أثر اختلاف عدد البدائل والبديل « لا شيء مما ذكر » على الخصائص السيكومترية لفقرات الاختيار من متعدد

إعداد
د. ابراهيم محمد العيسى اليعقوب
قسم علم النفس التربوي، كلية التربية
جامعة اليرموك، الأردن

خلفية الدراسة

تتعدد أشكال فقرات Item Format الاختبارات فمنها ما يتطلب من المفحوص أن يعطي الإجابة بلغته الخاصة كأسئلة المقال، والأسئلة التي تتطلب الإجابة القصيرة، وأسئلة التكميل. أما النوع الثاني من أشكال فقرات الاختبارات فيتطلب من المفحوص أن يختار الإجابة الصحيحة من بين مجموعة من البدائل كفقرات الاختيار من متعدد، وفقرات الصواب والخطأ، وفقرات الصواب والخطأ المتعدد، وفقرات المزاوجة. وتعتبر فقرات الاختيار من متعدد من أكثر أنواع الفقرات استعمالًا في وقتنا الحاضر.

إن جميع فقرات الاختيار من متعدد تتكون من نفس العناصر، وأول هذه العناصر هو متن أو أرومة (Stem) السؤال أو الفقرة والذي يمكن أن يكون على شكل سؤال، أو عبارة تتضمن قضية معينة تحتاج إلى إجابة. والعناصر الأخرى في فقرة الاختيار من متعدد هي الخيارات أو البدائل (Options or Choices or Alternatives) وهي بمثابة إجابات محتملة للسؤال الوارد في المتن، وأحد هذه البدائل هو الإجابة الصحيحة (Correct Response). أما بقية البدائل فتعرف بالمموهات أو المشتتات (Roids & Foils or distractors) (Haldyna, 1982 : 50).

إن أبسط صيغ فقرات الاختيار من متعدد هي فقرات الصواب والخطأ، أي في حالة ما يكون للفقرة بديلان فقط. ومن الصيغ الأخرى لفقرات الاختيار من متعدد أن يكون للفقرة ثلاثة بدائل أو أكثر من ذلك كأربعة أو خمسة ... وأن نسبة التخمين العشوائي على هذا النوع من الفقرات تقل بزيادة عدد بدائلها، فهي حوالي 20 % في حالة وجود خمسة بدائل و 25 % في حالة وجود أربعة بدائل و 33 % في حالة وجود ثلاثة بدائل للفقرة. أما في حالة فقرات الصواب والخطأ فتصل نسبة التخمين العشوائي فيها إلى 50 % تقريبًا.

ومع هذه الميزة لزيادة عدد بدائل فقرات الاختيار من متعدد ـــ وهي تقليل نسبة التخمين العشوائي فإنّه يؤخذ عليها أنها تزيد من الوقت الذي يحتاجه المفحوص للإجابة عن مثل هذه الفقرات. كما وأن زيادة عدد بدائل الفقرات تقلل من إمكانية وضع مموهات أو بدائل جذابة للمفحوصين، علمًا بأن قدرة هذه المموهات على جذب بعض المفحوصين، والتجانس بين هذه المموهات يعتبران من الخصائل الأساسية التي يجب أن تتمتع بها المموهات في هذا النوع من الفقرات. وأنه لا فائدة من وضع عدد كبير من البدائل للفقرات دون تمتعها بمثل هذه الخصائص.

ومنذ دخول اختبارات الاختيار من متعدد ميدان القياس النفسي والتربوي اهتم العاملون في الميدان بعامل التخمين، وأثره على نتائج المفحوصين على الاختبار، وبالتالي على الخصائص السيكومترية لمثل هذا النوع من الاختبارات. أما أثر اختلاف عدد بدائل فقرات هذا النوع من الاختبارات على خصائصها السيكومترية كالثبات والصدق وعلى خصائص فقراتها كالصعوبة والتمييز فقد لقي اهتمامًا أكبر في العقدين الأخيرين من هذا القرن. لذا جاءت هذه الدراسة لتبحث أثر اختلاف عدد البدائل والبديل « لا شيء مما ذكر » على الخصائص السيكومترية لفقرات الاختيار من متعدد.

ومن الدراسات التي تناولت أثر اختلاف عدد بدائل الفقرات على صعوبتها، وتمييزها، ومعامل ثباتها دراسة كوستين (Costin, 1970)، حيث استخدم في هذه الدراسة صورتين لاختبار تتكوّن كل منهما من 110 فقرات تقيس محتوى أربعة موضوعات رئيسة من مساق مقدمة في علم النفس هي : الإدراك، والتعلم، والدافعية، والذكاء. منها 25 فقرة على الإدراك و 30 فقرة على التعلم و 30 فقرة على الدافعية و 25 فقرة على الذكاء. وكان عدد بدائل فقرات الصورة الأولى أربعة بدائل، وعدد بدائل فقرات الصورة الثانية ثلاثة بدائل، تم الحصول عليها بحذف أحد بدائل فقرات الصورة الأولى عشوائيًا. وبعد تطبيق هذه

الاختبارات على 207 طلاب من الطلبة المسجلين في جامعة إلّينوي وينتمون إلى أربع شعب صفية منها شعبتان من الطلبة المسجلين في الفصل الأول، وشعبتان من الطلبة المسجلين في الفصل الثاني. ولأغراض التحليل الإحصائي استبعدت منها إجابات سبعة طلاب عشوائيًا، ثم حسبت معاملات صعوبة هذه الفقرات وتمييزها وكذلك معاملات الاتساق الداخلي لصورتي الاختبار. وبينت نتائج هذه الدراسة أن معاملات صعوبة وتميز الفقرات ذات الثلاثة بدائل أعلى من معاملات صعوبة وتمييز الفقرات ذات الأربعة بدائل. كما وبينت نتائج هذه الدراسة أيضًا أن معامل الاتساق الداخلي للفقرات ذات الثلاثة بدائل أعلى من معامل الاتساق الداخلي للفقرات ذات الأربعة بدائل.

وقام كوستين (Costin, 1972) بدراسة أخرى هدفت مقارنة خصائص اختبار لفقراته ثلاثة بدائل باختبار لفقراته أربعة بدائل، واستخدم في هذه الدراسة 100 فقرة اختيرت عشوائيًا من مجمع من الفقرات تقيس موضوعات في علم النفس هي الذكاء، والشخصية، والتفاعل الاجتماعي، والاضطراب السلوكي. واختيرت هذه الفقرات بما يتناسب وحجم محتوى هذه الموضوعا، ثم اختير من هذه الفقرات 50 فقرة حذف أحد بدائلها عشوائيًا لتشكل الفقرات ذات الثلاثة بدائل، وأما الفقرات الخمسون الباقية فأبقي على بدائلها كما هي. ثم طبق الاختبار بكامله على 1566 طالبًا كاختبار نهائي للمساق. ولأغراض التحليل اعتبر كل نوع من هذه الفقرات اختبارًا منفصلًا. وحُسِب معامل ثبات الاتساق الداخلي لكل منها بمعادلة كودر–ريشاردسون رقم 20 فكان 0،75 للفقرات ذات الثلاثة بدائل و 0،78 للفقرات ذات الأربعة بدائل وتبين أنه لا يوجد فرق دال احصائيًا بين قيم معاملات الثبات هذه. كما حسبت معاملات تمييز هذه الفقرات وهي عبارة عن معاملات ارتباط الفقرات مع الاختبار ككل، ثم حسب الوسط لمعاملات تمييز الفقرات في كلا النموذجين بعد تحوليها إلى درجات فشر الزائية فكانت 0،29 للفقرات ذات الثلاثة بدائل و 0،30 للفقرات ذات الأربعة بدائل. كما حسب وسط عدد الفقرات التي اجيبت من كل نوع فكان الوسط لعدد الفقرات التي اجيبت من الفقرات ذات الثلاثة بدائل 36،8 ومن الفقرات ذات الأربعة بدائل 36،4، وتبيّن أنّه لا يوجد فرق دال احصائيًا بين هذه الأوساط. كما بينت نتائج هذه الدراسة أن لا فرق بين خصائص الفقرات ذات الثلاثة بدائل والفقرات ذات الأربعة بدائل، وأنّه لا يوجد فرق دال احصائيًا بين معاملات ثبات الاتساق الداخلي لكلا النموذجين من الاختبار.

