أثر طريقة التصحيح لفقرات الاختيار من متعدد على الخصائص السيكومترية للاختبار وفقراته

إعداد

أ/ فادية علي إبراهيم بني عيسى

طالبة دكتوراة تخصص القياس والتقويم – جامعة اليرموك

د/ أحمد سليمان عودة

أستاذ بقسم القياس والتقويم – كلية التربية – جامعة اليرموك

أثر طريقة التصحيح لفقرات الاختيار من متعدد على الخصائص السيكومترية للاختبار وفقراته

أ/ فادية علي إبراهيم بني عيسى و د/ أحمد سليمان عودة *

الملخص:

هدفت هذه الدراسة إلى تقصي أثر طريقة التصحيح لفقرات الاختيار من متعدد على الخصائص السيكومترية للاختبار وفقراته. تألف مجتمع الدراسة من جميع طلبة الصف السادس من جميع مدارس قصبة إربد الأولى، وبلغت عينة الدراسة من (٦٠٦) طالباً وطالبة، والتي تم اختيارهم باستخدام العينة العشوائية الطبقية. ولتحقيق أهداف الدراسية قامت الباحثة بتطبيق اختبار قدرات عقلية (اوتيس–لينون) من نوع الاختيار من متعدد، وكشف نتائج التحليل عن دقة التصحيح باستخدام (قيمة معامل الصعوبة) حيث تم تصحيح إجابات الطلبة بإعطاء الإجابة الصحيحة (قيمة معامل الصعوبة) المصححص باستخدام معادلة عودة (٢٠١٤)، والإجابة الخاطئة (٠) حيث لم تظهر أي فقرات سهلة أو صعبة وجميها وقع ضمن المدى. كما توصلت الدراسة النتائج إلى أن قيم معاملات الثبات متقاربة بين طرق التصحيح، وتحققت البنية العاملية للاختبار على البيئة الأردنية.

الكلمات المفتاحية: طرق التصحيح، فقرات اختيار من متعدد، الخصائص السيكومترية.

* أ/ **فادية علي إبراهيم بني عيسى:** طالبة دكتوراة تخصص القياس والتقويم – جامعة اليرموك.
د/ أحمد سليمان عودة: أستاذ بقسم القياس والتقويم – كلية التربية – جامعة اليرموك.

Abstract

This study aimed to investigate the effect of scoring method of multiple choice test items on the psychometric properties of the test and its items. The study population consisted of all sixth-grade students from all Qasaba Irbid first schools, and the study sample consisted of (606) male and female students, who were selected using the stratified random sample. To achieve the objectives of the study, the researcher applied a test of mental abilities (Otis-Lennon) of the multiple choice type,

The results of the analysis revealed the accuracy of the correction using (the value of the difficulty coefficient), where the students' answers were corrected by giving the correct answer (the value of the difficulty coefficient) corrected using the return equation (2014), and the wrong answer (0), where no easy or difficult paragraphs appeared, and all of them fell within the range. The study also found that the values of the stability coefficients are close between the correction methods, and the factorial structure of the test was achieved on the Jordanian environment.

Keywords: Scoring Method, Multiple Choice Items, Psychometric Properties.

المقدمة:

حظي موضوع الاختبارات باهتمام كبير من قبل علماء القياس النفسي والتربوي، والاختبارات أكثر أدوات القياس استخداما وشيوعا في قياس سمة أو قدرة عقلية من خلال استجابة لعينة من المثيرات التي يفترض أنها تمثل تلك السمة أو القدرة، ويتوقف عليها نسبة كبيرة من التقدير الكلي لدرجة المفحوص من اجل إصدار أحكام حول مستويات المعرفة عند المفحوصين لأغراض تربوية أو تعليمية أو مهنية، لذلك حرص المختصين التربويون على أن تكون هذه الاختبارات ذات كفاءة عالية في عملية القياس والتقويم، وبالتالي فإن أثر طرق التصحيح للاختبارات على أداء المفحوصين وعلى الخصائص السيكومترية للفقرات يجب أن تكون ملائمة في تحليلها من أجل تحري الدقة في تقدير ما يستحقه الطلاب من تقديرات.

وتعتمد دقة القرارات على سلامة المعلومات والمعطيات التي تقوم عليها تلك القرارات (Hopkins & Antes, 1978). وتؤدي أدوات القياس بشكل عام، والاختبارات على وجه التحديد دورا أساسيا في توفير هذه المعلومات التي تسبق عملية التقويم واتخاذ القرارات، فكلما ازدادت درجة الثقة بأدوات القياس المستخدمة في جمع المعلومات، ازدادت دقة القرارات (الشايب، ٢٠٠٧).

ويهتم مطوّرو الاختبارات عادة ببعض الخصائص السيكومترية للوثوق بنتائج السمات التي صممت الاختبارات لقياسها. ومن هذه الخصائص صدق وثبات الفقرات، وفاعليتها بدلالة صعوبة تلك الفقرات وتمييزها وفعالية مموهاتها (Osterlind, 2006).

