

**المصادر السيكومترية لاختبارات الاختيار من متعدد والتكامل ”دراسة مقارنة باستخدام نظرية الاستجابة للمفردة“**

د. حمدي يونس أبو جراد

قسم علم النفس

كلية التربية -جامعة القدس المفتوحة

## **المصادر السيكولوجية لاختبارات الاختيار من متعدد والتكميل دراسة مقارنة باستخدام نظرية الاستجابة للمفردة\***

د. حمدي يونس أبو جراد  
قسم علم النفس  
كلية التربية -جامعة القدس المفتوحة

### **الملخص**

هدفت هذه الدراسة المقارنة بين المصادر السيكولوجية لاختبار الاختيار من متعدد واختبار التكميل في ضوء نظرية الاستجابة للمفردة ولتحقيق غرض الدراسة تم إعداد اختبارين في مادة العلوم لطلبة الصف السادس الابتدائي أحدهما اختيار من متعدد والأخر من نوع التكميل وكلاهما مكون من (٣٠) مفردة وباستخدام نموذج راش تم تدريب ومعايرة مفردات كلا الاختبارين على عينة مكونة من (٨٠) طالباً وطالبة والحصول على تقديرات الصعوبة مقدرة بوحدة اللوجيت وكشفت نتائج الدراسة عن حذف (٧) مفردات من اختبار التكميل و(٦) مفردات من اختبار الاختيار من متعدد وأشارت النتائج أيضاً إلى أن معاملات ثبات تقديرات الصعوبة للمفردات والأفراد في اختبار التكميل أعلى منها لاختبار الاختيار من متعدد.

**الكلمات المفتاحية:** نظرية الاستجابة للمفردة، اختبارات التكميل، اختبارات الاختيار من متعدد.

## The Psychometrics Tests Properties of Multiple Choice and Completion Test “A comparison Study by Using Item Response Theory”

**Dr. Hamdy Y. Abu Jarad**

College of Education  
AL-Quds Open University

### Abstract

The aim of this study was to compare the characteristics of the psychometric of multiple choice test with the complete test items in the light of the Item Response Theory of Measurement. For achieving the purpose of this study, two tests were prepared for the sixth grade in science and both the tests consisted of (30) items (one test is a multiple choice test and the other is complete test items). Calibration of the items for sample of (80) students and the degree of items difficulty (logit degree ) of both the tests were based on Rasche's model. The study results revealed the necessity of omitting (7) items of the complete test items and omitting (6) items of the multiple choice test. In addition, the results showed that the reliability of the difficulty degree for items and persons in the complete test items were higher than that in the multiple choice test.

**Key words:** item response theory, completion test, multiple choice test.

## **المصادر السيكولوجية لاختبارات الاختيار من متعدد والتكميل**

### **دراسة مقارنة باستخدام نظرية الاستجابة للمفردة**

د. حمدي يونس أبو جراد

قسم علم النفس

كلية التربية -جامعة القدس المفتوحة

#### **المقدمة**

تستعمل الاختبارات بأشكالها المختلفة لقياس تحصيل الطلبة في مختلف المواد الدراسية وتفيد نتائج الاختبارات في معرفة مدى حقوق الأهداف التعليمية، أو في مقارنة الطلبة بعضهم ببعض، أو في تشخيص جوانب القوة والضعف لدى التلاميذ، أو في انتقاء الأفراد للوظائف المختلفة.

وبعد الاختبار موقفاً تعليمياً، إذ يتوقع أن يتضمن الأسئلة التي تقيس الأهداف الأساسية، ومن ثم فهو يغطي أهم أجزاء المحتوى الدراسي، ويتوقع أن يفكر الطالب بعمق في موقف الاختبار مقارنة بالواقف الأخرى، ولذلك خطط الأفكار والمعلومات الواردة في الاختبار بنصيب مميز من التفكير؛ وبالتالي إمكانية الاحتفاظ بالمعلومات.

وتتنوع أشكال المفردات التي يمكن استخدامها في قياس تحصيل الطلبة فمنها ما يتطلب من المفحوص أن ينتقى الإجابة من بين عدة بدائل (المفردات ذات الإجابة المتقدمة) كاختبارات الاختيار من متعدد، والصواب والخطأ، والصواب والخطأ المتعدد، والمزاوجة، ومنها ما يتطلب من المفحوص أن يعطى الإجابة بلغته الخاصة (المفردات ذات الإجابة المصاغة) كاختبارات المقال، والإجابة القصيرة والتكميل.

وتحكم في اختيار نوع معين من المفردات دون الآخر عدّة عوامل منها: طبيعة المادة المتعلمة، ومستوى العبارة الهدافية، والغرض من التقويم، وأعمار الطلبة، ومستوياتهم العقلية، ومدى تأثير المصادر السيكولوجية للاختبار بطريقة التصحيح.

وأشار عودة (1999) إلى أن نوعاً من المفردات لا يختص بمستوى عقلي معين أو بخصائص سيكولوجية معينة كالصدق والثبات إذ من الممكن أن تحصل على اختبار بخصائص مرضية بصرف النظر عن نوع المفردات إذا أحسن إعدادها ولكن هذا لا يعني أن بإمكان أي نوع من المفردات أن يحل محل نوع آخر، بينما أشار أiken (Aiken, 1987) (المشار إليه في Allen & Ten, 1979) إلى أن اختبارات الاختيار من متعدد أصبحت الأكثر انتشاراً بين كل

أشكال الاختبارات في التربية. وما زاد في شيوع وانتشار هذا النوع من المفردات تفوقها على جميع أشكال المفردات الموضوعية الأخرى بسبب كفاءتها وتنوع استخداماتها حتى يمكن بواسطتها قياس أهداف بسيطة وأخرى مركبة في مختلف المواقف الدراسية.

ويعد ثورندايك وهيجن (١٩٨١) أن أسئلة الاختيار من عدة إجابات أكثر أنواع الأسئلة الموضوعية مرنة، إذ تصلح لتقديم التحصيل لأى من الأهداف التعليمية التي يمكن قياسها باختبارات كتابية باستثناء تلك الأهداف التي تتطلب مهارة في التعبير الكتابي أو أصالحة في التفكير. وتقول أنساتازى (Anastasi) إنّ مفردات الاختيار من متعدد أثبتت أنها أكثر أنواع المفردات استخداماً بالإضافة إلى أنها أسهل في التصحيح من أنواع المفردات الأخرى وأن فرص الحصول على الإجابة عن طريق التخمين يمكن تقليله عن طريق إضافة عدد أكبر من البديل (Anastasi, 1982).

وذكر جرونلندر (Gronlund, 1977) أن اختبارات الاختيار من متعدد أصعب من اختبارات الإجابة القصيرة، وعندما رتب أنواع الأسئلة حسب صعوبتها جاءت أسئلة الإجابة القصيرة قبل أسئلة الاختيار من متعدد.

ومن جهة أخرى يرى الكحلوت (٢٠٠٢) أنه على الرغم من شيوع استخدام مفردات الاختيار من متعدد في قياس العديد من الأغراض التربوية إلا أنها لا تزال تواجه نقداً متزايداً فيما يزال عامل التخمين من العوامل التي تهدد صدقها وثباتها ولا يوجد أدلة خبرية كافية تشير إلى أن هذا النوع من المفردات هو أكثر ثباتاً وصدقًا من جميع الاختبارات الموضوعية الأخرى.

بالمقابل فإنّ مفردات التكميل لا تتأثر بالتخمين العشوائي لأنها تتطلب من المفحوص استدعاء الاستجابة وليس انتقاء الإجابة كما أنها سهلة الإعداد ولا تحتاج إلى إجابات مطولة، إلا أن ما يعاب عليها أنها تشجع على الحفظ الصم للمعلومات وأنّها تقيس مستويات عقلية بسيطة إضافة إلى أنها أقل أنواع الأسئلة الموضوعية ثباتاً (مراد وسليمان، ٢٠٠٢).

وتتطلب عملية اتخاذ القرارات تطبيق المقاييس التي يمكن من خلال نتائجها تقدير قدرات الفرد بدقة، ويطلب ذلك بناء أدوات قياس تتصف بالدقة والموضوعية في القياس. وقد اعتمدت الدراسة الحالية على النظرية الحديثة في القياس، أو ما يعرف بنظرية الاستجابة للمفردة باعتبارها تشكل إطاراً للتوجه الحالي والمستقبل المنظور في اختيار المفردات (Anastasi, 1982).

وتتوفر نظرية الاستجابة للمفردة مؤشرات إحصائية ثابتة للاختبار ومفرداته لا تختلف باختلاف عينة الأفراد، مما يوفر قياساً أكثر ثباتاً مع أخطاء أقل للقياس. (Nitko, 2001).

- وتتيح هذه النظرية تكوين صور اختباريه فرعية مختلفة متعادلة القياس مسحوبة من مفردات التدريج الخاص بالقياس الكلي (كاظم، ١٩٩٦).
- وتميز نظرية الاستجابة للمفردة بأربع ميزات :
- ١- تمكننا من تقدير قدرة المفحوص بحيث تكون مستقلة إحصائيا عن عينة المفردات الخاصة التي طبقت على المفحوصين طالما أنها مفردات مناسبة. وذلك بشرط أن تكون جميع المفردات مدرجة معا على نفس المتصل بحيث تقيس نفس السمة.
  - ٢- تمكننا من الحصول على مؤشرات للمفردة مثل (معامل الصعوبة، ومعامل التمييز) تكون مستقلة عن عينة المفحوصين الخاصة المستخدمة في معایرة المفردة طالما أنها عينة مناسبة .
  - ٣- تمكننا من الحصول على معامل إحصائي يبين مدى دقة تقدير قدرة كل فرد بواسطة مفردات الاختبار وهذا المعامل يمكن أن يختلف من مفحوص لآخر إن خُّقيق هذه الميزات الثلاثة يعتمد على مدى الانسجام بين مجموعة بيانات الاختبار والنموذج المستخدم (أى خُّحقق شرطه). وإذا لم يتحقق هذا الانسجام أو خُّحقق بشكل ضعيف فإن هذه الميزات لا يمكن الحصول عليها .
  - ٤- استبدال مفهوم الثبات من خلال الصور المتوازية بمفهوم التقدير الاحصائي والأخطاء العيارية المترافقه (Associated standerd errors) (Hambleton & Swaminthan,) (1985).