ومن الدراسات في هذا المجال أيضًا دراسة لاندرم وكاشن وثيس (Landrum, Cahin & Theis, 1993) التي كان من أهدافها معرفة مدى التغير في صعوبة الفقرات عندما يتغير عدد بدائلها من أربعة بدائل إلى ثلاثة بدائل. ولتحقيق أهداف هذه الدراسة اختير 46 طالبًا مسجلين في مساق علم النفس العام في الفصل الدراسي الأول خريف عام 1990 في جامعة وسكونسون بلات إيفيل وطبّق عليهم الاختبار النهائي لهذا المساق. وتكوّن هذا الاختبار من 250 فقرة غطّت جميع الوحدات المكونة لهذا المساق وعددها خمس وحدات (50 فقرة لكل وحدة) هي : الأولى المقدمة وطرق البحث والثانية التعلم والذاكرة والمعرفة واللغة، والثالثة الأساس البيولوجي للسلوك والرابعة الشعور والإدراك، والخامسة في علم النفس الاجتماعي. كما طبّق اختبار نهائي لنفس المساق على مجموعة أخرى مكونة من 67 طالبًا من الطلبة المسجلين في نفس المساق في الفصل الدراسي الثاني ربيع عام 1991. ويتكون هذا الاختبار أيضًا من 250 فقرة (50 فقرة لكل وحدة) منها 144 فقرة مشتركة مع الاختبار الأول. ولغايات الدراسة حذف أحد بدائل فقرات هذا الاختبار بناءًا على الخبرة الذاتية لمدرس المساق ليصبح عدد بدائل فقراته ثلاثة بدائل فقط. وتمت مقارنة أداء المفحوصين على الفقرات المشتركة بين الاختبارين. وبينت النتائج أن أداء المفحوصين على الفقرات ذات الثلاثة بدائل في جميع الوحدات عدا الوحدة الرابعة أفضل من أدائهم على الفقرات ذات الأربعة بدائل لنفس الوحدات، وهذا يعني أن الفقرات ذات الثلاثة بدائل كانت أسهل من الفقرات ذات الأربعة بدائل. وعند تعديل صعوبة الفقرات لأثر عامل التخمين وذلك بطرح احتمال اجابة الفقرة بالصدفة من معامل صعوبتها تبين أنه لا يوجد فرق دال احصائيًا بين صعوبة الفقرات ذات الثلاثة بدائل وصعوبة الفقرات ذات الأربعة بدائل بالنسبة إلى فقرات الوحدات الأولى والثانية والثالثة. أما فقرات الوحدتين الرابعة والخامسة وفقرات الاختبار ككل فقد كانت الفروق بينها دالة احصائيًا.

كما قام ستراتون وكاتس (Straton & Catts, 1980) بدراسة هدفت المقارنة بين خصائص اختبارات اختيار من متعدد تحوي عددًا متساويًا من البدائل ولكنها تختلف في عدد بدائل فقراتها. ولتحقيق هدف الدراسة قام الباحثان ببناء ثلاثة نماذج لاختبار في مادة الاقتصاد للصف الثاني عشر. وتضم فقرات كل نموذج منها 120 بديلًا. ويتكون النموذج الأول من 60 فقرة لكل منها بديلان، والنموذج الثاني من 40 فقرة لكل منها ثلاثة بدائل. أما النموذج الثالث فيتكون من 30 فقرة لكل منها أربعة بدائل.

طبقت الاختبارات في ظروف صفية عادية حيث وزع كل نموذج منها عشوائيًا على 65 تلميذًا من تلاميذ الصف الثاني عشر. وبينت نتائج هذه الدراسة أن أسهل الفقرات كانت ذات البديلين وأصعبها ذات الأربعة بدائل، حيث كانت الأوساط لمعاملات الصعوبة هي : (0،75) و (0،67) و (0،59) للفقرات ذات البديلين والثلاثة والأربعة بدائل على التوالي.

أما فيما يتعلق بمعاملات تمييز الفقرات فكان أقلها تميزًا الفقرات ذات البديلين وأعلاها تمييزًا الفقرات ذات الأربعة بدائل. حيث كانت أوساط معاملات التمييز لهذه الفقرات هي : (0،24) و (0،35) و (0،41) على التوالي.

أما معاملات الثبات فكان أدناها لاختبار الفقرات ذات البديلين وأعلاها لاختبار الفقرات ذات الأربعة بدائل وكانت قيم معاملات الثبات هي : (0،47) و (0،66) و (0،68) على التوالي.

أما دراسة وليامز وايبل (Williams & Ebel, 1957) فقد هدفت إلى مقارنة اختبارات فقراتها ذات الأربعة بدائل باختبارات فقراتها ذات الثلاثة بدائل، ولتحقيق أغراض هذه الدراسة طبق اختبار يتكون من 150 فقرة تقيس مفردات لغوية على 700 من طلبة المرحلة الثانوية. وقد تم الحصول على الفقرات ذات الثلاثة بدائل من نفس الفقرات ذات الأربعة بدائل بعد حذف البدائل الأقل تمييزًا من هذه الفقرات. وقد بينت النتائج أن وسط معاملات صعوبة الفقرات ذات الأربعة بدائل هو (0،496)، بينما كان وسط صعوبة الفقرات ذات الثلاثة بدائل (0،684)، أي أن الفقرات ذات الثلاثة بدائل هي أسهل من الفقرات ذات الأربعة بدائل. كما بينت النتائج أن وسط معاملات التمييز للفقرات ذات الأربعة بدائل هو (0،487)، وللفقرات ذات الثلاثة بدائل هو (0،412)، وهذا يعني أن الفقرات ذات الثلاثة بدائل أقل تمييزًا من الفقرات ذات الأربعة بدائل. كما بينت النتائج أنه لا يوجد فرق بين معاملات الاتساق الداخلي للصورتين من الاختبار، حيث كانت قيمة معامل الاتساق الداخلي لنموذج الأربعة بدائل (0،95) بينما كانت لنموذج الثلاثة بدائل (0،93).

ومن الدراسات التي هدفت إلى معرفة العدد الأمثل من البدائل في فقرات اختبارات الاختيار من متعدد، ولمعرفة أثر وجود البديل « لا شيء مما ذكر » على صعوبة الفقرات

وتمييزها دراسة كريهان وهالادينا وبروير (Crehan, Haladyna & Brewer, 1993) ولتحقيق أهداف هذه الدراسة تم اختيار 48 فقرة من دليل المدرّس لكتاب علم النفس لمؤلفه مايرز (Myers, 1986) وأعيدت كتابة هذه الفقرات في أربع صور، الأولى منها لها أربعة بدائل والثانية لها أربعة بدائل أحدها البديل « لا شيء مما ذكر »، والصورة الثالثة لها ثلاثة بدائل، والرابعة لها ثلاثة بدائل أحدها البديل « لا شيء مما ذكر ». وعند تقليل عدد المموّهات كان يحذف أقل هذه المموهات فاعلية أي أقلها تمييزًا. وبعد ذلك وزعت هذه الفقرات على أربعة نماذج بحيث اشتمل كل نموذج منها على 48 فقرة تقيس نفس المحتوى وكتب كل 12 فقرة منها بواحدة من الصور الأربع السابقة.

طبقت هذه النماذج عشوائيًا على شعبتين من الطلبة المسجلين في مساق مقدمة في علم النفس بواقع 55 طالبًا لكل نموذج وقورنت معاملات صعوبة وتمييز الفقرات ذات الثلاثة بدائل مع صعوبة وتمييز الفقرات ذات الأربعة بدائل. كما قورنت صعوبة وتمييز الفقرات التي تحتوي على البديل « لا شيء مما ذكر » مع الفقرات التي لا تحتوي على هذا البديل. وبينت النتائج أن الفقرات ذات الثلاثة بدائل أسهل من الفقرات ذات الأربعة بدائل. وكذلك كانت الفقرات التي تحتوي على البديل « لا شيء مما ذكر » أصعب من الفقرات التي لا تحتوي على مثل هذا البديل. أما بالنسبة إلى معاملات التمييز فلم يكن هناك فرق دال احصائيًا بين الفقرات ذات الثلاثة بدائل والفقرات ذات الأربعة بدائل. ولم يكن هناك فرق دال احصائيًا بين تمييز الفقرات التي تحتوي على البديل « لا شيء مما ذكر » وتلك التي لا تحتوي على مثل هذا البديل.

ومن الدراسات في هذا المجال أيضًا دراسة توليفسون (Tollefson, 1987) التي هدفت إلى مقارنة صعوبة وتمييز فقرات الاختبار واتساقه الداخلي عندما يكون « لا شيء مما ذكر » أحد مموهات الفقرات في صورة منها وكبديل صحيح في صورة أخرى من هذه الفقرات.

ولتحقيق أهداف هذه الدراسة اختير اختبار يتكون من 73 فقرة من نوع الاختيار من متعدد من أربعة بدائل كاختبار نهائي لمساق مقدمة في الإحصاء، كان قد طوّره أربعة من الأساتذة درّسوا هذا المساق في فصول سابقة، أي تم تطبيقه ميدانيًا قبل اجراءات هذه الدراسة عليه. ولغايات هذه الدراسة اختير 12 فقرة من فقرات هذا الاختبار لتقيس

مفاهيم احصائية ولا تحتوي أي نوع من المسائل العددية وكتبت هذه الفقرات بثلاث صور دون أي تغيير في متونها. وكان لفقرات الصورة الأولى أربعة بدائل عادية. أما الصورة الثانية فكان أحد مموهات فقراتها « لا شيء مما ذكر » والذي حل محل أكثر المموهات اختيارًا من قبل المفحوصين في التطبيقات السابقة لهذا الاختبار. والصورة الثالثة لهذه الفقرات كان البديل « لا شيء مما ذكر » هو الجواب الصحيح والذي حل محل الجواب الصحيح لهذه الفقرات في وضعها السابق.