واختبارات الاختيار من متعدد من أكثر الاختبارات انتشاراً في التربية، والمجالات التنظيمية الأخرى، وهذا النوع من الاختبارات قد تفوقت على كل اشكال الفقرات الموضوعية الأخرى، وذلك بسبب كفاءتها وتعد استعمالاتها، إذ يمكن بواسطتها قياس أهداف بسيطة وأخرى مركبة في مختلف المواضيع الدراسية (Aiken, 2003). ويعد اختبار الاختيار من متعدد هو الاكثر استخداما في الاختبارات الموضوعية حيث يتكون من مشكلة وعدة بدائل، حيث تصاغ المشكلة اما في صيغة استفهامية او على شكل جمل ناقصة، وتسمى المشكلة (الجذر)، أما الحلول البديلة فهي عبارة عن الاجابات المحتملة والاجابات البديلة (البدائل)، وتتضمن البدائل اجابة واحدة صحيحة وعدد من الاجابات الخاطئة تسمى المموهات (المشتتات). حيث تمتاز الاختبارات من نوع الاختيار من متعدد بأنها تمكن الفاحص من قياس مدى تحقيق جميع الاهداف التعليمية بما فيها العمليات العقلية العليا كالتحليل والتركيب والتقويم، كما أنها تقلل من عملية تخمين للجواب الصحيح بزيادة عدد المموهات، ويمكن استخدام الكمبيوتر في تصحيح وتحليل نتائج هذا النوع من الاسئلة، وتساعد على قياس قدرة الطالب في التمييز بين الاحكام

الصحيحة والاحكام الخاطئة. وهي صادقة وثابتة بدرجة أكبر بكثير من بقية الانواع من الاختبارات (العجلوني، ٢٠١٦).

وتعتبر اختبارات الاختيار من متعدد من أصعب أنواع الاختبارات الموضوعية من حيث الإعداد وتتكون فقرة الاختيار من المتن ومجموعة من الإجابات، واحدة منها تسمى الإجابة الصحيحة والبقية تسمى مموهات (الإجابات الخاطئة)، ويطلب من المفحوص أن يختار إجابة واحدة صحيحة من الإجابات المعروضة عليه أو يضع عليها علامة أو يكتب رقم الإجابة الصحيحة في المكان المحدد لذلك (ربيع واحمد، ٢٠٠٨).

ويطلق على البدائل غير الصحيحة في فقرات الاختيار من متعدد مصطلح مموهات أو مشتتات (Distractors). وتعتمد جودة فقرة الاختيار من متعدد على جودة مموهاتها، فالمفهوم التقليدي للقوة التمييزية للفقرة هو المعكوس الجمعي لمجموع قيم التمييز لمموهاتها. وهذا يدلل على تأثير خصائص مموهات الفقرة بشكل كبير في صدقها وثباتها؛ فقد تكون الفقرة ضعيفة بسبب عدم فعالية مموهاتها؛ فالمموه الجيد هو المموه الجذاب وبالذات للأفراد الذين لا يعرفون اختيار الإجابة الصحيحة، أي أنه يميز باتجاه معاكس لتمييز الفقرة، أي يكون عدد المفحوصين الذين يختارونه من الفئة العليا أقل من عددهم في الفئة الدنيا (الجلبي، ٢٠٠٥). وهذا يعني أنه يمكن السيطرة على خصائص فقرات الاختبار السيكومترية والتحكم فيها من خلال زيادة الاهتمام ببدائل الإجابة لتلك الفقرات (عوده، ٢٠٢٢).

وفيما يتعلق بأثر شكل الفقرات في الخصائص السيكومترية للاختبار فقد أشار فريزبي (Frisbie, 1973) أن اختلاف شكل الفقرة قد يؤدي إلى أن تصبح الفقرة سهلة جداُ، أو صعبة جداُ، أو متوسطة الصعوبة حسب ما تتطلبه الفقرة من قدرة للإجابة عليها بشكل صحيح، مما يؤدي إلى حصول الطالب على علامات لا تمثل قدرته الحقيقية، وهذا بدوره يؤثر في صدق الاختبار وثباته. وصف كرونباخ (١٩٧١) التحقق من الصدق بأنه العملية التي يقوم من خلالها مطور الاختبار أو مستخدم الاختبار بجمع الأدلة التي تدعم أنواع الاستدلالات التي سيتم استخلاصها من علامات الاختبار. للتخطيط لدراسة الصدق، يجب تحديد الاستنتاج المطلوب بوضوح. ثم تصميم دراسة تجريبية لجمع الأدلة اللازمة لملائمة وفائدة العلامات لمثل هذه الاستنتاج (Crocker & Algina, 1986).

ويرى مارشال وهيلز (Marshall & Hales, 1971) أن الاختبارات من نوع الإجابة المصاغة اقل ثباتا من الاختبارات ذات الإجابة المنتقاة وهذا النقص في الثبات يعود لسببين:

- أخطاء ناتجة عن التصحيح.
- عينة الفقرات تكون غير كافيه ولا تمثل المحتوى.

أما جيه (Gay, 1880) فقد أشاد بالاختبارات الموضوعية وقال انها تستطيع أن تقيس نتاجات تعلم في جميع المستويات العقلية ويمكن أن تغطي مجالاً واسعاً من السلوكيات وبالتالي تكون أكثر صدقا وثباتاُ من الاختبارات المقالية.

وصعوبة فقرة الاختيار من متعدد تزادا بزيادة تجانس البدائل، وعندما تكون الفقرة غامضة وضعيفة لغوياً فإن ذلك يؤدي إلى ارباك الطلبة ويزيد من احتمال أن يستجيبوا للفقرة بشكل خاطئ وحتى وإن كانوا يمتلكون المعرفة اللازمة لإجابة الفقرة بشكل صحيح، وفي هذا الصدد يرى جرونلند ولين (Gronlund & linn, 1990) أنه إذا كان هناك غموضاً في متن فقرة الاختيار من متعدد، فإن ذلك قد يؤدي إلى أن كل مفحوص يمكن أن يفهم السؤال بطريقة مختلفة عن الآخر. وإذا كانت مموهات الفقرة ضعيفة وغير جذابة فإن ذلك قد يؤدي إلى تخفيض صعوبة الفقرة وتصبح قدرتها التميزية ضعيفة حيث أن الفقرات التي معامل صعوبتها قريب من الصفر أو الواحد صحيح تكون قدرتها التميزية ضعيفة (Murphy & Hofer, 1994).

