السمة المقيسة، وحكمته في إجابة الفقرة، إضافة إلى عوامل أخرى قد يكون لها تأثير في مستوى الأداء في الفقرة، وفيما يتعلق بمستوى قدرة المفحوص يرى هامبلتون وسواميناثن (Hambleton & Swaminathan, 1985) أنّ احتمال إجابة المفحوص عن الفقرة إجابة صحيحة، يزداد بازدياد مستوى القدرة، أي أنّ المفحوص الذي يمتلك مستوى عال من

القدرة على السمة المقيسة، يكون احتمال إجابته لفقرة معطاة إجابة صحيحة، أعلى من مفحوص آخر يمتلك مستوى قدرة أقل من السمة المقيسة.

وفيما يتعلق بالخصائص السيكومترية لفقرات الاختبار فقد أجرى ماركوس (Marcus, 1963, pp. 48 - 51) في جامعة مسيج يوست دراسة بعنوان: " أثر موقع الإجابة الصحيحة في مستوى صعوبة أسئلة فقرات الاختيار من متعدد". إذ هدفت إلى فحص أثر موقع الاستجابة في فقرات الاختيار من متعدد على مستوى صعوبة الفقرات، ولتحقيق ذلك جرى إعداد أربعة نماذج من نوع الاختيار من متعدد، إذ جرى توزيع الإجابات الصحيحة والمموهات عشوائيا في نماذج الاختبار، وذلك وفق مخطط يسمح لكل موقع بعدد متساوي من الإجابات الصحيحة، إذ ظهر الجواب الصحيح لكل فقرة في مواقع مختلفة في كل النماذج. جرى تطبيق نماذج الاختبار على (434) طالبا من طلبة الجامعة. وأشارت النتائج إلى أنّ اختبارات الاختيار من متعدد متحررة نسبيا من أفضلية الموقع، بمعنى أنّ موقع الاستجابة كان مهملا وغير دال إحصائيا على مستوى صعوبة الفقرة، وأوضحت النتائج فعالية موقع المشتت القوي على التأثير في استجابات الطلبة.

وقام كاسلز وجونستون (Cassels & Johnstone, 1984, pp. 613 - 615) بإجراء دراسة بعنوان: " أثر اللغة في أداء المفحوص في اختبارات الاختيار من متعدد في مبحث الكيمياء ". إذ قام الباحثان بالتلاعب بتعقيدات اللغة في الأزواج المتناظرة في فقرات الكيمياء، وقد شملت التغيرات في مصطلحات الكميات، والتغيرات في التعقيد العام لجذور الفقرة باستعمال الفقرات المصاغة إيجابا وسلبيا، وقد أظهرت النتائج آثار صغيرة في صعوبة الفقرات عند استعمال المبني للمعلوم أو المجهول، ولوحظت آثار كبيرة في صعوبة الفقرات عند تغيير الكلمات المفتاحية في جذور الفقرة.

وقام لين وبول(Lane & Bull, 1987, PP. 865 - 879) بإجراء دراسة بعنوان: "أثر معرفة ترتيب فقرة الاختبار والجنس وصعوبة الفقرة الإحصائية والمعرفية في الأداء على الفقرة". وكان هدف الباحث من ذلك الإجابة عن الأسئلة التالية: هل هناك أثر لترتيب الفقرات في الأداء؟ وهل يوجد أثر للجنس في الأداء؟. تكونت عينة الدراسة من(١٣٧) طالبا و(٢٦٥) طالبة من طلاب الجامعة، قاموا بالإجابة عن نموذجين من نوع الاختيار من متعدد مكون كل منهما من(٤٠) فقرة.

أشارت النتائج إلى وجود فروق ذات دلالة إحصائية في الاداء بين طرق ترتيب الفقرات إحصائيا، إذ وجد أنّ مستوى الأداء كان أفضل عندما كان الترتيب من السهل إلى الصعب منه في حالة ترتيب الفقرات من الصعب إلى السهل، وأشارت النتائج إلى أنّ الأداء كان أفضل عندما كانت صعوبة الفقرات المعرفية منسجمة مع صعوبتها الإحصائية منه في حالة كون الصعوبة المعرفية والإحصائية متعاكسة.

وأجرى سيزك(Cizek, 1991, pp. 8 - 20) دراسة بعنوان: " أثر تغير موقع البدائل في اختبارات الاختيار [illegible] هدفت إلى تحديد فيما إذا كانت

إعادة ترتيب بدائل فقرة الاختيار من متعدد تنتج أي أثر دال إحصائياً على صعوبة الفقرة. جمعت البيانات من (٧٥٩) شخصًا (خريجي برامج تدريب الإقامة الطبية الخاصة) لفحص الشهادة في حقل تخصص الطب، إذ أجاب كل شخص عن(٢٠) فقرة من نوع الاختيار من متعدد، وكان لابدّ على المفحوصين أن يختاروا من (٣٠) بديلاً للبديل الصحيح، وقد أظهرت البدائل الـ(٣٠) بصريا كمثيرات لكل فقرة.

تمّ إعداد كراستين للاستجابة، وتختلف الكراستان فقط في موقع البدائل(الخيارات) لـ(٢٠) فقرة، وقد أجاب(٣٨٠) مفحوصًا على الكراسة(A)، و(٣٧٩) مفحوصًا أجابوا على الكراسة(B). وقد أظهرت نتائج الدراسة أنّ إعادة ترتيب بدائل الفقرات غالبا ما يكون دالا إحصائيا على صعوبة الفقرة، لكن آثاره متقلبة على الأداء في الفقرة.

ومن الدراسات العربية في هذا المجال دراسة أجراها أمبوسعيدي والعفيفي(٢٠٠٤، ص ص: ١٦٩ - ١٩١) بعنوان: "أثر تغيير موقع المشتت (Strong Distractor) في بعض الخصائص السيكومترية لأسئلة الاختيار من متعدد في مادة الفيزياء"، إذ هدفت إلى تقصي أثر تغيير موقع المشتت القوي وهو أحد أشكال التحيز في موقع الاستجابة في أسئلة الاختيار من متعدد في تحصيل الطلاب، وفي بعض الخصائص السيكومترية للفقرات كالصعوبة والتمييز. ولتحقيق أهداف الدراسة تم تصميم اختبار في مادة الفيزياء للصف الأول الثانوي مكون من(٣٠) فقرة في صورته النهائية، إذ جرى تصميم نموذجين متشابهين في كل شيء من هذا الاختبار ما عدا موقع المشتت القوي بالنسبة للبديل الصحيح. تكونت عينة الدراسة من(١٩٧) طالبا وطالبة، جرى اختيارهم عشوائيا من ثلاث مدارس في قطاع ولاية السيب بمحافظة مسقط بسلطنة عُمان خلال العام الدراسي ٢٠٠٢/٢٠٠٣م. أظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لدرجات الطلبة في نموذجي الاختبار، ووجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لمعاملات الصعوبة ومعاملات التمييز لنموذجي الاختبار.