وقد وزعت فقرات كل صورة من صور هذه الفقرات على الاختبار النهائي بشكل عشوائي وبهذا تشكل ثلاث صور للاختبار النهائي. تضمنت الصورة الأولى 12 فقرة لكل منها أربعة بدائل عادية. والصورة الثانية تضمنت (12) فقرة بأربعة بدائل أحدها « لا شيء مما ذكر » كأحد المموهات. والصورة الثالثة تضمنت 12 فقرة بأربعة بدائل أحدها « لا شيء مما ذكر » كجواب صحيح. وبعد ذلك وزعت هذه النماذج من الاختبار عشوائيًا على 81 مفحوصًا في ظروف صفية عادية. أجاب 26 منهم على الصورة الأولى و 28 على الصورة الثانية و 27 على الصورة الثالثة. وبينت النتائج أن أصعب الفقرات كانت تلك التي فيها « لا شيء مما ذكر » كجواب صحيح، وأن الفرق كان دالًا بين صعوبتها وصعوبة الفقرات ذات البدائل العادية. أما الفرق بين صعوبة هذه الفقرات والفقرات ذات البديل « لا شيء مما ذكر » كمموه لم يصل لمستوى الدلالة الإحصائية. كما أن الفرق بين صعوبة الفقرات ذات البديل « لا شيء مما ذكر » كمموه والفقرات ذات البدائل العادية لم يصل لمستوى الدلالة الإحصائية أيضًا.

كما بينت النتائج أن أعلى معاملات تمييز كانت للفقرات ذات البدائل العادية، ولكن الفروق بين معاملات التمييز لم تصل لمستوى الدلالة الإحصائية بين جميع الأشكال المختلفة لهذه الفقرات. أما بالنسبة إلى معاملات الاتساق الداخلي للنماذج المختلفة محسوبة بمعادلة كودر–ريشاردسون 20 فقد كان اعلاها (0،74) للفقرات ذات البدائل الأربعة العادية وأقلها (0،51) للفقرات التي تحوي البديل « لا شيء مما ذكر » كمموه. وأوسطها (0،56) كان للفقرات التي تحوي البديل « لا شيء مما ذكر » كجواب صحيح.

وفي مراجعة لـ 46 مرجعًا تتعلق بكتابة فقرات الاختيار من متعدد وجد هالادينا وداوننج (Haladyna & Downing, 1989) أن 34 (73 %) مرجعًا من هذه المراجع تعرضت لقضية استخدام البديل « لا شيء مما ذكر ». وانقسم المؤلفون ما بين مؤيدين

ومعارضين لاستخدام هذا البديل. حيث أيد استخدامه (19) منهم بينما عارض استخدامه الباقون. كما أن هذه المراجعة بينت أن قواعد استخدام البديل « لا شيء مما ذكر » كأحد الخيارات كان ترتيبها العاشر من بين جميع قواعد كتابة فقرات اختبارات الاختيار من متعدد. وهذا مؤشر على تدني أهمية هذه القاعدة بالنسبة إلى بقية قواعد كتابة الفقرات الأخرى. كما أن التعارض بين المؤيدين والمعارضين لاستخدام هذه القاعدة هو دليل أيضًا على مدى صدق استخدام هذه القاعدة. كما يذكر الباحثان أن عدد الأبحاث الامبريقية المتعلقة بهذه القاعدة هو محدد بعشر دراسات فقط هي دراسات Boyntn, 1950; Dudycha & Carpenter, 1973; Forsyth & Spratt, 1980; Hughes & Trimble, 1965; Mueller, 1975; Oosterhof & Coats, 1984; Rimland, 1960; Schrneiser & Whitney, 1975; Wesman & Bennett, 1946; Williamson & Hopkins, 1967 وأن جميع هذه الدراسات تناولت خاصية صعوبة الفقرات، بينما تناولت خمس دراسات منها فقط تمييز الفقرات ومعاملات الثبات، وتناولت دراستان فقط معامل الصدق. وفي نتائج جميع هذه الدراسات كان البديل « لا شيء مما ذكر » يجعل الفقرات أكثر صعوبة وأقل تمييزًا كما أن وجود هذا البديل يُحسّن من ثبات نتائج الاختبار.

وفي مراجعة تحليلية لصعوبة وتمييز فقرات اختبارات الاختيار من متعدد تحتوي البديل « لا شيء مما ذكر » قامت بها نوليس وويلش (Knowles & Welch, 1992) تبين أن استخدام هذا البديل يعطي فقرات ذات خصائص سيكومترية (صعوبة وتمييز) لا تقل في جودتها عن الفقرات العادية. وقد أشار الباحثان في دراستهما هذه إلى أن نتائج الدراسات التي تناولاها بالتحليل كانت متعارضة فيما يتعلق بصعوبة الفقرات وتمييزها، وأن الجدل لا يزال مستمرًا حول استخدام البديل « لا شيء مما ذكر » في اختبارات الاختيار من متعدد وذلك للاعتقاد بأن هذا البديل يجعل من الفقرات أكثر صعوبة وأقل تمييزًا. ويذكر الباحثان أن الدراسات الامبريقية التي تناولت استخدام البديل « لا شيء مما ذكر » محدودة العدد، منها دراسة ويسمان وبنيت (Wesman & Bennett, 1946) التي استخدم فيها البديل « لا شيء مما ذكر » في اختباري المفردات والحساب لدى المتقدمين لكلية التمريض حيث بينت نتائج هذه الدراسة عدم اختلاف أثر هذا البديل على ثبات الاختبار وصعوبته مقارنًا بالبدائل العادية. بينما أظهرت نتائج دراسة ريملاند (Rimland, 1960) أن البديل « لا شيء مما ذكر » زاد من صعوبة الفقرات. ويذكر الباحثان أن نتائج دراسة تولفسون وتريب (Tollefson & Tripp, 1983) بينت أن الفقرات التي تحوي البديل « لا شيء مما ذكر »

كجواب صحيح يكون تمييزها أقل وبشكل دال احصائيًا من تمييز تلك الفقرات التي لا تحتوي هذا البديل. أما صعوبة الفقرات التي احتوت هذا البديل كواحد من بدائلها فلم تختلف عن صعوبة بقية الفقرات الأخرى. أما دراسة اوسترهوف وكوتس (Oosterhof & Coats, 1984) فلم تظهر نتائجها أثرًا للبديل « لا شيء مما ذكر » على صعوبة الفقرات إلا إذا كان هذا البديل هو الجواب الصحيح.

ويذكر نوليس وويلش في دراستهما التحليلية أن نتائج دراسات وليامسون وهوبكنز (Williamson & Hopkins, 1967) التي أجريت على اختبارات تحصيلية فيها البديل « لا شيء مما ذكر » بينت أن الفرق بين صعوبة الفقرات التي تحتوي هذا البديل وصعوبة الفقرات التي لا تحتوي هذا البديل كانت دالة احصائيًا في دراستين منها ولم تصل هذه الفروق لمستوى الدلالة الإحصائية في دراستين اخريين. أما فيما يتعلق بنتائج دراسة دوديشا وكاربنتر (Dudycha & Carpenter, 1973) فيذكر الباحثان أن صعوبة الفقرات التي تحتوي البديل « لا شيء مما ذكر » تزيد بدلالة احصائية على صعوبة الفقرات التي لا تحتوي على هذا البديل. أما الفرق بين تمييز هذه الفقرات فلم يصل إلى مستوى الدلالة الإحصائية. وكذلك الحال بالنسبة إلى نتائج دراسة فورسيث وسبرات (Forsyth & Spratt, 1980) التي بينت أيضًا أن الفقرات التي تحتوي البديل « لا شيء مما ذكر » كانت أعلى صعوبة وأقل تمييزًا من الفقرات ذات البدائل العادية.

ويذكر الباحثان أيضًا في دراستهما التحليلية أن نتائج دراسة كريهان وهالادينا (Crehan & Haladyna, 1989) أظهرت فروقًا دالة احصائيًا في صعوبة الفقرات عند استخدام البديل « لا شيء مما ذكر » ولا فروق دالة في تمييز هذه الفقرات. أما دراسة فراري (Frary, 1991) التي قارن فيها صعوبة وتمييز الفقرات التي تحتوي البديل « لا شيء مما ذكر » بفقرات لا تحتوي هذا البديل وجد أن الزيادة طفيفة في صعوبة الفقرات التي تحتوي هذا البديل مقارنة بالفقرات التي لا تحتوي هذا البديل. أما تمييز الفقرات فكان متساويًا في كلا النوعين من الفقرات. أما فيما يتعلق بنتائج دراسة نوليس وويلش (Knowles & Welch, 1992) فإنّها تشير إلى وجود فروق دالة احصائيًا بين صعوبة الفقرات وتمييزها عند استخدام البديل « لا شيء مما ذكر » سواء كان هو الجواب الصحيح أو كان أحد البدائل أو واحدًا من الإجابات الصحيحة.