وتقاس جودة فقرة الاختيار من متعدد من جودة مموهاتها، فالقوة التمييزية للفقرة بمفهومها التقليدي هي المعكوس الجمعي لمجموع قيم التمييز لمموهاتها. وهذا يؤكد ان لمموهات الفقرة اهمية كبيرة في تحديد خصائص الفقرة حيث تلعب دورا مهما في ثباتها وصدقها. فقد تكون الفقرة ضعيفة بسبب عدم فعالية مموهاتها. ويفترض في المموهات ان تكون جذابة وبالذات للأفراد الذين لا يعرفون الاختيار الصحيح، أي يكون عدد المفحوصين الذين يختارونه من الفئة العليا اقل من عددهم في الفئة الدنيا، أي انه يميز في اتجاه معاكس لتمييز الفقرة (Penfield, 2008). وهذا يعني انه يمكن التحكم في الخصائص السيكومترية للاختبار وفقراته من خلال المزيد من الاهتمام ببدائل الاجابة لتلك الفقرات. فمثلا بازدياد التجانس بين البدائل للفقرات تزداد صعوبتها، وكلما قل التجانس بين البدائل كانت الفقرة أسهل (عودة، ٢٠١٠).

ويرى عودة (٢٠٢٢) بأن المموهات يجب أن تكون جذابة للطلبة، أي أنه من الأفضل أن لا يقل عدد الطلبة الذين اختاروا المموه عن (٠.٠٥) من الطلبة ككل، وبما أن المموه يعتبر إجابة خاطئة فمن البديهي أن يكون عدد الطلبة الذين يختارون أي منها في الفئة العليا أقل منه من الفئة الدنيا.

ويمكن القول أنه إذ كان عدد الطلبة الذين اختاروا المموه قليلاً يجب حذفه، وبالتالي، فإنه من الممكن اعتبار الطريقة التي يتم بها تصفيه (حذف) المموه متغيراً يؤثر على خصائص الفقرة من صعوبة وتمييز. وهذا التأثير أكدت عليه بعض الدراسات مثل دراسة ظاظا (٢٠٠٠) التي درست أثر أربع طرق لحذف المموهات على خصائص الفقرة والاختبار والتي أظهرت نتائجها

بوجود أثر لطريقة حذف المموهات على معاملات التمييز. وكذلك دراسة أبو جراد (١٩٩٧) في معرفة أثر طريقتي الحذف التجريبي والعشوائي على خصائص الفقرة والاختبار، وقد أشارت النتائج بوجود أثر لطريقة حذف المموهات على معاملات الصعوبة لفقرات اختبار الاختيار من متعدد.

كما تتحدث ادبيات القياس والتقويم عن طرق عديدة للتصحيح في اختبارات الاختيار من متعدد من نوع أقصى أو متعدد Maximal perform، وأبرزها الطريقة التقليدية (١،٠) دون أي تعديل للعلامة علماً بأن هذه الأدبيات تشير إلى أن العلامة الظاهرية بهذه الطريقة أعلى من الحقيقية لأنها تختلف جزءً، يعود إلى التخمين، كما أنها لا تميز بين التخمين العشوائي والتخمين الذكي، وغالباً ما تستخدم هذه الطريقة لتجنب أي انعكاسات سلبية لافتراض قد لا يتحقق. (عودة، ٢٠٢٢).

اما الطريقة الثانية في التصحيح فتقوم على فرضية أن المفحوص يميل إلى التخمين الذكي، وان إجابته للسؤال حتى لو كانت خاطئة لا تخلو من المعرفة، بمعنى أن الصفر على مستوى السؤال الواحد هو صفر افتراضي Arbitrary Zero ولذلك تتعامل مع الإجابة الخاطئة على أنها جزئياُ أو نسبياً صحيحة، كأن يعطي علامة على البديل الأعلى نسبة في الاختبار بما يتناسب مع نسبة الطلبة الذين اختاروا ذلك البديل. (عودة، ٢٠٢٢).

في ضوء هذا العرض، تقوم الدراسة الحالية على ثلاثة طرق في التصحيح (التقليدية Traditional، معامل الصعوبة حسب معادلة عودة $d_c^* = \frac{Ld-1}{L-1} + \frac{1-d}{L}$ ، ووزن البديل Wight Distractor).

وهناك طرق أخرى في التصحيح ولكنها تقوم على أسس إحصائية، وترتكز إلى الطريقة التقليدية، مع بعض الافتراضات ويستخدم المصحح للدرجات على الاختبار معادلة أو أكثر للتعديل سواء للعلامة الكلية، أو للعلامة على مستوى السؤال في ضوء صعوبة ذلك السؤال، ثم تعديل معامل الصعوبة بمعادلة أو أكثر.

