مقارنة طريقتين لتصحيح تقدير القدرة من أثر التخمين باستخدام محك نسب الترجيح

إعداد

د/ شادية عبد العزيز مهتدي منتصر

مدرس علم النفس التعليمي - كلية البنات للآداب والعلوم والتربية - جامعة عين شمس

د/ حسان مخلوف خلاف مخلوف

مدرس علم النفس التربوي - كلية التربية - جامعة حلوان

مقارنة طريقتين لتصحيح تقدير القدرة من أثر التخمين باستخدام محك نسب الترجيح

شادية عبد العزيز مهتدي منتصر1، حسان مخلوف خلاف مخلوف2

1 قسم علم النفس التعليمي، كلية البنات للآداب والعلوم والتربية ، جامعة عين شمس، القاهرة، مصر.

2قسم علم النفس التربوي، كلية التربية، جامعة حلوان، حلوان، مصر.

1البريد الالكتروني: dr.shadia@yahoo.com

2 البريد الالكتروني: hassaanmakhlouf@gmail.com

المستخلص:

هدفت الدراسة إلى الكشف عن دقة تصحيح تقدير القدرة من أثر التخمين باستخدام طريقتين من طرق نموذج راش: الحذف المتتالي، والمواءمة بعد تعديلها، ومدى اختلاف هذه الدقة باختلاف مستوى كل من القدرة والتخمين. ولتحقيق هذا الهدف اقترحت الدراسة سبع طرق للتقدير الجزئي للدرجة الكلية للفرد تبعا لنسب ترجيحه لصحة بدائل الإجابة، بالإضافة إلى الطريقة التي قدمها Taylor and Gardner (1999)، وقد اعتبرت أكثر هذه الطرق الثمانية تحقيقا لثبات القياس هي المحك المرجعي للتصحيح من أثر التخمين؛ حيث تم في ضوئها حساب التقديرات المرجعية للقدرة في الدراسة الحالية. وتكونت عينة الإجابة عن أسئلة الدراسة من 109 طالبة بالفرقة الأولى علم النفس بكلية البنات جامعة عين شمس، ممن تدرسن مقرر الإحصاء النفسي الوصفي في العام الجامعي 2015-2016، بالإضافة إلى مجموعة منتقاة من مجتمع الدراسة لتدريج أداة الدراسة وتضم 227 طالبا وطالبة. وتمثلت أداة الدراسة في اختبار الأعمال الفصلية لهذا المقرر (30 مفردة، 4 بدائل، منها بديل واحد صحيح). وقد استخدم برنامج WINSTEPS لتدريج الاختبار تبعا لنموذج راش. وبعد إجراء المعالجات الإحصائية باستخدام برنامج SPSS، توصلت الدراسة للنتائج التالية: اختلفت طرق التقدير الجزئي فيما بينها، كما اختلفت عن الدرجة الخام في تقديرها لدرجة الفرد الكلية على الاختبار. وكانت أكثر طرق التقدير الجزئي تحقيقا لثبات القياس هي تقدير درجة الفرد بما يكافئ نسبة ترجيحه للبديل الصواب، بشرط أن يكون البديل الصواب هو أعلى البدائل ترجيحا وألا يكون قد اعتمد على التخمين الصريح في الاستجابة. اختلف تقدير القدرة المصحح من أثر التخمين باختلاف طريقة التصحيح؛ حيث كانت تقديرات كل من الحذف المتتالي والمواءمة للقدرة أعلى من الطريقة المرجعية، كما كانت تقديرات الحذف المتتالي للقدرة أعلى من المواءمة. اختلفت دقة تصحيح تقدير القدرة من أثر التخمين - مقارنة بالمحك المرجعي- باختلاف مستوى القدرة والتخمين الصريح والجزئي؛ حيث حققت كل من طريقتي الحذف المتتالي والمواءمة أعلى دقة للتصحيح من أثر التخمين في حالة انخفاض مستوى القدرة، وكذا في حالة انخفاض مستوى التخمين. بينما كانت الطريقتان أقل دقة في التصحيح من أثر التخمين لدى متوسطي القدرة، وكذا لدى مرتفعي التخمين. وقد تكافأت الطريقتان في دقة التصحيح من أثر التخمين في حالتي منخفضي ومرتفعي القدرة، كما تكافأتا في حالة انعدام التخمين الصريح وحالتي مستوى التخمين الجزئي المنخفض والمتوسط، إلا أن الطريقتين قد اختلفتا عن المحك المرجعي في جميع مستويات القدرة والتخمين. وترجح النتائج دقة طرق التقدير الجزئي المقترحة، والدقة النسبية للتعديل الذي أدخلته الدراسة على طريقة المواءمة، وتوصي باختباره في دراسات أخرى، مع ضرورة إجراء المزيد من الدراسات حول طرق نماذج الاستجابة للمفردة للتصحيح من أثر التخمين قبل توظيفها في الميدان التربوي.

الكلمات المفتاحية: اختبارات الاختيار من متعدد، التخمين، التصحيح من أثر التخمين، نسب الترجيح، التقدير الجزئي، طريقة المواءمة، طريقة الحذف المتتالي، نظرية الاستجابة للمفردة، نموذج راش.

Comparing Two Ability Estimate Correction for Guessing Methods Using the Correctness Probability Reference

Shadia Abd-Elaziz Mohtadi Montasir, Hassaan Makhlouf Khallaf Makhlouf

1 Department of Psychology, Women College, Ain Shams University, cairo, Eygbt.

2 Department of Psychology, Faculty of Education, Helwan University, Helwan, Eygpt.

1Email: dr.shadia@yahoo.com

2Email: hassaanmakhlouf@gmail.com

Abstract:

This study aimed at exploring the accuracy of correction for guessing using two of Rasch model methods (Jackknife and Tailored methods after modifying the last one), and its relation to the ability and guessing level. The study developed seven partial Knowledge scoring methods according to the examinee estimates of the correctness probability ratio of each item option, besides Taylor & Gardner (1999) method. The most reliable method of them was used as the reference scoring method for the corrected ability estimates. The study sample consisted of two groups, the first group (227 students) data was used to calibrate the test and calculate the reference scale, and the other group (109 students) data was used to answer the study questions. The study tool was a thirty MCQ mid-term test in descriptive statistics course. Data was analyzed using WINSTEPS and SPSS computer programs. Results: The eight scoring methods differed from each other, and differed from the conventional (0,1) method, in both of the total scores and measurement reliability; the most restricted suggested scoring method gave the lowest total scores and the highest reliability. The ability estimates of the two correction for guessing methods were higher than the reference method ($p < 0.01$), and Jackknife estimates were higher than Tailored estimates. The accuracy degree of correction for guessing was found to be a function of the ability and guessing level. The two methods were most accurate for lower levels of ability and guessing, less accurate for average ability and high guessing levels, and differed from the partial knowledge reference in all levels of ability and guessing.

Keywords: MCQ Tests, Guessing, Correction for Guessing, Partial Knowledge Scoring, Correctness Probability Estimates, Item Response Theory, Rasch Model, Jackknife Method, Tailored Method.

مقدمة الدراسة:

لماذا يُمنح الفرد درجة سؤال خمّن إجابته الصحيحة عشوائياً؟ ولماذا يعاقب الفرد الذي لم يقم بتخمين هذه الإجابة؟ تساؤلان ظلا في الذهن منذ تعرضنا لدراسة تصحيح الاختبارات الموضوعية من أثر التخمين باستخدام المعادلة التقليدية؛ فحساب الدرجة بأسلوب عد الاستجابات الصحيحة قد يخل بدقة القياس باحتساب درجة للاستجابة التي لم تعتمد على معرفة حقيقية، كما أن التصحيح بخصم نسبة ثابتة من الدرجة بناء على عدد الاستجابات الخطأ يتغاضى عن احتمالية أن يكون هذا الخطأ ناتج عن المعرفة الجزئية أو الخطأ في المفاهيم أو ضعف الحكمة الاختبارية. وتمثل دقة تقدير درجة الفرد أهمية كبيرة؛ ليس فقط في تحسين الخصائص السيكومترية للاختبار، ولكن أيضا في شعوره بعدالة التقييم.

ولأنه لا يوجد قياس بدون أخطاء، ولأن التخمين هو أحد مصادر الخطأ المؤثرة في دقة تقدير القدرة، ومع انتشار الاختبارات الدولية والمحلية التي تستخدم الأسئلة الموضوعية ولا سيما أسئلة الاختيار من متعدد، ولأهمية ما يترتب عليها من قرارات تصنيف وتشخيص المفحوصين أو قبول التحاقهم بالبرامج أو انتقالهم من مرحلة لأخرى؛ يسعى المتخصصون في مجال القياس التربوي إلى حذف الخطأ في تقدير درجة الفرد الناتج عن سلوك التخمين، فيقل الفرق بين درجته الملاحظة ودرجته الحقيقية، ويتحسن ثبات وصدق القياس، فتتحقق دقة تحديد الموقع الحقيقي له على متصل السمة موضوع القياس؛ ومن ثم تزيد الثقة في نتائج الاختبار، وفي القرارات المترتبة عليها.

وتتنوع طرق التصحيح من أثر التخمين في التوجهين الكلاسيكي والحديث. ويمكن تصنيف الطرق الكلاسيكية في الدراسات المرتبطة إلى نوعين، الأول يقوم على افتراض أن التخمين يتم بشكل عشوائي وهو طرق التصحيح السلبي، والثاني يأخذ في الاعتبار أن التخمين يمكن أن يُبنى على معرفة جزئية (مثل: أحمد غنيم، 1986، 87-95، يحيى الأحمدي، 1990، 241، أحمد عودة، 2014، 252، طه الخرشة، 2016، 615؛ Frary, 1989, 87-91; Lesage, Valcke, and Sabbe, 2013, 192-189). ويستخدم النوع الأول في بعض الاختبارات الدولية مثل اختبار الدبلومة الثانوية الأمريكية المستوى الثاني *SAT-II* (The College Board, 2017, 4).

أما طرق نماذج نظرية الاستجابة للمفردة فمنها ما يعتمد على تقدير القدرة عندما تتحقق شروط النموذج فقط (Wainer and Wright, 1980; Bond and Fox, 2015, 77)، ومنها ما يستند إلى محك التباعد بين تقدير القدرة والصعوبة في انتقاء الاستجابات الصالحة لتقدير القدرة لجميع الأفراد (Wright and Stone,

1979; Andrich and Marais, 2014). ومنها ما يقوم على إنشاء مصفوفة تدريج مرجعي للقدرة بعد حذف كل مفردة على حده (Wainer & Wright, 1980). وقد تضاربت نتائج الدراسات التي استخدمت بعض الطرق السابقة بشأن مدى الدقة النسبية لطرق التصحيح من أثر التخمين (مثل: يحيى الأحمدي، 1990؛ طه الخرشة، 2016؛ Wainer & Wright, 1980; Taylor and Gardner, 1999; Karabatsos, 2003; Lesage, et al., 2013; Andrich & Marais, 2014).

وتتحدد الفروق بين هذه الطرق بالعوامل التي ترتبط بالتخمين. حيث يرتبط التخمين بخصائص كل من الفرد والمفردة معا؛ فقد ينتج عن السمات الثابتة نسبيا للفرد ومنها عادات التخمين (سامي ملحم، 2015، 247، 249)، والنزعة للتخمين والمخاطرة (Wainer & Wright, 1980, 379)، وبخاصة إن لم تحذر تعليمات الاختبار من عقوبة التخمين بالمعادلة التقليدية للتصحيح كما بينت دراسة (صلاح الدين علام، 1991)، وارتفاع صعوبة المفردة كما اتضح من دراستيّ (شادية منتصر وحسان مخلوف، 2017؛ Paek, 2015). كما قد ينتج عن أخطاء في بناء الاختبار تسبب أخطاء غير منتظمة في تقدير الدرجة نتيجة للتشجيع على التخمين مثل غموض المفردات والتعليمات (سامي ملحم، 2015، 248؛ Bond & Fox, 2015, 79-80).

وقد يكون التشجيع على التخمين ناتجا عن نوع الأسئلة مثل أسئلة الصواب والخطأ والاختيار من متعدد (صلاح الدين علام، 2015، 139)، أو طبيعة المشتتات (Wainer & Wright, 1980, 379)، أو إدارة الاختبار. كما أن تأثير التخمين في دقة القياس يرتبط بطريقة حساب الدرجة (صلاح الدين علام، 2015، 141؛ سامي ملحم، 2015، 249؛ رينولدز وليفنجستون، 2013، 167)، وعدد بدائل الإجابة؛ حيث تتزايد أهمية التصحيح من أثر التخمين عندما يقل عددها عن خمسة، وهذا ما اتضح من دراسة يحيى الأحمدي (1990)، وما أشار إليه أحمد عودة (2014، 250).

ويستدعي ذلك مراعاة معايير جودة إعداد المفردة والاختبار ككل، وتقنين إجراءات تطبيقه، واختيار أفضل الطرق الممكنة في تصحيح تقدير القدرة من أثر التخمين؛ ولذا تهتم الدراسة الحالية بمقارنة دقة طريقتين لتصحيح تقدير القدرة من أثر التخمين في ضوء محك واقعي، ومدى اختلاف هذه الدقة باختلاف كل من القدرة والتخمين.

مشكلة الدراسة وأدبياتها:

تتم دراسة التخمين كميا باعتباره أحد مصادر خطأ القياس، فمن المنظور الكلاسيكي يعطي نموذج الدرجة الحقيقية قيمة تقديرية كلية ثابتة لمصادر أخطاء القياس –وبينها التخمين– تساوي الفرق بين الدرجة الملاحظة للفرد ودرجته الحقيقية في إحدى نماذج الاختبار (صلاح الدين علام، 2015، 131، 706)، إلا أن افتراض تساوي الخطأ المعياري للقياس لجميع الأفراد يمثل أكثر العيوب التي تؤثر في تقدير القدرة؛ لأن أداء الفرد قد لا يتسق عبر مفردات الاختبار، كما قد لا يتكافأ الأفراد في تعرضهم لمصادر أخطاء القياس وتأثرهم بها (محمد عبد الوهاب، 2019، 63)، ويمثل معامل الثبات في هذه النظرية مؤشرا لدقة القياس ومدى تحرره من الخطأ.

بينما تفترض نظرية الاستجابة للمفردة بنماذجها الرياضية المتنوعة عدم ثبات قيمة تباين هذه الأخطاء، وأن الخطأ المعياري للقياس هو دالة للتغير في قيمة تباين التقديرات المتوقعة بالنموذج، فتتيح حساب قيم الخطأ المعياري لتقدير قدرة كل فرد على حده، ويمكن أن تكون دالة معلومات الاختبار –التي تساوي مقلوب تباين هذا الخطأ عند كل مستوى قدرة– أن تكون بديلا للثبات الكلاسيكي (صلاح الدين علام، 2005، 56، 113؛ Toland , 2014, 143-144).

وترتكز هذه النماذج على معادلة لوغاريتمية لحساب احتمال الإجابة الصواب للفرد على إحدى المفردات بدلالة الفرق بين قدرته وصعوبة هذه المفردة، وتتضمن بارامترات للمفردة منها التمييز والتخمين والإهمال، فيفترض نموذج راش Rasch Model أحادي البارامتر انعدام التخمين والإهمال وتساوي معاملات التمييز لمفردات الاختبار، فيصبح التخمين تبعا للنموذج دالة للصعوبة النسبية للمفردة بالنسبة لقدرة الفرد، ويضاف بارامتر التمييز في النموذج اللوغاريتمي الثنائي، والتمييز والتخمين في النموذج الثلاثي (أحمد التقي، 2013، 18، 22-25؛ Andrich & Marais, 2014, 443 ؛ Bond & Fox, 2015, 77)، والتمييز والتخمين والإهمال في النموذج الرباعي (Reckase, 2009, 111).

ويهتم نموذج راش بأثر التخمين في تقدير قدرة الفرد أكثر من النموذج الثلاثي؛ لأن الأخير يتناول التخمين كدالة في المفردة فقط، في حين أن التخمين من الناحية الواقعية هو تفاعل بين نزوع المفردة لإثارة التخمين وميل الفرد للتخمين، فهو دالة للصعوبة النسبية للمفردة بالنسبة لقدرة الفرد، وأن وضع التخمين كأحد البارامترات يطمس التمييز بين التخمين العشوائي والمستنير (Smith, 1993, 262)، بينما نموذج راش لا يُستخدم في إنشاء التدريج النهائي لكل من القدرة والصعوبة إلا بعد حذف الحالات غير

الملائمة للنموذج من الأفراد والمفردات ومن بينها من خمن الإجابة. ولأن الواقع العملي يفرض أحيانا ضرورة الإعلان عن نتيجة كل فرد في الاختبار بما فيهم المستبعدين، مثل اختبارات التحصل الدراسي؛ جاءت ضرورة تطوير طرق لتصحيح هذه التقديرات.

ويمكن تصنيف طرق التصحيح من أثر التخمين في اختبارات الاختيار من متعدد ذات البديل الواحد الصواب، كما يلي:

أولاً: طرق تصحيح درجات اختبارات الاختيار من متعدد من أثر التخمين من المنظور الكلاسيكي: تنقسم إلى:

1- طرق التصحيح السلبي للتخمين ***Negative Marking***: تعتمد في التقليل من التخمين على إعلام الطالب بأنه سيعاقب بخصم نسبة من عدد استجاباته الخطأ من درجته الكلية إن لجأ لتخمين إجابة الأسئلة التي لا يعرف إجابتها، أو سيكافأ بإضافة نسبة من عدد هذه الأسئلة إن تركها بدون إجابة، أو بالعقاب والمكافأة معا (Lesage et al., 2013, 189)، وبالتالي قد يحصل الطالب على درجة سالبة، ومن هنا جاءت تسمية هذا الأسلوب، وتحسب الدرجة المصححة من أثر التخمين كما يلي:

أ- التصحيح بالعقاب: تستخدم فيها المعادلة التقليدية للتصحيح من أثر التخمين، وهي معادلة مشتقة من نموذج التخمين العشوائي الذي يفترض أن اختيار المفحوص لأحد البدائل يعتمد على معرفة تامة بالبديل الصحيح فيختاره، أو على جهل تام فيلجأ للتخمين العشوائي فتتساوى احتمالات اختياره لبدائل الإجابة؛ ولأن عدد التخمينات الصائبة مجهول؛ يتم طرح نسبة من عدد الأخطاء باعتبارها تخمينات صائبة، وهذه النسبة تتحدد بمقلوب عدد بدائل الإجابة مطروحا منه واحد صحيح (أحمد عودة، 2014، 252).

وقد ينظر إلى هذه الطريقة على أنها وسيلة تقلل من التخمين لدى الممتحنين حرصا على ألا يخصم منهم درجات، وهذا ما أيدته نتائج صلاح الدين علام (1991) من أن تعليمات الاختبار التي تتضمن تحذير الممتحنين من عقوبة التخمين- باستخدام معادلة التصحيح بالعقاب- قللت من سلوك المخاطرة ومستوى التخمين لديهم، في حين كانت التعليمات التي لا تتضمنها مشجعة على التخمين.

ب- التصحيح من أثر التخمين بالإثابة: يضاف لدرجة المفحوص نسبة من عدد المفردات التي يتركها دون إجابة إن كان يجهل إجابتها، وهذه النسبة تتحدد بمقلوب عدد بدائل الإجابة (يحيى الأحمدي، 1990، 241).

ج- التصحيح من أثر التخمين بالعقاب والإثابة معا: يعاب على أسلوب الإثابة بأنه يبالغ في تقدير قدرة المفحوص، بعكس أسلوب العقاب الذي يعاب عليه بأنه يبخس قدرة المفحوص؛ ولذلك جاء التفكير في التصحيح باستخدام كل من العقاب والإثابة؛ بحذف نسبة من الاستجابات الخطأ وإضافة نسبة من الاستجابات المتروكة (أحمد عودة، 2014، 252).

وينتقد محمد عبد الوهاب (2013، 174) طرق التصحيح السلبي للتخمين بأنها تعتمد على الدرجة الكلية في الاختبار كتقدير للقدرة، وهي بذلك تفترض تساوي جميع المفردات في مؤشرات الصعوبة والتمييز والتخمين، وهذا مالا يمكن تحقيقه عمليا.

ويرى الباحثان أن معادلات التصحيح من أثر التخمين بالعقاب أو الإثابة أو كليهما تفترض أن هناك نسبة من التخمين العشوائي الذي يحالف الحظ فيه الفرد فيجيب صوابا عن المفردات، وأن هذه النسبة ثابتة لجميع الأفراد بغض النظر عن خصائصهم، ولا تتوقف على صعوبة الاختبار ومفرداته، وهذا أمر غير منطقي. كما أن هناك تناقض صريح بين منطق العقوبة الذي يحذف الدرجات التي يفترض أن الفرد حصل عليها عن طريق التخمين، ومنطق الإثابة بإضافة درجات يفترض أن المفحوص كان بإمكانه الحصول عليها لو أنه خمن إجابة المفردات المتروكة، الأمر الذي يدعو إلى التساؤل عن مدى دقة هذه الطرق. هذا بالإضافة إلى أن طريقة حساب نسبة التخمين العشوائي للإجابة الصواب تفترض تساوي احتمالات اختيار البدائل المختلفة كإجابة عن المفردة، أي أن الاختيار يأتي نتيجة لعدم المعرفة المطلقة، وهذا الافتراض لم يعتد بالاحتمالات الأخرى للإجابة الخطأ مثل المعرفة الجزئية غير الكافية وخطأ المفاهيم وخطأ تسجيل الاستجابة.

وتؤكد نتائج دراسة شادية منتصر وحسان مخلوف (2017) عدم منطقية تعميم افتراض عشوائية التخمين لجميع الأفراد وفي جميع المفردات، حيث ارتفعت نسبة الاستجابات الناجحة من إجمالي الاستجابات المخمنة بارتفاع مستوى القدرة، كما تباينت نسب ترجيح البدائل الأربعة في كل مفردة من المفردات المخمنة في أغلب الحالات، بل أن بعض هذه البدائل قد استبعدت (نسبة ترجيحها صفرا). ولذلك كانت الحاجة إلى تطوير طرق أخرى للتصحيح من أثر التخمين ومنها طرق التقدير الجزئي، والطرق المستندة إلى نماذج الاستجابة للمفردة.

