أثر مخالفة قواعد صياغة فقرات الاختيار من متعدد على الخصائص السيكومترية للاختبار وفقراته

الدكتور إبراهيم محمد يعقوب الدكتور باسل خميس أبوفودة

قسم علم النفس الإرشادي والتربوي

جامعة اليرموك - الأردن

الملخص

هدفت الدراسة إلى معرفة أثر مخالفة قواعد صياغة فقرات الاختيار من متعدد في الخصائص السيكومترية للفقرات (الصعوبة، والتمييز)، والخصائص السيكومترية للاختبار (الثبات، والصدق). ولتحقيق ذلك تمّ بناء اختبار من نوع الاختيار من متعدد في مبحث الرياضيات، تكون من (40) فقرة ذات أربعة بدائل، ثم جرى صياغة فقرات الاختبار في نموذج آخر مخالف لقواعد الصياغة. جرى تطبيق نموذجي الاختبار على عينة مكونة من (300) طالب، بواقع (150) طالباً لكل نموذج، وفي ظروف صفية عادية.

وبعد إجراءات التطبيق حسبت قيم معاملات صعوبة الفقرات وتمييزها في نموذجي الاختبار، وكذلك معاملات صدقها وثباتها. فأظهرت النتائج أن الفرق كان دالاً إحصائياً بين صعوبة وتمييز الفقرات التي لم تحتو المخالفات في قواعد الصياغة، أو

التي احتوتها، وبينت النتائج أن الفرق بين قيم معاملي ثبات الاتساق الداخلي في نموذجي الاختبار كان دالاً إحصائياً، وكان الفرق بين قيم معاملي صدقها دالاً أيضاً.

الكلمات المفتاحية:

مخالفة قواعد صياغة فقرات الاختيار من متعدد، صعوبة الفقرة، تمييز الفقرة، معامل ثبات الاتساق الداخلي، معامل صدق المحك التلازمي.

مقدمة:

يتعرض الطالب في المراحل التعليمية المختلفة إلى عدد كبير من الاختبارات باختلاف أشكالها؛ إذ تتعدد أشكال فقرات الاختبارات التحصيلية من حيث صياغتها، فمنها ما يتطلب من المفحوص أن يعطي الإجابة بلغته الخاصة مثل: الأسئلة المقالية، 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: مستوى الهدف المراد قياسه، وعمر المفحوص، وغرض الاختبار (Gronlund & Linn, 1990,PP.166 - 189).

وتعّد فقرات الاختيار من متعدد من أفضل أنواع الفقرات الموضوعية، وأكثرها شيوعاً واستخداماً في الاختبارات التحصيلية (Aiken,1987,PP.44 - 57). وتتكون جميع فقرات الاختيار من متعدد من العناصر نفسها، وأول هذه العناصر هو المتن أو أرومة السؤال (Stem)، والذي 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 بمثابة إجابات محتملة للسؤال الوارد في المتن، وأحد هذه البدائل هو الإجابة الصحيحة، أما بقية البدائل فتعرف بالمموهات أو المشتتات (Distracters or Foils) (Roid & Haladyna, 1982,PP.49 - 55).

ويتوقف مستوى أداء المفحوص في الاختبار، على خصائص الاختبار والمفحوص، ومن خصائص فقرة الاختيار من متعدد المؤثرة في مستوى أداء المفحوص ما يعرف بصعوبة الفقرة، وقدرتها التمييزية، وجاذبية المموهات، ومدى تأثرها بعامل التخمين (Plake, Thompson & Lowry, 1981, PP. 214 - 219). وتتأثر القدرة اللازمة لإجابة الفقرة، بعوامل فنية تتعلق بدرجة الوضوح في صياغة المتن، ومدى تجانس

بدائل الفقرة، ودرجة علاقتها مع متن السؤال، ومدى تحرر الفقرة بشكل عام من خلال في صياغتها (Lane & Bull,1987,PP. 865-879).

وقد كشف منتزر(Mentzer,1982,PP.437-446)عدم وعي المعلمين لقواعد صياغة فقرات الاختيار من متعدد، وأوضح سارناكي (Sarnacki, 1979, PP. 252 - 279) عدم دراية بعضهم بالنتيجة، ومن ثمَّ تكون اختباراتهم عرضةً للتلوث بفقرات ضعيفة الصياغة، ووجود مثل هذا الخلل قد يؤثر في الخصائص السيكومترية للاختبار وفقراته.

الدراسات السابقة:

بنية الفقرة لها تأثير على جودة الفقرة ومن ثمَّ على الاختبار الذي يتضمن هذه الفقرة، وقد كان هذا الموضوع مدار بحث بعض الدراسات التي تساءلت عن مدى تأثير الخلل الناتج عن مخالفة قواعد صياغة فقرة الاختيار من متعدد على الخصائص السيكومترية للفقرة والاختبار، ففي دراسة أجراها شيس(Chase,1964, PP. 861– 866) في جامعة انديانا في الولايات المتحدة الأمريكية بعنوان "طول البديل النسبي والاستعداد الخاص للاستجابة في فقرة الاختيار من متعدد"، إذ هدفت إلى دراسة تحيز المفحوصين الخاص نحو البديل الأطول. تمّ بناء اختبار مكون من(25) فقرة من نوع الاختيار من متعدد من أربعة بدائل في مساق علم النفس التربوي، وجرى تطبيق الاختبار على عينة مكونة من(48) من الطلبة الجامعيين في جامعة أنديانا يدرسون مساق علم النفس التربوي. ومن أهم نتائج الدراسة أنّ التحيز يظهر نحو البديل الأطول، وأنّ عامل التفاوت بين بديل معين وبين أطول باقي بدائل الفقرة، أدى إلى اختلاف احتمالات الإجابة.

