مقارنة طرق ضبط أثر التخمين في اختبارات الاختيار من متعدد وأثرها على خصائص الاختبار

شريف السعودي، معين النصراوين *

ملخص

هدفت هذه الدراسة إلى مقارنة ثلاث طرق لضبط أثر التخمين في اختبارات الاختيار من متعدد، هي: طريقة التصحيح لأثر التخمين التقليدية، طريقة تصحيح المموهات، طريقة استخدام البديل "لا أعرف الإجابة"، من خلال أثرها على خصائص الاختبار، باستخدام اختبار رياضيات تحصيلي على عينة مكونة من (159) طالباً وطالبةً، وأشارت النتائج إلى وجود فروق دالة احصائياً عند مستوى دلالة ($\alpha \leq 0.05$) بين طريقة تصحيح المموهات وطريقة التصحيح لأثر التخمين لصالح طريقة تصحيح المموهات على وفق محكّات، هي: الأداء على الاختبار، الصدق، والثبات؛ نظراً لاعتمادها على المعرفة الجزئية لدى الطلبة. في حين لم تظهر النتائج فروق ذات دلالة احصائية بين طريقتي التصحيح لأثر التخمين وطريقة البديل لا أعرف الإجابة.

الكلمات الدالة: اختبارات الاختيار من متعدد، تصحيح الاختبارات.

المقدمة

تعدُ الاختبارات التحصيلية من أهم الأدوات التي يلجأ من خلالها العاملون في المؤسسات التربوية لتقييم ما حققه طلابهم من نتاجات التعلّم، أو لتحديد مسار الطلاب الأكاديمي بما يتناسب مع قدراتهم ومهاراتهم، وتلجأ المؤسسات التربوية ضمن نظامها التربوي إلى عكس مدى النجاح الذي حققته من خلال نتائج الاختبارات التحصيلية، كما تستخدم الاختبارات التحصيلية كوسيلة لتشخيص الصعوبات وتصميم البرامج العلاجية للطلبة، ولعل تحقّق كل ما سبق في وظائف الاختبارات التربوية واستخداماتها مرهوناً بدقة هذه الأدوات، وصدق وثبات وموضوعية النتائج المتمخضة عنها.

وتُعدّ الدقة والموضوعية والتمتّع بصدق، وثبات النتائج المتمخضة عن أدوات قياس التحصيل مرهونة بأنماط وأنواع هذه الأدوات، وتوجد أنماط مختلفة من الأساليب الاختبارية التحصيلية التي يمكن للعاملين في المؤسسات التربوية استخدامها، فهناك الاختبارات الشفوية واختبارات المقال، والاختبارات الموضوعية بمختلف أنواعها ولكل منها خصائصها وميزاتها، والمواقف التي يُفضل أن تُستخدم فيها، وتُعدّ الاختبارات الموضوعية الأحدث ظهورًا مقارنة مع الاختبارات الإنشائية حيث يُشير ميريت (Merritt, 2006) إلى أنّ أول من وضع هذا النمط من الاختبارات هو (فريديرك كيلي) في جامعة كنساس في الولايات المتحدة الأميركية عام (1914م)؛ بهدف تقييم القدرات العقلية للجنود المشاركين في الحرب العالمية الأولى، وقد لاقت قبولاً واسعاً من قبل العاملين في المجال التربوي، وانتشر استخدامها في المؤسسات التربوية، وتُعدّ هي اليوم من أكثر أدوات قياس التحصيل استخداماً.

لقد أدى شيوع فكرة الاختبارات الموضوعية إلى التعدد في الأشكال التي تتخذها هذه الاختبارات، فهنالك اختبارات الصواب والخطأ، واختبارات المزاوجة، واختبارات ملء الفراغ، واختبارات الاختيار من متعدد التي تُعدّ كما يشير هالدينا وداوننج (Haladyna & Downing, 2004) أكثر أنواع الأسئلة الموضوعية استخداماً في المؤسسات التعليمة، لاسيما في المواقف التي تهدف لاختبار إعدادٍ كبيرة من المفحوصين، حيث يتطلب هذا النوع من الاختبارات من المستجيب أن يختار إجابة واحدة فقط من مجموعة من الإجابات المطروحة، ويتكون السؤال من قسمين رئيسين هما: متن السؤال (Stem) الذي يحتوي على المشكلة، يليها مجموعة من البدائل أو المموهات بالإضافة للإجابة الصحيحة، وعموماً يُعد هذا النمط من الاختبارات الموضوعية الأكثر استخدامًا عند بناء بنوك الأسئلة، أو إجراء الاختبارات التكيفية (Adapting Tests) أو غيرها من الاختبارات المختلفة في شتى المجالات؛ وذلك لما يتمتع به هذا النمط من الاختبارات الموضوعية من مرونة وقدرة على اختبار عددٍ كبيرٍ من المفحوصين، كما أنها تصلح لتقويم معظم نتاجات التعلم (Merritt, 2006 ; Poham, 2003).

على الرغم من الميزات التي تمتاز بها فقرات اختبار الاختيار من متعدد، إلا أنها تفتقرُ لإمكانية الكشف عن قدرة الطلبة على

qþ2¢kÄ Ü2016/03/04 ł o℠!ƒ ¤ž¤ƙ¶ƒ qþ2¢k .(2) ł¤℠2z!ƒ LJ¢¤žZ ł z¤¤¢ɳ Ü ł¤ƙ¶sU!ƒÄ ł¤Ä℠2k!ƒ ¤žÄ¤z!ƒ ł¤¤‡ Ý (1)LJþ2Ũƒ Ü¤ž¤¤z k!ƒÄ ł¤℠2k!ƒ I2ƒ¤Ä * .2016/07/29 ₪!Ä℠p

© 2018 عمادة البحث العلمي/ الجامعة الأردنية. جميع الحقوق محفوظة.

الإبداع، أو الإتيان بأفكار عديدة ومتنوعة وأصيلة، ولا تستطيع فقراتها قياس قدرة الطلبة على التركيب والتقويم كعمليات عقلية عليا، كما أنه يصعُب الوقوف من خلالها على قدرة الطلبة على تنظيم المعلومات وربطها ببعضها البعض، إضافة إلى أنها تدفع المفحوصين للتركيز على تفاصيل صغيرة متناثرة وغير مترابطة من المحتوى الدراسي، إذ يُشير كل من شنايدر، وايجن، وجوليان (Schneider, Egan and Julian, 2013) في هذا السياق إلى أنه حينما يختار الطالب بديلاً خاطئاً عند إجابته عن فقرات اختبار الاختيار من متعدد، فإنّ المفهوم الخطأ يتعزّز لديه؛ وبهذا قد يكتسب الطالب معرفة غير صحيحة، وهذا ما يُسمى بحسب كل من شنايدر وايجن بـ "تأثير البديل الخطأ"(Negative Suggestion Effect)، كما أن بنية أسئلة فقرات الاختيار من متعدد تؤدي إلى لجوء الطلبة للغش والتخمين العشوائي، ولعلّ لجوء الطالب للتخمين العشوائي عندما لا تتوافر لديه المعلومات التي تُمكنه من الوصول إلى الإجابة الصحيحة من أكبر المشكلات التي تواجه هذا النوع من الاختبارات، حيث يقوم الطالب بالاختيار العشوائي لأحد البدائل حتى لو طُلب منه أن يترك الفقرة بدون إجابة إذا لم تكن لديه معرفة بالإجابة الصحيحة، فبعض الطلبة يميلون إلى المجازفة أو المخاطرة بدرجات مختلفة حسب درجة التزامهم بالتعليمات، وحكمتهم في الإجابة، فأحياناً قد توجد منبهات أو مفاتيح في الفقرة أو في البدائل تساعد الطالب على التوصل للإجابة الصحيحة، وبهذا لا يمكن للمصحح أن يقرر فيما إذا كانت الإجابة الصحيحة للطالب عن فقرةٍ ما تعكس قدرته الحقيقية على الإجابة وتمّكنه من المفهوم، أو التعميم، أو العلاقة المقاسة بالفقرة، وخاصة عندما يكون الهدف من الاختبار مثلاً تشخيص صعوبات التعلّم أو الوصول إلى درجة الإتقان لمهارة معينة (Annie & Chan, 2012).

ويُشير كبنجر وهولشر وريف (Kubinger, Holocher & Reif, 2010) إلى أن التخمين في الاختبارات الموضوعية ولاسيما الاختيار من مُتعدد، يؤدّي إلى غياب خاصية العدالة عن نتائج الاختبار، فمثلاً إذا ما تقدم مفحوصان من مستوى قدرة متوسطة لأداء اختبار موضوعي من نوع الاختيار من متعدد فإن هناك احتمالاً بأن يجتاز المفحوص الذي يلجأ للتخمين الاختبار أو يحصل على درجة أعلى مقارنة بالمفحوص الذي أجاب عن فقرات الاختبار بدون تخمين، وأيضا إذا كان هناك مفحوصان من مستوى قدرة متفاوتةٍ فقد ينجح الطالب ذو القدرة المنخفضة نتيجة لجوئه للتخمين، ويفشل الطالب ذو القدرة المرتفعة الذي أجاب عن فقرات الاختبار بناءً على معرفته الحقيقية، مما يؤدي إلى انتهاك أهم الخصائص الأساسية للاختبار وهي العدالة فتصبح نتائج الاختبار متحيزة.

فالتخمين العشوائي يعُدّ من المُهددات لكل من صدق وثبات اختبارات الاختيار من متعدد، حيث يرى ميسك (Messick, 1995) أنّ لجوء الطلبة لعملية التخمين خلال إجابتهم عن فقرات اختبار الاختيار من متعدد يُضيف تبايناً لا يرتبط بالتباين في القدرة الحقيقية (أي ذلك التباين الذي يعكس قدرة الاختبار على قياس ما وُضع لقياسه)؛ ممّا يترتب على ذلك تخفيض صدق الاختبار بشكل ملموس؛ والذي يُقاس في إطار النظرية الكلاسيكية في القياس بنسبة التباين الحقيقي المرتبط بالسمة المقاسة من التباين الكلي، وإلى جانب ذلك فإن التخمين يُعّد نوعاً من الخطأ العشوائي في القياس، فهو أيضاً عاملاً مؤثراً في دقة تقدير القدرة للمفحوصين، كما ويؤثر على بناء الاختبار حيث يتوقف قبول الفقرات أو رفضها على معاملي الصعوبة، والتمييز اللذان يتأثران بدورهما بعامل التخمين، ويرى كل من بارهلل وبودسكيو واتالي (Bar-Hillel, Budescu & Attali, 2005) أنّ التخمين قد يؤثر أيضاً في ثبات الاختبار من ناحيتين، الأولى: إنّ ضيق الوقت في نهاية بعض الاختبارات قد يدفع الطالب إلى التخمين العشوائي السريع للإجابة عن باقي فقرات الاختبار، وهذا قد يؤدي إلى عدم اتساق إجابته عن هذه الفقرات المتبقية مع الإجابة عن الفقرات التي استغرق فيها وقتاً كافياً للإجابة في بداية الاختبار؛ الأمر الذي يؤثر على ثبات ودرجة اتساق الإجابة عن الاختبار، أما الثانية، فتتمثل في أنه عند تطبيق الاختبار مرة أخرى فمن المحتمل ألاّ يحصل المفحوص الذي يلجأ للتخمين العشوائي على الدرجة نفسها؛ لأنه قد يُغير إجابته عن فقرات الاختبار التي قام بتخمينها على أساسٍ غير منطقي (Camili, 2006; McMillan, 2013).