وتقوم هذه النظرية على مجموعة من الافتراضات هي:

### **أولاً : أحادية البعد (Unidimensionality)**

يشير اصطلاح الأبعاد (Dimensions) إلى عدد السمات الكامنة المسئولة عن أداء الأفراد لمفردات الاختبار، والسمة مفهوم يستخدم لوصف الأفراد وهي عبارة عن مجموعة من السلوكيات المتراقبة معاً والمترابطة معاً وتغبل للحدوث معاً. ومن ثم فإن السمة مفهوم نظري غير ملموس. لذا يُعَدُّ تعريف السمة من الخطوات الأساسية في القياس السلوكي . وأشار إيلوت (Elliot, 1983): إلى أنه لكي تكون مفردات الاختبار أحادية البعد يجب أن يكون للمفردة معلم واحد فقط (معلم الصعوبة ) يؤدي إلى الفروق بين احتمالات الاستجابة الصواب للأفراد (في نفس مستوى القدرة ) على المفردات المختلفة الصعوبة. أي تختلف المفردات فيما بينها في مستوى صعوبتها فقط .

وترى كاظم (١٩٩٤) أن أحادية البعد تعنى أن مفردات القياس تكون متجانسة فيما

بينها وتقسيس في أساسها نفس الصفة، وهذا يعني أن أي مفردة من هذه المفردات المتدرجة الصعوبة تتطلب في حلها النوع نفسه من الإجراءات والعمليات السلوكية ولكنها تختلف فيما بينها من حيث تدرج صعوبتها فقط.

أما إذا كانت البنود تقسيس أكثر من متغير "قدرة" فيمكن جمجمة مفردات كل فدراة في مجموعة متGANسة باستخدام التحليل العاملـي، ثم استخدام أحد نماذج السمات الكامنة مع كل مجموعة متGANسة من المفردات تدريجـاً. (مراد، وسليمان، ٢٠٠٢).

ويعد هامبلتون وسواميـنـثان (Hambleton & Swaminthan, 1985) أن افتراض أحـاديـة الـبعدـ شيءـ مرغـوبـ فيهـ عندـ جـمـجمـةـ مـطـوريـ الاختـبارـاتـ وـذـلـكـ منـ أـجـلـ تـطـوـيرـ وـتعـزيـزـ عـملـيةـ التـفـسـيرـ لـدـرـجـاتـ الاختـبارـ، إـلاـ أـنـ هـذـاـ اـفـتـرـاضـ يـنـتـهـيـ فـيـ الـأـبـحـاثـ التـرـبـوـيـةـ، إـذـ تـوـجـدـ عـوـاـمـلـ كـثـيـرـةـ تـؤـثـرـ فـيـ أـدـاءـ الـمـفـحـوسـ مـثـلـ الشـخـصـيـةـ، وـعـوـاـمـلـ تـنـعـلـقـ بـتـطـبـيقـ الاختـبارـ مـثـلـ مـسـتـوـيـ الـدـافـعـيـةـ، وـفـلـقـ لـلـاـختـبارـ، وـالـقـدـرـةـ عـلـىـ الـعـمـلـ بـسـرـعـةـ، وـمـعـرـفـةـ الاـسـتـخـدـامـ الصـحـيـحـ لـأـورـاقـ الـإـجـابـةـ...ـالـخـ.

ولـكـ ماـ هوـ مـقـصـودـ بـذـلـكـ اـفـتـرـاضـ أـنـ يـكـونـ هـنـاكـ عـامـلـاـ وـاحـدـاـ سـائـداـ عـلـىـ الـعـوـاـمـلـ الـأـخـرـىـ بـحـيثـ يـكـونـ هـذـاـ عـاـمـلـ هـوـ السـمـةـ التـيـ يـقـيـسـهـاـ الاـختـبارـ.

### ثانياً : الاستقلال المحلي (Local Independence)

يـقـصـدـ بـالـاسـتـقـلـالـ الـمـحـلـيـ (Local Independence) أـنـ استـجـابـةـ الـمـفـحـوسـ عـلـىـ مـفـرـدـاتـ الاـختـبارـ مـسـتـقـلـةـ إـحـصـائـيـاـ عـنـدـمـاـ يـؤـخذـ مـسـتـوـيـ الـمـفـحـوسـ بـعـينـ الـاعـتـباـرـ، أـيـ أـنـ استـجـابـةـ الـمـفـحـوسـ عـلـىـ مـفـرـدـةـ مـاـ يـجـبـ أـنـ لاـ يـؤـثـرـ فـيـ مـفـرـدـةـ أـخـرـىـ، أـيـ أـنـ تـقـدـيرـ صـعـوبـةـ أـيـ مـفـرـدـةـ لـاـ يـعـتـمـدـ عـلـىـ تـقـدـيرـاتـ صـعـوبـةـ الـمـفـرـدـاتـ الـأـخـرـىـ وـلـاـ يـعـتـمـدـ عـلـىـ قـدـرـةـ الـأـفـرـادـ الـذـيـنـ يـجـبـونـ عـلـيـهـاـ، وـكـذـلـكـ لـاـ يـعـتـمـدـ عـلـىـ تـقـدـيرـ قـدـرـةـ أـيـ مـجـمـوـعـةـ أـخـرـىـ مـنـ الـأـفـرـادـ الـذـيـنـ طـبـقـ عـلـيـهـمـ الاـختـبارـ وـلـاـ يـعـتـمـدـ عـلـىـ قـيـمـ صـعـوبـةـ الـمـفـرـدـاتـ التـيـ يـجـبـونـ عـلـيـهـاـ، وـهـذـاـ اـفـتـرـاضـ يـوـضـحـ أـنـ قـدـرـةـ الـمـفـحـوسـ وـخـصـائـصـ الـمـفـرـدـةـ هـمـاـ وـحـدهـمـاـ اللـتـانـ تـؤـثـرـانـ فـيـ الـأـدـاءـ أوـ الـإـسـتـجـابـةـ.

وبـتـحـقـقـ هـذـاـ اـفـتـرـاضـ فـإـنـ اـحـتـمـالـ الـحـصـولـ عـلـىـ أـيـ تـسـلـسـلـ مـنـ الـعـلـامـاتـ لـجـمـوـعـةـ مـنـ الـمـفـرـدـاتـ هـوـ بـبـسـاطـةـ حـاـصـلـ ضـرـبـ الـاحـتـمـالـاتـ لـكـلـ هـذـهـ الـمـفـرـدـاتـ. (Allen & Yen, 1979). ويـعـدـ عـلـامـ (1995) أـنـ تـحـقـقـ هـذـاـ الشـرـطـ يـجـعـلـ لـلـدـرـجـةـ الـكـلـيـةـ لـلـفـرـدـ عـلـىـ الاـختـبارـ مـعـنـىـ، بـحـيثـ يـكـنـ اـسـتـخـدـامـهـاـ فـيـ التـقـدـيرـ الـإـحـصـائـيـ لـلـسـمـةـ الـمـقاـسـةـ تـقـدـيراـ يـتـسـمـ بـالـكـفـاـيـةـ، وـعـنـدـئـذـ يـكـنـ اـعـتـباـرـ الاـختـبارـ يـقـيـسـ سـمـةـ أـحـاديـةـ الـبـعـدـ (Sufficient estimate) (1995).

ومن المهم ملاحظة أن افتراض الاستقلال المحلي (Local Independence) لا يؤدي إلى أن تكون المفردات غير مرتبطة من خلال المجموعة الكلية للمفحوصين. حيث أن الارتباطات الموجبة بين زوج المفردات Pairs of Items هو نتيجة للتباين بين المفحوصين على القدرة المفاسدة لمفردات الاختبار، في حين أن درجات المفردات غير مرتبطة عند مستوى ثابت من القدرة (Hambleton & Swaminthan, 1985).

ويشير كروكر و الجينا (Crocker & Algina, 1986) إلى نقطتين مهمتين:

١- الاستقلال المحلي، وأحادية البعد مفهومان غير مترادفين.

٢- في حالة وجود سمة واحدة فهذا لا يعني بالضرورة أن المفردات مستقلة محليا. وبישار إلى أن هذا الافتراض ينافق ظاهرياً الفكرة التقليدية عن التجانس الداخلي والذي يفترض بأن تكون الاستجابات على جميع المفردات مترابطة، بينما فعلياً لا يوجد تناقض فعلى الرغم من أن الاستقلال المحلي مشروطاً (conditional statement) إلا أن التجانس الداخلي ليس كذلك. بمعنى خت افتراض الاستقلال المحلي فإن المفردات المختلفة غير مرتبطة للموضوعات أحادية القدرة. بينما خت مبدأ التجانس الداخلي فإن الاستجابات للمفردات المختلفة مرتبطة ثنائياً عبر جميع الموضوعات المختلفة ذات القدرات المختلفة (Suen, 1990).

### **ثالثاً: منحنى خصائص المفردة (ICC)**

تفق كل من النماذج الكلاسيكية ونماذج الاستجابة للمفردات الاختبارية في افتراض وجود متصل للسمة، وأنه يمكن تقدير احتمال إجابة فرد إجابة صحيحة عن مفردة اختبار إذا علمنا موقعه على هذا المتصل.

ويؤكد هذا الاتجاه فانكستو (Fanxita, 1998) إذ يقول: أن النماذج السيكومترية سواء الكلاسيكية أو الحديثة تعتمد على فرضية مؤداها أن موقع الفرد على متصل سمة كاملة يكون مؤشراً على احتمال إجاته إجابة صحيحة عن أي مفردة في اختبار يقيس هذه السمة، ولكنهما يختلفان في كيفية تحديد هذا الموقع وارتباطه باحتمال الإجابة الصحيحة عن المفردة أي أن الفرق بينهما يتعلق بشكل وخصائص الدالة التي تحدد هذه العلاقة.

ويعده منحنى خصائص المفردة (ICC) أحد المفاهيم الأساسية في نظرية الاستجابة للمفردة وهو يمثل احتمالية إجابة المفحوص إجابة صحيحة عن مفردة كونها دالة في القدرة  $\theta$ ). وتزداد احتمالية إجابة المفحوص على المفردة بازدياد قدرة المفحوص (Crocker & Algina, 1986).

وتتميز هذه الدالة اللوغاريتمية بخصائص يمكن الإفادة منها في جعل ميزان درجات

الاختبار خطياً و في تقدير الخصائص السيكومترية للاختبارات تقديراً مستقلاً عن خصائص عينة المختبرين (علام، ١٩٩٥).