وأجرى ميولر وشروك (Mueller & Schrock, 1982, PP. 314 - 318) في جامعة مونتانا الرسمية وجامعة انديانا في الولايات المتحدة الأمريكية دراسة بعنوان " أثر

مخالفة ثلاث قواعد في صياغة فقرة الاختيار من متعدد"، إذ هدفت إلى معرفة أثر مخالفة قواعد صياغة فقرة الاختيار من متعدد على الخصائص السيكومترية للاختبار وفقراته، بوصفها محاولة لتحديد ضرورة التقيد بهذه القواعد، وقد تضمنت الدراسة ثلاث مخالفات هي: صياغة المتن على شكل سؤال أو جملة غير مكتملة، وموضع الفراغ في المتن، هل هو في البداية أم في النهاية، ووجود أو غياب المؤشرات الدالة على الإجابة الصحيحة في الفقرة. وتكونت عينة الدراسة من(480) طالباً من طلبة المرحلة الجامعية الأولى والثانية في أوهايو وانديانا في الولايات المتحدة الأمريكية، إذ شاهدوا فيلماً، ثمّ أجابوا عن اختبار من نوع الاختيار من متعدد يتضمن المخالفات سابقة الذكر. ومن أهم نتائج الدراسة أنّ تأثير المخالفات كان ضئيلاً، وغير دال إحصائياً على نتائج الاختبار وإحصائياته.

وأجرى ايفانس(Evans, 1984, PP. 141 – 144) في جامعة نوتردام في ولاية انديانا الأمريكية دراسة بعنوان "حنكة الإجابة: فحص استراتيجيات مفاتيح الإجابة"، إذ هدفت إلى معرفة أثر ثلاثة مخالفات في بنية فقرة الاختيار من متعدد في تحصيل الطلبة، وهذه المخالفات هي: البديل الصحيح وهو البديل الأطول بين بدائل الفقرة، والاتساق القواعدي بين المتن والبديل الصحيح للفقرة، وفي حال كون البدائل قيم رقمية، يكون البديل الصحيح هو البديل الأصغر قيمة. شملت عينة الدراسة (73) متخصصاً في الإسعاف الفوري للنوبات القلبية والمنتظمين في دورة تدريبية، وقد قسمت العينة إلى مجموعتين تجريبية طبق عليها خمسة اختبارات من نوع الاختيار من متعدد أثناء الدورة، ومجموعة ضابطة لم يطبق عليها أي اختبار. وفي نهاية الدورة جرى تطبيق على المجموعتين اختباراً تحصيلياً يتكون من(67) فقرة من نوع الاختيار من متعدد لكل فقرة من فقراته أربعة بدائل، منها(52) فقرة تتعلق بالدروة التدريبية متحررة من المخالفات في قواعد الصياغة، و(15) فقرة ضمّنت المخالفات الثلاثة سابقة الذكر بواقع خمس فقرات لكل مخالفة.

ومن أهم نتائج الدراسة وجود فروق ذات دلالة إحصائية بين متوسط أداء المجموعتين لصالح المجموعة التجريبية على كل مخالفة من المخالفات الثلاثة، وعلى الـ(15) فقرة ككل، ولم تكن هناك فروق ذات دلالة إحصائية في أداء المجموعتين على الفقرات الـ(52) المتحررة من المخالفات والخاصة بالدروة التدريبية للعينة ككل، واستنتج من ذلك أنّ حنكة الطالب في الإجابة هي صفة مميزة بوجود مؤشرات في الفقرات.

وقام ويتن (Weiten, 1984, pp. 78 - 174) في كلية ديوبيج في جامعة إلينويز في الولايات المتحدة الأمريكية، بدراسة عنوانها " المخالفات في قواعد صياغة الفقرة في القياس التربوي"، إذ هدفت إلى معرفة أثر ارتكاب أربع مخالفات لِصياغة فقرات الاختيار من متعدد، على صعوبة الفقرة وتمييزها، وثبات الفقرة وصدقها، وكانت المخالفات المُتضمّنة في الاختبار هي: الكلمات المفتاحية المتكررة في متن الفقرة والبديل الصحيح الذي يوحي إلى الإجابة الصحيحة، وعدم الاتساق القواعدي بين متن الفقرة ومموهاتها، واستخدام بدائل غير محتملة التصديق، وجعل البديل الصحيح أطول من باقي بدائل الفقرة.

ولتحقيق ذلك تمّ كتابة اختبار من نوع الاختيار من متعدد يتضمن (64) فقرة، لكل فقرة أربعة بدائل في علم نفس النمو، وذلك بافتراض أنّ هذه الفقرات متحررة من أي مخالفة في صياغتها، ثمّ عمل صورتين للاختبار، الصورة الأولى تتكون من (32) فقرة تتضمن المخالفات الأربعة سابقة الذكر، والصورة الثانية تتكون من (32) فقرة متحررة من المخالفات، ثمّ جرى تطبيق الاختبارين على عينة مؤلفة من (54) طالباً وطالبة من طلبة البكالوريوس في علم النفس في كلية ديوبيج. ومن أهم نتائج الدراسة أنّ الفقرات المتضمنة للمخالفات السابقة، أقل صعوبة وبدرجة دالة إحصائياً مقارنة بالفقرات المتحررة من هذه المخالفات، أما بالنسبة لمؤشرات التمييز ومعاملات الثبات

والصدق فلم تظهر النتائج فروقاً دالة إحصائياً بين الفقرات المتضمنة للمخالفات والمتحررة منها.

وقامت كارتر(Carter,1986, P P. 20 – 23) في جامعة أريزونا في الولايات المتحدة الأمريكية بدراسة بعنوان "الحنكة في الإجابة لكل من الطلبة والمعلمين " ، إذ جاءت الدراسة نتيجة انتشار برامج تعلم الطلبة مهارة أخذ الاختبار. وجرى إعداد اختبار مؤلف من خمسة فقرات قام بكتابتها مجموعة من الخبراء عالجوا بها خمس معضلات اكتشفوها نتيجة مراجعة اختبارات الاختيار من متعدد التي يقوم المعلمون بصياغتها، وقد صيغت الفقرات الخمس على نص أدبي في مستوى الصف السابع، ولا يستطيع الطالب الإجابة عن هذه الأسئلة دون قراءة النص الأدبي، أو استغلال المؤشرات الثانوية الموجودة في كل فقرة من الفقرات الخمس، والدالة على الإجابة الصحيحة. جرى تطبيق الفقرات على عينة مؤلفة من(312) طالباً من طلبة الصف السابع في أريزونا، ثمّ عملت مقابلة مع هؤلاء الطلبة ليقرروا ما الاستراتيجيات التي تم اتباعها في الإجابة عن الفقرات الخمس.