وقام أمبوسعيدي (٢٠٠٥، ص ص: ٨٦ - ١٠٥) بإجراء دراسة بعنوان: "أثر تغيير موقع المشتت(Strong Distractor) عند استخدام تقنية إبعاد البدائل الخطأ في صعوبة الفقرات لأسئلة الاختيار من متعدد في الفيزياء وتمييزها "، إذ هدفت إلى تقصي أثر تغيير موقع المشتت القوي وهو أحد أشكال التحيز في موقع الاستجابة، عند استخدام تقنية إبعاد البدائل الخطأ في أسئلة الاختيار من متعدد في التحصيل الدراسي، وفي بعض الخصائص السيكومترية للاختبار كمعاملات الصعوبة ومعاملات التمييز.

ولتحقيق أهداف الدراسة تمّ استخدام اختبار في مادة الفيزياء للصف العاشر من التعليم العام مكون من(٢٤) فقرة من نوع الاختيار من متعدد، إذ جرى تصميم نموذجين متشابهين في كل شيء من هذا الاختبار ما عدا موقع المشتت القوي بالنسبة للبديل الصحيح، كما تمّ استخدام تقنية إبعاد البدائل الخطأ من الورقة الامتحانية في النموذجين. تكونت عينة الدراسة من(١٣٣) طالبة من

طالبات الصف العاشر من التعليم العام، جرى اختيارهن عشوائياً من مدارس التعليم العام في قطاع ولاية السيب بمحافظة مسقط بسلطنة عُمان أواخر العام الدراسي ٢٠٠٣/٢٠٠٤م. أظهرت نتائج الدراسة عدم وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لدرجات الطالبات في نموذجي الاختبار، وأشارت النتائج أيضا عدم وجود فروق ذات دلالة إحصائية بين المتوسطات الحسابية لمعاملات الصعوبة ومعاملات التمييز لنموذجي الاختبار.

- **مشكلة الدراسة**:

اتضح من العرض السابق أن بعض الدراسات تناولت أثر تغيير موقع المشتت القوي في صعوبة الفقرات في أسئلة الاختيار من متعدد، وبعضها تناول أثر ترتيب فقرات الاختبار في الأداء، وبعضها الآخر بحث في أثر موقع الإجابة الصحيحة في صعوبة أسئلة فقرات الاختيار من متعدد، وذلك من خلال التركيز على مجموعة من الاستجابات مكونة من أربعة أو خمسة بدائل لفقرة الاختبار، ولكن البحث الحالي يركز على فحص التأثيرات المحتملة من تغير موقع الاستجابة الصحيحة في صعوبة فقرة الاختيار من متعدد، وذلك بعرض المثيرات(بدائل الاستجابة) في كراسة منفصلة أكثر من المتوقع نظريا، ثمّ إعادة ترتيبها في كراسة أخرى، إذ لم تتطرق الدراسات بصورة مباشرة إلى البحث في أثر إعادة ترتيب بدائل الاستجابة في صعوبة فقرة الاختيار من متعدد، إذ أنّ التلاعب في ترتيب بدائل الاستجابة قد يؤثر في مستوى صعوبة الفقرة، وتُعدّ هذه الدراسة بإجراءاتها الحديثة أول دراسة عربية - حسب علم الباحث - في هذا المجال.

- **أهداف الدراسة**:

سعت الدراسة إلى تحقيق الهدف الرئيس وهو: الكشف عن أثر إعادة ترتيب بدائل الاستجابة في معاملات صعوبة فقرات الاختيار من متعدد.

- **أسئلة الدراسة**:

حاولت الدراسة الإجابة عن السؤال الرئيس الآتي:

- هل هناك أثر في إعادة ترتيب بدائل الاستجابة في صعوبة فقرة الاختيار من متعدد؟

- **أهمية الدراسة**:

يتضح في ضوء ما سبق أنّ موقع الاستجابة الصحيحة وأثر ذلك في صعوبة الفقرة ما يزال موضع تساؤل، إذ أنّ المعرفة بمدى التأثيرات التي يمكن أن تنشأ من التلاعب التجريبي في ترتيب بدائل الاستجابة في صعوبة فقرة الاختبار ما تزال مدار البحث، كما أنّ الدراسات التي أجريت في هذا المجال تناولت الموضوع من ناحية موقع الاستجابة الصحيحة في فقرات الاختيار من متعدد وأثر ذلك في صعوبة الفقرة معتمدة في ذلك على الفقرات ذات الأربعة أو الخمسة بدائل، كما تناولت بعض الدراسات أثر موقع المشتت القوي في صعوبة فقرات الاختيار من متعدد، ولم تتطرق الدراسات إلى البحث في أثر إعادة ترتيب بدائل الاستجابة في صعوبة فقرة الاختيار من متعدد، فما يبرر أهمية هذه الدراسة هو أهميتها النظرية والعملية على حد سواء، فمن الناحية النظرية تبين فيما إذا كان هناك أثر في إعادة ترتيب بدائل الاستجابة في صعوبة فقرة الاختيار من متعدد، أما أهميتها العملية فتكمن في [illegible] لأوضاع قد تمارس في الميدان

التربوي، ومحاولتها التوصل إلى تبريرات عملية تقدم إلى الباحثين لتثري المجال التربوي.