ويعد منحنى خصائص المفردة الممثل ب\_daالة تراكمية (Ogive Function) هو المقبول حالياً بوصفه تمثيلاً جيداً للعلاقة بين القدرة ( $\theta$ )، واحتمالية الإجابة عن مفردة إجابة صحيحة ( $\pi$ ). ويتم اعتماده من خلال العديد من الدراسات التجريبية وأن هذه الدالة التراكمية هي الأساس النظري لنظرية الاستجابة للمفردة . (Suen, 1990).

#### رابعاً: السرعة Speediness

إن الافتراض الضمني عند جميع مستخدمي نماذج الاستجابة للمفردة هو أن الاختبارات الملائمة للنموذج لا تطبق تحت شرط السرعة. إذ إنّ أن المفحوصين الذين أخفقوا في الإجابة عن مفردات الاختبار كانت بسبب القدرة المحدودة وليس بسبب فشلهم في عدم الوصول لفردات الاختبار .

وربما لا يعلن عن هذا الافتراض إذ أنه متضمن ضمن افتراض أحاديه البعد . فعندما تؤثر السرعة في الأداء على الاختبار فإنه على الأقل هناك سمتان تقاسان بهذا الاختبار هما سرعة الأداء والسمة المقاسة بحتوى الاختبار (Hambleton & Swaminthan, 1985).

وقد اثبتت عن هذه النظرية مجموعة من النماذج والتي تعرف بنماذج السمات الكامنة (Latent Trait Models) وتهدف جميعها تحديد علاقة بين أداء الفرد على الاختبار- وهو ما يمكن ملاحظته ملاحظة مباشرة- وبين السمات أو القدرات التي تكمن وراء هذا الأداء وتفسره.

وهناك العديد من الاستخدامات العملية لنماذج الاستجابة للمفردة فعلى سبيل المثال : يمكن تقدير قيمة السمة الكامنة لمفحوص معين من خلال أي مجموعة من المفردات الملائمة، وبالتالي يقلص زمن الاختبار، كما يمكن أيضاً تقدير بارامترات المفردة من خلال أي مجموعة من المفحوصين بحيث تغطي مدى مناسبها من السمة الكامنة، كما يمكن أيضاً الكشف عن التحيز في مفردات الاختبار للعرق أو الجنس (Allen & Yen, 1979).

ويعد نموذج راش أحادي البارامتر من أكثر نماذج نظرية الاستجابة للمفردة شيوعاً في تصميم وبناء الاختبارات والمقاييس التربوية والنفسية. وبهتم بتحديد موقع المفردة الاختبارية على ميزان صعوبة جميع المفردات التي تشكل الاختبار، كما يهتم بتدريب مستويات قدرة الفرد باختبار معين على نفس مستويات القدرة المقاسة. وقد بدأ استخدام نموذج راش في الاختبارات التي تكون الإجابة عن مفرداتها ثنائية يعني إما صفر أو واحد. ولم تقتصر تطبيقات نموذج

(راش) على مجال القياس النفسي أو على التحصيل الدراسي في مجال القياس التربوي، ولكن امتدت تطبيقات هذا النموذج لتشمل الاستجابات المتردجة بعد اقتصاره في أول الأمر على الاستجابة الثنائية (Tissen, Wainer & Bowang, 2001) ويعتمد نموذج راش على نقطتين أساسيتين:

١- يزداد احتمال الإجابة الصحيحة عن المفردة الاختبارية بزيادة قدرة الفرد التي تقيسها تلك المفردة.

٢- احتمال أن يجيب فرد عن مفردة سهلة إجابة صحيحة أكبر من احتمال إجابتة عن مفردة صعبة.

ويتميز نموذج راش عن النموذجين ثنائي المعلم وثلاثي المعلم بأن الدرجة الكلية على الاختبار (عدد الاستجابات الصحيحة) كافية إحصائياً لتقدير قدرة الفرد ( $\theta$ ). وأن عدد الأفراد الذين أجابوا على المفردة إجابة صحيحة كافٍ إحصائياً لتقدير معلم الصعوبة (b). كما أن الأفراد الذين لهم نفس الدرجات الخام سيكون لهم نفس القدرة ( $\theta$ ) (Harris, 1989).

كما يشير ألن وبن (Allen & Yen, 1979) إلى أنه في نموذج راش ليس هناك حاجة إلى مفردات موزونة للحصول على الدرجة الكلية والتي تعطي أعلى قيمة ممكنة من المعلومات عن السمة المقاسة إذ إن عدد الدرجات الصحيحة هو أفضل درجة كلية ممكنة للاستخدام بينما في النموذجين ثنائي وثلاثي المعلم فإن عدد الدرجات الصحيحة لا بعد أفضل درجة كلية ممكنة للاستخدام.

ويتميز نموذج راش بإمكانية تقدير صعوبة كل مفردة من مفردات الاختبار بطريقة مستقلة عن الأفراد الذين طبق عليهم الاختبار، وهذه الخاصية تسمح باختيار أي مفردة بوصفها نقطة أصل لميزان صعوبة المفردات، حيث يتم تدريج صعوبات جميع المفردات الأخرى التي يشتمل عليها الاختبار بدءاً بهذه النقطة على الميزان.

ولما كان معلم الفرد يقيسه ما يقيسه معلم البند نفسه، ويغير عنه على نفس المقاييس، لذا ينبغي أن ترد جميع التقديرات سواء الخاصة بالفرد، أو الخاصة بالبند، إلى نقطة أصل واحدة، من الممكن تحديدها بصورة مستقلة. (كاظام، ١٩٩٦).

وبناء على ذلك فالمحظوظ الذي يتلك قدرة تساوى ضعف قدرة مفحوص آخر تكون احتمالية إجابتة على مفردة ما إجابة صحيحة تساوى ضعف احتمالية إجابة المحظوظ الآخر عن المفردة إجابة صحيحة، كما أن في حالة وجود مفردة سهولة تها تساوى ضعف سهولة مفردة ما، وأن المحظوظ يتلك ضعف فرصه النجاح للإجابة عن المفردة الأسهل.

- وهناك عدد من الخصائص التي يتميز بها منحنى خصائص المفردة (ICC) المتعلق بنموذج (راش) يمكن إيجازها فيما يلي:
- ١- تزداد احتمالية الإجابة عن المفردة إجابة صحيحة بازدياد مستوى القدرة.
  - ٢- تختلف المفردات في صعوبتها فقط.
  - ٣- ميل المنحنى لجميع المفردات متساو (القدرة التمييزية لجميع المفردات متساوية)
  - ٤- تقارب المنحنيات إلا أنها لا تتقاطع
  - ٥- نقطة انقلاب المنحنى (ICC) تظهر عندما تكون احتمالية الإجابة عن المفردة إجابة صحية (Embretson & Reise, 2000).
- ويفترض النموذج أن الفرد لا يلجأ إلى التخمين العشوائي في إجابته على مفردات الاختبار، (علم، ٢٠٠٢؛ Hashway, 1998). وحتى توافر متطلبات الموضوعية في القياس يجب أن تستوفي فروض نموذج راش وهي:
- ١- أحادية البعد : أي لا تختلف بنود الاختبار فيما بينها إلا من حيث مستوى الصعوبة فقط، كما يكون الأفراد ذوي قدرة أحادية البعد خدد وحدها مستوى أدائهم على الاختبار.
  - ٢- استقلالية القياس، وتعنى:
- عدم اعتماد تقدير صعوبة البنود على تقديرات صعوبة البنود الأخرى المكونة للاختبار ولا على تقديرات قدرة الأفراد الماسبة الذين يحبون عنها.
  - عدم اعتماد تقدير قدرة الفرد على تقديرات قدرة أي مجموعة من الأفراد الذين يؤدون الاختبار أو على تقديرات صعوبة البنود التي يؤدونها.
- ٣- توازي المنحنيات المميزة للبنود أي أن القدرة التمييزية لجميع البنود متساوية (كاظام، ١٩٩٦).

ويتميز نموذج راش بإمكانية تقدير صعوبة كل مفردة من مفردات الاختبار بطريقة مستقلة عن الأفراد الذين طبق عليهم الاختبار، وهذه الخاصية تسمح باختيار أي مفردة كونها نقطة أصل لميزان صعوبة المفردات، حيث يتم تدريج صعوبات جميع المفردات الأخرى التي يشتمل عليها الاختبار بدءاً بهذه النقطة على الميزان. ويرى ثورندايك (Thorndike, 1982) أن هذه الفروض تكون مقبولة تحت الشروط التالية :

- ١- جميع المفردات متجانسة تماماً في الشكل والمحتوى.
- ٢- جميع المفردات تم غربلتها من خلال التجربة التمهيدية بحيث يتم حذف المفردات الغامضة وتلك غير المميزة.

٣- أن تكون المفردات من نوع الاستجابة المصاغة وليس الاختيار من عدة إجابات معطاة بحيث يقترب معامل التخمين من الصفر

٤- أن تعكس المفردات نموا عاما في سمة معينة وليس في تعليم محدد، فمن المتوقع أن تتأثر الصعوبة النسبية لتلك المفردات التي تعرف موضوعا قد تم حديثا الانتهاء من تدرسيه. ونظراً لحداثة نظرية الاستجابة للمفردة وحداثة مفاهيمها وقلة توفر البرامج الإحصائية لتقدير برامحها ماذجها المختلفة فمعظم الدراسات السابقة التي وقعت لدى الباحث استخدمت النظرية التقليدية في القياس في المقارنة بين اختبارات الاختيار من متعدد واختبارات التكميل فقد هدفت دراسة الأحمدي (٢٠٠٩) إلى معرفة أثر اختبارات الاختيار من متعدد والإجابة القصيرة في الاحتفاظ بنوائح التعلم ولتحقيق هدف الدراسة تم إعداد اختبار خصيلي في مادة العلوم لطلبة الصف الثاني المتوسط، وتم تكوين نموذجين من الاختبار متشابهين في المتن و في عدد المفردات .النموذج الأول من نمط الاختيار من متعدد وعدد مفرداته (٣٠) مفردة لكل منها أربعة بدائل. والنموذج الثاني من نمط الإجابة القصيرة وعدد مفرداته (٣٠) مفردة وتكونت عينة البحث من (٦٧) طالبا من طلاب الصف الثاني المتوسط موزعين على ثمانية مدارس بواقع ثلاث فصول في كل مدرسة ويدرسها نفس المعلم وبينت النتائج تفوق مجموعة الإجابة القصيرة على مجموعة الاختيار من متعدد في الاحتفاظ بنوائح التعلم.

