

2016

The Effect of Processing Guessing Method on the Estimating Item/Person Statistics for Multiple-Choice Test According to Item Response Theory (IRT)

Taha Alkursheh

Tabouk university, taha_oglah@yahoo.com

Follow this and additional works at: https://digitalcommons.aaru.edu.jo/anu jr_b

Recommended Citation

Alkursheh, Taha (2016) "The Effect of Processing Guessing Method on the Estimating Item/Person Statistics for Multiple-Choice Test According to Item Response Theory (IRT)," *An-Najah University Journal for Research - B (Humanities)*: Vol. 30 : Iss. 12 , Article 2.

Available at: https://digitalcommons.aaru.edu.jo/anu jr_b/vol30/iss12/2

This Article is brought to you for free and open access by Arab Journals Platform. It has been accepted for inclusion in An-Najah University Journal for Research - B (Humanities) by an authorized editor. The journal is hosted on [Digital Commons](#), an Elsevier platform. For more information, please contact rakan@aar u.edu.jo, marah@aar u.edu.jo, dr_ahmad@aar u.edu.jo.

أثر طرق معالجة أثر التخمين على تقدير إحصائيات الأفراد والفقرات في اختبارات الاختيار من متعدد وفق النظرية الحديثة في القياس

The Effect of Processing Guessing Method on the Estimating Item/Person Statistics for Multiple-Choice Test According to Item Response Theory (IRT)

طه الخرشه

Taha Alkursheh

قسم علم النفس، كلية التربية والآداب، جامعة تبوك، السعودية

بريد الالكتروني: taha_oglah@yahoo.com

تاريخ التسليم: (2015/12/21)، تاريخ القبول: (2016/7/31)

ملخص

هدفت الدراسة إلى التحقق من أثر طرق معالجة أثر التخمين في تقدير إحصائيات الأفراد والفقرات في اختبارات الاختيار من متعدد. ولتحقيق الغرض من الدراسة تم بناء اختبار تحصيلي من نوع الاختيار من متعدد في مادة مبادئ القياس والتقويم التي تدرس لطلاب كلية التربية بجامعة مؤتة مكون من 30 فقرة بصورته النهائية، وقد اشتمل على ثلاثة نماذج حسب طرق المعالجة. وباستخدام برنامج (Bilog – Mg3) تم تحليل استجابات المفحوصين البالغ عددهم (200) طالبا وطالبة لجميع نماذج الاختبار الثلاث وفق النموذج اللوجستي ثنائي المعلمة. كشفت نتائج تحليل التباين الاحادي بأنه يوجد فروق ذات دلالة إحصائية بين متوسطات معالم الصعوبة للفقرات تعزى لطريقة معالجة. وقد جاءت تلك الفروق لصالح طريقة المعالجة الأولى. وكذلك أظهرت النتائج أيضا بأنه لا يوجد فروق ذات دلالة إحصائية بين متوسطات معالم التمييز للفقرات تعزى لمتغير الدراسة. وبينت نتائج التحليل كذلك بأنه لا يوجد فروق دالة إحصائية بين متوسطات معلمة القدرة للأفراد تعزى لمتغير الدراسة.

الكلمات المفتاحية: التخمين، نظرية استجابة الفقرة.

Abstract

This study aimed at verifying the effect of the guessing on the items parameters, person's ability and information function. To achieve this aim, a 30 item multiple – choice achievement test. The test had three different forms according Processing Guessing Effect. The responses of

200 examinees on the three forms were analyzed by Bilog–Mg3 programs according to the three parameter logistic model. The results of two-way ANOVA revealed significant statistical differences between the means of item difficulty parameters. No Differences were found, however, between the means of item discrimination and the means person's ability parameter.

Keywords: Guessing, Item Response Theory.

المقدمة

تتوقف كثير من القرارات التربوية والتعليمية في المراحل المختلفة على نتائج الاختبارات التحصيلية مثل: اجتياز الطالب مقرر ما أو رسوبه، أو انتقاله من صف إلى الصف الذي يليه، أو من مرحلة تعليمية إلى التي تليها. كما أن نجاح الطالب أو رسوبه أو حصوله على تقدير مرتفع أو منخفض في الجامعة يتوقف بدرجة كبيرة على نتائجه في هذه الاختبارات التحصيلية، وعملية تقويم الأفراد تكتسب أهمية كبيرة بقدر أهمية القرارات المبنيّة عليها، وبقدر خطورة القرارات الخاطئة التي يمكن أن تترتب على ذلك في المواقف والمجالات المتعددة على مستوى الفرد والمجتمع إلى حد قد يصعب معالجته، أو قد يحتاج علاجه إلى وقت طويل مما يعيق عملية التنمية ومواكبة تطور المجتمعات الأخرى، وللحصول على قرارات أكثر دقة يجب توفر معلومات صادقة ودقيقة من خلال التخطيط والإعداد الجيد للاختبار، ولذلك كان هناك اهتمام بالغ من قبل المتخصصين في القياس النفسي والتربوي تجاه هذه الاختبارات، فوضعوا شروطاً للاختبار التحصيلي الجيد، واهتموا بكيفية بنائه، والتأكد من صلاحيته لقياس التحصيل الدراسي وتحقق الشروط السيكومترية من صدق وثبات وموضوعية، وعكفوا على العوامل المختلفة التي قد تؤثر في قدرته على تحديد قدرات الطلاب فيما يراد قياسه.

وهناك أشكال عدة لهذه الاختبارات منها الاختبارات ذات الإجابة المختارة والاختبارات ذات الإجابة المنتجة. ولفقرات الإجابة المختارة أشكال متعددة تستخدم لقياس تحصيل الطلبة، ولكن فقرات الاختيار من متعدد تتمتع بانتشار واسع، وتستخدم لأغراض متنوعة، كما أشارت نتائج الأبحاث التي قارنت فقرات الاختيار من متعدد مع غيرها من أنواع الفقرات مثل فقرات الصواب والخطأ وفقرات التكميل أن فقرات الاختيار من متعدد تستطيع قياس التحصيل بشكل أكثر صدقاً وثباتاً وأكثر كفاءة وفاعلية ومما لاشك فيه أن فقرات الاختيار من متعدد من أكثر أشكال الأسئلة التي لها استخدام واسع في الاختبارات التحصيلية واختبارات الاستعداد (الشريفيين وطعمة، 2009). غير أن هذه الاختبارات، لا تخلو من أخطاء القياس، والتي تؤثر بشكل كبير على عملية التقييم والقرارات المترتبة عليها والتي تكون أحياناً غير صحيحة نتيجة هذه الأخطاء (Ban-Simon, 1997).

وقد يتساءل البعض عن مصادر هذه الأخطاء من أجل التحكم فيها أثناء إعداد الاختبار وعند تطبيقه وتصحيح فقراته، وذلك لجعل تأثيرها في العلامات الحقيقية أقل ما يمكن، وبذلك تقترب العلامات الملاحظة للأفراد في الاختبار من علاماتهم الحقيقية التي تعبر عن مدى امتلاكهم للسمة موضوع القياس، وبالتالي تزداد دقة القرارات المترتبة على ذلك (علام، 2002).

