Journal of the Association of Arab Universities for Research in Higher Education (مجلة اتحاد الجامعات العربية (للبحوث في التعليم العالي

Volume 42 | Issue 1 Article 6

2022

Correction for Guessing of Multiple-Choice Tests Between Partial Knowledge and Risk-taking: A Suggested Model for Bi-directional Scale

Ahmad Suleiman A. Bani Ahmad
Yarmouk University-Jordan, audeh@yu.edu.jo

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

Part of the Curriculum and Instruction Commons

Recommended Citation

Bani Ahmad, Ahmad Suleiman A. (2022) "Correction for Guessing of Multiple-Choice Tests Between Partial Knowledge and Risk-taking: A Suggested Model for Bi-directional Scale," *Journal of the Association of Arab Universities for Research in Higher Education (مجلة اتحاد الجامعات العربية (للبحوث في التعليم العالي*: Vol. 42: Iss. 1, Article 6.
Available at: https://digitalcommons.aaru.edu.jo/jaaru_rhe/vol42/iss1/6

This Article is brought to you for free and open access by Arab Journals Platform. It has been accepted for inclusion in Journal of the Association of Arab Universities for Research in Higher Education (مجلة اتحاد الجامعات العربية (للبحوث في التعليم العالي by an authorized editor. The journal is hosted on Digital Commons, an Elsevier platform. For more information, please contact rakan@aaru.edu.jo, marah@aaru.edu.jo, u.murad@aaru.edu.jo.

Journal of the Association of Arab Universities for Higher Education Research (JAARU-RHE)
Vol.(42) – No. (1) – March 2022
مجلة اتحاد الجامعات العربية للبحوث في التعليم العالي
المجلد(42) – العدد (1)– آذار/ مارس 2022

التصحيح لأثر التخمين في اختبارات الاختيار من متعدد بين المعرفة الجزئية والمخاطرة: نموذج مقترح لتدريج ثنائي الاتجاه

Correction for Guessing of Multiple-Choice Tests Between Partial Knowledge and Risk-taking: A Suggested Model for Bi-directional Scale

Ahmad Suleiman Audeh Bani Ahmad
Faculty of Education
Yarmouk University
The Hashimate Kingdom of Jordan
audeh@yu.edu.jo

أحمد سليمان عودة بني أحمد
كلية التربية
جامعة اليرموك
audeh@yu.edu.jo

Abstract

The correction for guessing in tests with multiple-choice items still a controversial issue, the contribution of literature reveals that more research is needed to investigate and interpret the relationship between the achievement based on partial knowledge and the risk-taking variable in terms of relevant or irrelevant validity when subset selection and penalty for partial guessing takes place. The inclusion of sample questions that have no answer within the test questions is considered as a source of an irrelevant variance which affect the relevant validity. This study suggested a solution for this negative impact by dealing with bidirectional scale. The Operational definition of this scale comes from linking the achievement and risk variables in 7-points. Real data of 64 students and 50 items based on subset selectin method were presented for statistical and logical analysis. The results support the suggested approach for dealing with derived observed scores as if they are two-in-one scale. The statistical results, interpretations, and derived conclusions revealed that this approach is acceptable and promising theoretically and practically in the domain of test Theory, adding a new formula for adjusting the observed scores of the cognitive variable based on the scores of the risk variable.

Key words: Partial Knowledge, Intelligent Guessing, Risk-taking, Bi-directional Scaling, Multiple Choice Items.

الملخص

كانت وما زالت قضية التصحيح لأثر التخمين قضية جدلية، وتتحدث أدبيات القياس والتقويم عن شبكة من العلاقات بين مفاهيم التخمين الذكي والتخمين العشوائي والمخاطرة، والتصحيح لأثر التخمين بالعقاب أو المكافأة. ومن ضمن ما طرحته تلك الأدبيات أسلوب الاختيار المتدرج لأكثر من بديل في إطار المعرفة الجزئية، وإدراج عينة أسئلة ليس لها إجابة ضمن أسئلة الاختبار للكشف عن المخاطرة، ولم تعالج الأدبيات التأثير السلبي للتعامل مع العلامات السالبة التي تفرزها عملية التصحيح وفق أسلوب الاختيار المتدرج، ولم تعالج أيضا مخالفة أسس التقويم بالاختبارات عند إدراج فقرات ليس لها إجابه بغرض الكشف عن المخاطرة. وجاءت هذه الدراسة بتصور مقترح للربط بين متغيري التحصيل والمخاطرة في تدريج ثنائي الاتجاه بسبع فئات، يكون فيها الصفر مشتركاً بين التدريجين، باعتماد بيانات واقعية لعينة من 64 طالباً و50 فقرة في موقف اختباري حقيقي لمساق جامعي، ثم تحليل تلك البيانات لتقديم المؤشرات الإحصائية والمنطقية الداعمة لهذا المدخل المقترح والمصمم للتعامل مع متغير المخاطرة كمصدر تباين يضيف ثباتا زائفا على حساب الصدق. وكانت النتائج داعمة لهذا التصور. وفي ضوء النتائج وتفسيراتها والاستنتاجات المستخلصة منها، قدمت الدراسة متغير المخاطرة خياراً واقعياً في التعامل مع العلامات السالبة وربطها بمتغير التحصيل. كما قدمت الدراسة معادلة لتصحيح العلامات وفقا لمتغير المخاطرة.

كلمات مفتاحية: المعرفة الجزئية، التخمين الذكي، المخاطرة، التدريج ثنائي الاتجاه، الاختيار من متعدد.

Received: 15/12/ 2021, Accepted: 23/2/ 2022
Doi:10.36024/1248-042-001-006
تأريخ الإستلام: 15/12/ 2021، تأريخ القبول: 2022/2/23
ISSN: 1680-6549

مقدمة

من القرارات الصعبة التي تواجه المختصين في القياس والتقويم التربوي تقديم الاستراتيجية المناسبة للتصحيح لأثر التخمين وحساب العلامات في الاختبارات من نوع الاختيار من متعدد، فما زال الجدل واضحاً مع تزايد الاهتمام بالحديث عن التخمين الذكي، والتخمين العشوائي، والمخاطرة، والمعرفة الجزئية، والحكمة في التصرّف أثناء الإجابة Testwiseness. وقد كانت هذه الزيادة أكثر وضوحاً في التقويم عن بعد Online Assessment، وخاصة في الجامعات التي أصبح فيها التعليم عن بعد خياراً بديلاً للتعليم التقليدي بسبب جائحة كورونا. فقد كان التركيز واضحاً على الاختيار من متعدد لأسباب متعلقة بمزايا هذا النوع من الاسئلة، على الرغم من بعض السلبيات الجوهرية، ومن أبرزها العلاقة بين التخمين والمخاطرة، وتأثيرهما على علامة الطالب في الاختبار. وقد أدى الرفض الضمني لمبدأ التصحيح لأثر التخمين بالعقاب الذي يقوم على العشوائية في التخمين أو التخمين الأعمى Bling Guessing إلى توجيه الجهود البحثية للمختصين في القياس والتقويم لتقديم ما قد يقلل من سلبيات هذا النوع من الأسئلة، من خلال تعدد استراتيجيات التصحيح، مع تقديم تعليمات الإجابة المتوافقة مع هذه الاستراتيجيات؛ ولذلك تعددت محاور البحث في هذا الموضوع حيث تناولت جدلية التصحيح لأثر التخمين، ونوع التصحيح، ومعادلاته وامكانية استخدامها. محاولة بذلك الربط بين التخمين العشوائي والتخمين الذكي من جهة، والمخاطرة Risk-Taking مقابل التحفظ Risk Averse من جهة أخرى. وما زالت الجدلية قائمة؛ ولذلك فإن من القرارات الصعبة التي تواجه التربويين هو اختيار أو تقديم الوصفة المناسبة في هذا الإطار.

ولتوضيح أبرز مداخل البحث في التعامل مع التخمين، يمكن الاشارة إلى أن منها ما بحث في علاقة معادلات التصحيح لأثر التخمين بخصائص الفقرات، حيث أكدت بعض الدراسات (Dodeen, 2005; Guilford, 1936)على ضرورة مراعاة أثر التخمين في معادلات حساب صعوبة الأسئلة وقدرتها التمييزية، سواء تم حسابها وفق المعادلات المبنية على نظرية الاختبار Test Theory المعروفة بالنظرية التقليدية في القياس، والقائمة على العلاقة بين العلامات الحقيقية (T) والعلامات الظاهرية (X) حسب المعادلة X = T + E. أو وفق نظرية الفقرة Item Theoryالمعروفة بنظرية الاستجابة للفقرة، والقائمة على العلاقة بين قدرة المفحوص (θ) وصعوبة الفقرة (b) وتمييزها (a) والتخمين (c) حسب المعادلة $P(\theta) = c + (1 - c)/[1 + e^{-1.7a(\theta - b)}]$ التي تجمع المعالم Parameters الثلاثة الخاصة بالفقرة ومعلمة القدرة. وقد قدمت دراسة عودة (Audeh,2014) في إطار نظرية الاختبار صورة معدلة لمعادلة جيلفورد القائمة كليا على التخمين العشوائي. وقارنت بين معاملات الصعوبة للفقرات في الاختبارات باختلاف أُسلوب العقاب المستخدم، وتصحيح هذه المعاملات بمعادلة قائمة على التخمين الذكي لتقدير صعوبة الفقرة.

