"أثر اختلاف عدد البدائل الصحيحة في فقرات الصواب ـ خطأ المتعدد في دقة تقدير صعوبة الفقرات وقدرات الأفراد وثبات الاختبار"

حمدي يونس أبو جراد

جامعة القدس المفتوحة – غزة

Hamdy Younis Abu jarad

Al-Quds Open University-Gaza

hjarad@qou.edu

ملخص

هدفت الدراسة الحالية إلى الكشف عن أثر اختلاف عدد البدائل الصحيحة في فقرات الصواب - خطأ المتعدد في دقة تقدير صعوبة الفقرات وقدرات الأفراد وثبات الاختبار، ولتحقيق هدف الدراسة تم بناء اختبار تحصيلي من نوع صواب- خطأ متعدد ، في الرياضيات لطلبة الصف الثامن مكون من ٢٥ فقرة ، واشتمل الاختبار على ثلاثة نماذج حسب عدد البدائل الصحيحة ، وباستخدام برنامج Winsteps تم تحليل استجابات الطلبة البالغ عددهم ٢٧٨٠ طالباً وطالبة لجميع نماذج الاختبار الثلاثة وفق نموذج التقدير الجزئي . كشفت نتائج تحليل التباين الأحادي أن هناك فروق دالة إحصائيا في الأوساط الحسابية للأخطاء المعيارية الخاصة بصعوبة الفقرات بين النموذج الثاني وكلٍ من النموذجين الأول والثالث لصالح النموذج الثاني(عدد البدائل الصحيحة = ٢) ، وبينت النتائج أن هناك فروق دالة إحصائيا في الأوساط الحسابية الخاصة بقدرات الأفراد بين النموذج الثاني وكلٍ من النموذجين الأول والثالث لصالح النموذج الثاني، وكشفت النتائج أيضاً عن عدم وجود فروق دالة إحصائيا(α= ٠,٠٥) بين معاملات الثبات للنماذج الثلاثة.(الكلمات المفتاحية: فقرات الصواب – خطأ المتعدد ، عدد البدائل الصحيحة ، نموذج التقدير الجزئي ، الثبات، نظرية الاستجابة للمفردة)

Abstract

The present study aimed to examine the effect of the different number of correct alternatives in multiple true- false items on the estimation's accuracy of item's difficulty, individuals' abilities and test reliability. To achieve the study aims, a 25 item multiple true - false achievement test in 8th grade mathematics was constructed . The test included three forms according to the number of correct alternatives. By using Winsteps program, responses of 2780 male and female students were analyzed for all the three test forms according to partial credit model. The results of One-Way ANOVA revealed a significant statistical difference in the standard errors of means for difficulty of the items between the second form and both the first and third ones in behalf of the second form (the number of correct alternatives=2). The results revealed significant statistical differences in the means for the individuals' abilities between the second form and both the first and third ones in behalf of the second form. Also the results revealed no significant statistical differences ($\alpha = 0.05$) between the reliability coefficients for the three forms.(Key words: multiple true- false items, Number of correct alternative, partial credit model, Reliability, Item Response theory)

مقـــدمـــة

تستعمل الاختبارات بأشكالها المختلفة لقياس تحصيل الطلبة في مختلف المواد الدراسية، وتشكل الدرجات وما ينبثق عنها من تقديرات أساساً مهما للكثير من الإجراءات، والقرارات الهامة التي ترتبط بوضع الفرد وتؤثر عليه. فأهلية الفرد للاستمرار بالدراسة، أو القبول في برنامج معين، أو الحصول على بعثة دراسية، أو وظيفة معينة تتقرر بالمستوى الأكاديمي الذي يحققه ممثلاً في الدرجات أو التقديرات التي يحصل عليها.

وتؤدي الدرجات وظائف فعلية ترتبط بحاجات الطلبة وأولياء الأمور والمدرسين وأصحاب العمل؛ فهي تلعب دوراً مهماً في تكوين الطالب صورة عن ذاته، وستبقى من أفضل عوامل التنبؤ بالأداء اللاحق(ثورندايك وهيجن، ١٩٨٩).

وتصنف فقرات الاختبارات التحصيلية وفقاً لنوع الاستجابة المطلوبة إلى فقرات ذات إجابة مصاغة أى تلك التي يعطي فيها المفحوص إجابته بلغته الخاصة، وفقرات ذات إجابة منتقاة وفيها ينتقي الطالب الإجابة من بين عدة بدائل معطاة، ويتوقف استخدام نوع الفقرة على عدة عوامل منها: مستوى الهدف المراد قياسه، وعمر المفحوص، وغرض الاختبار، وعدد الطلبة، وظروف التطبيق، ومهارة المعلم في صياغة الأسئلة(عودة ، ٢٠١٠).

وتعد فقرات الاختيار من متعدد هي الأكثر استعمالا لقياس تحصيل الطالب في العديد من الأغراض التربوية (ثورندايك وهيجن، ١٩٨٩ ؛ Gronlund, 1982 ; Aiken, 1987) حيث تتسم النتائج المتوفرة عنها بدرجة عالية من الصدق والثبات مقارنة بأشكال الفقرات الأخرى .

وقد تزايد الاهتمام بصناعة فقرات الاختيار من متعدد لكثرة المزايا التي تتمتع بها فهي دقيقة البنية، سهلة التصحيح، تطبق على أعداد كبيرة من المفحوصين، قدرتها على تحقيق درجة عالية من صدق المحتوى التعليمي؛ وعلى الرغم من تلك المميزات إلا أنها تتأثر بعامل التخمين، كما أنها لا تراعي المعرفة الجزئية. وقد كان هناك عدد من المحاولات لمعالجة قصور فقرات الاختيار من متعدد في مراعاة المعرفة الجزئية فقد قدم بوش(Bush ,2001) نوعاً آخر من الفقرات سميت بفقرات الاختيار من متعدد المتحررة Liberal multiple choice items وفي هذا النوع من الفقرات يسمح للطلبة باختيار أكثر من إجابة في حالة عدم تمكنهم من اختيار الإجابة الصحيحة وتصحح استجاباتهم على النحو التالي:

يعطى الطالب العلامة (١) في حالة اختياره للبديل الصحيح من بين ثلاثة بدائل

يعطى الطالب (٦٧%) من العلامة في حالة اختياره لبديلين أحدهما البديل الصحيح

يعطى الطالب (٣٣%) من العلامة في حالة اختياره لثلاثة بدائل أحدها البديل الصحيح

يعطى الطالب (٠) غير ذلك

كما اقترح أكرويد (Akeroyd ,1982) اسلوباً آخر لمعالجة المعرفة الجزئية يشبه الأسلوب السابق حيث يستطيع الطالب اختيار أكثر من بديل إلا أن اسلوب التصحيح في هذا الأسلوب مختلف عما قبله ففي الفقرة ذات الأربع بدائل يعطى الطالب العلامة (١) فى حالة اختياره البديل الصحيح، والعلامة (٠,٥) في حالة اختياره لبديلين أحدهما البديل الصحيح، و(٠,٢٥) في حالة استبعاده لأحد المموهات، و(٠) غير ذلك.

وتعتبر فقرات الصواب والخطأ المتعدد أحد أشكال الفقرات التي تعالج المعرفة الجزئية لدى الطلبة، وتتكون فقرة الصواب والخطأ المتعدد من متن يكون على شكل جملة خبرية لا على شكل سؤال ، ومجموعة من البدائل(بديلين أو أكثر) كل بديل فيها يمكن أن يكون صواباً أو خطأ ، وكل بديل يتم التعامل معه كسؤال مستقل عن بقية البدائل ويضع الطالب أمام كل بديل إشارة صواب أو خطأ وليس اختيار بديل واحد من بين البدائل كما في فقرات الاختيار من متعدد.

ولم تحظَ فقرات الصواب والخطأ المتعدد بالبحث والدراسة بالمقارنة مع أشكال الفقرات الأخرى كفقرات الاختيار من متعدد ، أو فقرات الاختيار من متعدد المركب، لكنه من الممكن من الناحية النظرية على الأقل، أن تعطي نفس النتائج المرغوبة التي تعطيها فقرات الاختيار من متعدد(Frisbie & Sweeny, 1982).

وقد زاد الاهتمام بفقرات الصواب والخطأ المتعدد في السنوات العشرة الأخيرة من القرن الماضي وبالأخص في المجالات الطبية، ويعود ذلك إلى أن أغلب الطلبة يفضلون فقرات الصواب والخطأ المتعدد كونها تراعي المعرفة الجزئية ويمكنهم أن يحصلوا من خلالها على علامات أعلى، ولتفوقها على فقرات الاختيار من متعدد في الخصائص السيكومترية(Frisbie, 1992).

وتقتضي مسألة التوسع في استعمال هذا النوع من الفقرات، الإلمام بها والوعي بقواعد بنائها، واكتساب مهارات إعدادها وللأسف، فإنه في الوقت الذي تؤدي فيه كتب القياس والتقويم المتوافرة حالياً دوراً أساسياً في التعريف بالأنواع المختلفة للأسئلة والإرشادات الخاصة بكتابتها، يلاحظ أنها عاجزة عن أداء الدور نفسه فيما يتعلق باختبارات الصواب والخطأ المتعدد(سوالمة، وقواسمه، ٢٠٠٠).