ومن أهم هذه الأخطاء التخمين العشوائي Random Guessing أو ما يطلق عليه التخمين الأعمى Blind Guessing (Andra & Magnano, 2011) الذي يشير إلى السلوك الذي يقوم به المفحوص عندما لا تكون لديه الإجابة الصحيحة؛ أي عندما لا تكون لديه المعرفة التامة والأكيدة بالإجابة الصحيحة للفقرة. ففي مثل هذه الحالة؛ إما أن تكون لدى المفحوص معرفة جزئية (غير تامة أو أكيدة) Partial Information تجعله قادراً على استبعاد بعض البدائل وحصص الإجابة الصحيحة في البعض المتبقي من البدائل الموضوع للفقرة، فيختار أحدها كإجابة للفقرة، وقد لا تكون لديه مثل هذه المعرفة الجزئية فيختار أحد البدائل عشوائياً كإجابة للفقرة (الخرشه، 2004).

وتعد مشكلة التخمين في اختبارات الاختيار من متعدد من بين المشكلات التي اهتم خبراء القياس النفسي والتربوي بدراساتها وبحثها للتعرف على مدى تأثيرها سواء في أداء المفحوص أو في مختلف الخصائص السيكومترية لهذه الاختبارات. ويعود ذلك إلى كون التخمين يمثل أحد مصادر الخطأ في القياس، وعليه فقد تم اقتراح عدداً من الطرق للتغلب على هذه المشكلة والتقليل من تأثيرها، وأكثر هذه الطرق شيوعاً، طريقة عدد الإجابات الصحيحة - Number right scoring وطريقة التصحيح من أثر التخمين (العقاب) وطريقة التصحيح من أثر التخمين (المكافأة) وطريقة التصحيح من أثر التخمين (العقاب والمكافأة) (Abu-Sayf, 1977).

وجميع هذه الطرق تقوم على مجموعة من الافتراضات فأولها: أن المفحوص يعرف الإجابة الصحيحة، وثانيها: أن المفحوص لا يعرف الإجابة الصحيحة ويترك الفقرة دون إجابة، وثالثها: أن المفحوص لا يعرف الإجابة على الفقرة، ولكنه يجيب على الفقرة بشكل عشوائي (Ben-Shakhar & Sinai, 1991).

وقد لوحظ أن بعض المستخدمين لاختبارات الاختيار من متعدد يستخدمون طريقة التصحيح التقليدية، ويستخرجون نتائج يبنون عليها قرارات وتوصيات متعددة، رغم أن جميع الدراسات في هذا الموضوع أثبتت عدم كفاية هذه الطريقة لتصحيح اختبارات الاختيار من متعدد من حيث قدرتها على ضبط عامل التخمين.

وقارنت دراسة النبهان (Alnabhan, 2002) تناول بين ثلاث طرق لتصحيح اختبارات الاختيار من متعدد على الخصائص السيكومترية للاختبار حيث تمت المقارنة بين طريقة عدد الإجابات الصحيحة وطريقة استخدام معادلة التصحيح وطريقة الاختيار الجزئي، ولتحقيق هدف الدراسة تم إعداد اختبار تحصيلي من نوع الاختيار من متعدد لقياس المصطلحات والمفاهيم في مادة مناهج البحث في علم النفس، وتم تطبيق الاختبار على أفراد العينة البالغ عددهم 120 طالباً

من الطلبة المسجلين في مبحث "مناهج البحث في علم النفس" للفصل الدراسي الثاني من العام الدراسي 2000\2001، أظهرت نتائج هذه الدراسة إلى ان تأثير معادلة التصحيح وطريقة الاختيار الجزئي على معامل ثبات الاختبار تأثير متساوي، وتعطيان معامل ثبات أعلى من الطريقة الأولى وهي طريقة عدد الإجابات الصحيحة.

وفي دراسة لـ (Muijtjens et al, 1999) أجرى فيها مقارنة بين طريقة عدد الإجابات الصحيحة (Number – Right) وطريقة استخدام معادلة التصحيح لأثر التخمين (العقاب) (Scoring Formula) وبينت نتائج هذه الدراسة إلى أن معامل ثبات الاختبار يكون أعلى عند تصحيح الاختبار باستخدام معادلة التصحيح لأثر التخمين (العقاب) مقارنة بطريقة عدد الإجابات الصحيحة.

كما أجرى (jaradat & Tollefson, 1988) دراسة هدفت إلى التعرف على أثر طرق تصحيح اختبارات الاختيار من متعدد على الخصائص السيكومترية للاختبار. حيث تم المقارنة بين ثلاث طرق للتصحيح هي معادلة التصحيح من أثر التخمين (العقاب) وطريقة الاختيار الجزئي من بين مجموعة البدائل الموضوعية للسؤال وطريقة استبعاد البدائل غير الصحيحة. وتكونت عينة الدراسة من 54 طالبا والذين يدرسون مادة القياس النفسي، وأظهرت النتائج فيما يخص معامل الثبات تفوق طريقة الاختيار الجزئي.

إن المتصفح لنتائج الدراسات السابقة يجد التباين في نتائجها حول تأثير طرق التصحيح الاختبار من اختيار متعدد في ثبات الاختبار ولم تتطرق الدراسات إلى خصائص الفقرات وكذلك لم تتطرق لطرقتي التصحيح المكافأة وطريقة المكافأة والعقاب في معالجة أثر التخمين، وتجدر الإشارة كذلك بان غالبية هذه الدراسات اعتمدت على النظرية التقليدية في القياس في تحليل النتائج.

والنظرية التقليدية تنطلق من فرض أساسي في بناء الاختبارات والمقاييس وتحليل فقراتها من كون توزيع درجات الأفراد في السمة أو الخصيصة التي يقيسها الاختبار تتوزع توزيعاً اعتدالياً (Normal Curve)، والذي يتأثر بطبيعة خصائص عينة الأفراد وخصائص عينة الفقرات. وتستند هذه النظرية على مبدأ الفروق الفردية، لذا تقارن درجة الفرد فيها بدرجات المجموعات التي ينتمي إليها (Crocker & Algian, 1986).

لذلك تتأثر الخصائص السيكومترية للاختبار بطبيعة خصائص عينة الأفراد التي يطبق عليها الاختبار وبخصائص عينة الفقرات التي يتألف منها الاختبار (Brown, 1983). وترى النظرية الكلاسيكية أن درجة الفرد في الاختبار، هي دالة خطية مطردة، بمعنى كلما زادت درجة الفرد على الاختبار زاد مقدار وجود السمة لديه (Cronbach, Gleser, 1970).

وقارنت دراسة هاميلتون وجونز (Hambleton & Jonse, 1993) بين النظريتين التقليدية ونظرية الاستجابة للفقرة، وبينت جوانب القصور في النظرية التقليدية، وكيف جاءت

نظرية الاستجابة للفقرة لتعالج هذه الجوانب، وأشارت إلى التفوق الذي تظهره نظرية الاستجابة للفقرة في حل المشكلات التي تعجز عن حلها النظرية التقليدية.

وتقوم نظرية الاستجابة للفقرة (Item Response Theory) على مجموعة من الافتراضات الأساسية وهي: افتراض أحادية البعد (Unidimensionality) وافتراض الاستقلال الموضوعي Local Independence والعلاقة الوتيرية بين القدرة والأداء على الفقرة، (Item Characteristic Curve) وافتراض التحرر من السرعة في الأداء (Speededness). وقد انبثق عنها مجموعة من النماذج منها

1. النموذج اللوجستي أحادي المعلمة (نموذج راش) Rasch Model

الذي استخدم في تحليل استجابات أفراد عينة الدراسة. يعد هذا النموذج أبسط نماذج استجابة الفقرة ثنائية التدرج، ويفترض هذا النموذج أن جميع الفقرات لها القدرة التمييزية نفسها للتمييز بين المفحوصين، أي تكون جميع المنحنيات المميزة لفقرات الاختبار متوازية (متساوية في الميل)، وتختلف عن بعضها البعض في نقاط التقائها بالمحور الأفقي، الذي يمثل متصل القدرة أو السمة الكامنة، كما يفترض عدم لجوء المفحوصين للتخمين العشوائي عند اجابته عن فقرات الاختبار، وهذا يعني إن الفقرات تختلف فقط في صعوبتها، وتكون معادلة هذا النموذج كما يأتي:

$$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1 + e^{D(\theta-b_i)}}, (i = 1 \dots \dots \dots n)$$

حيث $P_i(\theta)$ احتمال إجابة الفقرة (i) إجابة صحيحة من قبل مفحوص له القدرة θ تم اختياره عشوائياً من بين المفحوصين الذين لهم القدرة نفسها.