وبالمقابل، فقد أوصت دراسات مبكرة ومنها دراسة فاري Frary, (1988) بعدم العقاب لأثر التخمين على الرغم من إشارة دراسات أخرى لدور العقاب لأثر التخمين في تحسين صدق الاختبار (Thomas et al., 2006)، لأن الخصائص الشخصية للطالب تؤثر على قراره في التعامل مع المعرفة الجزئية Partial Knowledge ، وتُعاقب المتحفظين، علما بأن تعريف أدبيات القياس لمعامل صعوبة الفقرة يتضمن فرصة التخمين العشوائي حسب عدد البدائل، فقد أشارت دراسة ماريا وجيفر María & Javier, (2010) إلى أن العقاب لأثر التخمين يقلل خطأ القياس، على الرغم من عدم وضوح العلاقة بين تقدير الخسارة في العلامة الناتجة عن العقاب مقابل تقدير الخسارة الناتجة عن المخاطرة. فقد أكدت الأدبيات الحديثة على أهمية عدالة القياس بالاختبارات، والتي تقتضي أن تكون محصلة أي إجراءات في صالح المفحوص (Ypsilandis & Mouti, 2019)، بمعنى ألا يتحمل المفحوص تبعات أخطاء القياس غير المكشوفة، والتي يشكل التخمين أبرزها. فقد أدركت بعض الدراسات المبكرة Little, (1962) خطورة المبالغة في العقاب القائم على التخمين العشوائي، وما يرافقه من قلق على العلامة وامكانية معاقبة المتحفظين، وقدّم معادلة تقوم على تخفيض الجزء الخاص بالعقاب في المعادلة التقليدية إلى النصف، بالقسمة على (2).

كما أشارت دراسة بورتون (Burton, 2002) إلى أنه لا ينظر إلى العقاب على أنه مجرد تخفيض علامة الطالب بسبب التخمين، حيث أشار في دراسته إلى رفض الطلبة لأي طريقة تصحيح قائمة على رصد علامات سالبة، وعدم وضوح الفرق في العلامات الذي قد يتحدد من خلال فهم العلاقة بين التحفظ والمخاطرة وتباين الطلبة على متصل العلاقة بين التحفظ والمخاطرة في الإجابة. وأن الطلبة يدخلون في دوامة Dilemma التوفيق بين جدوى الإجابة

مجلة اتحاد الجامعات العربية 42(1) لسنة 2022، 81-95 Doi:10.36024/1248-042-001-006

عن الفقرة أو تركها دون إجابة، بالإضافة الى عدم وضوح الفرق بين الإجابات الخاطئة التي قد تعزى إلى المعلومات الخاطئة Misinformation مقابل ما قد يعزى إلى سوء الفهم Misconception (Jaradat & Tolefson, 1988)، حيث أشارت هذه الدراسة إلى أن نسبة من يمارسون التخمين الأعمى Blind Guessing لا تزيد عن 8%. كما أن تعليمات الاختبار الخاصة بطريقة الإجابة بغرض الحد من التخمين، ومنها القائمة على العلامات السلبية، تنعكس على الاختلاف بين الطلبة في علامة المخاطرة Risk-Taking وطريقة تفاعلهم مع تعليمات الاختبار ذات الصلة بطريقة الإجابة، واشارت نتائج الدراسة إلى أن معادلة الربح والخسارة في العلامات بالنسبة للطالب تعتمد على نسبة الشك في صحة الإجابة.

اهتمت دراسات أخرى في البحث عن أساليب للكشف عن ميل الطلبة للتخمين وعلامة المخاطرة لديهم في ضوء مدى التزامهم بتعليمات الإجابة. فقد اهتمت دراسة جرادات وسواقد (Jaradat and Sawaged, 1986) في الكشف عن متغير المخاطرة بتضمين الاختبار عشرة (10) أسئلة ليس لها جابة في اختبار مكون من 40 سؤالاً على عينة مكونة من 160 طالباً، تم اختيارهم عشوائيا، لقياس تحصيلهم في مبحث مدرسي للصف التاسع. وقد تم تقسيم الطلبة إلى مجموعتين حسب علامة المخاطرة، اعتماداً على وسيط العلامات في أسئلة المخاطرة، وإعادة تقسيم الطلبة إلى ثلاث مجموعات وفق متغير التحصيل متقاطعة مع متغير المخاطرة، وقد تم حساب العلامات بثلاث طرق (عدد الاجابات الصحيحة، وعقاب بالمعادلة التقليدية، والاختيار الجزئي -3 إلى +3)، مشيرة في نتائجها إلى تفوق الطرق التي تحد من التخمين في توفير صدق وثبات أعلى للاختبار. وبالمقابل، فقد استخدمت دراسة النبهان (Alnabhan, 2002) طرق التصحيح الثلاث (عدد الاجابات الصحيحة، وعقاب بالمعادلة التقليدية، والاختيار الجزئي -3 إلى +3)، في اختبار تحصيلي لمساق جامعي، الا أنها اعتمدت أُسلوباً مختلفا في تقسيم الطلبة حسب علامة المخاطرة، وهو عشرة أسئلة ليس لها معنى Nonsense Questions بعلامة قطع حددها الباحث بمن أجابوا عن سبعة (7) أسئلة من أصل عشرة (10). إلا أن وضع مثل هذه الأسئلة يشكل حمولة زائدة في الموقف الاختباري، ويخالف أسس تطبيق الاختبارات من نوع أقصى أداء Maximal Performance ومنها الاختبارات في المساقات الجامعية.

يتضح من الدراسات وأدبيات القياس والتقويم السابقة أن المجال مفتوح للإضافة العلمية في إطار العلاقة بين التخمين والمخاطرة من جهة، ومعادلات التصحيح من جهة أخرى. وتأتي الدراسة الحالية في هذا الإطار لتضيف بعداً آخر، ومدخلاً جديداً مقترحاً، لم تتطرق اليه أي من الدراسات السابقة ويقوم على فكرتين هما:

- أن وضع أسئلة لا إجابة لها بقصد الكشف عن المخاطرة تعد حمولة زائدة على الاختبار، وقد تؤثر بصورة سلبية على الخصائص السيكومترية للاختبار، وإن وضع أسئلة لغرض غير الغرض الذي وضع الاختبار من أجله أمر غير مقبول تربويا، فالاختبار مصمم لقياس تحصيل الطلبة في المساق، وليس لقياس المخاطرة.
- أن الاختيار الجزئي Subset Selection أسلوب معمول به كمدخل للتعامل مع التخمين الذكي ومراعاة المعرفة الجزئية، ولكنّه بنفس الوقت قد يؤدي إلى حصول الطالب على علامة سالبة على مستوى السؤال الواحد، وربما على مستوى العلامة الكلية؛ وبالتالي لم يضع هذا المدخل حلاً للعلامات السالبة التي تفرزها معادلات التصحيح القائمة على العقاب، سواء راعت أو لم تراع المعرفة الجزئية. ولم يكن هناك قناعة من الناحية النظرية أو التطبيقية بوجود علامات سالبة (Ndu et al. 2016; Vanderoost et al., 2018) لصعوبة تفسيرها معرفياً أو تحصيلياً لدى الطالب، سواء ارتبطت بنقص المعرفة أو بسوء الفهم؛ ولذلك عادة ما تتم معالجة هذه العلامات ببساطة وسطحية بتحويلها إلى صفر (وهو ما تعتمده برمجية موودل Moodle التي تستخدمها الكثير من الجامعات في التعليم الالكتروني https://docs.moodle.org/311/en/Multiple_Choice_question_type ودون التمييز بين مقادير هذه القيم أو الاستفادة من دلالاتها ومضامينها. وباختصار، لم تقدم الدراسات أساساً نظرياً أو عملياً مقنعاً لقياس المخاطرة مقابل العلامات السالبة.

مشكلة البحث

في ضوء التصور المشار إليه في نهاية المقدمة لطبيعة العلاقة التي مازالت جدلية بين مداخل التعامل مع التصحيح لأثر التخمين والمعرفة الجزئية والمخاطرة، جاءت هذه الدراسة لإلقاء الضوء على الجزء السالب من التدريج في التصحيح القائم على المعرفة الجزئية، انطلاقاً من دلالات الفروق الفردية التي تعكسها القيم السالبة في نظرية القياس، حيث تطرح الدراسة الحالية فكرة

Published by Arab Journals Platform, 2022

التعامل مع تدريج ثنائي متعلق بمتغيري التخمين والمخاطرة، وبالتالي التعامل مع التباين في القيم السالبة كمتغير خاص بعلامة الامعان في المخاطرة؛ ولذلك فإن الإضافة لهذه الدراسة تقوم على محاولة توفير أساس علمي مقترح لإبراز متغير العلامات السالبة بدلا من تسطيح معناه أو طمسه بمثل هذا التحويل غير المنطقي، أو احتسابه على أنه عمق في غياب المعرفة، والتعامل معه على أنه تباين غير صادق True Irrelevant Variance. ويقوم هذا التصور على إجراء مقارنات بين خصائص الامتحان وتوزيع العلامات التي يتم فرزها بأساليب التصحيح التقليدية قبل وبعد التعديل لأثر التخمين (بالمعادلة التقليدية، وبمعادلة لتل)، ثم بالأسلوب القائم على الاختيار الجزئي من بين البدائل، بتدريج رباعي مرة، وآخر سباعي. بالإضافة إلى تقدير الثبات على المصفوفات المختلفة لتوزيع العلامات، وحساب العلامة للطالب على متغير المخاطرة (كمتغير متصل)، ثم تحويلها إلى منفصل ثنائي باعتماد المتوسط الحسابي كعلامة قطع، والتي قد تشكل بمجموعها إجابة عن الأسئلة العامة الآتية: هل يمكن التعامل مع الجزء السالب من متغير العلامات (متغير المعرفة أو التحصيل) على أنه متغير لسمة أخرى؟ وما المؤشرات التي تدعم التعامل مع متغير واحد مقابل التعامل مع متغيرين على التدريج نفسه؟ وهل هناك قيمة مضافة لهذا التصور يمكن استخلاصها من الإجابة عن السؤال المركب الآتي:

- هل يحسّن (أي المتغير المضاف) في معامل الثبات؟ وما صفة هذا التحسن (إن وجد)؟ وما علاقته بمؤشر الصدق وفقا لنظرية القياس التقليدية (نظرية الاختبار)؟ وهل يغيّر في متوسط علامات الطلبة على الاختبار؟ وما هو اتجاه هذا التغير (إن وجد)؟ ويمكن النظر إلى مشكلة الدراسة بصورة عامة من زاوية أخري من خلال الإجابة عن السؤال العام الآتي: كيف يمكن الربط بين علامة مخاطرة الطالب في الإجابة وعلامته القائمة على المعرفة الجزئية وفق تدريج ثنائي الاتجاه؟ والسؤالين الفرعيين الآتيين المنبثقين عنه:
- هل يوجد فرق ذو دلالة إحصائية بين متوسط علامات الطلبة ذوي المخاطرة المرتفعة وذوي المخاطرة المنخفضة باعتماد المتوسط الحسابي كعلامة قطع لتوزيع علامات المخاطرة؟
- كيف تتأثر خصائص فقرات الاختبار (الصعوبة والتمييز)، وصدق الاختبار وثباته بسبب التعامل مع التدريج المزدوج وفقا للتصحيح القائم على وجود علامات سالبة مرة وفق تدريج سباعي، وأخرى بعد تحويل العلامات السالبة إلى صفر وفق تدريج رباعي، مع التحويل للمحافظة على مدى العلامات، وعدم التغاير Scale Invariant في التدريج Jelte & Conor, (2010).