2- طرق التقدير الجزئي ***Partial Credit***: تعتمد في التقليل من أثر التخمين على إعلام المفحوص بأنه سيمنح جزءا من درجة المفردة إن كان غير متأكد من الإجابة الصواب وكانت لديه معرفة جزئية تمكنه من استبعاد بعض البدائل الخطأ، أو اختيار أكثر من بديل يرى أحدها هو الصواب، وقد علل **Lesage et al. (2013, 190)**

قلة انتشار هذه الطرق لصعوبة تعليماتها، وخطورة نتائجها في بعض المجالات الأكاديمية ولا سيما الطبية؛ إذا كان الأمر متعلق بحياة مريض. ومن هذه الطرق:

أ- الاستبعاد Elimination: هنا لا يضطر المفحوص إلى تخمين الإجابة، ولكنه يستبعد البدائل التي يتأكد أنها غير صحيحة، وبذلك يستغل المعرفة الجزئية لديه ويبتعد عن التخمين (يحيى الأحمدي، 1990، 241)، أو يجب عليه التخمين إذا تمكن من استبعاد بديل واحد (Lesage et al., 2013, 192)، ويمنح المفحوص درجة تتراوح بين الدرجة العظمى للمفردة إذا استبعد جميع البدائل الخطأ، والصفر إذا لم يستطع استبعاد أي بديل خطأ؛ فتزيد درجته بزيادة عدد البدائل الخطأ التي قام باستبعادها.

وقسم Wu, De Laet, and Janssen (2018) أنماط الإجابة المتوقعة للمفحوصين على المفردة (4 بدائل) باستخدام هذه الطريقة إلى: ذوي المعرفة الكاملة ممن يستبعدون ثلاثة بدائل، وذوي المعرفة الجزئية ممن يستبعدون بديلين، أو بديلا واحدا، ومن ليس لديهم معرفة فلا يستبعدون أي بديل، ومن لديهم خطأ في المفهوم فيستبعدون البديل الصحيح، وكانت الدرجة المصححة بالاستبعاد لهم هي (1، 3/1، 9/1، 0، -3/1) على الترتيب، والمصححة بالعقاب (1، 0، -3/1) للنمط الأول والرابع والأخير على الترتيب.

وقد قاموا بمقارنة الطريقتين باستخدام بيانات محاكاة وفقا لمؤشرات القدرة ومستوى العزوف عن المخاطرة والصعوبة، واتضح أنه عندما ازداد كل من العزوف عن المخاطرة وصعوبة المفردة، تزايدت نسبة الأسئلة المتروكة بدون إجابة مع التصحيح بالعقاب، بينما قل ترك الطالب للأسئلة مع التصحيح بالاستبعاد، وظهر التأثير الكبير للعزوف عن المخاطرة في استجابات الطلاب ذوي القدرة المتوسطة عن زملائهم المرتفعين أو المنخفضين، لأن لديهم معرفة جزئية تجعلهم في شك؛ لأن احتمال حصولهم على الدرجة يتساوى مع احتمال عقابهم عن التخمين؛ ولذلك فإنهم يحصلون على درجات أكبر عند استخدام أسلوب الاستبعاد عن أسلوب التصحيح بالعقاب.

ب- الاختيار الجزئي Partial Selection: فيه يختار المفحوص أحد البدائل فقط إذا استطاع استبعاد بديل أو أكثر وتمكن من ترجيح البديل الصواب من بين البدائل الأخرى، وإلا يتركها بدون إجابة (طه الخرشة، 2016، 615)، أو يختار أكثر من بديل إذا شعر بعدم التأكد من أيها الصواب (Lesage et al., 2013, 190)، ويمنح الطالب درجة تتراوح بين الدرجة العظمى للمفردة إذا اختار البديل الصواب، والصفر إذا لم يستطع اختياره؛ فتقل درجته بزيادة عدد البدائل الخطأ التي قام باختيارها مع البديل الصواب.

واتبع كل من مروان الصمادي ويوسف السوالمة (2009)، محمد عبد الوهاب (2019) أربع طرق لتصحيح المفردة (4 بدائل، يحدد الطالب فيها صحة أو خطأ كل بديل على حده)، واحتسبت درجة الطالب وفقا لعدد استجاباته الصحيحة في هذه المفردة، فيمنح في الطريقة الأولى (1، 0)، والثانية (0، 2/1، 1)، والثالثة (0، 3/1، 3/2، 1)، والرابعة (0، 4/1، 2/1، 4/3، 1). وأسفرت نتائج المقارنة بين هذه الطرق عن أن طريقة التصحيح الأكثر مراعاة للمعرفة الجزئية هي الأكثر دقة (الأقل في قيم الخطأ المعياري) في تقدير قدرات الأفراد وصعوبة المفردات.

وقد وجد طه الخرشة (2016) أن أسلوب الاختيار الجزئي كان الأكثر دقة في تقدير قدرة الأفراد يليه أسلوب التصحيح بالعقاب ثم الطريقة العادية التصحيح. كما أظهرت نتائج بعض الدراسات أن طريقتي الاستبعاد والاختيار الجزئي تؤدي إلى ارتفاع ثبات القياس مقارنة بالدرجة الخام والتصحيح بالعقاب (Frary, 1989, 88–91; Lesage et al., 2013, 190).

ج- نسب الترجيح Correctness Probability: هنا يحدد المفحوص مستويات ثقته في صحة البدائل، فإما يطلب منه أن يرتب البدائل من حيث ترجيحه لصحتها (Frary, 1989, 88)، وإما أن يضع وزنا نسبيا لكل بديل يعبر عن درجة ترجيحه لصحة هذا البديل، بحيث يكون مجموع هذه الدرجات يساوي عشرة (أو 100%) للمفردة الواحدة (Taylor & Gardner, 1999). وبذلك تختلف نسب الترجيح عن طرق الوزن النسبي للبدائل، التي يرجحها المحكمون لصحة كل بديل في المفردة، أو التي يحسبها المصحح لكل بديل من خلال استجابات عينة التقنين (مثل: Ypsilandis and Mouti, 2019؛ شريف السعودي ومعين النصراوين، 2018؛ Doden, 2005; Frary, 1989).

وتستخدم أساليب متنوعة لحساب درجة الممتحن لكل مفردة وفقا لنسب الترجيح، فمنها ما تعالج هذه النسب بمعادلة رياضية تراعي أحد الشروط: الإجابة الصحيحة فقط، أو البدائل كلها حتى لو لم يختر البديل الصواب، أو جميع نسب البدائل وصحة الإجابة، أو الفرق بين نسبتي الترجيح الواقعية والمتوقعة (مثل: أحمد غنيم، 1986، 88–95)، ومنها المنطقية التي تمنح درجة تكافئ نسبة ترجيح البديل الصواب، أو الدرجة كاملة إن كانت أعلى نسبة ترجيح أمام البديل الصواب، وصفرا إن كانت أمام البديل الخطأ، ونصف الدرجة إذا تساوت نسبة ترجيح البديل الصحيح مع بديل آخر خطأ (مثل: Taylor & Gardner, 1999, 359)، التي أسفرت عن عدم وجود اختلاف هاتين الطريقتين والدرجة الخام غير المصححة من أثر التخمين في اختبار تحصيلي في الفيزياء (29 مفردة، 5 بدائل)، وذلك في كل من: الدرجات الكلية للطلاب، أو تكرارات اختيارهم للبدائل في السؤال الواحد، وأكدت الدراسة على أهمية أسلوب نسب الترجيح في تشخيص مدى

فهم الطلاب للمفاهيم الفيزيائية والكشف عن التخمين؛ حيث كانت نسب عدم التأكد التام شائعة بين أفراد العينة. وقد اعتمدت الدراسة الحالية على منطق هذه الطريقة في تطوير طرق التقدير الجزئي المقترحة.

بينما وجد حمزة دودين (2005) Dodeen أن الطالب متوسط القدرة هو الأكثر استفادة من استخدام طريقة وزن البدائل في التصحيح مقارنة بالتصحيح بالعقاب؛ لأنها تهتم أكثر بما لديه من معرفة جزئية لم تحتسب في التصحيح بالعقاب، بينما كان تأثيرها أقل لدى كل من: مرتفعي القدرة (الذين لديهم عدد قليل من الإجابات الخطأ)، ومنخفضي القدرة (الذين لا يميزون بين مستوى صحة البدائل). ووجد أيضا أن قيم الثبات بمعامل ألفا أكبر قليلا لصالح طريقة وزن البدائل. واختلفت عنها نتائج شريف السعودي ومعين النصراوين (2018) حيث وجدت فروقا بين الطريقتين لصالح طريقة العقاب في كل من الدرجة الكلية والصدق، بينما لم يختلف الثبات باختلاف طريقة التصحيح. وأظهرت نتائج Ypsilandis & Mouti (2019) أفضلية طرق الوزن النسبي للبدائل عن أسلوب عد الاستجابات الصحيحة بما يسهم في تحقيق عدالة التقييم.

وتُنتقد طريقة نسب الترجيح في أمرين، الأول هو تكلفتها العالية في إدخال البيانات وتصميم برامج الحاسوب؛ فلا تستخدم إلا مع التمويل الكبير مثل اختبارات الجيش والطب الأمريكية، والثاني هو أن البعض يعتبر أن الثقة سمة شخصية، فقد تعتمد دقة تحديد الطالب لنسب ترجيح البدائل على خصائصه الشخصية المرتبطة بموقف الامتحان (Frary, 1989, 81,88)، وسماته النفسية العامة مثل الثقة بالنفس والدقة والدافعية والانتباه.

في حين يرى أحمد غنيم (1986، 87) أن هذه الطريقة يمكن أن تعطي تقديرا أدق لقدرات المفحوصين مقارنة بطرق التصحيح السلبي؛ حيث تقلل من التخمين بدرجة كبيرة، فلا يظهر إلا إذا كانت درجة الثقة في صحة البديل منخفضة. وقد أشار Frary (1989, 87-89) إلى أن بعض البحوث أظهرت نتائجها تحسن الثبات في هذه الطريقة عنه في أسلوب عد الاستجابات الصحيحة. ويرى الباحثان أن توجيه المفحوص لتحديد نسب ترجيحه للبدائل يتيح له الفرصة للتأني في الاستجابة مما يقلل من احتمالية الخطأ الناتج عن التسرع والإهمال، ومن ثم تزيد دقة القياس.

وقدم Lesage et al. (2013) مراجعة شاملة لطرق تصحيح اختبارات الاختيار من متعدد المستخدمة في التعليم العالي منذ (1979) وحتى (2011): عد الاستجابات الصحيحة، والتصحيح السلبي، والطرق التي تعتد بالمعرفة الجزئية كالاختيار الجزئي والاستبعاد ونسب الترجيح، والصواب والخطأ المتعدد، حيث قاموا بتحليل نقاط القوة

والضعف لكل طريقة، ووجدوا أنه لا توجد طريقة تعد الأفضل بشكل موحد من حيث: قاعدة أو معادلة التصحيح، استراتيجية الإجابة المثالية، ثباتها وصدقها، سلوك التخمين لدى الطالب فيها وميله للمخاطرة، اهتمامها بالمعرفة الجزئية، ومضمونها التربوي.

ومن خلال نتائج شادية منتصر وحسان مخلوف (2017) أوصت الدراسة أنه عند الاعتماد على نسب الترجيح كمحك مرجعي، ينبغي مراعاة قيم نسبة الترجيح وعدد المفردات المستبعدة كمتغيرات لحساب الدرجة، بما يعكس المعرفة الجزئية في تقدير درجة الفرد لكل مفردة، الأمر الذي قد يزيد حساسية الاختبار للتمييز بين مستويات القدرة.

وفي محاولة للباحثيْن تحري دقة التقدير بناء على ما سبق، طبقا في الدراسة الحالية ثمان طرق للتقدير الجزئي لدرجة الفرد وفقا لنسب الترجيح، منها ما ورد ذكره في طرق التصحيح من المنظور الكلاسيكي، وأخرى طورها الباحثان، وتم اختيار أكثر هذه الطرق تحقيقا لثبات القياس كمحك مرجعي واقعي؛ للمقارنة بين طريقتين من طرق نظرية الاستجابة للمفردة لتصحيح تقدير قدرة الفرد من أثر التخمين.

فنسب الترجيح تتيح دقة تصحيح درجة الفرد من أثر التخمين؛ من خلال تقدير ما لدى كل فرد من معرفة جزئية في كل مفردة. ويتسق هذا مع توجه طرق التصحيح من أثر التخمين في نظرية الاستجابة للمفردة. والتي تهتم نماذجها بتحليل استجابة كل فرد لكل مفردة وفقا لقدرته وخصائص المفردة (Chiu, 2010, 9)، ومع ما أكدته نتائج الدراسات التي استخدمت بعض طرق التصحيح تبعا لنموذج راش (مثل: شادية منتصر وحسان مخلوف، 2017؛ Andrich & Marais, 2014) من أن للتخمين مستويات، وليس مجرد أن يظهر أو لا يظهر.

ثانيا: طرق تصحيح تقديرات القدرة من أثر التخمين في اختبارات الاختيار من متعدد تبعا لنظرية الاستجابة للمفردة ***Item Response Theory***

يقتصر العرض التالي على بعض الأساليب المستخدمة لتصحيح تقدير قدرة الفرد من أثر التخمين باستخدام نموذج راش أحادي البارامتر، وفيه يحسب احتمال الإجابة الصواب(P_{ij}) للفرد(i) على المفردة(j) بدلالة الفرق بين قدرة هذا الفرد (b_i)، وصعوبة هذه المفردة (d_j) من المعادلة العامة التالية:

$$P_{ij} = [\exp (b_i - d_j)] / [(1+\exp (b_i - d_j)] \quad \text{.....(1)}$$

(أحمد التقي، 2013، 18؛ Wright & Stone, 1979, 15).

وبذلك فإن التخمين تبعا للنموذج دالة للصعوبة النسبية للمفردة بالنسبة لقدرة الفرد (Andrich & Marais, 2014, 443)، ويعتبر عشوائيا فقط عندما تكون

المفردة شديدة الصعوبة بالنسبة للفرد (Bond & Fox, 2015, 77). ويعتمد حساب تقدير القدرة بطريقة الأرجحية القصوى Maximum Likelihood على التحليل العددي للحصول على أقل فرق ممكن بين الدرجة الخام للفرد، ومجموع احتمالات أن يجيب هذا الفرد إجابة صحيحة عن كل مفردة بالاختبار (صلاح الدين علام، 2005، 93-94؛ Wainer & Wright, 1980, 374, 388).

وتستخدم هذه الطريقة لتقدير القدرة للأفراد الملائمين للنموذج فقط بعد حذف الأفراد غير الملائمين والمفردات غير الملائمة، وذلك باستخدام إحصاءي الملاءمة التقاربي (Infit) والتباعدي (Outfit)، أي عندما تتحقق شروط النموذج ومنها انعدام التخمين؛ وبالتالي فإن تقديرات القدرة لهؤلاء الأفراد تعتبر متحررة من أثر التخمين. ويوجد اتجاه آخر لتصحيح تقدير قدرة الفرد -المحسوب بالطريقة السابقة- من أثر التخمين، ومن هذه الطرق:

1- طريقة تحليل البواقي المعيارية Standardized Residual Analysis:

اقترحها Wright & Stone (1979)، وتعتمد هذه الطريقة على محك ارتفاع تقدير صعوبة المفردة عن قدرة الفرد، حيث يتم حساب التقدير المبدئي لقدرة الفرد تبعا للنموذج بطريقة الأرجحية القصوى السابقة، ثم يجرى تحليل البواقي المعيارية وحساب إحصاء (t) والتحليل البياني اللازم لتشخيص نمط استجابته، فإذا تبين أن هذا النمط مخمن، يتم تحديد الاستجابات المخمنة بناء على قيم البواقي المعيارية وبشروط تتعلق بالفرق بين تقدير القدرة المبدئي وتقدير الصعوبة وقيمة إحصاء (t)، ويتم استبعاد هذه الاستجابات من بيانات الفرد وكأنه لم يتعرض لها في الاختبار. ثم يعاد التحليل الإحصائي لهذه البيانات تبعا لنموذج راش للحصول على تقدير القدرة لكل فرد مخمن والخطأ المعياري لها. كما يعاد تحليل البواقي المعيارية لهؤلاء الأفراد لمعرفة نمط استجابة كل فرد بعد حذف هذه الاستجابات، ويكرر هذا الإجراء إلى أن يسفر التحليل عن نمط استجابة طبيعي لكل فرد، وبذلك يتم الحصول على تقديرات القدرة المصححة من أثر التخمين. (Wright & Stone, 1979, 165-169, 188).

وقد عدل Wainer & Wright (1980) هذه الطريقة، وقاما بحذف كل المفردات التي تزيد عن التقدير المبدئي لقدرة هذا الفرد بنموذج راش على كامل الاختبار بمقدار 2 لوجيت، وكررا الخطوات السابقة حتى يصلا إلى قيم مقبولة للإحصاء (t)، أو حين يصبح عدد المفردات قليل جدا.

وقد أثبت Karabatsos (2003) كفاءة هذه الطريقة في الكشف عن الاستجابات الشاذة في بيانات محاكاة، بينما أسفرت دراسة (شادية منتصر وحسان مخلوف، 2017) عن انخفاض كفاءتها في الكشف عن الاستجابات المخمنة في بيانات حقيقية؛ حيث لم تكشف عن أغلب الاستجابات المخَّمنة التي حددها المحك المرجعي، واختلفت هذه الدقة باختلاف مستوى كل من القدرة والتخمين والصعوبة. ولذا لم تستخدم طريقة تحليل البواقي المعيارية للتصحيح من أثر التخمين في الدراسة الحالية.

2- **طريقة الحذف المتتالي Jackknife**: تعتمد هذه الطريقة على حذف كل مفردة بطريقة منفصلة وإعادة تقدير القدرات (Andrich & Marais, 2014, 433)، حيث يتم تكوين مصفوفة لقيم قدرات الأفراد المقدرة وفقا لطريقة الأرجحية القصوى العادية السابقة المقابلة لكل درجة خام، وقيم هذه التقديرات بنفس الطريقة ولكن بعد تصحيحها بإعادة حسابها بعد حذف كل مفردة بعينها على حده، ثم تستخدم المعادلات التالية لتقدير قدرة الفرد المصححة من أثر التخمين:

$$b_{v(i)}^{*} = L(\text{ability based on full test}) - (L-1)(\text{ability based on omitting item i}) \quad \dots (2)$$

$$b_{v}^{*} = L\, b_{v(r)} - ((L-1)/L) \sum(x_i\, b_{v(r-1,i)} + (1-x_i)\, b_{v(r,i)}) \quad \dots (3)$$

حيث أن: (bv*) القدرة المصححة للفردv ، (bv(i)*) قدرته المصححة على المفردة i، ((bv(r)) قدرته المقدرة بطريقة الأرجحية القصوى العادية على كامل الاختبار المقابلة للدرجة الخام (r) التي حصل عليها،(xvi) درجته الخام على هذه المفردة (0، 1)، ((bvi(r-1)) قدرته المقابلة للدرجة الخام الأقل (r-1) بنفس الطريقة للاختبار محذوفا منه هذه المفردة، (L) طول الاختبار. ويحسب الخطأ المعياري لهذا التقدير من المعادلة التالية:

$$S.E = [(\sum (b_{v(i)}^{*} - b_{v}^{*})^2) / (L(L-1))]^{(0.5)} \quad \dots(4)$$

ووفقا لهذه الإجراءات رأى الباحثان أن تسمية (الحذف المتتالي) أكثر تعبيرا عن هذه الطريقة من الترجمة الحرفية (المطواة Jackknife) للتسمية التي أطلقها (Wainer & Wright, 1980).

وهذه الطريقة في التصحيح حساسة لموقع الاستجابات الصواب والاستجابات الخطأ في (بيان الاستجابة) للفرد، وبذلك تتغلب على مشكلة تساوي تقديرات القدرة

المقابلة لدرجة خام معينة تبعا لنموذج راش بغض النظر عن نمط الاستجابة -كما يقدمها برنامج WINSTEPS المستخدم في الدراسة، إلا أنها تعطي تقديرات مبالغ فيها لقيم الخطأ المعياري، لا سيما إذا كان طول الاختبار أقل من(40) مفردة (Wainer & Wright, 1980, 375-377, 388-389). وقد استخدمها الباحثان في الدراسة الحالية كإحدى طريقتي التصحيح من أثر التخمين. وقد تم عرض مثال تفصيلي لتوضيح هذه الطريقة في التعريف الإجرائي لمصطلحات الدراسة.

3- طريقة الحذف المتتالي المعدلة

اقترحها Wainer & Wright (1980)، وتعتمد على المزج بين كل من طريقة الحذف المتتالي العادية السابقة في تقدير قدرة الفرد على كل مفردة على حده، وطريقة دالة الجيب التقريبية Andrew's Sine M-estimator التي يتم التعويض فيها بكل من: القيم الموجبة (التي نتجت عن المفردات التي أجابها إجابة صحيحة)، والقيم السالبة (التي نتجت عن المفردات التي لم يجب عنها إجابة صحيحة) على حده للحصول في النهاية على قيمتي نزعة مركزية مقربة بديلة عن المتوسط أو الوسيط، ويتم تقديرهما وزنيا للحصول على التقدير المصحح لقدرة هذا الفرد.

وقد قارن Wainer & Wright (1980) بين طريقة الأرجحية القصوى في تقدير الدرجة الخام لقدرة الفرد المقدرة بنموذج راش، وطريقة المعادلة التقليدية لتصحيح التخمين بالعقاب، وطريقة الحذف المتتالي، وطريقة الحذف المتتالي المعدلة، وطريقة تحليل البواقي المعيارية، وذلك وفقا لنموذج محاكاة لعوامل التخمين (منها: توزيع الصعوبة، وطول الاختبار، والقدرة، وعدد البدائل، والتخمين، الميل للمخاطرة). ووجدا أن طريقة الحذف المتتالي المعدلة هي الأفضل كفاءة بصفة عامة تليها مباشرة طريقة الحذف المتتالي العادية، ولكن عندما ترتفع النزعة للتخمين (90% فأكثر) تتراجع كفاءة الطريقتين قليلا مقارنة بالطرق الأخرى، خاصة مع زيادة طول الاختبار (40 مفردة فأكثر).

وبالرغم من أن درجة دقة طريقة الحذف المتتالي المعدلة بوجه عام أعلى قليلا من طريقة الحذف المتتالي، إلا أنه بالأخذ في الاعتبار سهولة إجراء التصحيح بطريقة الحذف المتتالي العادية تكون المفاضلة -كما أشار Wainer & Wright (1980)- لصالح هذه الطريقة، علاوة على أن طريقة الحذف المتتالي كانت أكثر دقة من الطريقة المعدلة أو مكافئة لها عند مستويات التخمين المنخفضة والمتوسطة ومستويات القدرة المنخفضة وطول الاختبار أقل من (40) مفردة ومدى الصعوبة (4) لوجيت. ولذا استخدم الباحثان طريقة الحذف المتتالي في الدراسة الحالية.