ومن أهم نتائج الدراسة أنّ نسبة الطلبة الذين اختاروا البديل المتوسط كانت(69.23%)، وفي مقابلة مع الطلبة أكدوا أن البديل(ج) غالبا ما يكون صحيحاً، وهذا ما تعودوه من مدرسيهم، كما وجد أنّ(86%) من الطلبة قد اختاروا البديل الأطول بين بدائل الفقرة، لذا جعل الاختيار للبديل الأطول أحد عناصر حنكة الطالب في الإجابة، حتى يتسنى على الطالب اختيار البديل الأطول، عندما لا يعرف الإجابة عن الفقرة.

وقام كريهان وهالادينا (Crehan & Haladyna, 1991, PP. 183 - 192) في جامعة نيفادا وجامعة أريزونا في الولايات المتحدة الأمريكية، بدراسة عنوانها "صدق قاعدتين لقواعد كتابة الفقرات"، إذ هدفت إلى معرفة أثر إرشادين من إرشادات صياغة فقرة

الاختيار من متعدد في الخصائص السيكومترية للفقرة، والإرشادان هما: استخدام البديل "لا شيء مما ذكر"، مقابل التحرر منه، إذ صيغت أرومة الفقرة على شكل سؤال، أو جملة غير مكتملة، واستخدم اختبار مكون من (32) فقرة من نوع الاختيار من متعدد تم اختيارها من دليل علم النفس التربوي.

وجرى تطبيق أداة الدراسة على عينة مؤلفة من (228) طالباً من طلبة مساق علم النفس(115 طالباً للنموذج الاول، و113 طالباً للنموذج الثاني) في كلا الجامعتين. وأشارت أهم نتائج الدراسة إلى عدم وجود فروق ذات دلالة إحصائية من حيث الصعوبة والتمييز تعزى إلى بنية الفقرة، سواء أكان المتن على شكل سؤال أم جملة غير مكتملة، وأظهرت النتائج فروقاً غير دالة إحصائياً في صعوبة الفقرات المتضمنة للبديل "لا شيء مما ذكر" على أنه بديل صحيح، مقابل الفقرات التي لم تتضمن هذا البديل، ولم تكن هناك فروق دالة إحصائياً في متوسطات معاملات التمييز.

مشكلة البحث:

اتضح من العرض السابق أن الدراسات التي تناولت أثر مخالفة قواعد صياغة فقرات الاختيار من متعدد على الخصائص السيكومترية للاختبار وفقراته هي محدودة، واستناداً إلى أنه إذا دخل عنصر غموض في أرومة فقرة الاختيار من متعدد، فإنّ ذلك قد يؤدي إلى أنّ كل مفحوص يمكن أن يفهم السؤال بطريقة مختلفة عن الآخر، ونظراً إلى أنّ الأسئلة الواردة في الكتب المدرسية المقررة للطلبة، لا تخلو من بعض الثغرات في صياغتها، فمن المتوقع ألا تخلو أسئلة المعلمين الصفية من مثل هذه الثغرات، والتي قد تؤثر في الخصائص السيكومترية للاختبار (الثبات، والصدق)، وفقراته (الصعوبة، والتمييز).

أهمية البحث:

على الرغم من الألفة المتزايدة بالنظرية الكلاسيكية في القياس، إلا أنّ المعرفة بمدى التأثيرات التي يمكن أن تنشأ من مخالفة قواعد صياغة فقرات الاختيار من متعدد ما تزال موضع تساؤل، كما أنّ الدراسات التي أجريت في هذا المجال تناولت عدداً قليلاً من الأنماط للبنى المخالفة لقواعد صياغة فقرة الاختيار من متعدد، أما هذه الدراسة فقد تناولت ست مخالفات شائعة. وما يبرر أهمية هذه الدراسة هو أهميتها النظرية والعملية على حد سواء، فمن الناحية النظرية تبين فيما إذا كان هناك أثر في مخالفة قواعد صياغة فقرات الاختيار من متعدد على الخصائص السيكومترية للاختبار وفقراته، أما أهميتها العملية فتكمن في دراستها التطبيقية لأوضاع ممارسة في الميدان التربوي، ومحاولتها التوصل إلى تبريرات عملية تقدم إلى الباحثين، تثري المجال التربوي. ومما يزيد في أهمية الدراسة أنها اعتمدت بيانات حقيقية، تم الحصول عليها من مواقف تعليمية واقعية. وتبرز أهمية الدراسة في زيادة عدد الفقرات التجريبية وطريقة تطبيقها، إذ كان عدد الفقرات (40) فقرة، طبقت في موقف صفي عادي، دون أن يظهر للتجريب أي أثر في أثناء التطبيق، والذي 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 في أثناء الاختبار، فإجراءات هذه الدراسة في مواقف تعليمية واقعية يسهل إعادتها في تخصصات ومستويات دراسية مختلفة، مما يمكن الوقوف على صحة ما يتوصل إليه من نتائج وإمكانية تعميمها.

أهداف البحث:

هدفت الدراسة إلى البحث في أثر مخالفة قواعد صياغة فقرات الاختيار من متعدد في الخصائص السيكومترية للفقرات (الصعوبة، والتمييز)، والخصائص السيكومترية للاختبار (الثبات، والصدق).