إنّ الخاصية المثيرة جداً والمهمة في هذه الدراسة تتمثل في كراستي الاستجابة، اللتين تحتويان على(٣٣) مثيرا موزعة بصورة مختلفة في الكراستين، اللتين تمثلان الاستجابات لفقرات الاختبار، إذ جرى إعدادهما بحيث تكون المسافة بين البدائل أكثر من المتوقع نظريا، أي أكثر من أربعة أو خمسة بدائل في فقرة الاختيار من متعدد، بمعنى أنّ هناك فصلا بين متون الفقرات التي وضعت بصورة منفصلة في نموذج الاختبار، وبدائل الاستجابة التي وضعت في كراسة الاستجابة، ذلك أنّ البدائل للفقرة ذات خمسة البدائل مثلا، يكون الاختلاف فقط في إعادة ترتيب بدائل الاستجابة ضمن أربعة مواقع فقط، أما في هذه الدراسة فيوجد مدى واسع من الاحتمالات لموقع البديل الصحيح من (١ ؛ ٣٣) موقعا، إذ تسمح هذه الخاصية للبحث أن يكون أكثر حساسية، خصوصا للفروق بين أماكن الاستجابة الصحيحة التي قد تؤثر في صعوبة فقرة الاختيار من متعدد .ومما يزيد في أهمية الدراسة أنها اعتمدت بيانات حقيقية، تم الحصول عليها من مواقف تعليمية واقعية، وتبرز أهمية الدراسة من الناحية العملية أيضا في أن إعادة ترتيب بدائل الاستجابة له أثر دال في صعوبة الفقرة، فليس من الضروري تجنب ذلك، إذ يجب على معدي الاختبار أن يكونوا على حذر في مثل هذه الممارسة خصوصا ضمن سياق الاختبارات المتكافئة. وتبرز أهمية البحث في طبيعة عدد الفقرات المشتركة وطريقة تطبيقها، إذ كان عددٍ الفقرات(٢٠) فقرة، وكان سبب اختيار الباحث لهذا العدد من الفقرات اعتمادا على القاعدة الاجتهادية المقدمة من قبل أنجوف(Angoff, 1971, p.578) التي مفادها: " إنّ مجموعة الفقرات المشتركة يجب أن لا تقل عن(٢٠) فقرة، أو(٢٠٪) من فقرات الاختبار الكلي، أيهما أكبر" . إذ طبقت هذه الفقرات في موقف صفي عادي، دون أن يظهر للتجريب أي أثر في أثناء التطبيق، الذي قد يؤثر في أداء المفحوصين في حالة ظهوره، فموقف الاختبار في الظروف الصفية العادية، يثير دافعية جميع المفحوصين للإجابة عن فقرات الاختبار بدرجة عالية ومتقاربة إلى حد ما، وهذا يزيل أثر اختلاف مستوى دافعية المفحوصين في أدائهم أثناء الاختبار، فإجراءات هذه الدراسة في مواقف تعليمية واقعية يسهل إعادتها في تخصصات ومستويات دراسية مختلفة، مما يمكن من الوقوف على صحة ما يتوصل إليه من نتائج وإمكانية تعميمها.

- **محددات الدراسة**:

اقتصرت الدراسة على بيان أثر إعادة ترتيب بدائل الاستجابة في صعوبة فقرة الاختيار من متعدد، وهذا يحدّ من إمكانية تعميمها على باقي الخصائص السيكومترية كمعامل تمييز الفقرة ومعامل تخمينها من منظور النظرية الكلاسيكية في القياس.

- **مصطلحات الدراسة**:

- معامل صعوبة فقرة الاختبار(حسب نظرية القياس الكلاسيكية): ويشير إلى نسبة المفحوصين الذين أجابوا عن الفقرة إجابة صحيحة ممن حاولوا الإجابة عنها.(معامل السه [illegible] لصعوبة)

- ترتيب بدائل الاستجابة: ويقصد بذلك إعادة ترتيب بدائل الاستجابة لفقرات الاختبار التحصيلي في كراستين مختلفتين للاستجابة، بحيث يكون الاختلاف في موقع البدائل وبواقع (٣٣) بديلا في كل كراسة استجابة، وذلك بترقيم المثيرات - بدائل الاستجابة - من (١ : ٣٣) بديلا في كراستي الاستجابة، أي أنّ هنالك متصلا رتبت عليه بدائل الاستجابة في كراستي الاستجابة؛ وذلك من أجل مقارنة الفقرات المتماثلة، حيث إنّ الاختلاف الوحيد هو موقع الاستجابة الصحيحة في كراستي الاستجابة.

- **منهجية الدراسة وإجراءاتها**.
- **المنهج**:

استخدم الباحث المنهج التجريبي في البحث، ويمكن تعريف البحث التجريبي بأنه تغيير عمدي ومضبوط للشروط المحددة لحدث ما، مع ملاحظة التغيرات الواقعة في ذات الحدث وتفسيرها. ويعدّ أقرب مناهج البحوث لحل المشكلات بالطريقة العلمية، ويعبر عن محاولة للتحكم في جميع المتغيرات قدر الإمكان والعوامل الأساسية باستثناء متغير واحد، حيث قام الباحث بتطويعه أو تغييره وهو في هذه الدراسة إعادة ترتيب بدائل الاستجابة، بهدف تحديد وقياس تأثيره في صعوبة فقرة الاختيار من متعدد.

- **مجتمع الدراسة**:

تكون مجتمع الدراسة من جميع طلبة الصف التاسع الأساسي المنتظمين في المدارس الحكومية التابعة لمديرية التربية والتعليم في محافظة جرش بالمملكة الأردنية الهاشمية، والبالغ عددهم (٢١٣٢) طالبا وطالبة، حسب إحصائيات قسم التخطيط للعام الدراسي ٢٠١٢/٢٠١٣م.

- **عينة الدراسة**:

تكونت عينة الدراسة من (٦٠٠) طالبٍ وطالبة، إذ اختيرت (١٠) مدارس عشوائيا، وكان هدف الباحث من اختيار هذه المدارس؛ هو تمثيلها لمستويات مختلفة من قدرات الطلبة، وتشكل عينة الدراسة ما نسبته (٢٨٪) من مجتمع الدراسة.

- **أداة الدراسة**:

تمّ إعداد اختبار تحصيلي في مادة الرياضيات لطلبة الصف التاسع الأساسي، وقد اقتضى بناء أسئلة الاختبار ووضعه في صيغته النهائية، الاسترشاد بالأسس المتبعة في تصميم اختبارات التحصيل الصفية (Gronlund & Linn, 1990, PP. 189 - 166). وبعد تحديد الغرض من الاختبار، تم تحديد الموضوعات المراد قياسها، ومن ثمّ تحليل المحتوى للوحدة المعنية، ثم صياغة النتاجات التعليمية للوحدة الدراسية سابقة الذكر، ثم إعداد جدول المواصفات، وبعد ذلك تمّ صياغة (٢٠) فقرة من نوع الاختيار من متعدد بأربعة بدائل، لقياس النتاجات المقصودة.