تعريف المصطلحات

من أبرز المتغيرات أو المصطلحات التي وردت في العناوين السابقة وتحتاج إلى بلورة أو تعريف إجرائي هي: متغير المعرفة، ومتغير المخاطرة، وما يرتبط بهما من مصطلحات مثل: علامة القطع، والتخمين العشوائي مقابل المبني على المعرفة الجزئية. وقد اتضحت مثل هذه المصطلحات بنوع من التفصيل في المقدمة، ويمكن اختزال هذه التعريفات بالصورة الآتية:

المخاطرة: ميل الطالب للإجابة بدون معرفة، وذلك باختيار بديل أو أكثر ليس من بينها الإجابة الصحيحة في أسئلة الاختيار من متعدد، ويُعطى الطالب علامة سالبة لكل اختيار حسب الطريقة المعتمدة في تصحيح الإجابات القائمة على الاختيار التدرجي، ويشكل مجموع العلامات السالبة تقديراً كمياً لعلامة المخاطرة، كما وتشكل الزيادة في العلامات السالبة إمعاناً بالمخاطرة.

علامة القطع للمخاطرة: تشير أدبيات القياس والتقويم إلى أساليب متعددة في تحديد علامة أو علامات القطع Sharma & (Jain, 2014). وتعرّف في هذه الدراسة بأنها الوسط الحسابي لتوزيع العلامات السالبة، وهي العلامة المستخدمة في تصنيف الطلبة إلى مخاطرين بعلامة عالية ومخاطرين بعلامة منخفضة (متحفظين نسبياً).

متغير المعرفة: سمة عقلية تتشكل بفعل برنامج تعليمي موجه لتحقيق نواتج تعلم Learning Outcomes محددة يطلق عليها التحصيل، وهو في هذه الدراسة عبارة عن مجموع العلامات التي يحصل عليها الطالب في اختبار بمساق جامعي، وتقابل في نظرية الاختبار العلامة الظاهرية Observed Score التي يتوقع أن تكون مُقدّراً للعلامة الحقيقية True Score المتحررة نسبيا من المتغيرات الدخيلة مثل المخاطرة.

المعرفة الجزئية: مقدار الزيادة في علامة الطالب الناتجة عن زيادة فرصة اختياره للبديل الصحيح كلما زاد عدد البدائل التي يختارها، والعكس صحيح، أي النقص في علامة الطالب الناتجة عن زيادة اختياره لعدد البدائل الخاطئة، ولا يكون البديل الصحيح من بينها. ويشير هذا التعريف إلى الصراع أو المعضلة في

مجلة اتحاد الجامعات العربية 42(1) لسنة 2022، 81-95 Doi:10.36024/1248-042-001-006

الاقدام او الاحجام لدى الطالب عند اختيار أكثر من بديل في السؤال.

التخمين العشوائي: التخمين المبني على الاحتمال الرياضي، ولا ينطلق من أي أساس معرفي، حتى لو كان جزئياً Partial Knowledge، سواء كان عائداً إلى سوء الفهم، أو نقص في المعرفة أو نقص في التصرّف بحكمة Testwiseness.

أهمية البحث

تنبثق أهمية هذا البحث من المدخل المقترح الذي يتوقع أن يقدم حلاً مقترحاً للسلبيات التي تتضمنها الأساليب الأخرى في ربط متغير المخاطرة بمتغير المعرفة، وأساليب التصحيح المختلفة، والتعديل لأثر التخمين المبنية على العقاب أو التخمين الذكي والعشوائي، أو المعرفة الجزئية؛ وذلك في ضوء ما يمكن أن تقدمه نتائج الدراسة في إطار تلك السلبيات التي يمكن تلخيصها على النحو الاتي:

- يتطلب أسلوب الكشف عن المخاطرة وضع أسئلة لا يوجد لها إجابة، وهذه مخالفة صريحة في تطبيق الاختبارات التي تقوم على أن يكون هناك إجابة لكل سؤال دون استثناء، وخاصة في الاختبارات الواقعية بالمساقات الجامعية، وليس المستخلصة من بيانات افتراضية simulated Scores لأغراض بحثية، مما يضيف وجود مثل هذه الأسئلة صعوبة في تقدير تأثيرها على عدالة العلامات وخصائص الاختبار.
- يؤدي عدم التصحيح لأثر التخمين إلى تضخيم العلامات Overestimate، وأن الطريقة التقليدية في العقاب لأثر التخمين تُبخسها Underestimate، ولذلك انتقل بعضه هذه الطرق إلى التأكيد على أهمية الجمع بين العقاب والثواب التي تأخذ بالاعتبار ضمنا الاختلاف في المخاطرة أو التحفظ عند محاولة الاستفادة من المعرفة الجزئية.
- الاختلاف في تفسير العلامات عندما تتاح للطالب الفرصة لاختيار عدد من البدائل التي يتوقع أن تتضمن الإجابة الصحيحة، سواء تم عزو الإجابة إلى المعرفة الخاطئة Misinformation أو إلى الفهم الخاطئ Misconception أو إلى المخاطرة.
- تُجمع الدراسات على أن طريقة الحذف الجزئي Subset Elimination ليست عملية أو واقعية، وأنها مربكة للطالب، ومخالفة للقواعد العامة في التعليمات التي تقوم على تفتيش الطالب عن الإجابة الصحيحة وليس عن الإجابة الخاطئة.

في ضوء هذه السلبيات الضمنية المستخلصة من الأدبيات والدراسات السابقة ذات الصلة، جاءت الدراسة الحالية لتبحث في كيفية الربط بين المعرفة الجزئية وعلامة المخاطرة بمنهجية مختلفة، أو بمدخل مختلف لا يعتمد على إضافة أسئلة ليس لها إجابة، مع الاحتفاظ بطريقة تصحيح قائمة على المعرفة الجزئية، والتعامل مع أكثر من استراتيجية تقوم على أسس إحصائية ومنطقية في تحويل التدريج نفسه؛ للخروج باستنتاجات وتوصيات لمستخدمي الاختبارات، بغرض توجيه قراراتهم المبنية على التقويم بالاختبارات من جهة، وللباحثين لتقديم المزيد من الدراسات الداعمة لنتائج هذه الدراسة من جهة أخرى، أو تقديم مداخل جديدة للربط بين التخمين والمخاطرة.

المحددات والافتراضات

تقوم الدراسة على مجموعة من المحددات Limitations والافتراضات Assumptions، وبالتالي تتحدد مصداقية النتائج في ضوء المؤشرات التي تكشف عن هذه المحددات والمؤشرات الدالة على صحة الافتراضات، وفق التصوّر الآتي:

- لا تميز أدبيات القياس بين الأوزان النسبية للبدائل الخاطئة باختلاف موقع الإجابة الصحيحة عندما يختار الطالب بديلاً أو بديلين خاطئين إلى جانب الإجابة الصحيحة في الفقرات من أربعة بدائل، بمعنى أن المعرفة الجزئية ليست مرتبطة بترتيب اختيار البديل، وهذا ما تفترضه هذه الدراسة.
- تفترض الدراسة ثبات تأثير تفاعل معرفة الطالب بطريقة إجابته (عند اختيار بديل أو أكثر من بدائل الفقرة) بنوع طريقة التصحيح أو عدم التصحيح لأثر التخمين، أو نوع التدريج الذي يراعي أو لا يراعي المعرفة الجزئية على درجته في الاختبار، إلاّ إذا خضع لبرنامج مكثف لتدريب الطلبة على ما يعرف بحكمة الاختبار (Stenlund et al., 2018).
- تفترض الدراسة أن الإجابة غير الصحيحة (بصرف النظر عن عدد خياراته الخاطئة) تحمل معنى التخمين المُحتمل في حالة الطلب من الطالب في التعليمات ألّا يخمّن (حسب المعنى التقليدي لهذا الطلب، والذي يعني أن هناك عقاباً لأثر التخمين يتناسب مع عدد البدائل).
- تم تعريف الطلبة (أفراد الدراسة) بتعليمات الإجابة في لقاء سابق معهم قبل الامتحان، حتى لا تكون مصدراً جوهريا من مصادر الأخطاء في القياس لهذا المستوى من التعقيدات

النسبية في الإجابة، فهي تفترض أن الطلبة استوعبوا طريقة الإجابة متعددة المراحل قبل التطبيق.

- تم تطبيق الامتحان في ظروف منضبطة تضمن تقديم مؤشرات على تحقيق شرط أقصى أداء Maximal Performance من جهة، ومؤشرات كافية للحصول على بيانات صادقة ومتوائمة مع بحث صفي إجرائي Action Research من جهة أخرى.