ومن المعادلات المستخدمة في التصحيح للعلامة الكلية على افتراض أن المفحوص مطلوب منه أن لا يخمن أو أن يستفيد من التخمين وترك تقدير أسلوب التعديل للباحث أو تستخدم الاختيار حسب الغرض، ومن هذه الأساليب والمعادلات الأقل تحفظاً في استخدامها (المعادلة التقليدية لا تشجع الأدبيات استخدامها) ولكن هناك أساليب يعتقد أنها أكثر عدالة منها

على مستوى الاختبار وأبرزها معادلة لتل ($X^* = x - \frac{W}{L-1}$)، وقد اعتبر لتل (Little, 1962) أن هذه المعادلة متشددة أو قدم معادلة أقل تشدداً، وذلك بتقليل التخفيض في الدرجة إلى النصف لتصبح ($X^* = x - \frac{W}{2(L-1)}$)، مفترضاً أن ما يقوم به المفحوص في الإجابة لا يمكن أن يكون عشوائياً تاماً أو تخميناً ذكياً تاماً، وافتراض أن التخمين يتم مناصفة بين النوعين. أما على مستوى الفقرة بطريقتي جيلفورد ومعادلة عودة (٢٠١٤). قدم جيلفورد تصحيحاً لمعامل الصعوبة (d_c) قائما على أنه نسبة المفحوصين الذين أجابوا إجابة صحيحة بدون تخمين (S/N) معتبراً أن ($S = R - \frac{W}{L-1}$)، وانتهت عمليات الاشتقاق بالتعويض للتوصل إلى أن معامل الصعوبة المصحح لأثر التخمين ($d_c = \frac{S}{N} = \frac{Ld-1}{L-1}$)، وهي معادلة تصحيح قائمة على العقاب لأثر التخمين دون التمييز بين الفقرات المحذوفة (O)، والفقرات التي أجاب عنها إجابة خاطئة (W)، مفترضاً أن لدى جميع المفحوصين نفس الدرجة من المخاطرة، وكأنه لا علاقة بين التصحيح وتعليمات الاختبار، ولذلك فهي ممارسة غير منطقية من ناحية، والافتراضات القائمة عليها لا تتحقق من ناحية اخرى، ولا تتجاوب مع التحذيرات الخاصة بانتهاك (Violation) الافتراضات التي تقوم معادلة تصحيح الدرجات بالعقاب، ولذلك فهي (اي معادلة التصحيح بطريقة جيلفورد) تتجاوز افتراض عدم وجود تباين في المخاطرة بين المفحوصين، وتتجاوز حقيقة التخمين الذكي والمعرفة الجزئية، حيث إن (W) في معادلة جيلفورد غير تلك التي ترد عادة في معادلة التصحيح لأثر التخمين بالعقاب انسجاما مع التعليمات التي تنص على أن لا يخمن أو لا يخاطر، فهو يفترض (اي جيلفورد) أن جميع المفحوصين لا يتقيدون بالتعليمات كما لو أنه طلب منهم الإجابة عن جميع الفقرات.

التعديل المقترح لمعادلة جيلفورد:

ينطلق التعديل المقترح من أن:

١. التعليمات تنص على أنه يطلب من المفحوص أن لا يخمن أو لا يخاطر في الإجابة العشوائية، ويتحمل بالتالي مسؤولية عدم الالتزام بهذه التعليمات.
٢. عدد الإجابات الخاطئة (w) في التعديل المقترح تختلف عنها عند جيلفورد، فهي عند جيلفورد تساوي (N–R)، إلا أنها في التعديل المقترح (N–R–O).
٣.معامل الصعوبة المصحح d_c^* قائم على نسبة الإجابات الصحيحة المعدّلة لأثر التخمين التي توازن بين المكافأة $\frac{1-d}{L}$ والعقاب، $\frac{Ld-1}{L-1}$ لأن التصحيح الذي يجمع بين المكافأة والعقاب يتوقع ان يحقق التوازن النسبي في التحيز بتقدير معامل صعوبة الفقرة وفقا للمعادلة المقترحة. $d_c^* = \frac{Ld-1}{L-1} + \frac{1-d}{L}$ ويبين الجدول التالي مقارنة ثنائية لقيم معاملات الصعوبة المعدلة بالمعادلتين d_c، d_c^* المقابلة لمعاملات الصعوبة على متصل الصعوبة المحسوب من نسبة الإجابات الصحيحة (d).

ولذلك فقد استخدم الباحثان في الدراسة الحالية أسلوب التصحيح الأولي ثم التعديل ما بعد التصحيح لمعامل صعوبة الفقرة حسب الوزن النسبي للإجابة الصحيحة بما يعادل صعوبة المقدر للفقرة.

ويكون بذلك المتغير المستقل هو طريقة التصحيح (A,B,C) والمتغيرات التابعة في الخصائص السيكو مترية للفقرات (صعوبة، تمييز) وخصائص توزيع العلامات بالأساليب الثلاثة، تم الثبات وفقاً لمعامل كرونباخ الفا كونه الاختبار يتم فحص أحادية البعد هنا بالتحليل العاملي لعامل واحد وحذف الأوزان الأقل من (20.) (عودة، 2014).