D: ثابت مقداره 1.7، (Sampling Factor) وهو عامل تدرج يجعل القدرة المستنتجة من استعمال هذا النموذج مساوية للقدرة المستنتجة من استعمال النموذج، الذي يعتمد على المنحنى الطبيعي التراكمي.

b_i : معلمة الصعوبة للفقرة i

e: الأساس اللوغاريتمي الطبيعي (2.718)

2. النموذج اللوغاريتمي ثنائي البارامتر (Lord Model)

كان لورد 1952 أول من طور النموذج ثنائي البارامتر معتمداً على التوزيع الاعتدالي التراكمي. ثم قام بيرنيوم 1968 باستبدال دالة اللوغاريتم التراجيحي لمعلمين بدلا من دالة التوزيع الاعتدالي كصيغة للدالة المميزة للمفردة. وتتميز دالة اللوغاريتم التراجيحي بأنها أكثر إقناعاً للملاءمة مع الدالة الاعتدالية. ويعتبر النموذج اللوغاريتمي سهل التطويع رياضياً أكثر من

النموذج الاعتدالي لأنه يحتوي بداخله بعض صيغ التكامل، بينما يعتبر النموذج الاعتدالي دالة صريحة لمعلمي المفردة والقدرة ولديه خصائص إحصائية هامة .

الصورة الرياضية للنموذج اللوغاريتمي ثنائي البارامتر

تمثل المعادلة التالية المنحنى المميز للمفردة للنموذج اللوغاريتمي ثنائي المعلم والذي طوره بيرنباوم

$$p_i(\theta) = \frac{e^{Dai(\theta - b_i)}}{1 + e^{Dai(\theta - b_i)}}$$

حيث أن:

(i) $p_i(\theta)$: احتمال إجابة المفحوص الذي اختير عشوائياً من مستوى القدرة (θ) على المفردة (i) إجابة صحيحة، b_i معلم الصعوبة، ai : معلم التمييز، (θ) معلمة القدرة، (D) عامل التدرج ويساوي (1.7).

من الواضح أن النموذج اللوغاريتمي ثنائي البارامتر يشبه النموذج اللوغاريتمي أحادي البارامتر فيما عدا وجود عنصرين جديدين هما:

– العنصر الأول: عامل D وهو عامل القياس أو التدرج وتم إدخاله للمعادلة لكي يجعل الدالة اللوغاريتمية أقرب ما يكون للدالة الاعتدالية. وقد وجد انه حينما تكون ($D = 1.7$) فإن قيم $p_i(\theta)$ للنموذج اللوغاريتمي ثنائي البارامتر والنموذج الاعتدالي ثنائي البارامتر تختلف بقيمة مطلقة أقل من 0.01 لجميع قيم θ .

– العنصر الثاني الإضافي للنموذج اللوغاريتمي ثنائي البارامتر هو البارامتر ai والذي يسمى معلم تمييز المفردة. ويتناسب معلم التمييز ai مع ميل منحنى خاصية المفردة عند النقطة b_i على مقياس القدرة. فالمفردات التي تتميز منحنياتها بميل شديد الانحدار تكون أكثر فائدة في تصنيف المفحوصين إلى مستويات مختلفة القدرة مقارنة بالمفردات الأقل انحداراً في ميل منحنياتها، وفي الحقيقة فإن فائدة المفردة تكمن في قدرتها على التمييز بين المفحوصين القريبين من القدرة θ وهذا يتناسب مع ميل منحنى خاصية المفردة عند نقطة θ .

3. النموذج اللوغاريتمي ثلاثي البارامتر (Birnbaum Model)

في هذا النموذج أضاف بيرنباوم معلم ثالث هو التخمين، لأن بعض مفردات الاختبار تسمح أحياناً لبعض المفحوصين من ذوي القدرة المنخفضة جداً، بالتوصل إلى الإجابة الصحيحة عن طريق التخمين، والقيمة التقديرية لهذا المعلم، تمثل احتمال توصل هؤلاء المفحوصين إلى الإجابة الصحيحة عن المفردة، مع وجوب أخذ ذلك بعين الاعتبار، عند مطابقة البيانات المستمدة

من الاختبار للنموذج. وهنا تكون المنحنيات المميزة للمفردات متقاطعة؛ أي تختلف عن بعضها البعض في الجزء الذي تلتقي فيه بالمحور الأفقي الممثل لمتصل السمة الكامنة.

الصورة الرياضية للنموذج اللوغارتمي ثلاثي البارامتر

التعبير الرياضي للنموذج اللوغارتمي ثلاثي البارامتر كما يلي:

$$P\left(X = \frac{1}{\theta_S}, \beta_i, \alpha_i, \gamma_i\right) = \gamma_i + (1 - \gamma_i) \frac{e^{[\alpha_i(\theta_S - \beta_i)]}}{1 + e^{[\alpha_i(\theta_S - \beta_i)]}}, X = 0, 1$$

وقد تكتب بصيغة أخرى هي :

$$P_i(\theta) = C_i + (-C_i) \frac{e^{Dai(\theta - \beta_i)}}{1 + e^{Dai(\theta - \beta_i)}}$$

حيث :

$P_i(\theta)$: احتمال اجابة المفحوص ذو القدرة θ اجابة صحيحة على المفردة i .

b_i : معلم صعوبة المفردة.

a_i : معلم تمييز المفردة.

C_i : معلم تخمين المفردة الذي يمثل المقطع الرأسي (الصادي) لخط التقارب الأدنى لمنحنى خاصة المفردة، كما ويعبر عن احتمالية ان يجيب المفحوص ذو القدرة المنخفضة اجابة صحيحة عن المفردة بالتخمين، وان هناك علاقة طردية بين احتمالية ان يجيب المفحوص اجابة صحيحة عن المفردة، وبين قيمة مؤشر التخمين، حيث تزداد هذه الاحتمالية بزيادة معامل التخمين.

وتأتي أهمية هذه الدراسة في أنها تهدف إلى معرفة أثر طرق معالجة التخمين في تقدير إحصائيات الأفراد والفقرات في اختبارات الاختيار من متعدد وفق النظرية الحديثة في القياس، حيث تم تحليل البيانات المستمدة من صور الاختبار بطريقة أكثر موضوعية وهي استخدام احد نماذج نظرية الاستجابة للفقرة وهو النموذج اللوجستي ثنائي المعلمة two – Parameter Logistic Model). الذي يكسب الدراسة أهمية أخرى لما يتميز به من إمكانية تقدير صعوبة وتميز الفقرات بطريقة مستقلة عن قدرات الأفراد، وكذلك تقدير قدرات الأفراد بطريقة مستقلة عن صعوبة الفقرات.