- **صدق الأداة**:

وللتأكد من صدق محتوى الاختبار، قام الباحث بعرض الصورة الأولية للاختبار، وجدول المواصفات، وتحليل المحتوى، والنتاجات التعليمية، على مجموعة من المحكمين؛ لأخذ وجهات نظرهم في مدى صدق الفقرة الاختبارية في قياس الهدف السلوكي الم

وقد طبقت صورة الاختبار الأولية على عينة استطلاعية مكونة من(٨٠) طالباً من غير مدارس عينة الدراسة، وأعطي الطلبة الوقت الكافي للإجابة عن فقرات الاختبار بصورته الأولية، الذي قدر وقته بحصة دراسية واحدة (٤٥) دقيقة. ثمّ جرى تحليل أولي لفقرات الاختبار؛ وذلك للكشف عن الفقرات التي تحتاج إلى تعديل، أو حذف في ضوء معاملات صعوبتها وتمييزها، وقد تراوحت قيم معاملات الصعوبة في نموذج الصورة الأولية للاختبار بين(٠,٣٤ – ٠,٧٥) بوسيط حسابي مقداره (٠,٤٥)، بينما تراوحت قيم معاملات التمييز بين(٠,٢٥ – ٠,٦٥)، بوسيط حسابي مقداره(٠,٥٠)، وعند النظر بالإحصائيات المقترحة من قبل ايبل(Eble, 1972, PP. 187 - 210)، التي تتلخص بإمكانية قبول الفقرات التي تزيد معاملات تمييزها عن(٠,١٩)، وقبول الفقرات التي تتراوح معاملات صعوبتها بين(٠,٣٠ – ٠,٨٠)، قام الباحث بقبول جميع فقرات الاختبار. وقد تمّ التحقق من صدق أداة الدراسة من خلال الصدق المرتبط بمحكٍ، عن طريق حساب معامل ارتباط بيرسون بين علامات الطلبة في الاختبار المعّد وعلاماتهم المدرسية، إذ بلغت قيمته(٠,٩٣).

- **ثبات الأداة**:

جرى حساب معامل ثبات الاختبار التحصيلي بمعادلة كودر – ريتشاردسون ٢٠ (KR – 20) فكانت قيمته(٠,٨٩)، وهذا مؤشر على تحقق ثبات مقبول لأداة الدراسة.

وبعد الخروج بالصورة النهائية للاختبار، قام الباحث بتجريد نموذج الصورة الأولية للاختبار من جميع البدائل، ليصبح نموذج الاختبار مكون من متون فقط، ثم جرى إعداد كراستين للإجابة مختلفتين في موقع الـ(٣٣) بديلا، التي أخذت من بدائل الاستجابة لنموذج الصورة الأولية للاختبار، إذ جرى إعادة ترتيب بدائل الاستجابة في كراستي الاستجابة، وتمّ ترقيم المثيرات(بدائل الاستجابة) من (١ : ٣٣) في كراستي الاستجابة، بمعنى أنّ هناك متصلا رتبت عليه بدائل الاستجابة (المثيرات Stimuli)، وكان هدف الباحث من هذا التصميم لكراستي الاستجابة؛ من أجل المقارنة المباشرة للفقرات المتماثلة، حيث الاختلاف الوحيد في الفقرات يكون في موقع الاستجابة الصحيحة في كراستي الاستجابة.

وتجدر الإشارة إلى أنّ بعض الفقرات في نموذج الاختبار لها نفس موقع الاستجابة الصحيحة في كراسة الاستجابة الواحدة، فمثلاً الفقرة الأولى والفقرة العشرون في نموذج الاختبار تختلفان فقط في طبيعة المشكلة التي تطرحها كل من هاتين الفقرتين، إلا أنّ لهما موقع الاستجابة الصحيحة نفسه في كراسة الاستجابة الأولى وهو الموقع السادس، وكان غرض الباحث من ذلك، هو مقارنة صعوبة فقرات نموذج الاختبار ضمن كراسة الاستجابة الواحدة نفسها.

- **جمع البيانات**:

بعد إعداد نموذج الاختبار بصورته النهائية وكراستي الاستجابة، وفي موعد امتحان الشهر الأول للامتحانات المدرسية، جرى تطبيق أداة الدراسة على أفراد عينة الدراسة وبشكل جماعي في الصف، وقد طلب من المفحوصين

الاختيار من قائمة مكونة من(٣٣) بديلاً، بحيث كان الخيار الصحيح مطابقًا للتصور البصري لفقرة الاختبار، وقد تمّ توزيع نموذج الاختبار وكراستي الاستجابة المختلفتين في موقع الاستجابة الصحيحة، ليأخذ الطالب الأول في الجلوس نموذج الاختبار وكراسة الاستجابة الأولى، والطالب الثاني يأخذ النموذج نفسه وكراسة الاستجابة الثانية، وهكذا حتى وزع نموذج الاختبار وكراستي الاستجابة على جميع الطلبة، وأعطي الطلبة الوقت الكافي للإجابة عن جميع أسئلة الاختبار، وكان عدد الطلبة الذين أجابوا عن نموذج الاختبار في كراسة الاستجابة الأولى(٣٠٠) طالبٍ وطالبة، وكان عدد الطلبة الذين اجابوا عن نموذج الاختبار مستخدمين كراسة الاستجابة الثانية(٣٠٠) طالبٍ وطالبة. وبعد أن تمت إجراءات التطبيق صححت الأوراق، وأعطي صفر للفقرة الخطأ، ودرجة واحدة للفقرة الصحيحة، وبذلك تكون العلامة الكلية للمفحوص هي مجموع الإجابات الصحيحة عن فقرات الاختبار، ثمّ أدخلت البيانات في ذاكرة الحاسوب باستخدام برنامج (SPSS) لتحليلها، والحصول على الإحصائيات التي ساعدت الباحث في الإجابة عن سؤال الدراسة.

- **الأساليب الإحصائية**:

قام الباحث بحساب معامل الصعوبة لكل فقرة من فقرات نموذج الاختبار ضمن كراستي الاستجابة، والذي هو نسبة الذين أجابوا عن الفقرة إجابة صحيحة ممن حاولوا الإجابة عنها(معامل السهولة = ١ – معامل الصعوبة). واستخدم اختبار (Independent Sample t – test) لفحص فرضية تساوي أوساط معاملات الصعوبة، واستخدم اختبار ولكوكسون المعروف باسم (Wilcoxon Matched Pairs Signed Ranks Test)، وسبب استخدام هذا الاختبار؛ هو لفحص مدى تطابق توزيع قيم معاملات الصعوبة لفقرات نموذج الاختبار في كراستي الاستجابة، واستخدمت الفروق بين القيم المستخرجة والرسومات التوضيحية لعرض ومناقشة نتائج الدراسة.