وبناءً على ما تقدم، ونظراً لأهمية معالجة أثر التخمين في اختبارات الاختيار من متعدد، فقد برزت جهود كثيرة لإيجاد طرق تنطلق من أطر ومفاهيم نظرية متباينة، وتعتمد أساليب احصائية مختلفة، للمساعدة في الحد من أثر عامل التخمين على دقة وصدق وثبات نتائج الاختبارات الموضوعية، سواءً كان ذلك في إطار نظرية القياس الكلاسيكية، أو في نظرية الاستجابة للفقرة، وتأخذ طرق ضبط أثر التخمين في إطار النظرية الكلاسيكية في القياس بعين الاعتبار ثلاثة مواقف عند إجابة المفحوص على فقرات الاختيار من متعدد هي: إن المفحوص يعرف الإجابة الصحيحة، أو أن المفحوص لا يعرف الإجابة الصحيحة، ويترك الفقرة دون إجابة، أو أن المفحوص لا يعرف الإجابة الصحيحة على الفقرة، ولكنه يُجيب عن الفقرة بشكل عشوائي، وقد تعددت

الطرق المستخدمة لضبط أثر التخمين في اختبارات الاختيار من متعدد، إذ تستخدم هذه الطرق أساليب إحصائية مختلفة أثناء تصحيح الاختبار، من شأنها أن تحد من أثر التخمين العشوائي وما يسببه من أثرٍ على خصائص الاختبار، وفيما يأتي عرض لطرق ضبط أثر التخمين الكلاسيكية التي تم استخدامها في هذه الدراسة، وهي:

أولاً: طريقة معادلة التصحيح لأثر التخمين/ الطريقة التقليدية (Scoring Formula-SF):

تعد هذه الطريقة من أكثر الطرق شيوعاً، وتقوم على معاقبة المفحوص جراء لجوئهللتخمين خلال إجابته على فقرات الاختيار من متعدد، عن طريق حذف الدرجات التي يُقدر أنه قد حصل عليها من خلال التخمين، إذ تفترض هذه الطريقة أن كل استجابة خطأ هي نتيجة تخمين خطأ قام به المفحوص، حيث لا تأخذ هذه الطريقة بعين الاعتبار المعلومات الجزئية التي ربما يكون قد استند إليها المفحوص في إجابته عن فقرات الاختبار، وتبعاً لهذه الطريقة تقدّر علامة المفحوص على الاختبار من خلال المعادلة الآتية:

$$F = R - \frac{W}{A-1} \dots\dots\dots\dots\dots (1)$$

حيث:

F: الدرجة النهائية المصححة.

R: عدد الإجابات الصحيحة.

W: عدد الإجابات الخاطئة.

A: عدد البدائل.

فعلى سبيل المثال لو أجاب مفحوص على (20) فقرة من أصل (30) فقرة اختيار من متعدد ذات أربعة بدائل فإنّ درجته النهائية هي:

$$F = 20 - \frac{10}{3} = 16.67$$

ويشير لسابي وفالكي ولسج (Sabbe, Valcke & Lesage, 2013) إلى أنّ هذه الطريقة لا تُشجع على التخمين العشوائي الذي يُضخم درجات المفحوصين، كما إنها تعطي تقديراً غير متحيز للمعرفة الحقيقية التي يمتلكها المفحوص بناءً على أدائه في الاختبار، علاوةً على أنها تزيد من صدق الاختبار وثباته مقارنة مع طريقة التصحيح القائمة على عدد الإجابات الصحيحة، بينما يرى كل من مهرنز ولهمن (Mehrens & Lehmann, 1991) أنّ استخدام المعادلة السابقة يميل لمُعاقبة المفحوصين ذوي المعرفة الجزئية أكثر من ذوي المعلومات غير الصحيحة، وفي هذا السياق يقترح فراي (Fray, 1988) لمعالجة ميل هذه الطريقة لمعاقبة التخمين المترتب على المعرفة الجزئية أن يعُطى المفحوص تعليمات تُفيد أن عليه أن يجيب عن كل فقرة يشعر أن لديه عنها معرفة جزئية بها ترشده لاختيار البديل الصحيح، أو تمكّنه من استبعاد أيٍ من البدائل التي يعتقد جازماً أنها غير صحيحة، ثم يقوم بالتخمين من بين ما تبقى من البدائل، ويجب أن تكون تعليمات الاختبار واضحة لإعلام المفحوص، وإقناعه باتباع هذه الاستراتيجية عند أخذ الاختبار، وعندها يرى فراي أن استخدام معادلة التصحيح السابقة في ظل هذه التعليمات يُمَكّن المفحوص من استغلال المعرفة الجزئية لديه، إذ إنه سيحذف فقط الفقرات التي يعتقد أن إجاباته عنها ناتجة عن التخمين العشوائي.

ثانياً: طريقة تصحيح المموهات (Distraction-Scoring Formula-DFS):

لقد اقترح هذه الطريقة دودين (Dodeen, 2005) وما يُميّز هذه المعادلة أنها تأخذ بعين الاعتبار مستوى جاذبية المموه (البديل) في حساب درجة السؤال النهائية، وتفترض أنّ التخمين ليس عشوائياً دائماً، وأن الاختيار الأكثر جاذبية هو الاختيار الأقرب إلى الصواب من ضمن قائمة البدائل في سؤال الاختيار من متعدد، حيث يعطى كل بديل (مموه) خاطئ وزناً حسب نسبة اختياره من قبل المفحوصين، ثم يتم ترتيب البدائل حسب نسب اختيارها (جاذبيتها) من الأعلى نسبةً إلى الأقل، ثم تطبيق الصيغة الرياضية الآتية:

$$F = R - \frac{\sum_{r=1}^{k} M_r}{A-1} \dots\dots\dots\dots\dots (2)$$

حيث:

F: الدرجة النهائية المصححة.

R: عدد الإجابات الصحيحة.

K: عدد الإجابات الخاطئة.

A: عدد البدائل.

M_r: (A-1/1) إذا اختار المفحوص البديل الخاطئ ذي الرتبة الأولى، (A-1/2) إذا اختار المفحوص البديل الخاطئ ذي الرتبة الثانية، (A-1/3) إذا اختار المفحوص البديل الخاطئ ذي الرتبة الثالثة، وهكذا.

وعلى سبيل المثال يوضح الجدول (1) (10) فقرات أجاب عليها مفحوص ما بطريقة خاطئة، من أصل (30) فقرة اختيار من متعدد بأربعة بدائل.

الجدول (1)

تحليل البدائل الخاصة بالاستجابات الخاطئة

الفقرة	1	2	3	4	5	6	7	8	9	10	مجموع Mr
ترتيب البدائل الخاطئة حسب نسب اختيارها	a b c	c b d	a b D	d a b	c a d	d b a	a b c	c b d	c a d	d b a	
البديل الذي اختاره المفحوص	a	b	D	d	c	a	a	b	c	a	
قيمة Mr	**3/1**	**3/2**	**1**	**3/1**	**3/1**	**1**	**3/1**	**3/2**	**3/1**	**1**	**6**

وعليه تكون درجة الطالب النهائية هي:

$$F = 20 - \frac{6}{3} = 18$$

ويرى دودين (Dodeen, 2005) أن خصائص الفقرات (صعوبة، تمييز، فعالية المموهات) هي التي تقرر خصائص الاختبار، وبالتالي موقع المفحوص النسبي على متصل السمة في ضوء النظرية الكلاسيكية في القياس، وأن طرقاً أخرى مثل طريقة التصحيح لأثر التخمين لا تأخذ بعين الاختبار مموهات الفقرة وجاذبيتها أثناء التصحيح لأثر التخمين، التي تعبّر بطريقة ما عن المعرفة الجزئية لدى المفحوصين، وليست الإجابات الخاطئة دائماً هي تخمين عشوائي كما تفترضها طريقة التصحيح لأثر التخمين، فأحياناً يكون اختيار البديل الخاطئ ذي الجاذبية الأعلى بناءً على أفكار أو معرفة جزئية.

ثالثاً: طريقة استخدام البديل "لا أعرف الإجابة" (Don't Know Option):

تقوم هذه الطريقة على استخدام البديل "لا أعرف الإجابة " بين بدائل فقرات الاختبار، ومن ثم يتم استخدام طريقة التصحيح لأثر التخمين من خلال المعادلة (1)، حيث يتم تصحيح التخمين للفقرات الخاطئة فقط، أما الفقرات التي يجيب عليها المفحوص بلا أعرف الإجابة، فإنه يأخذ عليها الدرجة صفراً، ولكن لا يتم معاقبته عليها من خلال تصحيح التخمين، فعلى سبيل المثال لو أجاب مفحوص على (20) فقرة إجابة صحيحة من أصل (30) فقرة اختيار من متعدد، وأجاب على (7) فقرات بـ "لا أعرف الإجابة"، وأجاب بشكل خاطئ على (3) فقرات فإن درجته النهائية هي:

$$F = 20 - \frac{3}{3} = 19$$

وتستخدم هذه الطريقة بشكل كبير في الكليات الطبية، وهي تُسهم بشكل فعّال في الحد من أثر التخمين على درجات اختبار الاختيار من متعدد، ويرى كل من مجتجنز ومامرنبم ووهجنبوم وايفرز (Muijtjens, Mamerenp, Hoogenboom and Evers, 1999)

أن الاختبارات الموجهة لقياس مدى تحسن الطلبة عبر المحتوى الدراسي كاملاً لا بدّ أن تتضمن مثل هذا البديل؛ لأنه من المتوقع عدم تمكّن بعض الطلبة من الوقوف على كافة النتاجات التعليمية التي يسعى الاختبار لتقييمها، وتوصّلواد وهاريسون وهولاند وماتك وريكتس (Wade, Harrison, Hollands, Mattick, Ricketts, 2011) من خلال دراسة تجريبية قاموا بها إلى أهمية البديل "لا أعرف الإجابة " في التحكم بأثر التخمين، كما وجدوا أن الطلبة الذين تقدموا لاختبار لا تتضمن فقراته البديل "لا أعرف الإجابة " يرون أن الحظ والتخمين ساهما بشكل كبير في نجاحهم في الاختبار أكثر من المعرفة التي يمتلكونها، وفي هذا السياق يرى بورتون (Burton, 2002) أن استخدام البديل "لا أعرف الإجابة" يزوّد المفحوصين، والمعلمين بالتغذية الراجعة المناسبة حول طبيعة الاختبار، وحول جوانب القصور لدى المفحوصين؛ ليصار لاحقاً إلى معالجتها في المواقف التعليمية داخل الغرفة الصفيّة، بينما يرى كل من اسبونزا وجاردازبل(Espinoza & Gardeazabal,2010) أن إضافة البديل "لا أعرف الإجابة" إلى بدائل فقرات الاختبار يقود إلى خطأ قياس من خلال التحيز ضد المفحوصين الذين يميلون للمخاطرة والذين يختارون بديلاً آخر غير البديل "لا أعرف الإجابة" حتى عندما لا يمتلكون المعرفة التامة بالإجابة الصحيحة.