مشكلة الدراسة وأهميتها

من أبرز المشاكل التي يعاني منها القياس النفسي والتربوي هي وجود خطأ القياس ويشكل التخمين أحد مصادر الخطأ الأساسية في القياس. حيث أشار (Lau & Usop, 2011) إلى أن التخمين يؤثر في الدرجة التي يستحقها الطالب في الاختبارات التحصيلية، مما يؤثر في صدق

هذه الاختبارات. كما يؤثر في بناء الاختبار حيث يتوقف قبول أو رفض المفردات على معاملي الصعوبة والتمييز اللذين يتأثران بدورهما بعامل التخمين. كما يؤدي التخمين إلى تضخيم قيمة التباين الناتج عن الخطأ العشوائي Random error variance وحدث نوع من التحيز (Camilli, 2006: 19). ويؤثر التخمين أيضا في ثبات الاختبار من ناحيتين؛ الأولى: أن ضيق الوقت في نهاية اختبارات السرعة قد يدفع الطالب إلى التخمين العشوائي للإجابة عن الفقرات المتبقية، وهذا يؤدي إلى عدم اتساق إجابته عن هذه الفقرات مع الإجابة عن الفقرات التي استغرق فيها وقتا كافيا في بداية الاختبار (Attali, 2005)، أما الثانية: أنه عند تطبيق الاختبار مرة أخرى فإنه من المحتمل ألا يحصل الطالب الذي يلجأ للتخمين العشوائي على نفس الدرجة لأنه قد يغير إجابته عن المفردات التي قام بتخمينها على غير أساس (عبدالوهاب، 2013). إذ أن المفحوص قد يحصل باستخدام التخمين على علامة أكبر من العلامة التي تمثل قدرته الحقيقية على السمة موضوع القياس. والسؤال الذي يطرح: هل توجد طريقة تصحيح أكثر دقة للفقرات التي يؤثر فيها التخمين بحيث تمثل العلامة على الاختبار تقديراً دقيقاً لقدرة المفحوص في السمة المقاسة. وهل طريقة التصحيح تؤدي إلى تحسين الخصائص السيكومترية للاختبار؟

وللإجابة على هذه الأسئلة استخدم العاملون في الميدان طرقا ومعالجات إحصائية اعتمدت جميعها على طريقة تصحيح الاختبار. من بينها توظيف معادلات خاصة للتصحيح من أثر التخمين، إلا أن دراسات عديدة شككت بفعالية استخدام معادلة عقاب التخمين في تحسين ثبات الاختبار. ولما كانت الدراسات التي توفرت للباحث لا تشير إلى وجود دراسات تجريبية لاختبار طريقة المكافأة على عدم التخمين أو طريقة المكافأة والعقاب، أتت هذه الدراسة لتقارن بين ثلاث طرق لمعالجة أثر التخمين على تقدير إحصائيات الأفراد والفقرات في اختبارات الاختبار من متعدد هي:

1. استخدام معادلة التصحيح لأثر التخمين (العقاب) والتعليمات المنسجمة معها.
2. استخدام معادلة التصحيح لأثر التخمين (المكافأة) والتعليمات المنسجمة معها.
3. معادلة التصحيح من أثر التخمين (المكافأة والعقاب) والتعليمات المنسجمة معها.

هدف الدراسة وأسئلتها

تهدف الدراسة إلى بحث أثر طرق معالجة أثر التخمين في تقدير إحصائيات الأفراد والفقرات باستخدام نموذج ثنائي المعلم وتحديد انسب طريقة لتصحيح اختبارات الاختبار من متعدد لمعايرة فقرات الاختبار. وبالتحديد ستم الإجابة عن الأسئلة التالية:

1. هل تختلف دقة تقدير معلمة صعوبة الفقرة باختلاف طريقة المعالجة لأثر التخمين؟
2. هل تختلف دقة تقدير معلمة تميز الفقرة باختلاف طريقة المعالجة لأثر التخمين؟
3. هل تختلف دقة تقدير معلمة القدرة للفرد باختلاف طريقة المعالجة لأثر التخمين؟

تعريف المصطلحات

طريقة معالجة أثر التخمين: تحديد القاعدة التي تعطى على أساسها قيم عددية لفقرات الاختبار أي أنها طريقة لتحديد مقياس معين للمعلومات التي نحصل عليها من إجراءات القياس، يعكس مستوى الأداء عند المفحوصين على الاختبار. والطرق الثلاث المستخدمة في الدراسة هي:

1. الطريقة الأولى: طريقة تصحيح الاختبار من نوع الاختيار من متعدد باستخدام معادلة التصحيح من اثر التخمين (العقاب): تكون علامة المفحوص على الاختبار المصحح بهذه

$$\text{الطريقة تساوي: عدد الإجابات الصحيحة} - \frac{\text{عدد الإجابات الخاطئة}}{\text{عدد البدائل} - 1}$$

2. الطريقة الثانية: تصحيح الاختبار من نوع الاختيار من متعدد باستخدام معادلة التصحيح من اثر التخمين (المكافأة): تكون علامة المفحوص على الاختبار المصحح بهذه الطريقة

$$\text{تساوي: عدد الإجابات الصحيحة} + \frac{\text{عدد الفقرات المحذوفة}}{\text{عدد البدائل}}$$

3. الطريقة الثالثة: تصحيح الاختبار من نوع الاختيار من متعدد باستخدام معادلة التصحيح من اثر التخمين (المكافأة والعقاب): تكون علامة المفحوص على الاختبار المصحح بهذه

$$\text{الطريقة تساوي: عدد الإجابات الصحيحة} + \frac{\text{عدد الفقرات المحذوفة}}{\text{عدد البدائل}} - \frac{\text{عدد الإجابات الخاطئة}}{\text{عدد البدائل} - 1}$$

صعوبة الفقرة: القيمة التي تدرج على متصل السمة ويكون احتمال إجابة الفقرة عندها إجابة صحيحة مساويا 50%.

تمييز الفقرة: معدل التغير في احتمال الاستجابة الصحيحة للأفراد على المفردة بالنسبة لمستوى القدرة، وتقدر هذه القوة بمعامل التمييز الذي يقصد به الميل النسبي للمنحنى المميز للمفردة على محور القدرة.

قدرة الفرد: قيمة يتم تقديرها تعظم أرجحية استجابات الفرد عن فقرات الاختبار.

الطريقة والإجراءات

عينة الدراسة

تكونت عينة الدراسة من جميع الطلاب الذين درسوا مادة مبادئ في القياس والتقويم بكلية التربية في جامعة مؤتة، وقد بلغ عددهم (276) طالبا وطالبة موزعين على خمس شعب، وقد تم اختيار شعبة بطريقة عشوائية من بين الشعب الخمسة وكان عدد المسجلين فيها (55) طالبا وطالبة شكلوا عينة الدراسة الاستطلاعية. والباقي وعددهم (221) شكلوا عينة الدراسة التجريبية.

أدوات الدراسة

تم تطوير اختبار تحصيلي يشتمل على 40 فقرة من نوع الاختبار من متعدد ذات أربع بدائل لقياس المصطلحات والمفاهيم الأساسية في مقرر "مبادئ القياس والتقويم" الذي يُدرّس لطلاب كلية التربية بجامعة مؤتة. وذلك بالاستفادة من الكتب المتخصصة بهذا المجال.

