281
المجلد 13 العدد 3 سبتمبر 2012

أثر طريقة تصحيح اختبار الصواب - الخطأ المتعدد على الخصائص السيكومترية للاختبار وفقراته

د. مروان صالح الصمادي
قسم التربية الخاصة
كلية العلوم التربوية-جامعة إربد الأهلية

أثر طريقة تصحيح اختبار الصواب - الخطأ المتعدد على الخصائص السيكومترية للاختبار وفقراته

د. مروان صالح الصمادي
قسم التربية الخاصة
كلية العلوم التربوية-جامعة إربد الأهلية

الملخص

هدفت هذه الدراسة للكشف عن أثر طريقة تصحيح اختبار الصواب والخطأ المتعدد، في معاملات الصدق و الثبات للاختبار، ومعاملات الصعوبة والتمييز لفقرات الاختبار، وكذلك أثرها في تحصيل الطلبة، ولتحقيق هذه الأهداف تم بناء اختبار تحصيلي لمساق القياس والتقويم، يتكون من (٢٥) فقره صواب وخطأ متعدد لكل منها أربعة بدائل. وطبق الاختبار على (١٢٢) طالب وطالبة في جامعة اربد الأهلية، تم تصحيح الاختبار باستخدام أربع طرق مختلفة للتصحيح متدرجة في مراعاتها للمعرفة الجزئية، وباستخدام برمجية (SPSS) تم حساب معاملي صدق للاختبار، ومعامل الثبات كرونباخ ألفا، وكذلك حسبت معاملات الصعوبة والتمييز لفقرات الاختبار، ومتوسطات التحصيل، ولكل طريقة من طرق التصحيح المختلفة. وأظهرت نتائج هذه الدراسة عدم وجود فروق ذات دلاله إحصائية بين معاملات الصدق والثبات والتمييز، في حين أشارت النتائج لوجود فروق ذات دلاله 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 تعطي العلامة الأعلى.

الكلمات المفتاحية: طرق التصحيح، اختبار الصواب- الخطأ المتعدد، الخصائص السيكومترية، اختبار، فقرة، ثبات، صدق، تمييز، صعوبة.

* تاريخ تسلم البحث: ٢٠١٠/٣/١٦م
* تاريخ قبوله للنشر: ٢٠١٠/١٢/١٣م

The Effect of Multiple True - False Tests Scoring Method on the Psychometric Properties of a Test and its Items

Dr. Marwan S. Al-Smadi
Irbid National University
Faculty of Educational Since

Abstract

The purpose of the study was to examine the effect of the multiple true-false test scoring method on test validity, reliability, item-difficulty and discrimination coefficients, as well as on examinees achievement. To achieve this goal, a multiple true- false measurement and evaluation test, consisting of (25) items of four choices each, was developed and administered to (122) students at Irbid National University. The test was scored by using four different methods that varied in the degree of considering partial knowledge. By using SPSS, two validity coefficients, Cronbach alpha, item-difficulty and discrimination coefficients, and achievement means were computed for each scoring method. The findings of the study indicated that there were no statistically significant differences in the test validity, reliability, and discrimination coefficients. Whereas, there were statistically significant differences in the item-difficulty and examinees achievement: The fourth scoring method in which partial knowledge was most considered was the most effective method for reference test achievement, followed by the third method, then the second and the last was the first method. This down ranking showed that the fourth method was the easiest one since it gave the highest score.

Key words: scoring method, multiple true-false, test, psychometric properties, item, reliability, validity, discrimination, difficulty.

أثر طريقة تصحيح اختبار الصواب - الخطأ المتعدد على الخصائص السيكومترية للاختبار وفقراته

د. مروان صالح الصمادي
قسم التربية الخاصة
كلية العلوم التربوية-جامعة إربد الأهلية

المقدمة

تعد الاختبارات من أدوات القياس التربوي المهمة فهي متعددة الأغراض والأهداف ،وتؤدي إلى تحسين عملية اتخاذ القرارات التدريسية ،وتحسن دافعية التلاميذ، وتزيد مستوى الاحتفاظ بالمعلومات، وانتقال أثر التعلم، وتوفر التغذية الراجعة عن فعالية العملية التدريسية. لذلك ينبغي أن تتسق الاختبارات التحصيلية مع الأهداف التدريسية، وأن تمثل المادة الدراسية، وأن تتمتع بدرجة أكبر من الثبات، وأن تصاغ بطريقة تخدم الغرض من خلال انتقاء النوع المناسب من الأسئلة. من هنا فإن قياس التحصيل بشكل فعال يتأثر باختيار نوعية الأسئلة التي من شأنها أن تستثير الإجابة المطلوبة وتستبعد الإجابات الأخرى التي ليس لها صلة بالإجابات الصحيحة (عدس، ١٩٨٨).

وللاختبارات التحصيلية أشكال متعددة، منها فئة الاختبارات الكتابية، وفئة الاختبارات غير الكتابية، وتتضمن الاختبارات الكتابية، الاختبارات ذات الإجابة القصيرة (المقيدة) والاختبارات ذات الإجابة الحرة الطويلة (المفتوحة) واختبارات المزاوجة Matching، واختبارات الصواب والخطأ (ص-خ) True-False، واختبارات الاختيار من متعدد (خ-م) Multiple-Choice، واختبارات الصواب - الخطأ المتعدد (ص-خ-م) Multiple True-False.

ونظراً لتعدد أنواع الاختبارات فإن من الأهمية بمكان تحديد الأسس التي يقوم عليها اختيار النوع المناسب منها. ويعتمد تحديد نوع الأسئلة الذي يعتبر ملائما للاختبار إلى حد كبير على طبيعة المحتوى المراد قياسه، فهناك بعض المواد الدراسية يناسبها نوع معين من الأسئلة ولا يناسبها نوع آخر، كما يعتمد نوع الأسئلة أيضاً على الهدف المراد قياسه، فإذا كان الهدف يتطلب من الطالب، أن يعبر بلغته الخاصة عن مفهوم معين، فهنا يكون من الملائم استخدام الأسئلة الحرة (المفتوحة)، أما إذا كان الهدف يتطلب من الطالب أن يقارن بين عدد من الإجابات المعطاة وأن يختار من بينها الإجابة الصحيحة، فهنا يكون من الملائم استخدام أسئلة الاختيار من متعدد، كما أن مهارة المعلم في كتابة نوع من الأسئلة، لها دور

في تحديد نوعية الأسئلة المناسبة (ثورندايك وهيجن، ١٩٨٦).

ويرى عودة (٢٠٠٥) بأن اختيار نوعية الأسئلة، يعتمد كذلك على عدد الطلبة الذين سيطبق عليهم الاختبار، والزمن المتوفر لإعداد الاختبار، والزمن المتوفر للتصحيح، فعندما يتساوى تأثير هذه العوامل، فإن اختيار نوع الأسئلة الملائم يعتمد على الخصائص السيكومترية للاختبار، التي تتعلق بكل نوع من الأسئلة، فعلى المعلم أن يختار الأسئلة التي يتوفر لها معاملات صدق وثبات مرتفعة والتي هي من الصفات الأساسية المرغوب توفرها في الاختبار، حيث يزيد توفر هذه الصفات من تحقيق الاختبارات لأهداف القياس التربوي، كما أن عمر المفحوص وسهولة التطبيق، والتصحيح، وعامل الكلفة للاختبار، وتوفر الصور المتكافئة للاختبار ووجود المعايير هي من العوامل التي تدفع باتجاه اختيار نوع معين من الأسئلة دون نوع آخر.

ويتكون اختبار الصواب والخطأ المتعدد-موضوع الدراسة الحالية-من مجموعة من الفقرات التي تشبه فقرات الاختيار من متعدد من حيث الشكل، لكن بدل اختيار إجابة واحدة صحيحة من بين عدة بدائل، يقوم المفحوص بوضع إشارة صح أو خطأ أمام كل بديل من البدائل في الفقرة الواحدة، ويمكن أن لا يكون هناك أي بديل صحيح أو أن تكون جميع البدائل صحيحة، أو أن بعض البدائل صحيحة وبعضها خطأ ضمن الفقرة الواحدة (Eble & Frisbie, 1986). وتدعى أحياناً بالفقرات العنقودية، وتتكون الفقرة العنقودية من عدة فقرات من نوع الصواب والخطأ التي تشترك بنفس المتن، وبذلك فهي أسلوب آخر من الفقرات التي تجمع بين خصائص فقرات الاختيار من متعدد وخصائص الصواب والخطأ وخصائص أخرى لا تتوفر في كليهما (Frisbie & Sweeny, 1982)، و يمكن أن تعالج فكرة واحدة أو عدة أفكار مترابطة (سوالمه، ١٩٩٤).