ويلاحظ من خلال الأمثلة السابقة أن عدد الإجابات الخاطئة التي حصل عليها المفحوص متساوية، إلا أن الدرجة النهائية المعدّلة (المصحّحة) في كل طريقة كانت مختلفة، إذ يعود هذا الاختلاف إلى الأساس النظري الذي تنطلق منه كل طريقة، والإجراء الإحصائي المتبع في استخراج الدرجة المعدّلة.

مشكلة الدراسة وأسئلتها

بناءً على ما سبق يمكن القول إنّضبط أثر التخمين في اختبارات الاختيار من متعدد عملية بالغة الأهمية، وذات تأثير مباشر على الخصائص السيكومترية للاختبار، ومن ثمّعلى دقة وموضوعية النتائج المتأتية منها، وبالتالي دقة القرارات التربوية المتّخذة في ضوء هذه النتائج وعدالتها، وأن لكل طريقة من هذه الطرق مزاياها وسلبياتها وأثرها المختلف على خصائص الاختبار، وأن الجهود لا زالت مستمرة للحد من أثر التخمين من خلال إيجاد طرق حديثة تعتمد على المعرفة الجزئية في تصحيح الاختبارات الموضوعية، ومن هنا جاءت الدراسة الحالية استكمالاً للجهود التي بُذلت في إطار كل من الأدب التجريبي والنظري الذي تناول موضوع التخمين، ومعالجته أو التخفيف من آثاره في مجال القياس النفسي والتربوي، من خلال مقارنة ثلاث طرق لضبط أثر التخمين، هي: طريقة تصحيح أثر التخمين (طريقة قديمة لا تعتمد المعرفة الجزئية)، طريقة تصحيح المموهات (طريقة حديثة تعتمد المعرفة الجزئية)، وطريقة البديل "لا أعرف الإجابة" (طريقة حديثة تعتمد المعرفة الجزئية)، باستخدام اختبار رياضيات تحصيلي للصف الأول الثانوي العلمي.

وتسعى هذه الدراسة للإجابة على الأسئلة الآتية:

1. هل يوجد أثر ذو دلالة احصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في متوسط أداء الطلبة على الاختبار يعزى لطريقة ضبط أثر التخمين ؟
2. هل يوجد أثر ذو دلالة احصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في صدق الاختبار يعزى لطريقة ضبط أثر التخمين ؟
3. هل يوجد أثر ذو دلالة احصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في ثبات الاختبار يعزى لطريقة ضبط أثر التخمين ؟

أهمية الدراسة

تكتسب الدراسة جانباً من أهميتها في كونها استهدفت مقارنة طريقة تقليدية شائعة هي طريقة تصحيح أثر التخمين، وطريقتين حديثتين هما طريقة تصحيح المموهات وطريقة البديل "لا أعرف الإجابة"، علماً بأن الطريقتين الحديثتين لم يسبق – في ضوء علم الباحثين– مقارنة نتائجهما وأثرهما على خصائص الاختبار، كما ويمكن أن تسهم هذه الدراسة في توضيح معالم جديدة لمصممي الاختبارات في المجال التربوي، بأسلوب جديد في قياس تحصيل الطلبة، وتصحيح الاختبارات الموضوعية، واتخاذ القرارات التربوية المختلفة.

حدود الدراسة ومحدداتها:

1- اقتصرت طرق ضبط أثر التخمين على ثلاث طرق هي: طريقة تصحيح أثر التخمين، طريقة تصحيح المموهات، وطريقة البديل لا أعرف الإجابة.

2- اقتصرت عينة الدراسة على طلبة الصف الأول الثانوي العلميفي المدارس الحكومية الأردنية، أثناء الفصل الدراسي الأول

من العام الدراسي (2016/2015م).

3- اقتصر المحتوى الدراسي على مبحث الرياضيات.

4- اعتماد نتائج الدراسة على مدى صدق وثبات أداة الدراسة.

التعريفات المفاهيمية والاجرائية:

- **التخمينGuessing:**

اختيار الطالب للبديل الصحيح من خلال الصدفة، ويسمى تخميناً عشوائياً، أو من خلال المعرفة الجزئية ويسمى تخميناً ذكياً، ويستدل على الفقرة التي تم تخمينها إجرائياً من خلال الإجابة الخاطئة.

- **ضبط أثر التخمين Control the Effect of Guessing**

طرق إحصائية يتم من خلالها تصحيح (تعديل) درجة الطالب حسب معادلات رياضية، للحد من أثر التخمين بناءً على عدد الفقرات التي تم تخمينها.

- **الدرجة المصححة Correct Score**

درجة الطالب النهائية بعد تصحيحها من أثر التخمين.

- **طريقة التصحيح لأثر التخمين Scoring Formula-SF**

طريقة تقوم على حذف الدرجات التي يُقدر أن المفحوص حصل عليها من خلال التخمين، إذ تفترض هذه الطريقة أن كل استجابة خطأ هي نتيجة تخمين قام به المفحوص، حيث لا تأخذ هذه الطريقة بعين الاعتبار المعلومات الجزئية التي ربما يكون قد استند إليها المفحوص في إجابته عن فقرات الاختبار(Mehrens and Lehmann, 1991).

- **طريقة تصحيح المموهاتDistraction-Scoring Formula-DFS**

طريقة تأخذ بعين الاعتبار مستوى جاذبية المموه (البديل) في حساب درجة السؤال النهائية، وتفترض أنّ التخمين ليس عشوائيًا دائمًا، وأن الاختيار الأكثر جاذبية يعبّر بطريقة ما عن المعرفة الجزئية لدى المفحوصين (Dodeen, 2005).

- **طريقة البديل لا أعرف الإجابةDon't Know Option**

تقوم هذه الطريقة على استخدام البديل "لا أعرف الإجابة " بين بدائل فقرات الاختبار، ومن ثم يتم استخدام طريقة التصحيح لأثر التخمين، حيث يتم تصحيح التخمين للفقرات الخاطئة فقط، أما الفقرات التي يجيب عليها المفحوص بـ "لا أعرف الإجابة" فإنه يأخذ عليها الدرجة صفراً، ولكن لا يتم معاقبته عليها من خلال تصحيح التخمين (Muijtjens et al, 1999).

الدراسات السابقة

هناك كثير من الدراسات التي تناولت موضوع مقارنة ضبط أثر التخمين في اختبارات الاختيار من متعدد، وأثرها على خصائص الاختبار، إلا أنه في هذه الدراسة سيتم استعراض الدراسات السابقة التي تناولت الطرق الكلاسيكية المستهدفة في هذه الدراسة ومنها:

أجرى سواقد (1983) هدفت إلى مقارنة أثر ثلاث طرق لتصحيح اختبارات الاختيار من متعدد على الخصائص السيكومترية للاختبار وعلى أداء المفحوصين من مستويات تحصيل ودرجات مخاطرة مختلفة، والطرق هي: التصحيح التقليدي دون ضبط أثر التخمين، طريقة معادلة التصحيح لأثر التخمين، وطريقة الاختيار الجزئي، باستخدام ثلاث صور متكافئة لاختبار تحصيلي في مبحث العلوم للصف التاسع الأساسي في الأردن، طبّق على عينة مكونة من (173) طالباً، حيث أظهرت النتائج فروق ذات دلالة احصائية عند مستوى دلالة ($\alpha \leq 0.05$) في قيم المتوسطات الحسابية للمفحوصين ذوي المخاطرة العالية والمنخفضة عند التصحيح بالطريقة التقليدية وطريقة الاختيار الجزئي، ووجود فروق لصالح ذوي المخاطرة العالية عند استخدام معادلة التصحيح لأثر التخمين، كما وأشارت النتائج إلى زيادة ذات دلالة احصائية عند مستوى دلالة ($\alpha \leq 0.05$) في قيم الثبات عند استخدام طريقة الاختيار الجزئي، وزيادة ذات دلالة احصائية عند مستوى دلالة ($\alpha \leq 0.05$) في قيم صدق المحك التلازمي عند استخدام طريقة الاختيار الجزئي أيضاً، وأوصت الدراسة باستخدام طريقة الاختيار الجزئي في تصحيح اختبارات الاختيار من متعدد؛ لقدرتها على ضبط أثر التخمين، واستغلال المعرفة الجزئية للطلبة.

أجريمجتجنز وآخرون (Muijtjens et al, 1999) دراسة هدفت إلى المقارنة بين طريقة استخدمت البديل "لا أعرف الإجابة" بين المموهات وطريقة التصحيح لأثر التخمين التقليدية، بدون استخدام البديل "لا أعرف الإجابة"، وذلك في جامعة ماسترشت

(Maastricht) في بلجيكيا، حيث أعد الباحثان اختباراً تحصيلياً مكوناً من (180) فقرة، وتكونت عينة الدراسة من (212) طالباً وطالبةً من طلاب السنة الثانية في كلية الطب، و(151) طالباً وطالبةً في السنة الثالثة، وأظهرت نتائج الدراسة أن طريقة "لا أعرف الإجابة"، أظهرت قيماً أعلى للمتوسطات الحسابية، وقيماً أقل للانحراف المعياري مقارنةً بطريقة تصحيح أثر التخمين، كما وأظهرت طريقة "لا أعرف الإجابة" قيماً أعلى للثبات باستخدام العينة الكلية، وعينات فرعية أخرى مشتقة منها، ويرى الباحثون أن هذه الطريقة تقود إلى معلومات أدق عن جودة الفقرة وثباتها، ويوصي الباحثون بأن الطريقة المثلى لتصحيح أثر التخمين هي الطريقة الأكثر ثباتاً، والأقل تحيزاً، مع الأخذ بعين الاعتبار الأسباب النفسية والتربوية.

كما وأجرى النبهان (Alnabhan, 2002) دراسة هدفت إلى تقصّي أثر ثلاث طرق للتصحيح ضد أثر التخمين، وهي: (عدد الإجابات الصحيحة، ومعاقبة التخمين، ومكافأة التخمين الجزئي) على الخصائص السيكومترية (الصدق والثبات) للاختبار، حيث تكوّنت عينة الدراسة من (120) طالباً وطالبة في مرحلة ما قبل التخرج في جامعة أبوظبي في دولة الإمارات العربية المتحدة، وتألّف اختبار الدراسة من (40) فقرة اختيار من متعدد بواقع أربعة بدائل لكل فقرة في مساق منهجية البحث النفسي، كما استخدم الباحث اختباراً قصيراً مكونًا من (10) فقرات للمحتوى التعليمي نفسه الذي يغطيه الاختبار المكون من (40) فقرة وببدائل جميعها خاطئة لكل فقرة؛ من أجل تصنيف المفحوصين إلى: ذوي مخاطرة مرتفعة وذوي مخاطرة متدنية، فالطالب الذي أجاب عن عدد أكبر من الفقرات عُدّ ذا مخاطرة مرتفعة، والذي أجاب عن عدد أقل من الفقرات عُدّ ذا مخاطرة منخفضة، وقد أظهرت نتائج الدراسة أن الطرق الثلاثة المختلفة أعطت معاملات صدق وثبات مختلفة عن طريقة المعرفة الجزئية –التي تقوم على افتراض أن المفحوص يمتلك معرفة تمكنه من حذف بعض البدائل عند إجابته عن الفقرة–، حيث كان معامل الثبات للاختبار المصحح ضد أثر التخمين باستخدام طريقة المعرفة الجزئية (0.81) بينما كان (0.65) عند استخدام طريقة التصحيح باستخدام عدد الإجابات الصحيحة، كما بلغ (0.73) عند استخدام طريقة معاقبة التخمين، وأكدت الدراسة أن التخمين يُقلل من صدق، وثبات الاختبار من خلال إعطاء تباين غير حقيقي في نتائج الاختبار.