بعد ذلك تم عرض فقرات هذا الاختبار على سبعة من أساتذة علم النفس (خمسة منهم يُدرّسون مادة مبادئ في القياس) لمراجعة فقرات الاختبار للتحقق من وضوح عباراتها وسلامة صياغتها واستيفائها لشروط هذا النوع من الفقرات وملائمة الفقرات للمحتوى المقرر. وبعد مراجعة الفقرات بناءً على مقترحات المحكمين ومناقشتهم تم تعديل بعض فقرات الاختبار وحذف مجموعة أخرى لعدم مطابقتها لمحتوى الخطة الدراسية لهذا المقرر في جامعة مؤتة، ليتكون الاختبار في شكله النهائي من 30 فقرة.

الإجراءات

أولاً: تطبيق الاختبار التحصيلي على عينة الدراسة الاستطلاعية

طُبِّق الاختبار على أفراد عينة الدراسة الاستطلاعية في ضوء تعليمات تنص على ان تصحيح الاختبار سوف يكون على أساس مجموع الإجابات الصحيحة. وحُسب الوسط الحسابي، والانحراف المعياري، ومعامل الثبات باستخدام معادلة كرونباخ (α) والجدول رقم (1) ببيان نتائج المفوضين في عينة الدراسة الاستطلاعية على الاختبار التحصيلي.

جدول (1): الوسط الحسابي والانحراف المعياري ومعامل الثبات للاختبار التحصيلي، عند عينة الدراسة الاستطلاعية.

عدد المفوضين	الوسط الحسابي	الانحراف المعياري	معامل الثبات
55	15.5	5.68	0.82

ويتضح من الجدول أن معامل ثبات الاختبار هو 0.82 وهو مؤشر كافي الاستخدام هذا الاختبار لأغراض الدراسة، خاصة إذا أخذنا بعين الاعتبار كون الاختبار قصير عدد فقراته 30 فقرة فقط.

ثانياً: تطبيق الاختبار على عينة الدراسة التجريبية

1. تم تقسيم فقرات الاختبار إلى ثلاث مجموعات، بحيث تشتمل كل مجموعة منها على (10) فقرات. رتبنا أسئلة هذه المجموعات الثلاث في ثلاثة نماذج.
2. تم توزيع النماذج الثلاثة من الاختبار عشوائياً حسب الطريقة التي سيتم بها تصحيح كل نموذج فكانت على النحو التالي:

– **النموذج الأول:** استخدام معادلة التصحيح لأثر التخمين (العقاب) والتعليمات المنسجمة معها.

- **النموذج الثاني:** استخدام معادلة التصحيح لأثر التخمين (المكافأة) والتعليمات المنسجمة معها.
- **النموذج الثالث:** استخدام معادلة التصحيح لأثر التخمين (المكافأة والعقاب) والتعليمات المنسجمة معها.
3. قُسم أفراد الدراسة عشوائياً إلى ثلاث مجموعات.
4. أعطيت كل مجموعة واحداً من نماذج الاختبار مع التعليمات المنسجمة مع طريقة تصحيحه، وتم توضيح التعليمات للمشاركين، كما تم تذكيرهم بالاستراتيجية التي يفضل اتباعها عند اخذ كل نموذج، وبعد التأكد من فهم المشتركين للتعليمات، طلب إليهم البدء بالإجابة على الاختبار.
5. وبعد أسبوع، تم تطبيق نموذج آخر من الاختبار على كل مجموعة بحيث يكون مختلفاً عن النموذج الذي تم تطبيقه قبل أسبوع.
6. بعد مرور أسبوع آخر. تم تطبيق النموذج الأخير لكل مجموعة.
- وبذلك تكون المجموعات الثلاث قد جلست للاختبار ثلاث مرات لاستكمال النماذج الثلاث والجدول رقم (2) يوضح الترتيب الذي اتبع في التطبيق.
- جدول (2):** ترتيب تطبيق نماذج الاختبار على المجموعات الثلاث.

المجموعة			الترتيب
الثالثة	الثانية	الأولى	
النموذج الثاني	النموذج الثالث	النموذج الأول	أولاً
النموذج الثالث	النموذج الأول	النموذج الثاني	ثانياً
النموذج الأول	النموذج الثاني	النموذج الثالث	ثالثاً

وقد بلغ عدد الذين اخذوا الأشكال الثلاثة من الاختبار التحصيلي (211) طالبا وطالبة، فقد تغيب عشرة من أصل عينة الدراسة التجريبية، والتي كانت عددها 221. كما وجد بعد تصحيح الاختبار ان هناك 11 ورقة قد كانت إجابة المفحوصين عليها غير جادة، فأسقطت هذه الورقات، ليصبح عدد أفراد عينة الدراسة التجريبية، التي خضعت نتائجهم للتحليل الإحصائي (200) طالبا وطالبة.

المعالجات الإحصائية

قبل معالجة البيانات إحصائياً كان لا بد من فحص الافتراضات الأساسية التي يجب توفرها في البيانات عند تطبيق نموذج راس، وأهم هذه الافتراضات:

1- أحادية البعد للسمة المقيسة. وقد تم التأكد من ذلك من خلال التحليل العاملي للبيانات وحساب معامل الارتباط بين الأداء على الفقرة والعلامة الكلية لجميع فقرات الاختبار في كل نموذج. حيث تم إجراء تحليل عاملي للبيانات باستخدام طريقة المكونات الأساسية (Principal Components Analysis). وفي التحليل العاملي، تعتمد أحادية البعد على أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني تكون نسبة كبيرة لا تقل عن (2)، وأن ما يفسره العامل الأول من التباين أكثر من 20% تقريباً (Hambleton & Swaminathan, 1985) ويبين الجدول (1) قيم الجذر الكامن، ونسب التباين المفسر للعامل الأول والعامل الثاني، ونواتج قسمة قيمة الجذر الكامن للعامل الأول على العامل الثاني لكل نموذج من النماذج الثلاثة للاختبار.

جدول (3): قيم الجذر الكامن ونسبة التباين المفسر للعاملين الأول والثاني، ونسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني لكل نموذج من نماذج الاختبار الثلاثة.

نسبة الجذر الكامن للعامل الأول إلى الثاني	العامل		النموذج
	الثاني	الأول	
4.511	2.313	10.433	الجذر الكامن
	3.045	20.063	نسبة التباين المفسر
4.327	2.432	10.523	الجذر الكامن
	3.847	20.286	نسبة التباين المفسر
5.171	2.033	10.512	الجذر الكامن
	3.910	20.215	نسبة التباين المفسر

ويتبين من جدول (3) أن الاختبار التحصيلي كان أحادي البعد في النماذج الثلاثة، حيث أن نسبة الجذر الكامن للعامل الأول إلى العامل الثاني كانت كبيرة وتزيد على (2)، وأن نسبة التباين المفسر من العامل الأول في النماذج الثلاثة هي نسبة عالية (أكبر أو تساوي 20% تقريباً). كما تم حساب معامل الارتباط بين الأداء على كل فقرة والعلامة الكلية للاختبار في كل طريقة من طرق معالجة أثر التخمين، حيث تشير الارتباطات المرتفعة بين الفقرة والعلامة الكلية (في حالة تحققها لمعظم الفقرات) إلى أن فقرات الاختبار تقيس بعداً واحداً فقط تعبر عنه العلامة الكلية. كما يمثل معامل الارتباط بين الأداء على الفقرة والأداء على الاختبار القدرة التمييزية للفقرة. ويبين الجدول (4) معاملات الارتباط بين العلامات المستحقة على الفقرات والعلامة الكلية في كل نموذج من النماذج الثلاثة. وقد تبين أن جميع معاملات الارتباط ذات دلالة إحصائية ($\alpha = 0.05$) وأن معظمها يتخذ قيمةً متوسطة لكنها تعد عالية عندما يتعلق الأمر بالفقرات.