وفي هذا السياق، أجريت العديد من الدراسات التي اهتمت بأثر طرق التصحيح كدراسة عوده (٢٠٢٢) " وأشارت هذه الدراسة إلى التصحيح لأثر التخمين بالعقاب أو المكافأة، والعلاقات بين مفاهيم التخمين الذكي والتخمين العشوائي والمخاطرة. وجاءت هذه الدراسة بتصور مقترح للربط بين متغيري التحصيل والمخاطرة في تدريج ثنائي الاتجاه بسبع فئات، يكون الصفر مشتركاً بين التدريجين، باعتماد بيانات واقعية لعينة من ٦٤ طالبا و٥٠ فقرة في موقف اختباري حقيقي لمساق جامعي، تم تحليل تلك البيانات لتقديم المؤشرات الإحصائية والمنطقية الداعمة لهذا المدخل المقترح والمصمم للتعامل مع متغير المخاطرة كمصدر تباين يضيف ثباتا زائفاً على حساب الصدق. وكانت النتائج داعمة لهذا التصور. وفي ضوء النتائج وتفسيراتها والاستنتاجات المستخلصة منها، قدمت الدراسة متغير المخاطرة خياراً واقعياً في التعامل مع

العلاقات السالبة وربطها بمتغير التحصيل. كما قدمت الدراسة معادلة لتصحيح العلامات وفقاً لمتغير المخاطرة.

دراسة خصاونة (٢٠١٢) هدفت للكشف عن أثر اختلاف الأوزان النسبية لقواعد انتهاك صياغة فقرات الاختيار من متعدد في الاختبارات على خصائصها السيكومترية. ولأغراض الدراسة تم اختيار أربعة من مخالفات قواعد صياغة فقرات الاختيار من متعدد ذات الأوزان النسبية الأعلى بعد اشتقاقها باستخدام أسلوب التحكيم على مقياس ثيرستون، بحيث تم تطبيق كل نموذج على عينة مكونة من (٢٠٠) طالب من طلبة الصف الثامن الأساسي. وأظهرت النتائج وجود فروق ذات دلالة إحصائية في الخصائص السيكومترية للاختبار وفقراته بين نموذج R وكل النماذج (A,B,C,D) التي تتضمن فقرات مخالفة، ولصالح هذه النماذج في معاملات الصعوبة، بينما لصالح نموذج R في معاملات التمييز ومعاملات الثبات ومستوى الأداء على الاختبار. كذلك كان هناك أثر لاختلاف الأوزان النسبية لقواعد انتهاك صياغة فقرات الاختيار من متعدد بحيث كلما زاد الوزن النسبي كلما زادت صعوبة وتمييز الاختبار بينما قل مستوى الأداء على الاختبار، في حين لم يكن هناك أثر لاختلاف هذه الأوزان النسبية على معاملات ثبات الاختبار. كما أظهرت النتائج أن خلو الفقرات من المخالفات يحقق أحادية البعد لهذه الفقرات، والعكس صحيح، وأن اختلاف الأوزان النسبية لقواعد انتهاك صياغة فقرات الاختيار من متعدد يؤثر بدرجات مختلفة على أحادية البعد للاختبار، وعلى نسبة التباين المفسر بواسطة العوامل الرئيسة المستخلصة، وعلى عدد هذه العوامل.

كما قام عيلبوني (٢٠٠٧) بدراسة هدفت عدد البدائل وتمييز المموهات في فقرات الاختيار من متعدد على التوافق مع النموذج ثلاثي المعلم، ولتحقيق أهداف الدراسة تم بناء اختبار تحصيلي في مادة الرياضيات للصف التاسع الأساسي، تكوين ثلاثة أشكال اختبارية لكل منها (٥٠) فقرة، بحيث يحتوي الشكل الأول على خمسة بدائل للفقرة، والثاني على ثلاثة بدائل وذلك بحذف مموهين بشكل عشوائي من الشكل الأول، والشكل الثالث له ثلاث بدائل وذلك بحذف مموهين بالطريقة التجريبية (حذف المموهين الأقل تمييزاً) من الشكل الأول. طبق الاختبار على عينة مكونة من (١٦٥٦) طالباً وطالبة من طلاب الصف التاسع الأساسي التابعين لمديرية التربية والتعليم لمنطقة إربد الثانية. وبعد تحليل البيانات أظهرت الدراسة في جزء من نتائجها بأنه لا يوجد أثر لطريقة حذف المموهات في فقرات الاختيار من متعدد للأشكال ذات الثلاثة بدائل على خصائص الاختبار وفقراته وفق النموذج ثلاثي المعلم.

أجرى دودين (Dodeen, 2005) بدارسة هدفت إلى طرح معادلة جديدة للتصحيح لأثر التخمين في أسئلة الاختيار من متعدد Distraction-Scoring Formula (DSF). ومما

يميز المعادلة الجديدة (DSF) أنها تأخذ بعين الاعتبار مستوى جاذبية البديل في حساب درجة السؤال النهائية، وتفترض أن التخمين ليس عشوائياً دائماً، وأن الاختيار الأكثر جاذبية هو الاختيار الأقرب إلى الصواب من ضمن قائمة البدائل في سؤال الاختيار من متعدد، وقد تمت في هذه الدراسة مقارنة المعادلة الجديدة (DSF) بالمعادلة التقليدية للتصحيح لأثر التخمين الشائعة الاستعمال Scoring Formula (SF)، كذلك تمت مقارنة فعالية المعادلتين خصوصاً من حيث تأثير كل منهما على درجات الاختبار وثباتها من خلال تطبيق المعادلتين على نتائج اختبارين معياريين في جامعة الإمارات العربية المتحدة. وقد دلت النتائج على أفضلية استخدام المعادلة الجديدة (DSF) في مكافأة الطلبة على حسن اختيارهم للبديل الأقرب إلى الصواب أو الأكثر صحة عندما يكونون غير متأكدين تماما من الجواب الصحيح، وبالإضافة إلى ذلك فقد أدى تطبيق هذه المعادلة إلى زيادة ثبات نتائج الاختبارين عندما حسبت على عينات فرعية مشتقة من العينة الكلية لكل اختبار.