- **عرض النتائج ومناقشتها**:

وللإجابة عن سؤال الدراسة: هل هناك أثر في إعادة ترتيب بدائل الاستجابة في صعوبة فقرة الاختيار من متعدد؟. قام الباحث بإيجاد قيم معاملات الصعوبة للفقرات في نموذج الاختبار وفق كراستي الاستجابة، وتمّ تحديد موقع الاستجابة الصحيحة لكل فقرة في كراستي الاستجابة، ثمّ جرى أخذ الفروق المطلقة بين مواقع الاستجابة الصحيحة في كراستي الاستجابة، وجرى أخذ الفروق بين قيم معاملات الصعوبة للفقرات من استجابة الطلبة عن نموذج الاختبار وفق كراستي الاستجابة، ويبين الجدول(١) ذلك:

ويلاحظ من الجدول أن قيم معاملات الصعوبة لفقرات نموذج الاختبار وفق استجابة الطلبة حسب كراسة الاستجابة الاولى قد تراوحت بين(0.20 – 0.97) بوسيط حسابي مقداره(0.64)، أي أنّ معامل السهولة يساوي(0.36)، على اعتبار أنّ (معامل السهولة = ١ – معامل الصعوبة)، بينما تراوحت قيم معاملات الصعوبة لفقرات نموذج الاختبار وفق استجابة الطلبة حسب كراسة الاستجابة الثانية بين(0.17 – 0.94) بوسيط حسابي مقداره(0.62) أي أنّ معامل السهولة يساوي(0.38)، بمعنى أنّ فقر[illegible] من جرّاء استجابة الطلبة وفق

كراسة الاستجابة الأولى أسهل، مقارنةً بفقرات نموذج الاختبار نفسه وفق استجابة الطلبة من كراسة الاستجابة الثانية. وتم استخدام اختبار(t – test) للكشف عن الفروق بين قيم معاملات الصعوبة لفقرات نموذج الاختبار وفق استجابة الطلبة من كراستي الاستجابة، ويبين الجدول(2) نتائج التحليل.

جدول (١) قيم معاملات الصعوبة لفقرات نموذج الاختبار ومواقع الاستجابات الصحيحة في كراستي الاستجابة.

رقم الفقرة	معاملات الصعوبة لفقرات نموذج الاختبار وفق:		موقع الاستجابة الصحيحة في:		الفرق بين قيم معاملي الصعوبة في نموذج الاختبار حسب كراستي الاستجابة	الفرق المطلق بين مواقع الاستجابة الصحيحة في كراستي الاستجابة.
	كراسة الاستجابة الأولى	كراسة الاستجابة الثانية	كراسة الاستجابة الأولى	كراسة الاستجابة الثانية		
١	0.95	0.94	٦	١١	0.01	٥
2	0.73	0.72	١٧	١٥	0.01	٢
3	0.83	0.81	٢٧	٢٢	0.02	٥
4	0.70	0.64	١٣	١٦	0.06	٣
5	0.97	0.93	٢٨	٣٠	0.04	٢
6	0.67	0.66	٥	٩	0.01	٤
7	0.65	0.51	٣	٢٦	0.14	٢٣
8	0.77	0.76	٢	٢٣	0.01	٢١
9	0.87	0.87	٧	٢١	0.00	١٤
10	0.33	0.35	٢٢	١٩	-0.02	٣
11	0.24	0.17	٢٨	٣٠	0.07	٢
12	0.37	0.35	٣٠	١	0.02	٢٩
13	0.20	0.22	١٨	٦	-0.02	١٢
14	0.95	0.93	١٤	١٠	0.02	٤
15	0.75	0.78	١٣	١١	-0.03	٢
16	0.43	0.40	٤	١١	0.03	٧
17	0.37	0.32	٢٩	٢٨	0.05	١
18	0.70	0.65	٤	١٢	0.05	٨
19	0.94	0.91	٨	٣	0.03	٥
20	0.45	0.46	٦	٨	-0.01	٢
الوسط الحسابي	٠.٦٤	٠.٦٢			٠.٠٢٥	٧.٧٠

ويلاحظ من الجدول عدم وجود فرق ذي دلالة إحصائية(α=0.05) بين قيم معاملات صعوبة فقرات الاختيار من متعدد يعزى إلى إعادة ترتيب بدائل الاستجابة في كراستي الاستجابة، إذ بلغت قيمة(t) (0.31) باحتمال(0.76). واستخدم اختبار ولكوكسون لفحص فرضية تطابق توزيع قيم معاملات الصعوبة لفقرات نموذج الاختبار في كراستي الاستجابة، ويبين الجدول (٣) نتائج التحليل:

جدول (٢) نتائج اختبار (t) لاختبار الفرق بين وسطي معاملات صعوبة الفقرات وفق استجابة الطلبة من كراستي الاستجابة.

نموذج الاختبار وفق الاستجابة من:	المتوسط الحسابي	الإنحراف المعياري	قيمة (t) المحسوبة	درجات الحرية	قيمة الاحتمال
كراسة الاستجابة الاولى	0.644	0.252	0.31	38	0.76
كراسة الاستجابة الثانية	0.619	0.255			

جدول (٣) نتائج اختبار ولكوكسون Wilcoxon لمقارنة قيم معاملات الصعوبة لفقرات نموذج الاختبار وفق الاستجابة من كراستي الاستجابة.

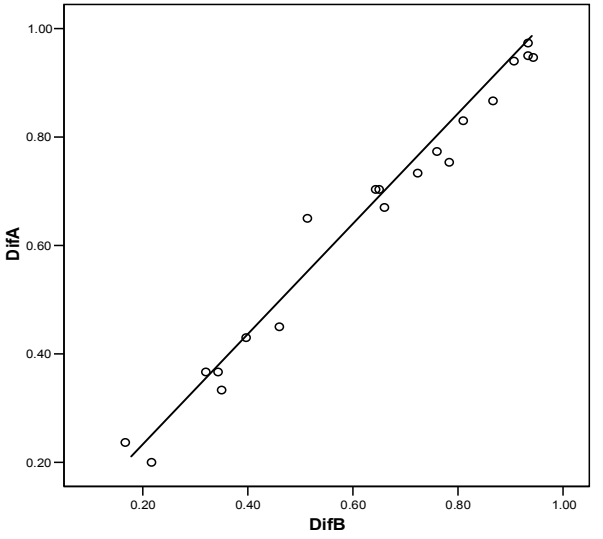
نموذج الاختبار وفق الاستجابة من:	الوسط للرتب	عدد الرتب السالبة	عدد الرتب الموجبة	عدد الرتب المتشابهة	قيمة الإحصائي (Z)	احتمال (Z)
كراسة الاستجابة الاولى	10.8	15	4	١	2.70	0.007
كراسة الاستجابة الثانية	7.0					

ويلاحظ من نتائج التحليل أنّ توزيع قيم معاملات صعوبة فقرات نموذج الاختبار جرّاء استجابة الطلبة من كراستي الاستجابة مختلفة، إذ بلغت قيمة(Z) (2.70) باحتمال(0.00٧)، وهي دالة إحصائيا عند مستوى الدلالة(α=0.05).