كما أجرى الخرشة (2004) دراسة هدفت إلى المقارنة بين ثلاث طرق(المكأفاة، والعقاب والمكافأة والعقاب معاً) للتصحيح ضد أثر التخمين في اختبارات الاختيار من متعدد، حيث قام الباحث بإعداد اختبار تحصيلي مكون من (60) فقرة من نوع الاختيار من متعدد، بواقع أربعة بدائل للفقرة الواحدة بهدف قياس المصطلحات الأساسية في مقرر مبادئ القياس والتقويم، وتكونت عينة الدراسة من (197) طالبًا وطالبة في كلية التربية في جامعة مؤتة في العام الدراسي (2004/2003م)، وللحكم على فاعلية الطرق المُستخدمة، قام الباحث باستخدام المؤشرات الإحصائية الآتية: اختبار (ت) للعينات المستقلة، ومعاملات الثبات، ومقاييس النزعة المركزية، ومقاييس التشتت، والخطأ المعياري للقياس لعلامات المفحوصين على وفق كل طريقة من طرق التصحيح لأثر التخمين، وفي ضوء نتائج الإحصاءات المستخرجة أوصت الدراسة باستخدام معادلة التصحيح من أثر التخمين (العقاب والمكافأة) في تصحيح اختبارات الاختيار من متعدد.

وفي دراسة أجراها الزبون (2004) هدفت إلى المقارنة بين أثر ثلاث طرق لتصحيح الإجابات على الاختبارات الموضوعية/ الاختيار من متعدد لضبط التخمين على الخصائص السيكومترية للاختبار والفقرة في ضوء نظرية الاستجابة للفقرة، حيث جرى استخدام الطريقة التقليدية، ومعادلة التصحيح لأثر التخمين وطريقة الاختيار الجزئي، حيث تمّ إعداد ثلاث صور متكافئة لاختبار تحصيلي لطلبة مساق (الإحصاء التربوي) في جامعة مؤتة، وجرى تطبيقها على عينة مكونة من (406) طلاب في كلية التربية في جامعة مؤتة، جرى تصحيح الإجابات على كل صورة من صور الاختبار بإحدى طرق التصحيح الثلاثة، ثم حُلّلت البيانات باستخدام برمجية ونستب (Winstep) وجرى تقدير مَعلمة القدرة (θ) والخطأ المعياري في القياس لقدرة كل فرد، كما تم استخراج مؤشرات على معاملات الثبات لتباعد الأفراد، ومعاملات صدق المحك التلازمي، ومعاملات صعوبة الفقرة، ومعاملات تمييز الفقرة.

وأظهرت نتائج الدراسة وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) بين معاملات الثبات لصالح استخدام كل من طريقة (التصحيح لأثر التخمين) وطريقة (الاختيار الجزئي)، كما أشارت النتائج إلى وجود فروق ذات دلالة إحصائية بين معاملات صدق الاختبار لصالح الطريقة التقليدية، وطريقة الاختيار الجزئي، إضافة إلى ذلك أظهرت الدراسة أن هناك فروقاً ذات دلالة إحصائية بين معاملات صعوبة الفقرة، لصالح معادلة التصحيح لأثر التخمين، وأن هناك فروقاً ذات دلالة إحصائية بين معاملات تمييز الفقرة عند استخدام الطريقة التقليدية وطريقة (الاختيار الجزئي).

وأجرى دودين (Dodeen, 2005) دراسة قارن فيها بين طريقة معادلة تصحيح المموهات التي تأخذ بعين الاعتبار مستوى جاذبية

المموه (البديل) في حساب درجة السؤال النهائية، والطريقة التقليدية لتصحيح أثر التخمين، كما قام بمقارنة فعالية المعادلتين، لاسيما تأثير كلّ منهما على أداء الطلبة على الاختبار، وقيمة الثبات، من خلال تطبيق المعادلتين على نتائج اختبارين معياريين في الرياضيات واللغة الإنجليزية، على عينة مكونة من (148) طالباً وطالبةً من طلبة جامعة الإمارات العربية المتحدة، وقد دلّت النتائج على أفضلية استخدام معادلة تصحيح المموهات في مكافأة الطلبة على حسن اختيارهم للبديل الأقرب إلى الصواب أو الأكثر صحة عندما يكونون غير متأكدين تماماً من الجواب الصحيح، حيث أظهرت طريقة تصحيح المموهات متوسط درجات أكبر من الطريقة التقليدية، كما وأدت إلى خفض الانحراف المعياري لدرجات الاختبار، وكانت نسب التخمين أعلى للطلبة ذوي التحصيل المتوسط، بالإضافة إلى ذلك، فقد أدى تطبيق هذه المعادلة إلى زيادة ثبات نتائج الاختبارين عندما حُسبت على العينة الكلية، وعينات فرعية مشتقة من العينة الكلية لكل اختبار، في حين كانت درجات ثبات الاختبار تتناقص مع تناقص حجم العينة للعينات الفرعية.

وقام ذيابات (2008) دراسة هدفت إلى مقارنة طريقتي الحذف الجزئي والاختيار الجزئي لضبط أثر التخمين مع طريقة الثواب والعقاب في اختبارات الاختيار من متعدد، وذلك من خلال دقة تقدير العلامات، وقيم معاملات صدق المحك، وخصائص الفقرات السيكومترية (الصعوبة، التمييز)، باستخدام اختبار تحصيلي في مادة الرياضيات للصف التاسع الأساسي في الأردن مكوّن من (30) فقرة، طبّق على عينة مكونة من (124) طالباً، إذ طبّق الاختبار ثلاث مرات مع اختلاف التعليمات وطريقة الاستجابة في كل مرة، وأظهرت نتائج الدراسة عدم وجود فروق ذات دلالة احصائية بين طريقتي التصحيح المستخدمة في كل من معاملات صدق المحك، وصعوبة الفقرات، وتمييز الفقرات، وأنه يوجد فروق بين الطريقتين مع طريقة الثواب والعقاب، وكانت الفروق لصالح طريقة الثواب والعقاب، حيث أوصت الدراسة باستخدام طريقة الثواب والعقاب في تصحيح اختبارت الاختيار من متعدد للحصول على علامات أكثر صدقاً وثباتاً.

أمّا راشد (2011) فقد أجرى دراسة هدفت إلى معرفة أثر التعليمات الخاصة بالتخمين في استجابة الطلبة لفقرات اختبار موضوعي/ اختيار من متعدد على درجاتهم، وعلى الخصائص السيكومترية للاختبار (الصدق والثبات)، وعلى خصائص الفقرات (الصعوبة والتمييز)، ولتحقيق أهداف دراسته قام ببناء اختبار تحصيلي في وحدتين دراسيتين من مادة الحضارة العربية الإسلامية، وتكّون الاختبار من (40) فقرة من نوع الاختيار من متعدد، بأربعة بدائل لكل فقرة، طُبّق على عينة مكونة من (666) طالباً وطالبةً من الملتزمين وغير الملتزمين بتعليمات التخمين من كلية التربية الأساسية في جامعة الكويت، وأظهرت نتائج الدراسة وجود فروق دالة إحصائياً بين درجات التحصيل للمجموعتين، وذلك لصالح مجموعة الملتزمين بتعليمات التخمين، حيث بلغ المتوسط الحسابي لدرجاتهم (0.66) في حين كان المتوسط الحسابي لدرجات الطلبة غير الملتزمين بتعليمات التخمين (0.40).

يتضح من خلال استعراض الدراسات السابقة أن كل طريقة من طرق ضبط أثر التخمين في اختبارات الاختيار من متعدد تمتعت بخصائص مختلفة، كما أنه لا يوجد اتفاق في نتائج الدراسات حول أي الطرق أكثر فاعلية، إلا أن الطرق الحديثة المعتمدة على المعرفة الجزئية مثل طريقة استخدام البديل "لا أعرف الاجابة"، وطريقة معادلة تصحيح المموهات قد أثبتت فاعليتها أكثر من الطرق التقليدية الشائعة، وهذا يثبت دور تعليمات الاختبار واجراءاته في الحد من أثر التخمين بالإضافة إلى الاجراءات والمعالجات الإحصائية المتبعة في ضبط أثر التخمين، زيادةً على أنه لا يوجد دراسة حاولت الجمع بين الطرق المستهدفة في هذه الدراسة، وهي طريقة التصحيح لأثر التخمين، وطريقة تصحيح المموهات، وطريقة استخدام البديل "لا أعرف الإجابة"، التي ينطلق كل منها من أساس نظري، وإجراء إحصائي مختلفين.

أفراد الدراسة:

تكوّن أفراد الدراسة من (159) طالباً وطالبةً، من طلبة الصف الأول الثانوي العلمي الملتحقين بالدراسة في المدارس الحكومية الأردنية في محافظة العاصمة عمّان/ مديرية تربية لواء الجامعة للعام الدراسي (2016/2015م)، منهم (78) من الذكور، و(81) الإناث، حيث تم اختيار ستة مدارس ثانوية من مدارس المديرية، ثلاثة للذكور وثلاثة أخرى للإناث بطريقة عشوائية، ثم اختيار شعبة الأول الثانوي العلمي بصورة عشوائية في حال كان هناك أكثر من شعبة في المدرسة.

أدوات الدراسة:

لأغراض الدراسة تم بناء اختبار تحصيلي في مبحث الرياضيات للصف الأول الثانوي العلمي، باستخدام فقرات الاختيار من متعدد، يقيس قدرات رياضية (وحدات دراسية) محددة، وهي: المتتاليات والمتسلسلات، والاقترانات الأسية واللوغاريتمية، على وفق خطوات بناء الاختبار التحصيلي العلمية، وقواعد كتابة الفقرات الصحيحة، بداية من تحديد النتاجات الخاصة، ثم تحديد الأهمية

النسبية لكل وحدة دراسية، ثم تم عمل جدول المواصفات وفق الأهمية النسبية والمستويات المعرفية، وأخيراً كتابة الفقرات، وتم عرض الصورة الأولية من الفقرات والبالغ عددها (20) فقرة، (11) فقرة منها لوحدة المتتاليات والمتسلسلات، و(9) فقرات لوحدة الاقترانات الأسية واللوغاريتمية على محكمين متخصصين في مبحث الرياضيات (معلّمي الصفوف المستهدفة، ومشرفي المبحث)، للتحقق من الصدق الظاهري للاختبار، وقد طلب منهم إبداء الرأي في مدى وضوح الفقرات، وشمولها، وملائمتها للفئة العمرية المستهدفة، ثم تمّ عمل صورتين من الاختبار تحتوي الصورة الأولى على أربعة بدائل لكل فقرة واحدة منها صحيحة، في حين تم إضافة بديل خامس في الصورة الثانية لكل فقرة، وهو البديل "لا أعرف الإجابة"، وذلك للتمكّن من استخدام طريقة استخدام البديل لا أعرف الإجابة.