قام (Yaşar & Aybek, 2021) بدراسة هدفت إلى الكشف عن "طرق التصحيح للاختبارات من نوع (الاختيار من متعدد): كيف يغيّر التصحيح القائم على الأوزان النسبيّة والمتعلّق بصعوبة الفقرة نتائج اختبار الطّالب. ولقد هدفت هذه الدراسة إلى مقارنة الدّرجات التي أحرزها الطلبة في اختباراتهم مع الإحصاءات المتعلّقة بهذه الاختبارات وصولاً لحساب صعوبة الفقرة (the item difficulty) وذلك بالاستناد على طرق وأساليب التصحيح (scoring methods) التي تم استخدامها في الاختبار. ولقد أظهرت نتائج التحليل تسجيل النسب التالية: (725. %) و (721.%) وذلك على نحو متتالي. كذلك، فقد اشتمل التحليل على قياس معامل ارتباط النتاج الزمني لبيرسون (Pearson's Product Moment Correlation Coefficient)، اختصارا: (r, R)، والذي سجّل نسبة (٩١٦.%). كما وأشتمل التحليل كذلك على قياس معامل الارتباط حسب الرّتب لسبيرمان (Spearman's Rank-order Correlation Coefficient)، إختصاراً: (rs)، والذي سجّل نسبة (٩٢٦.%) بين درجات الطلاب التي تم الحصول عليها وفقًا لطريقتا التصحيح المذكورتان آنفاً (الطريقة الأولى والمستندة على معيار الدرجة (١-٠)، والطريقة الثانية المستندة معيار طريقة التصحيح القائم على الوزن النسبي (Qj – 0). علاوة على ذلك، فقد تم إجراء تقييم معياريّ، والذي إستند على معيارين لدرجات الإختبار وهما: (درجة ٥٠) و (درجة ٦٠). كما تم تحديد أعداد الطلبة الناجحين وغير الناجحين في المساق وفقًا لطريقتا التصحيح المذكورتان آنفاً. وبناء على التحليل النهائي لبيانات الدراسة، فقد أكّدت النتائج التي تم التوصّل إليها أنه وفي حال تم اعتماد طريقة (Qj – 0) للتصحيح، فإن ذلك سيؤدي في لاعتبار نسبة كبيرة من الطلبة غير ناجحين في المساق. ولكن،

وبالرّغم من النتيجة المترتّبة على استخدام هذه الطريقة، والتي تعتبر غير محبّذة في معظم الأحيان، إلّا أن هذه الطريقة لا تزال تمتلك أفضليّة دون غيرها. وتتمثّل تلك الأفضليّة بإمكانيّة طريقة (Qj – 0) من كشف الاختلافات المتواجدة بين الأفراد على اختلاف مستوياتهم وبشكل أكبر مقارنةً مع طريقة التصحيح المستندة على الأوزان غير نسبيّة.

أما دراسة عبدالله (٢٠١٩) والتي هدفت إلى قياس أثر التخمين في الاختبارات التحصيلية من نوع الاختيار من متعدد في ضوء النظرية الكلاسيكية لدى طلبة الصف التاسع الأساسي في مبحث الرياضيات. وقد تكونت عينة الدراسة من ١٥٩ طالباً وطالبة من طلبة الصف التاسع الأساسي لدى مدارس الأونروا (شمال) (عمان) تم اختيارهم عشوائياً. واستخدم الباحث المنهج الوصفي التحليلي في الدراسة. ولتحقيق هدف الدراسة قام الباحث بتطبيق اختبار تحصيلي من نوع الاختيار من متعدد بأربعة بدائل الرياضيات. وتوصلت الدراسة إلى وجود فروق ذات دلالة إحصائية عند مستوى0.05=a بين طريقة تصحيح المموهات وطريقة استخدام البديل "لا أعرف" ولصالح طريقة المموهات. كما بينت النتائج وجود فروق دالة إحصائياً عند مستوى (α=0.05) بين طريقة معادلة التصحيح لأثر التخمين (العقاب) وطريقة تصحيح المموهات ولصالح طريقة تصحيح المموهات ذات المتوسط الأعلى.

أجرى البركات (٢٠٢٢) دراسة هدفت هذه إلى التحقق من ثبات الصدق العاملي في اختبارات الاختيار من متعدد، وذلك باختلاف طرق التصحيح وقد تكونت عينة الدراسة من ١٤٩ طالباً وطالبة من الصف العاشر الأساسي مدارس في محافظة إربد. ولتحقيق أهداف الدراسة قام الباحث بتطبيق اختبار في الرياضيات نوع الاختيار من متعدد بعد أن قدمت للطلبة التعليمات الخاصة بطريقة الإجابة عن أسئلة الاختبار. وقد كشفت نتائج التحليل العاملي الاستكشافي عن وجود ثلاثة عوامل في كل من طريقتي التصحيح ، كما أشارت النتائج أن هناك فرقاً ذا دلالة إحصائية بين المتوسطين الحسابيين لمعاملات الصعوبة المحسوبة وفقا لطريقتي التصحيح، حيث بلغت قيمة اختبار ولكوكسون ٣.١٩٥ (Z) وهذه القيمة دالة إحصائياً عند مستوى 0.05=a كما بينت النتائج أيضاً أن هناك فرقاً دالاً إحصائياً بين المتوسطين الحسابيين لمعاملات التمييز المحسوبة وفقا لطريقتي التصحيح حيث بلغت قيمة اختبار ولكوكسون (Z) للفروق بين المتوسطات ٣.٠٠٨- وهذه القيمة دالة إحصائياً عند مستوى .0.05=aوكانت هذه الفروق لصالح فقرات الاختبارات المصححة بالطريقة التقليدية. أما النتائج المتعلقة بثبات الاختبار فقد أشارت إلى عدم وجود فرق ذا دلالة إحصائية بين معاملي ثبات الاتساق الداخلي (كرونباخ ألفا) لفقرات الاختبار يعزى لطريقة التصحيح.