وهدفت دراسة الردادي (٢٠٠٦) معرفة أثر اختلاف نمط الاختبار على المصائص السيكومترية للاختبار ومفرداته ولتحقيق أهداف الدراسة قام الباحث بإعداد اختبار خصيلي في مادة العلوم لطلبة الصف الثالث المتوسط وتم تكوين ثلاثة ماذج من الاختبار متشابهه في المتن وعدد المفردات وتكون النموذج الأول من (٣٠) مفردة من نوع الاختيار من متعدد لكل مفردة منها أربعة بدائل والنماذج الثاني كان من نمط الصواب والخطأ وعدد مفرداته (٣٠) مفردة والنماذج الثالث كانت مفرداته من نمط التكميل وعدد مفرداته (٣٠) مفردة وطبقت النماذج الثلاثة على (١٢٠) طالبا وطالبة من طلبة الصف الثالث المتوسط تم اختيارهم من (١٥) مدرسة حكومية وأشارت النتائج إلى أن وجود فروق ذات دالة إحصائية في معاملات الثبات بين الاختبارات الثلاثة لصالح اختبار التكميل وجود فروق دالة إحصائيا في معاملات الصعوبة والتمييز لصالح اختبار التكميل.

وهدفت دراسة الكحلوت (٢٠٠٢) المقارنة بين المصائص السيكومترية لكل من اختبارات الاختيار من متعدد واختبارات التكميل وقام الباحث بإعداد اختبارين أحدهما من نوع الاختيار

من متعدد والآخر من نوع التكميل بحيث تكون مفردات التكميل هي مفردات الاختبار من متعدد نفسها بعد حذف البدائل وتم تطبيق الاختبارين على عينة مكونة من (٤٥١) طالباً وطالبة من طلبة الصف السادس الابتدائي تم اختبارهم عشوائياً من مجتمع الدراسة الذي ضم مدارس وكالة الغوث الدولية في الأردن. وتوصلت نتائج الدراسة إلى أن معامل الثبات لفردات التكميل أعلى منه لفردات الاختبار من متعدد وأن عدد العوامل ذات المعنى الجوهري المستخلصة من التحليل العاملی لاختبار التكميل (أربعة) وهي متساوية لعدد العوامل التي خطط لقياسها بينما كان عدد العوامل لاختبار الاختبار من متعدد (سبعة) وأن متوسط عوامل الصعوبة لفردات الاختبار من متعدد أكبر من متوسط عوامل الصعوبة لفردات التكميل.

وهدفت دراسة الموموري (١٩٩٩) معرفة أثر اختلاف نمط الاختبار حسب نوعية الاستجابة المطلبة من مفرداته: منشأة أو مختاراة في صدقه العاملی والتبنؤی و تكونت عينة الدراسة من (٤٥٤) طالباً وطالبة من طلبة الصف السادس الأساسي فقد طبق عليهم اختباران في مفاهيم العمليات الأربعية على الكسور والأعداد الكسرية أحدهما من نمط الاختبار من متعدد والآخر من نمط التكميل وتم إجراء التحليل العاملی لنمطي الاختبار بتحليل المكونات الأساسية ولاستقصاء قدرة النمطين على ترتيب الأفراد حسب قيم (z) الناجة عن اختبار ولكسون (Wilcoxon) للأزواج المتزابطة كما أجرى تحليل الانحدار المتعدد على العوامل الناجة من التحليل العاملی لكلا النمطين وحسب اختبار (t) لعينتين متزابطتين لمعرفة دلالة الفروق في متوسطات العلامات المتنبئ بها من النمطين ودللت النتائج على أن النمطين يقيسان عوامل مختلفة وبفسران نسباً مختلفة من التباين لصالح نمط التكميل الذي يقيس عدداً أقل من العوامل ولا يرتبطان الطلبة وفق قدراتهم بالطريقة ذاتها وإن للنمطين القدرة ذاتها على التنبؤ.

وبينت دراسة كدادا (Gadala) أن هناك فروقاً دالة إحصائياً بين متوسط علامات الطلبة على اختبار الاختبار من متعدد الذي يقيس مهارة إجراء العمليات الحسابية واختبار التكميل الذي يقيس نفس المهارة وذلك في الصفين الثاني والثالث الأساسيين (Gadalla, 1999). وأجرى مانهارت (Manhart) دراسة هدفت استقصاء عدد العوامل التي يقيسها نمطان من اختبارات العلوم: الاختبار من متعدد، والاستجابة المنشأة وطبقت الاختبارات على عينة مكونة من (٨٧٢) طالباً من الصفوف التاسع إلى الثاني عشر وأسفرت نتائج التحليل العاملی عن أن كلا النمطين يقيسان نفس العامل (Manhart, 1996).

وأجرى بريجمان (Brigeman) دراسة هدفت المقارنة بين نتائج الطلبة على اختبار الاختيار من متعدد مع نتائجهم على اختبار الإجابة المنشأة وتوصلت النتائج إلى أن متوسط علامات الطلبة في اختبار الإجابة المنشأة أقل من متوسط علاماتهم في اختبار الاختيار من متعدد، وأن ثبات اختبارات الإجابة المنشأة أعلى من ثبات اختبار الاختيار من متعدد (Brigeman, 1994).

كما بينت نتائج دراسة ثايسن وأخرين التي استخدم فيها اختباران غير متساوين في عدد المفردات الاختبارية أحدهما في علم الحاسوب والثاني في الكيمياء أن عدد العوامل المستخلصة من مفردات الاختيار من متعدد أكبر من عدد العوامل المستخلصة من مفردات الإجابة المنشأة (Tissen, Wainer & Bowang, 1994).

من خلال العرض السابق تبين هناك تناقض وعدم اتساق في النتائج التي تم التوصل إليها وجميع الدراسات السابقة اعتمدت المؤشرات التقليدية في انتقاء المفردات والمقارنة بينها ومن هنا جاءت الدراسة الحالية لإضافة مؤشرات جديدة تعتمد على الموضوعية في القياس من خلال ما يوفره نموذج راش أحدى المعلم تمثل في بنود صادقة في تعريفها للمتغير موضوع القياس وأنماط استجابات صادقة تتحدد مواضع الأفراد على متصل المتغير.

## **مشكلة الدراسة**

يدور جدل كبير حول الكيفية التي تطور بواسطتها عملية بناء الاختبارات وهل الأفضل استخدام مفردات من نوع الاستجابة المنتقدة أو الاستجابة المصاغة فهناك أساس منطقى للافتراض القائل أن المتطلبات المعرفية لمفردات الاستجابة المنتقدة تختلف عنها لمفردات الاستجابة المصاغة ونظراً لشيوخ نتائج الدراسات في تقديم أدلة نوعية على أن نوعي المفردات يقدمان عمليات عقلية مختلفة.

## **أهداف الدراسة**

تهدف الدراسة الحالية مقارنة المصائص السيكومترية لمفردات الاختيار من متعدد مع مفردات اختبار التكميل التي تناظرها في المحتوى وذلك باستخدام أشهر نماذج نظرية الاستجابة للمفردة (نموذج راش).

### **أسئلة الدراسة**

- ١- ما مدى اختلاف تدرج مفردات اختبار الاختيار من متعدد عن تدرج مفردات اختبار التكميل باستخدام نموذج راش؟
- ٢- ما مدى اختلاف متوسط تقديرات القدرة للأفراد على اختبار الاختيار من متعدد عن متوسط تقديرات القدرة للأفراد على اختبار التكميل؟
- ٣- ما مدى اختلاف القيمة التقديرية لثبات درجات اختبار الاختيار من متعدد عن القيمة التقديرية لثبات درجات اختبار التكميل؟

### **أهمية الدراسة**

- ١- تأتي أهمية هذه الدراسة من خلال استخدام نظرية الاستجابة للمفردة في المقارنة بين الخصائص السيكومترية لمفردات الاختيار من متعدد ومفردات التكميل وفقاً لنموذج راش أحادي المعلم حيث يوفر هذا النموذج وحدة قياس واحدة لكل من الأفراد والمفردات مما يعطي دقة أعلى في إجراء المقارنات.
- ٢- كما تأتي أهمية هذه الدراسة من خلال رفع صدق البناء للاختبار من خلال ما يوفره نموذج راش من أحادية البعد وتتمثل أحادية البعد في مفردات صادقة تتدرج من حيث صعوبتها على ميزان تدرج واحد مشترك، وصفراً واحد مشترك.

### **محددات الدراسة**

اقتصرت الدراسة الحالية على:

- عينة من طلبة الصف السادس الابتدائي بمدارس الحكومة في محافظة شمال غزة التعليمية
- الوحدة الأولى والثانية من كتاب العلوم العامة المقرر على الصف السادس الابتدائي في الفصل الأول.

### **مصطلحات الدراسة**

**دالة المعلومات للمفردة:** تمثل هذه الدالة بعلاقة منحنية بين متغيرين هما مستويات القدرة التي يمثلها المحور الأفقي والمعلومات المقدمة من خلال المفردة التي يمثلها المحور الرأسى. وتعبر هذه الدالة عن كمية المعلومات التي تقدمها المفردة عن مستوى القدرة التي تقيسها. (Hambleton & Swaminthan, 1985)

دالة المعلومات للاختبار؛ تمثل هذه الدالة بعلاقة منحنية بين متغيرين هما مستويات القدرة التي يمثلها المحور الأفقي والمعلومات المقدمة من الاختبار كله، وتعبر عن مقدار المعلومات المقدم من المجموع الكلى لفردات الاختبار عند أى مستوى للقدرة (Hambleton & Swaminthan, 1985).

### **منهجية الدراسة وإجراءاتها:**

#### **مجتمع الدراسة**

تكون مجتمع الدراسة من جميع طلبة الصف السادس الابتدائي في مدارس الحكومة في منطقة شمال غزة التعليمية في العام الدراسي (٢٠١٠-٢٠٠٩) (٢٨٤٧) طالباً وطالبة منهم (١٤١٩) ذكور و (١٤٢٨) إناث.