وعلى الرغم من أن المعرفة الشاملة بقواعد كتابة فقرات الصواب والخطأ المتعددة لم تكتمل بعد إلا أن الدراسات السابقة في المجال وفرت إجابات معقولة لعدد من الأسئلة المتعلقة بذلك النوع من الفقرات فقد قدم فرزبي(Frisbie, 1992) من خلال مراجعته للأدب التربوي المرتبط بفقرات الصواب والخطأ المتعدد إجابات لعدد من الأسئلة المتعلقة بهذا النوع من الفقرات مقارنة بفقرات الاختبار من متعدد من حيث الزمن اللازم لإجابتها، والطرق المناسبة لتقدير الثبات، ومؤشرات صدقها، وصعوبتها، وإدراك التلاميذ وتقبلهم لها، وخلصت المراجعة إلى أن الزمن اللازم لإجابة فقرة صواب وخطأ متعدد أطول من الزمن اللازم لإجابة فقرة اختيار من متعدد مكافئة لها في المحتوى، فالزمن الذي يكفي لإجابة ثلاث فقرات صواب وخطأ متعدد قد يكفي لأكثر من أربع فقرات اختبار من متعدد، وفيما يتعلق بالثبات فقد تفوقت فقرات الصواب والخطأ المتعدد على فقرات الاختبار من متعدد بغض النظر عن طريقة التصحيح، وبالنسبة للصدق التلازمي، تبين أن معاملات الارتباط بعد التصحيح لأثر الانخفاض أنه يمكن استخدام فقرات الصواب والخطأ المتعدد وفقرات الاختبار من متعدد على نحو تبادلي حيث تراوحت معاملات الارتباط المصححة بين (٠,٩١) و(١,٠)، وبخصوص صعوبة الفقرات كانت النتائج متفاوتة وغير مستقرة.

وأجرى الصمادي (٢٠١٢) دراسة هدفت إلى الكشف عن أثر طريقة تصحيح اختبار الصواب والخطأ المتعدد في معاملات الصدق والثبات للاختبار ومعاملات الصعوبة والتمييز لفقرات الاختبار وكذلك أثرها في تحصيل الطلبة، ولتحقيق هذه الأهداف تم بناء اختبار تحصيلي لمساق القياس والتقويم، يتكون من (٢٥) فقرة صواب وخطأ متعدد لكل منها أربعة بدائل، وطبق الاختبار على (١٢٢) طالب وطالبة في جامعة إربد الأهلية، وتم تصحيح الاختبار باستخدام أربع طرق مختلفة للتصحيح متدرجة في مراعاتها للمعرفة الجزئية، وباستخدام برمجة (SPSS) تم حساب معاملي صدق للاختبار(الأول: معامل الارتباط بين درجات الطلبة على الاختبار حسب طرق التصحيح المختلفة والدرجات على الاختبار الأول لمساق القياس والتقويم والثاني: معامل الارتباط بين درجات الأفراد على الاختبار حسب طرق التصحيح المختلفة والمعدل التراكمي للطلبة) كما تم حساب معامل الثبات بطريقة كرونباخ ألفا، وكذلك حسبت معاملات

الصعوبة والتمييز لفقرات الاختبار، ومتوسطات التحصيل ، ولكل طريقة من طرق التصحيح المختلفة، وأظهرت نتائج الدراسة عدم وجود فروق ذات دلالة إحصائية بين معاملات الصدق والثبات والتمييز، في حين أشارت النتائج بوجود فروق ذات دلالة إحصائية في معاملات صعوبة الفقرات وفي تحصيل الطلبة؛ حيث كانت طريقة التصحيح الرابعة – الأكثر مراعاة للمعرفة الجزئية- هي الأكثر فاعلية بالنسبة لاختبارات التحصيل محكية المرجع- وتلتها طريقة التصحيح الثالثة ثم الثانية فالأولى. وهذا الترتيب يظهر أن طريقة التصحيح الرابعة هي الأسهل حيث تُعطي العلامة الأعلى.

وهدفت دراسة الصمادي، وسوالمة(٢٠٠٩) إلى الكشف عن أثر طريقة تصحيح فقرات الصواب والخطأ المتعدد في دقة تقدير قدرات الأفراد وصعوبة الفقرات باستخدام نموذج التقدير الجزئي ولتحقيق ذلك، تم بناء اختبار تحصيلي يتكون من ٣٥ فقرة صواب وخطأ متعدد في وحدة الإحصاء المقررة لطلبة الصف الأول الثانوي في الأردن ، وطبق الاختبار على ٨٧٣ طالباً وطالبة بعد دراستهم للوحدة في عشر مدارس اختيرت عشوائياً من مدارس مديرية تربية إربد الأولى. وقد صححت أوراق الإجابة باستخدام أربع طرق للتصحيح متدرجة في مراعاتها للمعرفة الجزئية. وباستخدام برمجية BigSteps تم الحصول على إحصاءات المطابقة وتقديرات لقدرات الأفراد وصعوبة الفقرات والأخطاء المعيارية في تقديرها بحسب طريقة التصحيح بعد استبعاد الأفراد والمفردات غير المطابقين لنموذج التقدير الجزئي. أظهرت النتائج أن طريقة التصحيح الرابعة – الأكثر مراعاة للمعرفة الجزئية – أكثر دقة في تقدير قدرات الأفراد وصعوبة الفقرات ، تلتها طريقة التصحيح الثالثة ، ثم الثانية ، فالأولى.

وهدفت دراسة دودلي Dudley (٢٠٠٦) إلى المقارنة بين فقرات الاختيار من متعدد، وفقرات الصواب والخطأ المتعدد في اللغة الثانية (الانجليزية العامة) في كل من المفردات ، والقراءة من حيث الاستقلالية ، والثبات ، والصدق التلازمي ، ولتحقيق غرض الدراسة تم تطبيق اختبارين لتحديد المستوى في موضوعات المفردات والقراءة على عينة من طلبة السنة الأولى في إحدى الجامعات اليابانية الحكومية بلغ حجمها (١٤٣) طالباً وطالبة منهم (٧٨) طالباً و (٦٥) طالبة ، وأظهرت النتائج أن متوسط أداء الطلبة في اختبارات الصواب والخطأ المتعدد أفضل من فقرات الاختيار من متعدد ، كما بينت النتائج أن كلا النوعين من الفقرات حقق الاستقلالية في القياس ، وبينت النتائج أيضاً عدم وجود فروق جوهرية في معاملات ثبات الأفراد والمفردات ، إلا أن قيم معاملات الثبات كانت أعلى في فقرات الصواب والخطأ المتعدد .

وهدفت دراسة سوالمة وقواسمة (٢٠٠٠) إلى بحث أثر توزيع عدد البدائل الصحيحة والخاطئة في اختبارات الصواب والخطأ المتعدد في خصائصها السيكومترية، وتألفت عينة الدراسة من (٩٣) طالباً وطالبة في مساق التقويم في التربية الابتدائية و(١٤٥) طالباً وطالبة في مساق مقدمة في البحث والإحصاء التربوي، واشتملت أدوات الدراسة في كل مساق على اختبارين من إعداد مدرس المساق: تكون الأول من ٢٠ فقرة اختيار من متعدد، وتكون الثاني من ٢٠ فقرة صواب - خطأ متعدد.

وتم بناء ثلاثة نماذج للاختبار الثاني، وتختلف النماذج الثلاثة فقط في توزيع عدد البدائل الصحيحة والخاطئة، فجميع فقرات النموذج الأول تنتمي إلى فئة واحدة من حيث البدائل الصحيحة، حيث تحتوي كل فقرة بديلين صحيحين من بين أربعة بدائل، وتنتمي فقرات النموذج الثاني إلى فئتين من الفقرات، حيث تحتوي كل فقرة من الفئة الأولى بديلاً واحدً صحيحاً وثلاثة بدائل خاطئة، وتحتوي كل فقرة من الفئة الثانية ثلاثة بدائل صحيحة وبديلاً واحداً خاطئاً. أما فقرات النموذج الثالث فتنتمي إلى ثلاث فئات من حيث عدد البدائل الصحيحة، تحتوي كل فقرة من الفئة الأولى بديلاً واحداً وتحتوي كل فقرة من الفئة الثانية بديلين صحيحين، وتحتوي كل فقرة من الفئة الثالثة ثلاثة بدائل صحيحة. وقد وزعت النماذج الثلاثة بالإضافة إلى اختبار الاختيار من متعدد عشوائياً على الطلبة بحيث طبق كل نموذج على ثلث طلبة المساق تقريباً. وقد حسبت مؤشرات الثبات والصدق التلازمي في النماذج الثلاثة، وأشارت النتائج وجود

اختلافات في قيم الثبات، والصدق التلازمي لصالح النموذج الذي يحتوي على أكبر عدد من الفئات لفقرات الصواب أو الخطأ المتعدد من حيث البدائل الصحيحة.