مثال (١): من أمثلة الفقرات ذات الإجابة المنتقاة.

ص خ أ- فقرات التكميل

ص خ ب- الإجابة القصيرة

ص خ ج- الإنشائية المحددة

ص خ د- الإنشائية المفتوحة

مثال (٢): من الأمور التي قد تساعد الطالب على الغش

ص خ أ- المبالغة بأهمية الامتحان

ص خ ب- السماح بإدخال أوراق الملاحظة مع الطالب

ص_خ ج- سوء التنظيم لإدارة الاختبار

ص_خ د- شعور الطالب بعدم المعاقبة المرتبه على الغش

وتمتاز أسئلة الصواب والخطأ المتعدد بأنها سهلة الأعداد، وتقلل من درجة غموض الفقرة وفرص الحصول على علامات مرتفعة نتيجة مراعاة بعض طرق التصحيح للمعرفة الجزئية، مما يؤثر إيجابا على صدق الاختبار وثباته، وهناك إمكانية للتنويع في عدد البدائل الصحيحة في الفقرة، مما يمكن من التغلب على بعض عيوب فقرات الاختيار من متعدد (Eble & Frisbie, 1986).

وقد زاد الاهتمام باختبارات الصواب-الخطأ 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 (Frisbie, 1992). إذ أشارت الدراسات (Albanese & Sabers, 1988; Eble & Frisbie, 1986) إلى أن اختبار الصواب والخطأ المتعدد، يتمتع بصدق تلازمي وثبات أعلى مقارنة باختبار الاختيار من متعدد. كما يتفوق عليه في الاحتفاظ بالتعلم لمدة أطول (المومني، ١٩٩٦).

وقد أجريت العديد من الدراسات التي هدفت مقارنة اختبار الصواب والخطأ المتعدد بأنواع أخرى من الاختبارات كالاختيار من متعدد والصواب والخطأ وغيره (Frisbie & Sweeney, 1982), (Albanese & Sabers, 1988), (Erickson, 1988), (Kreiter & Frisbie, 1989) وقدمت مراجعة فرزبي (Frisbie, 1992) للأدب التربوي المرتبط بفقرات الصواب والخطأ المتعدد تلخيصاً لخصائص هذه الفقرات، كما وفرت إجابات لعدد من الأسئلة المتعلقة بهذا النوع من الفقرات مقارنة بفقرات الاختيار 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 (ص خ م) قد يكفي لأكثر من أربع فقرات (خ م). وفيما يتعلق بالثبات فقد تفوقت فقرات (ص خ م) على فقرات (خ م) بغض النظر عن طريقة التصحيح، وبالنسبة للصدق التلازمي؛ كشفت مراجعة معاملات الارتباط بعد التصحيح لأثر الانخفاض أنه يمكن استخدام فقرات (ص خ م) وفقرات (خ م) على نحو تبادلي إذْ تراوحت معاملات الارتباط المصححة بين (٠,٩١) و(١,٠). وبخصوص صعوبة الفقرات كانت النتائج متفاوتة وغير مستقرة.

وهدفت دراسة أمريزيق (١٩٩٢) مقارنة اختبار الاختيار من متعدد واختبار الصواب والخطأ واختبار الصواب و الخطأ المتعدد من حيث الثبات ومتوسط الصعوبة للاختبار ومتوسط تمييزه وشكل منحنى توزيع أداء المفحوصين على الاختبار. وقد استخدم طريقة الاستجابة المتعددة لتصحيح فقرات الصواب والخطأ المتعدد، أي يُعطى المفحوص العلامة الكاملة (١) على الفقرة عندما يجيب بصورة صحيحة عن جميع بدائلها، ويعطى العلامة صفر عندما يخفق في إجابة أي بديل من تلك البدائل. وقد أشارت النتائج إلى وجود فروق ذات دلالة إحصائية بين معاملات الثبات لصالح اختبار الاختيار من متعدد، كما لم توجد فروق بين معاملي ثبات اختبار الصواب والخطأ، واختبار الصواب و الخطأ المتعدد. كما أشارت النتائج إلى عدم وجود فروق ذات دلالة إحصائية في معامل الصدق. أما من حيث الصعوبة فقد كان الفرق دالاً إحصائياً بين معاملي صعوبة اختبار الاختيار من متعدد واختبار الصواب والخطأ المتعدد ولم يكن هناك فرق دال بين معاملي الصعوبة لاختبار الصواب والخطأ واختبار الاختيار من متعدد. وبالنسبة للتمييز، لم يكن هناك فروق بين متوسطات تمييز الاختبارات الثلاثة. أما من حيث معاملات الالتواء، فقد كانت جميع معاملات الالتواء موجبة وكان معامل الالتواء لاختبار الاختيار من متعدد هو الأقرب إلى الصفر مقارنة بالاختبارين الآخرين.

وأجرى سوالمة (١٩٩٤) دراسة هدفت مقارنة فقرات الصواب والخطأ المتعدد مع فقرات الاختيار من متعدد الموازية لها في المحتوى وعدد البدائل من حيث الثبات والصدق التلازمي والصعوبة. وأخذت الدراسة متغير عدد البدائل الصحيحة في فقرات الصواب والخطأ المتعدد بالاعتبار. وأشارت نتائج الدراسة إلى تمتع اختبارات الصواب والخطأ المتعدد بفعالية عالية في قياس تحصيل الطلبة للمهارات والقدرات التي تقيسها اختبارات الاختيار من متعدد الموازية لها في المحتوى. وقد كانت فقرات الاختيار من متعدد أسهل من فقرات الصواب والخطأ المتعدد وكانت معاملات الارتباط بين الاختبارين عالية نسبياً، وتميل العلامات على اختبارات الصواب والخطأ المتعدد لتكون أعلى ثباتاً من العلامات على اختبارات الاختيار من متعدد ويكون الفرق في الثبات له دلالته الإحصائية وبشكل خاص عندما يكون هناك أكثر من بديل صحيح في فقرة الصواب والخطأ المتعدد.

وقد استخدم داوننغ وبرانوسكي وجرس ونو رسني(Downing, Baranowski, Gross & Norcini, 1995) فقرات الصواب والخطأ المتعدد في المجالات الطبية، وقاموا بمقارنة الخصائص السيكومترية لكل من فقرات الصواب والخطأ المتعدد وفقرات الاختيار من متعدد. وأظهرت النتائج أن فقرات الصواب والخطأ المتعدد أكثر ثباتاً من فقرات الاختيار من متعدد.

بينما تحقق لفقرات الاختيار من متعدد درجة أعلى من الصدق المرتبط بمحك فقد كانت أكثر ارتباطاً بمقياس مستقل للأداء.

وقد أجرى المومني (١٩٩٦) دراسة هدفت مقارنة اختبارات الاختيار من متعدد مع اختبار الصواب والخطأ المتعدد من حيث تأثيرها في احتفاظ طلاب الصف الثاني الثانوي الأكاديمي بمادة قواعد اللغة الإنجليزية في مستويي المعرفة والفهم. وأظهرت نتائج الدراسة أن متوسط علامات مجموعة الصواب والخطأ المتعدد كانت أعلى رقمياً من متوسط مجموعة الاختيار من متعدد إلا أن هذا الفرق لا يرقى لمستوى الدلالة الإحصائية.