أسئلة البحث: حاولت الدراسة الإجابة عن الأسئلة التالية:

1. هل تختلف قيم معاملات صعوبة فقرات الاختيار من متعدد التي تتضمن المخالفات في قواعد الصياغة اختلافاً دالاً (α=0.05) عن قيم معاملات صعوبة الفقرات التي لا تتضمن المخالفات؟
2. هل تختلف قيم معاملات تمييز فقرات الاختيار من متعدد التي تتضمن المخالفات في قواعد الصياغة اختلافاً دالاً (α=0.05) عن قيم معاملات تمييز الفقرات التي لا تتضمن المخالفات؟
3. هل تختلف قيمة معامل ثبات الاتساق الداخلي للاختبار الذي تتضمن فقراته المخالفات في قواعد الصياغة اختلافاً دالاً (α=0.05) عن قيمة معامل ثبات نموذج الاختبار غير المخالف لقواعد الصياغة؟
4. هل تختلف قيمة معامل صدق المحك التلازمي للاختبار الذي تتضمن فقراته المخالفات في قواعد الصياغة اختلافاً دالاً (α=0.05) عن قيمة معامل الصدق لنموذج الاختبار غير المخالف لقواعد الصياغة؟

حدود البحث:

يقتصر البحث الحالي على وحدة " تحليل المقادير الجبرية " من منهاج الرياضيات المقرر للصف التاسع الأساسي.

الطريقة والإجراءات.

مجتمع البحث:

تكون مجتمع الدراسة من جميع طلبة الصف التاسع الأساسي الذكور، المنتظمين في المدارس الحكومية التابعة لمديرية التربية والتعليم لمنطقة إربد الأولى، والبالغ عددهم (3826) طالباً، حسب إحصائيات قسم التخطيط للعام الدراسي 2008/2007م.

عينة البحث:

تكونت عينة الدراسة من (300) طالب، إذ اختيرت (8) مدارس عشوائياً، وكان هدف الباحثين من اختيار هذه المدارس؛ هو تمثيلها لمستويات مختلفة من قدرات الطلبة، وتشكل عينة الدراسة ما نسبته (8%) من مجتمع الدراسة.

أداة البحث:

لتحقيق أهداف الدراسة تمّ إعداد اختبار تحصيلي في مادة الرياضيات، لطلبة الصف التاسع الأساسي في وحدة " تحليل المقادير الجبرية "، وقد اقتضى بناء أسئلة الاختبار ووضعه في صيغته النهائية، الاسترشاد بالأسس المتبعة في تصميم اختبارات التحصيل الصفية (Gronlund & Linn, 1990,PP.166 - 189)، وفيما يلي الإجراءات التي اتبعها الباحثان.

بعد تحديد الغرض من الاختبار، تم تحديد الموضوعات المراد قياسها، ومن ثمّ تحليل المحتوى للوحدة المعنية، إلى مفاهيم، ومصطلحات، ورموز، وتعميمات، ومهارات، وتطبيقات، ثم صياغة النتاجات التعليمية للوحدة الدراسية سابقة الذكر، ثم إعداد جدول المواصفات، وبعد ذلك تمّ صياغة (40) فقرة من نوع الاختيار من متعدد بأربعة بدائل، لقياس النتاجات المقصودة، وللتأكد من صدق محتوى الاختبار، قام الباحثان 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 المحدد.

وقد طبقت صورة الاختبار الأولية، على عينة استطلاعية مكونة من(100) طالبٍ من غير مدارس عينة الدراسة، وأعطي الطلبة الوقت الكافي للإجابة عن فقرات الاختبار بصورته الأولية، والذي قدر وقته بساعة ونصف. 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 (0.30 – 0.68) بوسط حسابي مقداره (0.49)، بينما تراوحت قيم معاملات التمييز بين (0.20 – 0.65)، بوسط حسابي مقداره (0.45)، وعند النظر بالإحصائيات المقترحة من قبل (Eble,1972,PP. 187 - 210)، والتي تتلخص بإمكانية قبول الفقرات التي تزيد معاملات تمييزها عن (0.19)، وقبول الفقرات التي معاملات صعوبتها بين (0.30 – 0.80)، قام الباحثان بقبول جميع فقرات الاختبار. وقد تمّ التحقق من صدق أداة الدراسة من خلال الصدق المرتبط بمحك، عن طريق حساب معامل ارتباط بيرسون بين علامات الطلبة في الاختبار المعّد وعلاماتهم المدرسية، إذ بلغت قيمته (0.90)، وحسب معامل ثبات الاختبار بمعادلة (KR – 20) فكانت قيمته (0.920). وبعد الخروج بالصورة النهائية للاختبار، تمّ كتابة هذه الفقرات في نموذج آخر مخالف لقواعد صياغة فقرات الاختيار من متعدد، ويحتوي على العدد نفسه من البدائل، ولم يتم تغيير موقع الإجابة الصحيحة، وقد ضُمنتْ المخالفات الأتية في نموذج الاختبار: استخدام بدائل غير منطقية، والكلمات المفتاحية المتكررة في كل من متن الفقرة والبديل الصحيح، وجعل الإجابة الصحيحة هي الأكثر تفصيلاً وتوضيحاً، وغالباً ما تكون الأطول بين باقي البدائل، واستخدام الكلمات الدالة على الإطلاق مثل: دائماً، باستمرار، لا شيء، فقط، ونحو ذلك في بعض بدائل الفقرات مما يوحي بخطأ هذه البدائل، وأن يكون أحد البدائل متضمناً في بديل آخر، أو أن يكون مشابهاً له، مع الاختلاف في الصورة المستخدمة للتدليل عليه، ووضع في الأرومة معلومات كثيرة لا ضرورة لها، فتصبح طويلة دون فائدة.