إجراءات الدراسة:

تضمّنت إجراءات الدراسة الخطوات الآتية:

أولاً: تطبيق الاختبار على عينة استطلاعية:

للتأكد من وضوح الصياغة اللغوية للفقرات، ووضوح تعليمات الاختبار، تم تطبيق الاختبار على عينة استطلاعية مكونة من (17) طالباً وطالبةً، من خارج العينة التجريبية وعينة التطبيق، وطلب منهم الاستجابة وإبداء ملاحظاتهم واستفساراتهم حول فقرات الاختبار، وتم الأخذ بها حتى أصبحت بصورتها النهائيةالمكوّنة من (18) فقرة، وتم حساب متوسط زمن الإجابة والذي بلغ (33) دقيقة.

ثانياً: تطبيق الاختبار على العينة التجريبية:

تمّ تطبيق الاختبار على أفراد عينة تجريبية مكوّنة من (54) طالباً وطالبةً من خارج أفراد الدراسة، للتحقق من خصائص الفقرات والاختبار، التي كانت على النحو الآتي:

صعوبة وتمييز الفقرات:

تم حساب معاملات الصعوبة من خلال معدل الاجابات الصحيحة لكل فقرة، حيث تراوحت قيم معاملات الصعوبة بين (0.27–0.76) وبمتوسط حسابي (0.53)، وهي معاملات صعوبة ملائمة، إذ غطت معظم متصل السمة للقدرة المستهدفة، كما تم حساب معاملات تمييز الفقرات من خلال معامل ارتباط بوينت بايسيريال بين الدرجة على الفقرة والدرجة الكلية للاختبار، إذ تراوحت معاملات التمييز بين (0.39–0.67) وبمتوسط حسابي (0.49) وهي معاملات تمييز جيدة وملائمة لأغراض الدراسة.

صدق الاختبار:

تمّالتحقّق من صدق الاختبار من خلال طريقتين، هما: صدق المحتوى، والصدق المرتبط بالمحك، وفيما يأتي عرض لهذه الطرق:

صدق المحتوى Content Validity:

تمّ عرض فقرات الاختبار على عدد من المحكمين ممن يمتلكون الكفاءة والخبرة العلمية، في مجال القياس والتقويم والرياضيات، حيث قام المحكمون بإجراء بعض التعديلات على صياغة بعض فقرات الاختبار، وتم الأخذ بالملاحظات التي أجمع عليها (80%) منهم.

الصدق المرتبط بالمحك Criterion-related Test Validity:

وتم ذلك من خلال إيجاد معامل ارتباط (بيرسون) بين الدرجات على الاختبار المعد لهذه الدراسة، ومتوسط اختباري الشهرين الأول والثاني المعدّين من قبل معلم المبحث، حيث بلغ معامل الارتباط بينهما (0.81)، وهو معامل ارتباط مرتفع يدل على صدق الاختبار الذي تم بناؤه في هذه الدراسة.

ثبات الاختبار:

للتحقق من ثبات الاختبار تم استخدام طريقة معامل ثبات الاتساق الداخلي (Internal Consistency Reliability)، وذلك من خلال استخدام معادلة كرونباخ الفا، إذ بلغ معامل ثبات الاختبار بهذه الطريقة (0.83) وهو معامل ثبات مرتفع يدل على تمتّع الاختبار بثباتٍ عالٍ.

ثالثاً: التطبيق على العينة النهائية:

بهدف مقارنة طرق التخمين تم تطبيق الاختبار بصورتيه على أفراد الدراسة، والبالغ عددهم (159) طالباً وطالبةً، وبفاصل زمني مقداره ثلاثة أسابيع، وتضمنت تعليمات الاختبار في الصورة الأولى أن يجيب الطالب على جميع الفقرات، واللجوء إلى

التخمين في حالة عدم معرفة الإجابة الصحيحة. أما الصورة الثانية، فنصّت تعليماتها على إجابة الأسئلة التي يعرفها الطالب فقط، واختيار البديل (لا أعرف الإجابة) في حالة عدم معرفة الإجابة الصحيحة.

رابعاً: تصحيح الاختبار:

1- الصورة الأولى: تم تصحيح الصورة الأولى للاختبار من خلال إعطاء الدرجة (1) في حال اختيار البديل الصحيح، والدرجة (0) في حال اختيار البديل الخاطئ، ثم تمّ تطبيق طريقتي ضبط أثر التخمين (تصحيح أثر التخمين، وتصحيح المموهات) باستخدام المعادلتين (2، 1) على هذه الدرجات بتحليلين منفصلين، حيث خضعت جميع الفقرات الخاطئة للتصحيح لأثر التخمين.

2- الصورة الثانية: تمّ إعطاء الدرجة (1) في حال اختيار البديل الصحيح، والدرجة (0) في حال اختيار البديل الخاطئ، والدرجة (0) في حال اختيار البديل "لا أعرف الإجابة"، ثم تمّ تطبيق المعادلة (1) لتصحيح أثر التخمين على الفقرات الخاطئة فقط، وتم استثناء الفقرات التي تم الإجابة عليها بـ "لا أعرف الإجابة" من التصحيح لأثر التخمين.

خامساً: معامل التوافق بين رتب البدائل:

تم ترتيب البدائل الخاطئة لكل فقرة من فقرات الاختبار حسب مستوى جذبها للطلبة من الأكثر إلى الأقل من قبل ثلاثة معلمين ومشرفين تربويين، بناءً على خبرتهم العلمية والعملية، وتم اعتماد الترتيب الذي اتفق عليه ثلاثة محكمين من أصل خمسة، وذلك لغاية استخدامها في حساب معامل التوافق، ثم تمّ ترتيب هذه البدائل الخاطئة حسب نسبة الطلبة الذين اختاروا كل بديل من هذه البدائل، بطريقة مماثلة للمثال في الجدول (1)، وحُسب معامل التوافق بين التصنيفين (الترتيبين) باستخدام معامل ارتباط سبيرمان، إذ بلغ معامل الارتباط (0.69) وكان ذا دلالة إحصائية عند مستوى دلالة ($\alpha \leq 0.01$)، وتدعم هذه القيمة دقة استخدام طريقة تصحيح المموهات.

المعالجات الإحصائية:

للإجابة عن أسئلة الدراسة تم استخدام المعالجات الإحصائية الآتية:

1- المتوسطات الحسابية والانحرافات المعيارية لأداء الطلبة على الاختبار.
2- معامل ارتباط بوينت بايسيريال لحساب معاملات تمييز الفقرات.
3- معامل ارتباط سبيرمان لحساب معامل التوافق.
4- معامل ارتباط (بيرسون) لحساب صدق المحك التلازمي.
5- معادلة كرونباخ الفا لحساب ثبات الاتساق الداخلي.
6- تحليل التباين الثلاثي للقياسات المتكررة (Repeated Measurement Design).
7- اختبار أقل فرق دال [Least Square Differences (LSD)].
8- اختبار (T- Test) لمقارنة معاملات صدق المحك التلازمي.

نتائج الدراسة ومناقشتها

هدفت هذه الدراسة إلى مقارنة ثلاث طرق لضبط أثر التخمين هي: طريقة تصحيح أثر التخمين، طريقة تصحيح المموهات، وطريقة البديل "لا أعرف الإجابة"، باستخدام اختبار رياضيات تحصيلي للصف الأول الثانوي العلمي، وفيما يأتي أهم النتائج التي تم التوصّل إليها، ومناقشة هذه النتائج.

النتائج المتعلقة بالسؤال الأول:

هل يوجد أثر ذو دلالة احصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في متوسط أداء الطلبة على الاختبار يُعزى لطريقة ضبط أثر التخمين؟

للإجابة عن هذا السؤال تم حساب مجموعة من الإحصاءات الوصفية لأداء الطلبة على فقرات الاختبارعلى وفق الطرق المستخدمة لضبط أثر التخمين، ويبيّن الجدول (2) طبيعة الأداء على فقرات الاختبار على وفق كل طريقة.

الجدول (2)

الإحصاءات الوصفية لأداء الطلبة على فقرات الاختبار

الطريقة	المتوسط الحسابي	الانحراف المعياري	الالتواء	التفلطح	المئينات				
					10	25	50	75	90
تصحيح تقليدي بدون ضبط	10.45	3.18	-0.07	-0.61	6.00	8.00	10.00	13.00	15.00
تصحيح أثر التخمين	7.82	4.12	0.03	-.59	2.00	4.70	7.30	10.00	14.00
طريقة تصحيح المموهات	9.06	3.39	-0.03	-0.54	4.46	6.78	9.12	11.79	13.23
طريقة "لا أعرف الإجابة"	8.45	3.93	-0.16	-0.42	2.70	6.00	8.70	11.00	14.00

يتضح من خلال الجدول (2) وجود فروق في المتوسطات الحسابية لأداء الطلبة على الاختبار، حيث أظهرت طريقة تصحيح أثر التخمين أقل متوسط حسابي والبالغ (7.82)، تليها طريقة لا أعرف الإجابة التي بلغ متوسطها الحسابي (8.45)، بفارق مقداره (2.5%) عن طريقة تصحيح أثر التخمين، أما طريقة تصحيح البدائل، فقد أعطت أعلى متوسط حسابي ومقداره (9.06)، بفارق مقداره (6%) عن طريقة ضبط أثر التخمين، و(3%) عن طريقة "لا أعرف الإجابة".

أما الانحراف المعياري، فقد أظهرت الطرق الثلاثة تفاوتاً في قيمه، حيث أعطت طريقة تصحيح البدائل أقل انحراف معياري، تليها طريقة "لا أعرف الإجابة"، وأخيراً طريقة تصحيح أثر التخمين التي أظهرت أعلى قيمة للانحراف المعياري.

ويلاحظ من خلال قيم الالتواء، والتفلطح أن التوزيعات الأربعة كان لها قيم التواء، وتفلطح سالبة قريبة من الصفر، ويدل الالتواء السالب على أن الأفراد ذوي التحصيل المرتفع خمّنوا الإجابات بشكلٍ أقل من ذوي التحصيل المنخفض، أما التوزيعات المئينية للطرق الأربعة، فأشارت إلى أن طريقة تصحيح أثر التخمين أعطت درجات كلية أكبر للطلبة من ذوي التحصيل المنخفض (المقابلة للمئين 10)، وذوي التحصيل المتوسط (الواقعة بين المئين 25 والمئين 75)، أما الدرجات الكلية للطلبة ذوي التحصيل المرتفع (المقابلة للمئين 90)، فكانت أقل في هذه الطريقة، وفيما يخص طريقة "لا أعرف الإجابة" فأظهرت درجات أقل لذوي التحصيل المنخفض والمتوسط ودرجات أعلى لذوي التحصيل المرتفع من طريقة تصحيح المموهات. وأخيراً أظهرت طريقة تصحيح أثر التخمين أقل تقدير لدرجات الطلبة ذوي التحصيل المنخفض والمتوسط، – وتعد هذه الفئة الأكثر ميلاً للتخمين العشوائي –، وتقديراً لدرجات ذوي التحصيل المرتفع مماثلاً لطريقة لا أعرف الإجابة.