الطريقة

تشمل الطريقة والاجراءات توضيحا لعينة الدراسة وأداتها وتصميمها، حيث تكونت عينة الدراسة من 64 طالبا بمرحلة الماجستير في مساق للبحث التربوي بجامعة اليرموك، وطبق عليهم اختبار منتصف الفصل (أداة الدراسة)، حيث كان عدد الأسئلة 50 سؤالاً من نوع الاختيار من متعدد بواقع أربعة بدائل لكل سؤال، وقد تم توزيع الأسئلة على الوحدات في المساق الداخلة ضمن اختبار منتصف الفصل. بما يضمن صدق محتوى الاختبار والتمثيل حسب مستويات الأهداف في المجال المعرفي. حيث تم استيراد أسئلة ذلك الامتحان من بنك أسئلة لذلك المساق وفق برمجية كاي- برو مايلستون Chi-Pro Milestone، ثم انتاج صورة ذلك الامتحان وتصميم ورقة الإجابة، ومفتاح التصحيح، باستخدام برمجية كاي سوفت Chi-swift (ويمكن للمهتم الاستفادة من هاتين البرمجيتين بتنزيلهما مجانا من الرابط http://sulieman.net/xget) . كما تم تحديد تعليمات الإجابة حسب طريقتي الإجابة التقليدية باختيار الإجابة الصحيحة، وطريقة الاختيار الجزئي (باختيار اجابة واحدة في حالة الثقة العالية بصحة الإجابة، واختيار بديلين إذا كان الطالب غير قادر على ترجيح أحدهما والثقة بخطأ البدائل الاخرى، أو اختيار ثلاثة بدائل إذا كان الطالب غير قادر على الترجيح لأي من البدائل الثلاثة. كما هو موضح في النموذج التالي لتعليمات الامتحان على ورقة الإجابة المنفصلة عن أوراق الأسئلة (كما في المقطع التالي) هكذا: (كل سؤال له أربعة بدائل فقط)

رقم الفقرة	A	B	C	D
1	◯	◯	◯	◯

- إذا كنت متأكدا (تماماً) من صحة إجابتك لا تختار إلاّ دائرةً واحدةً، وظللها هكذا ● وليكن C

هام جدا: لا تغير إجابتك؛ ولذلك تأكد من أنّ خيارك هو قرارك النهائي (لا تتسرع)

- إذا كنت ترجح وجود بديل أخر منافس لاختيارك الأول (أي أنك محتار بين بديلين ضع الرقم 2 داخل دائرة ذلك البديل (الخيار الثاني) هكذا (2) وليكن A.
- إذا كنت ترجح وجود بديل ثاني منافس (أي أنك محتار بين ثلاثة بدائل من أصل أربعة) ضع الرقم 3 داخل دائرة ذلك البديل (أي الخيار الثالث هكذا (3) وليكن D.
- إذا كنت لا ترجح أي إجابة من الأربعة أتركها فارغة.

وفي ضوء هذه التعليمات، تم التصحيح لإجابات الطلبة في عينة الدراسة، وتحويلها بصورة مصفوفات، بدأت بالحروف حسب ترميز بدائل الإجابة للفقرة الواحدة في ملف نصيNotepad للتدريجين الرئيسين (الثنائي والسباعي)، ثم تحويل هذه الملفات من حروف إلى أرقام وتخزينها بصورة ملفات SPSS. حيث تمت عملية التصحيح وإنشاء ملفات البيانات لتتناسب مع التداريج scales المعتمدة في التصحيح، ومتصل العلامات على مستوى الفقرة، وحساب العلامات الكلية على متغير المعرفة، والعلامات على متغير المخاطرة، على النحو الآتي:

- **ملف العلامات على مستوى الفقرات:**
  - تدريج ثنائي (1،0) حيث يختار الطالب إجابةً واحدةً قد تكون صحيحة أو خاطئة بصرف النظر أن كانت قائمة على التخمين أو على المعرفة.
  - تدريج رباعي (3،2،1،0) ثم تحويل هذا التدريج إلى مستويات مقابلة 0.00، 0.33، 0.67، 1.00 بعد تحويل القيم السالبة إلى صفر على افتراض أنها لا تحتمل إمعاناً في عدم المعرفة.
  - تدريج سباعي (-3، -2، -1، 0، 1، 2، 3) ثم تحويل التدريج إلى مستويات مقابلة 0.00، 0.17، 0.34، 0.50، 0.67، 0.83، 1.00، وقد تمت الاشارة إلى أن هذا التدريج مبني على افتراض أن التقديرات السالبة قد تحتمل تفسيراً معرفياً، إلى جانب التفسير الأقوى وهو المخاطرة كسمة كامنة Construct Irrelevant Variable.
- ملف العلامات الكلية المستخلصة من الملفات الثلاثة السابقة، بمعنى أن لكل طالب خمس علامات (العلامة وفق الثنائي، وعلامة معدلة بالمعادلة التقليدية في التصحيح لأثر التخمين، وعلامة معدلة بمعادلة لتل التي يفترض بأنها جاءت للتقليل من العقاب الجائر، وعلامة باعتماد التدريج الرباعي، وعلامة باعتماد التدريج السباعي)

مجلة اتحاد الجامعات العربية 42(1) لسنة 2022، 81-95 Doi:10.36024/1248-042-001-006

- ملف واحد لخصائص الفقرات (الصعوبة والتمييز) المحسوبة وفق كل تدريج (ثنائي، رباعي، سباعي)

وفيما يأتي توضيح لرموز المتغيرات في ملف العلامات الأصلية حسب طرق التصحيح والمشتقة وفقاً لمعادلة التصحيح:

X1: متغير علامات الطلبة وفق التدريج الثنائي التقليدي (1,0) (لا يوجد تصحيح لأثر التخمين).

X2: متغير علامات الطلبة وفق التدريج الثنائي التقليدي (1,0) بعد التصحيح لأثر التخمين بمعادلة لتل (little) التي تقوم على أساس نظري مدعوم من أدبيات القياس والتقويم وتجمع بين التخمين الذكي والتخمين العشوائي.

X3: متغير علامات الطلبة وفق التدريج الثنائي التقليدي (1,0) بعد التصحيح لأثر التخمين بالمعادلة التقليدية التي تفترض بأن التخمين يتم عشوائيا، ولا تتعامل مع التخمين الذكي أو المعرفة الجزئية.

X4: متغير علامات الطلبة المحسوبة وفق التدريج السباعي (من -3 إلى 3، والمحول خطيا من -1 إلى 1 بالقسمة على 3)

X5: متغير علامات الطلبة المحسوبة وفق التدريج الرباعي من 0 إلى 3 بعد أن تم التعامل مع جميع العلامات السالبة على أنها تساوي صفر ، ثم تحويلها خطيا بالقسمة على ثلاث ليصبح التدريج من 0 إلى 1)

X6: علامات الطلبة على متغير المخاطرة، والمحسوبة على أساس مجموع العلامات بالاتجاه السالب من التدريج، وتتراوح من صفر إلى -3 على مستوى الفقرة الواحدة، ثم جمع القيم السالبة وتحويلها إلى قيم مطلقة لأنها لم تعد تحمل إشارة سالب أو المعنى السالب على متغير المخاطرة، والتعامل مع المتوسط كعلامة قطع في تحويل متغير المخاطرة من متصل إلى منفصل ثنائي (Dichotomized)

وللإجابة عن أسئلة الدراسة، وفحص فرضياتها الخاصة بعلامات الطلبة أو خصائص الفقرات، فقد تم استخدام الأساليب الإحصائية الآتية حسب طبيعة السؤال أو الفرضية وعلى النحو الآتي:

- حساب متغير المخاطرة وفق التعريف الإجرائي (مجموع العلامات السالبة على جميع الفقرات في ملف التدريج الرباعي، وهو نفسه في ملف التدريج السباعي)، واعتماد المتوسط الحسابي كعلامة قطع (أي تحويله إلى متغير ثنائي)
- فحص الفرضية الصفرية للفروق بين متوسطات العلامات التي قد تعزى إلى متغير المخاطرة، والتحقق من تجانس التباين لاختيار الاحصائي المناسب في الفحص.
- حساب معاملات الصعوبة، ومعاملات التمييز حسب التداريج الثلاث، وفحص الفروق بين متوسطات هذه المعاملات باختلاف التداريج وفقا لمتغير المخاطرة.
- حساب الارتباطات الثنائية بين العلامات حسب طريقة التصحيح للكشف عن التوافق الرتبي Rank Isomorphism للعلامات وفقا لاختلاف طرق التصحيح.
- فحص الفروق في متوسطات العلامات للكشف عن التوافق القيمي Value Isomorphism باختلاف طرق التصحيح.
- حساب معاملات الارتباط بين متغير المخاطرة ومتغير المعرفة وفقا لكل طريقة من طرق التصحيح في تقدير العلامات.
- حساب معاملات الثبات بطريقة كرونباخ ألفا للاتساق الداخلي حسب التداريج الثلاث التي لم تخضع للتحويل الخطي حسب معادلة التصحيح، وإجراء المقارنات وفق مؤشر الإشارة والتشويش signal/noise

النتائج ومناقشتها

يبين الجدول 1 الاحصائيات الوصفية للمتغيرات التي توضح طرق حساب العلامات، وفق التدريج الثنائي، والرباعي، والسباعي، والاحصائيات الوصفية للعلامات بعد التصحيح لأثر التخمين بطريقتين، بالإضافة إلى متغير المخاطرة، وفحص اعتدالية كل توزيع حسب إحصائي كولمجروف - سماينروف (K-S)، وقيمة الاحتمال (الدلالة الإحصائية) لكل منها.

الجدول 1

المتوسطات الحسابية والانحرافات المعيارية لعلامات 64 طالب حسب الطرق الثلاث للتصحيح (X1، X2، X3) وطريقتين للتصحيح لأثر التخمين (X4 ،X5) ومتغير المخاطرة (X6) بالإضافة إلى أعلى وأوطأ علامة لكل متغير ومؤشر الالتواء.