وفيما يتعلق بقواعد كتابة الفقرات يذكر كريهان وهالادينا (Crehan & Haladyna, 1991) أن هالادينا وداوننج في مراجعتهما (96) بحثًا امبريقيًا تتعلق بكتابة الفقرات أن (53) بحثًا منها تناولت قاعدتين فقط من قواعد كتابة الفقرات، الأولى تتعلق بعدد الخيارات المفضلة عند كتابة الفقرات، والثانية تتعلق بتوازن توزيع مفتاح الإجابة على البدائل. ويذكر الباحثان أيضًا أن معظم طرق كتابة الفقرات درست بأقل من (10) مرات، وأن الأسس التي تبنى عليها العديد من قواعد كتابة الفقرات تعتمد في معظمها على ما جاء في الكتب والمؤلفات والمحاضرات ولا تعتمد في معظمها على دراسات وأبحاث امبريقية، مما يدلل على أن صدق طرق كتابة الفقرات لا يزال ضعيفًا من ناحية تجريبية.

هدف الدراسة وأسئلتها

يتضح من العرض السابق تناقض نتائج الدراسات التي تناولت أثر اختلاف عدد بدائل فقرات اختبارات الاختيار من متعدد على مستوى صعوبة هذه الفقرات وتمييزها. كما وأن مستوى صعوبة وتمييز الفقرات التي تحتوي البديل « لا شيء مما ذكر » لا تزال موضع نقاش وأن الجدل لا يزال قائمًا بين المعارضين والمؤيدين لاستخدام هذا البديل. إضافة إلى أن الدراسات الأمبريقية التي تناولت هذه القاعدة من قواعد كتابة الفقرات لا تزال محدودة. لذا جاءت هذه الدراسة لتلقي مزيدًا من الضوء على مدى اختلاف مستوى صعوبة الفقرات وتمييزها لاختلاف عدد بدائلها، وكذلك لمعرفة أثر وجود البديل « لا شيء مما ذكر » على مستوى صعوبة وتمييز هذه الفقرات عندما يكون هذا البديل واحدا من بدائل فقرات الاختبار أو عندما يكون هو البديل الصحيح فيها. وبالتحديد فإن هدف هذه الدراسة هو مقارنة صعوبة وتمييز فقرات ذات ثلاثة بدائل بصعوبة وتمييز فقرات ذات أربعة بدائل. وكذلك مقارنة صعوبة وتمييز فقرات ذات أربعة بدائل بصعوبة وتمييز فقرات ذات أربعة بدائل تحتوي البديل « لا شيء مما ذكر » كبديل في نموذج منها، وكجواب صحيح في نموذج آخر. ثم مقارنة صعوبة وتمييز الفقرات التي تحتوي « لا شيء مما ذكر » كبديل بتلك التي تحتويه ويكون فيها هو الجواب الصحيح.

ولذا تحاول هذه الدراسة الإجابة عن التساؤلات التالية :

1 ــ هل تختلف قيم معاملات صعوبة وتمييز فقرات الاختيار من متعدد ذات الثلاثة بدائل عن قيم معاملات صعوبة وتمييز الفقرات ذات الأربعة بدائل ؟

2 ــ هل تختلف قيم معاملات صعوبة وتمييز فقرات الاختيار من متعدد التي تحتوي البديل « لا شيء مما ذكر » عن قيم معاملات صعوبة وتمييز الفقرات التي لا تحتوي هذا البديل ؟

3 ــ هل تختلف قيم معاملات صعوبة وتمييز فقرات الاختيار من متعدد التي تحتوي البديل « لا شيء مما ذكر » كجواب صحيح عن قيم معاملات وتمييز الفقرات التي لا تحتوي هذا البديل ؟

4 ــ هل تختلف قيم معاملات صعوبة وتمييز فقرات الاختيار من متعدد التي تحتوي البديل «لا شيء مما ذكر » كبديل عن قيم معاملات صعوبة وتمييز الفقرات التي تحتوي هذا البديل كجواب صحيح ؟

الطريقة والإجراءات

أفراد الدراسة

شارك في هذه الدراسة (165) طالبًا وطالبة منهم (97) طالبًا و (68) طالبة، منهم (76) من طلبة التخصص الفرعي في التربية و (89) من طلبة دبلوم التربية الذين كانوا مسجلين في (4) شعب لمساق مقدمة في القياس والتقويم خلال الفصلين الأول والثاني من العام الدراسي 1995/94 في جامعة اليرموك. وهذا المساق من المساقات الإجبارية التي يطرحها قسم علم النفس التربوي في كلية التربية في جامعة اليرموك لطلبة التخصص الفرعي في التربية ولطلبة دبلوم التربية. وهؤلاء الطلبة لا يختلفون عن بقية الطلبة الذين سجلوا للمساق في السنوات السابقة.

أداة الدراسة

تكونت أداة الدراسة من الاختبار النهائي لمساق مقدمة في القياس والتقويم والبالغ عدد فقراته 75 فقرة من نوع الاختيار من متعدد وتغطي المادة الدراسية المقررة للمساق والمكونة من ست وحدات دراسية هي : مبادىء ومفاهيم أساسية في القياس والتقويم، الأهداف التدريسية، اعداد الاختبارات التحصيلية، صدق الاختبار، ثبات الاختبار، تحليل وتفسير نتائج الاختبار. وكانت فقرات هذا الاختبار قد أعدت من قبل عدد من مدرسي المساق

وطبقت في فصول سابقة. وهي من نوع الاختيار من متعدد من أربعة بدائل. ولأغراض هذه الدراسة أعدت أربعة نماذج من هذا الاختبار دون إجراء أي تغيير على متون الفقرات، ففي النموذج الأول ابقيت الفقرات على حالها بأربعة بدائل، أما في النموذج الثاني فجعل للفقرات ثلاثة بدائل فقط، وذلك بحذف البديل الأقل جاذبية أي الأقل تمييزًا من بين بدائل فقرات النموذج الأول. أما في النموذج الثالث والرابع فكان عدد بدائل الفقرات فيهما أربعة. أحدها البديل « لا شيء مما ذكر ». وقد اختيرت الفقرات التي استخدمت قيم معاملات صعوبتها وتمييزها في التحليلات الإحصائية للإجابة عن أسئلة الدراسة من بين فقرات الاختبار بحيث لا يحتاج أي من هذه الفقرات المختارة إلى عمليات حسابية، وكان عدد هذه الفقرات 32 فقرة أي ما يشكّل 42،6 % من فقرات الاختبار. وزعت هذه الفقرات عشوائيًا في كل نموذج من نماذج الاختبار الأربعة، بحيث كان عدد بدائلها في النموذج الأول أربعة وفي النموذج الثاني ثلاثة، أما في النموذج الثالث أربعة أحدها البديل « لا شيء مما ذكر » كواحد من الخيارات الأربعة أما في النموذج الرابع فكان للفقرات المختارة أربعة بدائل أحدها البديل « لا شيء مما ذكر » بحيث يكون هذا البديل هو الجواب الصحيح في جميع هذه الفقرات المختارة، أما في بقية الفقرات من هذا النموذج فكان البديل « لا شيء مما ذكر » كواحد من بدائل هذه الفقرات أي يكون هو البديل الصحيح في بعض هذه الفقرات وغير صحيح في بعضها الآخر. والمثال التالي نموذج لفقرة كتبت في الصور الأربع سابقة الذكر.

ـــ يسمى الانحراف المعياري لتوزيع اخطاء التنبؤ بالعلامات الفعلية على المحك :

أ) الانحراف المعياري للعلامات على المحك.
ب) الخطأ المعياري للقياس.
*ج) الخطأ المعياري للتقدير.
د) الانحراف المعياري للعلامات على المتنبىء.

ـــ يسمى الإنحراف المعيـري لتوزيع أخطاء التنبؤ بالعلامات الفعلية على المحك :

أ) الانحراف المعياري للعلامات على المحك.
ب) الخطأ المعياري للقياس.
*ج) الخطأ المعياري للتقدير.

ـــ يسمى الانحراف المعياري لتوزيع أخطاء التنبؤ بالعلامات الفعلية على المحك :

أ) الانحراف المعياري للعلامات على المحك.

ب) الخطأ المعياري للقياس.

*ج) الخطأ المعياري للتقدير.

د) لا شيء مما ذكر.

ـــ يسمى الانحراف المعياري لتوزيع أخطاء التنبؤ بالعلامات الفعلية على المحك :

أ) الانحراف المعياري للعلامات على المحك.

ب) الخطأ المعياري للقياس.

ج) الانحراف المعياري للعلامات على المتنبىء.

*د) لا شيء مما ذكر.