وهدفت دراسة سوالمة (١٩٩٤) إلى مقارنة فقرات الصواب والخطأ المتعدد (ص خ م) مع فقرات الاختيار من متعدد (خ م) التي توازيها في المحتوى وعدد البدائل من حيث الثبات والصدق التلازمي والصعوبة. وأخذت الدراسة بالاعتبار عدد البدائل الصحيحة في فقرة الصواب والخطأ المتعدد، وتألفت عينة الدراسة من (٦٥) طالباً وطالبة هم مجموع الطلبة في شعبتين لمساق مقدمة في القياس والتقويم خلال الفصل الصيفي لعام ١٩٩٢ في جامعة اليرموك. اشتملت أداة الدراسة على اختبارين متوازيين في المحتوى، يتكون الأول من ٢٥ فقرة من نوع الاختيار من متعدد، ويتكون الثاني من ٢٥ فقرة من نوع الصواب والخطأ المتعدد، وقد طبق الاختباران على أفراد الدراسة بمقتضى خطة تتفق مع غرض الدراسة، وللدلالة على صعوبة الاختبارين حسبت نسبة الصعوبة النسبية (RDR)، وللدلالة على التجانس الداخلي للاختبارين حسبت قيم α وللدلالة على الصدق التلازمي حسبت معاملات الارتباط بين علامات الاختبارين وصححت المعاملات المحسوبة لأثر الانخفاض. أشارت نتائج الدراسة إلى تمتع اختبارات الصواب والخطأ المتعدد بفعالية عالية في قياس تحصيل الطلبة. تقيس هذه الاختبارات نفس المهارات والقدرات التي تقيسها اختبارات الاختيار من متعدد الموازية لها في المحتوى، وتميل العلامات على اختبارات الصواب والخطأ المتعدد لأن تكون أعلى ثباتاً من العلامات على اختبارات الاختيار من متعدد ، ويكون لفرق الثبات دلالته الإحصائية، وبشكل خاص عندما يكون هناك أكثر من بديل صحيح في فقرة الصواب والخطأ المتعدد.

من خلال نتائج الدراسات السابقة يتبين أن هناك تباين في الخصائص السيكومترية لفقرات الصواب والخطأ المتعدد بالمقارنة مع فقرات الاختيار من متعدد، كما في دراسة سوالمة(١٩٩٤) كما بينت نتائج بعض الدراسات (سوالمة، وقواسمة ٢٠٠٠ ؛ سوالمة ، ١٩٩٤) أن هناك تباين في الخصائص السيكومترية يرجع لعدد البدائل الصحيحة في الفقرة ، وتجدر الإشارة إلى أن غالبية تلك الدراسات اعتمدت على النظرية التقليدية في القياس في تحليل النتائج باستثناء دراسة الصمادي (٢٠١٢) التي اعتمدت نظرية الاستجابة للمفردة Item Response Theory (I.R.T) والتي تعد ثورة في القياس النفسي والتربوي ، حيث ساعدت في التخلص من أوجه القصور التي اكتنفت النظرية التقليدية في القياس ، وتفترض نظرية الاستجابة للمفردة أنه يمكن التنبؤ بأداء الأفراد أو تفسير أدائهم في اختبار نفسي، أو تربوي في ضوء خصائص تميز هذا الأداء تسمى سمات، أو قدرات، وهذه السمات غير ملاحظة، ولكن يمكن قياسها من خلال أداء الأفراد على مفردات الاختبار. كما تسمى هذه النظرية أيضا بنظرية المنحنى المميز للمفردة Item Characteristic Curve Theory (Suen, 1990) حيث يستخدم المنحنى المميز للمفردة في عمل استقراءات حول السمة الكامنة من خلال الاستجابات الملاحظة، حيث يمثل هذا المنحنى العلاقة بين احتمال الاستجابة الصحيحة لمفردة من مفردات الاختبار، والسمة أو القدرة موضوع القياس(Kline, 1993). وتوفر نظرية الاستجابة للمفردة مؤشرات إحصائية ثابتة للاختبار، ومفرداته لا تختلف باختلاف عينة الأفراد، مما يوفر قياساً أكثر ثباتاً مع أخطاء أقل للقياس (Nitko, 2001).

ولقد انبثق عن نظرية الاستجابة للمفردة مجموعة من النماذج التي استخدمت في بناء وتطوير المقاييس والاختبارات التي يمكن من خلالها الحصول على مؤشرات إحصائية للمفردة لا تعتمد على خصائص المفحوصين وتقديراتهم، ولا تعتمد على صعوبة مفردات المقياس، ومن هنا أعطيت أهمية اللاتغير(Invariance) في تقدير معالم المفردات بين مختلف مجموعات الأفراد، ووصفت هذه الأهمية بأنها أكبر الصفات المهمة في نظرية الاستجابة للمفردة (Lord, 1980).

وقدُطورت عبر السنوات الماضية مجموعة من نماذج نظرية الاستجابة للمفردة وسميت بأسماء وظيفتها المقترحة مثل: مقاييس التقدير أو التقدير الجزئي، وتهدف هذه النماذج إلي تحديد العلاقة بين أداء الفرد

على مفردات الاختبار وبين السمات أو القدرات الكامنة وراء هذا الأداء وتفسيره ، ولكل منها معادلة رياضية تحدد العلاقة بين أداء الفرد والقدرة أو السمة الكامنة وراء هذا الأداء وتفسيره.

وُيعد نموذج راش الأحادى المعلم أبسط نماذج الاستجابة للمفردة وأكثرها استخداماً في بناء الاختبارات وبنوك الأسئلة (Bachkman,1997) ، واستخدم نموذج راش قبل تطويره في تحليل مفردات الاختبارات التي تعطى فيها درجة خام واحدة عن كل استجابة صحيحة على الفقرة ودرجة خام صفر عن كل استجابة خاطئة، وطور من هذا النموذج نماذج متعددة ليلائم كل منها نوعاً خاصاً من البيانات مثل :

نموذج الاستجابة المتدرجة Graded Response Model، ونموذج سلم التقدير Rating Scale Model، ونموذج التقدير الجزئى Partical Credit Model

ويعد نموذج التقدير الجزئي الذي قدمه ماسترز (Masters, 1982) كتعميم لنموذج "راش" أحادي البارامتر، حيث تتحول مفرداته من التصحيح الثنائي إلى مفردات متعددة التصحيح، وتصنف الاستجابة على المفردات في فئات مرتبة بدرجات متتابعة من 0,1,2,3 ..الخ. وقد افترض ماسترز Masters (1982) درجة لكل مهمة أو خطوة ينجزها الفرد بنجاح.

ويمكن أن يعبر عن احتمالية أن يجيب المفحوص ذو مستوى قدرة معين على فئة معينة من الدرجات بالمعادلة التالية:

$$P_{xi}(\theta) = \frac{exp\left[\sum_{j=0}^{xi}(\theta - b_{xi})\right]}{\sum_{k=0}^{m} exp\left[\sum_{j=0}^{k}(\theta - b_{xi})\right]}$$

(Hambleton, 2004:181)

وهذه المعادلة هي الشكل العام للحصول على المنحنى المميز لدرجة أي فئة في نموذج التقدير الجزئي، والرمز b_{xi} هو بارامتر صعوبة الخطوة المتزامن مع درجة الفئة x_i وθ هي مستوى السمة أو السمة الكامنة، ففي سياق حل المشكلات – على سبيل المثال – فإن θ هنا مرتبطة بمستوى قدرة المفحوص. وعلى هذا فهناك b واحدة متزامنة مع كل خطوة أنجزت للاستجابة على الفقرة، ولذلك أطلق Master (1982) على البارامتر b صعوبة الخطوة، وليس صعوبة الفقرة، لأنها تشير إلى مدى إتمام المفحوص لكل خطوة بنجاح في سبيل حل المشكلة. فالخطوة هي مرحلة مطلوبة لإتمام الفقرة، فعلى سبيل المثال $(\{٣/٦\}+٢)^{٢}$ تحتوي على ثلاث خطوات، لأن هناك ثلاث مراحل مختلفة يجب أن تنجز بترتيب معين للإجابة على المسألة بصورة صحيحة. (DeAyala, Dodd & Koch, 1992).