المتغير (حسب طريقة التصحيح وحساب العلامات)	المتوسط	الانحراف المعياري	اعلى علامة	أقل علامة	الالتواء	K-S	P-value
X1: متغير علامات الطلبة وفق التدريج الثنائي التقليدي (1,0)	31.3	5.1	47	17	.3	.107	.068
X2: متغير علامات الطلبة المحسوبة وفق التدريج الرباعي من 0 إلى 3	31.1	4.7	46	22	.6	.064	.200
X3: متغير علامات الطلبة المحسوبة وفق التدريج السباعي (من -3 إلى)	25.4	6.4	45	11	.5	.080	.200
X4: متغير علامات الطلبة وفق التدريج الثنائي التقليدي (1,0) بعد التصحيح لأثر التخمين بالمعادلة التقليدية	25.1	6.8	46	6	.3	.107	.068
X5: متغير علامات الطلبة وفق التدريج الثنائي التقليدي (1,0) بعد التصحيح لأثر التخمين بمعادلة لتل (little)	28.2	5.9	47	12	.3	.107	.068
X6: علامات الطلبة على متغير المخاطرة، والمحسوبة على أساس مجموع العلامات بالاتجاه السالب من التدريج.	5.63-	2.02	-11.0	- 0.67	-.6	.142	.003

وتشير النتائج في الجدول إلى الآتي:

- إن خصائص توزيع العلامات (الوسط الحسابي، والانحراف المعياري) للمتغير X3 مساوية لخصائص العلامات على المتغير X4، مما يشير إلى أن هذه الطريقة (اي ذات العلامات السالبة بالتدريج السباعي) لا تختلف عن الطريقة التقليدية التي تقوم على معاقبة الطالب وتتعامل مع التخمين على أنه عشوائي، ولا تراعي التخمين الذكي أو المعرفة الجزئية، ويبقى الغرض الذي جاءت من أجلة موضع تساؤل.
- يتضح أن X1، X2 تؤديان الغرض نفسه من حيث تقدير العلامات وفقا لنظرية الاختبار، الا أن X2 يضيف متغير المخاطرة، مما يعفي مستخدمي الاختبارات من إضافة أسئلة لا إجابة لها، فالطالب لا يتوقع وجود هذا النوع من الأسئلة، فهو لا يعرف غرضها، ثم أنها تقتطع من وقت الطالب، ولا تدخل في حساب العلامات، فهي مربكة لموقف الاختبار وملوثة لنتائجه احصائيا Statistically Contaminated وخاصة إذا جاءت أو بعضها في بداية الامتحان.
- تشير الفروق بين المتوسطات (31 مقابل 25 تقريبا) إلى أنها ليست مجرد فروق رقمية، وخاصة عند مقارنة X1، X2 من جهة، مع X3، X4 من جهة أخرى. أما X5 فجاء مختلفاً، ويمكن تفسير ذلك في إطار التشدد والليونة لأساليب التصحيح، وتقديرها لعدالة العلامات في ضوء تقديرها لطبيعة العلاقة بين متغيري المعرفة والمخاطرة، حيث جاء تقدير العلامات حسب معادلة لتل بين التقديرين (بمتوسط 28 تقريبا) السابقين. وهي معادلة ترتكز إلى أساس منطقي في التقدير للعلامات.
- يشير الوصف الاحصائي الخاص بمتغير المخاطرة X6 إلى خصوصية التوزيع لمتغير المخاطرة الذي جاء بالتواء سالب، وأن العلامات على متغير المعرفة يمكن تعديلها وفقا لعلامة المخاطرة. حيث يلاحظ أن الوسط الحسابي للعلامات السالبة = -5.63 بانحراف معياري=2.02 (أعلى علامة = 11.0- وأقل علامة 0.67 -) ولذلك تقترح الدراسة المعادلة التالية X+ =X* (Negative Score/2) في التصحيح لأثر المخاطرة؛ لأنها تعطي نفس التقدير بمعادلة لتل، فالمتوسط حسب هذه المعادلة = 28.28 وانحراف معياري5.51، وهذا مؤشر أخر يدعم التعامل مع متغير ثنائي الاتجاه.
- اما الثبات الذي تم تقديره بطريقة كرونباخ ألفا بعد تحرير العلامات من متغير المخاطرة بتحويل العلامات السالبة إلى صفر، أي بتحرير الثبات من تباين غير مرتبط بمتغير المعرفة،

مجلة اتحاد الجامعات العربية 42(1) لسنة 2022، 81-95 Doi:10.36024/1248-042-001-006

وهو متغير المخاطرة (والذي يتغير من 70. إلى 65.) مع ملاحظة أن x- Negative هي العلامات على متغير المخاطرة Risk Variable الذي يوفر تبايناً حقيقياً، ولكن ليس له صلة بمتغير المعرفة True Irrelevant Variable (التي تتوزع علاماته من صفر إلى 3- ، ثم المحولة لتصبح من صفر إلى 1.00-، للمحافظة على التدريج الأصلي في فهم العلامات على أسئلة الاختيار من متعدد).

كما يبين الجدول 2 المقارنات الثنائية والدلالات الإحصائية للفروق بين المتوسطات للعلامات الواردة في الجدول 1 بعد إجراء تحليل التباين لتصميم القياسات المتكررة، حيث كانت قيمة ولكس لمبدأ 054. باحتمال أقل من 001. ودلالة عملية عالية أيضا، حيث كان مربع إيتا (Eta square) = 946.

الجدول 2

المقارنات الثنائية باستخدام بينفيروني Benferouni للفروق بين المتوسطات للعلامات المحسوبة وفقا للطرق الخمسة في التصحيح

المتغير	X2	X3	X4	X5
X1	.24	**5.91	** 6.22	**3.11
X2		**5.67	** 5.98	**2.87
X3			.31	** -2.81
X4				** -3.11

** دالة احصائيا باحتمال اقل من 01.

X1 :متغير علامات الطلبة وفق التدريج الثنائي التقليدي (1,0) (لا يوجد تصحيح لأثر التخمين)

X2 :متغير علامات الطلبة المحسوبة وفق التدريج الرباعي من 0 إلى 3 بعد أن تم التعامل مع جميع العلامات السالبة على أنها تساوي صفر ، ثم تم تحويلها خطيا بالقسمة على ثلاث ليصبح التدريج من 0 إلى 1)

X3 :متغير علامات الطلبة المحسوبة وفق التدريج السباعي (من -3 إلى 3، والمحول خطيا من -1 إلى 1 بالقسمة على 3)

X4 :متغير علامات الطلبة وفق التدريج الثنائي التقليدي (1,0) بعد التصحيح لأثر التخمين بالمعادلة التقليدية إلى تفترض بأن التخمين يتم عشوائيا، ولا تتعامل مع التخمين الذكي أو المعرفة الجزئية

X5 :متغير علامات الطلبة وفق التدريج الثنائي التقليدي (1,0) بعد التصحيح لأثر التخمين بمعادلة لتل (little)

حيث تشير هذه النتائج إلى أن جميع الفروق دالة احصائياً باحتمال أقل من 01. باستثناء الفرق بين المتوسطين للمتغيرين (x1، x2) وللمتغيرين (x3، x4)، حيث لم يكن الفرق دالاً احصائياً، ويمكن تفسير ذلك على النحو الآتي:

أولاً: تشير النتائج المتعلقة بالمتغيرين X1، X2 إلى أن القيم السالبة عند تحويلها إلى الصفر على أساس منطقي تفرز توزيعا للعلامات مماثلا في خصائصه (المتوسط والانحراف المعياري) للتوزيع الذي توفره طريقة التصحيح القائمة على التدريج الثنائي عندما تطلب تعليمات الاختبار الإجابة عن جميع الأسئلة، مما يشير إلى أنه لا معنى للمعرفة السالبة من هذه الزاوية، ولا تضيف معلومات أكثر مما توفره الطريقة التقليدية غير القائمة على التعامل مع التخمين. ومن هنا جاء استثمار الإجابات السالبة على أنها متغيرٌ دالٌ على المخاطرة، وهو امتداد لمتغير المعرفة؛ فالتفسير المرجح للطالب الذي يحصل على صفر هو أنه لا يعرف Lack of Knowledge، أو أن لديه سوء فهم Misconception، وبالتالي فمن الطبيعي أن يفسر حصول الطالب على أقل من الصفر بأنه مخاطر، وأن أقوى تفسير محتمل للزيادة في العلامة السالبة بأنه إمعان في المخاطرة. ويشكل الصفر نقطة انقلاب بين المعرفة والمخاطرة.

ثانيا: تشير النتائج المتعلقة بالمتغيرين X3، X4 إلى أن التصحيح وفق التدريج السباعي (من -3 إلى 3، والمحول خطيا من -1 إلى 1 بالقسمة على 3) يفرز توزيعاً للعلامات مشابهاً في خصائصه (المتوسط والانحراف المعياري) لما يفرزه التصحيح وفق التدريج الثنائي التقليدي (1،0) بعد التصحيح لأثر التخمين بالمعادلة التقليدية التي تفترض بأن التخمين يتم عشوائيا، ولا تتعامل مع التخمين الذكي أو المعرفة الجزئية. وبذلك، لم يقدم التدريج السباعي الذي يتضمن العلامات السالبة مدخلا مقنعا للتعامل مع التخمين الذكي والعشوائي، فهو أسلوب متشدد، دفع باتجاه البحث عن عدة مداخل في التصحيح بالعقاب لأثر التخمين ومنها معادلة لتل، حيث يلاحظ أن المتوسط (28.2) جاء بين الفئتين من المتوسطات. وهذه إشارة ضمنية إلى أن الطريقة الأفضل في تقدير العلامات تتم على خطوتين هما: الإجابة عن جميع الأسئلة والتصحيح وفق التدريج الثنائي ثم التعديل بمعادلة لتل Little، وأن الطريقة القائمة على التدريج الثنائي توفر عدة مزايا أبرزها ضبط أثر متغير المخاطرة في تقدير ثبات الاختبار؛ مما يشير إلى أن التباين في الشق السالب هو تباين غير صادق Irrelevant

Variance. حتى لو وفر زيادة في التباين الداخل في حساب الثبات، حيث يلاحظ تغير معامل الثبات بطريقة كرونباخ من 65. ليصل إلى70. خاصة وأن أي تحسن في معاملات الثبات قبل علامة القطع 0.70 لا معنى له وفقا لمؤشر نسب فروق التباين المفسر (Signsl/Noise) (Sijtsma, 2009) وأن دلالة الفروق في المعاملات محكومة بالغرض من الاختبار بعد تحريرها من التباين الحقيقي غير المرتبط بالسمة المقاسة (Gugiu, & Gugiu, 2017). فمؤشر الثبات يتعامل مع التباين الحقيقي True Varianceبالتعريف الاحصائي للثبات (Haladyna, & Downing, 2005)، وهذا يعني أن المخاطرة تعمل كمصدر تباين مهدد للصدق Threat of Validity، ولا يشكل الارتفاع في معامل الثبات (كما هو مبين في الجدول 3) الذي يعود إلى هذا التباين ارتفاعا حقيقيا، بل مؤشرا زائفا spurious على دقة القياس. مما يسوغ التعامل مع تدريج ثنائي الاتجاه بمؤشري ثبات (65.، 52.) في هذه الدراسة، أي باعتماد الثبات المتحرر من الزيادة الزائفة في معامل الثبات، والنظر إلى المخاطرة كمتغير يمكن قياسه وتقدير ثباته، وبذلك يكون الثبات 65. أكثر واقعيةMore Authentic.