جدول (4): معاملات الارتباط بين العلامة المستحقة على الفقرة والعلامة الكلية على الاختبار حسب النموذج.

رقم الفقرة	النموذج			رقم الفقرة	النموذج		
	الأول	الثاني	الثالث		الأول	الثاني	الثالث
1	0.56	0.45	0.48	16	0.42	0.39	0.39
2	0.55	0.56	0.48	17	0.38	0.49	0.49
3	0.51	0.58	0.56	18	0.48	0.61	0.49
4	0.45	0.36	0.48	19	0.53	0.51	0.47
5	0.52	0.58	0.43	20	0.45	0.36	0.48
6	0.28	0.55	0.55	21	0.52	0.58	0.43
7	0.45	0.56	0.49	22	0.28	0.55	0.55
8	0.53	0.47	0.48	23	0.45	0.56	0.49
9	0.41	0.48	0.53	24	0.37	0.46	0.58
10	0.59	0.54	0.59	25	0.48	0.49	0.55
11	0.39	0.69	0.33	26	0.41	0.41	0.33
12	0.59	0.57	0.49	27	0.27	0.56	0.58
13	0.38	0.62	0.49	28	0.55	0.66	0.49
14	0.53	0.51	0.47	29	0.37	0.46	0.58
15	0.56	0.57	0.49	30	0.56	0.54	0.59

النتائج المتعلقة بالسؤال الأول: هل تختلف دقة تقدير معلمة صعوبة الفقرة باختلاف طرق معالجة أثر التخمين وتعليماتها؟

للإجابة عن السؤال الأول تم تقدير قيم معاملات الصعوبة باستخدام النموذج اللوجستي ثنائي المعلمة لكل نموذج من نماذج الاختبار الثلاث حسب متغيري طريقة التصحيح، وتم أيضاً حساب المتوسطات الحسابية والانحرافات المعيارية لقيم هذه المعاملات. ويوضح الجدول (5) قيم معاملات الصعوبة والمتوسطات الحسابية لكل نموذج من نماذج الاختبار.

جدول (5): قيم معاملات الصعوبة والمتوسطات الحسابية والانحرافات المعيارية لقيم معاملات الصعوبة لكل نموذج من نماذج الاختبار حسب متغير طريقة معالجة أثر التخمين.

رقم الفقرة	طريقة المعالجة			رقم الفقرة	طريقة المعالجة		
	الأولى	الثانية	الثالثة		الأولى	الثانية	الثالثة
1	0.076	-1.181	1.297	16	0.12	1.465	0.249
2	0.024	-0.767	-1.101	17	-0.746	-1.154	-0.949

...تابع جدول رقم (3)

رقم الفقرة	طريقة المعالجة			رقم الفقرة	طريقة المعالجة		
	الأولى	الثانية	الثالثة		الأولى	الثانية	الثالثة
3	-0.412	-0.857	-0.589	18	0.461	1.349	-0.692
4	1.271	0.21	-0.881	19	0.19	0.169	-0.658
5	-0.153	-0.857	0.611	20	0.069	0.492	-1.62
6	0.256	0.184	0.318	21	1.448	0.821	0.631
7	0.829	-0.304	-0.912	22	0.388	0.337	0.906
8	0.772	1.056	0.606	23	-0.535	-0.142	0.2
9	-1.271	1.324	-0.312	24	1.139	0.096	1.061
10	0.057	-0.645	1.07	25	0.414	1.167	-0.957
11	0.589	-1.697	1.638	26	0.301	0.41	0.186
12	0.106	0.017	-1.797	27	1.11	0.55	1.397
13	0.153	-1.614	-0.688	28	1.769	-0.491	0.037
14	1.58	0.218	1.514	29	-1.322	0.429	1.88
15	2.07	0.665	1.519	30	1.312	0.951	-0.134
الوسط الحسابي							
الانحراف المعياري							
					-0.030	-0.131	0.533
					0.92	1.04	1.11

يتضح من النتائج الواردة في الجدول (5) تباينا ملحوظا في مدى صعوبة الفقرات، حيث كان المدى في صعوبة الفقرات هو الأعلى في النموذج الثالث الذي تراوحت قيمه من -1.62 إلى 2.07 في حين كان الأقل النموذج الأول الذي تراوحت قيمه من -1.697 إلى 1.465، إن التباين في مدى قيم صعوبة الفقرات أظهر تباينا ملحوظا في الأوساط الحسابية لقيم هذه المعالم، ولمعرفة الاختلافات بين معلم صعوبة الفقرات حسب متغير طريقة التصحيح فقد تم استخدام تحليل التباين الاحادي *One Way ANOVA* ويوضح الجدول (6) نتائج تحليل التباين الاحادي لمعرفة دلالة الفروق في معالم الفقرات لكل نموذج من نماذج الاختبار تبعا لمتغير طريقة المعالجة لأثر التخمين.

جدول (6): نتائج تحليل التباين الاحادي لمتوسطات الأخطاء المعيارية في تقدير معالم الصعوبة تبعا لطريقة المعالجة.

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف المحسوبة	مستوى الدلالة
بين المجموعات	7.682	2	3.841	3.662	.030
داخل المجموعات	91.248	87	1.049		
الكلية	98.929	89			

يلاحظ من الجدول (7) وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) بين تقديرات معالم الصعوبة تعزى لاختلاف طريقة المعالجة. وبهدف الكشف عن مواقع هذه الفروق بين طرق التصحيح، استخدم اختبار شيفية (Scheffe's Method) للمقارنات الثنائية، ويبين الجدول (7) المقارنات الثنائية بين أوساط الأخطاء المعيارية لتقديرات معالم الصعوبة حسب طريقة المعالجة.

جدول (7): نتائج المقارنات الثنائية بين متوسطات الأخطاء المعيارية لتقديرات صعوبة الفقرات بحسب طريقة المعالجة.

الطريقة	الأولى	الثانية	الثالثة
الأولى	-	*0.6639	0.5632
الثانية		-	0.1007
الثالثة			-