الإجراءات:

بعد إعداد نموذجي الاختبار بصورة نهائية، وفي موعد الامتحانات المدرسية، جرى تطبيق نموذجي الاختبار على أفراد عينة الدراسة وبشكل جماعي داخل غرفة الصف، وقد تمّ توزيع نموذجي الاختبار، ليأخذ الطالب الأول في الجلوس نموذج الاختبار

الأول، والطالب الثاني يأخذ النموذج الثاني وهكذا حتى وزع نموذجي الاختبار على جميع الطلبة، وأعطي الطلبة الوقت الكافي للإجابة عن جميع أسئلة الاختبار، وكان عدد الطلبة الذين أجابوا عن كل نموذج (150) طالباً. 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، وبعد أن تمت إجراءات تصحيح الأوراق، أدخلت البيانات في ذاكرة الحاسوب باستخدام برنامج (SPSS) لتحليلها، والحصول على الإحصائيات التي ساعدت الباحثين في الإجابة عن أسئلة الدراسة.

المعالجات الإحصائية:

حسب معامل الصعوبة لكل فقرة من هذه الفقرات في نموذجي الاختبار، والذي هو نسبة الذين أجابوا عن الفقرة إجابة صحيحة ممن حاولوا الإجابة عنه (معامل السهولة =1– معامل الصعوبة)، وحسب معامل التمييز لكل فقرة من هذه الفقرات في نموذجي الاختبار، والذي هو معامل الارتباط بين الأداء عن الفقرة والأداء عن جميع الفقرات التجريبية في النموذج الواحد (r_{it}). واستخدم اختبار (Independent Sample t – test) لفحص فرضية تساوي أوساط معاملات الصعوبة، وكذلك لفحص فرضية تساوي أوساط معاملات التمييز، واستخدام اختبار ولكوكسون المعروف باسم (Wilcoxon Matched Pairs Signed Ranks Test)؛ وسبب استخدام هذا الاختبار، وذلك لفحص مدى تطابق توزيع قيم معاملات الصعوبة، وقيم معاملات التمييز لفقرات نموذجي الاختبار. وحسب معامل ثبات الاتساق الداخلي لكل نموذج باستخدام معادلة كودر – ريتشاردسون(KR 20)، وقورنت القيم بالإحصائي (F) (Feldet,1969, PP. 363-373) لفحص فرضية تساوي معاملي ثبات مستقلين، وحسب معامل الصدق والذي هو معامل ارتباط بيرسون بين مجموع درجات

المفحوصين عن الفقرات التجريبية في كل نموذج، ودرجاتهم المدرسية، واختبر الفرق بين قيم معاملي الصدق باستخدام الإحصائي (Fisher – Z).

نتائج البحث:

بينت النتائج(ملحق 1) أنّ قيم معاملات الصعوبة لفقرات نموذج الاختبار غير المخالف لقواعد الصياغة، قد تراوحت بين(0.30–0.77) بوسط حسابي مقداره (0.43)، أي أنّ معامل السهولة لنموذج الاختبار يساوي(0.57)، على اعتبار أنّ (معامل السهولة = 1 – معامل الصعوبة)، بينما تراوحت قيم معاملات الصعوبة لفقرات نموذج الاختبار المخالف لقواعد الصياغة بين(0.35 – 0.82)، بوسط حسابي مقداره(0.57)، أي أنّ معامل السهولة لنموذج الاختبار يساوي(0.43)، بمعنى أنّ فقرات هذا النموذج أسهل مقارنةً بفقرات نموذج الاختبار الآخر، فالقيمة العالية للوسط الحسابي لمعاملات الصعوبة تعني أنّ فقرات النموذج أسهل. وللإجابة عن سؤال الدراسة الأول استخدم اختبار (t–test)؛ لاختبار الفرق بين وسطي معاملات صعوبة الفقرات في نموذجي الاختبار، ويبين الجدول (1) نتائج التحليل:

جدول 1

نتائج اختبار (t) لاختبار الفرق بين وسطي معاملات صعوبة الفقرات في نموذجي الاختبار.

نموذج الاختبار	المتوسط الحسابي	الانحراف المعياري	قيمة (t) المحسوبة	درجات الحرية	قيمة الاحتمال
1	0.43	0.12	5.54	78	0.001
2	0.57	0.12			

ملاحظة: (1) يعني: نموذج الاختبار غير المخالف لقواعد صياغة فقرات الاختيار من متعدد.

(2) يعني: نموذج الاختبار المخالف لقواعد صياغة فقرات الاختيار من متعدد.

ويلاحظ من نتائج التحليل وجود فرق ذي دلالة إحصائية (α=0.05) بين قيم معاملات صعوبة فقرات الاختيار من متعدد في نموذجي الاختبار. وتدل هذه النتيجة أن معاملات الصعوبة لفقرات نموذج الاختبار غير المخالف لقواعد الصياغة أصعب، إذ بلغت قيمة (t) (5.54)، باحتمال (0.001). واستخدم اختبار ولكوكسون المعروف باسم (Wilcoxon Matched Pairs Signed Ranks Test)، وسبب استخدام هذا الاختبار؛ وذلك لفحص فرضية تطابق توزيع قيم معاملات الصعوبة لفقرات نموذجي الاختبار، ويبين الجدول (2) نتائج التحليل:

جدول 2

نتائج اختبار ولكوكسون Wilcoxon لمقارنة قيم معاملات الصعوبة لفقرات نموذجي الاختبار.

نموذج الاختبار	الوسط للرتب	عدد الرتب السالبة	عدد الرتب الموجبة	عدد الرتب المتشابهة	قيمة الإحصائي (Z)	احتمال (Z)
1	15.10	5	35	0	4.50	0.001
2	21.27					

ملاحظة: (1) يعني: نموذج الاختبار غير المخالف لقواعد صياغة فقرات الاختيار من متعدد.

(2) يعني: نموذج الاختبار المخالف لقواعد صياغة فقرات الاختيار من متعدد.

ويلاحظ من نتائج التحليل وجود فرق ذي دلالة إحصائية (α=0.05) بين قيم معاملات صعوبة فقرات الاختيار من متعدد في نموذجي الاختبار؛ أي أنّ توزيع قيم معاملات صعوبة فقرات النموذجين مختلفة، إذ بلغت قيمة (Z) (4.50) باحتمال (0.001).