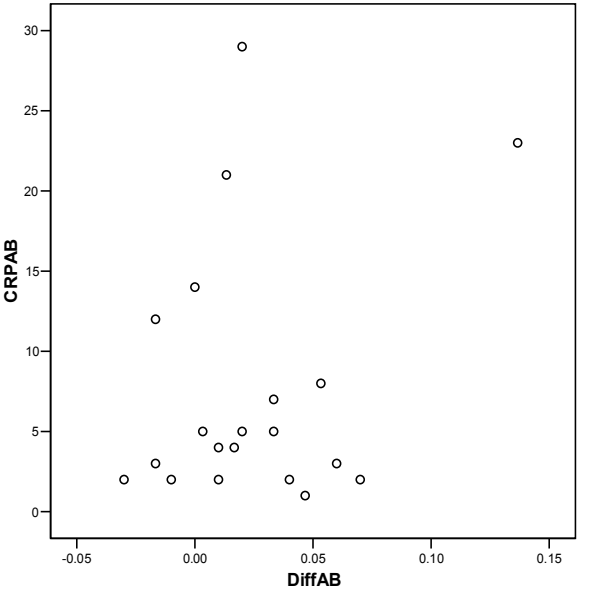
وجرى رسم شكل العلاقة بين أزواج معاملات الصعوبة للفقرات، إذ تمّ تمثيل قيم معاملات الصعوبة وفق الاستجابة من الكراسة الثانية على المحور السيني(DifB)، وقيم معاملات الصعوبة وفق الاستجابة من الكراسة الأولى على المحور الصادي(DifA)، ويلاحظ من خلال الشكل(١) أنّ العلاقة تقريبا خطية بين الأزواج، وقد كان معامل ارتباط بيرسون بين قيم معاملات الصعوبة لفقرات نموذج الاختبار وفق الاستجابة من الكراستين يساوي(0.989).

ولتقييم التغيرات المحتملة في قيم معاملات الصعوبة يعزى لتغير موقع الاستجابة، جرى رسم الفروق المطلقة في موقع الاستجابة الصحيحة مقابل الفروق في صعوبة الفقرات وفق استجابة الطلبة من كراستي الاستجابة، إذ تمّ إنشاء متغيرين: المتغير الأول يشير إلى الفروق في قيم معاملات الصعوبة لأزواج الفقرات، وذلك عن طريق طرح قيمة معاملات صعوبة الفقرات في نموذج الاختبار وفق الاستجابة من كراسة الاستجابة الثانية، من قيمة معاملات صعوبة الفقرات المناظرة لها في نموذج الاختبار وفق استجابة الطلبة من كراسة الاستجابة الأولى، وجرى تمثيل ذلك على محور السينات(المحور الأفقي)(DiffAB)، أما المتغير الثاني فيشير إلى الفرق المطلق في موقع الاستجابة الصحيحة لأزواج الفقرات المتناظرة، وذلك عن طريق طرح القيمة العددية لموقع الاستجابة الصحيحة في كراسة الاستجابة الثانية، من القيمة العددية لموقع الاستجابة الصحيحة في كراسة الاستجابة الأولى، وجرى تمثيل ذلك على محور الصادات(المحور العامودي)(CRPAB)، وجرى إيجاد معامل الارتباط بين المتغيرين السابقين، وكانت قيمته تساوي(0.243)، ويبين الشكل(٢) ذلك:

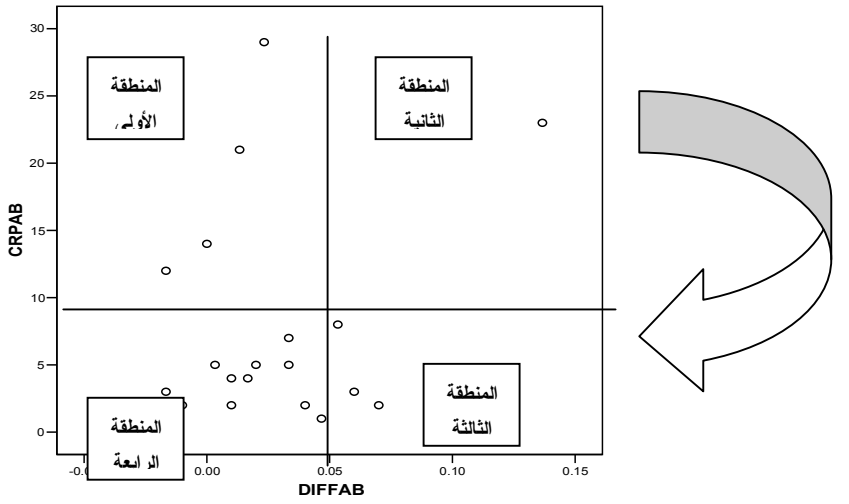
ويلاحظ من الشكل (٢) وفي غياب أي خط مرتبط بين المتغيرين السابقين، أنّ فحص العلاقة في هذا الشكل يقدم مجموعة من الملاحظات المثيرة، ولتفسير هذه العلاقة، تمّ إنشاء رسم يظهر صورة أخرى للشكل(2) وذلك بإضافة خطين متقاطعين لتقسيم الشكل إلى أربع مناطق؛ وذلك لعزل الفقرات، ويبين الشكل (٣) ذلك :



شكل (١) الانتشار بين قيم معاملات الصعوبة لفقرات نموذج الاختبار وفق استجابة الطلبة من كراستي الاستجابة.



شكل (٢) رسم الفروق المطلقة في موقع الاستجابة الصحيحة في كراستي الاستجابة مقابل الفروق في صعوبة الفقرات.



شكل (٣) رسم توضيحي يبين عزل فقرات الاختبار حسب المتغيرين السابقين.

ويلاحظ من الشكل(٣) وباتجاه عقارب الساعة ابتداءً بالاتجاه الأيسر العلوي الملاحظات التالية:

- المنطقة الأولى: وتقع فيها الفقرتان(٨ ، ١٢)، ويلاحظ أنّ الفقرة(٨) في نموذج الاختبار لاستجابة الطلبة في كراسة الاستجابة الثانية أصعب من استجابة الطلبة للفقرة نفسها في كراسة الاستجابة الأولى، إذ ظهرت الاستجابة الصحيحة متأخرة في كراسة الاستجابة الثانية مقارنة بموقعها في كراسة الاستجابة الأولى. أما الفقرة(١٢) فيلاحظ أنّ هذه الفقرة في نموذج الاختبار لاستجابة الطلبة في كراسة الاستجابة الثانية أصعب من استجابة الطلبة لهذه الفقرة حسب كراسة الاستجابة الأولى، رغم أنّ الاستجابةِ الصحيحة لهذه الفقرة جاءت متقدمة في كراسة الاستجابة الثانية مقارنة بموقعها في كراسة الاستجابة الأولى.
- المنطقة الثانية: وتقع فيها الفقرة(٧) فقط، ويلاحظ أنّ الفقرة(٧) في نموذج الاختبار من استجابة الطلبة في كراسة الاستجابة الثانية أكثر صعوبة من استجابة الطلبة حسب كراسة الاستجابة الأولى، إذ ظهرت الاستجابة الصحيحة متأخرة في كراسة الاستجابة الثانية مقارنة بموقعها في كراسة الاستجابة الأولى.
- المنطقة الثالثة: وتقع فيها الفقرات(٤ ، ١١ ، ١٨) ، ويلاحظ أنّ هذه الفقرات في نموذج الاختبار لاستجابة الطلبة في كراسة الاستجابة الثانية أكثر صعوبة من استجابة الطلبة حسب كراسة الاستجابة الأولى، إذ ظهرت الاستجابة الصحيحة متأخرة في كراسة الاستجابة الثانية مقارنة بموقعها في كراسة الاستجابة الأولى.
- المنطقة الرابعة: وتقع فيها الفقرات(١، ٢، ٣، ٥، ٦، ٩، ١٠، ١٣، ١٤، ١٥، ١٦، ١٧، ١٩، ٢٠)، وعلى الرغم من لاستجابات الصحيحة وصعوبة