يتضح من الجدول (7) وجود فروق ذات دلالة إحصائية ($\alpha \leq 0.05$) بين الأوساط الحسابية لمعاملات صعوبة فقرات الاختبار بين الطريقة الأولى والثانية، كانت لصالح الطريقة الأولى (معادلة العقاب). في حين لم تكن هناك الفروق الدالة إحصائياً بين الوسط الحسابي لمعاملات الصعوبة عند استخدام الطريقة الثالثة، والوسط الحسابي للطريقة الثانية والأولى. إن تفسير هذه النتائج، يمكن أن يكون في ضوء الاستراتيجيات التي يتبعها المفحوص عند أخذ الاختبار، والتعليمات المرافقة لكل من الطرق الثلاث لتصحيح أثر التخمين، وأثر ذلك على تباين علامات المفحوصين على الاختبار. فعندما يعطى المفحوص التعليمات المرافقة لطريقة التصحيح الأولى التي تركز على معاقبة المفحوص إذا قام بالتخمين سيكون هناك حد من ميل المفحوص للتخمين. ولكن ذلك يعتمد على طبيعة المفحوص، فطبيعة المفحوص تدفعه دائماً للحصول على علامة عالية، فالمفحوص تحت تعليمات المعادلة الأولى سيجيب على الفقرات التي يعرف إجابتها وتبقى الفقرات الأخرى التي لا يعرف المفحوص إجابتها أو لديه معرفة جزئية عن هذه الفقرات، فهو أمام حالتين، الأولى أن يترك الفقرات دون إجابة ويلتزم بتعليمات معادلة التصحيح وستبقى علامته هي مجموع الفقرات التي يعرف إجابتها. الثانية أن لا يلتزم بتعليمات المعادلة ويحاول تخمين بعض هذه الفقرات على أمل أن يكون تخمينه صحيحاً. وبما أن العقاب يولد أحيانا التمرد على التعليمات فان بعض المفحوصين سيعمل على تخمين بعض الفقرات، وبالتالي يزداد التباين الكلي في الأخطاء وهذا يؤدي إلى زيادة صعوبة الفقرة. حيث يجيب على الفقرات طلاب ذوي قدرات منخفضة نتيجة لاستخدامهم للتخمين مما يؤدي إلى التقليل من احتمالية الإجابة الصحيحة وهذا يزيد من صعوبة الفقرة. في حين إن تعليمات المعادلة الثانية تعطي المفحوص شعوراً بأنه سيكافئ على الفقرات المتروكة دون إجابة (التي لا يعرف إجابتها) وبالتالي سيزيد ذلك من العلامة الكلية له، سيدفع به للابتعاد عن التخمين مقارنة مع تعليمات المعادلة الأولى، وبالتالي سيقبل التباين الكلي في الأخطاء مما يؤدي إلى انخفاض في صعوبة الفقرة، ولكن بعض المفحوصين سيضحي بتخمين بعض الفقرات وخصوصاً إذا فكر

بالطريقة التالية "إذا جاءت الإجابة صحيحة على الفقرة سأحصل على علامة وان جاءت خاطئة فلن اخسر شيئاً". أما تحت تعليمات المعادلة الثالثة فإن العقاب والمكافأة يجتمعان وبالتالي لن يفكر المفحوص بالتفكير السابق وسيلتزم إلى حد كبير بالتعليمات ولن يلجأ للتخمين، وبالتالي نحصل على تقدير دقيق لمعلم صعوبة الفقرة.

النتائج المتعلقة بالسؤال الثاني: هل تختلف دقة تقدير معلمة تمييزه الفقرة باختلاف طرق معالجة أثر التخمين وتعليماتها؟

للإجابة عن السؤال الأول تم تقدير قيم معاملات التمييز باستخدام النموذج اللوجستي ثنائي المعلمة لكل نموذج من نماذج الاختبار الثلاث حسب متغير طريقة التصحيح. وتم أيضاً حساب المتوسطات الحسابية والانحرافات المعيارية لقيم هذه المعاملات. ويوضح الجدول (8) قيم معاملات التمييز والمتوسطات الحسابية لكل نموذج من نماذج الاختبار.

جدول (8): قيم معاملات التمييز والمتوسطات الحسابية والانحرافات المعيارية لقيم معاملات التمييز لكل نموذج من نماذج الاختبار حسب متغير طريقة المعالجة.

رقم الفقرة	طريقة المعالجة			رقم الفقرة	طريقة المعالجة			رقم الفقرة
	الأولى	الثانية	الثالثة		الأولى	الثانية	الثالثة	
1	1.945	2.21	1.266	16	1.643	1.117	1.241	1
2	3.93	1.738	2.611	17	2.228	2.544	2.549	2
3	4.235	3.277	1.48	18	1.106	4.545	0.981	3
4	3.189	3.026	2.716	19	3.881	2.377	2.557	4
5	3.312	1.105	2.976	20	2.951	2.105	2.761	5
6	1.942	1.314	2.956	21	2.908	2.606	2.106	6
7	2.431	2.859	1.325	22	2.036	1.072	3.557	7
8	1.308	1.747	1.553	23	2.108	2.264	3.154	8
9	2.194	2.266	3.013	24	3.227	4.032	1.254	9
10	2.962	2.163	1.735	25	3.717	1.95	3.115	10
11	2.731	1.712	1.347	26	2.482	3.304	1.385	11
12	2.52	1.164	1.259	27	2.097	2.603	2.102	12
13	1.336	2.596	2.435	28	2.392	3.323	1.301	13
14	1.497	3.818	2.75	29	2.347	2.963	0.803	14
15	1.534	3.206	2.427	30	2.431	1.832	2.513	15
	الوسط الحسابي				2.454	2.361	2.108	
	الانحراف المعياري				.820	.896	.784	

يتضح من النتائج الواردة في الجدول (8) تباينا ملحوظا في مدى صعوبة الفقرات، حيث كان المدى في صعوبة الفقرات هو الأعلى في النموذج الأول الذي تراوحت قيمه من 1.106 إلى 3.93 في حين كان الأقل النموذج الثاني الذي تراوحت قيمه من 1.072 إلى 3.818، إن التباين في مدى قيم تمييز الفقرات اظهر تباينا ملحوظا في الأوساط الحسابية لقيم هذه المعالم، ولمعرفة الاختلافات بين معلم تمييز الفقرات حسب متغير طريقة معالجة أثر التخمين، فقد تم استخدام تحليل التباين الاحادي *One Way ANOVA* ويوضح الجدول (9) نتائج تحليل التباين الاحادي لمعرفة دلالة الفروق في معالم الفقرات لكل نموذج من نماذج الاختبار تبعا لمتغير طريقة المعالجة.

جدول (9): نتائج تحليل التباين الأحادي لمتوسطات معالم التمييز تبعا لطريقة المعالجة.

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف المحسوبة	مستوى الدلالة
بين المجموعات	1.929	2	.965	1.599	.208
داخل المجموعات	52.486	87	.603		
الكلية	54.415	89			

يلاحظ من الجدول (9) عدم وجود فروق ذات دلالة إحصائية عند مستوى الدلالة ($\alpha = 0.05$) بين تقديرات معالم التمييز تعزى لاختلاف طريقة المعالجة. ويعزو الباحث عدم وجود الفروق الدالة إحصائياً تبعا لاختلاف طريقة المعالجة، إلى أن معلمة التمييز تقسم المفحوصين إلى متمكنين وغير متمكنين، وتعتمد إجابة الطالب على فقرة الاختبار على قدرة المفحوص، فالمفحوص المتمكن من محتوى الفقرة، يستطيع إتباع خطوات حل علمية بصورة متسلسلة دون الوقوع في أخطاء، بعكس المفحوص غير المتمكن، كما أن محتوى متن الفقرة واحد عند جميع المفحوصين.

النتائج المتعلقة بالسؤال الثالث: هل تختلف دقة تقدير معلمة قدرة الفرد باختلاف طرق معالجة أثر التخمين وتعليماتها؟

للإجابة عن هذا السؤال تم استبعاد بيانات الأفراد والفقرات التي لم تتلاءم مع توقعات النموذج في أي طريقة من طرق معالجة أثر التخمين المعتمدة. وتم إيجاد تقدير لقدرة كل مفحوص في كل طريقة من طرق التصحيح المعتمدة، وكذلك الخطأ المعياري في تقدير هذه القدرة. و يبين جدول (10) الأوساط الحسابية والانحرافات المعيارية لتقديرات القدرة والاختفاء المعيارية في تقديرها لكل طريقة من طرق معالجة أثر التخمين.

جدول (10): الأوساط الحسابية والانحرافات المعيارية لتقديرات القدرة والأخطاء المعيارية في تقديرها بحسب طرق معالجة أثر التخمين.