وفيما يتعلق بمعاملات التمييز لفقرات نموذجي الاختبار، وذلك بإيجاد معامل الارتباط بين نتائج المفحوصين على كل فقرة من فقرات الاختبار، ونتائجهم على الاختبار

الكلي (Corrected item total correlation, r_{it})، فقد بينت النتائج (ملحق 1) أنّ قيم معاملات التمييز لفقرات نموذج الاختبار غير المخالف لقواعد الصياغة قد تراوحت بين (0.20 – 0.65) بوسط حسابي مقداره (0.45)، بينما تراوحت قيم معاملات التمييز لفقرات نموذج الاختبار المخالف لقواعد الصياغة بين (0.18 – 0.58)، بوسط حسابي مقداره (0.36). وللإجابة عن سؤال الدراسة الثاني استخدم اختبار (t – test)؛ لاختبار الفرق بين وسطي معاملات تمييز الفقرات في نموذجي الاختبار، ويبين الجدول (3) نتائج التحليل:

جدول 3

نتائج اختبار (t) لاختبار الفرق بين وسطي معاملات تمييز الفقرات في نموذجي الاختبار.

نموذج الاختبار	المتوسط الحسابي	الانحراف المعياري	قيمة (t) المحسوبة	درجات الحرية	قيمة الاحتمال
1	0.45	0.108	2.76	78	0.007
2	0.36	0.111			

ملاحظة: (1) يعني: نموذج الاختبار غير المخالف لقواعد صياغة فقرات الاختيار من متعدد.

(2) يعني: نموذج الاختبار المخالف لقواعد صياغة فقرات الاختيار من متعدد.

ويلاحظ من نتائج التحليل وجود فرق ذي دلالة إحصائية (α=0.05) بين قيم معاملات تمييز فقرات الاختيار من متعدد في نموذجي الاختبار، وتدل هذه النتيجة أنّ معاملات التمييز لفقرات نموذج الاختبار غير المخالف لقواعد الصياغة أعلى تمييزاً، إذ بلغت قيمة (t) (2.76) باحتمال (0.007). واستخدم اختبار ولكوكسون المعروف باسم (Wilcoxon Matched Pairs Signed Ranks Test)، وسبب استخدام هذا الاختبار؛

وذلك لفحص فرضية تطابق توزيع قيم معاملات التمييز لفقرات نمـــوذجي الاختبـــار، ويبين الجدول (4) نتائج التحليل:

جدول 4

نتائج اختبار ولكوكسون Wilcoxon لمقارنة قيم معاملات التمييز لفقرات نموذجي الاختبار.

نموذج الاختبار	الوسط للرتب	عدد الرتب السالبة	عدد الرتب الموجبة	عدد الرتب المتشابهة	قيمة الإحصائي (Z)	احتمال (Z)
1	22.77	26	13	1	2.821	0.005
2	14.46					

ملاحظة: (1) يعني: نموذج الاختبار غير المخالف لقواعد صياغة فقرات الاختيـــار من متعدد.

(2) يعني: نموذج الاختبار المخالف لقواعد صياغة فقرات الاختيار من متعدد.

ويلاحظ من نتائج التحليل وجود فرق ذي دلالة إحصائية (α=0.05) بين قيم معاملات تمييز فقرات الاختيار من متعدد في نموذجي الاختبار؛ أي أنّ توزيع قـــيم معـــاملات تمييز فقرات النموذجين مختلفة، إذ بلغت قيمة (Z) (2.821) باحتمال (0.005).

وفيما يتعلق بمعاملي ثبات الاتساق الداخلي لنموذجي الاختبار، فقد كانت قيمة معامـــل الثبات لنموذج الاختبار غير المخالف لقواعد الصياغة (0.91)، بينما لنموذج الاختبـــار الآخر (0.88). وللإجابة عن سؤال الدراسة الثالث، استخدم الإحصائي (F) (Feldet, 1969,PP.363-373)، للكشف عن الفرق بين معاملي الثبات لعينتين مستقلتين بدرجات حرية (N – 1 ، N – 1)، حيث N : عدد الفقرات، حسب القانون الآتي:

$$F = \frac{1 - \propto_1}{1 - \propto_2}$$

حيث:

$\propto_1$: التقدير العيني لمعامل ثبات النموذج الأول.

$\propto_2$: التقدير العيني لمعامل ثبات النموذج الثاني.

ويبين الجدول(5) نتائج التحليل:

جدول 5

نتائج الإحصائي (F) لفحص مدى الاختلاف بين معاملي ثبات الاتساق الداخلي في نموذجي اختبار.

نموذج الاختبار	(KR – 20)	قيمة (F) المحسوبة	قيمة (F) الحرجة
1	0.91	1.33	1.29
2	0.88		

ملاحظة: (1) يعني: نموذج الاختبار غير المخالف لقواعد صياغة فقرات الاختيار من متعدد.

(2) يعني: نموذج الاختبار المخالف لقواعد صياغة فقرات الاختيار من متعدد.

ويلاحظ من نتائج التحليل وجود فرق ذي دلالة إحصائية بين معاملي ثبات الاتساق الداخلي لنموذج الاختبار، إذ بلغت قيمة (F) (1.33) وهي دالة عند مستوى (α=0.05). أما معاملي الصدق لنموذجي الاختبار، فقد كانت قيمة معامل الصدق لنموذج الاختبار غير المخالف لقواعد الصياغة (0.92)، بينما لنموذج الاختبار الآخر (0.76). وللإجابة عن سؤال الدراسة الرابع استخدم الإحصائي (Z)، لفحص مدى

اختلاف العلاقة الارتباطية لمجموعتين مستقلتين، بعد تحويل قيم معـــاملات الارتبـــاط إلى قيم فيشر الزائية المناظرة لها، ويبين الجدول (6) نتائج التحليل:

جدول 6

نتائج المقارنات الثنائية بين معاملي صدق المحك التلازمي لنموذجي الاختبار.