4- طريقة المواءمة Tailored

اقترحها Andrich & Marais (2014)، واعتمد في تصحيح تقدير قدرة الأفراد من أثر التخمين العشوائي على مدخل Waller's في حذف استجابات الفرد على المفردات التي يحتمل أن تكون مخمنة، وتسير إجراءات التحليل بنموذج راش لهذه الطريقة وفقا للخطوات الأتية:

- إجراء التحليل الأول Initial بنموذج راش لجميع استجابات الطلاب على الاختبار للحصول على التقديرات المبدئية لصعوبة المفردات وقدرة الأفراد بطريقة الأرجحية القصوى.
- حساب نسبة احتمال الإجابة الصواب للفرد عن كل مفردة.
- حذف استجابة الفرد لأية مفردة يقل احتمال إجابته لها صوابا عن (0,30) سواء كانت الإجابة صحيحة أو خطأ، حيث تعامل على أنها متروكة missing. وقد عدل Lin (2018) في أسلوب حذف الاستجابات المحتمل أن تكون مخمنة لكل مفردة يزيد تقدير صعوبتها عن تقدير قدرة الفرد بمقدار (2 لوجيت) على الأقل، وتم اتباع هذا التعديل في الدراسة الحالية.
- إجراء التحليل الثاني Tailored للملف بعد الحذف، وتشبه هذه الخطوة فكرة الاختبارات الموائمة أو التكيفية المحوسبة التي لا تقدم للفرد أية مفردة تتوقع أن تكون صعبة جدا بالنسبة لقدرته المقدرة في كل مرة، ولذلك سُمى التحليل الثاني بالموائم، ولذا أطلق الباحثان الحاليان مسمى (المواءمة) على هذه الطريقة.
- إجراء التحليل الثالث Origin-Equated للملف الأصلي لتقدير الصعوبة، بشرط ربط Anchored متوسط صعوبة مجموعة من ست مفردات الأكثر سهولة-من بين 36 مفردة بالاختبار- بمتوسط صعوبتها في التحليل الثاني.
- إجراء التحليل الرابع All-Anchored للملف الأصلي لتقدير القدرة، ويجرى التحليل بشرط ربط متوسط صعوبة جميع المفردات بمتوسط صعوبتها في التحليل الثاني؛ وبالتالي تسهل المقارنة بين الثلاثة تحليلات الأخيرة، لأن لها نفس صفر التدريج (Andrich & Marias, 2014, 435-438; 2018, 287).

وأظهرت نتائج Andrich & Marais (2014) حين طبق طريقة المواءمة في تصحيح تقدير قدرة الأفراد من أثر التخمين العشوائي بنموذج راش لاستجابات حقيقية وأيضا لاستجابات مولدة وفقا لنموذج محاكاة على اختبار رافن للمصفوفات المتتابعة المتقدم (36 مفردة، 6 و 8 بدائل)، أن الفرق في تقدير الصعوبة بين التحليلين الثاني والثالث يزداد بتزايد مستوى صعوبة المفردة؛ وهذا يؤكد الفرض القائل بأن التخمين هو مسألة درجة أو مستوى، وليس مجرد أن يكون موجودا أو غير موجود، وتوصلا أيضا إلى

أنه بالرغم من أن تقدير الفرد منخفض القدرة –من المفترض أنه الأكثر تخمينا– في التحليل الرابع زاد عن تقديره في التحليل الثاني؛ إلا أن معظم مرتفعي القدرة زاد تقديرهم بهامش أكبر، وفسروا ذلك بأن مرتفع القدرة يحل مفردات أصعب بمعدل أكبر من منخفض القدرة (حتى لو خمن بعضها صوابا)، وبالتالي فإن تزايد مقدار صعوبة المفردات الصعبة رفع تقدير قدرته.

وتوصل Lin (2018) إلى نتائج مشابهة عندما استخدم طريقة المواءمة في اختبار اتقان اللغة الإنجليزية (34 مفردة، 4 و6 بدائل) على 728 تلميذا بالصف السادس الابتدائي، ولكنه عدل في أسلوب حذف الاستجابات المحتمل أن تكون مخمنة على كل مفردة يزيد تقدير صعوبتها عن تقدير قدرة الفرد بمقدار (2 لوجيت) على الأقل؛ حيث كشفت النتائج أن الحذف أدى إلى تزايد تقدير صعوبة المفردات الصعبة، وزيادة في تقدير قدرة الأفراد ذوي الأداء المرتفع. وعلل ذلك بأن احتمال ظهور التخمين المحظوظ في المفردات الصعبة أكبر من السهلة، وبالتالي عند حذف الاستجابات المحتمل أن تكون مخمنة عشوائيا ستقل نسبة الاستجابات الصحيحة، فتزيد صعوبة المفردات الصعبة. كما أظهرت النتائج أن الحذف أدى إلى زيادة نسبة المفردات الملائمة، وزيادة دقة القياس لذوي الأداء المرتفع فقط (2,6 لوجيت فأكثر).

ويرى الباحثان أن زيادة تقديرات القدرة بعد تصحيحها من أثر التخمين يعتبر محل تساؤل؛ حيث يتناقض مع منطق تصحيح هذه التقديرات من الدرجات غير المستحقة التي يتوصل لها المفحوص دون معرفة حقيقية، ولمحاولة معالجة هذه المشكلة أدخلت الدراسة الحالية تعديلا على هذه الطريقة بتثبيت تقديرات صعوبة المفردات مع التدريج المرجعي الذي تم إنشاؤه باستخدام استجابات عينة ملائمة وغير مخمنة.

وأجرى Waterbury & DeMars (2019) دراسة محاكاة توصلت إلى وجود علاقة غير خطية بين القيم المطلقة للتحيز والدقة في تقدير الصعوبة وكل من معلم التخمين للمفردة وصعوبتها وعتبة ضبط التخمين، وذلك عند استخدام طريقة المواءمة مقارنة بالتقديرات الأولية.

وقد استخدمت دراسات أخرى مداخل مختلفة للتصحيح من أثر التخمين في إطار نظرية الاستجابة للمفردة؛ مثل وليد القفاص (El-Kaffass, 2006) الذي طور طريقة تجمع بين نموذج راش وطريقة الوزن النسبي لتقدير صعوبة كل مفردة، وتوصل إلى اختلاف تقديرات القدرة باستخدام طريقته المقترحة وطريقة الأرجحية القصوى، وذلك لذوي أنماط الاستجابة الشاذة في اختبار رافن. ومحمد عبد الوهاب (2013) الذي استخدم النموذج ثلاثي البارامتر في معالجة أثر التخمين مقارنة بطريقة التصحيح بالعقاب، ووجد

اختلافا بين معاملات الارتباط بين الإحصاءات الكلاسيكية قبل وبعد التصحيح من أثر التخمين، ومعاملات الارتباط المناظرة لها باستخدام النموذج ثلاثي البارامتر، مما يشير إلى اختلاف الطريقتين في التصحيح من أثر التخمين، ولكنه لم يختبر دقة أي منهما.

كما بينت دراسات أخرى (مثل: يحيى الأحمدي، 1990؛ صلاح الدين علام،1991؛ أحمد عودة، 2014؛ Wainer & Wright, 1980; Karabatsos, 2003; Dodeen, 2005; Lesage et al., 2013; Andrich & Marais, 2014; Lin, 2018; Waterbury & DeMars, 2019) أن الاختلافات بين طرق التصحيح في دقة التقدير دالة لبعض المتغيرات مثل توزيع الصعوبة، وطول الاختبار، ومدى قدرة الفرد وعدد البدائل، ونسبة تشجيع المفردة على التخمين، وعتبة ضبط التخمين، نسبة ميل الفرد للمخاطرة.

مما سبق، يتضح أن دقة التصحيح من أثر التخمين تختلف بوجه عام باختلاف الطريقة المستخدمة في التصحيح. ولم تتم مقارنة طريقتي الحذف المتتالي والمواءمة من قبل في حدود اطلاع الباحثين؛ ولذا تهتم الدراسة الحالية بالمقارنة بينهما في دقة تقدير القدرة المصححة من أثر التخمين. والطرق السابقة مهما بلغت دقة تقديرها لقدرة الفرد، فإن الإحصاءات وحدها تحدد الاستجابات الأكثر احتمالا أن تكون مخمنة، ولكنها لا تستطيع تحديد ما إذا كان الفرد قد خمّن فعليا أم لا (Andrich & Marais, 2014, 435)؛ ولذلك استخدم الباحثان في الدراسة الحالية محكا واقعيا للتصحيح من أثر التخمين يعتمد على نسب ترجيح الطالب لصحة بدائل المفردة.

ويلاحظ ندرة الدراسات العربية التي اهتمت بمقارنة دقة طرق تقدير القدرة المصححة من أثر هذا التخمين في نظرية الاستجابة للمفردة، كما أن كثير من الدراسات الأجنبية اعتمدت على بيانات مولدة بالمحاكاة؛ وبالتالي فإن المحك المرجعي لدقة التقدير قد اعتمد على خصائص مصفوفة البيانات المولدة.

وفي هذه المصفوفة توضع لكل فرد استجابة لكل مفردة، ولكن في البيانات الحقيقية قد يترك بعض المفحوصين المفردات الصعبة بدون إجابة (Waterbury & DeMars, 2019, 6)، كما قد تصل نسبة عدد الأفراد الذين يستخدمون التخمين العشوائي إلى 50% في بيانات المحاكاه (مثل: Holster & Lake, 2016)، وهذا لا يحدث في الحقيقة؛ ولذلك اعتمدت الدراسة الحالية على بيانات واقعية.

ومن ثم تتحدد مشكلة الدراسة الحالية في الأسئلة التالية:

1- ما مدى اختلاف الدرجة الكلية للفرد في الاختبار باختلاف طريقة التقدير الجزئي؟

2- ما أكثر طرق التقدير الجزئي للدرجة تحقيقا لثبات القياس؟
3- ما مدى دقة تقدير القدرة المصححة من أثر التخمين باستخدام كل من طريقتي الحذف المتتالي والمواءمة؟
4- ما مدى تباين دقة تقدير القدرة المصححة من أثر التخمين باستخدام طريقتي الحذف المتتالي والمواءمة باختلاف مستوى القدرة؟
5- ما مدى تباين دقة تقدير القدرة المصححة من أثر التخمين باستخدام طريقتي الحذف المتتالي والمواءمة باختلاف مستوى التخمين؟

أهداف الدراسة:

- الكشف عن أكثر طرق التقدير الجزئي المستخدمة في الدراسة دقة في حساب الدرجة اعتمادا على نسب الترجيح، بحيث يمكن استخدامها كمحك مرجعي لتصحيح الدرجة من أثر التخمين، وذلك من خلال الكشف عن كل من:
  - مدى اختلاف الدرجة الكلية للفرد في الاختبار باختلاف طريقة التقدير الجزئي.
  - أكثر طرق التقدير الجزئي للدرجة تحقيقا لثبات القياس.
- استكشاف مدى دقة طريقتي الحذف المتتالي والمواءمة في تصحيح تقديرات القدرة من أثر التخمين.
- استكشاف مدى اختلاف دقة تصحيح تقديرات القدرة من أثر التخمين باستخدام طريقتي الحذف المتتالي والمواءمة باختلاف مستوى القدرة.
- استكشاف مدى اختلاف دقة تصحيح تقديرات القدرة من أثر التخمين باستخدام طريقتي الحذف المتتالي والمواءمة باختلاف مستوى التخمين.

أهمية الدراسة:

- قد تسهم الدراسة في زيادة فهم ظاهرة التخمين في الامتحانات الموضوعية، وطرق تصحيح الدرجة من أثرها.
- قد تلفت الدراسة الانتباه إلى أهمية الاعتماد على محكات موضوعية لتقييم كفاءة طرق التصحيح من أثر التخمين.
- قد تضيف الدراسة لمجال القياس التربوي بكل من:
  - تطوير طريقة الملاءمة في التصحيح من أثر التخمين تبعا لنموذج راش.
  - تقديم طرق بديلة للتقدير الجزئي للدرجة في الاختبارات التحصيلية الموضوعية من نوع الاختيار من متعدد بما يراعي المعرفة الجزئية للممتحنين ويصحح الدرجة من أثر التخمين.

- أن تقييم طريقتي التصحيح من أثر التخمين (الحذف المتتالي والمواءمة) في ضوء محك مرجعي واقعي، ودراسة علاقة ذلك بمستوى قدرة الفرد ومستوى تخمينه، قد يسهم في تحديد الظروف المثلى لاستخدام هاتين الطريقتين، أو العمل على تعديلهما بما يحقق دقة القياس، ومن ثم دقة اتخاذ القرار.
- قد تسهم الدراسة في العمل على سد حاجة الميدان التربوي بصفة خاصة لتحديد طرق دقيقة التصحيح من أثر التخمين باستخدام نظرية الاستجابة للمفردة.
- قد تلفت الدراسة انتباه المهتمين بالقياس النفسي والتربوي القائمين على عملية التقويم التربوي إلى ضرورة المزيد من الضبط لموقف القياس بهدف تحجيم مصادر أخطاء القياس ولاسيما التخمين.
- قد تلفت الدراسة انتباه الباحثين إلى ضرورة إجراء المزيد من الدراسات العلمية لتقييم طرق التصحيح من أثر التخمين وأثرها في تحقيق دقة القياس.

التعريف الإجرائي لمصطلحات الدراسة:

نسب الترجيح: هي نسب يحددها الفرد لكل بديل من بدائل الإجابة عن المفردة من نوع الاختيار من متعدد في الدراسة الحالية، لتعبر كل منها عن نسبة احتمال أن يكون البديل المتعلق بها هو الإجابة الصواب عن هذه المفردة، بحيث يكون مجموع هذه النسب للمفردة الواحدة يساوي 100%.

التخمين: هو سلوك يتبعه الفرد عندما لا تكون لديه معرفة تامة بالإجابة الصحيحة عن المفردة، من نوع الاختيار من متعدد (أربعة بدائل في الدراسة الحالية)؛ فقد تكون لديه معرفة جزئية تمكنه من استبعاد بديل واحد أو بديلين، واختيار أحد البدائل عشوائيا من بين البدائل المتبقية، وقد لا تكون لديه أية معرفة فيكون الاختيار عشوائيا تماما من بين البدائل الأربعة. كما قد تمكنه هذه المعرفة الجزئية من ترجيح أحد البدائل بنسبة أكبر من البدائل الأخرى.

وقد تناولت الدراسة الحالية تخمين الفرد للإجابة عن المفردة بطريقين، أولهما هو **التخمين الصريح**؛ حيث يعتبر الفرد قد اعتمد على التخمين في الإجابة الصواب عن المفردة إذا كانت نسبة ترجيحه لصحة البديل الصواب الذي اختاره للإجابة مساوية لنسبة ترجيحه لصحة بديل آخر (مشتت) أو أكثر. والتخمين الصريح بهذا المعنى يعد تخمينا عشوائيا من بين بديلين أو أكثر؛ أي أنه قد يكون تخمينا عشوائيا تاما (من بين جميع بدائل الإجابة الأربعة) أو تخمينا عشوائيا جزئيا (من بين بديلين أو ثلاث بدائل، مع

استبعاد البديل الرابع أو التقليل من نسبة ترجيحه). ويعرف التخمين الصريح إجرائيا بالنسبة المئوية للمفردات المخمنة صوابا، من بين إجمالي المفردات الصواب للفرد.

أما الطريقة الثانية فتتمثل في درجة التخمين أو التخمين الجزئي الذي يتحدد من خلال قيمة نسبة ترجيح الفرد للبديل الصواب للإجابة عن المفردة؛ حيث تقل درجة التخمين كلما ارتفعت هذه النسبة. ويعتبر الفرد قد اعتمد على التخمين جزئيا في الإجابة الصواب عن المفردة إذا كانت نسبة ترجيحه لصحة البديل الصواب الذي اختاره للإجابة تقل عن (100%). ويعرف التخمين الجزئي إجرائيا بالنسبة المئوية للمفردات الصواب المخمنة جزئيا من بين إجمالي المفردات الصواب للفرد.

تقدير القدرة: قدرة الفرد هي مقدار ما لدى هذا الفرد من السمة موضع القياس، وتتحدد من احتمالات استجابته الصحيحة لكل مفردة من مفردات الاختبار، وقد تم تقدير القدرة تبعا لنموذج راش بوحدة اللوجيت باستخدام برنامج الحاسب الآلي **WINSTEPS**، مع اعتبار صفر التدريج هو صفر التدريج المرجعي لمفردات الاختبار باستخدام استجابات عينة تدريج ملائمة لأسس القياس الموضوعي. وبذلك يعبر معلم الفرد أو تقدير قدرته في الدراسة الحالية عن موقع الفرد على متصل التحصيل الدراسي لمقرر الإحصاء النفسي الوصفي مدرجا بوحدة اللوجيت تبعا لنموذج راش.

وأفراد مجتمع الدراسة الحالية هم طلاب وطالبات الجامعة الدارسين لمقرر الإحصاء النفسي الوصفي كما يتضح تفصيلا في إجراءات الدراسة.

طرق تصحيح تقدير القدرة من أثر التخمين: يقصد بالتصحيح من أثر التخمين تنقية درجة الفرد أو تقدير قدرته من الدرجات غير المستحقة التي حصل عليها عن طريق التخمين، بحيث تعبر درجته أو تقدير قدرته تعبيرا أكثر دقة عن موقعه الحقيقي على متصل القدرة موضوع القياس. ويتم ذلك باتباع مجموعة من الإجراءات يطلق عليها طريقة التصحيح من أثر التخمين. وتتمثل طريقتي تصحيح تقدير القدرة من أثر التخمين في الدراسة الحالية في طريقتي الحذف المتتالي والمواءمة، كما استخدمت طريقة التقدير الجزئي اعتمادا على نسب الترجيح كمحك مرجعي لتحديد دقة التصحيح من أثر التخمين. وفيما يلي التعريف الإجرائي لهذه لطرق:

أولا: التصحيح من أثر التخمين باستخدام طريقة الحذف المتتالي Jackknife: قدم هذه الطريقة **Wainer & Wright (1980)** لضبط أثر التخمين العشوائي على تقدير قدرة الفرد، وذلك باتباع الإجراءات واستخدام المعادلات الإحصائية اللازمة (Wainer &

(Wright, 1980, 375-377, 388-389 كما سبق عرضها، وفيما يلي مثال -من واقع بيانات حقيقية من دراسة سابقة للباحثين- لتوضيح هذه الطريقة إجرائيا:

بافتراض أن الاختبار يضم خمس مفردات تقديرات صعوبتها على الترتيب هي (-1,04، -0,51، 0,04، 0,47، 1,4) لوجيت، وكان بيان درجات أحد الأفراد على هذه المفردات هو (1، 1، 0، 0، 1) على الترتيب، أي أنه حصل على الدرجة الخام (3)، ويرجح أنه قد خمن الإجابة الصواب عن أصعب مفردة بعد أن أخفق في إجابة مفردتين أسهل منها، وكان تقدير قدرته والخطأ المعياري يساوي (0,533) و(0,986) لوجيت على الترتيب. فتبعا لطريقة الحذف المتتالي، يتم تصحيح تقدير القدرة من أثر الاستجابة الشاذة (المخمنة) على المفردة الخامسة كما يلي:

أ- حذف المفردة الأولى من الاختبار، وحساب جدول تقديرات القدرة المقابلة لكل درجة كلية محتملة على الاختبار (Score tabl) بعد حذف هذه المفردة، وتكرار هذه الخطوة بعد حذف كل مفردة من مفردات الاختبار على التوالي، ورصد هذه التقديرات في جدول كالتالي:

جدول(١) مثال توضيحي لمصفوفة الحذف المتتالي

الدرجة الخام	تقدير القدرة في الاختبار ككل (b) (التقدير الأولي)	الخطأ المعياري	تقدير القدرة في الاختبار بعد حذف المفردة (i)				
			(١)	(٢)	(٣)	(٤)	(٥)
٠	٢,٩٥٥-	١,٩١١	٢,٣٥١-	٢,٦١-	٢,٧٦٢-	٢,٨٣٢-	٢,٩٠٨-
١	١,٥١٢-	١,١٦٨	٠,٨٦٩-	١,٠٦٢-	١,٢٣٣-	١,٣٢٦-	١,٤٣٨-
٢	٠,٤١١-	٠,٩٨٠	٠,٣٤٠	٠,٢٢٤	٠,٠٦٦	٠,٠٦١-	٠,٢٥٨-
٣	٠,٥٣٣	٠,٩٨٦	١,٥٦٥	١,٥	١,٣٨٨	١,٢٥٩	٠,٩١٩
٤	١,٦٥٩	١,١٨٤	٣,٠٧٣	٣,٠٣١	٢,٩٥٤	٢,٨٥١	٢,٣٨٤
٥	٣,١٣١	١,٩٢٣					

ب- من الجدول السابق يتم حساب متوسط تقدير قدرة الفرد بعد حذف كل مفردة من مفردات الاختبار، ويساوي ((0,34+0,224+1,388+1,259-0,258) /5) أي (0,5906). وقد روعي في ذلك أن الدرجة الخام بعد حذف مفردة ما تنخفض إلى (2) إذا كان الفرد قد أجاب هذه المفردة صوابا، وبالتالي تم أخذ تقدير القدرة المقابل للدرجة الخام (3) عندما كانت الإجابة عن المفردة المحذوفة خطأ، في حين تم أخذ تقدير القدرة المقابل للدرجة الخام (2) عندما كانت الإجابة عن المفردة المحذوفة صوابا، كما يلاحظ من الخلايا المحددة بالجدول السابق.

ج- حساب تقدير القدرة المصحح بالتعويض في المعادلة التالية: تقدير القدرة المصحح = طول الاختبار×(تقدير القدرة المبدئي) - (طول الاختبار- ١)×(متوسط تقديرات القدرة المحسوبة من الاختبار محذوفا منه كل مفردة من مفرداته على التوالي)

تقدير القدرة المصحح= 5×(0,533) – 4 ×(0,5906) = 0,3026 لوجيت.

ويلاحظ أن قيمة تقدير القدرة المصحح أكبر من تقدير القدرة (−0,411) الذي يمكن الحصول عليه إذا استبدلت درجته المناظرة لإجابته الصواب للمفردة الخامسة بالدرجة صفر.

د- حساب الخطأ المعياري لتقدير القدرة المصحح بالتعويض في المعادلة(4) كما يلي:

الخطأ المعياري = الجذر التربيعي{(مجـ(تقدير قدرته المصحح على المفردة – تقدير قدرته المصحح على الاختبار))^2 / (طول الاختبار× (طول الاختبار – 1))}=1,265 لوجيت.