وقد كانت أسئلة الصواب والخطأ المتعدد مداراً للاهتمام في مجال الاختبارات المتعلقة باللغة الثانية.فقد قارن دودلي (Dudley, 2006) بين اختبارات الصواب والخطأ المتعدد واختبارات الاختيار من متعدد في مجالي المفردات والقراءة باللغة الإنجليزية.وقد حللت البيانات باستخدام كل من النظرية التقليدية في القياس ونموذج راش. وقد بينت النتائج تحسناً في الأداء عند استخدام فقرات الصواب والخطأ المتعدد. وقد استجاب المشاركون لعدد أكبر من فقرات الصواب والخطأ المتعدد مقارنة بفقرات الاختيار من متعدد. كما كشف التحليل عن استقلالية الفقرات وتحسن الثبات في مجال القراءة. وقد وفرت نتائج هذه الدراسات دعماً إضافياً لاستخدام اختبارات الصواب والخطأ المتعدد في قياس التحصيل. كما أظهرت النتائج تفوقها من حيث الصدق والثبات ومتوسط صعوبة الفقرات وتمييزها، على أنواع اخرى من الفقرات.

أما من حيث طريقة التصحيح فيؤخذ على فقرات الصواب والخطأ المتعدد أنه لا يوجد اتفاق على طريقة تصحيحها، إذ يمكن تصحيح فقرات الصواب والخطأ المتعدد بعدة طرق ربما يكون لها أثار مختلفة على خصائص الاختبار وخصائص الفقرات نفسها.

ويقصد بطريقة التصحيح: تحديد القاعدة التي تعطى على أساسها قيماً عددية لاستجابة المفحوص لفقرات الاختبار ،أي أنها طريقة لتحويل المعلومات التي نحصل عليها من استجابات المفحوصين إلى مقادير كمية ،تعكس مستوى أداء المفحوصين على الاختبار (جرادات وسواقد، ١٩٨٥) (Lord & Novik, 1968).

وقد درس البانس وسابرس (Albanese & Sabers, 1988) أربع طرق يمكن استخدامها في تصحيح الفقرات ذات الأربعة بدائل، بهدف دراسة الارتباطات الداخلية لفقرات (ص خ م) لتحديد إلى أي مدى يمكن التعامل مع خيار الصواب والخطأ باعتباره مستقلاً.

كما قام تيساي وسوين (Tsai & Suen, 1993) بمقارنه ست طرق لتصحيح فقرات

الصواب – الخطأ المتعدد تعتمد على عدد البدائل k، وعدد الإجابات الصحيحة i، وعدد البدائل المتروكة بدون إجابة j، وهذه الأساليب هي:

١. Multiple-response (MR) method: (ط١): وتتمثل هذه الطريقة بإعطاء علامة كاملة لكل فقرة إذا أجاب الطالب بشكل صحيح عن البدائل جميعها، وصفر إذا لم يجب بشكل صحيح عن جميع البدائل، ويتم ذلك وفق القاعدة:

$$f(i)=\begin{cases}1 & \text{if } i=k;\\ 0 & \text{otherwise}\end{cases}$$

٢. Count for 3 options correct (Count 3) method: (ط٢): وتتمثل هذه الطريقة بإعطاء نصف علامة إذا أجاب بشكل صحيح عن ثلاثة بدائل وعلامة كاملة إذا أجاب عن البدائل الأربعة بشكل صحيح، وصفر فيما عدا ذلك. ويتم ذلك وفق القاعدة:

$$f(i)\begin{cases}0 & \text{if } i=0 \text{ orl or2};\\ \dfrac{i-2}{k-2} & \text{otherwise}\end{cases}$$

٣. Count for 2 options correct (Count 2) method: (ط٣): وتتمثل هذه الطريقة بإعطاء ثلث علامة إذا أجاب عن بديلين بشكل صحيح، وثلثي علامة إذا أجاب عن ثلاثة بدائل بشكل صحيح، وعلامة كاملة إذا أجاب عن البدائل الأربع بشكل صحيح، وصفر فيما عدا ذلك. ويتم ذلك وفق القاعدة:

$$f(i)=\begin{cases}0 & \text{if } i=0;\\ \dfrac{i-1}{k-1} & \text{otherwise}\end{cases}$$

٤. multiple true - false (MTF) method: (ط٤): وتتمثل هذه الطريقة بإعطاء ربع علامة لكل بديل يتم الإجابة عنه بشكل صحيح، ويتم ذلك وفق القاعدة الآتية:

$$f(i)=\frac{i}{k} \qquad (k\geq i\geq 0)$$

٥. Correction for guessing (CFG) method: وتتمثل هذه الطريقة باستخدام طريقة التصحيح لأثر التخمين، ويتم ذلك وفق القاعدة التالية:

$$f(i)=\frac{i-(k-i)}{k} \qquad (k\geq i\geq 0)$$

٦. Let - omit (LO) method: وهنا يحاسب الممتحن على الإجابة الصحيحة والإجابة الخاطئة فقط، ويتم ذلك وفق القاعدة الآتية:

$$f(i)=\frac{i-(k-i-j)}{k} \qquad (k\geq i\ j\geq 0)$$

وقد أظهرت نتائج هذه الدراسة (Tsai & Suen, 1993) أن الفروق في قيمة الثبات والتمييز لم تكن ذات دلالة إحصائية أما من حيث الصعوبة، فقد تبين أن أسلوب الصواب والخطأ المتعدد (MTF) وأسلوب تصحيح التخمين (CFG) وأسلوب دعها ـ احذفها (let-omit) تعطي أوساطاً حسابية أعلى للفقرات (أي أسهل) من الأساليب الأخرى.

وتعد دراسة براد (Brad, 2003) المتعلقة بتصحيح فقرات الصواب والخطأ المتعدد ضمن الدراسات التي اهتمت بدراسة طرق التصحيح الأربعة الأولى ـــ المستخدمة في دراسة تيساي وسوين ـــ فقد اهتمت دراسته بمقارنة هذه الطرق من خلال فحص العلاقة بين العلامات الخام الناتجة عن استخدام كل منها وتقديرات القدرة المقدرة بحسب النموذج الثلاثي من نماذج نظرية استجابة الفقرة (Item response theory). وتبين من النتائج أنه توجد علاقة موجبة شبه تامة بين العلامات الخام وتقديرات القدرة في الأساليب الأربعة. فقد تراوحت قيم معاملات ارتباط بيرسون بين (٠,٩٧ و٠,٩٩) على مستوى الاختبار المتكون من خمس فقرات، وبين (٠,٧٨ و ٠,٩٨) على مستوى الفقرة الواحدة. كذلك أشارت النتائج إلى عدم اختلاف الثبات للعلامات باختلاف أسلوب التصحيح.

وكذلك دراسة الصمادي وسوالمة (٢٠٠٩) التي هدفت 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 اربد الأولى. وقد صححت أوراق الإجابة باستخدام أربع طرق للتصحيح متدرجة في مراعاتها للمعرفة الجزئية-وهي الطرق الأربعة الأولى المستخدمة في تيساي وسوين (Tsai & Suen, 1993). وباستخدام برمجية (BIGSTEPS) تم الحصول على إحصاءات المطابقة وتقديرات لقدرات الأفراد و صعوبة الفقرات والأخطاء المعيارية في تقديرها بحسب طرق التصحيح المختلفة بعد استبعاد الأفراد والفقرات غير المطابقين لنموذج التقدير الجزئي. أظهرت النتائج أن طريقة التصحيح الرابعة – الأكثر مراعاة للمعرفة الجزئية – أكثر دقة في تقدير قدرات الأفراد و صعوبة الفقرات، تلتها طريقة التصحيح الثالثة، ثم الثانية، فالأولى.