يذكر أن جميع المفردات يجب أن تكون ذات تمييز متساو مثل جميع نماذج "راش" (كما هي في جميع الخطوات داخل كل مفردة)، لذا لا يوجد بارامتر تمييز الفقرة في هذا النموذج، ويتطلب نموذج التقدير الجزئي إنجاز الخطوات داخل كل مفردة بالتتابع (بنظام ثابت أو بترتيب معين)، بيد أن هذا لا يعني أن الخطوات متساوية الصعوبة، كما أنه ليس بالضرورة أن تكون آخر مفردة هي الأصعب في حلها، أو أن الخطوة الأولى هي الأسهل. فمدى صعوبة المفردات يتفاوت وليس له ترتيب محدد كما هو الحال مع نموذج الاستجابة المتدرجة. وإذا احتوت الفقرة على فئتين فقط تحول النموذج إلى نموذج "راش" (Dodd & Koch,1985)

وقد تم استخدام نموذج التقدير الجزئي في العديد من الدراسات العربية والأجنبية (التقي، ١٩٩٢؛ شاهين، ١٩٩٣ ؛ الشافعي، ونور الدين، ٢٠٠٧ ؛ الزعبي، ٢٠٠٥ ؛ الشريفين، وبني عطا ٢٠١٣ ؛ Beretvas,

2001) ومن مراجعة نتائج هذه الدراسات التي استخدمت نموذج التقدير الجزئي في تحليل البيانات تبين ما يلي:

١- أن نموذج التقدير الجزئي يقدم معلومات أكثر تفصيلاً من تلك التي يقدمها نموذج راش ومقياس جتمان وذلك بالنسبة لعدم الملاءمة و استقلالية القياس.

٢- فاعلية النموذج في معايرة الفقرات وتقديرات القدرة للأفراد وأوصى الباحثون باستخدامه عند تقنين الاختبارات التحصيلية متعددة الخطوات

٣- تزداد الكفاءة النسبية للاختبار بزيادة عدد خطوات الحل

٤- تختلف المعلومات التي تعطيها الأسئلة تبعاً لاختلاف الفاصل بين الخطوة الأولى والأخيرة فتكون المعلومات في حدها الأدنى عندما يكون الفاصل كبير ، بينما تكون في حدها الأقصى عندما يكون الفاصل صغير

٥- دالة المعلومات تتأثر بترتيب خطوات الحل فتكون أكثر تفلطحاً إذا كانت الخطوات مرتبة من الأسهل إلى الأصعب، وتكون أكثر حدة إذا كانت مرتبة من الأصعب إلى الأسهل

مشكلة الدراسة

لقد حظي استعمال فقرات الصواب والخطأ المتعدد باهتمام الباحثين ، واتسع مجال استخدامها، فقد أشارت نتائج مجموعة من الدراسات (Frisbie ,1992; Sweeney, 1982 , سوالمة ، ١٩٩٤) إلى توفر الأدلة التي تؤيد استعمال فقرات الصواب والخطأ المتعدد على نطاق واسع، فقد أشارت نتائج تلك الدراسات إلى الفعالية العالية لفقرات الصواب والخطأ المتعدد في قياس تحصيل الطلبة، كما أشارت النتائج أيضاً إلى تمتع هذا النوع من الفقرات بخصائص سيكومترية عالية.

إن عدد البدائل الصحيحة من الأمور التي يجب أن تؤخذ بعين الاعتبار عند بناء فقرات الصواب والخطأ المتعدد ، حيث يبدو أنه من العوامل الهامة التي تلزم للحصول على فقرات ذات خصائص سيكومترية جيدة. كما أن التوزيع الاحتمالي للإجابات العشوائية على فقرة الصواب والخطأ المتعدد لا يختلف نظرياً باختلاف عدد البدائل الصحيحة فيها ، إلا انه من الناحية العملية قد يكون هناك اختلاف في التوزيع الاحتمالي للإجابات، وفي ضوء ذلك كان لابد من مقارنة عدة صور لاختبار الصواب والخطأ المتعدد مختلفة العدد في بدائلها الصحيحة لمعرفة الأنسب منها للحصول على أفضل خصائص سيكومترية . ونظراً لما تتمتع به نظرية الاستجابة للفقرة من مميزات عديدة أبرزها عدم تغير معالم الفقرة عند تغيير العينة(Lord,1980 ; Embretson & Reise , 2000) فإن هناك حاجة لمعرفة عدد البدائل الصحيحة الأمثل لمعايرة فقرات الصواب والخطأ المتعدد وفق نماذج نظرية الاستجابة للفقرة ، ونظراً لأن فقرة الصواب والخطأ المتعدد هي فقرة عنقودية تتكون من عدة فقرات من نوع الصواب والخطأ التي تشترك في نفس المتن فإن نموذج التقدير الجزئي هو الأنسب للتطبيق بين نماذج نظرية الاستجابة للفقرة.

أسئلــة الدراســـة:

١- هل تختلف دقة تقدير صعوبة الفقرة (مقدرة بالخطأ المعياري في تقديرها) باختلاف عدد البدائل الصحيحة للفقرة؟

٢- هل تختلف تقديرات معلمة القدرة للأفراد باختلاف عدد البدائل الصحيحة للفقرة؟

٣- هل يختلف معامل الثبات لاختبار الصواب والخطأ المتعدد باختلاف عدد البدائل الصحيحة للفقرة؟

أهميـــــة الدراســـة:

- تعتبر الدراسة الحالية استكمالاً لما توصلت إليه الدراسات السابقة واستجابة لبعض توصياتها، للإسهام في تحديد معالم أخرى تتعلق بفقرات الصواب والخطأ المتعدد وقواعد بنائها

- قد توفر نتائج الدراسة معلومات امبريقية لمستخدمي هذا النمط من الفقرات حول العدد الأمثل للبدائل الصحيحة في الفقرة

- الإسهام العلمي المتواضع في التراكم المعرفي في هذا المجال حيث ستسهم نتائج الدراسة في قيام بعض الباحثين بالمزيد من الدراسات المشابهة على نطاق واسع .

مصطلحات الدراســـة:

نموذج التقدير الجزئي: أحد نماذج نظرية الاستجابة للفقرة ، ويعتبر امتداداً لنموذج راش أحادي المعلم ويناسب الأسئلة التي تتطلب خطوات متعددة Polytomous مثل: حل المسائل الحسابية التي تتطلب تعيين درجات جزئية لكل خطوة من خطوات الحل.

فقرة الصواب – خطأ المتعدد: هي فقرة عنقودية تتكون من متن ومجموعة من البدائل التي تكملها بصورة مستقلة وكل بديل عبارة عن فقرة صواب – خطأ .

محددات الدراســـة:

١- اقتصرت الدراسة على عينة من طلبة الصف الثامن الأساسي المسجلين في الفصل الأول من العام الدراسي ٢٠١٤ - ٢٠١٥ بمدارس الحكومة بمديرية غرب غزة .

٢- اقتصرت الدراسة على الوحدة الدراسية الأولى من كتاب الرياضيات للصف الثامن الأساسي.

الطريقـــــة والإجـــراءات:

مجتمع الدراســـة:

تكون مجتمع الدراسة من جميع طلاب وطالبات الصف الثامن الأساسي في المدارس الحكومية التابعة لمديرية التربية والتعليم – غرب غزة ، خلال العام الدراسي ٢٠١٣-٢٠١٤ والبالغ عددهم (٥٤٧٨) طالباً وطالبة منهم (٢٦٨٢) طالباً ، (٢٧٩٦) طالبة وفق التقرير الإحصائي التربوي السنوي في المديرية خلال العام الدراسي ٢٠١٤-٢٠١٥.

عينـــــة الدراســـــة:

تم حصر المدارس ، واختيرت عينة من المدارس بطريقة عشوائية بلغ عددها (٣٨) مدرسة منها (٢٠) مدرسة إناث و (١٨) مدرسة للذكور، ومن ثم جرى اختيار من كل مدرسة وبطريقة عشوائية عنقودية شعبتين وبلغ عدد الشعب الإجمالي (٧٦) شعبة منها (٤٠) شعبة إناث، و (٣٦) شعبة للذكور، ويوفر هذا الاختيار العشوائي ضمان عدم التحيز، وبذلك بلغ عدد أفراد عينة الدراسة (٢٧٨٠) طالباً وطالبة.

أداة الدراســــــــة:

من أجل تحقيق أهداف الدراسة، تم إعداد اختبار من (٢٥) فقرة صواب وخطأ متعدد تقيس تحصيل طلبة الصف الثامن الأساسي في الوحدة الأولى في مبحث الرياضيات، حيث تم الاسترشاد بالأسس العامة التي أوردها غرونلند ولين (Gronlund & linn,1990) في تصميم اختبارات التحصيل الصفي، إذ تم تحليل محتوى الوحدة الأولى من كتاب الرياضيات للصف الثامن الأساسي، وتم صياغة الأهداف السلوكية للوحدة، وإعداد جدول المواصفات، وكتابة (٣٥) فقرة من نوع الصواب والخطأ المتعدد جميعها بأربعة بدائل لقياس المستويات المعرفية الواردة في جدول المواصفات .