وهذه القيمة أكبر من قيمة الخطأ المعياري لتقدير القدرة قبل التصحيح(0,986).

ومن خلال الإجراءات السابقة لاحظ الباحثان أن طريقة الحذف المتتالي تصحح تقديرات القدرة من أثر الاستجابات الشاذة بوجه عام وليس الاستجابات المخمنة فقط.

ثانيا: التصحيح من أثر التخمين باستخدام طريقة المواءمة **Tailored**: قدم هذه الطريقة **Andrich & Marais (2014)** لتصحيح تقدير قدرة الأفراد من أثر التخمين العشوائي، وقد اعتمدت الدراسة تعديل **Lin (2018)** لمحك حذف الاستجابات التي يرجح أنها جاءت عن طريق التخمين؛ وذلك بحذف استجابة الفرد لكل مفردة يزيد تقدير صعوبتها عن تقدير قدرته بمقدار (2 لوجيت) سواء كانت الإجابة صحيحة أو خطأ، وكأن هذا الفرد لم تعرض له هذه المفردات في الاختبار، أي أن استجاباته لهذه المفردات ترصد (**Missing**). وقد تم إدخال تعديل آخر على طريقة المواءمة في الدراسة الحالية؛ حيث تم تدريج القدرة قبل وبعد حذف هذه الاستجابات مع تثبيت تقديرات صعوبة المفردات الخاصة مع التدريج المرجعي للاختبار، وقد تم الحصول على هذا التدريج المرجعي باستخدام بيانات عينة ملائمة وغير مخمنة. وبالتالي اقتصرت خطوات طريقة المواءمة على:

- التدريج المبدئي لقدرات الأفراد بناء على استجاباتهم لجميع مفردات الاختبار.
- حذف استجابات كل فرد لمفردات الاختبار التي تعتبر صعبة بالنسبة لمستوى قدرته (الصعوبة – القدرة ≤ 2 لوجيت) مع معادلة التدريج مع التدريج المرجعي للصعوبة.
- التدريج النهائي لقدرات الأفراد بناء على استجابات كل فرد للمفردات الموائمة لقدرته، مع معادلة التدريج مع التدريج المرجعي للصعوبة. ولذا أطلق الباحثان على هذا الإجراء: طريقة المواءمة.

ثالثا: التقدير المرجعي للقدرة باستخدام التقدير الجزئي وفقا لنسب الترجيح: هو تقدير القدرة المقابل للدرجة الكلية للفرد المصححة من أثر التخمين باستخدام التقدير الجزئي للدرجة بناء على قيم نسب الترجيح؛ وذلك باستخدام جدول معايرة تقديرات القدرة المقابلة للدرجة الكلية المحتملة في الاختبار بعد تدريجه باستخدام بيانات عينة ملائمة من مجتمع الدراسة (كما يتبين من عرض أداة الدراسة). وفي ضوء هذا التقدير المرجعي لقدرة الفرد، أجريت المقارنة بين طريقتي التصحيح من أثر التخمين (الحذف المتتالي والمواءمة) في الدراسة الحالية.

فقد استخدمت الدراسة ثمان طرق للتقدير الجزئي؛ اقترح Taylor & Gardner (1999) واحدة منها، واقترح الباحثان الحاليان الطرق السبع الأخرى. وتتيح هذه الطرق حساب درجة كل فرد في كل مفردة في ضوء نسب الترجيح بما يراعي معرفته الجزئية بالمحتوى الذي تتعلق به المفردة؛ ومن ثم تصحح درجة المفردة (1، صفر) من أثر التخمين. ثم تم تحديد أكثر هذه الطرق تحقيقا للثبات كمؤشر لدقة القياس؛ ومن ثم اعتمدت الدراسة على الدرجات التي حصل عليها أفراد العينة باستخدام هذه الطريقة باعتبارها الدرجات الكلية المصححة من أثر التخمين باستخدام التقدير الجزئي.

وقد تم تصنيف الطرق الثمانية للتقدير الجزئي في ثلاث مجموعات، كما يلي:

المجموعة الأولى لطرق التقدير الجزئي للدرجة: تضم ثلاث طرق، حيث تحسب درجة المفردة بمقلوب عدد البدائل التي حصلت على أعلى نسبة ترجيح على ألا يزيد عددها عن ثلاثة بدائل وأن يكون من بينها البديل الصواب للإجابة عن هذه المفردة؛ أي أن نسبة الترجيح التي قررها الفرد للبديل الصواب في هذه المفردة تحول إلى الدرجات (1) أو (0,5) أو (0,33) أو (صفر)، ثم تجمع درجات المفردات للحصول على الدرجة الكلية، وذلك وفقا لشروط كل طريقة، كما يلي:

أ- الطريقة الأولى (1، 0,5، صفر): اقترحها Taylor & Gardner (1999)، حيث:

- **الدرجة (1): تعني أن نسبة ترجيح البديل الصواب هي أعلى نسبة ترجيح بين بدائل الإجابة الأربعة.**
- **الدرجة (0,5): تعني أن نسبة ترجيح البديل الصواب هي أعلى نسبة ترجيح، لكن هناك بديل واحد آخر مكافئ له تماما في نسبة ترجيحه.**
- **وفي جميع الحالات الأخرى تحسب درجة المفردة (صفر).**

ويلاحظ أن الطريقة الأولى لم تميز بين حالة التخمين العشوائي التام عند تساوى نسب ترجيح جميع البدائل الأربعة، وحالة ترجيح ثلاث بدائل منها بالتساوي مع

استبعاد البديل الرابع أو التقليل من شأنه، ففي الحالتين حُسبت درجة المفردة صفرا. لذا طور الباحثان الطريقة الثانية.

ب- الطريقة الثانية (1، 0,5، 0,33، صفر): طورها الباحثان، حيث:

- الدرجة (1): تعني أن نسبة ترجيح البديل الصواب هي أعلى نسبة ترجيح بين بدائل الإجابة الأربعة، كما في الطريقة السابقة.
- الدرجة (0,5): تعني أن نسبة ترجيح البديل الصواب هي أعلى نسبة ترجيح، لكن هناك بديل واحد آخر مكافئ له تماما في نسبة ترجيحه. أي أن المستجيب قد تحير بين البديل الصواب وبديل آخر فقرر نفس نسبة الترجيح للبديلين، ثم خمن أحدهما كإجابة للمفردة، سواء كان هذا التخمين صوابا أم خطأ. أما البديلين الآخرين فتكون نسبة الترجيح الخاصة بهما أقل من نسبة ترجيح البديل الصواب، وقد تكون إحدى هاتين النسبتين أو كليهما صفرا؛ أي قد يكون هذا الفرد قد استبعد بديلا أو اثنين من البدائل الأربعة أو لم يستبعد أي بديل ولكنه رجح بديلين-من بينهما البديل الصواب- عن باقي البدائل. ويعني ذلك أن هذا الفرد لديه معرفة جزئية متوسطة، ولذا يعطى نصف درجة المفردة؛ حيث تكون أقصى نسبة ترجيح للبديل الصواب هي (50%).
- الدرجة (0,33): تعني أن نسبة ترجيح البديل الصواب هي أعلى نسبة ترجيح، لكن هناك بديلين آخرين مكافئان له تماما في نسبة ترجيحه. أي أن المستجيب قد تحير بين البديل الصواب وبديلين آخرين فقرر نفس نسبة الترجيح للبدائل الثلاث، ثم خمن أحدها كإجابة للمفردة، سواء كان هذا التخمين صوابا أم خطأ. أما البديل الرابع فتكون نسبة ترجيحه أقل من نسبة ترجيح البديل الصواب وقد تكون هذه النسبة صفرا؛ أي قد يكون هذا الفرد قد استبعد بديلا واحدا من البدائل الأربعة أو لم يستبعد أي بديل ولكنه رجح ثلاث بدائل-من بينهما البديل الصواب- عن باقي البدائل. ويعني ذلك أن هذا الفرد لديه معرفة جزئية -وإن كانت ضعيفة- مكنته من استبعاد مشتت واحد من المشتتات الثلاث، وقد لاحظ الباحثان أن نسب ترجيح البديل الصواب في هذه الحالة تتراوح بين (30%) و (33%)، ولذا يعطى ثلث درجة المفردة في هذه الحالة، تمييزا لقدر هذه لمعرفة الجزئية عن حالة التخمين العشوائي التام عند تساوي نسب ترجيح البدائل الأربعة.
- الدرجة (صفر): تعني أن نسبة ترجيح البديل الصواب إما أن تكون مساوية لنسب ترجيح البدائل الثلاث الأخرى، أو أقل من أحدها على الأقل، أو مساوية للصفر. ففي الحالة الأولى تتساوى نسب ترجيح جميع بدائل الإجابة، ومن ثم يكون التخمين عشوائيا تماما؛ مما يعني عدم وجود معرفة جزئية تساعد على ترجيح أي بديل لذا

تكون درجته صفرا. وفي الحالتين الثانية والثالثة يرجح الفرد واحدا أو أكثر من هذه المشتتات كإجابة عن المفردة، وقد يكون ذلك نتيجة للمعرفة الخطأ أو الخلط في المفاهيم أو التسرع، أو غير ذلك. لذا تعتبر هذه الطريقة أن هذا الفرد لا يكون لديه معرفة أو دقة كافية تساعده على ترجيح البديل الصواب على المشتتات الثلاث، ومن ثم تحسب درجته صفرا في المفردة.

ويلاحظ أن درجة المفردة (1، 0,5، 0,33، 0) تكافئ مقلوب عدد البدائل التي حصلت على أعلى نسبة ترجيح على أن يكون البديل الصواب من بينها، وينبغي الإشارة إلى أن هذه القاعدة لا تستقيم مع عدد البدائل الذي يختلف عن أربعة.

جـ- الطريقة الثالثة: يلاحظ أن كلا من الطريقتين السابقتين لم تميز بين كون الفرد قد اختار البديل الصواب للإجابة (التخمين صوابا) أم لم يخترها (التخمين خطأ)، وذلك عند احتساب الدرجة الجزئية (0,5 أو 0,33)، لذا طورت هذه الطريقة الثالثة بتعديل الطريقة الثانية بحيث تحتسب الدرجتان (0,5) أو (0,33) في حالة التخمين الصواب فقط. أما إذا كان البديل الذي وقع عليه الاختيار هو أحد المشتتات الخطأ تحول الدرجة ((0,5) أو (0,33)) إلى (صفر).

المجموعة الثانية لطرق التقدير الجزئي للدرجة: يلاحظ أن الطرق الثلاث السابقة لم تميز بين قيم نسب الترجيح التي قررها الفرد للبديل الصواب عند احتساب الدرجات (1) أو (0,5) أو (0,33)، لذا تم تطوير طرق التقدير الجزئي الثلاث التالية، حيث تحسب درجة المفردة بما يكافئ نسبة الترجيح التي قررها الفرد للبديل الصواب في هذه المفردة سواء كان قد اختار هذا البديل كإجابة للمفردة أم لا، ثم تجمع درجات المفردات للحصول على الدرجة الكلية، وذلك مع مراعاة شروط كل طريقة، كما يلي:

أ- الطريقة الرابعة: درجة المفردة تكافئ نسبة ترجيح البديل الصواب، فإذا كانت نسبة ترجيح البديل الصواب تساوي (70%)، فإن درجة المفردة تساوي (0,70). وحساب هذه الدرجة كما اقترحها Taylor & Gardner (1999) لا يأخذ في الاعتبار أن هناك حالات لا تعبر فيها نسب الترجيح عن وجود معرفة جزئية بالموضوع الذي تتعلق به المفردة، ولذا أضاف الباحثان شرط أن تكون نسبة ترجيح البديل الصواب ضمن أعلى ثلاث نسب ترجيح؛ أي أن يكون هناك بديل واحد على الأقل مستبعد أو أقل ترجيحا من البديل الصواب. وبناء على هذا الشرط فإن درجة المفردة تحسب صفرا إذا: تساوت نسب ترجيح البدائل الأربعة (25%)، أو كانت نسبة ترجيح البديل الصواب هي أقل نسبة ترجيح -حتى وإن كان هناك بديل أو اثنين من المشتتات لهما نفس نسبة الترجيح- أو تساوي صفرا.

ب- الطريقة الخامسة: تم تعديل الطريقة الرابعة بتشديد القيد على حساب درجة المفردة بما يكافئ نسبة ترجيح البديل الصواب، حيث يشترط احتساب هذه الدرجة في الطريقة الحالية أن تكون نسبة ترجيح البديل الصواب ضمن أعلى نسبتي ترجيح؛ أي أن تكون هي أعلى نسبة ترجيح –حتى وان كان هناك مشتت أو اثنين لهما نفس نسبة الترجيح- أو النسبة التالية لها مباشرة- حتى وان كان هناك مشتت واحد له نفس نسبة الترجيح؛ أي أن يكون هناك مشتتين على الأقل مستبعدين أو أقل ترجيحا من البديل الصواب في حالة اختلاف نسب ترجيح البدائل الأربعة، وغير ذلك تحتسب درجة المفردة صفرا.

جـ- الطريقة السادسة: تم تعديل الطريقة السابقة بتشديد القيد على حساب درجة المفردة بما يكافئ نسبة ترجيح البديل الصواب، حيث يشترط احتساب هذه الدرجة في الطريقة الحالية أن تكون نسبة ترجيح البديل الصواب هي أعلى نسبة ترجيح فقط–حتى وان كان هناك مشتت أو اثنين لهما نفس نسبة الترجيح- وغير ذلك تحتسب درجة المفردة صفرا.

وتقوم هذه المجموعة من طرق التقدير الجزئي على منطق أن التصحيح من أثر التخمين يتم من خلال احتساب جزء من درجة المفردة على قدر ترجيح المستجيب للبديل الصواب، وذلك استنادا إلى فلسفة أن التخمين يوجد بدرجات متفاوتة، فليست القضية هي أن هناك فرد مخمن وآخر غير مخمن.

المجموعة الثالثة لطرق التقدير الجزئي للدرجة: تركز هذه المجموعة على تنقية الدرجة الكلية في الاختبار تماما من أثر التخمين، وتضم هذه المجموعة طريقتين من طرق حساب الدرجة، هما:

أ- الطريقة السابعة ((1، صفر) المعدلة): حيث: تحسب درجة المفردة واحد صحيح إذا أجاب الفرد عنها صوابا وكانت نسبة ترجيحه للبديل الصواب هي أعلى نسبة ترجيح بين البدائل الأربعة ولم يحصل أي مشتت على نفس هذه النسبة، وبذلك تكون المفردة صواب وغير مخمنَّة. وغير ذلك تحسب الدرجة صفرا. ثم تجمع درجات المفردات للحصول على الدرجة الكلية في الاختبار بعد تصحيحها تماما من أثر التخمين بناء على نسب الترجيح التي قررها هذا الفرد. وقد تم تطوير هذه الطريقة في الدراسة الحالية كطريقة غير معقدة في إجراءاتها، ومن ثم يسهل تنفيذها إن ثبتت كفاءتها.

ب- الطريقة الثامنة (السادسة المعدلة): تم تعديل الطريقة السادسة بإضافة شرط آخر وهو أن درجة المفردة تحسب صفرا إذا لجأ الفرد للتخمين في الإجابة عن هذه

المفردة. أي أن درجة المفردة تحسب بما يكافئ نسبة ترجيح البديل الصواب، بشرط أن تكون نسبة ترجيح البديل الصواب هي أعلى نسبة ترجيح بين البدائل الأربعة ولم يحصل أي مشتت على نفس هذه النسبة، وبذلك تكون المفردة صواب وغير مخمنَّة، وغير ذلك تحتسب درجة المفردة صفرا.

وخلاصة الطرق السابقة أنه قد تم وضع قيود متدرجة في شدتها لحساب الدرجة الكلية الجزئية المصححة من أثر التخمين في هذه الطرق؛ بحيث يمكن الانتقاء من بينها بما يتلاءم ودرجة الإتقان المطلوبة للأهداف التي يقيسها الاختبار، والتي تتحدد بطبيعة القرارات المترتبة على نتائج عملية القياس ومدى حساسيتها. كما أن تحسن دقة القياس تدريجيا مع زيادة هذه القيود- والذي تسعى الدراسة لاختباره- يعتبره الباحثان مؤشراً على صدق هذا المدخل في التصحيح من أثر التخمين. ويلخص الجدول التالي طرق التقدير الجزئي الثمان السابقة:

جدول (٢) ملخص توصيف طرق التقدير الجزئي المستخدمة في الدراسة

المجموعة	طريقة تقدير الدرجة لكل مفردة	الشروط المميزة للطريقة (*)	ترتيب الطريقة
الأولى	مقلوب عدد البدائل التي حصلت على أعلى نسبة ترجيح على أن يكون من بينها البديل الصواب	عدد هذه البدائل لا يزيد عن (٢)، والدرجات المحتملة هي:(١، ٠,٥، ٠)	الأولى
		عدد هذه البدائل لا يزيد عن (٣)، والدرجات المحتملة هي: (١، ٠,٥، ٠,٣٣، ٠)	الثانية
		الطريقة الثانية (١، ٠,٥، ٠,٣٣، ٠) بشرط أن يكون التخمين صوابا إن وجد	الثالثة
الثانية	الدرجة تكافئ نسبة ترجيح البديل الصواب	ألا تكون أقل نسبة ترجيح	الرابعة
		أن تكون ضمن أعلى نسبتي ترجيح	الخامسة
		أن تكون هي أعلى نسبة ترجيح	السادسة
الثالثة	درجات الاستجابات الصواب المخمنة تخمينا صريحا تساوي صفر	الخام (١، ٠) معدلة السادسة معدلة	السابعة الثامنة

(*) درجة المفردة تساوي صفرا في حالة التخمين العشوائي التام، وعند مخالفة شروط الطريقة.

منهج الدراسة وإجراءاتها

منهج الدراسة:

استخدم المنهج الوصفي للإجابة عن أسئلة الدراسة السيكومترية الحالية. حيث استخدم المنهج الوصفي المقارن لدراسة الفروق في تقدير الدرجات وثباتها باختلاف طريقة التقدير الجزئي، بالإضافة إلى المقارنة بين طريقتي الحذف المتتالي والمواءمة

والطريقة المرجعية، ومدى اختلاف الفروق بينها باختلاف مستوى كل من القدرة والتخمين؛ ومن ثم الإجابة عن أسئلة الدراسة الخمسة. بينما استخدم المنهج الوصفي الارتباطي لدراسة الاتساق بين طرق التقدير الجزئي ودراسة الثبات، ومن ثم استكمال الإجابة عن السؤالين الأول والثاني من أسئلة الدراسة.

عينة الدراسة:

تكونت العينة الكلية من (373) طالبا وطالبة بالفرقة الأولى علم النفس (279 طالبة بكلية البنات للآداب والعلوم والتربية جامعة عين شمس شعبة الآداب انتساب، 94 طالبا وطالبة بكلية التربية جامعة حلوان)، ممن درسوا مقرر الإحصاء النفسي الوصفي في العام الجامعي (2015–2016). وقد تراوحت أعمارهم بين (18,7) و(20,3) عاما بمتوسط (19) عاما وانحراف معياري (0,8) عاما تقريبا. ومن هذه العينة الكلية تم انتقاء مجموعة التدريج المرجعي ومجموعة الإجابة عن أسئلة الدراسة، كما يلي:

مجموعة التدريج المرجعي لأداة الدراسة: تم انتقاء هذه المجموعة من أفراد العينة الكلية بشرط عدم اللجوء للتخمين(الصريح) في الإجابة عن مفردات الاختبار، وقد استثني من هذا الشرط الأفراد الذين اعتمدوا على التخمين في مفردة واحدة فقط على ألا يكون تخمينا عشوائيا تاما. وقد بلغ حجم هذه المجموعة (227) فردا، تراوحت تقديرات قدرتهم على التحصيل الدراسي، في مقرر الإحصاء الوصفي، بين (–1,88) و(3,77) لوجيت، بمتوسط قدره (0,79) لوجيت، وانحراف معياري (1,09) لوجيت.

مجموعة الإجابة عن أسئلة الدراسة: تم انتقاء هذه المجموعة من طالبات شعبة الانتساب فقط من أفراد العينة الكلية، بشرط أن تكون الطالبة قد أتمت الإجابة عن جميع مفردات الاختبار مع كتابة نسب الترجيح بطريقة صحيحة.

وقد وقع الاختيار على شعبة الانتساب نظرا لما لوحظ من انخفاض مستوى طلاب شعب الانتساب بوجه عام، بالإضافة إلى ما وجده الباحثان في دراسة سابقة (شادية منتصر وحسان مخلوف، 2017) من ارتفاع نسبة اللجوء للتخمين والاستجابات غير الجادة لأسئلة الاختيار من متعدد لدى طلاب شعبة الانتساب. ومن ثم توقع الباحثان أن تفي هذه المجموعة بمتطلبات الدراسة.

وقد بلغ حجم هذه المجموعة (109) فردا، تراوحت تقديرات قدرتهم على التحصيل الدراسي، لمقرر الإحصاء الوصفي، بين (–1,93) و(1,04) لوجيت، بمتوسط قدره (–0,34) لوجيت، وانحراف معياري (0,66) لوجيت.

أداة الدراسة:

تمثلت في اختبار تحصيلي لمقرر الإحصاء النفسي الوصفي، من إعداد الباحثين. يتكون الاختبار من (30) مفردة من نوع الاختيار من متعدد، لكل مفردة أربعة بدائل للإجابة، منها بديل واحد صحيح. وتتضمن تعليمات الإجابة عن كل مفردة ما يلي:

- **تحديد البديل الصواب للإجابة عن المفردة.**
- **كتابة نسبة ترجيح الطالب لصحة كل بديل من بدائل الإجابة، بحيث يكون مجموع النسب للبدائل الأربعة يساوي 100%.**

وتتضمن ورقة الأسئلة مثالا توضيحيا لتدريب الطالب على كيفية الإجابة قبل بدء الاختبار.

وقد تم انتقاء مفردات الاختبار من مفردات الاختبارات النهائية السابقة التي أعدها الباحثان لنفس المقرر لطلاب الفرقة الأولى علم نفس بكليتي التربية (جامعة حلوان) والبنات للآداب (جامعة عين شمس)، وذلك بحيث تتحقق في المفردات ما يلي:

1- أن تكون مناسبة لما تم تدريسه بالفعل لجميع أفراد العينة حتى وقت تطبيق الاختبار (وقت اختبار الأعمال الفصلية).
2- أن تتحقق في المفردات الملاءمة الإحصائية لأسس القياس الموضوعي كما تتمثل في نموذج راش.