وأجرى ليو وآخرون (Lau et al., 2011) دراسة هدفت إلى التعرف على أثر التخمين والمعرفة الجزئية على أداء الطلبة وثبات اختبار الاختيار من متعدد. وقد تألفت عينة الدراسة من ٤٤٩ طالباً من مدرسة ثانوية ماليزيا. ولتحقيق هدف الدراسة تمت مقارنة طريقة التصحيح بالطريقة التقليدية والتي يمكن للطالب أن يخمن خلالها، مقابل الطريقة الأخرى التي تشجع على استخدام المعرفة الجزئية واستبعاد بعض البدائل الصحيحة. وقد أشارت نتائج التحليل أن التخمين كان أقل عند استخدام طريقة التصحيح القائمة المعرفة الجزئية عن طريق استبعاد البدائل. وثبات الاختبار كان أعلى عند استخدام هذه الطريقة من ثبات الاختبار للدرجات المصححة بالطريقة التقليدية (اختيار البديل الصحيح فقط).

في ضوء مراجعة للدراسات ذات الصلة بموضوع الدراسة يمكن الإشارة إلى ندرة الدراسات العربية والأجنبية التي تناولت أثر طريقة التصحيح لفقرات الاختيار من متعدد على الخصائص السيكومترية للاختبار وفقراته، ومن خلال عرض الدراسات السابقة نجد أن هناك عدد من الدراسات التي اهتمت بدراسة تصفية المموهات في فقرات الاختيار من متعدد على معادلة الاختبار، وأثر اختلاف الأوزان النسبية لقواعد انتهاك صياغة فقرات الاختيار من متعدد في الاختبارات على خصائصها السيكومترية، أثر عدد البدائل وتمييز المموهات في فقرات الاختيار من متعدد. يتبين من الدراسات السابقة أن بناء الاختبارات مهمة جدا في قياس تحصيل الطلبة، ولذلك نتيجة ما يحققه المقياس من الوصول إلى مجموعة فقرات تكون على درجة عالية من الدقة والموضوعية في قياس السمة المراد قياسها، فمن خلال إطلاع الباحثة على الدراسات السابقة التي تم تناولها، ولم تجد الباحثة دراسات استخدمت طريقة أثر طريقة التصحيح لفقرات الاختيار من متعدد على الخصائص السيكومترية للاختبار وفقراته باستخدام طرق التصحيح (التقليدية، المموه، الصعوبة) مما يعطي أهمية لإجراء هذه الدراسة، والتي تتناول ما أغفلت عنه الدراسات الأخرى.

مشكلة الدراسة:

نظرا لما تحظى به الاختبارات من أهمية بين الباحثين والمختصين في المجال النفسي والتربوي، لقد بدأت تلقى أهمية وعناية في بنائها وتطويرها خاصة اختبارات الذكاء، مما أسفر عن دراسة الخصائص السيكومترية الذي يعد مؤشراً مهماً على دقة هذه الاختبارات، لما ما تقدمه من بيانات موضوعية في تحديد إمكانية الاستناد إليها في الحكم على قدرات الطلبة واتخاذ القرارات بحقهم.

وبغض النظر عن عدد البدائل الموضوعية لاختبارات الاختيار من متعدد، يتم تصحيحها كما هو مألوف بالطريقة التقليدية، بالرغم من أن هذه الطريقة تشجع المفحوصين على التخمين مما يترتب عليه من نتائج قد تكون مضللة في بعض الأحيان ولا تقيس الهدف المراد قياسه بدقة وخاصة عندما يتعلق الهدف من الاختبار التشخيص أو الاختيار أو التنبؤ. ولضبط هذا التخمين تم اقتراح عدة طرائق لتصحيح اختبارات الاختيار من متعدد، وأجريت العديد من الدراسات التي تناولت هذه الطرق دراسة (Yaşar & Aybek, 2021) استخدمت "طرق التصحيح (الأوزان النسبيّة، والصعوبة)، وكانت هناك دراسات استخدمت الأهمية النسبية لقواعد صياغة الفقرات الاختيار من متعدد بناء على تكرار اختيارها كقاعدة معتمدة من قبل خبراء القياس مثل دراسة فري بروس ورفاقة (Frey B. et al., 2005)، ودراسة (خصاونة، ٢٠١٢) لطريقة تصحيح باستخدام معاملات صعوبة الفقرات ومعاملات تميزه، ودراسة عيلبوني (٢٠٠٧) أثر عدد البدائل وتمييز المموهات في فقرات الاختيار من متعدد.