وهكذا يبدو واضحاً أن معظم الدراسات السابقة قد اهتمت بدراسة الخصائص السيكومترية النسبية لفقرات الصواب والخطأ المتعدد مقارنة مع أشكال أخرى للفقرات، وقد وفرت نتائج هذه الدراسات دعماً إضافياً لاستخدام اختبارات الصواب والخطأ المتعدد بديلاً

لاختبارات الاختيار من متعدد في قياس التحصيل. وأن هناك عدداً قليلاً جداً-في حدود ما توفر للباحث-من الدراسات التي اهتمت بطرق تصحيح فقرات الصواب والخطأ المتعدد. إذ اهتمت دراسة تيساي وسوين (Tsai & Suen, 1993) بأثر طريقة التصحيح في كل من الثبات للاختبار وصعوبة الفقرات وتمييزها بحسب النظرية التقليدية للقياس، لكنها لم تدرس أثرها في صدق الاختبار أو في مستوى تحصيل الطلبة، وهذا ما تضيفه الدراسة الحالية-بينما درس براد (Brad, 2003) أثر طريقة التصحيح لفقرات الصواب والخطأ على الصدق والثبات حسب النظرية الحديثة للقياس، كما درس الصمادي وسوالمة (٢٠٠٩) أثر طريقة تصحيح فقرات الصواب والخطأ المتعدد في دقة تقدير قدرات الأفراد وصعوبة الفقرات باستخدام نموذج التقدير الجزئي. كما يلاحظ تباين النتائج من حيث الثبات في دراستي (Albanese & Sabers, 1988; Tsai & Suen, 1993).

مشكلة الدراسة

في العقدين الماضيين زاد الاهتمام بفقرات الصواب والخطأ المتعدد، واتسع مجال استخدامها، فقد دعمت ذلك نتائج العديد من الدراسات التي اهتمت بدراسة خصائصها السيكومترية كالصدق والثبات ومعاملات الصعوبة والتمييز مقارنة بفقرات الاختيار من متعدد، ورافق استخدام فقرات الصواب والخطأ المتعدد، مشكلات تتعلق بطريقة تصحيحها حيث توجد أكثر من طريقه لتصحيح فقرة الصواب والخطأ المتعدد. وفي ظل تعدد طرق التصحيح لفقرات الصواب والخطأ المتعدد، كان لا بد من مقارنة هذه الطرق، لتحديد الطريقة الأنسب منها للحصول على اختبار يتمتع بخصائص أفضل من حيث الصدق والثبات، وعلى فقرات تتمتع بخصائص سيكومترية أفضل من حيث الصعوبة والتمييز.

أسئلة الدراسة

ستتم الإجابة من خلال هذه الدراسة عن الأسئلة الآتية بالتحديد:

١. هل تختلف معاملات الصدق المحكي لاختبار الصواب والخطأ المتعدد باختلاف طريقة التصحيح عند مستوى الدلالة (α =٠,٠٥)؟

٢. هل يختلف معامل الثبات لاختبار الصواب والخطأ المتعدد باختلاف طريقة التصحيح عند مستوى الدلالة (α =٠,٠٥)؟

٣. هل يختلف معامل التمييز لفقرات اختبار الصواب والخطأ المتعدد باختلاف طريقة

التصحيح عند مستوى الدلالة (α =٠,٠٥)؟

٤. هل يختلف معامل الصعوبة لفقرات اختبار الصواب والخطأ المتعدد باختلاف طريقة التصحيح عند مستوى الدلالة (α =٠,٠٥)؟

٥. هل يختلف تحصيل الطلبة في اختبار الصواب والخطأ المتعدد باختلاف طريقة التصحيح عند مستوى الدلالة (α =٠,٠٥)؟

هدف الدراسة

تهدف الدارسة تطبيق اختبار صواب وخطأ متعدد، على عينه من الأفراد، واستخدام أربع طرق لتصحيحه، لتحديد أثر كل من هذه الطرق في خصائص الاختبار وفقراته وكذلك في مستوى تحصيل الطلبة، ومحاولة التعرف إلى أنسب تلك الطرق التي تجعل من الاختبار أفضل ما يمكن من حيث الصدق والثبات. وتجعل من الفقرة أفضل ما يمكن من حيث الصعوبة والتمييز.

أهمية الدراسة

تأتي أهمية هذه الدراسة من أنها:

١- ستوفر بعض الأدلة التي قد تدعم وتبين الأسلوب الأنسب من هذه الأساليب للحصول على اختبار يتمتع بخصائص أفضل من حيث الصدق والثبات، وعلى فقرات تتمتع بخصائص سيكومترية أفضل من حيث الصعوبة والتمييز.

٢- كما يمكن أن توفر نتائج الدراسة المعلومات الكافية، لمستخدمي هذا النمط من الاختبارات سواءً في حقل التدريس أو في غيره، حول مزايا هذا النمط، وحول أنسب طرق تصحيح فقرات الصواب والخطأ المتعدد الضرورية لاختيار الفقرات وفق خصائصها السيكومتريه.

٣- وقد تقدم هذه الدراسة تأكيداً ودعماً إضافياً لاستخدام اختبارات الصواب والخطأ المتعدد وزيادة الوعي بها والتشجيع على استخدامها، بديلاً لاختبارات الاختيار من متعدد وغيرها في قياس التحصيل. والتي تؤدي في بعض الأحيان إلى تدني مستويات التقييم.

مصطلحات الدراسة

طريقة التصحيح: ويقصد بطريقة التصحيح، تحديد القاعدة التي تعطى على أساسها قيماً عددية لاستجابة المفحوص لفقرات الاختبار، أي أنها طريقة لتحويل المعلومات التي

نحصل عليها من استجابات المفحوصين إلى مقادير كمية، تعكس مستوى أداء المفحوصين على الاختبار (جرادات وسواقد، ١٩٨٥؛ Lord & Novik, 1968).

طرق التصحيح المستخدمة في الدراسة: استخدم الباحث أربع طرق لتصحيح استجابات الطلبة هي الطرق الأربعة الأولى المستخدمة في دراسة تيساي وسوين (Tsai & Suen, 1993)، والتي تتلخص بما يلي:

١- الطريقة الأولى: (ط١): تعطى العلامة (٠، ١)، وتتمثل هذه الطريقة بإعطاء علامة كاملة لكل فقرة إذا أجاب الطالب بشكل صحيح عن البدائل جميعها، وصفر إذا لم يجب بشكل صحيح على جميع البدائل.

٢- الطريقة الثانية: (ط٢): تعطى العلامة (٠،٠، ٠،٥، ١،٠)، وتتمثل هذه الطريقة بإعطاء نصف علامة إذا أجاب بشكل صحيح عن ثلاثة بدائل وعلامة كاملة إذا أجاب عن البدائل الأربعة بشكل صحيح، وصفر فيما عدا ذلك.

٣- الطريقة الثالثة: (ط٣): تعطى العلامة (٠، ٠,٣٣، ٠,٦٦، ١)، وتتمثل هذه الطريقة بإعطاء ثلث علامة إذا أجاب عن بديلين بشكل صحيح، وثلثي علامة إذا أجاب عن ثلاثة بدائل بشكل صحيح، وعلامة كاملة إذا أجاب عن البدائل الأربعة بشكل صحيح، وصفر فيما عدا ذلك.

٤- الطريقة الرابعة: (ط٤):تعطى العلامة (٠، ٠,٢٥، ٠,٥، ٠,٧٥، ١)، وتتمثل هذه الطريقة بإعطاء ربع علامة لكل بديل يتم الإجابة عنه بشكل صحيح.

الخصائص السيكومترية للاختبار: وتتمثل في خاصيتي الصدق والثبات للاختبار.

الخصائص السيكومترية لفقرات الاختبار: وتتمثل في خاصيتي الصعوبة والتمييز لفقرات الاختبار.

• معامل صدق المحك: استخدم الباحث في هذه الدراسة معاملي صدق محكي هما:

(م ص ١): وهو معامل الارتباط بين درجات الاختبار حسب طرق التصحيح المختلفة والدرجات على الاختبار الأول للمساق مبادئ القياس والتقويم.

(م ص ٢): وهو معامل الارتباط بين درجات الاختبار حسب طرق التصحيح المختلفة والمعدل التراكمي للطلبة.

متغيرات الدراسة

كان المتغير المستقل في هذه الدراسة، طريقة تصحيح فقرات الصواب والخطأ المتعدد، وله أربعة مستويات هي: (الطريقة الأولى، والثانية، والثالثة، والرابعة).

أما المتغيرات التابعة فكانت: (معاملي صدق المحك للاختبار، معامل الثبات للاختبار، معامل صعوبة الفقرة، معامل تمييز الفقرة، مستوى تحصيل الطلبة).

محددات الدراسة

١- اقتصرت عينة الدراسة على عينة من طلاب وطالبات جامعة اربد الأهلية، المسجلين لمساق القياس والتقويم للفصل الأول من العام الدراسي (٢٠٠٩/٢٠١٠).