الوساط الحسابية	الأولى	الثانية	الثالثة
الانحراف المعياري	0.604	0.612	0.528
الوسط الحسابي	0.464-	0.347-	0.343-

تشير النتائج الواردة في الجدول (10) إلى وجود فروق ظاهرية بين المتوسطات الحسابية لتقديرات معلمة القدرة للأفراد حسب متغير طريقة المعالجة، ولمعرفة الدلالة الإحصائية لتلك الفروق الظاهرية تبعاً متغير طريقة المعالجة، تم استخدام تحليل التباين الأحادي *One Way ANOVA* والجدول (11) يوضح ذلك.

جدول (11): نتائج تحليل التباين الأحادي لمتوسطات الأخطاء المعيارية في تقدير معالم القدرة تبعاً لطريقة المعالجة.

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	ف المحسوبة	مستوى الدلالة
بين المجموعات	1.953	2	.977	2.873	.057
داخل المجموعات	202.285	595	.340		
الكلية	204.238	597			

يتبين من الجدول (11) عدم وجود فروق ذات دلالة إحصائية عند ($\alpha = 0.05$) بين المتوسطات الحسابية لتقديرات معلمة القدرة للأفراد تعزى متغير طريقة المعالجة. ويمكن تفسير النتائج السابقة كون نماذج الاختبار المطبقة على الطلبة هي نفسها، إذ لم تتغير متون هذه الفقرات أو حتى عدد فقرات الاختبار، وإنما كان التغيير فقط في تعليمات الاختبار. ويمكن تفسير هذه النتائج أيضاً، في ضوء خصائص كل معادلة من معادلات التصحيح، فإذا كان هناك اختبار يتكون من 10 فقرات اختيار من متعدد (ذات بدائل أربع)، وكان أحد المفحوصين يعرف إجابة 6 فقرات، وتقدم لهذا الامتحان تحت تعليمات المعادلة الأولى (معادلة العقاب) والتزم بتعليماتها فإنه بتطبيق المعادلة على إجابته ستكون علامته 6، وان لم يلتزم وضمن الفقرات التي لا يعرف إجابته فإنه حسب الاحتمال النظري للتخمين العشوائي سيحصل على علامة واحدة زيادة على علامته الحقيقية ولكن اجراءات هذه المعادلة سوف تعاقبه على ذلك بخصم علامة وبالتالي فإن علامته ستبقى كما هي ان لم تنقص. إما إذا تقدم للامتحان تحت تعليمات المعادلة الثالثة والتزم بتعليماتها فإنه بتطبيق المعادلة على إجابته فستكون علامته 7، وان لم يلتزم وضمن الفقرات التي لا يعرف إجابته فلن يؤدي ذلك لزيادة علامته وربما يؤدي ذلك إلى نقصانها. إما إذا تقدم للامتحان تحت تعليمات المعادلة الثانية والتزم بتعليماتها فإنه بتطبيق المعادلة على إجابته فستكون علامته 7، وان لم يلتزم وضمن الفقرات التي لا يعرف إجابته، حسب الاحتمال النظري

للتخمين العشوائي سيحصل على علامة واحدة زيادة على علامته الحقيقية وبالتالي ستزداد علامته ويكون تحصيله افضل من المرتين السابقتين.

من خلال هذه النتائج مجتمعة يمكننا أن نصل إلى خلاصة مفادها أن طريقة التصحيح الأكثر ضبطاً لأثر التخمين هي معادلة التصحيح لأثر التخمين (العقاب والمكافأة)، وهي الأكثر دقة في قياس قدرات الأفراد، وتقدير قيم صعوبة الفقرات. لذلك يُنصح باستخدامها وتفضيلها على الطرق الأخرى. ويتوقع أن يترك استخدام المدرسين لهذه الطريقة أثراً إيجابياً لدى الطلبة يتمثل بدقة القراءة والاستعداد للاختبارات وقد يزيد ذلك من دافعيتهم للتعلم، كما يتيح استخدامها فرص زيادة دقة تقدير قدرات الأفراد مما ينعكس إيجاباً على صدق تلك التقديرات وثباتها.

لقد تبين من النتائج التي توصلت إليها الدراسة الحالية بأنه يوجد أثر لطرق معالجة أثر الخمين على معلم صعوبة الفقرة في حين لم يكن هناك أثر لطرق معالجة أثر الخمين على معلمة التمييز للفقرة وعلى معلمة القدرة للأفراد، وفي ضوء تلك النتائج فإن الباحث يوصي بإجراء دراسة مماثلة بحيث يتم فيها استخدام النموذج ثلاثي المعلم. ويوصي الباحث كذلك بإجراء دراسات مقارنة بين النظرية الكلاسيكية ونظرية الاستجابة للفقرة في معالجة وضبط أثر التخمين.

References (Arabic & English)

- Abdel wahab, Mohammed. (2013). Comparison of Classical Test Theory and Three Parameter Logistic Model in Processing Guessing Effect when Estimating Item/Person Statistics for Multiple-Choice Test. *Journal of Educational Sciences*, (109). 169-206.
- Alkursheh, taha. (2004). *The Effect of the Correction-for-Guessing Formulae on the Reliability of Tests, Standard Error of Measurement and Distribution's Characteristics*. Master Thesis. Mu'tah University.
- Allam, Salahuddin. *Educational and psychological measurement and evaluation*, Dar AL fikr. Cairo.
- Alnabhan, Mousa. (2002). An empirical investigation of the effects of three methods of handling guessing and risk taking on the psychometric indices of a test. *Social Behavior and Personalty*, 30(7), 645 – 652.
- Al-Shraifin, Nedal & Ta'mnah, Eman. (2009). The Effect of Multiple Choice Test Number of Alternatives on The Estimation of a person's Ability and the Psychometric Properties of a Test and its Items

According to Rasch Model in Item Response Theory (IRT). *Jordan Journal of Educational Sciences*. (4). 309-335.

- Andra, C. & Magnano, G. (2011). *Multiple-choice math tests: should we worry about guessing?* Quaderni di Ricerca in Didattica (Mathematics), 21, 235-243.
- Attali, Y. (2005). Reliability of Speeded Number-Right Multiple-Choice Tests. *Applied Psychological Measurement*, 29 (5), 357-368.
- Ben-Shakher, G. & Sinai, Y. (1991). Gender differences in multiple-choice: the role of differential guessing tendencies. *Journal of Educational Measurement*, Vol 28(1). 23-35.
- Ben-simon, A. & Budescu, D. & Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple – choice tests. *Applied psychological measurement*, 21(1). 65-88.
- Brown, F. (1983). *Principles of educational and psychological testing*. (3rded), Holt Rinehart and Winston, Inc, New York.
- Camilli, G. & Chiu, T. (2006). *Two New IRT-based Corrections for Guessing*. University of Colorado, Boulder.
- Crocker, L. & Algina, J. (1986), *Introduction to Classical and Modern Test Theory*. University of Florida, U.S.A.
- Hambleton, R.K., & Jonse, R.W. (1993). *Comparison of Classical Test Theory and Item Response Theory and their Applications to test development*. Educational Measurement. Issues and Practice, 38-47.
- Jaradat, D. & Tollefson, M. (1988). The impact of alternative scoring procedures for multiple-choice items on test reliability, validity, and grading. *Educational and psychological Measurement*, Vol. (48). 627-635.
- Muijtjens, A. & Vanmameren, H. & Hoogenboom, R. & Evers, J. & Vnder, V. (1999). The effect of a 'don't know' option on test scores: number-right and formula scoring compared, *Medical Education*, Vol 33(4). 267-275.