صـــــدق الأداة:

بعد الانتهاء من كتابة فقرات الاختبار تم عرض الاختبار بصورته الأولية (٣٥) فقرة على (٨) محكمين من ذوي الاختصاص من معلمي ومشرفي مبحث الرياضيات في مديرية التربية والتعليم، وطلب إليهم إبداء الرأي حول وضوح الفقرات ومدى ملاءمتها لقياس النتاجات التعليمية واقتراح ما يرونه مناسباً، وفي ضوء آراء المحكمين تم إجراء التعديلات اللازمة على فقرات الاختبار، وحذف بعض الفقرات التي أجمع المحكمين على أنها فقرات تؤدي إلى الهدف نفسه وقد بلغ عددها (٥) فقرات واستقر عدد فقرات الاختبار نتيجة لعملية التحكيم على (٣٠) فقرة، وتم تطبيق الاختبار على عينة استطلاعية مكونة من (٦٠) طالباً وطالبة من طلبة مدرسة عباد الرحمن الخاصة موزعين على شعبتين، شعبة ذكورة وشعبة إناث، وتم إبلاغ الطلبة من خلال معلميهم بموعد الاختبار، والمادة التي سيطبق عليها الاختبار، ومن خلال التجريب الأولى للاختبار تم تحديد الزمن المناسب للإجابة عن فقرات الاختبار بواقع (٤٥) دقيقة، وتم حساب معامل ثبات الاتساق الداخلي له بطريقة كرونباخ ألفا حيث بلغ (0.87) وهي قيمة عالية تدل على ثبات الاختبار.

الإجـــــــراءات:

بعد الخروج بالصورة النهائية للاختبار تم إعداد نماذج الاختبار الثلاثة:

نموذج الاختبار الأول: والذي تحوى كل فقرة من فقراته ثلاثة بدائل صحيحة، أما النموذج الثاني فتحوى كل فقرة من فقراته بديلين صحيحين بينما تحوى كل فقرة من فقرات النموذج الثالث على بديل واحد صحيح، وتم إعداد ورقة تعليمات الاختبار، وكذلك ورقة الإجابة النموذجية لكل نموذج من نماذج الاختبار، كما تم تعريف الطلبة بفقرات الصواب والخطأ المتعددة وتدريبهم على كيفية الإجابة عنها قبل موعد تطبيق الاختبار، إذ لم يسبق لهم التعرف عليها.

تم تطبيق نماذج الاختبار عشوائياً على أفراد عينة الدراسة والبالغ عددهم (٢٧٨٠) طالباً وطالبة بواقع نموذج لكل طالب، وبعد تطبيق نماذج الاختبار على أفراد عينة الدراسة، تم فرز كل نموذج على حدة وتصحيحه حسب عدد الإجابات الصحيحة للبدائل على النحو التالي:

صفر في حالة جميع الإجابات خطأ، (٠,٢٥) في حالة إجابة واحدة صحيحة، (٠,٥٠) في حالة إجابتان صحيحتان ، (٠,٧٥) في حالة ثلاث إجابات صحيحة، (١) في حالة جميع الإجابات صحيحة.

تم تحديد مستوى الأداء لكل طالب على كل فقرة من فقرات الاختبار في النماذج المختلفة بدلالة عدد الاستجابات الصحيحة له في تلك الفقرة والتي تعبر عن عدد المهمات الجزئية المنجزة لكل فقرة (٠ ، ١ ، ٢ ، ٣ ، ٤) ، وللتخلص من الكسور في العلامة على الفقرة فقد تم ضرب العلامة في (٤) وبالتالي تصبح العلامة المستحقة على الفقرة حسب طريقة التصحيح كما يلي:

(صفر) في حالة جميع الإجابات خطأ، (١) في حالة إجابة واحدة صحيحة، (٢) في حالة إجابتان صحيحتان، (٣) في حالة ثلاث إجابات صحيحة، (٤) في حالة جميع الإجابات صحيحة.

التحقق من افتراضات النموذج:

تم التحقق من افتراض أحادية البعد من خلال إجراء التحليل العاملي للبيانات لكل نموذج من نماذج الاختبار الثلاثة وتم استخدام طريقة المكونات الأساسية Principle Component Analysis وتم حساب قيم الجذر الكامن Eigen Values ونسب التباين المفسر Explained Variance للعامل الأول والثاني وناتج قسمة الجذر الأول على العامل الثاني في نماذج الاختبار الثلاثة كما في جدول ١

جدول ١

قيم الجذر الكامن ونسب التباين المفسر للعامل الأول والثاني وناتج قسمة قيمة الجذر الأول على العامل الثاني في النماذج الثلاثة.

نموذج الاختبار		العامل الأول	العامل الثاني	ناتج القسمة
الأول	الجذر الكامن	٦,٤٣١	٢,٠٢١	٣,١٨٢
	التباين المفسر	٢١,٣٤٢	٦,٤٢٣	
الثاني	الجذر الكامن	٧,٢٣٢	٢,٣٦٤	٣,٠٥٩
	التباين المفسر	٢٠,٨٤١	٥,٣٢٤	
الثالث	الجذر الكامن	٦,٣٣٤	٢,٣١٢	٢,٧٣٩
	التباين المفسر	٢٢١,٢١١	٤,٥٨١	

بالنظر إلى جدول ١ يمكن القول بأن افتراض أحادية البعد قد تحقق في نماذج الاختبار الثلاثة حيث كانت نسبة التباين المفسر من العامل الأول أكبر من (٢٠) (Hatties,1985)، كما أن مؤشر لورد (Lord,1980) يشير إلى تحقق افتراض أحادية البعد إذا كانت نسبة الجذر الكامن للعامل الأول إلى نسبة الجذر الكامن للعامل الثاني أكبر من (٢) وهذا ما يبدوا واضحاً حيث كانت نسبة الجذر الكامل للعامل الأول إلى نسبة الجذر الكامن للعامل الثاني أكبر من (٢) في جميع نماذج الاختبار.

أما فيما يتعلق بافتراض الاستقلال الموضعي Local Independence فقد تم التحقق منه من خلال حساب الاحصائي (Q3) وهو معامل الارتباط للبواقي لزوج من الفقرات ، وقد كانت قيمته (٠,٠٠٤) وكانت جميع معاملات الارتباط للبواقي بين أزواج الفقرات صغيرة جداً (قريبة من الصفر) .

كما أشار هامبلتون وسامنيثان (Hambleton & Swaminathan, 1985) إلى أن افتراض الاستقلال الموضعي يكافئ افتراض أحادية البعد، وهذا يعني أنه إذا تحقق افتراض أحادية البعد في المقياس، فإن المقياس يحقق افتراض الاستقلال الموضعي.

كما تم حساب معامل الارتباط بين درجات الأفراد على الفقرة والعلامة الكلية على الاختبار في كل نموذج من نماذج الاختبار، الثلاثة كما في جدول ٢

جدول ٢

معاملات الارتباط بين العلامة على الفقرة والعلامة الكلية على الاختبار

رقم الفقرة	النموذج		
	١	٢	٣
١	٠,٣٣	٠,٣٥	٠,٣٢
٢	٠,٤٢	٠,٥١	٠,٣١
٣	٠,٤٧	٠,٤٩	٠,٤٢
٤	٠,٤١	٠,٤٠	٠,٤٤
٥	٠,٣٣	٠,٢٥	٠,٢٩
٦	٠,٤٦	٠,٤٣	٠,٣٩
٧	٠,٤٦	٠,٤٥	٠,٤٢
٨	٠,٤٧	٠,٥١	٠,٤٩
٩	٠,٣٢	٠,٣٤	٠,٣٦
١٠	٠,٤٢	٠,٤٥	٠,٤٠
١١	٠,٥٠	٠,٥٢	٠,٥٣
١٢	٠,٤٤	٠,٤٣	٠,٤١
١٣	٠,٣٩	٠,٤٦	٠,٤٨
١٤	٠,٣١	٠,٤١	٠,٤٣
١٥	٠,٤٨	٠,٥٠	٠,٥٢
١٦	٠,٣٤	٠,٣٨	٠,٣٢
١٧	٠,٣٦	٠,٣٢	٠,٣٤
١٨	٠,٣٨	٠,٣٩	٠,٣٦
١٩	٠,٤٦	٠,٣٨	٠,٣٣
٢٠	٠,٤٤	٠,٣٧	٠,٣
٢١	٠,٤٦	٠,٤١	٠,٤٥
٢٢	٠,٤٨	٠,٤٩	٠,٤٦
٢٣	٠,٤٢	٠,٤٥	٠,٤٧
٢٤	٠,٥٢	٠,٥٤	٠,٥٠
٢٥	٠,٤٢	٠,٤٦	٠,٤٣
٢٦	٠,٤٤	٠,٤٧	٠,٤٢
٢٧	٠,٥١	٠,٥٣	٠,٤٩
٢٨	٠,٤٦	٠,٤٢	٠,٤٤
٢٩	٠,٤٨	٠,٤٦	٠,٤١
٣٠	٠,٥٠	٠,٤٩	٠,٥٣

يتضح من جدول٢ أن جميع معاملات الارتباط ذات دلالة إحصائية عند مستوى دلالة (٠,٠٥) كما أن معظم الفقرات كانت متكافئة في القدرة التمييزية حيث أن معظم معاملات التمييز للفقرات كان متوسطة.

أما فيما يتعلق بافتراض السرعة فقد لوحظ أن ما نسبته (٩٧%) من الطلبة قد أنهوا الإجابة عن الاختبار في جميع النماذج ضمن الوقت المخصص، وبالتالي لم يكن لعامل السرعة أي تأثير في محاولة الإجابة عن فقرات الاختبار ما يعني تحقق افتراض السرعة.