#### **عينة الدراسة**

أ- عينة التجربة: تم تطبيق الاختبار على عينة مكونة من (٨٠) طالباً وطالبة من طلبة الصف السادس الابتدائي في الفصل الدراسي الأول من العام الجامعي (٢٠١٠-٢٠٠٩). وكان الهدف من التجربة الاطمئنان على سلامته مفردات الاختبار، والتحقق من إمكانية استخدام موجز راش في تدريب الاختبار. كما تم حساب الزمن اللازم للإجابة عن المفردة الواحدة وقد بلغ (٤/٣) دقيقة لمفردة الاختبار من متعدد وحوالي دقيقة لمفردة اختبار التكميل.

ب- عينة التدريب: تم تطبيق الاختبار على عينة مكونة من (٢٥٠) طالباً وطالبة من طلبة الصف السادس الابتدائي بمدارس الحكومة بمحافظة شمال غزة موزعين على (٧) شعب وذلك في الفصل الدراسي الأول من العام الجامعي (٢٠١٠-٢٠٠٩) وتم اختيار العينة عن طريق اختيار (٤) مدارس إناث و (٣) مدارس ذكور بطريقة عشوائية وتم اختيار شعبية واحدة من كل مدرسة بطريقة عشوائية.

#### **أدوات الدراسة**

لتحقيق غرض الدراسة تم إعداد اختبارين في مادة العلوم لطلبة الصف السادس الابتدائي في الوحدتين الأولى والثانية من الكتاب المقرر أحدهما اختيار من متعدد لكل مفردة (٤) بدائل والأخر من نوع التكميل. كل منها مكون من (٣٠) مفردة وتم الحصول على اختبار التكميل من خلال حذف البديل لمفردات اختيار الاختبار من متعدد. ومن ثم اشتراك الاختباران (الاختيار من متعدد و التكميل) في المتن نفسه لمفردات لذلك لم يجد الباحث ضرورة لإرافق صورة

من هذا الاختبار واكتفى بإرفاق اختبار الاختبار من متعدد وتم عرض الاختبارين على مجموعة من (٤) من المدرسين و (٣) من الموجهين المتخصصين في المادة العلمية وطلب منهم إبداء الرأي في المفردات من حيث سلامتها اللغوية ودققتها في قياس الأهداف التي وضعـت من أجلها وفاعلية المـوهـات ولم تكن هناك اقتراحـات جوهرـية على الاختـبار إلا في بعض النواحي الشـكلـية وـتم معالجـتها.

### **إجراءات التطبيق**

طبقـت أدوات الـدرـاسـة وفقـا لـتصـمـيمـ الفـيـاسـاتـ المـتـكـرـرـةـ وـبعـدـ هـذـاـ التـصـمـيمـ أـقـوىـ منـ تصـمـيمـ المـجمـوعـاتـ العـشـوـائـيـةـ الـذـيـ اـعـتـمـدـتـهـ مـعـظـمـ الـدـرـاسـاتـ السـابـقـةـ ( Ferguson & Takane, 1989). فـبـعـدـ جـلوـسـ أـفـرـادـ الـدـرـاسـةـ لـلـاـخـتـيـارـ وـزـعـتـ أـورـاقـ أـسـئـلـةـ بـحـيثـ يـبـدـأـ طـالـبـ بـأـسـئـلـةـ الـاـخـتـيـارـ مـنـ مـتـعـدـدـ وـبـدـأـ الـآـخـرـ بـجـواـهـرـ بـأـسـئـلـةـ التـكـمـيلـ وـبـعـدـ الـاـنـتـهـاءـ مـنـ إـجـابـةـ هـذـهـ أـسـئـلـةـ أـعـطـىـ الـطـلـبـةـ الـذـينـ بـدـأـوـاـ بـأـسـئـلـةـ الـاـخـتـيـارـ مـنـ مـتـعـدـدـ أـسـئـلـةـ التـكـمـيلـ. وـأـعـطـىـ الـطـلـبـةـ الـذـينـ بـدـأـوـاـ بـأـسـئـلـةـ التـكـمـيلـ أـسـئـلـةـ الـاـخـتـيـارـ مـنـ مـتـعـدـدـ. وقد أـشـرـفـ الـبـاحـثـ نـفـسـهـ عـلـىـ جـمـيعـ إـجـارـاتـ الـتـطـبـيقـ وـسـاعـدـهـ فـيـ ذـلـكـ مـعـلـمـ الصـفـ وـمـ وـقـدـ أـشـرـفـ الـبـاحـثـ نـفـسـهـ عـلـىـ جـمـيعـ إـجـارـاتـ الـتـطـبـيقـ وـسـاعـدـهـ فـيـ ذـلـكـ مـعـلـمـ الصـفـ وـمـ التـطـبـيقـ بـشـكـلـ جـمـاعـيـ فـيـ غـرـفـةـ الصـفـ.

### **تصحيح الأداة**

نظـراـ لـأـسـئـلـةـ التـكـمـيلـ تـنـطـلـبـ مـنـ الـمـفـحـوصـ كـتـابـةـ الـعـبـارـةـ الصـحـيـحةـ وـهـىـ عـلـىـ الأـغـلـبـ إـمـاـ رـقـمـ أـوـ كـلـمـةـ أـوـ كـلـمـتـيـنـ عـلـىـ الـأـكـثـرـ أـعـطـيـتـ إـجـابـةـ الصـحـيـحةـ الـعـلـامـةـ (١)ـ وـإـجـابـةـ الـخـطـأـ (صـفـرـ). أـمـاـ بـالـنـسـبـةـ لـمـفـرـدـاتـ اـخـتـيـارـ الـاـخـتـيـارـ مـنـ مـتـعـدـدـ فـقـدـ أـعـطـيـتـ إـجـابـةـ الصـحـيـحةـ الـعـلـامـةـ (١)ـ وـالـعـلـامـةـ (صـفـرـ). فـيـ حـالـ إـجـابـةـ الـخـطـأـ وـمـنـ ثـمـ فـإـنـ عـلـامـةـ الـمـفـحـوصـ عـلـىـ أـيـ مـنـ النـمـطـيـنـ تـسـاـوـيـ مـجـمـوعـ الـمـفـرـدـاتـ الـتـيـ أـجـابـ عـنـهـ إـجـابـةـ صـحـيـحةـ.

### **المعاجـاتـ الإـحـصـائـيـةـ**

- ١- أـدـخـلـتـ الـبـيـانـاتـ لـبـرـنـامـجـ (SPSS)ـ الـإـصـدارـ (١٣)ـ فـقـدـ تـصـحـيـحـ الـأـوـرـاقـ تـبـعـاـ لـمـفـتـاحـ التـصـحـيـحـ المـعـدـ لـذـلـكـ.
- ٢- أـسـتـخـدـمـ بـرـنـامـجـ (SPSS)ـ لـتـحلـيلـ الـبـيـانـاتـ باـسـتـخـدـامـ نـمـوذـجـ (راـشـ)ـ الـاـحـتمـالـيـ الـلـوـغـارـيـتـمـيـ الـأـحـادـيـ الـعـلـمـ، وـتـدـريـجـ الـمـفـرـدـاتـ، وـتـقـدـيرـ كـلـ مـنـ صـعـوبـةـ الـمـفـرـدـاتـ، وـتـقـدـيرـاتـ الـقـدـرةـ لـلـأـفـرـادـ.

٣- تم حساب الإحصائي ( $t$ ) لعينتين مرتبطتين لدراسة الفروق في الأوساط الحسابية للتواترات تقديرات القدرة للأفراد.

٤- استخدمت الصورة المعدلة لاختبار فلدت (Feldt test) لفحص فرضية تساوى معامل ثبات مرتبطين (Feldt, 1980, Alswalmeh & Feldt, 1994). ويستعمل في هذه الصورة الإحصائي ( $W$ ) العطى في القانون التالي:

$$W = \frac{1 - \hat{\alpha}_1}{1 + \hat{\alpha}_2}$$

حيث

$\hat{\alpha}_1$  : التقدير العيني لمعامل الثبات الأول

$\hat{\alpha}_2$  : التقدير العيني لمعامل الثبات الثاني

وتقوم هذه الصورة على كون التوزيع العيني للاختبار الإحصائي ( $W$ ) يقترب من التوزيع الفائق ( $F$ ) بدرجات حرية خاصة تقدر استنادا إلى مجموعة من إحصاءات العينة (حجم العينة وعدد المفردات ومعامل الارتباط )

## عرض نتائج الدراسة ومناقشتها أولاً: عرض نتائج السؤال الأول

نص السؤال الأول في الدراسة الحالية على: هل يختلف تدريج مفردات اختبار الاختبار من متعدد عن تدريج مفردات اختبار التكميل باستخدام نموذج راش؟

ولتدريب اختبار الاختبار من متعدد والتكميل تم استخدام برنامج (Winsteps version 2.88) وفيما يلي نتائج التحليل:

- ١- حذف البيانات التامة والصفيرية من مصروفه التحليل ويتضمن:
- حذف كل فرد أجاب عن جميع المفردات إجابة صحيحة، إذ تعدد قدرته عندئذ أعلى من المدى الذي يغطيه المقياس.
- حذف كل فرد أخفق في حل جميع المفردات، إذ تعدد قدرته عندئذ أدنى من المدى الذي يغطيه المقياس، وتعتبر قدرة هؤلاء الأفراد خارج نطاق المقياس.
- وقد أسفر ذلك عن عدم حذف أي فرد من الأفراد.
- حذف كل مفردة أجاب عنها جميع الطلبة إجابة صحيحة.

• حذف كل مفردة أخفق في الإجابة عنها جميع الطلبة.

ولا يمكن لأي مفردة من هذه المفردات أن تميز بين مستويات التغير.

وقد أسف ذلك عن عدم حذف أي مفردة من مفردات المقياس. أي أسفت هذه الخطوة عن عدم حذف أي فرد من الأفراد أو أي مفردة من المفردات.

### ٢- استبعاد الأفراد غير الملائمين للنموذج:

بعد الانتهاء من الخطوة السابقة بـأـالـتـحـلـيل بـهـدـف حـذـف الأـفـرـاد غـيرـالـلـائـمـين لـلـنـمـوذـج، بـعـنـى حـذـف الأـفـرـاد غـيرـالـلـائـمـين لـعـمـلـيـة التـدـريـج، أي لـأـصـوـلـ الـقـيـاسـ، وـذـلـك تـبـعـاً لـلـمـحـكـاتـ التـالـيـةـ:

• استبعاد الأفراد الذين تقل قيمة إحصاءات الملاعمة لهم عن (-٢) إذ يعني ذلك تشابه إجابات هؤلاء الأفراد، ما يشير إلى عدم صدق الاستجابات.