نموذج الاختبار	معامل الصدق	Zr**	قيمة (Z) المحسوبة	قيمة(Z) الحرجة
1	0.92	1.589	5.08	1.96
2	0.76	0.996		

ملاحظة: (1) يعني: نموذج الاختبار غير المخالف لقواعد صياغة فقـــرات الاختيـــار من متعدد.

(2) يعني: نموذج الاختبار المخالف لقواعد صياغة فقرات الاختيار من متعدد.

(Zr**) يعني: العلامة الفشرية المقابلة لمعامل صدق المحك التلازمي.

ويلاحظ من نتائج التحليل وجود فرق ذي دلالة إحصائية بين معاملي صـــدق المحـــك التلازمي لنموذجي الاختبار، إذ بلغت قيمة(Z) (5.08) وهـــي دالـــة عنـــد مـــستوى (α=0.05)، إذ كان معامل الصدق لنموذج الاختبار غير المخالف أعلى وبـــشكل دال إحصائياً من معامل صدق الاختبار المخالف لقواعد صياغة الفقرات.

مناقشة النتائج:

أوضحت النتائج الخاصة بمقارنة قيم معاملات الـــصعوبة، وجـــود فـــرق ذي دلالـــة إحصائية بين قيم معاملات صعوبة فقرات نموذجي الاختبـــار، إذ كانـــت معـــاملات الصعوبة لفقرات نموذج الاختبار غير المخالف لقواعد الـــصياغة أصـــعب، مقارنـــةً بفقرات نموذج الاختبار المخالف لقواعد الصياغة؛ بمعنى أنّ فقرات النمـــوذج الـــذي تتضمن بدائل فقراته المخالفات في قواعد الصياغة أسهل مقارنـــةً بفقـــرات نمـــوذج

الاختبار الآخر، ويمكن تفسير ذلك إلى أنّ معرفة الطالب بجوانب الضعف المحتملة في صياغة فقرات الاختيار من متعدد، تسهم بصورة فاعلة في جعل فقرات النموذج المخالف أقل صعوبة(أسهل) على اعتبار أنّ معامل السهولة = 1 – معامل الصعوبة، الأمر الذي يقلل من فاعلية فقرة الاختبار، ومن ثمَّ تجعل مهمة إجابة الفقرة أقل صعوبة مقارنةً بنموذج الاختبار غير المخالف لقواعد الصياغة، والمصاغة فقراته بشكل محكم وفق قواعد صياغة فقرات الاختيار من متعدد. وتجدر الإشارة إلى أنّ الاختبار أو الفقرة التي تستطيع أن تظهر الفروق الفردية بين المختبرين هي التي تستطيع أن تقسم الطلاب إلى قسمين(50%) ناجح، و(50 %) راسب، أي يكون معامل السهولة يساوي معامل الصعوبة (معامل السهولة + معامل الصعوبة = 1).

وتتفق هذه النتيجة مع نتائج دراسة(Weiten, 1984)، التي أظهرت أنّ تضمين المخالفات يقود إلى فقرات أقل صعوبة وبدرجة دالة إحصائياً، مقارنةً بالفقرات المتحررة من هذه المخالفات، كما أشارت دراسة كل من(Chase, 1964؛ Carter, 1986) إلى وجود تحيز نحو البديل الأطول، مما أدى إلى اختلاف احتمالات الإجابة، كما أنّ حنكة الطالب في الإجابة قد تلعب دوراً في ذلك، إذ تعدّ صفة مميزة بوجود المخالفات في قواعد الصياغة وهذا ما أشارت له دراسة(Evans, 1984)، بينما تناقضت مع نتائج دراسة كل من (Mueller & Schrock, 1982؛ Crehan & Haladyna, 1991) التي لم تظهر أي تأثير لهذه المخالفات في بنية الفقرة. ويمكن القول إنّ إدخال المخالفات في فقرات الاختيار من متعدد، تعطي مؤشراً يستغله بعض الطلبة ممن يتمتعون بحكمة وبراعة في الإجابة في التعرف على الإجابة الصحيحة، من خلال هذه المؤشرات أو الثغرات في بناء الفقرات، بالرغم من جهل البعض منهم بمحتوى الفقرة، أو افتقارهم للمعرفة اللازمة للإجابة عنها، مما يرفع نسب احتمال الإجابة عنها مقارنةً بنظيرتها محكمة البناء.

أما النتائج المتعلقة بمقارنة قيم معاملات التمييز لفقرات نموذجي الاختبار، فقد أظهرت النتائج وجود فرق ذي دلالة إحصائية، إذ كانت معاملات التمييز لفقرات نموذج الاختبار غير المخالف لقواعد الصياغة أعلى مقارنةً بقيم معاملات التمييز لفقرات نموذج الاختبار المخالف لقواعد الصياغة. إنّ هذه النتيجة التي أظهرت أنّ أعلى الفقرات تمييزاً كانت الفقرات التي لا تتضمن مخالفات في قواعد الصياغة، يعود إلى اختلاف توزيع قيم معاملات التمييز لفقرات نموذجي الاختبار، فمدى قيم معاملات التمييز لفقرات نموذج الاختبار غير المخالف لقواعد الصياغة تراوحت بين (0.20–0.65)، بينما كان المدى لقيم معاملات التمييز لفقرات نموذج الاختبار المخالف لقواعد الصياغة بين (0.18–0.58)، وتعارضت نتيجة هذه الدراسة مع نتائج دراسات أخرى كدراسات (Crehan & Haladyna , 1991 ; Weiten, 1984 ; Mueller & Schrock, 1982;).