مطابقة البيانات للنموذج:

بهدف التحقق من درجة مطابقة الأفراد والفقرات لنموذج التقدير الجزئي، تم استخدام برنامج لتحليل البيانات لكل نموذج من نماذج الاختبار الثلاثة، وأظهرت نتائج التحليل عدم مطابقة Winsteps (١١٠) فرداً من أفراد عينة الدراسة الذين استجابوا على النموذج الأول، (١٢٥) فرداً من أفراد عينة الدراسة الذين استجابوا على النموذج الثاني و (١٣٨) فرداً من أفراد عينة الدراسة الذين استجابوا على النموذج الثالث، بمعنى أن قيمة إحصائي المطابقة الخارجية أو الداخلية المقابلة لتقديرات قدرات هؤلاء الأفراد تقع خارج حدود المطابقة (٠,٧٠– ١,٣)، أي أن أولئك الأفراد قد تجاوزوا الحد المقبول إحصائيا بإجابتهم صواباً عن المفردات التي تزيد في مستوى صعوبتها عن مستوى قدراتهم، أو يفشلون في الإجابة صواباً عن مفردات تتطلب مستوى قدرة أقل من قدراتهم ؛ مما يعنى أنهم قد اعتمدوا على التخمين، أو عدم جديتهم أو صدقهم في الاستجابة (Wright & Linacre,1998) .

كما بينت النتائج عدم مطابقة الفقرات ذات الأرقام (١٩ ، ٧، ٢٢) في النموذج الأول (لكل فقرة من فقراته ثلاث بدائل صحيحة) والفقرات (٧ ، ١٥) في النموذج الثاني (لكل فقرة من فقراته بديلين صحيحين)، وعدم مطابقة الفقرات (٧ ، ١٩، ٢٢، ٢٦) في النموذج الثالث (لكل فقرة من فقراته بديل واحد صحيح ، أي أن قيمة إحصائي المطابقة الخارجية أو الداخلية المقابلة لتقديرات معالم الصعوبة لهذه الفقرات تقع خارج حدود المطابقة الملائمة (٠,٧ – ١,٣). أي أن تلك الفقرات غير مستقلة عن باقي مفردات الاختبار ، أو أنها تقيس متغيرا آخر شديد التشابه مع المتغير المفترض قياسه.

وبعد حذف جميع الفقرات المشتركة في جميع نماذج الاختبار الثلاثة والبالغ عددها (٥) فقرات أصبح عدد فقرات الاختبار في النماذج الثلاثة ٢٥ فقرة . ويلاحظ أن عدد الفقرات غير المطابقة لنموذج التقدير الجزئي كانت أقل في النموذج الثاني عنها في النموذجين الآخرين.

المعالجات الإحصائية

من أجل الإجابة عن أسئلة الدراسة، تم استخدام المعالجات الإحصائية التالية:

- تم إيجاد معاملات الصعوبة للفقرات باستخدام نموذج التقدير الجزئي لكل نموذج من نماذج الاختبار الثلاثة ، كما تم حساب الأوساط الحسابية والانحرافات المعيارية لقيم هذه المعاملات .

- تم حساب تحليل التباين الأحادي للكشف عن الفروق في الأوساط الحسابية لكل من تقديرات القدرة للأفراد وصعوبة المفردات في نماذج الاختبار الثلاثة.

- تم استخدام الإحصائي Ux_i (Feldt , Woodruff & Salih, 1987, P98) حيث يتوزع هذا الاحصائى وفقا لتوزيع كاى تربيع بدرجات حرية (m-1).وذلك لحساب دلالة الفروق في معاملات الثبات.

$$\overline{\mu} = \sum_{0}^{m}\left(1-\hat{\xi}_i\right)^{-1/3} \Big/ m \qquad UX_i = \sum_{0}^{m}\left[\left(1-\hat{\xi}_i\right)^{-1/3} - \overline{\mu}\right]^2 \Big/ \overline{S}^2$$

حيث : ξ_i: معامل الثبات ، $\overline{S^2}$: متوسط التباين ، m : عدد الحالات

النتائـج ومناقشتها

أولاً: النتائج المتعلقة بالسؤال الأول" هل تختلف دقة تقدير صعوبة الفقرة (مقدرة بالخطأ المعياري في تقديرها) باختلاف عدد البدائل الصحيحة للفقرة؟"

للإجابة عن هذا السؤال تم تقدير قيم معاملات الصعوبة للفقرات لكل نموذج من نماذج الاختبار الثلاثة ، وتم أيضاً حساب المتوسطات الحسابية والانحرافات المعيارية لتلك القيم، ويوضح جدول ٣ قيم معاملات الصعوبة والأخطاء المعيارية لكل نموذج من نماذج الاختبار الثلاثة.

جدول ٣

قيم معاملات الصعوبة والأخطاء المعيارية والمتوسطات الحسابية والانحرافات المعيارية لنماذج الاختبار الثلاثة

نموذج الاختبار						
الأول		الثاني		الثالث		
رقم الفقرة	الصعوبة	الخطأ المعياري	الصعوبة	الخطأ المعياري	الصعوبة	الخطأ المعياري
١	١,١٠-	٠,٢٨	١,٢٨-	٠,٢٠	١,١٢-	٠,٢٧
٢	٠,٩٥-	٠,٢٧	١,١٧-	٠,٢١	٠,٩٨-	٠,٢٧
٣	٠,٧١	٠,٣١	٠,٣٩	٠,٢٥	٠,٦٨	٠,٣٠
٤	٠,٢٨-	٠,٢٨	٠,٥٠-	٠,٢٧	٠,١٩	٠,٢٧
٥	٠,٦٢	٠,٣٠	٠,٤٧	٠,٢٤	٠,٥٩	٠,٢٩
٦	٠,٣٢	٠,٢٩	٠,٢٧	٠,٢٨	٠,١٩	٠,٢٨
٧	٠,٤٤	٠,٢٩	٠,٣٠	٠,٢١	٠,٢٨	٠,٢٩
٨	١,١٠-	٠,٢٨	١,٢١-	٠,٢٨	١,١٧-	٠,٢٧
٩	١,٣٠-	٠,٢٨	٠,٨٩-	٠,١١	٠,٩٨-	٠,٢٧
١٠	٠,٣٦	٠,٢٩	٠,٣٠	٠,٢٩	٠,٤٢	٠,٢٩
١١	٠,٨٠	٠,٣١	٠,٧٤	٠,١٤	٠,٦٨	٠,٣١
١٢	٠,٣٦	٠,٢٩	٠,٤٥	٠,١٨	٠,٢٦	٠,٢٨
١٣	١,٤٧-	٠,٢٨	١,٤٠-	٠,١٢	١,٢٠-	٠,٢٧
١٤	٠,٨٢	٠,٣١	٠,٥٨	٠,٢٠	٠,٨٦	٠,٣١
١٥	٠,٨٠-	٠,٢٧	٠,٦٥-	٠,٢١	٠,٤٨-	٠,٢٦
١٦	٠,٤٧	٠,٢٩	١,٩٨	٠,١٩	٠,٢٦	٠,٢٨
١٧	٠,٩٠	٠,٣٢	١,٤٩	٠,٢١	٠,٨٦	٠,٣١
١٨	٠,٨٠	٠,٣١	٠,٧٤	٠,١٠	٠,٨٦	٠,٣١

١٩	٠,٢٠-	٠,٢٨	٠,١٠	٠,١٨	٠,٦٠-	٠,٢٧
٢٠	٠,٧٣-	٠,٢٧	٠,٤٠-	٠,١٧	٠,٢٤-	٠,٢٦
٢١	٠,٩٠	٠,٣٢	٠,٩٦	٠,٢٢	٠,٨٦	٠,٣١
٢٢	١,١١	٠,٣٣	١,١٣	٠,١٣	١,١٦	٠,٣٣
٢٣	٠,٢٧	٠,٢٨	٠,٣٥	٠,١٩	٠,٠٣	٠,٢٧
٢٤	٠,٩٤-	٠,٢٧	١,١٤-	٠,١٧	٠,٧٥-	٠,٢٦
٢٥	٠,١١	٠,٢٨	٠,٢٩	٠,١٩	٠,١٣	٠,٢٧
الوسط الحسابي	٠,٠٠	٠,٢٩	٠,٠٧	٠,١٩	٠,٠٣	٠,٢٨
الانحراف المعياري	٠,٨١	٠,٠٢	٠,٩١	٠,٠٢	٠,٧٤	٠,٠٢

يتضح من جدول ٣: أن هناك تباينا ملحوظا في مدي صعوبة الفقرات ، حيث كان مدى الصعوبة لفقرات النموذج الثاني (عدد البدائل الصحيحة لكل فقرة من فقراته يساوي ٢) هو الأعلى وتراوحت قيمته بين (-١,٤٠ و ١,٩٨) يليه النموذج الأول (تحوى كل فقرة من فقراته على ثلاثة بدائل صحيحة) وتراوح المدى لصعوبة الفقرات بين (-١,٤٧ و ١,١١) في حين كان الأقل في النموذج الثالث(تحوى كل فقرة من فقراته على بديل واحد صحيح) حيث تراوحت قيمته بين (-١,٢٠ و ١,١٦) ، كما أن الوسط الحسابي لمعالم الصعوبة في النماذج الثلاثة يساوي صفر ، إلا أن هناك اختلاف في الأوساط الحسابية للأخطاء المعيارية ولفحص دلالة الفروق في دقة تقدير معالم الصعوبة تم استخدام تحليل التباين الأحادي للأخطاء المعيارية لهذه التقديرات وجدول ٤ يبين ذلك.