فقرات الاختبار، فإنّ الفروق في صعوبة الفقرات المتناظرة ليست كبيرة، فمثلا الفقرة (١) في نموذج الاختبار لاستجابة الطلبة في كراسة الاستجابة الثانية أصعب من استجابة الطلبة حسب كراسة الاستجابة الأولى، وقد كان فارق الصعوبة(0.01)، رغم أنّ موقع الاستجابة الصحيحة في كراستي الاستجابة مختلفة، إذ كان الفرق المطلق بين موقعي الاستجابة يساوي(٥)، حيث ظهرت الاستجابة الصحيحة متأخرة في كراسة الاستجابة الثانية مقارنة بموقعها في كراسة الاستجابة الأولى. ويلاحظ أنّ في الفقرة(٩) كان معامل الصعوبة وفق الاستجابة في كراستي الاستجابة متساويًا، على الرغم من أنّ الاستجابة الصحيحة جاءت متأخرة في كراسة الاستجابة الثانية مقارنة بموقعها في كراسة الاستجابة الأولى. وعند النظر إلى الفقرة(٢٠) نلاحظ أنّ هذه الفقرة في نموذج الاختبار لاستجابة الطلبة في كراسة الاستجابة الثانية أسهل من استجابة الطلبة لهذه الفقرة حسب كراسة الاستجابة الأولى، إذ ظهرت الاستجابة الصحيحة متأخرة في كراسة الاستجابة الثانية مقارنة بموقعها في كراسة الاستجابة الأولى.

ويلاحظ من الشكل(٣) وجود بعض أزواج الفقرات المتطرفة(outliers) في الصعوبة، مثل زوج الفقرة(١١)، إذ تظهر هذه الفقرة فرقا قليلا من كراسة لكراسة أخرى في موقع الاستجابة، بالمقابل يلاحظ أنّ زوج هذه الفقرة يظهر قيمة عالية في تغير قيمة صعوبة الفقرة، ويلاحظ أنّ زوج الفقرة(١٣) يظهر فرقا كبيرا من كراسة لكراسة أخرى في موقع الاستجابة، لكن يلاحظ أنّ زوج هذه الفقرة يظهر قيمة قليلة في تغير صعوبة الفقرة. ويلاحظ أنّ أزواج الفقرات(٨ ، ١٢) تظهر فروقا قليلة في صعوبة الفقرة، على الرغم من أنّ هذه الأزواج تظهر تغيرات متطرفة في موقع الاستجابة الصحيحة، وعند النظر إلى زوج الفقرة(٧)، نلاحظ وجود تطرف واضح في موقع الاستجابة، إذ بلغ الفرق المطلق بين موقعي الاستجابة في الكراستين (٢٣)، وأظهرت هذه الفقرة قيمة متطرفة في صعوبة الفقرة، إذ بلغ الفرق بين معاملي صعوبة هذه الفقرة لاستجابة الطلبة في كراستي الاستجابة(0.14).

- **الاستنتاجات والتوصيات**:

أوضحت النتائج الخاصة بمقارنة قيم معاملات الصعوبة، عدم وجود فرق بين قيم معاملات صعوبة فقرات الاختبار من متعدد يعزى إلى إعادة ترتيب بدائل الاستجابة في كراستي الاستجابة، إذ كانت المتوسطات الحسابية لقيم معاملات الصعوبة متقاربة، وقد يعزى ذلك إلى أنّ الفقرات في نموذج الاختبار هي الفقرات نفسها، والاختلاف الوحيد كان في طريقة عرض بدائل الاستجابة الصحيحة بصريا على المفحوصين والمتمثلة في كراستي الاستجابة.

وعند استخدم اختبار ولكوكسون تبين أنّ توزيع قيم معاملات صعوبة فقرات نموذج الاختبار جرّاء استجابة الطلبةٍ في كراستي الاستجابة مختلفة، إذ كانت بعض أزواج الفقرات تختلف قليلا من ناحية التغيرات في قيم معاملات الصعوبة من كراسة استجابة إلى أخرى، وقد يكون بسبب إعادة تغير موقع الاستجابة الصحيحة في كراستي الاستجابة. ومن خلال شكل العلاقة بين أزواج معاملات الصعوبة الشكل(١) [illegible] تقريبا خطية بين أزواج الفقرات،

وهذا ما عززه معامل ارتباط بيرسون بين قيم معاملات الصعوبة لفقرات نموذج الاختبار حسب الاستجابة في الكراستين، إذ بلغت قيمته(0.989)، وهذا يعني وجود علاقة قوية بين تقديرات معاملات الصعوبة في فقرات الاختبار لاستجابة الطلبة في كراستي الاستجابة، ولعل هذا يعود إلى أنّ نموذج الاختبار المطبق هو النموذج نفسه، ولم يتغير على فقراته أي شيء، ولكن تغيرت مواقع الاستجابة الصحيحة في كراستي الاستجابة، مما أظهر تغيرات قليلة في قيم معاملات صعوبة بعض الفقرات.

وعند تقييم الباحث للتغيرات المحتملة في قيم معاملات الصعوبة يعزى إلى تغير موقع الاستجابة، وبالنظر في الشكل(٣)، يمكن القول أنّ تغيرات قليلة في موقع الاستجابة الصحيحة قد أظهر تأثيراً قليلا في صعوبة الفقرة. على أية حال إنّ ملاحظة التأثير الطفيف على صعوبة الفقرة قد يكون مربكا بعض الشيء، كون تأثيره متقلب من ناحية الاتجاه، وبالتالي لا يمكن الاستنتاج من هذه البيانات أن إعادة تغيير موقع الاستجابة الصحيحة في كراسة الاستجابة، يمكن أن يؤثر في مستوى صعوبة فقرة الاختبار.