الإجراءات

بعد إعداد نماذج الاختبار النهائي للفصول الدراسية التي أجريت فيها هذه الدراسة، وفي موعد الامتحانات النهائية لكل فصل دراسي وحسب الجدول المقرر من إدارة الجامعة لهذه الامتحانات وزعت النماذج الأربعة بشكل عشوائي على الطلبة في قاعات التدريس حيث أجاب كل واحد منهم على نموذج من هذه النماذج وأعطي الطلبة الوقت الكافي للإجابة عن جميع أسئلة الاختبار. وقد أشرف الباحث نفسه على جميع اجراءات التطبيق وساعده في ذلك ثلاثة من طلبة الماجستير في قسم علم النفس التربوي، وكان عدد الطلبة الذين أجابوا عن النموذج الأول من الاختبار (42) طالبًا وطالبة، بينما أجاب عن كل من النماذج الثلاثة الباقية (41) طالبًا وطالبة وبعد أن تمت اجراءات التطبيق صححت الأوراق وأعطي صفر للفقرة الخطأ ودرجة واحدة للفقرة الصحيحة وبذا تكون العلامة الكلية للمفحوص هي مجموع الإجابات الصحيحة على فقرات الاختبار. وبعد أن تمت اجراءات تصحيح الأوراق أدخلت البيانات في ذاكرة الحاسوب لتحليلها وتم الحصول على الاحصائيات التي ساعدت الباحث في الإجابة عن أسئلة الدراسة.

المعالجة الإحصائية

جرى حساب العديد من الإحصائيات لفقرات كل نموذج من نماذج الاختبار. فقد

حسب معامل الصعوبة لكل فقرة من هذه الفقرات وذلك بإيجاد نسبة من اجابوا عن الفقرة إجابة صحيحة ممن حاولوا الإجابة عنها. وحسب معامل التمييز لكل فقرة من فقرات نماذج الاختبار وذلك بإيجاد معامل الارتباط بين نتائج المفحوصين على هذه الفقرة ونتائجهم على الاختبار الكلي (Item-Total Correlation rit) وهو ما يعرف بمعامل ارتباط بوينت بايسريال (Point biserial correlation rpbi) واستخدم اختبار «ت» لوسطين مرتبطين Paired (t) test لفحص فرضية تساوي أوساط معاملات الصعوبة في النماذج الأربعة، ولتساوي أوساط معاملات التمييز بعد تحويلها إلى درجات فشر الزائية، على اعتبار أن معاملات التمييز والتي هي معاملات ارتباط بوينت بايسريال هي حالات خاصة من معامل ارتباط بيرسون. وكذلك استخدم اختبار ولكوكسون للأزواج المرتبطة والمعروف باسم Wilcoxon Matched-Pairs Signed-Ranks Test لفحص فرضية تطابق توزيع معاملات الصعوبة والتمييز للفقرات في نماذج الاختبار المختلفة. واستخدمت معادلة كودر–ريشاردسون (20) KR-20 لايجاد معاملات التجانس الداخلي لكل نموذج من نماذج الاختبار الأربعة.

النتائج

بينت النتائج أن قيم معاملات الاتساق الداخلي محسوبة بمعادلة كودر–ريشاردسون (20) هي (0،71) للنموذج الأول و (0،72) للنموذج الثاني و (0،82) للنموذج الثالث و (0،89) للنموذج الرابع أي أن أعلى معامل اتساق داخلي كان للنموذج الرابع الذي كان فيها البديل « لا شيء مما ذكر » هو الجواب الصحيح، يليه معامل الاتساق الداخلي للنموذج الثالث الذي كان أحد بدائل فقراته البديل « لا شيء مما ذكر ». ثم معامل الاتساق الداخلي للنموذج الثاني الذي لفقراته ثلاثة بدائل وأخيرًا يأتي معامل الاتساق الداخلي للنموذج الأول الذي لفقراته أربعة بدائل. ويتضح من هذه القيم لمعاملات الاتساق الداخلي أن نماذج الاختبار الأربعة المكوّن كل منها من (32) فقرة تتمتع بدرجة عالية من الاتساق الداخلي.

وفيما يتعلق بمعاملات صعوبة الفقرات فقد بينت النتائج أن قيم معاملات صعوبة الفقرات للنموذج الأول ذات الأربعة بدائل تراوحت من (0،267) إلى (0،978) بوسط مقداره (0،6236) وتراوحت هذه القيم لفقرات النموذج الثاني ذات الثلاثة بدائل من (0،202) إلى (0،902) بوسط مقداره (0،6089). أما قيم معاملات صعوبة فقرات

النموذج الثالث التي كان أحد بدائلها « لا شيء مما ذكر » فتراوحت من (0،220) إلى (0،902) بوسط مقداره (0،5534). وتراوحت قيم معاملات صعوبة فقرات النموذج الرابع التي كان فيها البديل « لا شيء مما ذكر » هو الجواب الصحيح من (0،079) إلى (0،763) بوسط مقداره (0،4180).

ويتضح من قيم أوساط معاملات صعوبة الفقرات أن أصعب هذه النماذج هو النموذج الرابع الذي كان فيه البديل « لا شيء مما ذكر » هو الجواب الصحيح، يليه في الصعوبة النموذج الثالث الذي كان احد بدائله « لا شيء مما ذكر »، ثم النموذج الثاني الذي لفقراته ثلاثة بدائل، ويأتي أخيرًا النموذج الأول الذي لفقراته أربعة بدائل.

ويبين الجدول (1) نتيجة اختبار «ت» للفرق بين وسطي زوجين مترابطين Paired (t) test أو ما يعرف باختبار «ت» للقياسات المتكررة Repeated Measures Contrast لمعاملات صعوبة الفقرات في النماذج المختلفة. ويبين الجدول (2) نتيجة اختبار ولكوكسون Wilcoxon المعروف بـ The Wilcoxon Matched-Pairs Signed-Ranks Test لمقارنة قيم هذه المعاملات في النماذج المختلفة أيضًا :

الجدول (1)

الأوساط الحسابية والانحرافات المعيارية وقيم «ت» لمعاملات صعوبة الفقرات في النماذج المختلفة

النموذج	الوسط الحسابي	الانحراف المعياري	الفرق بين الوسطين	الخطأ المعياري	قيمة «ت»	احتمال «ت»
1	0,6236	0,177				
			0,0147	0,021	0,69	0, 495
2	0,6089	0,191				
1	0,6236	0,191				
			0,0702	0,037	1,89	0, 068
3	0,5534	0,197				
1	0,6236	0,191				
			0,2056	0,025	8,22	0,0001
4	0,4180	0,174				
3	0,5534	0,197				
			0,1354	0,036	3,77	0,0001
4	0,4180	0,174				

الجدول (2)
نتائج اختبار ولكوكسون Wilcoxon لمقارنة قيم معاملات صعوبة الفقرات في النماذج المختلفة

النموذج	الوسط للرتب	عدد الرتب السالبة	عدد الرتب الموجبة	قيمة «ز»	احتمال «ز»
1	16,78	14		0, 71	0, 477
2	16,14		18		
1	18,13	20		1, 84	0, 065
3	13,79		12		
1	17,89	27		4,605	0,0001
4	3,25		4		
3	18,83	23		3, 16	0,0001
4	10,56		9		

يتضح من الجدول (1) أن الوسط لمعاملات صعوبة الفقرات في النموذج الأول (0,6236) أعلى من وسط معاملات صعوبة الفقرات في النموذج الثاني (0,6089) أي أن الفقرات في النموذج الثاني اصعب منها في النموذج الأول، ولكن الفرق بينهما (0,0147) لم يصل لمستوى الدلالة الإحصائية، حيث بلغت قيمة «ت» (0,69) باحتمال يساوي (0,495)، وهذه النتيجة اتفقت مع نتيجة اختبار ولكوكسون Wilcoxon كما يظهر ذلك في الجدول (2) حيث بلغت قيمة «ز» (0,71) باحتمال يساوي (0,477).

كما يتضح من الجدول (1) أيضًا أن وسط معاملات صعوبة الفقرات في النموذج الأول (0,6236) أعلى من وسط معاملات صعوبة الفقرات في النموذج الثالث (0,5534) أي أن الفقرات في النموذج الثالث أصعب منها في النموذج الأول، ولكن الفرق بينهما (0,702) لم يصل مستوى الدلالة الإحصائية، حيث أن قيمة (ت) تساوي (1,89) باحتمال يساوي (0,068) وهذه النتيجة لاختبار «ت» جاءت متفقة مع نتيجة اختبار ولكوكسون Wilcoxon كما يظهر ذلك في الجدول (2)، حيث بلغت قيمة (ز) 1,84 باحتمال يساوي (0,065).

ويتضح من الجدول (1) أيضًا أن وسط معاملات صعوبة الفقرات في النموذج الأول (0,6236) أعلى من وسط معاملات صعوبة الفقرات في النموذج الرابع (0,4180)، وأن الفرق بينهما (0,2056) دال احصائيًا حيث أن قيمة «ت» تساوي (8,22) باحتمال يساوي (0,0001). وهذا يعني أن الفقرات في النموذج الرابع أصعب من الفقرات في النموذج الأول. واتفقت هذه النتيجة لاختبار «ت» مع نتيجة اختبار ولكوكسون Wilcoxon كما يظهر ذلك في الجدول (2) حيث بلغت قيمة (ز) (4,605) باحتمال يساوي (0,0001) أيضًا.