جدول ٤

نتائج تحليل التباين الأحادي للأخطاء المعيارية لصعوبة الفقرات حسب نموذج الاختبار

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة f	الدلالة الإحصائية
بين الفقرات	٠,١٣٦	٢	٠,٠٦٨	٥٨,٦١	٠,٠٠
داخل الفقرات	٠,٠٨٣	٧٢	٠,٠٠١		
الكلي	٠,٢١٩	٧٤	٠,٠٠٢٩٦		

يتبين من جدول ٤ : وجود فروق دالة إحصائيا عند مستوى دلالة (٠,٠٥) في الوسط الحسابي للأخطاء المعيارية في تقدير معالم صعوبة الفقرات تعزى لعدد البدائل الصحيحة ، ولمعرفة الفروق بين نماذج الاختبار استخدم اختبار (LSD) للمقارنات الثنائية ، ويبين جدول ٥ المقارنات الثنائية بين أوساط الأخطاء المعيارية لتقديرات معالم الصعوبة حسب عدد البدائل الصحيحة

جدول ٥

نتائج اختبار (LSD) للمقارنات الثنائية بين أوساط الأخطاء المعيارية لتقديرات معالم الصعوبة حسب عدد البدائل الصحيحة .

النموذج	الأول (٠,٢٩١)	الثاني (٠,١٩٨)	الثالث(٠,٢٨٤)
الأول(٠,٢٩١)	-	*٠,٠٩٣٦	٠,٠٠٧
الثاني(٠,١٩٨)	-	-	*٠,٠٨٦٤

يتبين من جدول ٥: أن هناك فروق دالة إحصائيا في الأوساط الحسابية للأخطاء المعيارية الخاصة بصعوبة الفقرات بين النموذج الثاني وكلٍ من النموذجين الأول والثالث لصالح النموذج الثاني ولم تكن هناك فروق دالة إحصائيا بين النموذجين الثاني والثالث. ويعني ذلك أن أخطاء القياس في النموذج الثاني (عدد البدائل الصحيحة يساوي ٢) أقل من أخطاء القياس في النموذج الأول والثالث أي أنه في حالة تساوي عدد البدائل الصحيحة مع عدد البدائل الخطأ فإن تقدير معالم الفقرات يكون أكثر دقة عنه في الحالات الأخرى، وتتفق نتيجة الدراسة الحالية مع ما خلصت إليه دراسة سوالمة ، وقواسمة (٢٠٠٠) حيث اعتبر تساوي عدد البدائل الصحيحة في الاختبار مع عدد البدائل الخطأ أحد الإرشادات الواجب الالتزام بها عند إعداد اختبارات الصواب أو الخطأ المتعدد من أجل الحصول على خصائص سيكومترية مقبولة .

النتائج المتعلقة بالسؤال الثاني" هل تختلف تقديرات معلمة القدرة للأفراد باختلاف عدد البدائل الصحيحة للفقرة؟"

للإجابة عن هذا السؤال تم إيجاد الأوساط الحسابية والانحرافات المعيارية لمعلمة القدرة للأفراد على كل نموذج من نماذج الاختبار الثلاثة وجدول ٦ يبين ذلك

جدول ٦

الأوساط الحسابية والانحرافات المعيارية لمعلمة القدرة للأفراد على نماذج الاختبار الثلاثة

نماذج الاختبار	العدد	الوسط الحسابي	الانحراف المعياري
الأول : عدد البدائل الصحيحة =١	٧٨٩	٠,٨٦٦٩-	١,٢٩٢
الثاني : عدد البدائل الصحيحة =٢	٧٩٦	٠,٥٧٨٠-	١,٣١٢
الثالث : عدد البدائل الصحيحة = ٣	٨٢٢	٠,٧٥٣٥-	١,٢٠٨

يتبين من جدول ٦ وجود فروق رقمية في الأوساط الحسابية لمعلم القدرة للأفراد في النماذج الثلاثة تعزى لعدد البدائل الصحيحة، ولفحص دلالة الفروق في الأوساط الحسابية لمعالم القدرة تم استخدام تحليل التباين الأحادي لقدرات الأفراد على النماذج الثلاثة وجدول ٧ يبين ذلك.

جدول ٧

نتائج تحليل التباين الأحادي للأوساط الحسابية لقدرات الأفراد حسب نموذج الاختبار

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	قيمة f	الدلالة الإحصائية
بين المجموعات	٣٧,٧٢٦	٢	١٨,٨٦٣	١١,٦٧	٠,٠٠
داخل المجموعات	٣٨٨٥,٠٨	٢٤٠٤	١,٦١٦		
الكلي	٣٩٢٢,٨١	٢٤٠٦			

يتبين من جدول ٧ : وجود فروق دالة إحصائيا عند مستوى دلالة (٠,٠٥) في الوسط الحسابي لقدرات الأفراد تعزى لعدد البدائل الصحيحة ، ولمعرفة الفروق بين نماذج الاختبار استخدم اختبار (LSD) للمقارنات الثنائية ، ويبين جدول ٨ المقارنات الثنائية بين أوساط قدرات الأفراد حسب عدد البدائل الصحيحة .

جدول ٨

نتائج اختبار (LSD) للمقارنات الثنائية بين أوساط قدرات الأفراد حسب عدد البدائل الصحيحة .

النموذج	الأول (-٠,٨٦٦٩)	الثاني(-٠,٥٧٨٠)	الثالث (-٠,٧٥٣٥)
الأول (-٠,٨٦٦٩)	-	*٠,٢٨٨٩	٠,٠٥٢٤
الثاني (-٠,٥٧٨٠)	-	-	*٠,٢٣٦٥

يتبين من جدول ٨: أن هناك فروق دالة إحصائيا في الأوساط الحسابية الخاصة بقدرات الأفراد بين النموذج الثاني وكلٍ من النموذجين الأول والثالث لصالح النموذج الثاني ولم تكن هناك فروق دالة إحصائيا بين النموذجين الأول والثالث .

ويعني ذلك أن نموذج الاختبار الثاني (عدد البدائل الصحيحة = ٢) كان الأصعب من بين النماذج الأخرى ، ويعزو الباحث ذلك إلى أنه في النموذج الأول (عدد البدائل الصحيحة = ٣) يستطيع المفحوص الحصول على علامة مرتفعة (٠,٧٥ من علامة الفقرة) من خلال إجابة جميع البدائل للفقرة باعتبارها صحيحة حيث يخطئ المفحوص في إجابة البديل الخطأ فقط ويصيب في إجابة بقية البدائل ، وكذلك الأمر في النموذج الثالث (عدد البدائل الصحيحة = ١) حيث يستطيع المفحوص الحصول على علامة مرتفعة (٠,٧٥ من علامة الفقرة) من خلال إجابة جميع البدائل للفقرة باعتبارها خطأ في هذه الحالة يخطئ المفحوص في إجابة البديل الصحيح فقط ويصيب في إجابة بقية البدائل ، أما في النموذج الثاني(عدد البدائل الصحيحة = ٢) فنسبة الحصول على علامة مرتفعة لا تتجاوز (٠,٥) سواءً أجاب جميع البدائل للفقرة باعتبارها صحيحة أو خطأ.

النتائج المتعلقة بالسؤال الثالث" هل يختلف معامل الثبات لاختبار الصواب والخطأ المتعدد باختلاف عدد البدائل الصحيحة للفقرة؟" للإجابة عن هذا السؤال تم حساب معاملات الثبات لنماذج الاختبار الثلاثة من خلال ما يوفره برنامج الحاسب الآلي WINSTEPS وهو يكافئ معامل ثبات كودر ريتشاردسون-٢٠ (Linacre , 2003) وقد بلغت قيم معاملات الثبات للنماذج الثلاثة (٠,٩١) ، و(٠,٩٤) ، و(٠,٩٢) على الترتيب ولحساب دلالة الفروق في معاملات الثبات تم استخدام الإحصائي UX_i (Feldt , Woodruff

(& Salih, 1987, p98) حيث يتوزع هذا الإحصائى إلى توزيع كاى تربيع بدرجات حرية (m-1) وأشارت النتائج إلى عدم وجود فروق دالة إحصائيا(α= ٠,٠٥) بين معاملات الثبات للنماذج الثلاثة حيث بلغت قيمة الإحصائى Ux_i (٥,٤١) وهي أقل من القيمة الحرجة (٥,٩٩) لتوزيع كاي تربيع عند درجات حرية (df = 2) .