٢- اقتصرت الدراسة على مادة الاختبار الثاني من مساق القياس والتقويم المقرر لطلبة كلية العلوم التربوية.

٣- اقتصرت الدراسة على طرق التصحيح الأربعة التي تم استخدامها. وتم شرحها في مصطلحات الدراسة.

٤- اهتمت الدراسة بمقارنة الخصائص السيكومترية للاختبار وفقراته في ضوء اختلاف طريقة التصحيح.

منهجية الدراسة وإجراءاتها:

منهاج الدراسة

اعتمدت الدراسة تصميم القياسات المتكررة، إذْ يصبح لكل طالب أربع علامات مرتبة حسب طرق التصحيح الأربع، ويعتبر هذا التصميم أقوى من تصميم المجموعات العشوائية، وبذلك تم ضبط أثر جميع العوامل التي قد تؤثر في النتائج.

مجتمع الدراسة وعينتها

تضمن مجتمع الدراسة جميع الطلبة المسجلين لجميع أقسام كلية العلوم التربوية في جامعة اربد الأهلية، والتي تضم (قسم الإرشاد النفسي والتربوي، وقسم التربية الخاصة، وقسم معلم صف عام). وتألفت عينة الدراسة من (١٢٢) طالباً وطالبة، هم جميع الطلبة المسجلين لمساق القياس والتقويم خلال الفصل الأول للعام الدراسي (٢٠٠٩/٢٠١٠)، والذين تقدموا للامتحان الثاني في موعده المقرر ـــ ويعتبر هذا المساق متطلب تخصص إجباري لجميع طلبة الكلية.

أداة الدراسة

تكونت الأداة الرئيسية للدراسة من اختبار تحصيلي يتكون من (٢٥) فقرة صواب - خطأ

متعدد لكل منها أربعة بدائل. ويغطي المادة الدراسية الخاصة بالاختبار الثاني حسب ما تقتضيه خطة المساق. وقد طبق الاختبار على جميع أفراد الدراسة في آنٍ واحد، مع مراعاة توزيع عدد البدائل الصحيحة فيها على فقرات الاختبار بشكل متوازن. إذْ كان مجموع البدائل الصحيحة (٥١)، ومجموع البدائل الخاطئة (٤٩). واسترشد الباحث في بناء هذا الاختبار بالأسس العامة المتبعة في تصميم اختبارات التحصيل الصفية (ثورندايك وهيجن، ١٩٨٦)، إذ تم صياغة الأهداف السلوكية للمادة، وتحديد المحتوى والموضوعات المراد قياسها، و إعداد جدول المواصفات، وصياغة ٢٥ فقرة من نوع الصواب والخطأ المتعدد جميعها بأربعة بدائل لقياس الأهداف الواردة في جدول المواصفات.

إجراءات التنفيذ

بعد إعداد الاختبار بصورته النهائية، طبق الاختبار بتاريخ (٢٠٠٩/١٢/٨) على أفراد عينة الدراسة بإشراف الباحث، ومساعدة عدد من الزملاء أعضاء هيئة التدريس، إذْ جرى تعريف الطلبة بفقرات الصواب واخطأ المتعدد وتدريبهم على كيفية إجابتها، قبل موعد الاختبار ـــ إذْ لم يسبق لهم التعرف عليها. وبعد الانتهاء من إجراءات التطبيق، تم تصحيح الاختبار لأغراض الدراسة بإعطاء درجة للفقرة هي عبارة عن عدد الاستجابات الصحيحة، والتي تعبر عن عدد المهمات الجزئية المنجزة لكل فقرة (٠، ١، ٢، ٣، ٤)، وإدخال هذه العلامات للحاسوب بصفته ملف بيانات (Data File) يعتمد لجميع طرق التصحيح المستخدمة في الدراسة. وقد عولجت هذه البيانات لتتلاءم مع معادلة التصحيح المناسبة لكل طريقة من طرق التصحيح الأربعة المستخدمة في الدراسة، إذْ تعطى العلامة (٠، ١) للطريقة الأولى، و(٠،٠، ٠,٥، ١,٠) للطريقة الثانية، و(٠، ٠,٣٣، ٠,٦٦، ١) للطريقة الثالثة، و(٠، ٠,٢٥، ٠,٥، ٠,٧٥، ١) للطريقة الرابعة. وسجلت علامات الطلبة على هذا الاختبار ـــ مصححة بالطريقة الرابعة ـــ علامةً للاختبار الثاني، وشكلت نسبة (٢٥٪) من العلامة النهائية لهم، وكانت استجابة الطلبة بمنتهى الاهتمام والجدية، وبذلك أصبح لكل طالب أربع درجات مختلفة حسب طرق التصحيح المستخدمة.

الأساليب الإحصائية

اعتمدت الدراسة تصميم القياسات المتكررة، وتم حساب العديد من الإحصائيات لكل مجموعة من الدرجات حسب طرق التصحيح الأربع. وبالتحديد فقد حسبت المتوسطات

الحسابية، والانحرافات المعيارية، ومعاملات ارتباط بيرسون للصدق المحكي لدرجات الطلبة ومعدلاتهم التراكمية، ومعاملات (α) للثبات، ومعاملات الصعوبة والتمييز، واستخدم الإحصائي :(2 α-1) / (1 α-1)=W الصورة المعدلة لاختبار فلدت (feldt test) لفحص تساوي معاملات الثبات المرتبطة (Feldt,1980, Alsawalmeh & Feldt, 1994). كما تم استخدام تحليل التباين للقياسات المتكررة لمقارنة المتوسطات الحسابية لمعاملات الصعوبة والتمييز، واختبار بنفر وني للكشف عن الفروق ذات الدلالة إن وجدت.

عرض النتائج ومناقشتها

أولاً: عرض نتائج السؤال الأول

نص هذا السؤال على: هل تختلف معاملات الصدق المحكي لاختبار الصواب والخطأ المتعدد باختلاف طريقة التصحيح عند مستوى الدلالة (α =٠,٠٥)؟

للإجابة عن هذا السؤال تم حساب معاملي صدق لاختبار الصواب والخطأ المتعدد، الأول: معامل الارتباط بين درجات الاختبار حسب طرق التصحيح المختلفة والدرجات على الاختبار الأول للمساق مبادئ القياس والتقويم (م-ص ١)، والثاني: معامل الارتباط بين درجات الاختبار حسب طرق التصحيح المختلفة والمعدل التراكمي للطلبة (م ص ٢)، وهي كما في الجدول رقم (١).

الجدول رقم (١)

معاملات الصدق حسب طرق التصحيح المختلفة

طريقة التصحيح	ط١	ط٢	ط٣	ط٤
معامل صدق/ م ص١	٠,٤٩٥	٠,٥٠٠	٠,٤٧٣	٠,٤٦٢
معامل صدق/ م ص٢	٠,٦٩٢	٠,٦٧٦	٠,٦٦٣	٠,٦٥٥

يتبين من الجدول رقم (١) وجود فروق ظاهرية طفيفة في معاملات الصدق تبعا لاختلاف طريقة التصحيح وبهدف الكشف عن الدلالة الإحصائية للفروق في معاملي الصدق (م ص ١، م ص ٢)، تم استخدم اختبار (ت) للكشف عن الدلالة الإحصائية للفروق، للعينات المترابطة، وذلك كما في الجدول رقم (٢).