وفي تحليل آخر تم حساب معاملات ارتباط بيرسون بين الدرجة الكلية بدلالة عدد الإجابات الصحيحة، والفرق بين الدرجات المعدّلة لكل طريقتين، ويوضّح الجدول (3) قيم هذه الارتباطات.

الجدول (3)

نتائج الارتباطات بين فروق الدرجات المعدّلة والدرجة الكلية

الفرق بين الدرجات المعدلة	الدرجات الكلية
تصحيح أثر التخمين وتصحيح المموهات	-0.79**
"لا أعرف الإجابة" وتصحيح أثر التخمين	-0.63**
"لا أعرف الإجابة" وتصحيح المموهات	-0.59**

****معامل ارتباط دال إحصائياً عند مستوى دلالة (0.01)**.

ويتضح من خلال الجدول (3) أن جميع معاملات الارتباط كانت دالة عند مستوى دلالة ($\alpha \leq 0.01$)، وجميعها كانت ذات قيم سالبة، وتدل القيمة السالبة على أن الطلبة ذوي التحصيل المرتفع لجأوا إلى التخمين بشكلٍ أقل من ذوي التحصيل المنخفض، أي

أن الطلبة ذوي التحصيل المنخفض حصلوا على درجات تُعزى للتخمين أكثر من ذوي التحصيل المرتفع، حيث يقل التخمين بزيادة القدرة، وكانت القيمة الأكبر لمعامل الارتباط ناتجة عن الفروق بين طريقة تصحيح أثر التخمين، وطريقة تصحيح المموهات، كما يبيّن الجدول (4) أكبر ثلاث قيم فرق بين الدرجات المعدّلة لكل طريقتين، ومدى الدرجات الكلية المقابلة لها، وكذلك لأقل ثلاث قيم.

الجدول (4)

أكبر وأقل قيم فرق بين الدرجات المعدلة لكل طريقتين والدرجات الكلية المقابلة لها

الفرق بين الدرجات المعدلة	أكبر ثلاث قيم			مدى الدرجات الكلية المقابلة لها	أقل ثلاث قيم			مدى الدرجات الكلية المقابلة لها
تصحيح أثر التخمين وتصحيح المموهات	3.45	3.48	3.53	9−4	0.01	0.08	0.12	17-14
"لا أعرف الإجابة" وتصحيح أثر التخمين	7.00	8.30	10.60	9-5	0.00	0.30	0.40	17-9
"لا أعرف الإجابة" وتصحيح المموهات	5.43	5.85	8.74	8-5	0.03	0.10	0.12	17-8

ويتضح من خلال الجدول (4) أن قيم الفرق الكبيرة كانت مقابلة للدرجات الكلية التي تقع بين المئينين (10، 75)، وهي خاصة بالطلبة متوسطي ومتدني القدرة (التحصيل)، في حين كانت أقل الفروق تقابل درجات كلية تقع بين المئينين (90، 99) لطريقتي تصحيح أثر التخمين وتصحيح المموهات، أي الطلبة ذوي التحصيل المرتفع، أما المقارنة بين طريقتي تصحيح أثر التخمين و"لا أعرف الإجابة"، وطريقتي تصحيح المموهات و"لا أعرف الإجابة" كانت القيم الدنيا تقابل درجات كلية تقع بين المئين (50) والمئين (99)، أي أنها شملت الطلبة ذوي التحصيل المتوسط والمرتفع.

ولمعرفة فيما إذا كانت الفروق في متوسطات أداء الطلبة على الاختبار تبعاً لطريقة التصحيح المستخدمة ذات دلالة احصائية عند مستوى الدلالة ($\alpha \leq 0.05$) تم عمل تحليل التباين الثلاثي للقياسات المتكررة (Repeated Measurement Design)، ويوضح الجدول (5) نتائج تحليل التباين الثلاثي للقياسات المتكررة.

الجدول (5)

نتائج تحليل التباين الثلاثي للقياسات المتكررة

لأثر الطريقة المستخدمة لضبط أثر التخمين في متوسط أداء الطلبة على الاختبار

اختبار الآثار:	مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	ف	الدلالة الإحصائية
داخل الأفراد	الطريقة	100.661	2	50.330	15.178	0.000*
	الخطأ	1047.889	316	3.316		

يلاحظ من نتائج الجدول (5) وجود أثر ذو دلالة إحصائية في متوسطات أداء الطلبة على الاختبار باختلاف الطريقة المستخدمة للتصحيح من أثر التخمين، إذ كانت قيمة مستوى الدلالة الخاصة بـ(ف) أقل من (0.05)، ولمعرفة لمن تُعزى هذه الفروق، تم استخراج نتائج اختبار أقل فرق دال[Least Square Differences (LSD)]، والموضّحة في الجدول (6).

الجدول (6)

نتائج اختبار أقل فرق دال (LSD) للمقارنات الثنائية بين المتوسطات الحسابية لدرجات الطلبة باختلاف طريقة التصحيح.

طريقة التصحيح		متوسط الفروق بين الطريقتين	الخطأ المعياري	مستوى الدلالة
طريقة معادلة التصحيح لأثر التخمين	طريقة تصحيح المموهات	**-1.123***	0.088	0.000
	طريقة البديل لا أعرف الاجابة	-0.499	0.253	0.050
طريقة تصحيح المموهات	طريقة البديل لا أعرف الاجابة	**0.624***	0.231	0.008

*** فرق متوسط دال عند مستوى دلالة (0.05).**

يتضح من خلال الجدول (6) وجود فروق ذات دلالة إحصائية عند مستوى دلالة ($\alpha \leq 0.05$) بين طريقة معادلة التصحيح لأثر التخمين وطريقة تصحيح المموهات ولصالح طريقة تصحيح المموهات ذات المتوسط الحسابي الأعلى، كما يلاحظ وجود فروق ذات دلالة احصائية بين طريقة تصحيح المموهات وطريقة استخدام البديل "لا أعرف الإجابة" ولصالح طريقة تصحيح المموهات، كما ويتضح عدم وجود فروق دالة إحصائياً في متوسطات أداء الطلبة على الاختبار بين طريقة معادلة التصحيح لأثر التخمين وطريقة البديل "لا أعرف الإجابة".

مناقشة نتائج السؤال الأول:

يتضح من خلال نتائج السؤال الأول أن طريقة تصحيح المموهات، كان لها أعلى مستوى أداء بدلالة المتوسط الحسابي، وكان متوسطها هو الأقرب إلى متوسط درجات الطلبة الأصلية دون ضبط أثر التخمين، وكانت لها فروق دالة احصائياً عن طريقة استخدام معادلة التصحيح لأثر التخمين وطريقة استخدام البديل لا أعرف الإجابة، ويشير ذلك إلى مدى فاعلية هذه الطريقة من حيث الأخذ بعين الاعتبار المعرفة الجزئية للطلبة من خلال جاذبية المموه (البديل) الخاطئ الذي اختاره الطالب، حيث تستند إلى مبدأ أن التخمين ليس دائماً عشوائياً، وإنما يلجأ الطالب باختيار البديل الأقرب للصواب في حال عدم معرفته للبديل الصحيح بناءً على أفكاره السابقة ومعرفته الجزئية، وأدى ذلك إلى ارتفاع متوسط أداء الطلبة على الاختبار، كما وأدت هذه الطريقة إلى تناقص الانحراف المعياري، وبالتالي حافظت على تجانس عينة الدراسة، كما في الدرجات الأصلية، ويتفق هذا مع دراسة دودين (Dodeen, 2005)، وهي الدراسة الوحيدة التي استهدفت هذه الطريقة، التي أشارت إلى زيادة المتوسط الحسابي، وتناقص الانحراف المعياري في هذه الطريقة، مقارنة مع طريقة تصحيح أثر التخمين التقليدية، كما واتفقت هذه النتيجة مع دراسة مجتجنز وآخرون (Muijtjens et al,1999)، التي أشارت إلى أن طريقة استخدام البديل "لا أعرف الإجابة" أظهرت متوسطاً حسابياً أعلى من طريقة تصحيح أثر التخمين التقليدية، كما واتفقت هذه النتيجة مع نتائج دراسة سواقد (1983) التي أشارت إلى أفضلية طريقة الاختيار الجزئي على طريقة التصحيح لأثر التخمين، كما ويتضح عدم وجود فروق دالة احصائياً بين طريقة التصحيح لأثر التخمين وطريقة البديل "لا أعرف الإجابة" وقد يعزى ذلك إلى أن الطلبة لا يميلون إلى اختيار هذا البديل (لا أعرف الإجابة) وبالتالي تقترب درجاتهم من الدرجات الناتجة من معادلة التصحيح لأثر التخمين.

وفيما يخص الارتباطات بين فروق الدرجات المعدّلة، والدرجة الكلية بدلالة عدد الإجابات الصحيحة كان الفرق أقل بين طريقتي تصحيح المموهات، وطريقة "لا أعرف الإجابة"، ممّا يدل على تفوّق هاتين الطريقتين على طريقة تصحيح أثر التخمين التقليدية، إلا أن طريقة تصحيح المموهات أظهرت فاعلية أكبر، إذ كانت القيم المعدلّة المتدنية تقابل درجات كلية كبيرة، في حين تقابل القيم المعدلّة الكبيرة درجات كلية متدنية، ويعد هذا مؤشراً على أن التخمين الأكثر كان للطلبة متدني التحصيل، بينما كان التخمين لدى الطلبة مرتفعي التحصيل قليلاً، وبالتالي مدى نجاح هذه الطريقة في المحافظة على أثر المعرفة الجزئية وإلغاء التخمين العشوائي.

النتائج المتعلقة بالسؤال الثاني:

هل يوجد أثر ذو دلالة احصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في صدق الاختبار يُعزى لطريقة ضبط أثر التخمين؟

للإجابة عن السؤال الثاني تمت مقارنة الطرق الثلاثة باستخدام عينة التطبيق من خلال إجرائين للصدق، هما: صدق المحك التلازمي باستخدام معامل الارتباط بين درجات الطلبة على الاختبار المعد في هذه الدراسة على وفق كل طريقة، ودرجات الطلبة على الاختبارين الشهريين الأول والثاني المدرسيين، وصدق البناء من خلال إجراء التحليل العاملي لدرجات الطلبة على وفق كل

طريقة، ويوضّح الجدول (7) نتائج معاملات الارتباط بين الدرجات الناتجة من طرق التصحيح الثلاث والدرجات على الاختبار المحك (صدق المحك التلازمي)، في حين يوضّح الجدول (9) نتائج التحليل العاملي لكل طريقة من الطرق الثلاث.