وعلى الرغم من أنّ توزيع قيم معاملات صعوبة فقرات نموذج الاختبار جرّاء استجابة الطلبة في كراستي الاستجابة مختلفة، وهذا ما أظهرته نتائج اختبار ولكوكسون، فإنه لم يكن هناك نمط ثابت فيما يتعلق بتغيير موقع الاستجابة الصحيحة، فمثلا لوحظ أن الفقرتين (٥ ، ١١) تختلفان في موقع الاستجابة الصحيحة فقط في موقعين، إلا أنّ زوج الفقرة(٥) أسهل من زوج الفقرة (١١).

وتفضي هذه النتائج إلى أنّ التغيرات في موقع الاستجابة الصحيحة قد تُعدّ ممارسة خطيرة جدا خصوصا في الاختبارات المتكافئة، ويوصي الباحث بتوخي الحذر عندما يقوم باني الاختبار بتغيير مواقع الاستجابات في الاختبارات المتكافئة، وأظهرت نتائج هذه الدراسة أنّ صعوبة فقرة الاختيار من متعدد تتأثر بدرجة قليلة أو بدرجة كبيرة بتغير موقع الاستجابة الصحيحة، أي أنّ الآثار متقلبة، وقد تكون هذه النتيجة سببا للاهتمام الكبير، عندما نُعدّل الفقرات ونستعملها كفقرات متكافئة، وربما يكون من المرغوب تجنب إعادة ترتيب بدائل أي اختبار، عندما تكون مستويات صعوبة الأهداف هي المقصودة.

ونظراً للاتجاه غير المعروف في أثر إعادة ترتيب بدائل الاستجابة في صعوبة الفقرات، والتغيرات الحاصلة في قيم معاملات الصعوبة لفقرات الاختبار في هذه الدراسة، فإنّ الباحث يوصي بإجراء المزيد من الأبحاث المستقبلية في هذا المجال.

- **قائمة المراجع** :

١. الصراف، قاسم علي. (٢٠٠٢). **القياس والتقويم في التربية والتعليم**. الكويت: دار الكتاب الحديث.

٢. أمبو سعيدي، عبد الله والعفيفي، منى (٢٠٠٤) : أثر تغيير موقع المشتت القوي (Strong Distractor) على بعض الخصائص السيكومترية لأسئلة الاختيار من متعدد في مادة الفيزياء. مجلة العلوم التربوية والنفسية، جامعة البحرين، ٥(٢)، ١٦٩ – ١٩١.

٣. أمبو سعيدي، عبد الله (٢٠٠٥) : أثر تغيير موقع المشتت القوي (Strong Distractor) عند استخدام تقنية إبعاد البدائل الخطأ في صعوبة الفقرات لأسئلة الاختيار من متعدد في مادة الفيزياء. مجلة العلوم التربوية والنفسية، جامعة البحرين، ٦(٣)، ٨٦ – ١٠٥.

٤. دودين ، حمزة (٢٠٠٥): تدريس استراتيجيات تقديم الاختبارات. مجلة التربية. قطر ، ١٠٢ – ١١٧.

٥. عدس، عبد الرحمن (٢٠٠٢) : دليل المعلم في بناء الاختبارات التحصيلية. عمان: دار الفكر للطباعة والنشر والتوزيع.

٦. عودة ، أحمد سليمان (٢٠٠٥): القياس والتقويم في العلمية التدريسية. اربد، الأردن، دار الامل.

7. Aiken, L. (1987). Testing with multiple - choice items. Journal of Development in Education, 20 (4), PP. 44 – 57.

8. Angoff, H. (1971). Scales, norms, and equivalent scores. In R. L. Thorndike (Ed.). *Education Measurement* (2nd ed.). Washington Dc: American Council on Education. PP. 508 – 600.

9. Blunch, N. (1984). Positional bias in multiple – choice questions. Journal of Marketing Research, 21, PP. 216 – 220.

10. Brown, F. G. (1983). Principle of education and psychological testing (3nd ed.). New York: Holt, Rinehart and Winston.

11. Cassels, T. & Johnstone, A. (1984). The effect of language on student performance on multiple choice tests in chemistry. Journal of Chemical Education, 61, PP. 613 – 615.

12. Cizek, G.. J. (1991). The Effect of altering the position of options in a multiple – choice examination. *Educational and Psychological Measurement, 54*(1),8 – 20.

13. Crehan, K. & Haladyna, T.(1991). The validity of two items writing rules. *Journal of Experimental Education*, 59(2), PP.183 – 192.

14. Eble, R. (1972). *Essentials of educational measurement*. New Jersey: prentice – Hall, Inc, PP. 187 – 210.

15. Ebel, R. L. and Frisbie, D. A. (1986). Essentials of educational measurement (4th ed.). New Jersey: Prentice-Hall.

16. Gronlund, N. & Linn, R. (1990). *Measurement and evaluation in teaching* (6th ed.). New York: Macmillan Publishing Company, PP. 166 - 189.

17. Hambleton, R. & Swaminathan, H. (1985). *Item Response Theory: Principles and applications*. Boston: Kluwer Nijhoff publishing.

18. Hambleton, R. K. (1983). *Item response theory* .Boston: Kluwer – Nijhoff publishing.

19. Haladyna, T., Downing, S. & Rodriguez, M. (2002). A Review of multiple- choice item writing guidelines for classroom assessment. *Applied Measurement in Education, 15* (3), pp. 309 – 334.

20. Johnstone, A., MacGuire, P., Friel, S. & Morrision, E. (1983). Criterion referenced testing in science: Thought, worries, and suggestions. *School Science Review, 64* (229), PP. 626 – 634.

21. Lane, D. S. & Bull, K. S. (1987). The effects of Knowledge of item arrangement, Gender, and statistical and cognitive item difficulty on test performance. *Educational and Psychological Measurement, 47*, PP. 865 – 879.

22. Marcus, A. (1963). The effects of correct response location on the difficulty level of multiple – choice questions. *Journal of Applied Psychology, 47*(1), PP. 48 – 51.

23. Plake, B., Thompson, A. & Lowry, S. (1981). Effects of item arrangement and test anxiety on two scoring methods. *The Journal of Experimental Education, 49*, PP. 214 – 219.

24. Roid, G. & Haladyna, T. (1982). *A Teachnology for test item writing*. New York: Academic Press, PP. 49 – 55.

25. Straton,R. G. & Catts, R. M. (1980). A comparison of two , three and four choices item tests given a fixed total number of choices. *Educational and Psychological Measurement,40*, pp. 357 – 365.

❋❋❋❋❋❋❋