ويتضح من الجدول (1) أن وسط معاملات صعوبة الفقرات في النموذج الثالث (0,5534) أعلى من وسط معاملات صعوبة الفقرات في النموذج الرابع (0,4180) وأن الفرق بينهما (0,1354) دال احصائيًا حيث أن قيمة (ت) تساوي (3,77) باحتمال يساوي (0,0001) وهذا يعني أن الفقرات في النموذج الرابع اصعب من الفقرات في النموذج الثالث. واتفقت هذه النتيجة لاختبار (ت) مع نتيجة اختبار ولكوكسون Wilcoxon كما يظهر ذلك في الجدول (2) حيث بلغت قيمة (ز) (3,16) باحتمال يساوي (0,0001) أيضًا.

أما فيما يتعلق بمعاملات تمييز الفقرات والتي هي عبارة عن معاملات ارتباط أداء المفحوصين على هذه الفقرات مع أدائهم على الاختبار الكلي وهي هنا عبارة عن معاملات ارتباط بوينت بايسيريال (r_{pbis}) Point biserial correlation للفقرات. فقد بينت النتائج أن قيم معاملات التمييز لفقرات النموذج الأول ذات الأربعة بدائل تراوحت من (0,016) إلى (0,669) بوسط مقداره (0,2176). وتراوحت معاملات التمييز لفقرات النموذج الثاني ذات الثلاثة بدائل من (0,017) إلى (0,488) بوسط مقداره (0,2349). أما قيم معاملات تمييز فقرات النموذج الثالث التي كان أحد بدائلها « لا شيء مما ذكر » فتراوحت من (0,009) إلى (0,589) بوسط مقداره (0,3201). وتراوحت قيم معاملات تمييز فقرات النموذج الرابع التي كان فيها البديل « لا شيء مما ذكر » هو الجواب الصحيح من (0,068) إلى (0,719) بوسط مقداره (0,4156).

ويتضح من قيم أوساط معاملات تمييز الفقرات في النماذج الأربعة أن اعلاها تمييزا هي فقرات النموذج الرابع تليها فقرات النموذج الثالث ثم فقرات النموذج الثاني فففقرات النموذج الأول.

الجدول (3)
الأوساط الحسابية والانحرافات المعيارية وقيم «ت» لمعاملات تمييز الفقرات في النماذج المختلفة

النموذج	الوسط الحسابي	الانحراف المعياري	الفرق بين الوسطين	الخطأ المعياري	قيمة «ت»	احتمال «ت»
1	0,2176	0,117				
2	0,2349	0,164	0,0173	0,029	0,60	0, 553
1	0,2176	0,164				
3	0;3201	0,158	0,1025	0,038	2,69	0, 012
1	0,2176	0,164				
4	0,4156	0,221	0,1980	0,043	4,60	0,0001
3	0,3201	0,158				
4	0,4156	0,221	0,0955	0,051	1,88	0, 069

ويبين الجدول (3) نتيجة اختبار «ت» للفرق بين وسطي زوجين مترابطين Paired (t) test أو ما يعرف باختبار «ت» للقياسات المتكررة Repeated Measures Contrast لمعاملات تمييز الفقرات في النماذج المختلفة بعد تحويلها إلى درجات فشر الزائية. ويبين الجدول (4) نتيجة اختبار ولكوكسون Wilcoxon لمقارنة قيم هذه المعاملات في النماذج المختلفة أيضًا :

الجدول (4)
نتائج اختبار ولكوكسون Wilcoxon لمقارنة قيم معاملات تمييز الفقرات في النماذج المختلفة

النموذج	الوسط للرتب	عدد الرتب السالبة	عدد الرتب الموجبة	قيمة «ز»	احتمال «ز»
1	17,61	18		0,991	0, 322
2	15,07		14		
1	11,36	11		2,599	0, 009
3	19,19		12		
1	13, 6	5		3,665	0,0002
4	17,04		27		
3	16, 2	10		1,907	0,0565
4	16,64		22		

يتضح من الجدول (3) أن الوسط معاملات تمييز الفقرات في النموذج الثاني (0,2349) أعلى من وسط معاملات تمييز الفقرات في النموذج الأول (0,2176) ولكن الفرق بينهما (0,0173) لم يصل مستوى الدلالة الإحصائية فكانت قيمة «ت» (0,60) باحتمال يساوي (0,553)، واتفقت مع هذه النتيجة لاختبار «ت»، نتيجة اختبار ولكوكسون Wilcoxon كما يظهر ذلك في الجدول (4) حيث بلغت قيمة (ز) (0,991) باحتمال يساوي (0,322).

ويتضح من الجدول (3) أن وسط معاملات تمييز الفقرات في النموذج الثالث (0,3201) أعلى من وسط معاملات تمييز الفقرات في النموذج الأول (0,2176)، وأن الفرق بينهما (0,1025) دال احصائيًا حيث بلغت قيمة «ت» (2,69) باحتمال يساوي (0,012) وتتفق هذه النتيجة لاختبار «ت» مع نتيجة اختبار ولكوكسون Wilcoxon كما يظهر ذلك في الجدول (4)، حيث بلغت قيمة (ز) (2,599) باحتمال يساوي (0,009).

ويتضح من الجدول (3) أيضًا أن وسط معاملات تمييز الفقرات في النموذج الرابع (0,4156) أعلى من وسط معاملات تمييز الفقرات في النموذج الأول (0,2176)، وأن الفرق بينهما (0,1980) دال احصائيًا فكانت قيمة «ت» (4,60) باحتمال يساوي (0,0001). وتتفق هذه النتيجة لاختبار «ت» مع نتيجة اختبار ولكوكسون Wilcoxon كما يظهر ذلك في الجدول (4) حيث بلغت قيمة «ز» (3,665) باحتمال يساوي (0,0002).

وفي الجدول (3) يظهر أن وسط معاملات تمييز الفقرات في النموذج الرابع (0,4156) أعلى من وسط معاملات تمييز الفقرات في النموذج الثالث (0,3201)، ولكن الفرق بينهما (0,0955) لم يصل مستوى الدلالة الإحصائية، حيث بلغت قيمة «ت» (1,88) باحتمال يساوي (0,069) وتتفق هذه النتيجة لاختبار «ت» مع نتيجة اختبار ولكوكسون Wilcoxon كما يظهر ذلك في الجدول (4) حيث بلغت قيمة (ز) (1,907) باحتمال يساوي (0,565).

مناقشة النتائج

بينت النتائج أن الفرق بين وسط معاملات صعوبة الفقرات في النموذج الأول ذات الأربعة بدائل والفقرات في النموذج الثاني ذات الثلاثة بدائل لم يكن دالا احصائيًا وكذلك الحال بالنسبة إلى معاملات تمييزها. وكانت قيمة «ت» لاختبار الفرق بين وسطي معاملات الصعوبة لفقرات هذين النموذجين (0,69) باحتمال يساوي (0,495). وقيمتها لاختبار الفرق بين وسطي قيم معاملات تمييز فقراتهما (0,60) باحتمال يساوي (0,553). وجاءت هذه النتائج لاختبار «ت» لدلالة الفروق بين أوساط قيم معاملات الصعوبة والتمييز في النموذج الأول والثاني متفقة مع نتائج اختبار ولكوكسون Wilcoxon لمعرفة مدى تماثل توزيعي معاملات الصعوبة والتمييز لفقرات هذين النموذجين، فكانت قيمة «ز» (0,71) باحتمال يساوي (0,477) في حالة مقارنة توزيعي معاملات الصعوبة في النموذجين، وقيمتها (0,991) باحتمال يساوي (0,322) في حالة مقارنة توزيعي معاملات التمييز في هذين النموذجين من الفقرات. فقبول الفرضية الصفرية في هذه الحالة دليل على عدم اختلاف قيم معاملات صعوبة الفقرات في هذين النموذجين، وكذلك الحال بالنسبة إلى قيم معاملات تمييز فقراتهما.

ومن هذه النتائج التي أظهرت عدم اختلاف صعوبة وتمييز الفقرات ذات الثلاثة بدائل عن صعوبة وتمييز الفقرات ذات الأربعة بدائل أمكن الإجابة عن السؤال الأول من أسئلة الدراسة والمتعلق بمدى اختلاف صعوبة وتمييز فقرات الاختيار من متعدد ذات الثلاثة بدائل عن قيم معاملات صعوبة وتمييز الفقرات ذات الأربعة بدائل. حيث تبين أنه لا توجد فروق دالة احصائيًا بين أوساط قيم معاملات الصعوبة والتمييز فيهما. واتفقت هذه النتيجة مع نتائج دراسات سابقة أظهرت عدم وجود فروق دالة احصائيًا بين معاملات الصعوبة والتمييز لفقرات ذات ثلاثة بدائل وأربعة بدائل كدراسة كوستن (Costin, 1972) ولاندرم وكاشن وثيس (Landrum, Cashin & Theis, 1993) ولكنها تعارضت مع نتائج دراسات أخرى كدراسة كوستن (Costin, 1970) وستراتون وكاتش (Straton & Catts, 1980) ووليامز وايبيل (Williams & Ebel, 1957).