الجدول (7)

نتائج صدق المحك ومعاملات الارتباط بين درجات طرق التصحيح

الطريقة	تصحيح أثر التخمين	تصحيح المموهات	"لا أعرف الإجابة"	اختبار المحك
تصحيح أثر التخمين	1	0.90**	0.68**	**0.69****
تصحيح المموهات		1	0.67**	**0.79****
"لا أعرف الإجابة"			1	**0.73****
اختبار المحك				1

****معامل ارتباط دال إحصائياً عند مستوى دلالة (0.01)**.

ويلاحظ من خلال الجدول (7) أن معاملات صدق المحك جميعها كانت مرتفعة، ودالة عند مستوى الدلالة ($\alpha \leq 0.01$)، إلا أن طريقة تصحيح المموهات أظهرت أعلى مؤشر لصدق المحك، تليها طريقة "لا أعرف الإجابة"، وأظهرت طريقة تصحيح أثر التخمين أقل مؤشر لصدق المحك، وفيما يتعلق بمعاملات الارتباط بين درجات الطلبة باستخدام طرق التصحيح الثلاث فقد كانت جميعها مرتفعة ودالة احصائياً عند مستوى الدلالة ($\alpha \leq 0.01$)، وقد بلغ أعلى معامل ارتباط بين طريقتي تصحيح أثر التخمين وتصحيح المموهات.

ولمعرفة مدى دلالة الفروق بين معاملات صدق المحك للطرق الثلاث تمّ استخدام اختبار (T- Test) لاختبار الفروق بين معاملي ارتباط للبيانات غير المستقلة باستخدام المعادلة (3)، ويوضّح الجدول (8) نتائج اختبار (ت) لفحص الفروق بين معاملات صدق المحك للطرق الثلاث.

$$T = (r_{13} - r_{23}) \sqrt{\frac{(n-3)(1+r_{12})}{2(1-r_{13}^2-r_{23}^2-r_{12}^2+2r_{13}r_{23}r_{12})}} \dots\dots\dots [3]$$

حيث:

T: قيمة (ت) المحسوبة.

n: حجم العينة = 54.

r_{12}: معامل الارتباط بين درجات الطريقة الأولى مع الطريقة الثانية.

r_{13}: معامل الارتباط بين درجات الطريقة الأولى مع المحك.

r_{23}: معامل الارتباط بين درجات الطريقة الثانية مع المحك.

df: درجات الحرية = n-3 = 54−3 = 51.

الفرضية الصفرية (r_{13}=r_{23})

الجدول (8)

نتائج اختبار (T- Test) لاختبار الفروق بين معاملات صدق المحك

المقارنة	r_{12}	r_{13}	r_{23}	قيمة (ت) الجدولية	قيمة (ت) المحسوبة	النتيجة
التصحيح لأثر التخمين * تصحيح المموهات	0.90	0.69	0.79	1.671	2.61	رفض الفرضية الصفرية
التصحيح لأثر التخمين * البديل "لا أعرف الإجابة"	0.68	0.69	0.73	1.671	0.57	قبول الفرضية الصفرية
تصحيح المموهات * البديل "لا أعرف الإجابة"	0.67	0.79	0.73	1.671	0.96	قبول الفرضية الصفرية

ويتضح من خلال الجدول (8) وجود فروق ذات دلالة احصائية عند مستوى الدلالة ($\alpha \leq 0.05$) بين طريقتي التصحيح لأثر التخمين وتصحيح المموهات لصالح طريقة تصحيح المموهات، أي أن معامل صدق المحك لطريقة تصحيح المموهات أكبر منه لطريقة تصحيح أثر التخمين، في حين لم تظهر نتائج الاختبار فروق دالة احصائياً بين طريقتي تصحيح أثر التخمين والبديل لا أعرف الإجابة، وطريقتي تصحيح المموهات وطريقة لا أعرف الإجابة.

الجدول (9)

نتائج التحليل العاملي

العامل	طريقة تصحيح أثر التخمين			طريقة تصحيح المموهات			طريقة "لا أعرف الإجابة"		
	الجذر الكامن	نسبة التباين المفسر%	نسبة التباين المفسر التراكمي%	الجذر الكامن	نسبة التباين المفسر%	نسبة التباين المفسر التراكمي%	الجذر الكامن	نسبة التباين المفسر%	نسبة التباين المفسر التراكمي%
1	3.21	17.83	17.83	4.18	23.22	23.22	3.36	18.67	18.67
2	1.48	8.22	26.06	1.55	8.61	31.83	1.53	8.50	27.17
3	1.38	7.67	33.72	1.31	7.28	39.11	1.35	7.50	34.67
4	1.20	6.67	40.39	1.22	6.78	45.89	1.20	6.67	41.33
5	1.15	6.39	46.78	1.13	6.28	52.17	1.13	6.28	47.61
6	1.10	6.11	52.89	1.11	6.17	58.33	1.07	5.94	53.56
7	1.01	5.61	58.50	0.992	5.51	63.84	1.02	5.67	59.22

يتضح من خلال الجدول (8) أن الاختبار أحادي البعد بجميع الطرق، حيث كان الفارق بين الجذر الكامن للعاملين الأول والثاني يزيد عن الضعف، إذ فسّر العامل الأول مقداراً من التباين بلغ (17.83%، 23.22%، 18.67%) للطرق الثلاثة على التوالي، وأظهرت طريقة تصحيح أثر التخمين سبعة عوامل فسّرت (58.50%) من التباين الكلي، وكان الفارق بين العاملين الأول والثاني (1.73)، في حين أظهرت طريقة تصحيح المموهات ستة عوامل فسّرت (58.33%) من التباين الكلي وبلغ الفارق بين العاملين الأول والثاني (2.63)، أما طريقة "لا أعرف الإجابة"، فقد أظهرت سبعة عوامل فسّرت (59.22%) من التباين الكلي وبفارق مقداره (1.83) بين العاملين الأول والثاني، وبناءً على المحكّات التي أشار إليها ريكاس (Reckase, 1997) لقبول أحادية البُعد التي من أهمها: نسبة التباين المفسّر من العامل الأول أكبر من (20%)، نسبة الفرق بين الجذر الكامن للعامل الأول، والجذر الكامن للعامل الثاني أكبر من (2)، تكون طريقة تحليل المموهات هي فقط التي حققت المحكين (الشرطين) معاً.

مناقشة نتائج السؤال الثاني:

يتضح من خلال نتائج السؤال الثاني أن جميع الطرق كان لها معامل صدق محك تلازمي مرتفع، إلا أن طريقة تصحيح المموهات كانت الأفضل والأعلى صدقاً على وفق هذا المعامل، وفيما يخص نتائج صدق البناء من خلال التحليل العاملي كانت

جميع الطرق متقاربة، إلا أن طريقة تصحيح المموهات استطاعت تفسير مقدارٍ من التباين متقاربٍ مع الطريقتين الأخريين وبعدد أقل من العوامل، وبالتالي كان لهذه الطريقة صدق محكي وصدق بناء أفضل، وقد يعزى ذلك إلى أن هذه الطريقة كانت أقرب إلى قياس القدرة الحقيقية للطلبة؛ كونها تقوم بتصحيح التخمين العشوائي فقط، وتبقي على المعرفة الجزئية لدى الطلبة من خلال إجراءاتها الإحصائية، ومن ناحية أخرى جاءت هذه النتائج متفقة مع نتائج الصدق لدراسات كل من سواقد (1983)، الزبون (2004)، والنبهان(Alnabhan, 2002)، التي أشارت إلى تدني مؤشر الصدق في الطريقة التقليدية، وارتفاعه في طرق المعرفة الجزئية، وتعد طريقة تصحيح المموهات من الطرق المعتمدة على المعرفة الجزئية، أما الطريقة التقليدية، فإنها تلغي المعرفة الجزئية، وتعتبر جميع التخمين عشوائياً، أما طريقة "لا أعرف الإجابة" فهي تقلل من أثر التخمين، إلا أنها في نفس الوقت تحد من استخدام المعرفة الجزئية للطلبة.

النتائج المتعلقة بالسؤال الثالث:

هل يوجد أثر ذو دلالة احصائية عند مستوى الدلالة ($\alpha \leq 0.05$) في ثبات الاختبار يُعزى لطريقة ضبط أثر التخمين؟

للإجابة عن السؤال الثالث، تم مقارنة الطرق الثلاث باستخدام عينة التطبيق من خلال استخدام معامل ثبات كرونباخ الفا للاتساق الداخلي، حيث تم حساب قيم الفا للعينة الكلية ولثلاث عينات فرعية أخرى، هي: الأفراد الذين تقل درجاتهم عن (14)، الأفراد الذين تقل درجاتهم عن (11)، الأفراد الذين تقل درجاتهم عن (9)، وتم تحديد هذه المجموعات من خلال المئينات (90، 75، 50) على التوالي وبشكلٍ تقريبي، والمبينة في الجدول (2)، ويوضّح الجدول (10) نتائج معامل ثبات كرونباخ الفا لكل طريقة من الطرق الثلاث.

الجدول (10)

نتائج معاملات ثبات كرونباخ الفا لكل طريقة

الطريقة	العينة الكلية		أقل من 14		أقل من 11		أقل من 9	
	العدد	الثبات	العدد	الثبات	العدد	الثبات	العدد	الثبات
تصحيح أثر التخمين	159	0.79	132	0.64	85	0.40	45	0.16
طريقة تصحيح المموهات	159	0.81	130	0.76	83	0.68	42	0.59
طريقة "لا أعرف الإجابة"	159	0.80	132	0.64	85	0.48	45	0.22

ويتضح من خلال الجدول (10) أن قيم معاملات الثبات للعينة الكلية كانت مرتفعة، ومتقاربة لجميع الطرق، إلا أن طريقة تصحيح المموهات أظهرت معاملات ثبات أعلى للمجموعات الفرعية، تليها طريقة "لا أعرف الإجابة"، وأعطت طريقة تصحيح أثر التخمين أقل قيم لمعاملات ثبات المجموعات الفرعية، كما ويلاحظ أن قيم الثبات كانت تنخفض للمجموعات الفرعية بانخفاض عدد أفراد المجموعة في الطرق الثلاثة.

ولمعرفة فيما إذا كانت الفروق في قيم معاملات ثبات الاختبار تبعاً لطريقة التصحيح المستخدمة ذات دلالة إحصائية عند مستوى الدلالة ($\alpha \leq 0.05$) تم عمل تحليل التباين الثلاثي للقياسات المتكررة (Repeated Measurement Design)، ويوضح الجدول (11) نتائج تحليل التباين الثلاثي للقياسات المتكررة.

الجدول (11)

نتائج تحليل التباين الثلاثي للقياسات المتكررة لأثر الطريقة المستخدمة لضبط أثر التخمين في قيم معاملات ثبات الاختبار

اختبار الآثار:	مصدر التباين	مجموع المربعات	درجات الحرية	متوسط مجموع المربعات	ف	الدلالة الإحصائية
داخل الأفراد	الطريقة	0.103	2	0.052	5.421	**0.045***
	الخطأ	0.057	6	0.010		

يلاحظ من نتائج الجدول (11) وجود أثر ذو دلالة إحصائية قيم معاملات ثبات الاختبار باختلاف الطريقة المستخدمة للتصحيح من أثر التخمين، إذ كانت قيمة مستوى الدلالة الخاصة بـ(ف) تساوي (0.045) وهي أقل من مستوى الدلالة (0.05)،

ولمعرفة لمن تُعزى الفروق، تم استخراج نتائج اختبار أقل فرق دال [Least Square Differences (LSD)]، والموضّحة في الجدول (12).