وفيما يتعلق بمقارنة قيم معاملي ثبات الاتساق الداخلي، فقد كان أعلى قيمة لنموذج الاختبار غير المخالف لقواعد الصياغة، وكان الفرق بين معاملي الثبات ذو دلالة إحصائية، وجاءت هذه النتيجة انعكاساً للفرق الدال إحصائياً بين قيم معاملات تمييز فقرات نموذجي الاختبار، فالفقرات الأعلى تمييزاً تسهم أكثر من الفقرات الأقل تمييزاً في زيادة التباين الحقيقي للاختبار، وهذا بدوره يرفع من قيمة معامل ثبات الاتساق الداخلي للاختبار الذي صيغة فقراته وفق قواعد صياغة فقرات الاختيار من متعدد (محكم البناء) مقارنةً بنموذج الاختبار المخالف لقواعد الصياغة. ويمكن تفسير ارتفاع قيمة معامل ثبات نموذج الاختبار المخالف لقواعد الصياغة، على أساس أنّ تأثير المخالفة في بنية الفقرة كان على شكل خطأ ثابت(منتظم)، أدى إلى تضخيم معامل الثبات أكثر من قيمته الحقيقية، على أساس أنّ تباين الخطأ الثابت يعمل وكأنه جزء من التباين الحقيقي، وتعارضت نتيجة الدراسة مع نتائج بعض الدراسات مثل دراسات (Weiten, 1984; Mueller & Schrock, 1982).

وقد كان الفرق دالاً إحصائياً بين قيم معاملي صدق المحك التلازمي لنموذجي الاختبار، إذ كان أعلى قيمة معامل صدق لنموذج الاختبار الذي لم تتضمن فقراته أية مخالفات في قواعد الصياغة؛ وهذا يعود لقوة العلاقة الارتباطية بين علامات الطلبة الخام في الاختبار غير المخالف لقواعد الصياغة وعلاماتهم المدرسية في وحدة الرياضيات المعنية، مقارنةً بنموذج الاختبار المخالف لقواعد الصياغة. ويمكن القول أنّ الاختبار غير المخالف لقواعد الصياغة كان أكثر تشبعاً بعامل يقيسه المحك مقارنةً بالاختبار المخالف لقواعد الصياغة، ومن ثمَّ كان معامل الصدق بدلالة المحك للاختبار غير المخالف لقواعد الصياغة أكبر وبفارق دال إحصائياً من نظيره الاختبار المخالف، وقد تعارضت نتيجة الدراسة مع نتائج بعض الدراسات مثل دراسات (Mueller & Schrock, 1982 ; Weiten, 1984).

إنّ النتائج السابقة التي أظهرت وجود فرق ذي دلالة إحصائية بين قيم معاملات صعوبة فقرات نموذجي الاختبار، ووجود فرق ذي دلالة إحصائية بين قيم معاملات التمييز التي لم تحتوِ فقراتها مخالفات في قواعد الصياغة، وتمييز الفقرات التي احتوت المخالفات في قواعد الصياغة، وكذلك التمييز الأعلى لفقرات نموذج الاختبار غير المخالف لقواعد الصياغة، مقارنةً بنموذج الاختبار المخالف لقواعد الصياغة، وارتفاع معامل ثبات الاتساق الداخلي، ومعامل صدق المحك التلازمي لنموذج الاختبار غير المخالف لقواعد الصياغة بصورة دالة إحصائياً، مقارنةً بنموذج الاختبار المخالف لقواعد الصياغة، فكل هذه النتائج تعزز ضرورة التقيد بقواعد صياغة فقرات الاختيار من متعدد أثناء كتابة فقرات الاختبار، فمثلاً اختيار البدائل الفعالة والمتجانسة للفقرات هي من القضايا الهامة في إعداد فقرات اختبارات الاختيار من متعدد، وهي ليست مهمة سهلة على واضع الاختبار.

إضافةً إلى ما تقدم تبرز أهمية تطبيقية لضرورة التقيد بقواعد صياغة فقرات الاختيار من متعدد، كشفت عنها نتائج هذه الدراسة نابعة من تعدد أغراض الاختبارات

التحصيلية، فمن هذه الاختبارات ما يستخدم لأغراض القبول، ومنها ما يستخدم لأغراض تحديد المسار، ومنها ما يستخدم لتشخيص مناطق الضعف لدى الطلاب، فهذا التنوع في أغراض الاختبارات يتطلب اختبارات محكمة البناء، تتفق مع قواعد صياغة فقرات الاختيار من متعدد التي يطرحها المختصون في كتب القياس النفسي والتربوي.

وفي ضوء نتائج الدراسة يوصي الباحثان بضرورة التقيد بقواعد صياغة فقرات الاختيار من متعدد ما أمكن، وينصح بإجراء المزيد من الدراسات والبحوث الأمبريقية في هذا المجال، لتحديد أي المخالفات أكثر تأثيراً في الخصائص السيكومترية للاختبار وفقراته من منظور النظرية الكلاسيكية في القياس.

المراجع

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

Carter, K.(1986).Test wiseness for teachers and students. *Educational Measurement: Issues and Practice, 5*(4), PP. 20 – 23.

Chase,C.(1964).Relative length of option and response set in multiple choice items. *Educational and Psychological Measurement, 24*(4), PP. 861 – 866.

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

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

Evans, W (1984). Test wiseness: an examination of cue using strategies. *Journal of Experimental Education*, 52 (3), PP.141 – 144.

Feldet, I. s. (1969). A test of the hypothesis that cronbachs Alpha or Kuder – Richardson coefficient twenty is the same for two tests. *Psychometrika, 34*, PP. 363 – 373.

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

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.

Mentzer, T. L. (1982). Response bias in multiple – choice test item bias. *Educational and Psychological Measurement, 42* (1), PP. 437 – 446.

Mueller, D. & Schrock, T. (1982). Effects of violating three multiple - choice item construction principles. *Journal of Educational Research*, 75 (5), PP. 314 – 318.

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.

Roid, G. & Haladyna, T. (1982). *A Teachnology for Test Item Writing. New York: Academic Press,* PP. 49 – 55.

Sarnacki, R. (1979). An examination of testwiseness in the cognitive test domain. *Review of Educational Research,* 2, PP.252– 279.

Weiten, W. (1984). Violation of selected item construction principles in educational measurement. *Journal of Experimental Education,* 52 (3), PP. 78 – 174. .

تاريخ ورود البحث إلى مجلة جامعة دمشق 2008/10/28