الجدول رقم (٢)
قيم (ت) المحسوبة لاختبار بين معاملات الصدق المحكي
(م ص ١، م ص ٢) للعينات المترابطة

الاختبار			الأول	الثاني	الثالث
	الارتباط مع المحك م ص ١		٠,٤٩٥	٠,٥٠٠	٠,٤٧٣
الثاني	٠,٥٠٠	قيمة الإحصائي ت	٠,١٢٠-		
		الارتباط البيني	٠,٩٧٣		
		عدد أفراد العينة	٢٥		
		درجة الحرية	٢٢		
		الدلالة الإحصائية	٠,٩٠٦		
الثالث	٠,٤٧٣	قيمة الإحصائي ت	٠,٣٦٥	٠,٨٩٠	
		الارتباط البيني	٠,٩٤٩	٠,٩٨٧	
		عدد أفراد العينة	٢٥	٢٥	
		درجة الحرية	٢٢	٢٢	
		الدلالة الإحصائية	٠,٧١٩	٠,٣٨٣	
الرابع	٠,٤٦٢	قيمة الإحصائي ت	٢,٠٣٥	٢,٣١١-	٠,٨٧٠-
		الارتباط البيني	٠,٩٣٠	٠,٩٧٢	٠,٩٩٥
		عدد أفراد العينة	٢٥	٢٥	٢٥
		درجة الحرية	٢٢	٢٢	٢٢
		الدلالة الإحصائية	٠,٦٤٣	٠,٣٩٥	٠,٥٧٥
الاختبار			الأول	الثاني	الثالث
	الارتباط مع المحك م ص ٢		٠,٦٩٢	٠,٦٧٦	٠,٦٦٣
الثاني	٠,٦٧٦	قيمة الإحصائي ت	٠,٤٤٥		
		الارتباط البيني	٠,٩٧٣		
		عدد أفراد العينة	٢٥		
		درجة الحرية	٢٢		
		الدلالة الإحصائية	٠,٦٦١		
الثالث	٠,٦٦٣	قيمة الإحصائي ت	٠,٥٨٨	٠,٥٠١	
		الارتباط البيني	٠,٩٤٩	٠,٩٨٧	
		عدد أفراد العينة	٢٥	٢٥	
		درجة الحرية	٢٢	٢٢	
		الدلالة الإحصائية	٠,٥٦٣	٠,٦٢٢	
الرابع	٠,٦٥٥	قيمة الإحصائي ت	٠,٦٤٥	٠,٥٧٠	٠,٥٠٣
		الارتباط البيني	٠,٩٣٠	٠,٩٧٢	٠,٩٩٥
		عدد أفراد العينة	٢٥	٢٥	٢٥
		درجة الحرية	٢٢	٢٢	٢٢
		الدلالة الإحصائية	٠,٥٢٥	٠,٥٧٥	٠,٦٢٠

يلاحظ من الجدول رقم (٢) عدم وجود فروق ذات دلالة إحصائية بين معاملي الصدق المحكي (م ص١، م ص٢) باختلاف طريقة التصحيح.

ثانياً: عرض نتائج السؤال الثاني

نص هذا السؤال على: هل يختلف معامل الثبات لاختبار الصواب والخطأ المتعدد باختلاف طريقة التصحيح عند مستوى الدلالة (α =٠,٠٥)؟

للإجابة عن هذا السؤال تم حساب قيم معاملات الثبات باستخدام معادلة كرونباخ ألفا لاختبار الصواب والخطأ المتعدد حسب طرق التصحيح المختلفة وهي كما في الجدول رقم (٣).

الجدول رقم (٣)
معاملات الثبات حسب طرق التصحيح المختلفة

طريقة التصحيح	ط١	ط٢	ط٣	ط٤
معامل ثبات التجانس	٠,٧٧٦	٠,٨٠٣	٠,٧٩٤	٠,٧٩١

يتبين من الجدول رقم (٣) وجود فروق ظاهرية في معاملات الثبات تبعا لاختلاف طريقة التصحيح وبهدف الكشف عن الدلالة الإحصائية لهذه الفروق في معاملات الثبات فقد استخدمت الصورة المعدلة لاختبار فيلدت (feldt test). (Feldt, 1980,Alsawalmeh & Feldt,1994) (W=(1 α 1)/(1- α2).

وذلك كما في الجدول رقم (٤). حيث: (α ١) هو معامل ثبات الاختبار الأول، (α ٢) معامل ثبات الاختبار الثاني.

الجدول رقم (٤)
قيم (w) المحسوبة لاختبار الفروق بين معاملات الثبات للعينات المترابطة

الاختبار			الأول	الثاني	الثالث
	ثبات التجانس		٠,٧٧٦	٠,٨٠٣	٠,٧٩٤
الثاني	٠,٨٠٣	قيمة الإحصائي فيلدت	١,١٤٠		
		الارتباط البيني	٠,٩٧٣		
		عدد أفراد العينة	٢٥		
		درجة الحرية	٣٢٦,٣٤٥		
		الدلالة الإحصائية	٠,١١٨		
الثالث	٠,٧٩٤	قيمة الإحصائي فيلدت	١,٠٩١	٠,٩٥٧	
		الارتباط البيني	٠,٩٤٩	٠,٩٨٧	
		عدد أفراد العينة	٢٥	٢٥	
		درجة الحرية	١٧٩,٠٩٠	٦٤٣,٩٨٠	
		الدلالة الإحصائية	٠,٢٨١	٠,٧١٣	
الرابع	٠,٧٩١	قيمة الإحصائي فيلدت	١,٠٧٥	٠,٩٤٣	٠,٩٨٦
		الارتباط البيني	٠,٩٣٠	٠,٩٧٢	٠,٩٩٥
		عدد أفراد العينة	٢٥	٢٥	٢٥
		درجة الحرية	١٣٣,٤٢٦	٣١٦,١٥٧	١٦١٧,١٥٦
		الدلالة الإحصائية	٠,٣٣٨	٠,٦٩٩	٠,٦١٣

يلاحظ من الجدول رقم (٤) عدم وجود فروق ذات دلالة إحصائية بين معاملات الثبات باختلاف طريقة التصحيح.

ثالثاً: عرض نتائج السؤال الثالث

نص هذا السؤال على: هل يختلف معامل التمييز لفقرات اختبار الصواب والخطأ المتعدد باختلاف طريقة التصحيح عند مستوى الدلالة (α =٠,٠٥)؟

للإجابة عن هذا السؤال تم حساب المتوسطات والانحرافات المعيارية لمعاملات تمييز الفقرات حسب طرق التصحيح المختلفة .وذلك كما هو مبين في الجدول رقم (٥).

الجدول رقم (٥)

المتوسطات الحسابية والانحرافات المعيارية لمعاملات تمييز الفقرات ن: عدد الفقرات=٢٥

معامل التمييز للفقرة حسب الطريقة	المتوسط الحسابي	الانحراف المعياري
الأولى	٠,٣٠٠	٠,١٧
الثانية	٠,٣٣٢	٠,١٥
الثالثة	٠,٣٢٤	٠,١٤
الرابعة	٠,٣٢٦	٠,١٤

يتضح من الجدول رقم (٥) وجود فروق ظاهرية طفيفة بين المتوسطات الحسابية لمعاملات تمييز الفقرات حسب طرق التصحيح المختلفة وبهدف الكشف عن الدلالة الإحصائية لهذه الفروق الظاهرية استخدم تحليل التباين وفق تصميم القياسات المتكررة ويبين الجدول رقم (٦) نتائج هذا التحليل.

الجدول رقم (٦)

نتائج تحليل التباين لمتوسطات تمييز الفقرات باختلاف طريقة التصحيح

مصدر التباين		مجموع المربعات	درجة الحرية	متوسط المربعات	ف المحسوبة	الدلالة الإحصائية
بين الفقرات		١,٩٠٣	٢٤	٠,٠٧٩		
داخل الفقرات	بين الطرق	٠,٠١٥	٣	٠,٠٠٥	١,٦٦٧	٠,١٨٢
	الباقي	٠,٢١٠	٧٢	٠,٠٠٣		
	الكلي	٠,٢٢٥	٧٥	٠,٠٠٣		

وتبين من نتيجة هذا التحليل الجدول رقم (٦) عدم وجود فروق ذات دلالة إحصائية بين متوسطات معاملات التمييز للفقرات حسب طرق التصحيح المختلفة.

رابعاً: عرض نتائج السؤال الرابع

نص هذا السؤال على: هل يختلف معامل الصعوبة لفقرات اختبار الصواب والخطأ المتعدد باختلاف طريقة التصحيح عند مستوى الدلالة (α =٠,٠٥)؟

للإجابة عن هذا السؤال تم حساب المتوسطات الحسابية والانحرافات المعيارية لمعاملات صعوبة الفقرات حسب طرق التصحيح المختلفة، وذلك كما هو مبين في الجدول رقم (٧).