الجدول (12)

نتائج اختبار أقل فرق دال (LSD) للمقارنات الثنائية بين قيم الثبات باختلاف طريقة التصحيح.

طريقة التصحيح		متوسط الفروق بين الطريقتين	الخطأ المعياري	مستوى الدلالة
طريقة معادلة التصحيح لأثر التخمين	طريقة تصحيح المموهات	**-0.212***	0.090	0.022
	طريقة البديل لا أعرف الاجابة	-0.038	0.019	0.353
طريقة تصحيح المموهات	طريقة البديل لا أعرف الاجابة	**0.175***	0.076	0.039

*** فرق متوسط دال عند مستوى دلالة (0.05).**

يتضح من خلال الجدول (12) وجود فروق ذات دلالة احصائية عند مستوى دلالة ($\alpha \leq 0.05$) بين طريقة معادلة التصحيح لأثر التخمين وطريقة تصحيح المموهات ولصالح طريقة تصحيح المموهات ذات معاملات الثبات الأعلى، كما يلاحظ وجود فروق ذات دلالة احصائية بين طريقة تصحيح المموهات وطريقة استخدام البديل "لا أعرف الإجابة" ولصالح طريقة تصحيح المموهات، كما ويتضح عدم وجود فروق دالة احصائياً في قيم معاملات ثبات الاختبار بين طريقة معادلة التصحيح لأثر التخمين وطريقة البديل "لا أعرف الإجابة".

مناقشة نتائج السؤال الثالث:

يتضح من نتائج السؤال الثالث أن طريقة تصحيح المموهات قد أظهرت أعلى مستوى لثبات الاتساق الداخلي على مستوى العينة الكاملة، والعينات الفرعية المشتقة منها، تليها طريقة البديل "لا أعرف الإجابة"، ثم طريقة تصحيح أثر التخمين التقليدية، كما وأشارت النتائج إلى وجود فروق دالة إحصائياً بين طريقة تصحيح المموهات وطريقتي التصحيح لأثر التخمين وطريقة البديل "لا أعرف الإجابة"، وتتفق هذه النتيجة مع دراسة دودين(Dodeen, 2005)، ودراسة مجتجنز وآخرون (Muijtjens et al,1999) التي أشارت إلى تفوق طريقة تصحيح المموهات، وطريقة البديل "لا أعرف الإجابة" على الطريقة التقليدية من حيث قيم الثبات، كما واتفقت هذه النتيجة مع دراسة سواقد (1983) ودراسة الزبون (2004)، ودراسة النبهان (Alnabhan, 2002)، ودراسة الخرشة (2004)، ودراسة راشد (2011)، التي أشارت جميعها إلى أن ثبات طرق الاختيار الجزئي أكبر من ثبات الطريقة التقليدية، ويلاحظ هنا زيادة قيم الثبات بين الطرق حسب مراعاة المعرفة الجزئية لدى الطلبة، كما أشارت النتائج إلى انخفاض مستويات الثبات بين العينات الفرعية ممّا يدل على أن الطلبة من مستويات القدرة المختلفة يستفيدون من التخمين بشكلٍ تفاضلي، وكانت طريقة تصحيح المموهات الأقدر من بين الطرق على إظهار هذا التفاضل بين مستويات القدرة دون التدني الكبير في مستويات الثبات تبعاً لحجم العينة.

الاستنتاجات:

يمكن التوصّل من خلال نتائج هذه الدراسة إلى الاستنتاجات الآتية:

- تعدّ طريقة تصحيح المموهات هي الأفضل بين الطرق الثلاثة من حيث الأداء، والصدق، والثبات، إذ تحافظ هذه الطريقة على المعرفة الجزئية لدى الطلبة، وتلغي التخمين العشوائي فقط.
- طريقة استخدام البديل "لا أعرف الإجابة" تعطي مؤشرات جيدة للصدق والثبات، وتلغي التخمين العشوائي، إلا أنها تحد من استخدام المعرفة الجزئية، والمعلومات السابقة لدى الطلبة، وقد تؤدي إلى خطأ قياس ناتج عن عامل المخاطرة في اختيار بديلٍ آخر عن البديل "لا أعرف الإجابة" حتى عند عدم امتلاك المعرفة التامة للإجابة الصحيحة.
- طريقة تصحيح أثر التخمين التقليدية أظهرت اقل مؤشرات للصدق والثبات، وتعمل هذه الطريقة على إلغاء المعرفة الجزئية لدى الطلبة، إذ تعتبر جميع التخمين عشوائياً.

التوصيات:

بناءً على النتائج التي تمّ التوصل إليها من خلال هذه الدراسة يوصي الباحثان بما يأتي:

- استخدام طريقة تصحيح المموهات في تصحيح اختبارات الاختيار من متعدد لما لها من خصائص جيدة.

- الاستفادة من نتائج هذه الدراسة قبل الجهات المعنية، والمختصة مثل وزارة التربية والتعليم والجامعات.
- إجراء المزيد من الدراسات تستهدف طرق أخرى لضبط اثر التخمين لا سيما في ضوء نظرية الاستجابة للفقرة.

المراجع

المراجع العربية:

الخرشا، ط. (2004)، مقارنة ثلاثة طرق للتصحيح ضد أثر التخمين في النظرية الكلاسيكية. أطروحة ماجستير غير منشورة، جامعة مؤته، الكرك، الأردن.

ذيابات، ض. (2008)، دراسة مقارنة بين طريقتي الحذف الجزئي والاختيار الجزئي لضبط أثر التخمين في اختبارات الاختيار من متعدد، أطروحة ماجستير غير منشورة، جامعة اليرموك، اربد، الأردن.

راشد، أ. (2011)، أثر تعليمات التخمين في استجابة الطلبة لفقرات الاختيار من متعدد على درجاتهم وعلى الخصائص السيكومترية للاختبار وفقراته، أطروحة ماجستير غير منشورة، جامعة عمان العربية، عمّان، الأردن.

الزبون، ح. (2004)، أثر استخدام ثلاث طرق للتصحيح لضبط أثر التخمين في الخصائص السيكومترية لاختبارات الاختيار من متعدد في ضوء نظرية الاستجابة للفقرة، أطروحة ماجستير غير منشورة، جامعة مؤته، الكرك، الأردن.

سواقد، س. (1983)، المقارنة بين أثر ثلاث طرق لتصحيح اختبارات الاختيار من متعدد على الخصائص السيكومترية للاختبار وعلى مستويات تحصيل ودرجات مخاطرة مختلفة على الاختبار، أطروحة ماجستير غير منشورة، جامعة اليرموك، اربد، الأردن.

المراجع الاجنبية:

Alnabhan, Mousa. (2002). An Empirical Investigation of the Effect of three Methods of Handling Guessing and Risk Taking on the Psychometric Properties of a Test. Social Behavior and Personality, 30, 645-652.

Annie, W.Y and Chan, Alan H.S. (2012). Evaluation of the Three Multiple-Choice Assessment Methods in a human Factors Engineering Course. Journal of the Chinese Institute of Industrial Engineers, 29, (7): 466-476.

Bar-Hillel, Maya. Budescu, David and Attali, Yigal. (2005). Scoring and keying Multiple Choice Tests: A case Study in Irrationality. Mind & Society, 3-12.

Burton, Richard F.(2002). Misinformation, Partial Knowledge and Guessing in true/ False Tests. Medical education, 36: 805-811.

Camilli, G. (2006).Test Fairness. In R. L. Bernnan. (Ed), Educational Measurement. 4th .Westport, CT: Praeger.

Dodeen, Hamzeh M. (2005). Using Distractors in Correcting for Guessing in Multiple- Choice Tests. Dirasat, Educational Sciences. 32, (1): 192-198.

Espinosa, Maria Paz. and Gardeazabal, Javier. (2010). Optimal Correction for Guessing in Multiple-Choice Tests. Retrieved Oct. 26.2013. From http://www.researchgate.net.

Fray, Robert B. (1988). Formula scoring of multiple-Choice tests (Correction for guessing). Educational Measurement: Issues and Practice. Summer 1988.

Haladyna, Thomas M.; Downing, Steven M. (2004). Construct-Irrelevant Variance in High-Stakes Testing. Educational Measurement: Issues and Practice.

McMillan, James H. (2013). Research on Classroom Assessment. Sage Publication, USA.

Mehrens, William A. and Lehmann, Irvin J. (1991). Measurement and Evaluation in Education and Psychology. 4th. Thomson Learning, USA.

Merrit, Stephen. (2006). Mastering Multiple Choice: The Definitive Guide to Better Grades on Multiple Choice Exams. Brain Ranch, Canada.

Messick، S. (1995).Validation of Psychological Assessment: Validation of Inferences from person's response and performances as scientific inquiry into scoring meanings. American Psychologist, 50, 741-749.

Muijtjens, A. M M. Mameeren, H. Van, Hoogenboom, R. J I. and Evers J L. (1999). The Effect of a "don't know" Option on Test Scores: Number- Right and Formula Scoring Compared. Medical Education, 33, 267-275.

Poham, W. James. (2003). Educational Assessment, Pearson Education. Inc, USA.

Reckase, M.D. (1997). The past and future of multidimensional item response theory. Applied Psychological Measurement, 1, 25-36.

Reif, Manuel. Holocher-Ertl, Stefana. Klaus D. Kubinger, and Frebort, Martina. (2010). On Minimizing Guessing Effects on Multiple-Choice Items: Superiority of two solutions and three Distractors item format to a one Solution and Five Distractors Item Format. International Journal of Selection and Assessment. 18, (1): 110-115.

Sabbe Elien. Valcke, Martin. and Lesage Ellen. (2013). Scoring methods for multiple choice assessment in higher education is it still matter of number right scoring or negative marking?. Studies in Education Evaluation. 39, 188-193.

Schneider, M. Christian. Eagn, Karla L. and Julian, Marc W. (2013).Classroom Assessment in The Context of High-Stakes Testing. In McMillan. James H. (Ed), Research on Classroom Assessment. Sage Publication.

Wade, L. Harrison, C. Hollands, J. Mattick, K, Ricketts, C. and Wass, L. (2011). Student perception of the progress test in two setting and the implications for test deployment. Adv. Health Sci Educ, Springer.

Comparison of Methods to Control the Effect of Guessing in Multiple-Choice Tests and their Effects on the Characteristics of the Test

Sharif Al-Saoudi, Mo'en Salman Alnasraween *

ABSTRACT

This study aimed at comparing three methods to control the effect of guessing in multiple-choice tests: scoring formula, distraction-scoring formula the "don't know" option, through investigating their effects on the characteristics of the test, using math test achievement. The sample of the study consistsed of (159) male and female students. The results showed statistically significant differences at the level of significance ($\alpha \leq 0.05$) between the method of distraction-scoring formula and the method of scoring formula in favor of the method of distraction-scoring formula, in accordance with the following criteria: performance on the test, validity and reliability, due to its reliance on partial knowledge of the students. The results did not show statistaclly significant differences between the two methods: "don't know" option method and scoring formula method.

Keywords: Guessing, Multiple-choice tests, Test scoring.

* Ministry of Education, Jordan(1); Faculty of Educational and Psychological Sciences (2), Amman Arab University. Received on 04/03/2016 and Accepted for Publication on 29/07/2016.