• استبعاد الأفراد الذين تزيد إحصاءات الملاعمة لهم عن (+٢) إذ يعني ذلك أن هؤلاء الأفراد قد جازوا الحد المقبول إحصائياً 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 ما يعني أنهم قد اعتمدوا على التخمين، أو عدم جديتهم أو صدقهم في الاستجابة (Wright & Linacre, 1998).

وقد أسفت هذه الخطوة عن حذف (١٨) شخصاً في اختبار الاختيار من متعدد و (١٦) شخصاً في اختبار التكميل.

### ٣- حذف المفردات غير الملائمة للنموذج:

أعيد التحليل مرة أخرى لـحـذـفـ المـفـرـدـاتـ غـيرـالـلـائـمـينـ لـلـنـمـوذـجـ، بـعـنـىـ حـذـفـ المـفـرـدـاتـ التـىـ بـهـاـ بـعـضـ الـعـيـوبـ، وـالـتـيـ لـاـ جـعـلـهـاـ صـالـحـةـ لـتـدـريـجـ التـغـيـرـ مـوـضـوـعـ الـقـيـاسـ، وـذـلـكـ تـبـعـاًـ لـلـمـحـكـاتـ التـالـيـةـ:

• حذف المفردات التي تقل قيمة إحصاءات الملاعمة لها عن القيمة (-٢,٥) إذ يعني ذلك عدم استقلالية تلك المفردات عن باقي مفردات المقياس، أو أنها تقيس متغيراً آخر شديد التشابه مع المتغير المفترض قياسه.

• حذف المفردات التي تزيد قيمة إحصاءات الملاعمة لها عن (+٢,٥) إذ يعني ذلك أن هناك خلل في بناء المفردة، أو أنها تقيس متغيراً آخر.

وقد أسفت نتائج التحليل عن حذف (٦) مفردات من اختبار الاختيار من متعدد و (٧) مفردات من اختبار التكميل، والمجدول رقم (١) يبين أرقام المفردات المحذوفة في كل من الاختبارين وإحصائي الملاعمة لكل منها.

**الجدول رقم (١)**

**المفردات التي تم حذفها في كل من اختباري الاختيار من متعدد والتكميل وإحصائي الملاعمة لكل منها**

الاختبار	رقم المفردة	الصعوبة باللوجيت	الخطأ المعياري S.E	إحصاءات الملاعمة	التقاريبي	التباعدي
اختبار الاختيار من متعدد	٨	٠,٩٣	٢,٦٠	٢,١٠	٢,٦٠	٢,٦٠
	١٠	٠,٩٢-	٠,٦٠	٢,٧٩	٠,٦٠	٠,٦٠
	١٢	٠,٧٧	١,٩٠	١,٢٢	٢,٦٠	٢,٦٠
	١٨	٠,٨٥	١,٤٠	١,٢٢	٢,٧٩	٢,٧٩
	٢٠	٠,٦٩	٢,٤٠	٠,٨٥	٢,٨٠	٢,٨٠
	٢٤	٠,٤٧	٠,٨٠	١,١٠	٢,٢٢	٢,٢٢
	٣	٠,٤٤	٠,١٣	١,١٦	٢,٨٤	٢,٨٤
	٥	٠,٢٠	٠,١٨	١,٢٢	٢,٧٣	٢,٧٣
	٧	٠,٣٥	١,٢٠	١,٠٧	٢,٦٠	٢,٦٠
	٨	٠,٨٢	٠,٨٨	١,٤٢	٢,٧٨	٢,٧٨
اختبار التكميل	١٢	٠,٤٤	٠,١٦	١,٦٥	٢,٥٠	٢,٥٠
	١٦	٠,٣٤	٠,٩٥	٢,١٠	٢,٦٦	٢,٦٦
	٢٠	٠,٤٦	٠,٩٣	١,٢٠	٢,٢٢	٢,٢٢

يلاحظ من خلال الجدول رقم (١) أن المفردات التي افتح البرنامج حذفهاتجاوزت حدود الملاعمة التي حددها الباحث وهي (٢,٥) و (٢,٥). وبعد حذف الأفراد والمفردات غير الملاعمة لعملية التدريج تم الحصول على تدريج المفردات لكل من الاختبارين والجدول رقم (٢) يبين تدريج المفردات لاختبار الاختيار من متعدد، بينما يبين الجدول رقم (٣) تدريج المفردات لاختبار التكميل.

**الجدول رقم (٢)**

**تدريج مفردات الاختيار من متعدد باستخدام نموذج راش مرتبة حسب صعوبتها**

ترتيب المفردات بعد التدريج	الصعوبة باللوجيت	الخطأ المعياري S.E	إحصائي الملاعمة	التقاريبي	التباعدي
١٩	١,٥-	٠,٧٨	٠,٢٨	٠,٢١	٠,٢١
١٦	١,٢-	٠,٧٤	٠,٢٩	٠,٢٧-	٠,٢٧-
٢٢	١,٤-	٠,٧٩	٠,٢٩	٠,٨٧	٠,٨٧
١٤	١,٢-	٠,٧٤	٠,٣١	٠,٤٨	١,٤٨
١٧	١,٠-	٠,٨٤	٠,٢٨	٠,١٥	٠,١٥
٢٧	١,٠-	٠,٨٣	٠,٢٩	٠,١٢	١,١٢
١٣	٠,٧-	٠,٧٢	٠,٣٤	٠,١٥-	١,١٥-

تابع الجدول رقم (٢)

الرتبة بعد الترتيب	الصعوبة باللوجيت	الخطأ المعياري باللوجيت	النوع	النوع
النوع	النوع	النوع	النوع	النوع
٢	٠,٥٨	٠,٥٣	التقاري	البعادي
٦	٠,٩١	٠,٧٩	التقاري	ال البعادي
١	٠,٥٣	٠,٦	التقاري	ال البعادي
١٥	٠,٩٤	٠,٢	التقاري	ال البعادي
٤	٠,٩١	٠,٢٢	التقاري	ال البعادي
٢٩	٠,٩٥	٠,٢٣	التقاري	ال البعادي
٢٨	٠,٩٧	٠,٢	التقاري	ال البعادي
٢	٠,٩٥	٠,٣٦	التقاري	ال البعادي
٢٢	١,٠١	٠,٢٩	التقاري	ال البعادي
٩	١,٠٥	٠,٢٨	التقاري	ال البعادي
٢٥	١,٠٨	٠,٢	التقاري	ال البعادي
٢١	١,١٢	٠,٢٨	التقاري	ال البعادي
٧	١,١٨	٠,٢٩	التقاري	ال البعادي
٣٠	١,٢	٠,٢٩	التقاري	ال البعادي
١١	١,٣٨	٠,٣١	التقاري	ال البعادي
٥	١,٤٤	٠,٣٤	التقاري	ال البعادي
٢٦	١,٤٥	٠,٧٨	التقاري	ال البعادي

نلاحظ من خلال الجدول رقم (٢) أن قيم معاملات الصعوبة لفردات اختبار الاختيار من متعدد امتدت بين (-١,٥ - ٢,٧) باللوجيت ويعد هذا الذي ضيق نسبيا بينما امتدت قيم الخطأ المعياري لفردات اختبار الاختيار من متعدد بين (٠,٥٣ - ١,٤٥).

الجدول رقم (٣)

ترتيب مفردات اختبار التكميل باستخدام نموذج راش مرتبة حسب صعوبتها

الرتبة بعد الترتيب	الصعوبة باللوجيت	الخطأ المعياري باللوجيت	النوع	النوع
النوع	النوع	النوع	النوع	النوع
١٧	١,٠-	٠,٦١	التقاري	ال البعادي
٢٩	٠,٧-	٠,٩٥	التقاري	ال البعادي
٢٧	٠,٨-	٠,٦٥	التقاري	ال البعادي
١١	٠,٦-	٠,٧٣	التقاري	ال البعادي
٢٢	٠,٣-	٠,٥١	التقاري	ال البعادي
٢	٠,٣-	٠,٦١	التقاري	ال البعادي
٢٨	٠,٣-	٠,٦٨	التقاري	ال البعادي
١٥	٠,٣-	٠,٧٧	التقاري	ال البعادي

## (٣) تابع الجدول رقم

ترتيب المفردات بعد التدريج	الصعوبة باللوجيت	الخطأ المعياري باللوجيت	النقاري	احصائي الملاعة
التباعدي	التباعدي			
١٢	٠,٢-	٠,٩٢	١,٠٩	١,٦٢-
٢٤	٠,٢-	٠,٦٥	٠,٩٥	١,٠٢
١٩	٠,١-	٠,٩٣	٠,٩٢	٠,٨٤-
١٤	٠,١-	٠,٨٤	١,٠٢	٠,٤٩
٢٠	٠,١-	٠,٧٥	١,٠٣	٠,٨٨
٢٥	٠-	٠,٩١	١,١٩	٠,٣٧
١٨	٠,١	٠,٦٢	١,٠٣	٢,٧١
٢١	٠,٢	٠,٧٨	١,٠٤	٣,٧١
٢٣	٠,٢	٠,٩٨	٠,٨٨	٠,٤٩
١٠	٠,٤	١,٠٨	١,١٩	١,٠٢
٢٦	٠,٤	١,١١	١,٠٣	٠,١٤
١	٠,٥	١,٠٨	١,٠٩	٤,٩٥-
٤	٠,٥	١,١١	١,١١	١,٥١-
٦	١,٠	١,٢٤	١,٢٦	٠,٥٢-
٩	١,٤	١,٦٧	١,٢٩	٠,٣٠

نلاحظ من خلال الجدول رقم (٣) أن قيم معاملات الصعوبة لمفردات اختبار التكميل امتدت بين (١,٠ - ١,٤) باللوجيت ويعدها المدى ضيق إلى حد ما وهو أقل من مدى الصعوبة لمفردات اختبار الاختيار من متعدد بينما امتدت قيم الخطأ المعياري لمفردات اختبار التكميل بين (٠,٧٩ - ١,٢٩). ولمعرفة مدى الدقة في القياس لكل من اختبار الاختيار من متعدد واختبار التكميل تم حساب دوال المعلومات لجميع المفردات في الاختبارين والجدول رقم (٤) يبيّن ذلك.