وقد تم إدخال تعديلات على صياغة مفردات الاختبار، ثم تم التحقق من ملاءمتها وحساب تقديرات صعوبتها في صورتها النهائية، وذلك كما يلي:

- **تم إجراء التحليل البياني للمفردات، ولم يتم حذف أي مفردة بناء على نتائج هذا التحليل.**

- **تم إجراء التحليل الإحصائي للبيانات المأخوذة من أداء مجموعة التدريج المرجعي، وذلك لتدريج الاختبار تبعا لنموذج راش، باستخدام برنامج WINSTEPS (version 3.62.0). وقد حسبت درجة الاستجابات المتروكة صفرا. وقد تم حذف (13) فردا غير ملائم إحصائيا، وبذلك بلغ حجم عينة التدريج النهائي (214) فردا. ولم تحذف أي مفردة بناء على نتائج التحليل الإحصائي، حيث أن قيم إحصاءي الملاءمة (ZSTD) التقاربي والتباعدي للمفردات لم تتجاوز القيمة الحرجة للملاءمة (± 2,5). ومن نتائج هذا التحليل الإحصائي تم الحصول على:**
  - **الجدول المرجعي لتقديرات القدرة المقابلة لكل درجة خام كلية في الاختبار، وقد استخدم هذا الجدول لحساب التقديرات الأولية للقدرة المقابلة للدرجات الخام**

الكلية وكذلك التقديرات المرجعية المقابلة للدرجات الجزئية الكلية لمجموعة الإجابة عن أسئلة الدراسة.

- التقديرات المرجعية لصعوبة مفردات الاختبار، وقد استخدمت هذه التقديرات كقيم ثابتة تحدد صفر التدريج عند إجراء التحليل الإحصائي اللازم لتصحيح تقديرات القدرة باستخدام طريقتي الحذف المتتالي والمواءمة. وقد تراوحت قيم تقديرات الصعوبة بين (−2,30) و(1,96) لوجيت، بمتوسط قَدرُه (0,00) لوجيت، وانحراف معياري (1,03) لوجيت. ويبين جدول(3) تقديرات صعوبة مفردات الاختبار والخطأ المعياري لها مقدرا بوحدة اللوجيت:

جدول (٣) تقدير صعوبة مفردات الاختبار والخطأ المعياري لها

رقم المفردة	التقدير باللوجيت		تابع	التقدير باللوجيت	
	الصعوبة	الخطأ المعياري	رقم المفردة	الصعوبة	الخطأ المعياري
١	٠,٩٢-	٠,١٩	١٦	٠,٨٠	٠,١٥
٢	٠,٣١	٠,١٦	١٧	١,٣٦	٠,١٥
٣	٠,٠٢-	٠,١٦	١٨	٠,٢٧-	٠,١٧
٤	١,٣٤	٠,١٥	١٩	٠,٩٦	٠,١٥
٥	٠,٣٨-	٠,١٧	٢٠	٠,١٨-	٠,١٧
٦	٠,٩٤-	٠,١٩	٢١	٢,٣٠-	٠,٢٩
٧	٠,٠٦	٠,١٦	٢٢	٠,٢٤-	٠,١٧
٨	١,٩٣-	٠,٢٥	٢٣	١,٥٣-	٠,٢٢
٩	١,٣٤-	٠,٢١	٢٤	٠,٧٩-	٠,١٩
١٠	٠,٢١	٠,١٦	٢٥	٠,٣٥-	٠,١٧
١١	١,٧٠	٠,١٦	٢٦	٠,٥٢	٠,١٥
١٢	٠,٠٠	٠,١٦	٢٧	١,٢٧	٠,١٥
١٣	١,٩٦	٠,١٦	٢٨	١,٣١	٠,١٥
١٤	٠,٠٠	٠,١٦	٢٩	٠,٢١-	٠,١٧
١٥	٠,٠٨-	٠,١٦	٣٠	٠,٣٢-	٠,١٧

التحقق من الكفاءة السيكومترية للاختبار

1- صدق الاختبار: تم التحقق من صدق القياس بالطرق التالية:

- صدق البناء: بما يراعي شروط جودة المفردة من نوع الاختيار من متعدد كما وردت بأدبيات القياس والتقويم التربوي (مثل: رجاء أبو علام، 2014، 245-264؛ Haladyna & Rodriguez, 2013, 90-107؛ أمينة كاظم، شادية منتصر، رباب سيف، 2015)، والتحقق من ذلك من خلال مراجعة نص المفردات من قِبَل أعضاء هيئة التدريس المتخصصين المسئولين عن تدريس المقرر والإشراف على تدريسه في الكليتين.

- صدق التدرج: إن تدرج مفردات الاختبار الحالي معا على ميزان تدرج واحد مشترك وفقا لنموذج راش –بعد حذف الأفراد غير الملائمين، وتحقق ملاءمة جميع المفردات– يعنى تحقق شروط النموذج، ومنها شرط أحادية البعد. ويعنى ذلك تحقق صدق تدرج المفردات في قياس المتغير موضوع القياس، وكذا صدق تدرج قدرات الأفراد على متصل المتغير والذي يقوم على صدق استجاباتهم على المفردات (Loyd, 1988, 142).

- الصدق العاملي: يوفر برنامج WINSTEPS إحصاء للتحقق من أحادية البعد كبديل للتحليل العاملي، حيث تتضمن مخرجات البرنامج نسبة التباين المفسر باستخدام مفردات الاختبار، وقد أسفرت هذه النتائج عن تأكيد تحقيق الاختبار لشرط أحادية حيث فسر العامل الأول نسبة (48,8 %) من التباين، في حين كانت نسبة التباين الذي يفسره العامل الثاني (3,5%) أي أقل من (5 %).

وبتحقق كل من صدق البناء وصدق التدرج وأحادية القياس يتحقق صدق مفردات الاختبار في تعريف متغير واحد هو متغير التحصيل الدراسي لمقرر الإحصاء الوصفي.

2- **ثبات الاختبار**: اعتمد حساب ثبات الاختبار على كل من:

- معامل الثبات الحدي أو الهامشي: ويساوي (1- مربع متوسط الخطأ المعياري) (Toland, 2014, 145). وقد بلغت قيمة معامل الثبات الهامشي لتقديرات الصعوبة (0,97)، كما بلغت قيمة معامل ثبات تقديرات القدرة (0,77)، وهي قيم تدل على أن الاختبار يحقق درجة مرتفعة من الثبات.

- قيم الخطأ المعياري لتقديرات كل من صعوبات المفردات وقدرات الأفراد: تراوحت قيم الخطأ المعياري لتقديرات الصعوبة بين (0,15) و(0,29) لوجيت بمتوسط (0,18) لوجيت، كما تراوحت قيم الخطأ المعياري لتقديرات القدرة بين (0,40) و(1,03) لوجيت بمتوسط (0,48) لوجيت، وجميعها قيم منخفضة نسبيا مما يدل على استقرار تقديرات الصعوبة والقدرة ومن ثم الاطمئنان إلى ثبات القياس.

- معامل الثبات الذي يوفره برنامج WINSTEPS، وهذا المعامل مكافئ لمعامل كيودر ريتشاردسون-20 (Linacre, 2006, 193). وقد بلغت قيمتي معامل ثبات تقديرات كل من صعوبة مفردات الاختبار وقدرة أفراد عينة الدراسة (0,97) و(0,80) على الترتيب، وهي قيم تدل على أن الاختبار يحقق ثبات القياس.
- ثبات التدرج: إن تدرج مفردات الاختبار الحالي معا على ميزان تدرج واحد مشترك وفقا لنموذج راش –بعد أن ثبت ملاءمة جميع المفردات وتم حذف الحالات غير

الملائمة من الأفراد- يعنى تحقق شروط النموذج، ومنها شرط استقلالية القياس. ويعنى ذلك ثبات تقديرات كل من الصعوبة والقدرة وعدم تأثرها باختلاف أداة القياس، أو باختلاف عينة الأفراد (أمينة كاظم، 2000، 331).

3- الاتساق الداخلي للاختبار:

يتحقق الاتساق الداخلي للاختبار بتحقق كل من أحادية القياس وصدق التدرج، وبالإضافة إلى ذلك تم حساب معامل الارتباط الثنائي الأصيل بين كل مفردة والاختبار ككل، وكانت وجميعها موجبة ودالة إحصائيا، كما أن قيم معاملات الارتباط بوجه عام مقبولة، كما يتبين من جدول (4):

جدول (٤) معاملات الارتباط بين كل مفردة والاختبار ككل

رقم المفردة	معامل الارتباط	رقم المفردة	معامل الارتباط	رقم المفردة	معامل الارتباط
١	**٠,٣٤٨	١١	**٠,٣٢٥	٢١	**٠,٣٧٩
٢	**٠,٥٤٤	١٢	**٠,٤٧١	٢٢	**٠,٦٢٠
٣	**٠,٥٤٢	١٣	**٠,١٧٧	٢٣	**٠,٥٦٧
٤	**٠,٣٣٦	١٤	**٠,٥٠٧	٢٤	**٠,٥٣٩
٥	**٠,٦٣٤	١٥	**٠,٢٨٥	٢٥	**٠,٥٧٣
٦	**٠,٤١٣	١٦	**٠,٣٦٩	٢٦	**٠,٣١٦
٧	**٠,٤٨٤	١٧	**٠,٣٨٩	٢٧	**٠,٢٦٩
٨	**٠,٤٧٤	١٨	**٠,٥٤٢	٢٨	**٠,٢٠٦
٩	**٠,٥٧١	١٩	**٠,٤٨٢	٢٩	**٠,٢٧١
١٠	**٠,٤٩٣	٢٠	**٠,٣٢٤	٣٠	**٠,٣١١

(**) دال عند مستوى أقل من (٠,٠١)

أساليب المعالجة الإحصائية المستخدمة للإجابة عن أسئلة الدراسة:

- معاملات الثبات (ألفا، التجزئة النصفية).
- المقاييس الإحصائية الوصفية، ومقاييس حجم التأثير.
- تقديرات الصعوبة والقدرة، والخطأ المعياري لها.
- معامل ارتباط بيرسون والثنائي الأصيل.
- اختبار تحليل التباين داخل المجموعات (القياسات المتكررة)، واختبار Bonferroni.
- اختبار t-test، واختبار Mann-Whitney، واختبار Wilcoxon.
- اختبار Kruskal-Wallis.

نتائج الدراسة ومناقشتها:

أولا: النتائج الخاصة بتحديد طريقة التقدير الجزئي التي استخدمت في حساب المحك المرجعي لدقة التصحيح من أثر التخمين في الدراسة الحالية:

تطلب ذلك حساب درجة كل فرد في كل مفردة من مفردات الاختبار، وحساب درجته الكلية في الاختبار، وذلك باستخدام طرق التقدير الجزئي الثمان السابق توضيحها، ويتضمن الجدول التالي مقاييس النزعة المركزية والتشتت وشكل التوزيع لدرجات أفراد العينة في كل طريقة من الطرق الثمان للتقدير الجزئي، بالإضافة للدرجة الخام.

جدول (٥) الوصف الإحصائي للدرجات الكلية في كل طريقة من طرق التقدير الجزئي والدرجة الخام (ن=١٠٩)

طريقة تقدير الدرجة	المتوسط	الوسيط	الانحراف المعياري	معامل الالتواء (*)	معامل التفلطح (**)
الدرجة الخام	١٣,٠٣	١٣,٠٠	٣,٨٠	٠,١٢	-٠,٧٧
الأولي	١٢,٦١	١٢,٥٠	٣,٦٦	٠,١٣	-٠,٦٩
الثانية	١٢,٦٦	١٢,٥٠	٣,٦٦	٠,١٤	-٠,٦٤
الثالثة	١٢,٠٢	١٢,٠٠	٣,٨٠	٠,١٩	-٠,٧٥
الرابعة	١١,٤٧	١٠,٩٥	٣,٣٥	٠,٣٦	-٠,٦٦
الخامسة	١١,٣٦	١٠,٨٠	٣,٣٧	٠,٣٦	-٠,٦٩
السادسة	١٠,٦٦	٩,٨٠	٣,٥٢	٠,٣٥	-٠,٧١
السابعة	١١,٣٣	١١,٠٠	٤,٠١	٠,١٤	-٠,٦٣
الثامنة	٩,٤٤	٩,٠٠	٣,٦٩	٠,٢٦	-٠,٥٤

(*) الخطأ المعياري لمعامل الالتواء (٠,٢٣) (**) الخطأ المعياري لمعامل التفلطح (٠,٤٦)

يلاحظ من مقاييس النزعة المركزية بالجدول السابق، أن مستوى التحصيل الدراسي لعينة الدراسة في المقرر موضع الاهتمام يعد منخفضا حيث أن النهاية العظمى للدرجة في الاختبار هي (30) درجة. كما تشير قيم معاملي الالتواء والتفلطح إلى اعتدالية توزيع الدرجات الكلية التي حصل عليها أفراد العينة، في جميع طرق التقدير الجزئي، ومن ثم فقد استخدم الإحصاء البارامتري في معالجتها بهدف تحديد الطريقة المرجعية للتقدير الجزئي للدرجة من بين هذه الطرق الثمان، وذلك من خلال الإجابة عن السؤالين الأول والثاني من أسئلة الدراسة، كما يلي:

نتائج الإجابة عن السؤال الأول، ومناقشتها:

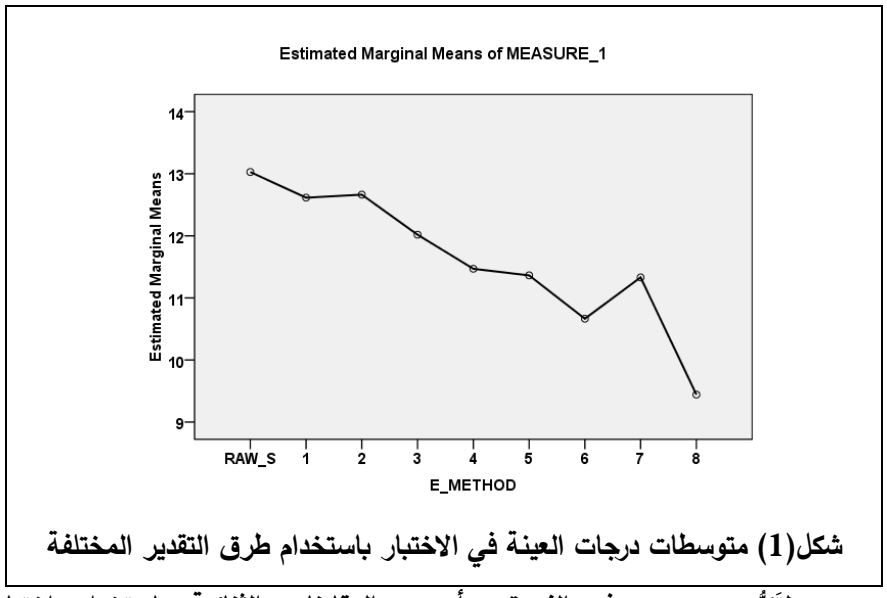
ينص السؤال الأول على: ما مدى اختلاف الدرجة الكلية للفرد في الاختبار باختلاف طريقة التقدير الجزئي؟ وللإجابة عنه تم: اختبار دلالة وحجم الفرق بين الدرجات التي حصل عليها أفراد العينة باستخدام طرق التقدير المختلفة، وذلك باستخدام تحليل التباين داخل المجموعات (القياسات المتكررة)، وقد تم الاعتماد على تصحيح Greenhouse-

Geisser لدرجات الحرية نظرا لانتهاك فرض الكروية، وكانت النتائج كما بالجدول التالي:

جدول (٦) دلالة وحجم الاختلاف بين درجات أفراد العينة في الاختبار باختلاف طريقة التقدير: الجزئي والخام (ن=١٠٩)

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	ف	مستوى الدلالة	مربع إيتا
طريقة تقدير الدرجة	١٠٩٥,٠٨	٢,٣٤	٤٦٨,٧٨	١٢٢,٥٣	٠,٠٠٠	٠,٥٣
الخطأ	٩٦٥,٢٢	٢٥٢,٢٩	٣,٨٣			

يتبين من الجدول السابق وجود فروق كبيرة دالة إحصائيا عند مستوى دلالة أقل من (0,01) في مستوى التحصيل الدراسي –كما يتمثل بمتوسط الدرجة الكلية للاختبار– باختلاف طريقة التقدير. ويؤكد ذلك النتائج الخاصة بالاختبارات المتعددة، حيث كانت قيمة اختبار Wilks' Lambda تساوي (0,12) وقيمة ومستوى دلالتها (0,000) أقل من (0,01) وقيمة مربع إيتا الجزئي الخاصة بها تساوي (0,88). ويتضح ذلك من التمثيل البياني التالي:



شكل(1) متوسطات درجات العينة في الاختبار باستخدام طرق التقدير المختلفة

ولتَبَيُّن مصدر هذه الفروق، أجريت المقارنات الثنائية باستخدام اختبار Bonferroni. ويتضمن الجدول التالي متوسط ودلالة الفروق بين كل طريقتين في تقدير الدرجة:

جدول (٧) نتائج المقارنات الثنائية بين طرق تقدير الدرجة الكلية للاختبار(ن=١٠٩)

طريقة التقدير	الأولى	الثانية	الثالثة	الرابعة	الخامسة	السادسة	السابعة	الثامنة
الدرجة الخام	٠,٤١*	٠,٣٦*	١,٠١**	١,٥٦**	١,٦٧**	٢,٣٦**	١,٧٠**	٣,٥٩**
الأولى		٠,٠٥-	٠,٦٠**	١,١٥**	١,٢٥**	١,٩٥**	١,٢٨**	٣,١٧**
الثانية			٠,٦٥**	١,١٩**	١,٣٠**	١,٩٩٩**	١,٣٣**	٣,٢٢**
الثالثة				٠,٥٥*	٠,٦٦**	١,٣٥**	٠,٦٩**	٢,٥٨**
الرابعة					٠,١١**	٠,٨٠**	٠,١٤	٢,٠٣**
الخامسة						٠,٧٠**	٠,٠٣	١,٩١**
السادسة							٠,٦٧-	١,٢٢**
السابعة								١,٨٩**

(*) الفروق دالة إحصائيا عند مستوى دلالة أقل من (٠,٠٥) (**) الفروق دالة إحصائيا عند مستوى دلالة أقل من (٠,٠١)

يتضح من الجدول السابق وشكل(1) ما يلي:

- **يختلف متوسط الدرجات الكلية المقدرة بكل طريقة من طرق التقدير الجزئي اختلافا دالا إحصائيا عن متوسط الدرجات الخام في الاختبار. حيث أدى تصحيح الدرجة من أثر التخمين باستخدام جميع طرق التقدير الجزئي إلى انخفاض درجات الأفراد في الاختبار بعد خصم الدرجات غير المستحقة التي حصل عليها الطالب نتيجة للتخمين بناء على ما قرره بنفسه من خلال نسب الترجيح. وتشير هذه النتائج إلى أهمية التصحيح من أثر التخمين لتحقيق دقة القياس.**
- **أن الطريقة الثامنة للتقدير الجزئي تعطي أقل تقدير للدرجة بعد تنقيتها من أثر التخمين، وذلك مقارنة بطريقة الدرجة الخام وبطرق التقدير الجزئي الأخرى. وقد اختلف متوسط الدرجات المتحصلة باستخدام هذه الطريقة اختلافا دالا إحصائيا عند مستوى دلالة أقل من (0,01) عن نظيره في جميع طرق التقدير الأخرى.**
- **لا تختلف الطريقة الثانية للتقدير الجزئي للدرجة –التي تراعي تخصيص ثلث درجة عندما يحصل البديل الصواب على أعلى نسبة ترجيح بالتساوي مع مشتتين اخرين– عن الطريقة الأولى التي اعتبرت الدرجة في هذه الحالة صفرا.**
- **لا تختلف الطريقة السابعة للتقدير الجزئي للدرجة –التي تخصم درجة الاستجابات المخمنة تخمينا صريحا من الدرجة الخام الكلية للفرد– عن طرق المجموعة الثانية للتقدير الجزئي التي تضم الطرق الرابعة والخامسة والسادسة من طرق التقدير الجزئي والتي تحسب درجة المفردة بما يكافئ نسبة ترجيح البديل الصواب بشرط عدم التخمين العشوائي التام.**

- **بوجه عام يختلف متوسط الدرجات الكلية المقدرة بكل طريقة من طرق التقدير الجزئي اختلافا دالا إحصائيا عن بعضها البعض، في متوسط الدرجات الكلية المعدلة من أثر التخمين، مما يستدعي الحاجة لتحديد أي طرق التقدير الجزئي أكثر دقة في تقدير الدرجة.**

وتختلف النتائج السابقة عن نتائج دراسة Taylor & Gardner (1999) التي توصلت إلى عدم اختلاف الدرجات التي تم الحصول عليها باستخدام التقدير الجزئي عن بعضها البعض وعدم اختلافها عن الدرجات الخام. وقد يرجع ذلك لاختلاف خصائص أفراد عينة الدراسة الحالية التي يرتفع نسبيا ميلها للجوء للتخمين نتيجة لانخفاض مستوى القدرة. كما قد يرجع هذا الاختلاف جزئيا إلى التطوير الذي أدخلته الدراسة الحالية على طرق التقدير الجزئي كما سبق بيانه.

ونظرا لأن ترتيب الفرد داخل مجموعة أقرانه من الأمور التي ينبغي أخذها في الاعتبار في نتائج اختبارات التحصيل الدراسي، فقد كان من الضروري اختبار حجم ودلالة الارتباط بين الدرجات الخام والدرجات المصححة من أثر التخمين التي حصل عليها أفراد العينة باستخدام الطرق الثمان للتقدير الجزئي، وذلك باستخدام معامل ارتباط بيرسون، وكانت النتائج كما بالجدول التالي:

جدول (٨) مصفوفة الارتباط بين الدرجات التي حصل عليها أفراد العينة باستخدام طرق التقدير المختلفة (ن=١٠٩)

طريقة تقدير الدرجة	الأولى	الثانية	الثالثة	الرابعة	الخامسة	السادسة	السابعة	الثامنة
الدرجة الخام	٠,٩٥**	٠,٩٦**	٠,٩٥**	٠,٨٨**	٠,٨٧**	٠,٨٨**	٠,٨٧**	٠,٨٣**
الأولى		٠,٩٩٨**	٠,٩٨**	٠,٩٢**	٠,٩٢**	٠,٩٢**	٠,٩٤**	٠,٩١**
الثانية			٠,٩٨**	٠,٩٢**	٠,٩١**	٠,٩٢**	٠,٩٤**	٠,٩٠**
الثالثة				٠,٩٠**	٠,٨٩**	٠,٨٩**	٠,٩٨**	٠,٩٢**
الرابعة					٠,٩٩٨**	٠,٩٩**	٠,٨٤**	٠,٩٣**
الخامسة						٠,٩٩**	٠,٨٤**	٠,٩٣**
السادسة							٠,٨٣**	٠,٩٤**
السابعة								٠,٩٣**

(**) الارتباط دال إحصائيا عند مستوى دلالة أقل من (٠,٠١)

يتبين من الجدول السابق وجود ارتباط قوي (قيمة ر = {0,83: 0,99}) ودال إحصائيا عند مستوى دلالة أقل من (0,01) بين الدرجات الكلية في الاختبار باستخدام طرق التقدير المختلفة، وكان هذا الارتباط تاما تقريبا بين كل من الطريقتين الأولى والثانية والطريقتين الرابعة والخامسة. مما يشير إلى اتساق كبير –وإن لم يكن تاما– بين هذه الطرق في ترتيب الأفراد تبعا لتحصيلهم الدراسي.