الجدول 3

معاملات الثبات (كرونباخ ألفا) المحسوب لمصفوفة العلامات على مستوى الفقرات (50 فقرة) وفق التصحيح بالطرق الثلاث (x1، x2، x3)

الاحصائي	طريقة التصحيح			المخاطرة
	X1	X2	X3	
معامل الثبات	0.63	0.65	00.7	0.52

x1: متغير علامات الطلبة وفق التدريج الثنائي التقليدي (1,0)
x2: متغير علامات الطلبة المحسوبة وفق التدريج الرباعي من 0 إلى 3، والمحول خطيا من صفر إلى 1
X3: X3متغير علامات الطلبة المحسوبة وفق التدريج السباعي (من -3 إلى 3، والمحول خطيا من -1 إلى 1 بالقسمة على 3)

أما المؤشر الأخر الذي يمكن أن يضاف إلى مؤشر الفروق بين المتوسطات، والفروق في مؤشرات الثبات وما تعكسه من دلالات، فهو مؤشر الارتباطات الذي يعكس مدى احتفاظ الطلبة بمواقعهم في توزيع العلامات سواءً في الرتبة او القيمة (Rank and Value Isomorphism)، فقد جاءت الارتباطات الخاصة بالعلامات (x1 إلى x5) كما هو مبين في الجدول 4 عالية نسبيا (لا يقل أي منها عن 70.)، كما جاءت الارتباطات بين متغير العلامات السالبة (متغير المخاطرة) عكسية وفوق المتوسطة في معظمها (Ratner, 2009)، مما يضيف مؤشراً إحصائياً على إمكانية التعامل مع متغير آخر غير مقصودIrrelevant (الذي اطلق عليه المخاطرة) إلى جانب متغير التحصيل (المشار اليه بمتغير المعرفة)

الجدول 4

معاملات الارتباط الثنائية بين المتغيرات حسب طرق التصحيح المختلفة بالإضافة إلى متغير المخاطرة

المتغيّر	x2	x3	x4	x5	المخاطرة**
x1	0.78	0.80	*1.00	*1.00	-0.66
x2		0.98	0.78	0.78	-0.62
x3			0.80	0.80	-0.72
x4				*1.00	-0.66
x5					-0.66

*معاملات ارتباط تامة لأن المتغيرين x4، x5 هما تحويل خطي على مستوى الفقرة الواحدة من المتغير الأصلي x1. وفقا لطريقة التصحيح لأثر التخمين

** جميع الارتباطات لعلامات الطلبة على جميع المتغيرات حسب طرق التصحيح مع متغير المخاطرة سالبة (متوسطة او قوية).

X1: متغير علامات الطلبة وفق التدريج الثنائي التقليدي (1,0) (لا يوجد تصحيح لأثر التخمين)

X2: متغير علامات الطلبة المحسوبة وفق التدريج الرباعي من 0 إلى 3 بعد أن تم التعامل مع جميع العلامات السالبة على أنها تساوي صفر ، ثم تم تحويلها خطيا بالقسمة على ثلاث ليصبح التدريج من 0 إلى 1)

X3: متغير علامات الطلبة المحسوبة وفق التدريج السباعي (من -3 إلى 3، والمحول خطيا من -1 إلى 1 بالقسمة على 3)

X4: متغير علامات الطلبة وفق التدريج الثنائي التقليدي (1,0) بعد التصحيح لأثر التخمين بالمعادلة التقليدية إلى تفترض بأن التخمين يتم عشوائيا، ولا تتعامل مع التخمين الذكي أو المعرفة الجزئية.

X5: متغير علامات الطلبة وفق التدريج الثنائي التقليدي (1,0) بعد التصحيح لأثر التخمين بمعادلة لتل Little

ولتحديد المؤشرات المضافة من مقارنة المعالم الإحصائية القائمة على التصنيف الثنائي للمخاطرة، فقد تم تجزئة ملف البيانات وفقا لمتغير المخاطرة، باعتماد المتوسط الحسابي كعلامة قطع،

مجلة اتحاد الجامعات العربية 42(1) لسنة 2022، 81-95 Doi:10.36024/1248-042-001-006

لتحويله من متغير متصل إلى متغير منفصل ثنائي (منخفض المخاطرة، وعالي المخاطرة) كمتغير مستقل. أما المتغيرات التابعة، فهي علامات الطلبة على كل طريقة تصحيح (x1،x2،x3)، والعلامات المصححة لأثر التخمين بالطريقتين (x4،x5) كما هو مبين في الجدول (5)، كما يبين هذا الجدول المقارنات الثنائية على كل متغير مستقل يتطلب مثل هذه المقارنة بعد اجراء التحليل المتعدد، حيث كانت قيمة ولكس لمبدا 0.611 باحتمال 001.

الجدول 5

نتائج فحص الفروق التي تعزى إلى المخاطرة بين متوسطات علامات الطلبة وفقاً لطرق التصحيح الثلاث x1،x2،x3، والمصححة لأثر التخمين x4،x5 بالمعادلة التقليدية ومعادلة لتل

طريقة التصحيح المتغير وفق	المخاطرة وفق التدريج الثنائي	المتوسط	الانحراف المعياري	العدد	قيمة F	الاحتمال	مربع ايتا
x1	منخفض	34.1	5.34	28	18.88	1.00	.233
	عالي	29.2	3.69	36			
x2	منخفض	33.1	5.25	28	10.79	.002	.148
	عالي	29.5	3.54	36			
x3	منخفض	28.8	6.58	28	17.80	1.00	.223
	عالي	22.8	4.85	36			
x4	منخفض	28.8	7.12	28	18.88	1.00	.233
	عالي	22.3	4.93	36			
x5	منخفض	31.5	6.23	28	18.88	1.00	.233
	عالي	25.7	4.31	36			

وتشير النتائج في الجدول 5 إلى أن المخاطرة متغير مؤثر في توزيع العلامات وخصائص الاختبار على مستوى الوصف الاحصائي وتحديداً الوسط الحسابي والانحراف المعياري، حيث يلاحظ أن جميع الفروق في المتوسطات دالة احصائياً وعملياً، حسب مؤشر قيمة الاحتمال ومربع إيتا.

ومن المؤشرات الداعمة للتعامل مع تدريج ثنائي الاتجاه، تلك المتعلقة بمعاملات الصعوبة ومعاملات التمييز على المتغيرات الأصلية (بدون تحويل خطي) وهي x1،x2،x3)، والمشار إليها في الجدول 6 (التدريج الثنائي، والرباعي، والسباعي)، حيث أشارت نتائج التحليل المتعدد إلى أن قيمة ولكس لمبدأ لمعاملات الصعوبة كانت 0.317 باحتمال 0.001، ولمعاملات التمييز0.790 باحتمال 0.012 كما يبين الجدول نفسه الدلالة الإحصائية للفروق (حسب قيمF)، وجميعها دالة إحصائيا بمعاملات صعوبة أعلى إحصائيا وبدلالة عملية أيضا(حسب مربع إيتا) لصالح مجموعة المخاطرة العالية. إلا أن الفروق في معاملات التمييز لم تكن بنفس العلامة؛ وهذا لا ينفي دور تأثير المخاطرة على تمييز الفقرة، حيث أن ثبات الفقرة ($s_i \ r_{ix}$) يتأثر جوهرياً بمعاملي الصعوبة والتمييز(Allen & Yen, 1979, 126).

الجدول (6)

المتوسطات الحسابية والانحر افات المعيارية لمعاملات الصعوبة ومعاملات التمييز باختلاف طرق التصحيح وفقا لعلامة المخاطرة على مستوى الفقرة

احصائي الفقرة	التدريج	المخاطرة	العدد	المتوسط	الانحراف المعياري	قيمة F	الاحتمال	مربع ايتا
الصعوبة	الثنائي (1،0)	منخفض	21	.43	.11	90.26	.000	.653
		عالي	29	.77	.13			
	الرباعي (0 إلى 3)	منخفض	21	.45	.10	90.33	.000	.653
		عالي	29	.75	.11			
	السباعي (-1 إلى +1)	منخفض	21	.26	.16	97.19	.000	.669
		عالي	29	.69	.14			
التمييز	الثنائي (1،0)	منخفض	21	.14	.13	.060	.808	.001
		عالي	29	.15	.14			
	الرباعي (0 إلى 3)	منخفض	21	.14	.13	4.217	.045	.081
		عالي	29	.22	.14			
	السباعي (-1 إلى +1)	منخفض	21	.13	.12	1.760	.191	.035
		عالي	29	.18	.13			

وتعزز النتائج في الجدول 7 العلاقة غير المباشرة بين متغير المخاطرة ومؤشر ثبات الفقرة، حيث يلاحظ قوة الارتباطات السالبة بين العلامات على متغير المخاطرة ومعاملات صعوبة الفقرات المحسوبة على العلامات المستخلصة من طرق التصحيح الثلاث، وبعلامة أقل وضوحاً بالنسبة لمعاملات التمييز، وخاصة للتدريجين الثنائي والسباعي. حيث تشير النتائج إلى تحسن معاملات تمييز الفقرات عند تحويل العلامات السالبة إلى صفر (أي التدريج الرباعي بارتباط سالب يساوي296. واحتمال 037.)، بعكس التدريج السباعي الذي يحتفظ بالعلامات السالبة.