## (٤) الجدول رقم

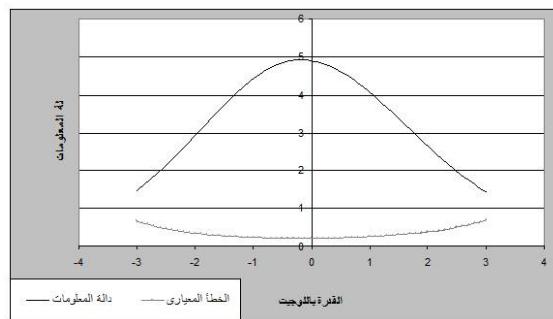
## دوال المعلومات لجميع المفردات في اختبار الاختيار من متعدد واختبار التكميل

مسلسل	رمز المفردة في اختبار	دالة المعلومات للمفردات في اختبار		
		الاختيار من متعدد	النكميل	الاختيار من متعدد
النكميل	الاختيار من متعدد	النكميل	الاختيار من متعدد	دالة المعلومات للمفردات في اختبار
١	١٩	١٧	٠,٠٥٦٥	٠,٠٦٠٣
٢	١٦	٢٩	٠,٠٥٨٢	٠,٠٦١٨
٣	٢٢	٢٧	٠,٠٥٧٤	٠,٠٦١٤
٤	١٤	١١	٠,٠٥٨٢	٠,٠٦٢٢
٥	١٧	٢٢	٠,٠٦٠٣	٠,٠٦٢٩
٦	٢٧	٢	٠,٠٦٠٣	٠,٠٦٢٩
٧	١٣	٢٨	٠,٠٦١٨	٠,٠٦٢٩
٨	٣	١٥	٠,٠٦٢٥	٠,٠٦٢٩

## تابع الجدول رقم (٤)

دالة المعلمات للمفردات في اختبار التكميل		رمز المفردة في اختبار الاختيار من متعدد		مسلسل
التكمل	الاختيار من متعدد	التكمل	الاختيار من متعدد	
٠,٠٦٢١	٠,٠٦٢٥	١٣	٦	٩
٠,٠٦٢١	٠,٠٦٢٧	٢٤	١	١٠
٠,٠٦٢١	٠,٠٦٢١	١٩	١٥	١١
٠,٠٦٢١	٠,٠٦٢١	١٤	٤	١٢
٠,٠٦٢٢	٠,٠٦٢١	٢٠	٢٩	١٣
٠,٠٦٢١	٠,٠٦٢٩	٢٥	٢٨	١٤
٠,٠٦٢١	٠,٠٦٢١	١٨	٢	١٥
٠,٠٦٢١	٠,٠٦٢١	٢١	٢٢	١٦
٠,٠٦٢٧	٠,٠٦٢٩	٢٢	٩	١٧
٠,٠٦٢٧	٠,٠٦٢٤	١٠	٢٥	١٨
٠,٠٦٢٥	٠,٠٦١٣	٢٦	٢١	١٩
٠,٠٦٢٥	٠,٠٥٩٧	١	٧	٢٠
٠,٠٦٢٥	٠,٠٥٩٠	٤	٢٠	٢١
٠,٠٦٠٣	٠,٠٥٧٤	٦	١١	٢٢
٠,٠٥٧٥	٠,٠٥٧٤	٩	٥	٢٢
-	٠,٠٤٠٥	-	٢٦	٢٤

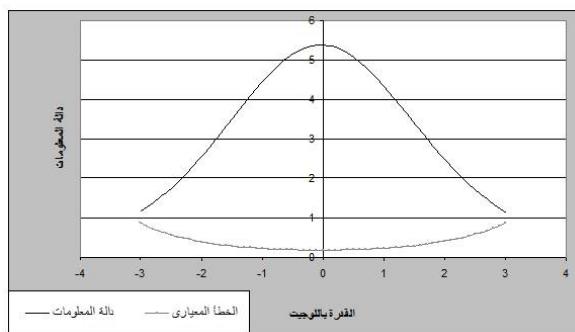
يتبيّن من خلال الجدول رقم (٤) أن هناك تفاوتاً في مقدار المعلمات التي تعطيها المفردات في الاختبارين، وتم حساب دالة المعلمات التي يعطيها الاختبار الكلي للاختبارين وقد بلغت (١,٤٤) و (١,٤٣) لاختبار الاختيار من متعدد والتكميل على الترتيب وتبيّن الأشكال (١) و (٢) دوال المعلمات لكل من الاختبارين



**الشكل رقم (١)**  
**دالة المعلمات لاختبار الاختيار من متعدد**

يتضح من خلال الشكل رقم (١) أن أعلى قدر من المعلمات يمكن الحصول عليه يتحدد عند مستوى قدرة (-٠,١) لوجيت وانخفاض الخطأ المعياري لتقدير القدرة في المدى (-٤,٠).

٤٤.). ويبين الشكل رقم (٢) دالة المعلومات لاختبار التكميل.



**الشكل رقم (٢)**  
**دالة المعلومات لاختبار التكميل**

يتضح من خلال الشكل رقم (٢) أن أعلى قدر من المعلومات يمكن الحصول عليه يتحدد عند مستوى قدرة (-١,٠) لوجيت وانخفاض الخطأ المعياري لتقدير القدرة في المدى (٠,٣-٠,٣+).

ومن خلال الشكلين رقم (١) و (٢) يتبيّن أن قيم دوال المعلومات لاختبار الاختيار من متعدد والتكميل كانت أكبر ما يمكن عند مستوى القدرة (-١,٠) لوجيت وهي قيمة قريبة جداً من الصفر بمعنى أن كلا الاختبارين يعطى معلومات أكثر فاعالية عند الأفراد ذوي القدرة المتوسطة بينما كانت قيم دوال المعلومات التي يقدمها اختبار الاختيار من متعدد والتكميل أقل ما يمكن عند مستويات القدرة العالية والمنخفضة كما أن قيمة دالة المعلومات تزداد تدريجياً بتزايد القدرة ( $\theta$ ) وتصل إلى أقصى قيمة لها عندما تكون قيمة القدرة صفر لوجيت أو قريبة من الصفر كما يلاحظ أن قيمة الخطأ المعياري للتقدير كانت أقل ما يمكن عند مستوى القدرة صفر إذ بلغت (٢,٠) لاختبار الاختيار من متعدد و (١٩,٠) لاختبار التكميل أي أن كمية المعلومات تزداد بنقصان الخطأ المعياري.

### ثانياً: عرض نتائج السؤال الثاني

نص السؤال الثاني على: ما مدى اختلاف متوسط تقديرات القدرة للأفراد على اختبار الاختيار من متعدد عن متوسط تقديرات القدرة للأفراد على اختبار التكميل؟ ولدراسة الفروق في متوازنات تقديرات القدرة للأفراد تم حساب متوسط تقديرات القدرة للأفراد على كل من اختبار الاختيار من متعدد واختبار التكميل وذلك بعد خوبل الدرجات الخام إلى تقديرات القدرة التي تناظرها مقدرة بوحدة اللوجيت إذ يقوم برنامج (Winsteps)

بحساب العلاقة بين الدرجة الخام والقدرة المحسوبة باللوجيت كما تم استخدام الإحصائي (t) لعينتين مرتبطتين لدراسة الفروق في تلك المتosteطات والجدول رقم (٥) يبين ذلك.

#### الجدول رقم (٥)

#### نتائج الاحصائي (t) لعينتين مرتبطتين لدراسة الفروق في متosteطات تقديرات الأفراد

مستوى الدلالة	درجات الحرية	(t)	قيمة (t)	الانحراف المعياري	الوسط الحسابي	العدد	الاختبار
٠,٠٠	٢٣١	٤,٤٧	٢,٠٦	٠,١٠٢	٢٢٢		التكامل
			١,٢٥	٠,٧٩٧	٢٢٢		ال اختيار من متعدد

يلاحظ من خلال الجدول أنه توجد فروق دالة إحصائية عند مستوى دلالة (٠,٠١٥) في متوسط تقديرات القدرة للأفراد على اختبار الاختيار من متعدد واختبار التكميل لصالح اختبار الاختيار من متعدد أى أن أداء الأفراد على اختبار الاختيار من متعدد كان أقل من أدائهم على اختبار التكميل ورما يكون ذلك نتيجة لعلامة الصدفة أو استخدام الطالب المعرفة الجزئية في الإجابة عن مفردات الاختيار من متعدد

### ثالثاً: عرض نتائج السؤال الثالث

نص السؤال الثالث على: ما مدى اختلاف القيمة التقديرية لثبات درجات اختبار الاختيار من متعدد عن القيمة التقديرية لثبات درجات اختبار التكميل؟ ولحساب دلالة الفروق في تقديرات معاملات الثبات تم الاعتماد في حساب الثبات على ما يوفره برنامج (Winsteps) وهو معامل مكافئ لمعامل كودر ريتشاردسون -٢٠ في النظرية التقليدية ويقوم حساب الثبات في النظرية التقليدية على أساس حساب خارج قسمة التباين المعيقي مقسوماً على التباين الكلي للأفراد على مفردات الاختبار، لذلك فهو يتعلّق بالأشخاص أكثر منه بالاختبار. (Linacre, 2003). أما مفهوم الثبات في نموذج راش فيشير إلى مدى الدقة في تقدير موقع كل من الأفراد والمفردات على متصل السمة الذي نهدف قياسها. وبين تحديد مدى دقة المفردات في تعريف هذا المتصل بإيجاد النسبة بين الانحراف المعياري للقيم التقديرية المتحررة لصعبوبة المفردات ومتوسط الخطأ المعياري لهذه القيم، وبطريق على هذه النسبة معامل الفصل بين المفردات (Item Separation Index) إذ يجب أن يكون قيمة هذا المعامل أكبر مما يمكن فإذا لم تزد هذه القيمة عن (٢) يصعب قياس التغيير بواسطة هذه المفردات. وقد وجد الباحث قيمة معامل الفصل بين المفردات في اختباري الاختيار من متعدد والتكميل بعد تدريجهما باستخدام نموذج راش (٢,٥) و (٣,٢٠) على الترتيب. ونظراً لأن هذه القيم تزيد