وقد كانت المجموعة الأولى من طرق التقدير الجزئي (باستخدام مقلوب عدد البدائل التي حصلت على أعلى نسبة ترجيح على أن يكون من بينها البديل الصواب بحالاتها الثلاث) أكثر اتساقا مع بعضها البعض، وأكثر اتساقا مع الدرجة الخام والطريقة السابعة مقارنة بالطرق الأخرى للتقدير الجزئي، وهذه نتيجة منطقية نظرا لأن المجموعة الأولى للتقدير الجزئي –وكذا الطريقة السابعة– لا تختلف عن الدرجة الخام إلا في تقدير درجة الاستجابات المخمنة تخمينا صريحا (عندما تعطى أعلى نسب ترجيح بالتساوي لبديلين أو ثلاثة من بينها البديل الصواب).

كما كانت الطريقة الثامنة للتقدير الجزئي هي الأقل اتساقا مع طريقة الدرجة الخام، وتعد هذه النتيجة أيضا منطقية، حيث تعتبر الطريقة الثامنة هي الأكثر اختلافا عن طريقة الدرجة الخام في مراعاة قدر المعرفة الجزئية التي استندت إليها الاستجابة عند تقدير الدرجة، كما أنها أكثر طرق المجموعتين الثانية والثالثة في وضع قيود على احتساب درجة المفردة كما سبق بيانه.

ولاتخاذ القرار بأي هذه الطرق توفر تقديرا أدق للدرجة، تم الإجابة عن السؤال الثاني كما يلي.

نتائج الإجابة عن السؤال الثاني، ومناقشتها:

ينص السؤال الثاني على: ما أكثر طرق التقدير الجزئي للدرجة تحقيقا لثبات القياس؟ وللإجابة عنه تم حساب معامل الثبات بطريقتي معامل ألفا، والتجزئة النصفية بطريقة جتمان، وذلك لكل طريقة من طرق التقدير الجزئي (درجة كل مفردة تتراوح بين صفر، 1). بالإضافة لحساب معامل الثبات للدرجات الخام (درجة كل مفردة هي 1 أو صفر) قبل التصحيح. وكانت النتائج كما بالجدول التالي.

جدول (٩) معاملي ثبات ألفا والتجزئة النصفية لكل طريقة من طرق التقدير الجزئي والدرجات الخام (ن=١٠٩)

طريقة حساب الدرجة	معامل الثبات	
	الفا	التجزئة النصفية
الدرجة الخام	٠,٥٦٣	٠,٦٤٠
الأولى	٠,٥٧٢	٠,٦٦٤
الثانية	٠,٥٧٤	٠,٦٦٣
الثالثة	٠,٥٩٦	٠,٦٨٧
الرابعة	٠,٦٥٤	٠,٧٠٩
الخامسة	٠,٦٥٥	٠,٧٠٤
السادسة	٠,٦٦٣	٠,٧١٦
السابعة	٠,٦٢٦	٠,٧١٧
الثامنة	٠,٦٧٨	٠,٧٤١

يتضح من الجدول السابق، ما يلي:

- انخفاض مؤشري ثبات القياس (0,563 , 0,640) عند الاعتماد على البيانات الخام للعينة التي تتضمن استجابات مخمنة متعددة، وذلك مقارنة بنتائج ثبات القياس الذي يوفره الاختبار كما سبق عرضها في أداة الدراسة؛ فقد تبين أن الاختبار يحقق درجة مقبولة من ثبات القياس عند حساب مؤشرات الثبات اعتمادا على البيانات الخام لعينة التدريج المرجعي (ن=227) التي استبعدت منها الحالات التي تدل نسب الترجيح الخاصة بها على أنها لجأت للتخمين الصريح في الإجابة عن مفردتين فأكثر. وبحساب قيمتي معاملي ألفا والتجزئة النصفية للدرجات الخام الكلية لعينة التدريج المرجعي وجد أنها تساوي (0,84) و(0,85) على الترتيب. مما يؤيد ضرورة التصحيح من أثر التخمين نظرا لتأثيره السلبي في دقة القياس.
- أن طرق التصحيح المختلفة رغم أنها تعتمد على نسب الترجيح كمحك واقعي في تحديد الاستجابات المخمنة وفي تقدير الدرجات، إلا أنها لا تصل في أفضل الحالات إلى تحقيق دقة القياس بنفس الدرجة التي يمكن أن تتحقق في حالة عدم الاعتماد على التخمين في الإجابة. فالطرق الإحصائية يمكن أن تلطف من التأثير السلبي لمشكلات التطبيق (أو مصادر أخطاء القياس) إلا أنها لا تتغلب عليها تماما. ومن هنا تتجلى أهمية ضبط مصادر أخطاء القياس قدر الإمكان في مرحلة جمع البيانات والمراحل الممهدة لها.
- أن الطرق الثمان للتصحيح من أثر التخمين تؤدي إلى تحسن ثبات القياس بدرجات متباينة مقارنة بثبات الدرجات الخام غير المصححة. وقد يرجع ذلك إلى أن إتاحة الفرصة للمفحوص لتقدير نسب الترجيح يتيح له التمعن والتدقيق في الاستجابة مما يقلل من أخطاء التسرع والإهمال، وبالتالي يزيد دقة القياس.

وتتسق هذه النتيجة بوجه عام مع ما وجده طه الخرشة (2016) وما أسفرت نتائج بعض الدراسات (Lesage et al., 2013, 190; Frary, 1989, 88-91) من ارتفاع ثبات القياس باستخدام طرق التقدير الجزئي مقارنة بالدرجة الخام. كما تتسق بوجه عام مع نتائج الدراسات السابقة التي أسفرت عن تحسن الثبات عند استخدام نسب الترجيح عنه في أسلوب عد الاستجابات الصحيحة، وقد فسر Frary (1989, 87-89) هذه النتيجة بزيادة زمن الاختبار نتيجة لإضافة الزمن اللازم لتحديد قيم الثقة. مع ملاحظة اختلاف طرق التقدير الجزئي المستخدمة في الدراسة الحالية عن نظيرتها في الدراسات المشار إليها، إلا أنها تقوم على نفس فكرة أخذ المعرفة الجزئية في الاعتبار عند حساب الدرجة.

- أن الثبات يتحسن بوجه عام كلما زادت القيود المفروضة على احتساب الدرجة لتحقق تصحيحا أكثر دقة من أثر التخمين، مما يدعم صدق طرق التقدير الجزئي في تصحيح الدرجة من أثر التخمين. وتتفق هذه النتيجة مع ما وجده مروان الصمادي

ويوسف السوالمة (2009) ومحمد عبد الوهاب (2019) من أن طريقة التصحيح الأكثر مراعاة للمعرفة الجزئية هي الأكثر دقة في تقدير قدرات الأفراد.

- أن الطريقة الثامنة –التي اقترحها الباحثان – هي أكثر هذه الطرق تحقيقا لثبات القياس. ويعني ذلك أن التقدير الجزئي لدرجة الفرد بما يكافئ نسبة ترجيحه للبديل الصواب بشرط أن يكون البديل الصواب هو أعلى البدائل ترجيحا وألا يكون الفرد قد اعتمد على التخمين في الاستجابة، يعد أفضل هذه الطرق للتصحيح من أثر التخمين. وقد يرجع ذلك إلى أن هذه الطريقة تأخذ في الاعتبار توفر المعرفة الكافية لدى الفرد لترجيح البديل الصواب بوضوح عن البدائل الأخرى (المشتتات)، بالإضافة إلى الأخذ في الاعتبار أن هذه المعرفة قد تكون جزئية. *ولذا استخدمت هذه الطريقة كمحك مرجعي لدقة التصحيح من أثر التخمين في الدراسة الحالية*.

يلي ذلك الطريقة السادسة أو التقدير الجزئي لدرجة الفرد بما يكافئ نسبة ترجيحه للبديل الصواب بشرط أن يكون البديل الصواب هو أعلى البدائل ترجيحا حتى وإن كان مخمنا تخمينا صريحا.

كما أن الطريقة السابعة أو الدرجات الخام المعدلة تلي هاتين الطريقتين في تحقيق دقة القياس وفقا لمؤشر الثبات بطريقة التجزئة النصفية. وحيث أن حساب هذه الدرجات الخام المعدلة يتم باستخدام الطريقة المعتادة لحساب الدرجة لكل استجابة (1 أو صفر)، لكن مع مراعاة تحويل درجة الاستجابة الصواب من (1) إلى صفر إذا كانت هذه الاستجابة مخمنة تخمينا صريحا، فإن هذا الإجراء يعد سهل نسبيا مقارنة بالطريقتين السابقتين، مما يجعله أكثر قابلية للاستخدام في التصحيح من قبل المعلمين والباحثين في حالة عدم توفر التصحيح الآلي لدرجات الاختبار.

وخلاصة هذه النتائج أن أدق الطرق لتصحيح الدرجة من أثر التخمين من خلال اعتبار المعرفة الجزئية عند تقدير درجة الفرد في الاختبارات المعرفية، تقوم على تقدير الدرجة الجزئية للاستجابات الصواب غير المخمنة تخمينا عشوائيا صريحا بما يكافئ نسب ترجيحها التي يقررها هذا الفرد، وعدم احتساب أية درجة جزئية في حالة الاستجابات الصواب المخمنة تخمينا عشوائيا صريحا، بالإضافة للقاعدة المعمول بها في اعتبار درجة الاستجابات الخطأ تساوي صفرا. أي أن تصحيح الدرجة من أثر التخمين يفضل أن يتعامل مع درجة الاستجابة الصواب فقط باعتبارها قد تتضمن قدرا من التخمين؛ بحيث إما أن تحسب هذه الدرجة صفرا عندما تكون هذه الاستجابة مخمنة تخمينا صريحا (أي إذا تساوت نسبة ترجيها مع بديل أو بدائل أخرى)، أو تكون مكافئة لنسبة ترجيح الفرد لها.

ثانيا: نتائج المقارنة بين طريقتي الحذف المتتالي والمواءمة في دقة تصحيح تقدير القدرة من أثر التخمين، ومناقشتها:

تضمن ذلك المقارنة بين تقديرات القدرة المصححة من أثر التخمين باستخدام كلتا الطريقتين والطريقة المرجعية (التقدير الجزئي)، واختبار ما إذا كانت دقة التصحيح تختلف باختلاف مستوى القدرة والتخمين، وذلك من خلال الإجابة عن الأسئلة الثالث والرابع والخامس من أسئلة الدراسة. وقد تطلب الإجابة عن هذه الأسئلة ما يلي:

- حساب التقديرات الأولية لقدرة أفراد العينة بناء على استجاباتهم للاختبار: بعد حساب الدرجة الخام الكلية لكل فرد باستخدام الطريقة التقليدية المعتادة كما سبق، تم الحصول على التقديرات الأولية للقدرة والخطأ المعياري لها، وذلك باستخدام الجدول المرجعي لتقديرات القدرة المقابلة لكل درجة خام كلية في الاختبار.
- تصحيح تقدير القدرة من أثر التخمين استنادا إلى التقدير الجزئي وفقا لنسب الترجيح (المحك المرجعي لتقدير القدرة): بعد حساب الدرجة الكلية المرجعية لكل فرد باستخدام الطريقة الثامنة من طرق التقدير الجزئي كما سبق، تم الحصول على التقديرات المرجعية للقدرة مصححةً من أثر التخمين والخطأ المعياري، وذلك باستخدام الجدول المرجعي لتقديرات القدرة المقابلة لكل درجة خام كلية في الاختبار.
- تصحيح تقدير القدرة من أثر التخمين باستخدام طريقة الحذف المتتالي: تم إعداد مصفوفة الحذف المتتالي التي تتضمن تقديرات القدرة الأولية المقابلة لكل درجة خام كلية في الاختبار ككل، وكذا تقديرات القدرة المقابلة لكل درجة خام كلية في الاختبار محذوفا منه المفردة الأولى، ثم المفردة الثانية وهكذا... وذلك مع توحيد صفر التدريج مع التدريج المرجعي لمفردات الاختبار. وباستخدام هذه المصفوفة تم حساب تقديرات القدرة المصححة من أثر التخمين والخطأ المعياري لها لكل فرد باستخدام المعادلات الخاصة بطريقة الحذف المتتالي السابق توضيحها تفصيلا.
- تصحيح تقدير القدرة من أثر التخمين باستخدام طريقة المواءمة: تم حساب الفروق بين تقدير القدرة الأولي لكل فرد وتقدير صعوبة كل مفردة من مفردات الاختبار، ومن ثم تم تحديد المفردات الصعبة بالنسبة لكل فرد والتي يزيد تقدير صعوبتها عن تقدير قدرة الفرد بمقدار (2) لوجيت على الأقل، وتم حذف استجابات كل فرد للمفردات الصعبة بالنسبة له من مصفوفة البيانات وكأنه لم يتعرض لها في الاختبار. ثم أعيد التحليل الإحصائي لهذه البيانات باستخدام برنامج WINSTEPS لتقدير القدرة لأفراد العينة، مع توحيد صفر التدريج مع التدريج المرجعي لمفردات الاختبار. وبذلك تم الحصول على تقديرات القدرة المصححة من أثر التخمين باستخدام طريقة المواءمة، ويصف الجدول التالي تقديرات القدرة التي تم الحصول عليها بالطرق الأربعة السابقة، والخطأ المعياري لها:

جدول (١٠) الوصف الإحصائي لتقديرات القدرة وخطئها المعياري قبل وبعد التصحيح من أثر التخمين (ن=١٠٩)

طريقة التقدير	المتوسط		الوسيط		الانحراف المعياري		معامل الالتواء (*)		معامل التفلطح (**)	
	القدرة	الخطأ المعياري	القدرة	الخطأ المعياري	القدرة	الخطأ المعياري	القدرة	الخطأ المعياري	القدرة	الخطأ المعياري
الأولية	-٠,٣٤	٠,٤٢١	-٠,٣٢	٠,٤١٠	٠,٦٦	٠,٠٢٣	-٠,٠٢	١,٨٧	-٠,٥٩	٥,٠١
المحك المرجعي	-١,٠١	٠,٤٥٦	-١,٠٣	٠,٤٤٠	٠,٧٥	٠,٠٦٤٦	-٠,٢٦	٢,٥٣	٠,٠٢	٨,٤٥
المواءمة	-٠,٣٨	٠,٤٣٣	-٠,٤٦	٠,٤٢٠	٠,٧٠	٠,٠٣٦	-٠,٠٨	١,٨٢	-٠,٥٥	٣,٦٥
الحذف المتتالي	-٠,٣٣	٠,٤٦١	-٠,٢٦	٠,٤٦٠	٠,٦٧	٠,٠٤٢	-٠,٠٦	٠,٤٣	-٠,٥٤	٠,٣٧

(*) الخطأ المعياري لمعامل الالتواء (٠,٠٢٣) (**) الخطأ المعياري لمعامل التفلطح (٠,٠٤٦)

يتضح من الجدول السابق ما يلي:

- أن قيم معاملي الالتواء والتفلطح تشير إلى اعتدالية توزيع تقديرات القدرة التي حصل عليها أفراد العينة باستخدام الطرق المختلفة للتقدير، بينما لا تتحقق اعتدالية توزيع الخطأ المعياري له باستثناء طريقة الحذف المتتالي.
- أن تقدير القدرة بعد التصحيح باستخدام طريقة المواءمة لم يرتفع عن التقدير الأولي (قبل التصحيح)، وهذه النتيجة المنطقية تختلف عن نتائج الدراسات السابقة التي استخدمت طريقة المواءمة في التصحيح من أثر التخمين (مثل: Andrich & Marais, 2014; Lin, 2018). وقد يرجح هذا الاختلاف جدوى التعديل الذي أدخلته الدراسة الحالية على طريقة المواءمة.
- أن متوسط قيم الخطأ المعياري لتقديرات القدرة باستخدام كل من المحك المرجعي وطريقتي التصحيح من أثر التخمين كان أكبر من نظيره قبل التصحيح من أثر التخمين. وبالرغم من أن ارتفاع قيم الخطأ المعياري بعد التصحيح من أثر التخمين- سواء كان ارتفاعا دالا إحصائيا أو غير دال إحصائيا- يعني انخفاض دقة تقدير القدرة بعد تصحيحه أو عدم تحسنه على أفضل تقدير، إلا أنه من الملاحظ ما يلي:
  - أن مستوى تحصيل العينة يعتبر متوسط أو أقرب إلى الانخفاض بوجه عام، حيث أن متوسط تقديرات القدرة قبل التصحيح (-0,34 لوجيت) تقع أدنى من نقطة صفر التدريج المرجعي (صفر لوجيت) على متصل كل من القدرة والصعوبة. كما أن مدى قدرة العينة(109طالبة) في تقديره الأولي يمتد من (-1,93) إلى (1,04)، منهم (40) طالبة فقط تقديرها أكبر من الصفر وكان تقدير أغلبهم (38) طالبة أقل من واحد لوجيت، في حين أن الاختبار يتيح تقديرات للقدرة تمتد من (-5,12 إلى 5,07) لوجيت. وقد كان اختيار هذه العينة مقصودا للحصول على الاستجابات المخمنة موضوع الدراسة.
  - وحيث أن التصحيح من أثر التخمين يؤدي إلى انخفاض تقدير هذه القدرة، بما يزيد من ابتعادها عن نقطة صفر التدريج، فقد تلازم ذلك مع زيادة قيم الخطأ المعياري لهذه التقديرات. ومن ثم فإن زيادة قيم الخطأ المعياري ترتبط بموقع

الفرد على متصل التدريج وليس بأحقيته في هذا الموضع، لاسيما وأن مستوى صعوبة الاختبار (بمتوسط يساوي صفر لوجيت) ملائم بوجه عام لمستوى قدرة العينة مع الأخذ في الاعتبار متوسط الخطأ المعياري لها (0,42 لوجيت).

- هذا بالإضافة إلى أن انخفاض عدد المفردات التي تقدر على أساسها القدرة بعد حذف المفردات الصعبة في طريقة المواءمة يرفع قيمة الخطأ المعياري، ويتسق ذلك مع ما ورد في مراجع القياس (مثل: Wise and Plake, 2016, 23-24) من أن مؤشر ثبات القياس يتناسب طرديا مع زيادة طول الاختبار. كما أن المعادلات المستخدمة لحساب الخطأ المعياري في طريقة الحذف المتتالي تختلف عن تلك المستخدمة في المحك المرجعي والمواءمة، وتعطي تقديرات مرتفعة للخطأ المعياري لا سيما إذا كان طول الاختبار أقل من(40) مفردة (Wainer & Wright, 1980, 388-389).

وبناء على ما سبق استندت الدراسة في الحكم على دقة التصحيح من أثر التخمين على تقديرات القدرة التي يوفرها المحك المرجعي، ولم تستند إلى الخطأ المعياري لهذه التقديرات، كما يتضح من إجابة الأسئلة الثالث والرابع والخامس.

- أن الطريقة المرجعية تعطي أقل تقديرات لقدرة أفراد العينة مقارنة بالتقديرات الأولية وطريقتي التصحيح موضع اهتمام الدراسة. ولاختبار دلالة هذه الفروق أجريت المقارنات اللازمة للإجابة عن الأسئلة التالية:

نتائج الإجابة عن السؤال الثالث ومناقشتها:

ينص السؤال الثالث على: ما مدى دقة تقدير القدرة المصححة من أثر التخمين باستخدام كل من طريقتي الحذف المتتالي والمواءمة؟

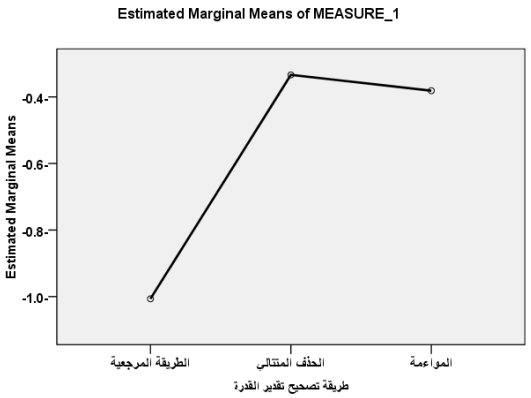
استند تقييم الدقة على اختلاف قيمة تقدير القدرة المصححة من أثر التخمين باستخدام كل من هاتين الطريقتين عن القيمة المصححة استنادا إلى المحك المرجعي (التقدير الجزئي).

ومن ثم فقد تضمنت الإجابة عن السؤال الثالث المقارنة بين كل من طريقتي التصحيح والمحك المرجعي، وكذا المقارنة بين هاتين الطريقتين، وذلك في قيم تقديرات القدرة المصححة من أثر التخمين. حيث تم اختبار دلالة وحجم الفروق بين هذه التقديرات باستخدام تحليل التباين داخل المجموعات (القياسات المتكررة)، وقد تم الاعتماد على تصحيح Greenhouse-Geisser لدرجات الحرية نظرا لانتهاك فرض الكروية، وكانت النتائج كما بالجدول التالي:

جدول (١١) دلالة وحجم الاختلاف بين تقديرات قدرة أفراد العينة في الاختبار باختلاف طريقة التصحيح من أثر التخمين(ن=١٠٩)

مصدر التباين	مجموع المربعات	درجة الحرية	متوسط المربعات	ف	مستوى الدلالة	مربع إيتا الجزئي
طريقة التصحيح من أثر التخمين	٣٠,٧٧	١,١٦	٢٦,٤٣	٢١٣,٣١	٠,٠٠٠	٠,٦٦
الخطأ	١٥,٥٨	١٢٥,٧٦	٠,١٢			

يتبين من الجدول السابق وجود فروق كبيرة ودالة إحصائيا عند مستوى دلالة أقل من (0,01) بين الطريقة المرجعية وطريقتي الحذف المتتالي والمواءمة في تصحيح تقديرات القدرة من أثر التخمين. ويؤكد ذلك النتائج الخاصة باختبارات المتغيرات المتعددة، حيث كانت قيمة اختبار Wilks' Lambda تساوي (0,32) بمستوى دلالة أقل من(0,001)، وقيمة مربع إيتا الجزئي الخاصة بها تساوي (0,68). وتتضح هذه الفروق من خلال التمثيل البياني التالي:



شكل (2) اختلاف متوسط تقديرات القدرة باختلاف طريقة التصحيح

ولتَبَيُّن مصدر هذه الفروق، أجريت المقارنات الثنائية بين طريقتي التصحيح من أثر التخمين والمحك المرجعي باستخدام اختبار Bonferroni، ويتضمن الجدول التالي متوسط ودلالة الفروق بين كل طريقتين في تقديرات القدرة.