إن هذه النتيجة التي أظهرت عدم وجود فروق دالة احصائيًا بين صعوبة وتمييز الفقرات ذات الثلاثة بدائل والفقرات ذات الأربعة بدائل يعزز دور الفقرات ذات الثلاثة بدائل في عملية القياس الصفي. فاستخدام ثلاثة بدائل للفقرات يجعل كتابتها أيسر والجهد المبذول في كتابتها أقل، لأن تقليل عدد البدائل يساعد على إعداد بدائل فعّالة ومتجانسة. فاختيار البدائل الفعّالة والمتجانسة للفقرات هي من القضايا الهامة في إعداد فقرات اختبارات الاختيار من متعدد. كما أن تقليل عدد البدائل في فقرات الاختبار يساعد على تقليل الوقت اللازم لتطبيقه، مما يمكّن المدرّس من زيادة عدد الفقرات المستخدمة في الاختبار الواحد. وهذه الزيادة في عدد الفقرات تساعد على تمثيل الفقرات لمحتوى المادة الدراسية بشكل أفضل، وهذا بدوره يحسن من صدق المحتوى للاختبار ومن ثبات درجات المفحوصين عليه.

وبينت النتائج أن الفرق بين وسطي قيم معاملات صعوبة الفقرات في النموذج الأول ذات الأربعة بدائل والفقرات في النموذج الثالث التي فيها البديل «لا شيء مما ذكر» كأحد بدائلها لم يكن دالًا. حيث كانت قيمة «ت» للفرق بين هذين الوسطين (1,89) باحتمال يساوي (0,068). واتفقت هذه النتيجة لاختبار «ت» مع نتيجة اختبار ولكوكسون Wilcoxon لمعرفة مدى تشابه توزيعي قيم معاملات الصعوبة في هذين النموذجين. فكانت قيمة «ز» (1,84) باحتمال يساوي (0,065).

أما الفرق بين وسطي قيم معاملات التمييز للفقرات في هذين النموذجين فكان دالًا،

حيث بلغت قيمة «ت» لهذا الفرق (2,69) باحتمال يساوي (0,012). واتفقت هذه النتيجة لاختبار «ت» مع نتيجة اختبار ولكوكسون Wilcoxon لمعرفة مدى تشابه توزيعي قيم معاملات التمييز في هذين النموذجين، فكانت قيمة «ز» (2,599) باحتمال يساوي (0,009).

ومن هذه النتائج أمكن الإجابة عن السؤال الثاني من اسئلة الدراسة والمتعلق بمدى اختلاف قيم معاملات صعوبة وتمييز فقرات الاختيار من متعدد التي تحتوي البديل « لا شي مما ذكر» عن قيم معاملات صعوبة وتمييز الفقرات التي لا تحتوي هذا البديل. فتبيّن أنه لا يوجد فرق دال احصائيًا بين قيم معاملات صعوبة الفقرات التي تحتوي البديل «لا شيء مما ذكر» وقيم معاملات صعوبة الفقرات التي لا تحتوي هذا البديل. أما الفرق بين وسطي قيم معاملات التمييز فكان دالًا، حيث كانت قيم معاملات تمييز الفقرات التي تحتوي البديل «لا شيء مما ذكر» أعلى من قيم تمييز الفقرات التي لا تحتوي هذا البديل. فتساوي قيم معاملات صعوبة الفقرات والتمييز الأعلى للفقرات التي تحتوي على هذا البديل يعزز دور استخدام البديل «لا شيء مما ذكر» في كتابة فقرات الاختبارات لأن استخدامه يقلل الجهد المبذول في إعداد بدائل فعّالة ومتجانسة، حيث يمكن استخدام هذا البديل كواحد من بدائل فقرات الاختيار من متعدد.

إن هذه النتيجة التي أظهرت عدم وجود فرق دال احصائيًا بين صعوبة الفقرات التي تحتوي البديل «لا شيء مما ذكر» كمموّه وتلك الفقرات التي لا تحتوي هذا البديل قد اتفقت مع نتائج عدد من الدراسات السابقة كدراسة (Oosterhof & Coats, 1984) ودراسة (Tolleffson, 1987) ودراسة (Wesman & Bennett, 1946) ولكنها تعارضت مع نتائج دراسات أخرى أيضًا كدراسات (Cerhan, Haladyna & Brewer, 1993) و (Dudycha & Charpenter, 1973) و (Rimland, 1960).

كما بينت النتائج أن الفرق بين وسط قيم معاملات صعوبة الفقرات في النموذج الأول ذات الأربعة بدائل والفقرات في النموذج الرابع التي فيها البديل «لا شيء مما ذكر» كجواب صحيح كان دالا احصائيًا فكانت قيمة «ت» للفرق بين وسطي قيم معاملات صعوبة الفقرات في هذين النموذجين (8,22) باحتمال يساوي (0,0001)، واتفقت هذه النتيجة لاختبار «ت» مع نتيجة اختبار ولكوكسون Wilcoxon أيضًا، فكانت قيمة «ز» (4,605) باحتمال يساوي (0,0001).

وفيما يتعلق بالفرق بين وسطي قيم معاملات التمييز في هذين النموذجين فكان دالًا أيضًا، إذ بلغت قيمة «ت» لهذا الفرق (4,6) باحتمال يساوي (0,0001) واتفقت هذه النتيجة لاختبار «ت» مع نتيجة اختبار ولكوكسون Wilcoxon أيضًا حيث بلغت قيمة «ز» (3,665) باحتمال يساوي (0,0002).

ومن هذه النتائج أيضًا أمكن الإجابة عن السؤال الثالث من أسئلة الدراسة والمتعلق بمدى اختلاف صعوبة وتمييز فقرات الاختيار من متعدد التي تحتوي البديل «لا شيء مما ذكر» كجواب صحيح في هذه الفقرات عن قيم صعوبة وتمييز الفقرات التي لا تحتوي هذا البديل. حيث تبين أن الفقرات التي تحتوي هذا البديل كجواب صحيح أصعب وأعلى تمييزًا من الفقرات التي لا تحتوي هذا البديل وبشكل دال احصائيًا. فالتمييز العالي للفقرات التي تحتوي البديل «لا شيء مما ذكر» كجواب صحيح يعزز صلاحية استخدام هذا البديل في فقرات الاختيار من متعدد. كما أن هذه النتائج التي أظهرت وجود فروق دالة احصائيًا بين صعوبة وتمييز الفقرات التي تحتوي البديل «لا شيء مما ذكر» كجواب صحيح وتلك الفقرات التي لا تحتوي هذا البديل قد اتفقت مع نتائج دراسات، (Knowles & Wellch, 1990) و (Oosterhof & Coats, 1984) وتعارضت مع نتائج دراسات (Tolle & Fson & Tripp, 1983; Tollefson, 1987).

كما بينت النتائج أن الفرق بين وسط قيم معاملات صعوبة الفقرات في النموذج الثالث التي فيها «لا شيء مما ذكر» كبديل ووسط قيمها في النموذج الرابع التي فيها «لا شيء مما ذكر» كجواب صحيح هو فرق دال احصائيًا. فكانت قيمة «ت» (3,77) باحتمال يساوي (0,0001) واتفقت هذه النتيجة لاختبار «ت» مع نتيجة اختبار ولكوكسون Wilcoxon أيضًا فكانت قيمة «ز» (3,16) باحتمال يساوي (0,0001). أما الفرق بين وسطي قيم معاملات التمييز في هذين النموذجين فانه لم يصل مستوى الدلالة الاحصائية. فكانت قيمة «ت» للفرق بين الوسطين (1,88) باحتمال يساوي (0,069). وقيمة «ز» (1,907) باحتمال يساوي (0,0565).

ومن هذه النتائج أمكن الإجابة عن السؤال الرابع من أسئلة الدراسة والمتعلق بمدى اختلاف قيم معاملات صعوبة وتمييز فقرات الاختيار من متعدد التي تحتوي البديل «لا شيء مما ذكر» كبديل عن قيم معاملات صعوبة وتمييز الفقرات التي تحتويه كجواب صحيح. حيث تبين أن الفرق بين وسطي قيم معاملات الصعوبة في النموذجين دال

احصائيًا. فكانت الفقرات في النموذج الرابع التي تحتوي البدائل «لا شيء مما ذكر» كجواب صحيح أصعب من الفقرات التي تحتوي هذا البديل كواحد من بدائلها.