عن (٢) فإنّ مفردات كلا الاختبارين تعد كافية لتعريف متصل السمة الذي تقيسه. كذلك وجد أن قيمة معامل الفصل للأفراد في اختبار الاختيار من متعدد والتكميل (٢,٣٣) و (٢,٨) لهذا فإنّ عينة الأفراد تعد أيضاً كافية في الفصل بين المفردات ويمكن الحصول على معامل الثبات لكل من الأفراد والمفردات من خلال الصيغة الرياضية التالية:

$$R = \frac{G_p^2}{1+G_p^2}$$

حيث ( $G_i$ ) ترمز إلى معامل الفصل (Wright & Masters, 1982). وقد تم حساب قيمة معامل الثبات لكل من الأفراد والمفردات على اختبار الاختيار من متعدد والتكميل وفقاً للعلاقة السابقة والجدول رقم (١) يبين ذلك

### الجدول رقم (١)

**قيم معاملات الثبات لتقديرات صعوبة المفردات وتقديرات الأفراد  
لكل من اختباري الاختيار من متعدد والتكميل**

مستوى الدلالة	معامل الثبات لتقديرات القدرة للأفراد	معامل الثبات لتقديرات صعوبة المفردات	الاختبار
٠,٠٠	٠,٨٤	٠,٨٦	الاختبار من متعدد
٠,٠٠	٠,٨٩	٠,٩١	التكميل

يتضح من خلال الجدول رقم (١) أن قيمة معامل الثبات لتقديرات صعوبة المفردات لاختبار التكميل أعلى من قيمة معامل الثبات لتقديرات الصعوبة لاختبار الاختيار من متعدد والفرق بين معاملى الثبات (٠,٩١ - ٠,٨٦) مقابل (٠,٨٩ - ٠,٨٤) فقد بلغت قيمة الاختبار الإحصائي (W) المحسوبة (١,٥١) وهي أعلى من قيمة (F) المحرجة (١,٢) وكذلك فإن الفرق بين معاملى ثبات قدرات الأفراد على اختباري التكميل والاختبار من متعدد (٠,٨٦ - ٠,٨٤) دال إحصائياً حيث بلغت قيمة الإحصائي (W) المحسوبة (١,٤٥) وهي أعلى من قيمة (F) المحرجة . وبshire ذلك إلى أن الدقة في القياس لاختبار التكميل أعلى منها لفردات اختبار الاختيار من متعدد وربما يعود ذلك إلى مصادر الخطأ في مفردات الاختيار من متعدد أكثر منها في مفردات التكميل وعلى رأس تلك المصادر التخمين العشوائي . وتتفق نتائج هذه الدراسة مع نتائج كل من دراسة الكحلوت (٢٠٠٢) والردادي (٢٠٠١) وبريجمان (Bridgeman, 1994) التي بينت أن ثبات اختبارات التكميل أعلى من ثبات اختبارات الاختيار من متعدد وينعكس هذا على قيمة الخطأ المعياري للقياس فهو في مفردات الاختيار من متعدد أعلى منه في مفردات التكميل

وبينج هذا النقصان في الخطأ المعياري عن أمرین هما : انخفاض قيمة الانحراف المعياري للدرجات على مفردات التكميل. وارتفاع قيمة معامل الثبات لمفردات التكميل.

### الاستنتاج والتوصيات

في ضوء ما أسفرت عنه هذه الدراسة من نتائج، فإنها توفر الأدلة التي تؤيد استعمال مفردات التكميل على نطاق أوسع مما هو عليه الآن. وتشير النتائج إلى الفعالية العالية لمفردات التكميل في قياس تحصيل الطلبة وتنزع العلامات على مفردات التكميل لأن تكون أعلى ثباتاً من العلامات على مفردات الاختيار من متعدد. وعليه يوصى الباحث بإجراء دراسة ماثلة باستخدام النموذج ثلاثي المعلم لدراسة بعض العوامل المؤثرة في الخصائص السيكومترية للمفردة مثل التخمين العشوائي والقدرة التمييزية لمفردات.

### المراجع

الأحمدى، عبد الرحمن (٢٠٠٩). *أثر اختبارات الاختيار من متعدد والإجابة القصيرة في الاحتفاظ بنوافذ التعلم*. رسالة ماجستير غير منشورة. جامعة أم القرى، مكة المكرمة، المملكة العربية السعودية.

الحموري، هند عبد المجيد (١٩٩٩). *أثر اختلاف نمط الاختبار حسب نوعية الاستجابة المطلبة من مفرداته: منشأة أو مختاراة في صدقه العامل والتنبؤ*. مجلة كلية التربية، جامعة الإمارات العربية المتحدة، ١١(١٤)، ١٠٣-١٤٠.

الردادى، عبد العالى (٢٠٠٦). *أثر اختلاف نمط الاختبار على الخصائص السيكومترية للاختبار ومفرداته*. رسالة ماجستير غير منشورة. جامعة أم القرى، المملكة العربية السعودية.

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

علام، صلاح الدين محمود (١٩٩٥). *الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية* (ط١). القاهرة: دار الفكر العربي.

علام، صلاح الدين محمود (٢٠٠٠). *القياس والتقويم التربوي والنفسي. أساسياته وتطبيقاته وتوجهاته المعاصرة* (ط١). القاهرة: دار الفكر العربي.

عودة، أحمد سليمان (١٩٩٩). *القياس والتقويم في العملية التدرسية* (ط١). الإصدار الثالث. إربد: دار الأمل.

كاظم، أمينة محمد، وإسماعيل، عماد الدين، ورمزي ناهد، وكرم، ليلى، وناشف، هدى (١٩٩٤). *تدريج ومعايير المقاييس. معايير نمو طفل ما قبل المدرسة*. ص ١١٤-٢٣٢. القاهرة: المجلس القومي للأمومة والطفولة.

كاظم، أمينة محمد (١٩٩٦). استخدام نموذج راش في بناء اختبار خصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج. في أنور الشرقاوي، وسليمان الخضري الشيخ، وأمينة محمد كاظم، ونادية محمد عبد السلام، ص ٤٧٥-٥٤٥. إنجاهات معاصرة في القياس والتقويم النفسي والتربوي. القاهرة: مكتبة الأبللو المصرية.

كاظم، أمينة محمد (١٩٩٦). استخدام نموذج راش في بناء اختبار خصيلي في علم النفس وتحقيق التفسير الموضوعي للنتائج. في أنور الشرقاوي، وسليمان الخضري الشيخ، وأمينة محمد كاظم، ونادية محمد عبد السلام، إنجاهات معاصرة في القياس والتقويم النفسي والتربوي. القاهرة: مكتبة الأبللو المصرية.

الكلحوت، أحمد إسماعيل (٢٠٠٢). مقارنة بين المصادر السيكومترية لكل من اختبارات الاختبار من متعدد واختبارات التكميل. مجلة مركز البحوث التربوية، جامعة قطر، (٢٢)، ١٥٣-١٢٧

مراد، صلاح محمد، وسليمان، أمين على (٢٠٠٢). الاختبارات والمقياس في العلوم النفسية والتربوية- خطوات إعدادها وخصائصها. القاهرة : دار الكتاب الحديث.

Allen, M. & Yen, W. (1979). Introduction To measurement theory. **California Education**, 20(4), 44-57. Brooks Cole Publishing Company Monterey.

Alswalmeh, Y. & Feldt, L. (1994). Testing the equality of two related intraclass reliability coefficients. **Applied Psychological Measurement**, 18(3), 183-190

Anastasi, A. (1982). **Psychological testing** (6<sup>th</sup> ed.). New York: Macmillan.

Bridgeman, B. (1994). A comparison of quantitative questions in open –ended and multiple choice format. **Journal Of Educational Measurement**, 29(3), 253-271.

Crocker, L. & Algina, J. (1986). **Introduction to classical and modern test theory**. New York: Holt, Rinehart And Winston.

Elliot, C. (1983). **Britishability scale manual, introductory handbook**, Windsor. England: National Foundation For Education Research (NFEER).

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

Fanxita O. (1998). Item Response Theory and classical test theory. an empirical comparaison of their item/ person statics. **Educational and Psychological Measurement**, 58(3), 357- 381.

Feldt, L. (1980). A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. **Pschometrika**, 49, 99-105.

- Ferguson, G. & Takane, Y. (1989). **Statistical analysis in psychology and education** (6<sup>th</sup> ed.). New York: McGraw-Hill
- Gadalla, T. (1999). **Multiple choice versus constructed response tests. in the assessment of mathematics computation skills.** (ERIC Document Reproduction Service .ED431813).
- Gronlund, N. (1977). **Constructing achievement tests** (2nd ed.). London: Prentice Hall.
- Hambleton, R. & Swaminthan, H. (1985). **Item response theory: Principles & applications.** Norwell, MA: Kluwer Nijhoff Publishing
- Harris, D. (1989). Comparison of 1, 2, and 3 parameter I.R.T. models. **Educational measurement**, (8), 35-41
- Hashway, R. (1998). **Assessment and evaluation of developmental learning: Qualitative individual assessment and evaluation models.** London, West port: Connecticut's Praiger Publishers.
- Linacre, J. (2003). **User's guide & program manual. winsteps minstep. rasch -model computer programs.** Retrieved on March 20 2010, From: [www.winsteps.com](http://www.winsteps.com).
- Manhart, J. (1996). **Factor analytic methods for determining whether multiple choice and constructed response tests measure the same construct.** Paper presented at the Annual meeting of the National Council on Measurement in Education, New York, April 9-11
- Nitko, A. (2001). **Educational assessment of student**, (3<sup>rd</sup> ed.). New Jersy: Merrill Pentice Hall.
- Suen, H. (1990). **Principles of test theories.** Hillsdale: New Jersey.
- Thorndike, R. L. (1982). **Applied psychometrics.** London: Houghton Mifflin Company Boston.
- Tissen, D., Wainer, H. & Bowang, X. (1994). Are test comprising both multiple choice and free response items necessarily less uni-dimensional than multiple choice tests? An analysis of two tests. **Journal of educational measurement**, 31(2).113-123.
- Thissen, D. & Wainer, H. (2001). **Test scoring.** New Jersey: Lawrence Erlbaum Associates publishers.
- Wright, B. & Linacre, J. (1998). **A User's guide to Winsteps/bigsteps, Version 2.88.** Chicago, U.S.A: MESA Press .

Wright, B. & Masters, G. (1982). **Rating scale analysis: Rasch measurement.**  
Chicago: Meta. Press.