جدول (١٢) نتائج المقارنات الثنائية بين تقديرات القدرة تبعا لطريقة التصحيح(ن=١٠٩)

طريقة التصحيح	الحذف المتتالي	المواءمة
المحك المرجعي	-٠,٦٧**	-٠,٦٣**
الحذف المتتالي	-	٠,٠٥*

(*) دال عند مستوى أقل من (٠,٠٥) (**) دال عند مستوى أقل من (٠,٠١)

يتضح من الجدول السابق وشكل(2) وجدول(10) للوصف الإحصائي:

- ارتفاع تقديرات القدرة المصححة من أثر التخمين باستخدام طريقة الحذف المتتالي عن نظيرتها في المحك المرجعي.
- ارتفاع تقديرات القدرة المصححة من أثر التخمين باستخدام طريقة المواءمة عن نظيرتها في المحك المرجعي.
- ارتفاع تقديرات القدرة المصححة من أثر التخمين باستخدام طريقة الحذف المتتالي عن نظيرتها في طريقة المواءمة. وحيث أن طرح نفس القيم (المرجعية) من تقديرات الطريقتين والمقارنة بينها يعطي نفس نتائج المقارنة بين تقديرات الطريقتين، فإن هذه النتيجة تعني أن طريقة الحذف المتتالي أكثر اختلافا عن المحك المرجعي مقارنة بطريقة المواءمة. وقد أكد استخدام اختبار t-test وجود فرق متوسط (مربع ايتا=0,092) ودال إحصائيا عند مستوى دلالة أقل من (0,01) بين الفرقين (الفرق بين تقديرات القدرة المصححة باستخدام طريقة الحذف المتتالي والتقديرات المناظرة لها المحسوبة استنادا إلى المحك المرجعي، والفرق بين تقديرات القدرة المصححة باستخدام طريقة الحذف المتتالي والتقديرات المناظرة لها المحسوبة استنادا إلى المحك المرجعي).

وتفيد هذه النتائج بأن كلا من طريقتي التصحيح باستخدام نموذج راش تعطي تقديرات للقدرة أكبر مما يستحقه الفرد. إلا أن طريقة الحذف المتتالي تعطي تقديرات للقدرة المصححة من أثر التخمين أعلى من نظيرتها في طريقة المواءمة، بل أنها أعلى قليلا من التقديرات الأولية غير المصححة من أثر التخمين في جدول(10)، وإن لم تختلف عنها اختلافا دالا إحصائيا (مستوى الدلالة=0,495)، في حين أنه من المنطقي أن يؤدي تصحيح تقدير القدرة من أثر التخمين إلى انخفاض هذا التقدير بدرجة ما قد ترقى أو لا ترقى لأن تكون دالة إحصائيا، طالما أن هناك استجابات مخمنة، وهذا ما تحقق بالفعل عند استخدام كل من المحك المرجعي وطريقة المواءمة.

ويرجح الباحثان أن هذه النتيجة ترجع إلى أن طريقة الحذف المتتالي –وفقا لإجراءاتها السابق توضيحها– يمكن القول بأنها تصحح تقدير القدرة من كل من استجابات التسرع والتخمين معا، في حين تركز طريقة المواءمة على التصحيح من أثر الاستجابة الصواب للمفردات شديدة الصعوبة بالنسبة للفرد والتي يرجح أنها جاءت نتيجة

للتخمين. كما أن الطريقة المرجعية تصحح الدرجة ومن ثم تقدير القدرة من أثر التخمين فقط، كما يقرره الفرد نفسه.

وبذلك يمكن القول بأن هناك اختلاف بين كل من طريقتي الحذف المتتالي والمواءمة والمحك المرجعي في تصحيح تقدير القدرة من أثر التخمين، مما قد يشكك في كفاية الاستناد للمحكات الإحصائية فقط في قرارات التصحيح من أثر التخمين. كما أن هناك اختلاف بين الطريقتين في تصحيح تقدير القدرة من أثر التخمين.

ولكن لا يمكن القول بأن طريقة المواءمة أكثر دقة في التصحيح من أثر التخمين؛ لاقتراب تقديراتها من تقديرات المحك المرجعي مقارنة بطريقة الحذف المتتالي. فربما إذا أُخِذ في الاعتبار المفردات السهلة (التي يقل تقدير صعوبتها عن قدرة الفرد بمقدار 2 لوجيت فأكثر) بالإضافة للمفردات الصعبة بالنسبة له عند التصحيح باستخدام طريقة المواءمة، ربما يؤدي ذلك إلى تكافؤ الطريقتين في التصحيح من أثر الاستجابات التي لا تتواءم مع مستوى الفرد سواء بسبب الاهمال والتسرع أو بسبب التخمين أو غير ذلك.

وربما يكون هذا الإجراء أكثر موضوعية في تقدير القدرة؛ حيث يعمل على تعويض الفرد عن الدرجات التي خسرها نتيجة للتسرع رغم أن معرفته تؤهله للحصول عليها، جنبا إلى جنب مع خصم الدرجات التي اكتسبها على غير علم نتيجة للتخمين. إلا أنه من ناحية أخرى؛ يرجح أن حذف جميع المفردات التي يتجاوز الفرق بين صعوبتها وقدرة الفرد الحدود (± 2) لوجيت قد يؤدي إلى تلاشي الاختبار أو قصره بدرجة تؤثر سلبا في دقة القياس. ويمكن أن يخلص الباحثان من ذلك إلى تأكيد الحاجة إلى الاختبارات التواؤمية في القياس النفسي للجانب المعرفي؛ حيث يتم مواءمة صعوبة مفردات الاختبار على قدرة الفرد أثناء التطبيق.

نتائج الإجابة عن السؤال الرابع ومناقشتها:

ينص السؤال الرابع على: ما مدى تباين دقة تقدير القدرة المصححة من أثر التخمين باستخدام طريقتي الحذف المتتالي والمواءمة باختلاف مستوى القدرة؟ وللإجابة عنه تم ما يلي:

- تقسيم العينة الكلية وفقا لمستوى القدرة إلى ثلاث مجموعات (أكثر قدرة /متوسط القدرة/ أقل قدرة)، وذلك باستخدام الإرباعين الثالث والأول لتقديرات القدرة ($0,17$، $-0,84$) لوجيت على الترتيب.
- حساب الفرق المطلق بين تقدير القدرة المصحح من أثر التخمين باستخدام طريقة الحذف المتتالي والمحك المرجعي، وذلك لكل فرد.

- **حساب الفرق المطلق بين تقدير القدرة المصحح من أثر التخمين باستخدام طريقة المواءمة والمحك المرجعي، وذلك لكل فرد.**

وقد استخدمت هذه الفروق المطلقة كمؤشر لدقة التصحيح من أثر التخمين باستخدام كلتا الطريقتين.

- **الوصف الإحصائي لهذه الفروق في كل مجموعة من المجموعات الثلاث المصنفة وفقا لمستوى القدرة.**
- **المقارنة بين المجموعات الثلاث في الاختلاف أو الفرق المطلق بين تقديرات القدرة المصححة باستخدام كل من طريقتي الحذف المتتالي والمواءمة والمحك المرجعي. وكانت النتائج كما يلي:**

1- نتائج الوصف الإحصائي للفرق المطلق بين تقديرات القدرة المصححة من أثر التخمين باستخدام كل من طريقتي التصحيح والمحك المرجعي في كل مجموعة من المجموعات المصنفة وفقا لمستوى القدرة، وذلك كما بالجدول التالي.

جدول (١٣) الوصف الإحصائي للفرق المطلق بين تقدير القدرة المصحح من أثر التخمين ونظيره المرجعي في المجموعات الفرعية لمستوى القدرة

طريقة التصحيح	المجموعة	ن	المتوسط	الوسيط	الانحراف المعياري	معامل الالتواء		معامل التفلطح	
						قيمته	الخطأ المعياري	قيمته	الخطأ المعياري
الحذف المتتالي	أكثر قدرة	٣٠	٠,٦٣	٠,٥٣	٠,٤١	٠,٧١	٠,٤٣	-٠,٢١	٠,٨٣
	متوسط القدرة	٤٦	٠,٨١	٠,٧٧	٠,٥٠	٠,٩٧	٠,٣٥	١,٠٦	٠,٦٩
	أقل قدرة	٣٣	٠,٥٣	٠,٥٣	٠,٤٢	١,٦٨	٠,٤١	٣,١٢	٠,٨٠
المواءمة	أكثر قدرة	٣٠	٠,٦٥	٠,٥٧	٠,٣٨	٠,٦٦	٠,٤٣	-٠,٠٢	٠,٨٣
	متوسط القدرة	٤٦	٠,٧٢	٠,٥٩	٠,٤٧	١,٣٤	٠,٣٥	٢,٣٧	٠,٦٩
	أقل قدرة	٣٣	٠,٥٠	٠,٤٠	٠,٤٠	١,٤١	٠,٤١	٢,٠٣	٠,٨٠

يتضح من الجدول السابق عدم تحقق اعتدالية توزيع مؤشر دقة التصحيح من أثر التخمين (الفرق المطلق) في مجموعتي متوسط القدرة والأقل قدرة في كل من طريقتي التصحيح؛ حيث تجاوزت قيم معامل الالتواء ضعف الخطأ المعياري لها.

2- نتائج اختبار دلالة وحجم اختلاف دقة تقدير القدرة المصححة من أثر التخمين باستخدام طريقتي الحذف المتتالي والمواءمة باختلاف مستوى القدرة:

نظرا لعدم تحقق الشروط المتعلقة باعتدالية التوزيع في أغلب المجموعات كما تبين من جدول(13)، وأيضا لعدم تحقق شرط تجانس التباين بين المجموعات الفرعية، لم يتسنى استخدام تحليل التباين المختلط أو البسيط للإجابة عن السؤال الرابع، ومن ثم تم اختبار دلالة الفرق في دقة التصحيح من أثر التخمين باختلاف مستوى القدرة لكل طريقة

من طريقتي التصحيح، وذلك باستخدام اختبار Kruskal–Wallis، وكانت النتائج كما بالجدول التالي:

جدول (١٤) دلالة وحجم اختلاف دقة تصحيح تقدير القدرة من أثر التخمين باختلاف مستوى القدرة

طريقة التصحيح	المجموعة	ن	متوسط الرتب	د.ح	كا^2	مستوى الدلالة	حجم التأثير
الحذف المتتالي	أكثر قدرة	٣٠	٥٣,٣٧	٢	٧,٧٣	٠,٠٢١	٠,٣٣
	متوسط القدرة	٤٦	٦٣,٩٣				
	أقل قدرة	٣٣	٤٤,٠٣				
المواءمة	أكثر قدرة	٣٠	٥٧,١٣	٢	٦,٢٨	٠,٠٤٣	٠,٣١
	متوسط القدرة	٤٦	٦١,٦٢				
	أقل قدرة	٣٣	٤٣,٨٣				

يتضح من الجدول السابق أن مستوى القدرة يمثل عاملا فارقا بالنسبة لدقة التصحيح من أثر التخمين، سواء استخدمت طريقة الحذف المتتالي أو طريقة المواءمة لإجراء هذا التصحيح. حيث كانت الفروق وفقا لمستوى القدرة متوسطة ودالة إحصائيا عند مستوى دلالة أقل من (0,05).

ولتبين مصدر هذه الفروق، أجريت المقارنات الثنائية بين كل مجموعتين باستخدام اختبار Mann–Whitney، ويبين الجدول التالي نتائج هذه المقارنات؛ حيث يتضمن مستوى دلالة الفرق بين كل مجموعتين في دقة التصحيح من أثر التخمين.

جدول (١٥) نتائج المقارنات الثنائية بين المجموعات الفرعية لمستوى القدرة في دقة التصحيح من أثر التخمين

طريقة التصحيح	المجموعة	أكثر قدرة	متوسط القدرة
الحذف المتتالي	متوسط القدرة	٠,١٤٢	
	أقل قدرة	٠,٢٢١	٠,٠٠٧
المواءمة	متوسط القدرة	٠,٥٦٢	
	أقل قدرة	٠,١٠٣	٠,٠١٣

يتضح من الجدول السابق وجدول(13) للوصف الإحصائي ما يلي:

- أن كلتا طريقتي التصحيح من أثر التخمين كانتا أكثر دقة في تقدير قدرة المجموعة الأقل قدرة مقارنة بمتوسطي القدرة، كما أن هناك اتجاه –لم يرق إلى أن يكون دالا إحصائيا– لارتفاع دقة التصحيح في الطريقتين في حالة منخفضي القدرة مقارنة بمرتفعي القدرة، وأيضا في حالة مرتفعي القدرة مقارنة بمتوسطي القدرة. أي أن العلاقة بين دقة التصحيح من أثر التخمين وتقدير القدرة تعتبر علاقة منحنية كالتي ظهرت في نتائج Waterbury & DeMars (2019)، وذلك في كل من طريقتي

التصحيح من أثر التخمين، ويؤيد ذلك التفاوت الكبير بين معامل ارتباط بيرسون ومعامل إيتا بين تقدير القدرة ودقة التصحيح من أثر التخمين (كما قدرت بالفرق المطلق) في طريقتي الحذف المتتالي (0,122، 0,514) والمواءمة (0,173، 0,452).

- وقد يرجع ذلك لكون المؤشرات الإحصائية تكون أكثر كفاءة في الكشف عن التخمين –ومن ثم أكثر اتساقا مع تقرير الأفراد– في المستويات المنخفضة من القدرة (شادية منتصر وحسان مخلوف، 2017)؛ حيث يكون من السهل تمييز مفردات صعبة بالنسبة لقدرة الفرد مقارنة بالمستويات المتوسطة والمرتفعة من القدرة.
- ولكن بناء على ذلك كان من المتوقع أن تكون التقديرات المصححة من أثر التخمين لمجموعة الأكثر قدرة هي الأكثر اختلافا عن المحك المرجعي، وأن تختلف اختلافا جوهريا عن الأقل قدرة، خاصة عند استخدام طريقة المواءمة في التصحيح من أثر التخمين، إلا أن تقديرات متوسطي القدرة على غير المتوقع كانت الأقل دقة (الأكثر اختلافا) في كل من طريقتي التصحيح! وقد ترجع هذه النتيجة إلى أن متوسطي القدرة يميلون في تقريرهم لنسب الترجيح إلى عدم القطع بترجيح صحة بديل ما للإجابة؛ لأن لديهم معرفة جزئية أكثر من منخفضي القدرة تجعلهم أكثر منهم قدرة على تحديد بديلين للإجابة في حين يكونون أقل من مرتفعي القدرة في القطع بصحة هذه البدائل مما يرفع من نسبة ما يخصم من الدرجة نتيجة التخمين الجزئي مقارنة بالمجموعتين الدنيا والعليا. إلا أن المجموعة الوسطى لم تختلف عن المجموعة العليا نظرا لأن المجموعة العليا –في الدراسة الحالية– لا تعتبر مرتفعة القدرة نظرا لانخفاض مستوى قدرة العينة بوجه عام كما سبق الإشارة. وبحساب الفروق بين تقديرات القدرة المصححة من أثر التخمين باستخدام الطريقة المرجعية والتقديرات الأولية قبل التصحيح، وحساب متوسط هذه الفروق للمجموعات الثلاث العليا والوسطى والدنيا وجد أنها تساوي (0,65، 0,77، 0,55) لوجيت على الترتيب، وقد اختلفت المجموعة الوسطى عن المجموعة الدنيا اختلافا دالا إحصائيا عند مستوى الدلالة (0,039) مما يرجح التفسير السابق.

وبالرغم من أن كلا من طريقتي الحذف المتتالي والمواءمة كانتا أكثر دقة في التصحيح من أثر التخمين في حالة منخفضي القدرة، إلا أن كلتا الطريقتين تظل مختلفة اختلافا جوهريا عن المحك المرجعي؛ فقد تم استخدام اختبار **Wilcoxon** لاختبار دلالة الفرق بين تقديرات القدرة المصححة من أثر التخمين باستخدام كل من الطريقتين والتقديرات المناظرة باستخدام المحك المرجعي، وذلك في كل مجموعة فرعية من المجموعات الثلاث السابقة، وقد كانت الفروق دالة إحصائيا في جميع المقارنات.

- كما يلاحظ أن قيم متوسط ووسيط الفروق المطلقة كانت أقل في طريقة المواءمة مقارنة بطريقة الحذف المتتالي، وذلك باستثناء مجموعة الأكثر قدرة. وبإجراء المقارنة بين الطريقتين في هذه الفروق باستخدام اختبار t-test (أو اختبار Wilcoxon في حالة عدم الاعتدالية)، تبين اختلاف المجموعتين في دقة التصحيح من أثر التخمين لمجموعة متوسطي القدرة فقط، بينما تكافأت الطريقتين في دقة التصحيح من أثر التخمين لكل من الأقل والأكثر قدرة.

مما سبق يمكن القول بأن دقة التصحيح من أثر التخمين بكل من طريقتي الحذف المتتالي والمواءمة تختلف باختلاف مستوى القدرة.

نتائج الإجابة عن السؤال الخامس ومناقشتها:

ينص السؤال الخامس على: ما مدى تباين دقة تقدير القدرة المصححة من أثر التخمين باستخدام طريقتي الحذف المتتالي والمواءمة باختلاف مستوى التخمين؟ وللإجابة عنه تم ما يلي:

- تقسيم العينة الكلية وفقا لمستوى التخمين الصريح إلى ثلاث مجموعات (أكثر تخمينا/ أقل تخمينا/ غير مخمِّن): تم حساب النسبة المئوية للمفردات المخمنة تخمينا صريحا من بين إجمالي المفردات الصواب لكل فرد، ثم صنف الفرد في فئة (غير مخمن) إذا كانت هذه النسبة تساوي صفرا؛ حيث تكون نسبة ترجيحه للبديل الصواب الذي نجح فيه في مفردة ما أعلى من نسبة ترجيحه للبدائل الثلاثة الأخرى، وغير ذلك صنف في فئة (مخمِّن). ثم قسمت مجموعة المخمن بدورها إلى مجموعتين (أقل تخمينا / أكثر تخمينا) وذلك باستخدام وسيط هذه النسب (17,65%) لمجموعة المخمِّن من عينة الدراسة.
- تقسيم العينة الكلية وفقا لمستوى التخمين الجزئي إلى ثلاث مجموعات (أكثر تخمينا/ متوسط التخمين/ أقل تخمينا): تم حساب النسبة المئوية للمفردات الصواب المخمنة جزئيا (التي تقل نسبة ترجيح البديل الصواب فيها عن 100%) من بين إجمالي المفردات الصواب لكل فرد، وقد كان هناك فرد واحد فقط لم يخمن أية مفردة جزئيا. ثم تم تحديد المستويات الثلاثة للتخمين الجزئي باستخدام الإرباعيَّيْن الأعلى والأدنى لهذه النسبة (80%، 39%) على الترتيب.
- وقد تم الاعتماد على الفروق المطلقة بين كل من طريقتي التصحيح والمحك المرجعي كمؤشر لابتعاد تقديرات الأفراد المحسوبة بكل منها عن التقديرات المناظرة المحسوبة بالمحك المرجعي، ومن ثم كمؤشر لدقة التصحيح من أثر التخمين، كما تبين من إجراءات إجابة السؤال الرابع.

- **الوصف الإحصائي لهذه الفروق في كل مجموعة من المجموعات المصنفة وفقا لكل من مستوى التخمين الصريح والتخمين الجزئي.**
- **المقارنة بين المجموعات المصنفة وفقا لمستوى التخمين في دقة تصحيح تقديرات قدرتها من أثر التخمين. وكانت النتائج كما يلي:**

1- نتائج الوصف الإحصائي للفرق المطلق بين تقديرات القدرة المصححة باستخدام كل من الحذف المتتالي والمواءمة والمحك المرجعي، في كل مجموعة من المجموعات المصنفة وفقا لمستوى التخمين، وذلك كما بالجدول التالي:

جدول (١٦) الوصف الإحصائي للفرق المطلق بين تقدير القدرة المصحح من أثر التخمين ونظيره المرجعي في المجموعات الفرعية لمستوى التخمين

متغير التصنيف	طريقة التصحيح	المجموعة	ن	المتوسط	الوسيط	الانحراف المعياري	معامل الالتواء		معامل التفلطح	
							قيمته	الخطأ المعياري	قيمته	الخطأ المعياري
التخمين الصريح	الحذف المتتالي	أكثر تخمينا	٣٣	١,١٠	١,٠٦	٠,٤٧	٠,٧٨	٠,٤١	٠,٣٦	٠,٨٠
		أقل تخمينا	٣٤	٠,٥٩	٠,٥٣	٠,٣١	٠,٧٠	٠,٤٠	٠,٦٥-	٠,٧٩
		غير مخمن	٤٢	٠,٤١	٠,٣٠	٠,٢٩	١,٢٥	٠,٣٧	٠,٩٩	٠,٧٢
	المواءمة	أكثر تخمينا	٣٣	١,٠٢	٠,٩٩	٠,٤٧	٠,٨١	٠,٤١	٠,٥٦	٠,٨٠
		أقل تخمينا	٣٤	٠,٥٣	٠,٤٩	٠,٣١	٠,٤٨	٠,٤٠	٠,٦٣-	٠,٧٩
		غير مخمن	٤٢	٠,٤١	٠,٣٨	٠,٢٥	٠,٧١	٠,٣٧	٠,٣٣	٠,٧٢
التخمين الجزئي	الحذف المتتالي	أكثر تخمينا	٣١	١,٠٣	١,٠٦	٠,٥١	٠,٧٥	٠,٤٢	٠,٤٠	٠,٨٢
		متوسط التخمين	٤٨	٠,٦٣	٠,٥٤	٠,٣٧	٠,٦٩	٠,٣٤	٠,٤٥-	٠,٦٧
		أقل تخمينا	٣٠	٠,٣٨	٠,٣٣	٠,٢٧	١,٦٤	٠,٤٣	٤,٠٥	٠,٨٣
	المواءمة	أكثر تخمينا	٣١	٠,٩١	٠,٨٣	٠,٥٢	٠,٨٩	٠,٤٢	٠,٦٩	٠,٨٢
		متوسط التخمين	٤٨	٠,٦٣	٠,٥١	٠,٣٤	٠,٦٣	٠,٣٤	٠,٤٤-	٠,٦٧
		أقل تخمينا	٣٠	٠,٣٥	٠,٣٥	٠,٢٥	١,٢٦	٠,٤٣	٢,٩٧	٠,٨٣

يتضح من الجدول السابق عدم تحقق اعتدالية توزيع مؤشر دقة التصحيح من أثر التخمين (الفرق المطلق) في مجموعة غير المخمن في طريقة الحذف المتتالي، ومجموعة الأقل تخمينا في كل من طريقتي التصحيح؛ حيث تجاوزت قيم معامل الالتواء ضعف الخطأ المعياري لها.