الجدول رقم (٧)
المتوسطات الحسابية والانحرافات المعيارية لمعاملات صعوبة الفقرات ن: عدد الفقرات=٢٥

معامل الصعوبة للفقرة حسب الطريقة	المتوسط الحسابي	الانحراف المعياري
الأولى	٠,٣٣٧١٦	٠,١٩٨
الثانية	٠,٤٧٤٠٤	٠,١٧٨
الثالثة	٠,٥٩٣٥٦	٠,١٥٢
الرابعة	٠,٦٨٢١٦	٠,١٢٥

يتضح من الجدول رقم (٧) وجود فروق ظاهرية بين المتوسطات الحسابية لمعاملات صعوبة الفقرات باختلاف طريقة التصحيح وبهدف الكشف عن الدلالة الإحصائية لهذه الفروق الظاهرية استخدم تحليل التباين وفق تصميم القياسات المتكررة ويبين الجدول رقم (٨) نتائج هذا التحليل.

الجدول رقم (٨)
جدول تحليل التباين وفق تصميم للقياسات المتكررة لمتوسطات معاملات صعوبة الفقرات حسب طرق التصحيح المختلفة

مصدر التباين		مجموع المربعات	درجة الحرية	متوسط المربعات	ف المحسوبة	الدلالة الإحصائية
بين الفقرات		٢,٣٩٤	٢٤	٠,١٠٠		
داخل الفقرات	بين الطرق	١,٦٨١	٣	٠,٥٦٠	١٦٧,٣٧٣	٠,٠٠٠
	الباقي	٠,٢٤١	٧٢	٠,٠٠٣		
	الكلي	١,٩٢٢	٧٥	٠,٠٢٦		

وتبين نتيجة التحليل، جدول رقم (٨) وجود فروق ذات دلالة إحصائية عند مستوى الدلالة (α =٠,٠٥) بين متوسطات معاملات الصعوبة تعزى لاختلاف طريقة التصحيح وبهدف الكشف عن مصدر هذه الفروق استخدم اختبار بنفروني للمقارنات البعدية، وذلك كما في الجدول رقم (٩).

الجدول رقم (٩)
نتائج اختبار بنفر وني للفروق بين معاملات الصعوبة حسب طرق التصحيح المختلفة

الطريقة		الأولى	الثانية	الثالثة	الرابعة
Bonferroni	المتوسط الحسابي المُقدّر	٠,٣٣٧	٠,٤٧٤	٠,٥٩٤	٠,٦٨٢
الأولى	٠,٣٣٧				
الثانية	٠,٤٧٤	٠,١٣٧			
الثالثة	٠,٥٩٤	٠,٢٥٦	٠,١٢٠		
الرابعة	٠,٦٨٢	٠,٣٤٥	٠,٢٠٨	٠,٠٨٩	

يتضح من الجدول رقم (٩) أن جميع الفروق الثنائية بين متوسطات معاملات الصعوبة دالة إحصائيا عند مستوى الدلالة (α =٠,٠٥). إذْ جاءت هذه الفروق لصالح الطريقة الرابعة ثم الثالثة ثم الثانية.

خامساً: نتائج السؤال الخامس

نص هذا السؤال على: هل يختلف تحصيل الطلبة في اختبار الصواب والخطأ المتعدد باختلاف طريقة التصحيح عند مستوى الدلالة (α =٠,٠٥)؟

للإجابة عن هذا السؤال تم حساب المتوسطات الحسابية والانحرافات المعيارية لعلامات أفراد العينة على اختبار الصواب والخطأ المتعدد حسب كل طريقة من طرق التصحيح المختلفة وذلك كما هو مبين في الجدول رقم (١٠).

الجدول رقم (١٠)
المتوسطات الحسابية والانحرافات المعيارية لعلامات اختبار الصواب والخطأ المتعدد

العلامة حسب طريقة التصحيح	العدد	المتوسط الحسابي	الانحراف المعياري
الأولى	١٢٢	٨,٤٢٦	٤,٢٨
الثانية	١٢٢	١١,٨٥٢	٤,٠٧
الثالثة	١٢٢	١٤,٨٣٩	٣,٤٠
الرابعة	١٢٢	١٧,٠٤٩	٢,٧٨

يلاحظ من الجدول رقم (١٠) وجود فروق ظاهرية بين المتوسطات الحسابية لعلامات أفراد العينة حسب طريقة التصحيح. وبهدف الكشف عن الدلالة الإحصائية لهذه الفروق في المتوسطات الحسابية لدرجات أفراد العينة، استخدم تحليل التباين وفق تصميم القياسات المتكررة، كما في الجدول رقم (١١).

الجدول رقم (١١)
نتائج تحليل التباين وفق تصميم للقياسات المتكررة لمتوسطات درجات أفراد العينة حسب طرق التصحيح المختلفة

مصدر التباين		مجموع المربعات	درجة الحرية	متوسط المربعات	ف المحسوبة	الدلالة الإحصائية
بين الأفراد		٦٢٢٩,٩٣٤	١٢١	٥١,٤٨٧		
داخل الأفراد	بين الطرق	٥١٢٤,٧٧١	٣	١٧٠٨,٢٥٧	١٩٠٢,٢٠٢	٠,٠٠٠
	الباقي	٣٢٥,٩٨٩	٣٦٣	٠,٨٩٨		
	الكلي	٥٤٥٠,٧٦٠	٣٦٦	١٤,٨٩٣		
الكلي		١١٦٨٠,٦٩٤	٤٨٧	٢٣,٩٨٥		

يتضح من الجدول رقم (١١) وجود فرق ذو دلالة إحصائية في متوسطات علامات أفراد عينة الدراسة يعزى لاختلاف طريقة التصحيح وبهدف الكشف عن مصدر هذه الفروق استخدم اختبار بنفر وني للمقارنات البعدية وذلك كما هي في الجدول التالي رقم (١٢).

الجدول رقم (١٢)
نتائج اختبار بنفر وني للفروقات بين متوسطات علامات الاختبار باختلاف طريقة التصحيح

المعادلة		الأولى	الثانية	الثالثة	الرابعة
Bonferroni	المتوسط الحسابي المُقدّر				
الأولى					
الثانية		٣,٤٣			
الثالثة		٦,٤١	٢,٩٩		
الرابعة		٨,٦٢	٥,٢٠	٢,٢١	

يتضح من الجدول رقم (١٢) أن جميع الفروق الثنائية بين متوسطات العلامات لأفراد العينة داله إحصائيا عند مستوى الدلالة (α =٠,٠٥). فقد جاءت هذه الفروق لصالح الطريقة الرابعة ثم الثالثة فالثانية.

مناقشة النتائج

تشير نتائج هذه الدراسة إلى عدم وجود أدلة تبين أن هناك اختلافاً في معاملات الثبات أو الصدق المحكي أو أن هناك مستويات مختلفة من التمييز، إلا أن هناك فروق ذات دلالة إحصائية في صعوبة الفقرات ومستوى تحصيل الطلبة تعزى إلى اختلاف طريقة التصحيح، إذْ أدت طريقة التصحيح الرابعة—الأكثر مراعاةً للمعرفة الجزئية—إلى معاملات صعوبة