أما الفرق بين وسطي قيم معاملات التمييز في هذين النموذجين فلم يكن دالًا. فتساوي قيم معاملات التمييز في الفقرات التي تحتوي البديل «لا شيء مما ذكر» كجواب صحيح أو كواحد من البدائل يشجّع استخدام هذا البديل في فقرات الاختبار من متعدد سواء كان هو أحد البدائل أو الإجابة الصحيحة فيها.

إن هذه النتيجة التي أظهرت وجود فرق دال احصائيًا بين صعوبة الفقرات التي تحتوي البديل «لا شيء مما ذكر» كبديل وتلك الفقرات التي تحتويه كجواب صحيح اتفقت مع نتائج دراسة (Oosterhof & Coats, 1984) ودراسة (Tollefon, 1987) فيما يتعلق بتمييز هذا النوع من الفقرات ولكنها تعارضت مع نتائج دراسة (Tollefson, 1987) فيما يتعلق بصعوبة هذا النوع من الفقرات ومع دراسة (Tollefson & Tripp, 1983) فيما يتعلق بتمييزها.

لقد بينت النتائج عدم وجود فروق دالة احصائيًا بين صعوبة وتمييز الفقرات ذات الثلاثة بدائل والأربعة بدائل مما يقدّم الدعم والتأييد لوجهة النظر المنادية باستخدام الفقرات ذات الثلاثة بدائل في اختبارات الاختيار من متعدد. فهي الأسهل عند الإعداد، والمفضلة لدى التلاميذ عند التطبيق لأنها لا تسبب كثيرًا من التشتت والإرباك لهم مقارنة بالفقرات ذات الأربعة بدائل عند الإستجابة عليها.

فتقليل عدد البدائل يساعد على تقليل الجهد المبذول في كتابة الفقرات واعدادها، كما أنه يساعد على تحاشي وجود بدائل من شأنها تقليل فاعلية الفقرات لوجود بدائل غير متجانسة أو غير فعّالة في هذه الفقرات. فالتخلي عن المموّهات غير الفعّالة يجعل من المموّهات المتبقية في الفقرات أكثر تجانسًا وفاعلية. كما أن تطبيق اختبارات لفقراتها ثلاثة بدائل يحتاج من الوقت أقل مما تحتاجه اختبارات لفقراتها أربعة بدائل أو أكثر من ذلك، وهذا بدوره يسمح بزيادة عدد فقرات الاختبار وزيادة عدد المفاهيم المقيسة بهذا الاختبار، مما يزيد من تمثيل الفقرات لمحتوى المادة الدراسية، وبالتالي يحسّن صدق المحتوى لهذا الاختبار وثبات درجات المفحوصين عليه.

كما أن هذه النتيجة التي بيّنت بشكل تجريبي عدم وجود فروق دالة بين صعوبة وتمييز

الفقرات ذات الثلاثة بدائل والأربعة بدائل جاءت لتدحض الاعتقاد النظري القائل بأن الفقرات ذات الأربعة بدائل هي أصعب من الفقرات ذات الثلاثة بدائل (على اعتبار أن احتمال الإجابة بالتخمين فيها 25 % بينما في الفقرات ذات الثلاثة بدائل 33 %) على اعتبار أن المفحوصين يخمنون عشوائيًا عند إجابتهم عن الأسئلة التي لا يعرفون اجاباتها. مع أن الواقع هو غير ذلك، أي أن التلاميذ لا يخمنون عشوائيًا عند إجابتهم ولكنهم يحاولون الاستفادة من معرفتهم الجزئية لها. وهذا ما أثبتته النتائج التي بينت أن صعوبة وتمييز الفقرات كان واحدًا في نموذجي الاختبار. وبالإضافة للمعرفة الجزئية فإن المفحوصين يعتمدون في اجاباتهم للفقرات على بعض المنبهات أو المؤشرات التي يمكن الحصول عليها من الفقرات نفسها. وهذه المنبهات تعتبر أكثر تهديدًا وخطورة من عدد البدائل لمعاملات الصعوبة والتمييز وحتى للصدق والثبات لهذا النوع من الاختيارات.

كما بينت النتائج أن وجود البديل «لا شيء مما ذكر» لا يزيد من صعوبة الفقرات إلّا إذا كان هو الجواب الصحيح. أما تأثيره على تمييز الفقرات فقد كان واضحًا، فتمييز الفقرات التي تحتوي هذا البديل، سواء كان كمموّه أو كجواب صحيح، أعلى من تمييز الفقرات التي لا تحتويه. وهذا دليل يؤيد وجهة النظر المنادية باستخدام هذا البديل في فقرات الاختيار من متعدد. حيث أنه من الأسس التي يتم في ضوئها اختيار فقرات الاختبار هو قيم معاملات تمييزها. وأما وجهة النظر التي لا تؤيد استخدام هذا البديل لكونه يجعل من الفقرات أكثر صعوبة وأقل تمييزًا فيبدو أنها غير دقيقة.

أما قضية زيادة صعوبة الفقرات باستخدام هذا البديل عندما يكون هو الجواب الصحيح فأمر قد يكون مرغوبًا فيه يحتاج إليه واضع الاختبار في بعض المواقف الاختبارية التي تتطلب فقرات صعبة. ويمكن حل قضية اختلاف صعوبة الاختبارات عن طريق اجراء معادلة Equating للاختبارات ذات المستويات المختلفة في الصعوبة.

وفي ضوء نتائج هذه الدراسة والدراسات السابقة التي اتفقت أو تعارضت معها فإن الباحث يوصي بإجراء المزيد من الدراسات الامبريقية تقارن الخصائص السيكومترية لاختبارات الاختيار من متعدد وفقراتها، وذلك باستخدام فقرات تختلف في عدد بدائلها، وباستخدام البديل «لا شيء مما ذكر» كمموّه في بعض هذه الفقرات، وكجواب صحيح في البعض الآخر.

كما يوصي الباحث باستخدام اختبارات لفقراتها ثلاثة بدائل واختبارات فيها البديل «لا شيء مما ذكر» كمموّه أو كجواب صحيح، سواء كان في الاختبارات المدرسية أو الجامعية منها. كما يوصي أيضا باستخدام فقرات ذات بدائل مختلفة ضمن الاختبار الواحد، إلّا إذا أريد استخدام معادلة التصحيح لأثر التخمين في تلك الاختبارات، وهذا ما يندر عمله من قبل المدرسين.

فوضع عدة خيارات أمام باني الاختبار أي أن يتضمن الاختبار فقرات ذات عدد مختلف من البدائل، ثلاثة أو أربعة بدائل، واستخدام البديل «لا شيء مما ذكر» كمموّه أو كجواب صحيح وغيرها من الخيارات، يسهّل عليه بناء مثل هذا الاختبار دون اضطراره لحشر بعض البدائل غير الفعّالة والتي يكون وجودها أكثر ضررا من عدم وجودها في فقرات هذا الاختبار.

المراجع

Costin, F. (1970). The optimal number of alternatives in multiple-choice achievement tests : Some empirical evidence for a mathematical proof. **Educational and Psychological Measurement,** 30, 353-358.

Costin, F. (1970). Three-choice versus four-choice items : Implications of reliability and validity of objective achievement tests. **Educational and Psychological Measurement,** 32, 1035-1038.

Crehan, K. D. and Haladyna, T. M. (1991). The validity of two item-writing rules. **Journal of Experimental Education,** 59, 183-192.

Crehan, K. D., Haladyna, T. M. and Brewer, B. W. (1993). Use of an inclusive option and the optimal number of options for multiple-choice items. **Educational and Psychological Measurement,** 53, 241-247.

Forsyth, R. A. and Spratt, K. F. (1980). Measuring problem solving ability in mathematics with multiple-choice items : The effects of item format on selected item test characteristics. **Journal of Education Measurement,** 17(1), 31-43.

Frary, R. B. (1991). The none-of-the above option : An empirical study. **Applied Measurement in Education,** 4(2), 115-124.

Haladyna, T. M. and Downing, S. M. (1989). A taxonomy of multiple-choice item wrinting rules. **Applied Measurement in Education,** (1), 37-50.

Knowles, S. L. and Welch, C. A. (1992). A meta-analytic review of item discrimination and difficulty in multiple-choice items using «none-of-the-above». **Educational and Psychological Measurement,** 52, 571-577.

Landrum, R. E., Cashin, J. R. and Theis, K. S. (1993) More evidence in favor of three-option multiple-choice tests. **Educational and Psychological Measurement,** 53, 771-778.

Oosterhof, A. C. and Coats, P. K. (1984). Comparison of difficulties and reliabilities of quantitative word problems in completion and multiple choice item formats. **Applied Psychological Measurement,** 8(3), 287-294.

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

Tollefson, N. (1987). A comparison of the item difficulty and item discrimination of multiple-choice items using the «none of the above» and one correct options. **Educational and Psychological Measurement,** 47, 377-383.

Williams, B. J. and Ebel, R. L. (1957). The effect of varying the number of alternatives per item and multiple-choice vocabulary test items. In Owen, S. V. and Froman, R. D. (1987). What's wrong with three-option multiple-choice items ? **Educational and Psychological Measurement,** 47, 513-522.