الجدول (7)

معاملات الارتباط الثنائية بين معاملات صعوبة الفقرات حسب طرق التصحيح الثلاث ومتغير المخاطرة وفق التعريف الإجرائي المعتمد في هذه الدراسة (والدلالة الإحصائية)

خاصية الفقرة	طريقة التصحيح حسب التدريج	الرباعي	السباعي	المخاطرة
الصعوبة	الثنائي	.986(.001)	.978(.001)	-.930 (.001)
	الرباعي		.996(.001)	-.943(.001)
	السباعي			-.964(.001)
التمييز	الثنائي	.656(.001)	.640(.001)	-.035(.810)
	الرباعي		.961(.001)	-.296(.037)
	السباعي			-.169(.240)

أما المؤشر الأخير في هذا العرض للنتائج الداعمة للتعامل مع تدريج ثنائي الاتجاه فهو مؤشر الثبات، حيث تشير معاملات الثبات حسب الطرق الثلاث في التصحيح (الثنائي والرباعي والسباعي) كما هو مبين في الجدول 7، والمحسوبة على علامات متغير المعرفة لذوي المخاطرة العالية، بأنها منخفضة جدا؛ لأن جميعها أقل من الحد الأدنى المقبول 70.-65. مقارنة بقيم معاملات الثبات لذوي المخاطرة المنخفضة لطرق التصحيح الثلاث المشار إليها بمتغير عدد فئات التدريج (الثنائي، والرباعي، والسباعي) حسب مؤشر نسب فروق التباين المفسر (Signsl/Noise) (Sijtsma, 2009) مع ملاحظة الاختلاف الجوهري بين معامل الثبات الخاص بالتدريج الرباعي.

الجدول 7

معاملات الثبات (كرونباخ ألفا) حسب طرق التصحيح الثلاث (التدريج الثنائي، والرباعي، والسباعي) وفق متغير المخاطرة الثنائي

التدريج	المخاطرة		نسبة التباين الحقيقي إلى الخطأ signal/noise		التحسن في النسبة (لأقرب عدد صحيح)
	عالية	منخفضة	عالية	منخفضة	
الثنائي (1،0)	.274	.692	.38	2.25	6
الرباعي (0 إلى 3)	.392	.809	.64	4.24	7
السباعي (-1 إلى +1)	.312	.734	.45	2.76	6

الاستنتاجات والتوصيات

من العرض السابق للنتائج والمناقشات الخاصة بكل نتيجة يمكن توضيح الترابط بينها تحت مظلة السؤال العام لمشكلة البحث والمتعلق بالتعامل مع متغيري المعرفة والمخاطرة على متصل واحد وأساسه النظري للخروج باستنتاجات وتوصيات مرتكزة عليها من خلال النقاط الآتية:

- إن حصول الطالب على علامة كلية ظاهرية سالبة (X-) ممكن نظرياً، أما واقعياً فهذا غير منطقي؛ لأن مصطلح "المعرفة السالبة" لم يتم تعريفه أو الإشارة اليه في أدبيات القياس والتقويم. وهو مختلف تماما عن سوء الفهم المعرفي Misconception، مع الأخذ بالاعتبار أن العلامة صفر على الفقرة في هذه الدراسة تنتج بعد إعطاء الطالب ثلاث فرص للاختيار، بمعنى أن الصفر الافتراضي في هذه الحالة قد يعكس معرفة صفرية Zero Knowledge؛ وهذا قد يعني صعوبة تحميل العلامة السالبة معنى يحتمل تفسيراً يُفيد بأن الطالب يقدم مؤشراً على التدرج في غياب المعرفة؛ ولذلك قد تكشف إجابات الطلبة على الأسئلة نفسها عند تحويلها من متعدد إلى تكميل أو إجابة قصيرة عن تعريف إجرائي أكثر وضوحاً للعلامات السالبة، حيث أن أدبيات القياس والتقويم لم تنظر إلى العلامات السالبة بهذه الصورة؛ وهذه دعوة للباحثين والمختصين في القياس والتقويم للتفكير بمسارين:

المسار الأول: وهو تفسير قائم على فكرة الصفر المطلق، لأن الإجابة الخاطئة تقابل صفراً في عملية التصحيح، وهو صفر افتراضي (لا يعني الغياب الكلي للمعرفة)؛ وبناء عليه، يقترح الباحث أن يتم تفسير إمعان الطالب في اختياره لبدائل خاطئة لا تتضمن الإجابة الصحيحة، وحصوله على علامات سالبة، على أنه مزيد من التأكيد في كشفه عن عدم معرفته، وبالتالي اقترابه التدريجي من الصفر المطلق، وهذا يشكل مساراً من المسارات البحثية في هذا الإطار، وتحديداً في إطار نظرية الفقرة Item theory كمحاولة متواضعة للرقي بمستوى القياس النفسي-التربوي.

المسار الثاني: وهو تفسير قائم على متغير المخاطرة Risk –Taking في الجزء السالب من التدريج، حيث ينظر إليه على أنه يقع في إطار المتغيرات النفسية في المجال التربوي (Bran & Vaidis, 2019). ومن هنا جاء الطرح لفكرة التدريج المزدوج، أي متغير تربوي وهو التحصيل Achievement Scale ليأخذ الشق الموجب

مجلة اتحاد الجامعات العربية 42(1) لسنة 2022، 81-95 Doi:10.36024/1248-042-001-006

من التدريج، ومتغير نفسي وهو المخاطرة ليأخذ الشق السالب منه، ويشكل الصفر النقطة المشتركة على متصل المتغيرين. وبذلك يمكن التعامل مع التباين في العلامات السالبة على انه تباين غير صادق True Irrelevant Variance؛ بمعنى أن الشق الموجب يتوقع أن يعكس تبايناً حقيقياً Relevant للجانب المعرفي (التحصيل)، ويشترك المصدران من التباين في تقدير الثبات (وهو0.70 حسب نتائج هذه الدراسة)، حيث لا تميز معادلة القياس التقليدية بين التباين الصادق وغير الصادق في تقدير الثبات (Downing, 2002; Panahi, 2014) وهذا يعني أن الانتقال من صفر الى سالب ثلاث (0- 3-) يقدم تفسيراً مختلفاً عن التفسير في المسار الأول. بمعنى أن الإمعان في الكشف عن عدم المعرفة التي تمثلها العلامات (من صفر إلى -3 مرورا بـ -1، -2) قد يعني تدخلا لمتغير نفسي حسب نظرية باندورا Fatemeh et al., (2011)؛ ولذلك يتوقع من الباحثين اعتماد تصاميم بحثية مشابهة أو معدّلة، تأخذ الاحتمالات والمتغيرات المؤثرة في العلامات على مستوى الفقرة، وعلى مستوى العلامة الكلية بما يتناسب مع النماذج الإحصائية لنظرية الاختبار التي تقوم على العلاقة بين العلامة الحقيقية T والظاهرية X وخطأ القياس (E) (X=T+E)، ولنظرية الفقرة التي تقوم على العلاقة بين القدرة θ وخصائص الفقرة بعدد المعالم Parameters منها (c, b, a) ونماذج التصحيح (مثل الثنائية، والعلامات الجزئية Partial Credit)، ومنها النموذج ثلاثي المعلمة المعبّر عنه بمعادلة تأخذ الصورة $P(\theta) = c + (1 - c)/[1 + e^{-1.7a(\theta-b)}]$.

- بالعودة إلى نتائج التحليل السابقة المشار اليها في الجدول1، وتحديداً المتوسطات للمتغيرات x1، x2 مقابل x3، x4، والمقارنات التي تلت ذلك الجدول من حيث الطرق التي أفرزت تلك العلامات؛ تقترح الدراسة المعادلة التالية X*= X+ (Negative Score/2) في تقدير العلامة عندما يتم التعامل مع التدريج السباعي، لأنها تعطي تقديراً مساوياً تقريباً للتقديرات التي تقدمها معادلة لتل، فالمتوسط حسب هذه المعادلة يساوي 28.28 وانحراف معياري 5.51، وبدون القسمة على 2 يكون المتوسط 25.47 وانحراف معياري 6.37. ومن هنا تأتي أهمية إجراء المزيد من البحوث التي قد تدعم صدق النتائج باعتماد بيانات واقعية من مساقات متعددة، وبعدد مختلف من البدائل، وباعتماد علامات قطع cut scores مناسبة لتصنيف الطلبة في فئات عند انخراطهم ببرامج علاجية، تقوم على تدريبهم للتصرف بحكمة Testwiseness في الإجابة وفقا لتعليمات الاختبار، وتقدير الأسباب الكامنة التي تدفع الطلبة إلى المخاطرة.
- يتجنب التربويون بشكل عام والمستخدمون للاختبارات التحصيلية ومنهم أعضاء هيئة التدريس في الجامعات ولأسباب تقنية متعلقة بتعقيدات التطبيق والتصحيح والتحليل استخدام طريقة الاختيار المتدرج Subset selection؛ والتدريج المزدوجBi-directional؛ ولذلك فإن الحاجة إلى تطوير برمجيات بالتعاون بين المختصين في التكنولوجيا والبرمجة من جهة والتربويين من جهة أخرى قائمة، ويشكل هذا التطوير المتوقع إضافة نوعية في تقويم نواتج التعلم والتعليم الجامعي.
- بالعودة أيضا إلى النتائج، وتحديدا المتعلقة بتقدير معاملات الثبات في الجدول 3، توصي الدراسة بأن يتم تحرير العلامات من متغير المخاطرة (أي العلامات المحسوبة وفق التدريج الرباعي)، وذلك بتحويل العلامات السالبة إلى صفر؛ بمعنى تحرير الثبات من تباين غير مرتبط بمتغير المعرفة، وهو متغير المخاطرة. مع ملاحظة أن العلامة السالبة X- negative هي تلك العلامات التي تمثّل متغير المخاطرة، والتي توفر تبايناً حقيقياً True Irrelevant Variableليس له صلة بمتغير المعرفة، وهذا يتطلب المزيد من مساهمات الباحثين لتوسيع دائرة النتائج المتعلقة بالثبات، أو لتعميمها في أطار العلاقة بين التباينات الداخلة في مكونات معادلة نظرية القياس التقليدية.