وقد يكون سبب ذلك أن تدرج مفردات المقياس على ميزان تدرج واحد مشترك وفقا لنموذج التقدير الجزئي بعد حذف الحالات غير الملائمة من الأفراد والمفردات يعنى تحقق شرط استقلالية القياس ويعنى ذلك ثبات تقديرات كل من الصعوبة والقدرة وعدم تأثرها باختلاف عدد البدائل الصحيحة . كما أن عدد الفقرات المحذوفة في كل من نماذج الاختبار الثلاثة وفقا لعدم ملاءمتها للنموذج كان متقاربا كما ونوعا.

وبالرغم من عدم ظهور فروق دالة إحصائياً في معاملات الثبات تعزى لعدد البدائل الصحيحة إلا أنه لوحظ أن معامل الثبات للنموذج الثاني (عدد البدائل الصحيحة = ٢) أعلى منه في النموذجين الآخرين ، ويمكن تفسير ذلك إلى أن احتمال حصول الطالب على علامة مرتفعة في غياب القدرة في النموذج الثاني أقل منها في النموذجين الآخرين، كما أن النموذج الثاني هو الأكثر انسجاماً مع الإرشادات الخاصة ببناء اختبارات الصواب والخطأ المتعدد والتي تقتضي أن يكون مجموع الفقرات الصحيحة في الاختبار مساوياً مجموع الفقرات الخطأ وتتفق نتائج الدراسة الحالية مع نتائج دراسة سوالمة (١٩٩٤) .

خلاصة وتوصيات

أشارت نتائج الدراسة إلى وجود فروق في تقديرات صعوبة الفقرات في اختبار الصواب والخطأ المتعدد تعزى لعدد البدائل الصحيحة ، أى أن الاختبار الذي تتساوى فيه عدد البدائل الصحيحة مع عدد البدائل الخطأ أكثر دقة في تقدير قدرات الأفراد ، وأكثر صعوبة ، وأكثر ثباتاً لذا يفضل عند بناء اختبارات الصواب والخطأ المتعدد أن يكون عدد الفقرات الصحيحة مساوياً عدد الفقرات الخطأ ، كما أثبتت الدراسة فعالية نموذج التقدير الجزئي في بناء اختبارات تتصف بالموضوعية في القياس ، كما أمكن للدراسة الحالية من تقديم أدلة إمبريقية عن تأثير عدد البدائل الصحيحة على الخصائص السيكومترية للفقرات، ويوصي الباحث بإجراء دراسة حول أثر عدد البدائل الصحيحة وعدد خطوات حل المسألة على الخصائص السيكومترية للاختبار وفقراته، كما يوصي الباحث باستخدام هذا النوع من الاختبارات لما وجده من ارتياح من قبل الطلبة والمدرسين لهذا النوع من الفقرات.

المراجع

التقي، أحمد محمد عيسى(١٩٩٢). **اللاتغير في معالم قدرات الأفراد ودرجات صعوبة أسئلة المقال من خلال نموذج التقدير الجزئي وسلم التقدير كحالتين خاصتين من نماذج راش**(رسالة دكتوراه غير منشورة). الجامعة الأردنية. الأردن.

ثورندايك ، روبرت، وهيجن، اليازبيت (١٩٨٦). **القياس والتقويم في علم النفس والتربية**. ترجمة عبد الله زيد الكيلاني وعبد الرحمن عدس. مركز الكتب الأردني.

الزعبي ، محمد (٢٠٠٥). فاعلية نموذج التقدير الجزئي في معايرة فقرات متعددة الخطوات . **مجلة جرش للبحوث والدراسات – الأردن**. ١٠(١). ٦٧ - ٨٨

سوالمة ، يوسف (١٩٩٤). اختبارات الصواب والخطأ المتعدد : الخصائص النسبية وعدد البدائل الصحيحة. **أبحاث اليرموك : سلسلة العلوم الإنسانية والاجتماعية**، ١٠ (٢) . ٥٧٣- ٥٩٥.

سوالمة ، يوسف، وقواسمة ، أحمد(٢٠٠٠) . أثر التنوع في عدد البدائل الصحيحة والخاطئة في اختبارات الصواب أو الخطأ المتعدد في خصائصها السيكومترية. **مجلة جامعة دمشق.** ١٦(١). ٦١- ٨٨.

الشافعي ، منصور، ونور الدين، أمين (٢٠٠٧). استخدام نموذج التقدير الجزئي الرياضي اللوغاريتمي في تطوير اختبار القدرات العقلية الأولية على عينة من البيئة السعودية . **مجلة كلية التربية** – الزقازيق. (٥٦) . ٢٤٥-٣٤٦.

شاهين، محمد (١٩٩٣). **أثر عدد خطوات تصحيح الأسئلة المقالية في تقديرات الأفراد وقيم صعوبة الأسئلة باستخدام نموذج التقدير الجزئي(** رسالة دكتوراه غير منشورة). عمان : الجامعة الأردنية.

الشريفين، نضال، وبني عطا، زايد (٢٠١٣). تقصي أثر عدد خطوات الأسئلة متعددة التدريج وشكل التوزيع لصعوبتها على تقديرات القدرة للأفراد ، والصعوبة للأسئلة، ودالة المعلومات للاختبار وفق نموذج التقدير الجزئي. *المجلة التربوية* . ٢٨(١٠٩). ٢١٣-٢٧٥.

الصمادي، مروان (٢٠١٢). أثر طريقة تصحيح اختبار الصواب – الخطأ المتعدد على الخصائص السيكومترية للاختبار وفقراته . **مجلة العلوم التربوية والنفسية** ،١٣(٣) .٢٨٢- ٣٠٦.

الصمادي، مروان ، وسوالمة، يوسف (٢٠٠٩).أثر طريقة تصحيح فقرات الصواب والخطأ المتعدد في دقة تقدير قدرات الأفراد وصعوبة الفقرات باستخدام نموذج التقدير الجزئي. **المجلة الأردنية في العلوم التربوية**. ٥(١). ٧١-٨٧.

عودة، أحمد(٢٠١٠). **القياس والتقويم في العملية التدريسية**، اربد ، الأردن: دار الأمل.

Aiken, J. (1987).Testing with multiple – choice Items: *Journal of Development in education*,(20),44-57.

Akeroyd, F. (1982). Progress in multiple – choice scoring methods . *Journal of Future and Higher education* , (6),87-105

Bachman , L. (1997*). Fundamentals Considerations in Language Testing*,4th edition , London : Oxford University Press.

Beretvas, S. (2000). *To meet or not to meet standards: Proficiency estimation using different polytomous IRT models* . University of Washington. Degree : PHD.

Bush, M. (2001). A multiple choice test that rewards partial knowledge . *Journal of Future and Higher education* , 25(2),157-164.

De Ayala ,R. , Dodd, B. ,& Koch ,W. (1992). An introduction to polytomous item response theory models. *Measurement & Evaluation in counseling & Development*,25(4),172-189.

Dodd, B. & Koch, w. (1985,April). *Item and scale information functions for the partial credit model*. Paper presented at the meeting of the American Educational Research Association, Chicago.

Dudley, A. (2006). Multiple dichotomous scored items in second language testing : Investigating the multiple true – false item type under norm – referenced conditions . *Language Testing* , 23(2),198-228.

Embretson, S. & Reise, S. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Eribaum Associates, Inc., Publishers

Feldt , L.; Woodruff, D.& Salih, F. (1987). Statistical Inference for Coefficient Alpha. *Applied Psychological Measurement,* 11(1), 93-103.

Frisbie, D. (1992). The Multiple True False item formats : A status review. *Educational Measurement* : Issues ND Practice,11(4),21-26.

Frisbie , D. & Sweeney, D. (1982). The relative merits of multiple true false achievement tests. *Journal of Educational Measurement*, (19), 29-35.

Gronlund, N.(1982). *Constructing Achievement Test* (3ed ed). New Jersey: Prentice – Hall.

Gronlund , N. & Linn , R. (1990). *Measurement and Evaluation in teaching* . New York : Macmillan publishing Co., INC.

Hambleton , R. (2004). A Rasch model for partial credit scoring. *Psychometrika* , 55, 170-196 .

Hambleton, R. & Swaminathan, H. (1985). *Item Response Theory "Principles & Applications"*, Kluwer Nijhoff Publishing.

Kline, P. (1993) .*The hand Book of the psychological testing* .London. Rout Ledge.

Linacre , J. (2003). *User's Guide & Program Manual . Winsteps Minstep .Rasch – model Computer Programs*. Retrieved December, 20, 2014,from: Www.winsteps.com.

Lord , F. (1980). *Application of Item response theory to practical testing problems*. Hillsdale ,NJ, Erbaum.

Masters , G. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 2, 149-174.

Nitko , A. (2001).*Educational assessment of student*, (3Ed), New Jersy, Merrill Pentice Hall.

Suen, H. (1990). *Principles Of Test theories.* Hillsdale, New Jersey.

Wright, B. & Linacre, j. (1998). *A User's Guide to Winsteps / Bigsteps* ,Version 2.88, Chicago, U.S.A : MESA Press .