ومتوسطات تحصيل أعلى (أي الأكثر سهولة والأعلى درجة)، في حين تلتها بذلك طريقة التصحيح الثالثة ثم الثانية فالأولى (أي الأكثر صعوبة والأقل درجة). وبذلك فان نتائج هذه الدراسة تختلف مع نتائج ألبانس وسابرس (Albanese & Sabers, 1988). التي أظهرت أن أساليب التصحيح التي تراعي المعرفة الجزئية كانت أكثر ثباتاً وصدقاً تلازمياً وتنبؤياً من تلك الأساليب التي تقلل من احتمال النجاح بالصدفة، واتفقت نتائج هذه الدراسة مع نتائج تيساي وسوين (Tsai & Suen, 1993) التي أظهرت أن الفروق في قيمة الثبات والتمييز لم تكن ذات دلالة إحصائية، وقد اتفقت كذلك مع نتائج براد (Brad, 2003) التي أظهرت عدم وجود فروق ذات دلالة إحصائية في معاملات الثبات باختلاف طريقة التصحيح. ولعل عدم وجود فروق ذات دلالة إحصائية في معاملات الصدق المحكي، ومعاملات الثبات، والتمييز، يعود إلى كون فقرات الاختبار هي ذاتها في طرق التصحيح المختلفة، مما جعل الأداء على فقرات الاختبار متقارباً، ومن ثمَّ تباين أداء متقارب إلى حد كبير، مما أدى إلى معاملات تمييز متقاربة ومعاملات صدق وثبات متقاربة أيضاً. وأن الفروق في قيمة معاملات الصعوبة ومتوسطات التحصيل كانت ذات دلالة إحصائية لصالح طريقة التصحيح الرابعة ويمكن تفسير ذلك بأن مراعاة المعرفة الجزئية في تصحيح فقرات الصواب والخطأ المتعدد تزيد من احتمالية النجاح بالصدفة، وكذلك احتمالية حصول الطلبة على درجات أعلى نتيجة هذه المعرفة الجزئية، بالتالي الوصول إلى قيمة معاملات صعوبة ومتوسطات تحصيل أعلى لصالح طريقة التصحيح الرابعة، تلتها بذلك طريقة التصحيح الثالثة ثم الثانية فالأولى.

الاستنتاجات والتوصيات

يتضح من خلال هذه النتائج أن اختيار طريقة التصحيح الأكثر ملائمة أمر يحتاج إلى التدبر والتفكير في قضايا عدة منها، الغرض من عملية التقويم، والتوجه والأسلوب المتبع في تفسير العلامة، فإذا كانت العلامة ستفسر بشكل نسبي (معيارية المرجع)، إذْ يقارن أداء الفرد بأداء المجموعة التي طبق عليها الاختبار، كما تشير نتائج هذه الدراسة إلى أن هناك القليل من الأدلة التي تدعم تفوق إحدى طرق التصحيح المستخدمة لفقرات الصواب والخطأ المتعدد -وبشكل مطلق- على الطرق الأخرى من ناحية القياس النفسي، أما في حال النتائج المحكية المرجع إذْ تفسر الدرجة بشكل مطلق يحدد مدى إتقان الطالب لمخرجات التعلم، وتقارن هذه الدرجة بعلامة قطع محددة سلفاً بوصفها حدًا أدنى للنجاح—كاختبارات التحصيل، أو إتقان مهارة معينة، أو اختبار منح رخصة قيادة مركبة—فإن اختيار طريقة

التصحيح هنا قد يؤثر وبشكل واضح في تفسير النتيجة والقرارات المتخذة بناءً عليها. ففي اختبارات التحصيل بوجهٍ عام تُعتمد العلامة (٥٠) من (١٠٠) على أنها الحد الأدنى المقبول للنجاح في المواد الدراسية المختلفة، وقد يرتفع الحد الأدنى في الجامعة عنه في المدارس، وقد يرتفع بعض الشيء في الماجستير والدكتوراه، وهنا فان اختيار طريقة التصحيح الأولى يتوقع أن تقدم علامة أقل للطالب.

وبذلك يوصي الباحث:

١- باستخدام طريقة التصحيح الرابعة في الاختبارات التحصيلية والتي من المتوقع أن تقدم العلامة الأعلى للطالب.

٢- أما في المهارات التي تحتاج درجة عالية من الإتقان، كاختبار منح رخصة قيادة مركبة، أو اختبارات منح رخص مهنية حساسة كالمهن الطبية، فيوصي الباحث باستخدام طريقة التصحيح الأولى، إذْ تمثل الإتقان التام والأكثر تشدداً، كونها تعتمد على المعرفة الكاملة ولا تراعي المعرفة الجزئية بتاتاً، في حين أن بقية الطرق تراعي المعرفة الجزئية بدرجات متفاوتة. من هنا فإنّ طريقة تفسير العلامة وتحديد علامة النجاح يجب أن يقابل مستوى الصعوبة النسبي للفقرة بطريقة التصحيح المستخدمة.

٣- كما يوصي الباحث بإجراء المزيد من الدراسات المماثلة لاختبارات محكية لمهارات تتطلب درجة عالية من الدقة والإتقان.

المراجع

أمريزيق، خالد (١٩٩٢). **المقارنة بين اختبار الاختيار من متعدد واختبار الصواب والخطأ المتعدد من حيث الخصائص السيكومترية**. رسالة ماجستير غير منشورة، الجامعة الأردنية، الأردن.

ثورندايك، روبرت وهيجن، إليزابيث (١٩٨٦). **القياس والتقويم في علم النفس والتربية**، (ترجمة عبد الله زيد الكيلاني وعبد الرحمن عدس). نيويورك: جون وايلي وأولاده.

جرادات، ضرار وسواقد، ساري (١٩٨٥). استقصاء أثر ثلاث طرق لتصحيح اختبارات الاختيار من متعدد على الخصائص السيكومترية للاختبار وعلى أداء المفحوصين عليه. **مجلة دراسات**، ١٢، ٢٧- ٤٢.

سوالمة، يوسف (١٩٩٤). اختبارات الصواب والخطأ المتعدد، الخصائص النسبية وعدد البدائل الصحيحة. أبحاث اليرموك، **سلسلة العلوم الإنسانية والاجتماعية**، ١٠(٣)، ٥٧٣-٥٩٥.

الصمادي، مروان وسوالمه، 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**، جامعة اليرموك، ٥(١)، ٧١- ٨٧.

عدس، عبد الرحمن (١٩٨٨). **دليل المعلم في بناء الاختبارات التحصيلية**. المنظمة العربية للتربية والثقافة والعلوم، تونس: إدارة البحوث التربوية.

عودة، أحمد (٢٠٠٥). **القياس والتقويم في العملية التدريسية**. اربد: دار الأمل.

المومني، محمد ضيف الله (١٩٩٦). **مقارنة أثر اختبارات الاختيار من متعدد، واختبارات الصواب والخطأ المتعدد في الاحتفاظ بقواعد اللغة الإنجليزية**. رسالة ماجستير غير منشورة، جامعة اليرموك، الأردن.

Albanese, M. A. & Sabers, D. L. (1988). Multiple true–false items: A study of interterm correlations scoring alternatives, and reliability estimation. **Journal of Educational Measurement, 25**(2), 111-123.

Alswalmeh, Y.M.& Feldt, L,S. (1994). Amodification of feldts' test of the equality of two dependent alpha coefficients. **Psychometrika 59**(1), 49-57.

Brad, W. (2003). **Scoring multiple true false items: A comparison of summed scores and response pattern scores at item and test levels.** (ERIC Document Reproduction Service No. ED476 148).

Downing, S.M., Baranoski, R. A., Gross, L. J. & Norcni. J. J. (1995). Item type and cognitive ability measured the validity evidence for multiple true-false items in medical specialty certification. **Applied Measurement in Education, 8**(2), 187-197.

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.

Ebel, R. L. & Frisbie. D. A. (1986). **Essentials of educational measurement** (4th ed). New Jersey: Prentice J Hall.

Erickson, I. L. (1988). The effect of item type and test format on level of difficulty and mean item discrimination index in achievement test for nurses. **Dissertation Abstract International, 49**(6), 1437.A.

Feldt, L. S. (1980). A test of hypothesis that Cranach's alpha reliability coefficient is the same for two tests administered to the same sample. **Psychometrika, 45,** 9-105 .

Frisbie, D. A. (1992). The multiple true false item formats a status review. **Education Measurement: Issues and Practice, 11**(4), 21–26.

Frisbie, D .A. & Sweeny, D. C. (1982). The relative merits of multiple true– false achievement tests. **Journal of Educational Measurement, 19**(1), 29-35.

Kreiter, C. D., & Frisbie, D. A. (1989). Effectiveness of multiple true-false item, **Applied Measurement in Education, 2,** 207-216.

Lord, F. M. & Novick, M. R. (1968). **Statistical theories of mental test scores.** Addison-Wesley publishing company: Reading Massachusetts, USA.

Tsai, F. & Suen, H. (1993). A Brief report on a comparison of six scoring methods for multiples true false item. **Educational and Psychological Measurement, 53,** 399-404.