References

Allen, M. & Yen, W. (1979). *Introduction to measurement theory*. California, Brooks/Cole Publishing Company.

Alnabhan, M. (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 Personality*, *30*(7), 645-652.

Audeh, A. S. (2014). Revision of Guilford formula to correct item difficulty for guessing in multiple choice test items. *Journal of Educational and Psychological Studies*, *8*(2), 248-257.

Ben-Simon, A., Budescu, D. V., & Nevo, B. (1997). A comparative study of measures of partial knowledge in multiple-choice tests. *Applied Psychological Measurement*, *21*(1), 65–88. https://doi.org/10.1177/0146621697211006

Blackey, R. (2009). So many choices, So Little time: strategies for understanding and taking multiple-choice exams in history. *The History Teacher*, *43*(1), 53-66.

Bran, A. & Vaidis, D. (2019). Assessing risk-taking: what to measure and how to measure it. *Journal of Risk Research*, *23*(1), 1-14. DOI: 10.1080/13669877.2019.1591489

Burton, R. F. (2002). Misinformation, partial knowledge and guessing in true ⁄false tests. *Medical Education*, *36*(9), 805–811. DOI: 10.1046/j.1365-2923.2002.01299.x

Bush, M. (2015). Reducing the need for guesswork in multiple-choice tests. *Assessment and Evaluation in Higher Education*, *40* (2), 218-231.

Chang, S.-H., Lin, P. -C., & Lin, Z. C. (2007). Measures of partial knowledge and unexpected responses in multiple-choice tests. *Educational Technology & Society*, *10*(4), 95-109.

Chiu, T. (2010). *Correction for guessing in the framework of the 3pl item response Theory*. (Unpublished Dissertation) Rutgers, The State University of New Jersey.

Claudia, A. & Christian, R. (2010). Misconceptions or missing conceptions? *Eurasia Journal of Mathematics, Science & Technology Education*, *6*(1), 3-18. DOI: 10.12973/ejmste/75223

Cronbach, L. J., & Gleser, G. C. (1964). The signal/noise ratio in the comparison of reliability coefficients. *Educational and Psychological Measurement*, *24*(3), 467–480 https://doi.org/10.1177/001316446402400303

Dodeen, H. M. (2005). Using distractors in Correcting for guessing in multiple-choice tests. *Dirasat*, *Educational Sciences*, *32*(1), 192-199.

Downing, S. M. (2002). Threats to the validity of locally developed multiple-choice tests in medical education: construct-irrelevant variance and construct underrepresentation. *Adv Health Sci Educ Theory Pract*., *7*(3), 235-241. doi: 10.1023/a:1021112514626.

Espinosa, M. P. & Gardeazabal, J. (2010). Optimal correction for guessing in multiple-choice tests. *Journal of Mathematical Psychology*, *54*, 415 - 425.

Fatemeh, S., Hassan, A., & Farhad, J. (2011). Testing Bandura's Theory in school. *Procedia - Social and Behavioral Sciences*, *12*, 426-435. DOI: 10.1016/j.sbspro.2011.02.053

Feldt, L. S. (1993). The relationship between the distribution of item difficulties and test reliability. *Applied Measurement in Education*, *6*(1), 37-48.

Frary, R. (1988). Formula scoring of multiple-choice tests (correction for guessing). *Educational Measurement: Issues and Practice*, *7*, 33-38. DOI:10.1111/J.1745-3992.1988.TB00434.X

Gugiu, C. & Gugiu, M. (2017). Determining the minimum reliability standard based on a decision criterion. *The Journal of Experimental Education*, 86(19),1-15. DOI: 10.1080/00220973.2017.1315712

Guilford, J. (1936). The determination of item difficulty when chance success is a factor. *Psychometrika*, *1*(4), 259-264.

Haladyna, T. & Downing, M. (2005). Construct irrelevant variance in high-stakes testing *Educational Measurement: Issues and Practice*, *23*(1), 17-27. DOI:10.1111/J.1745-3992.2004.TB00149.X

James, D. & William, E. (2016). The Correction for guessing. *Review of Educational Research*, *43*(2), 181-191.

Jaradat, D. & Sawaged, S. (1986). The Subset selection technique for multiple-choice tests: An empirical inquiry. *Journal of Educational Measurement*, *23*(4), 369-376.

Jelte, W. & Conor, D. (2010). Measurement invariance in confirmatory Factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement Issues and Practice*, *29*(3), 39 – 47. DOI: 10.1111/j.1745-3992.2010.00182.x

Kane, M. (2009). The Precision of measurements. *Applied Measurement in Education*, *9*(4), 355-379. https://doi.org/10.1207/s15324818ame0904_4

Klaus, D. K., & Christian, H. G. (2007). Item difficulty of multiple choice tests dependent on different item response formats – An experiment in fundamental research on psychological assessment. *Psychology Science*, *49*(4), 361-374.

Doi:10.36024/1248-042-001-006 مجلة اتحاد الجامعات العربية 42(1) لسنة 2022، 81-95

Lau, P. N., Lau, S. H., Hong, K. S., & Usop, H. (2011). Guessing, partial knowledge, and misconceptions in multiple-Choice tests. *Educational Technology & Society*, 14 (4), 99–110.

Little, E. (1962). Overestimation for Guessing in multiple choice test scoring. *Journal of Educational Research*, *2*, 245-252.

Lucy, R. B., Tracey, J. E., James, H., & Mark, T. (2009). Does correction for guessing reduce students' performance on multiple-*choice examinations? Yes? No? Sometimes? Assessment and Evaluation in Higher Education*, 34(1), 1-15.

María, P. E. & Javier, G. (2010). Optimal correction for guessing in multiple-choice tests. *Journal of Mathematical Psychology*, *54*. 415-425.

Martin, G. W., Erwin, K. T., & George W. H. (1972). The formula score -a correction for chance or a chance correction? *Personnel Psychology*, *25*, 75-93.

Mckenna, P. (2019). Multiple choice questions: answering correctly and knowing the answer. *Interactive Technology and Smart Education*, *16*(1), 59-73. https://doi.org/10.1108/itse-09-2018-0071.

Michael, K. & James, M. (1978) The Effect of guessing on item reliability under answer-until-correct scoring. *Applied Psychological Measurement*, *2*(1), 41-49.

Ndu, I. K., Ekwochi, U., DI Osuorah, C., Asinobi, I. N., Nwaneri, M. O., Uwaezuoke, S. N., Amadi, O. F., Okeke, I. B., Chinawa, J. M., & Orjioke, C.J. (2016). Negative marking and the student physician—A descriptive study of Nigerian medical schools. *Journal of Medical Education and Curricular Development*, *3*, 165–170. doi:10.4137/JMECD.S40705.

Osterlind, S. (2006). *Modern measurement: Theory, principles, and applications of mental appraisal*. Columbus: Pearson.

Panahi, A. (2014). Threats to validity: construct-irrelevant variances contributing to performance under-representation on Graduate Record Exam (GRE). *Journal of Education & Human Development*, *3*(1), 327-346.

Rajeeva L. K. (2010). On multiple choice tests and negative marking. *Current Science*, 99(8), 2010

Ratner, B. (2009). The correlation coefficient: Its values range between + 1 / − 1, or do they? *Journal of Targeting, Measurement and Analysis for Marketing*, *17*(2), 139 – 142. doi: 10.1057/jt.2009.5

Richard, F B. (2002). Misinformation, partial knowledge and guessing in true/false tests. *Medical Education*, *36*, 805–811.

Robert, B. F. (1988). Formula Scoring of multiple choice tests (correction for guessing). *Educational Measurement: Issues and Practice*,*7*(2), 33-38.

Sharma, B. & Jain, R. (2014). Right choice of a method for determination of cut-off values: A statistical tool for a diagnostic test. *Asian Journal of Medical Sciences, 5*(3), 30-34.

Sijtsma, K. (2009). The use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*,*74*(1), 107–120. DOI: 10.1007/S11336-008-9101-0

Stenlund, T., Per-Erik, L., & Eklöf, H. (2018). The successful test taker: exploring test-taking behavior profiles through cluster analysis. *European Journal of Psychology of Education*, *33*, 403–417. DOI 10.1007/s10212-017-0332-2

Thomas, P., Pinckard, R., McMahan, C., & Jones, A. C. (2006). Correcting for guessing increases validity in multiple-choice examinations in an oral and maxillofacial pathology course. *Journal of Dental Education*, *70*(4), 378-86. DOI: 10.1002/j.0022-0337.2006.70.4.tb04092.x

Vanderoost, J., Janssen, R., Eggermont, J., Callens, R.; & De Laet, T. (2018). Elimination testing with adapted scoring reduces guessing and anxiety in multiple-choice assessments, but does not increase grade average in comparison with negative marking. *PLoS One*, *13*(10):e0203931. doi: 10.1371/journal.pone.0203931.

Ypsilandis, G.S., & Mouti, A. (2019). Fairness and ethics in multiple choice (mc) scoring: An empirical study. *Journal of Language and Education*, *5*(1), 85- 97. doi: 10.17323/2411-7390-2019-5-1-85-97