2- نتائج اختبار دلالة وحجم اختلاف دقة تقدير القدرة المصححة من أثر التخمين باستخدام طريقتي الحذف المتتالي والمواءمة باختلاف مستوى كل من التخمين والتخمين الجزئي:

نظرا لعدم تحقق الشروط المتعلقة باعتدالية التوزيع في بعض المجموعات كما تبين من جدول(16)، وأيضا لعدم تحقق شرط تجانس التباين بين المجموعات الفرعية، لم يتسنى استخدام تحليل التباين المختلط أو البسيط للإجابة عن السؤال الخامس، ومن ثم تم اختبار دلالة الفرق في دقة التصحيح من أثر التخمين باختلاف مستوى التخمين الصريح أو الجزئي وذلك لكل طريقة من طريقتي التصحيح، وذلك باستخدام اختبار Kruskal–Wallis، وكانت النتائج كما بالجدول التالي.

جدول (١٧) دلالة وحجم اختلاف دقة تصحيح تقدير القدرة من أثر التخمين باختلاف مستوى التخمين

طريقة التصحيح	متغير التصنيف	المجموعة	ن	متوسط الرتب	د.ح	كا٢	مستوى الدلالة(*)	حجم التأثير
الحذف المتتالي	التخمين الصريح	أكثر تخمينا	٣٣	٨٣,٦٢	٢	٤٤,٥٦	٠,٠٠٠	٠,٦٨
		أقل تخمينا	٣٤	٥٢,٢٤				
		غير مخمن	٤٢	٣٤,٧٥				
	التخمين الجزئي	أكثر تخمينا	٣١	٧٧,٨٤	٢	٣٠,٩٦	٠,٠٠٠	٠,٥٧
		متوسط التخمين	٤٨	٥٤,٠٩				
		أقل تخمينا	٣٠	٣٢,٨٥				
المواءمة	التخمين الصريح	أكثر تخمينا	٣٣	٨٢,٥٠	٢	٣٨,٤٣	٠,٠٠٠	٠,٦٣
		أقل تخمينا	٣٤	٤٩,٥٣				
		غير مخمن	٤٢	٣٧,٨٢				
	التخمين الجزئي	أكثر تخمينا	٣١	٧٤,١٥	٢	٢٧,٣٥	٠,٠٠٠	٠,٥٤
		متوسط التخمين	٤٨	٥٦,٩٦				
		أقل تخمينا	٣٠	٣٢,٠٨				

يتضح من الجدول السابق أن مستوى التخمين يمثل عاملا فارقا بالنسبة لدقة التصحيح من أثر التخمين، سواء استخدمت طريقة الحذف المتتالي أو طريقة المواءمة لإجراء هذا التصحيح. حيث كانت الفروق وفقا لمستوى التخمين (الصريح والجزئي على حد سواء) كبيرة ودالة إحصائيا عند مستوى دلالة أقل من (0,01). وينبغي الإشارة إلى أن استخدام تحليل التباين البسيط للمقارنة بين مجموعات الأكثر تخمينا ومتوسطي التخمين وغير المخمن في دقة التصحيح من أثر التخمين باستخدام طريقة المواءمة – حيث تتحقق شروطه– قد أسفر عن نفس النتيجة السابقة.

ولتبين مصدر هذه الفروق؛ أجريت المقارنات الثنائية بين كل مجموعتين باستخدام اختبار t–test للعينتين المستقلتين في حالة تحقق شرط اعتدالية التوزيع في كلتا مجموعتي المقارنة كما اتضحت من نتائج الوصف الإحصائي في جدول(16)،

بالإضافة إلى تقارب حجمي المجموعتين، وفي حالة عدم تحقق هذه الشروط تم استخدام اختبار Mann–Whitney لإجراء المقارنة، ويبين جدول(18) نتائج هذه المقارنات الثنائية؛ حيث يتضمن مستوى دلالة الفرق بين كل مجموعتين في دقة التصحيح من أثر التخمين.

جدول (١٨) نتائج المقارنات الثنائية بين المجموعات الفرعية لمستوى التخمين في دقة التصحيح من أثر التخمين

متغير التصنيف	طريقة التصحيح	المجموعة			طريقة التصحيح	المجموعة	
التخمين الصريح	الحذف المتتالي		أكثر تخمينا	أقل تخمينا	المواءمة	أكثر تخمينا	أقل تخمينا
		أقل تخمينا	٠,٠٠٠(*)			٠,٠٠٠(*)	
		غير مخمن	٠,٠٠٠	٠,٠٠٤		٠,٠٠٠(*)	٠,٠٥٢(*)
التخمين الجزئي			أكثر تخمينا	متوسط التخمين		أكثر تخمينا	متوسط التخمين
		متوسط التخمين	٠,٠٠٠			٠,٠١٠	
		أقل تخمينا	٠,٠٠٠	٠,٠٠٢		٠,٠٠٠	٠,٠٠٠

(*) أجريت المقارنة باستخدام اختبار t-test

يتضح من الجدول السابق وجدول(16) للوصف الإحصائي، ما يلي:

- أن كلتا طريقتي التصحيح من أثر التخمين تكون أكثر دقة في تقدير قدرة الفرد كلما انخفضت نسبة عدد المفردات التي أجابها صوابا عن طريق التخمين سواء كان هذا التخمين جزئيا أو صريحا؛ فقد كانت الفروق بين كل مجموعتين من المجموعات المصنفة وفقا للتخمين فروقا دالة إحصائيا بوجه عام، وكانت قيم كل من المتوسط والوسيط للفروق المطلقة تنخفض كلما انخفض مستوى التخمين.

وبالرغم من أن كلا من طريقتي الحذف المتتالي والمواءمة كانتا أكثر دقة في التصحيح من أثر التخمين في حالة منخفضي التخمين، إلا أن كلتا الطريقتين تظل مختلفة اختلافا جوهريا عن المحك المرجعي؛ فقد تم استخدام اختباري t–test أو Wilcoxon لاختبار دلالة الفرق بين تقديرات القدرة المصححة من أثر التخمين باستخدام كل من الطريقتين والتقديرات المناظرة باستخدام المحك المرجعي، وذلك في كل مجموعة فرعية من المجموعات الست السابقة، وقد كانت الفروق دالة إحصائيا في جميع المقارنات.

وقد ترجع هذه النتيجة إلى أن انخفاض التخمين يعني انخفاض درجة تأثر التقدير الأولي للقدرة بالتخمين؛ أي أن التقدير الأولي للقدرة لا يرتفع ارتفاعا مضللا نتيجة للتخمين الصائب، ومن ثم تظل تقديرات صعوبة المفردات المخمنة بعيدة عن تقدير قدرة الفرد فيسهل الكشف عنها وتصحيح أثرها باستخدام الطرق الإحصائية، وتقترب بذلك التقديرات التي توفرها هذه الطرق مع نظيرتها التي يوفرها المحك المرجعي.

كما قد تشير هذه النتائج إلى أن المحك المرجعي للتصحيح من أثر التخمين ربما يبخس الفرد حقه في تقدير القدرة؛ نتيجة لاعتماده على تقدير الطالب لنسب الترجيح والذي قد يتأثر بعوامل عديدة منها: سمات الشخصية حيث يميل البعض إلى عدم الإقرار بنسبة (100%) حتى وإن اطمأن لإجابته فيظل لديه هاجس باحتمال الخطأ خاصة وأن زمن الامتحان يكون محدودا وليس لديه الفرصة ليتأكد من إجابته بأكثر من طريقة. ومن هذه العوامل أيضا ضعف الثقة نتيجة لضغط موقف الامتحان. ومنها أيضا أن تعليمات الامتحان لم توضح دور نسب الترجيح في حساب الدرجة، وبالتالي يختار الطالب الإجابة التي يثق بدرجة كبيرة في صحتها ثم يوزع نسب الترجيح بطريقة لا تعكس هذه الثقة طمعا في الحصول على جزء مناسب من الدرجة إذا كان قد أساء فهم السؤال واختار بديلا خطأ. ومن ثم قد لا تكون نسب الترجيح مؤشرا كافيا للتعبير الدقيق عن درجة التخمين.

- كما يتضح من الجداول السابقة (16، 17، 18) أن هناك تماثل كبير بين طريقتي التصحيح من أثر التخمين في دلالة واتجاه الاختلاف في دقة التصحيح من أثر التخمين باختلاف مستوى التخمين. وترجح هذه النتائج عدم وجود تأثير للتفاعل بين طريقة التصحيح ومستوى التخمين الصريح أو الجزئي، وذلك في دقة التصحيح من أثر التخمين.
- كما يلاحظ من قيم متوسط ووسيط الفروق المطلقة في المجموعات الفرعية المختلفة كانت أقل بوجه عام في طريقة المواءمة مقارنة بطريقة الحذف المتتالي، مما يؤكد نتائج السؤال الثالث.

مما سبق يمكن القول بأن دقة التصحيح من أثر التخمين بكل من طريقتي الحذف المتتالي والمواءمة تختلف باختلاف مستوى كل من القدرة والتخمين.

وتختلف نتائج الأسئلة الثلاث الأخيرة من الثالث إلى الخامس في الدراسة الحالية عن نتائج Wainer & Wright (1980) التي توصلت إلى أن طريقة الحذف المتتالي كانت من أدق طرق التصحيح من أثر التخمين، وذلك في جميع مستويات التخمين في نموذج المحاكاة المستخدم؛ وتجدر الإشارة إلى اختلاف محك الحكم على دقة التصحيح من أثر التخمين في الدراسة الحالية عن المحكات المستخدمة في تلك الدراسة (التحيز، تباين تقدير القدرة والخطأ المعياري). بيد أن نتائج الدراسة الحالية تتفق مع نتائج تلك الدراسة في تراجع دقة التصحيح من أثر التخمين باستخدام الحذف المتتالي وباستخدام الطرق المستندة إلى الفرق بين تقديري القدرة والصعوبة، وذلك عندما ترتفع النزعة للتخمين. كما تتفق معها في اقتراب كفاءة هذه الطرق في حالة انعدام التخمين.

من النتائج السابقة وتحليلها ومناقشتها، تخلص الدراسة إلى التوصيات والمقترحات التالية للباحثين والمهتمين بالقياس المعرفي والتربوي:

- تحقيق أقصى ضبط ممكن لمصادر أخطاء القياس من خلال الاهتمام بكل من الإعداد الجيد لأدوات القياس المناسبة وضبط موقف التطبيق؛ حيث أن الطرق الإحصائية يمكن أن تلطف من التأثير السلبي لمشكلات التطبيق، إلا أنها لا تتغلب عليها تماما.
- الاهتمام بتطوير برامج حاسوب محلية تيسر إجراءات التصحيح من أثر التخمين بطرق عديدة، بما يشجع على إجراء المزيد من البحوث في هذا المجال قبل تفعيل طرق التصحيح من أثر التخمين في الميدان التربوي.
- الاهتمام بحساب الدرجة في الاختبارات العقلية والمعرفية بما يعكس المعرفة الجزئية للفرد، ويمكن انتقاء طريقة التقدير الجزئي المناسبة وفقا لأهداف عملية القياس، ومدى حساسية هذه المعرفة في التطبيق الميداني للخريج أو من يؤدي الاختبار المعرفي.

كما توصي الدراسة المسئولين التربويين بما يلي:

- الاهتمام بتفعيل الاختبارات التكيفية المحوسبة في قياس الجانب المعرفي، وتهيئة البنية التحتية والتجهيزات اللازمة لها؛ نظرا لما لهذه الاختبارات مميزات من بينها التقليل من احتمال لجوء الفرد للتخمين العشوائي.
- توفير أجهزة التصحيح الآلي المهيئة لحساب الدرجة لأسئلة إنتاج الاستجابة الموضوعية (مثل الإكمال القصير وتصويب الخطأ) في المؤسسات التربوية، وذلك لتشجيع الأساتذة والمعلمين على تنويع أسئلة الامتحانات، لاسيما في ضوء تزايد أعداد المتعلمين. حيث يؤدي استخدام هذه الأسئلة للتغلب على مشكلة التخمين.
- توخي الحذر في تفعيل طرق نموذج راش في التصحيح من أثر التخمين، في اختبارات التحصيل الدراسي من نوع الاختيار من متعدد، إلى أن يتم تقديم أدلة علمية كافية تساعد على انتقاء أفضل هذه الطرق تحقيقا لعدالة وموضوعية القياس.
- استخدام الطريقة الثامنة للتقدير الجزئي التي اقترحها الباحثان في الدراسة الحالية لتصحيح الدرجة الخام الكلية من أثر التخمين.

موضوعات مطروحة للبحث والدراسة:

- المقارنة بين طرق التقدير الجزئي التي اقترحتها الدراسة في تحقيق ثبات القياس في الاختبارات العقلية.
- المقارنة بين طريقة المواءمة كما طورتها الدراسة الحالية وطريقة المواءمة في صورتها غير المعدلة، وذلك في دقة التصحيح من أثر الاستجابات الشاذة، باختلاف نمط الاستجابة، واختلاف عتبة التخمين (بفرق 2، 2,5، 3 لوجيت بين تقدير القدرة وتقدير الصعوبة).
- المقارنة بين طريقة المواءمة –بعد تعديلها بحذف كل من المفردات السهلة والصعبة لكل فرد– وطريقة الحذف المتتالي في دقة التصحيح من أثر التخمين.
- دراسة لدقة تقدير القدرة المصححة من أثر التخمين عند تدريج الأفراد وفقا لنمط استجابتهم (1، صفر) مقارنة بالتقدير الجزئي وفقا لنسب الترجيح.

المراجع

أحمد محمد التقي (2013). *النظرية الحديثة في القياس* (ط 2)، عمان: دار المسيرة.

أحمد الرفاعي غنيم (1986). بعض طرق تقدير الدرجات للمفردات ذات الاختيار من متعدد، *مجلة كلية التربية جامعة المنصورة*، 7(7)، 73-107.

أحمد سليمان عودة (2014). تصحيح معاملات صعوبة الفقرات لأثر التخمين في أسئلة الاختيار من متعدد: صورة معدلة لمعادلة جيلفورد، *مجلة العلوم النفسية والتربوية جامعة السلطان قابوس*، 8(2)، 248-257.

أمينة محمد كاظم (2000). اتجاهات معاصرة في بناء بنوك الأسئلة، في: *الأسس التربوية لإعداد المعلم الجامعي* (ط3)، القاهرة: جامعة عين شمس، 321-342.

أمينة محمد كاظم، شادية عبد العزيز منتصر، رباب عبد المنعم سيف (2015). *معايير الورقة الامتحانية*، ورقة تدريبية غير منشورة، جامعة عين شمس: كلية البنات للآداب والعلوم والتربية.

رجاء محمود أبو علام (2014). *تقويم التعلم*، عمان: دار المسيرة.

رينولدز، س. وليفنجستون، ر. (2013). *إتقان القياس النفسي الحديث، النظريات والطرق*. ترجمة صلاح الدين محمود علام. عمان: دار الفكر.

سامي محمد ملحم (2015). *القياس والتقويم في التربية وعلم النفس* (ط 7)، عمان: دار المسيرة.

شادية عبد العزيز منتصر، حسان مخلوف (2017). الكشف عن التخمين باستخدام نموذج راش وعلاقته بخصائص الفرد والمفردة، *مجلة دراسات تربوية واجتماعية جامعة حلوان*، 23(3 ج3)، 1223-1307.

شريف عبد الرحمن السعودي، معين سلمان النصراوين (2018). مقارنة طرق ضبط أثر التخمين في اختبارات الاختيار من متعدد وأثرها على خصائص الاختبار، *مجلة دراسات في العلوم التربوية جامعة الأردن*، 45(4 ملحق 6)، 597-615.

صلاح الدين محمود علام (1991). أثر المواقف الاختبارية جماعية المرجع ومحكية المرجع في مستويات التخمين وسلوك المخاطرة والأداء في علم النفس (دراسة تجريبية). *مجلة علم النفس*، 5 (18، 19)، 98-106.

صلاح الدين محمود علام (2005). *نماذج الاستجابة للمفردة الاختبارية: أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي*، القاهرة: دار الفكر العربي.

صلاح الدين محمود علام (2015). *القياس والتقويم التربوي والنفسي: أساسياته وتطبيقاته وتوجهاته المعاصرة* (ط 6)، القاهرة: دار الفكر العربي.

طه عقلة الخرشة (2016). أثر بعض طرائق تصحيح اختبارات الاختيار من متعدد في دقة صعوبة الفقرات وقدرات الأفراد وفق نموذج راش في نظرية الاستجابة للفقرة، *مجلة كلية الدراسات العليا في التربية جامعة القاهرة*، 24(1)، 607-627.

محمد محمود عبد الوهاب (2013). المقارنة بين النظرية الكلاسيكية والنموذج ثلاثي البارامتر في معالجة أثر التخمين عند تقدير إحصائيات الأفراد والمفردات في اختبار اختيار من متعدد، *المجلة التربوية جامعة الكويت*، 28(109)، 169 -211.

محمد محمود عبد الوهاب (2019). أثر طريقة تقدير الدرجات ونموذج الاستجابة للمفردة في دقة تقدير بارامترات الأفراد والمفردات لاختبار صواب وخطأ متعدد، *المجلة الدولية للأبحاث التربوية – جامعة الإمارات العربية المتحدة*، 43(2)، 59-85.

مروان صالح الصمادي، يوسف محمد السوالمة (2009). أثر طريقة تصحيح فقرات الصواب والخطأ المتعدد في دقة تقدير قدرات الأفراد وصعوبة الفقرات باستخدام نموذج التقدير الجزئي، *المجلة الأردنية في العلوم التربوية*، 5(1)، 71-87.

يحيى مرزوق الأحمدي (1990). أثر التخمين على البنية العاملية للاختبارات المعرفية، *مجلة كلية التربية جامعة المنيا*، 3(4)، 237-275.

Andrich, D., & Marais, I. (2014). Person proficiency estimates in the dichotomous Rasch model when random guessing is removed from difficulty estimates of multiple-choice items. *Applied Psychological Measurement*, 38(6), 432-449.

Andrich, D., & Marais, I. (2018). Controlling bias in both constructed response and multiple-choice items when analyzed with the dichotomous Rasch model. *Journal of Educational Measurement*, 55(2), 281-307.

Bond, T.& Fox, C. (2015). *Applying the Rasch Model: Fundamental measurement in the human sciences* (3rd ed.). New York: Routledge.

Chiu, T. (2010). *Correction for guessing in the framework of the 3PL item response theory* (Doctoral dissertation, Rutgers University-Graduate School-New Brunswick).

Dodeen, H. (2005). Using distractors in correcting for guessing in multiple- choice tests. *Jordan University Journal of Educational Sciences*, 32(1), 192-199.

El-Kaffass, W. (2006). Assessing measures of persons with aberrant response patterns and normal response patterns according to Rasch model using ratio weights (w-ratio) for each item. *Egyptian Psychological Studies*, 16(1), 137-160.

Frary, R. (1989). Partial-credit scoring methods for multiple-choice tests. *Applied Measurement in Education*, 2, 79-96.

Holster, T., & Lake, J. (2016). Guessing and the Rasch model. *Language Assessment Quarterly*, 13(2), 124–141.

Haladyna, T., & Rodriguez, M. (2013). *Developing and validating test items*. New York: Routledge.

Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, 16(4), 277-298.

Lesage, E., Valcke, M., & Sabbe, E. (2013). Scoring methods for multiple choice assessment in higher education: Is it still a matter of number right scoring or negative marking? *Studies in Educational Evaluation*, 39(3), 188-193.

Lin, C-K. (2018). Effects of removing responses with likely random guessing under Rasch measurement on a multiple-choice language proficiency test. *Language Assessment Quarterly*, 15(4), 406-422.

Linacre, J. M. (2006). *A user's guide to WINSTEPS & MINISTEPS: Rasch-model computer programs.* Chicago: MESA.

Loyd, B. (1988). Implications of item response theory for the measurement practitioner, *Applied Measurement in Education*, 1(2), 135-143.

Paek, I. (2015). An investigation of the impact of guessing on coefficient α and reliability. *Applied Psychological Measurement*, 39(4), 264-277.

Reckase, D. (2009). *Multidimensional item response theory*. New York: Springer.

Taylor, C., & Gardner, P. (1999). An alternative method of answering and scoring multiple choice tests. *Research in Science Education*, 29(3), 353-363.

The College Board. (2017). *The SAT Subject tests student-guide*. Retrieved from: https://collegereadiness.collegeboard.org/pdf/sat-subject-tests-student-guide.pdf

Toland, M. (2014). Practical guide to conducting an item response theory analysis. *Journal of Early Adolescence*. 34(1), 120-151.

Wainer, H., & Wright, B. (1980). Robust estimation of ability in the Rasch model. *Psychometrika*, 45(3), 373-391.

Waterbury, G. T., & DeMars, C. E. (2019). The effects of probability threshold choice on an adjustment for guessing using the Rasch model. *Journal of Applied Measurement, 20*(1), 1-12.

Wise, L., & Plake, B. (2016). Test design and development following the standards for educational and psychological testing. In S. Lane, M. Raymond, & T. Haladyna (Eds.), *Handbook of test development* (2nd ed., 19-39). New York: Routledge.

Wright, B. & Stone, M. (1979). *Best test design: A hand book for Rasch measurement. Chicago*: META Press.

Wu, Q., De Laet, T., & Janssen, R. (2018). Elimination scoring versus correction for guessing: A simulation study. In M. Wiberg, S. Culpepper, R. Janssen, J. González, & D. Molenaar (Eds.), *Quantitative Psychology: The 82nd Annual Meeting of the Psychometric Society* (183-193). Cham: Springer.

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