بالرغم من وجود العديد من الدراسات التي اهتمت بطرق التصحيح إلا انه لم يتم استخدام طرق التصحيح القائمة على (التقليدية، والصعوبة، المموه) في دراسة واحدة، وما زال الميدان النفسي والتربوي بحاجة إلى المزيد من إجراء الدراسات والأبحاث التي تركز على إيجاد الخصائص السيكومترية للاختبارات. وهنا تبلورت فكرة ومشكلة هذه الدراسة والتي كانت وليدة حاجة البحث العلمي، بضرورة تقصي أثر طريقة التصحيح لفقرات الاختبار من متعدد على الخصائص السيكومترية للاختبار وفقراته، لذا حاولت هذه الدراسة على وجه التحديد **في الإجابة عن الأسئلة التالية:**

السؤال الأول: هل يوجد أثر لطريقة التصحيح لفقرات الاختبار من متعدد على خصائص المقياس (معامل الصعوبة ومعامل التمييز)؟

السؤال الثاني: هل يوجد لطريقة التصحيح لفقرات الاختبار من متعدد على خصائص المقياس (الثبات، والصدق العاملي)؟

أهداف الدراسة:

- الكشف عن أثر طريقة التصحيح لفقرات الاختبار من متعدد على الخصائص السيكومترية للاختبار وفقراته (معامل الصعوبة ومعامل التمييز).
- الكشف أثر طريقة التصحيح لفقرات الاختبار من متعدد على الخصائص السيكومترية للاختبار وفقراته (الثبات، والصدق العاملي).

أهمية الدارسة:

جاءت أهمية هذه الدراسة من الناحيتين النظرية والعلمية على حد سواء، فمن الناحية النظرية تكمن أهميتها بتميزها على الدراسات السابقة بدراسة أثر طريقة التصحيح لفقرات الاختبار من متعدد على الخصائص السيكومترية للاختبار وفقراته، وكما تأتي أهميتها بدراسة موضوع متفرد لأثر استخدام طرق متنوعة لتصحيح فقرات الاختبار من متعدد. أما أهميتها من الناحية العملية فتكمن في دراستها التجريبية لأوضاع مختلفة عملياً، والتي يتوقع أن تعطي صورة أوضح عن احتمالية وجود أثر طريقة التصحيح لفقرات الاختبار من متعدد على الخصائص السيكومترية للاختبار وفقراته، ومحاولتها التوصل إلى تبريرات عملية تقدم إلى الباحثين والمهتمين في مجال الاختبارات، ويمكن اعتمادها على أساس تجريبي.

مصطلحات الدراسة:

اختبار الاختيار من متعدد: يعد هذا النوع من الاختبارات الموضوعية من أفضل أنواع الأسئلة، لأنها تقيس أهدافاً عقلية عليا يصعب على الاختبارات الموضوعية الأخرى قياسها. وتتألف من جزئين الأول هو الأورمة، والذي يكون عن حلول أو إجابات محتملة للسؤال الوارد في الأورمة (الظاهر وآخرون، ٢٠٠٢).

وتعرف في هذه الدراسة إجرائياً: بأنها اختبار قدرات عقلية يضم عددا من أسئلة الاختيار من متعدد، بحيث يتألف كل سؤال من أرومة أو متن، و(٤) بدائل من بينها الصحيحة.

تصحيح الاختبار بالطريقة التقليدية: وتعني وضع الدرجات بعد الانتهاء من التطبيق بالكامل للاختبار. فالطريقة التي فيها تسجيل إجابات الطلبة على فقرات الاختبار تحدد إلى حد كبير سهولة عملية التصحيح. في هذه الطريقة يحصل الطالب على الدرجة ١ إذا كانت إجابته صحيحة. أما إذا أجاب الطالب إجابه خاطئة يعطى الدرجة. (ألكين، ٢٠٠٧).

المموهات (Distractor): هي بدائل الإجابات غير الصحيحة للفقرات التي تكون من نوع الاختيار من متعدد.

اختبار اوتيس لينون للقدرات العقلية: هو اختبار مصمم للاستخدام مع التلاميذ وقياس العمليات اللفظية وغير اللفظية والكشف عن التشابهات والاختلافات وتعريف الكلمات وإتباع التعليمات واسترجاع الكلمات والأعداد وإجراء التصنيف والتسلسل وحل مسائل رياضية وينتج الاختبار مؤشرات كلية لفظية وغير لفظية للقدرة العقلية. (رينولدز وليفنقستون، ٢٠١٣: ٣٨٩).

تصحيح الاختبار بالطريقة القائمة على معامل الصعوبة: هي نسبة إجابة الصحيحة على الفقرة من اختيار من متعدد حسب معادلة عودة.

حدود الدراسة:

اقتصرت هذه الدراسة على أثر طريقة التصحيح لفقرات الاختبار من متعدد على الخصائص السيكومترية للاختبار وفقراته في تحليل بيانات عينة من طلاب وطالبات الصف السادس الأساسي في المدارس الحكومية في قصبة أربد.

- **الحدود البشرية:** اقتصرت الدراسة الحالية على طلاب وطالبات الصف السادس الأساسي في المدارس الحكومية في قصبة أربد.
- **الحدود المكانية:** تم تطبيق الدراسة في المدارس الحكومية في قصبة أربد.
- **الحدود الزمانية:** تم تطبيق الدراسة في الفصل الدراسي الأول من العام الدراسي ٢٠٢٢/٢٠٢٣. **الحدود الموضوعية:** اقتصرت الدراسة على المتغيرات المستقلة لأثر طريقة التصحيح لفقرات الاختبار من متعدد على الخصائص السيكومترية للاختبار وفقراته.

منهج الدراسة:

من أجل تحقيق أهداف الدراسة تم استخدام المنهج الوصفي التحليلي، وهو أسلوب من أساليب التحليل والتفسير العلمي المنظم لوصف ظاهرة أو مشكلة محددة وتصويرها بالأرقام من خلال جمع البيانات والمعلومات وتحليلها وتفسيرها بطريقة علمية دقيقة.

مجتمع الدراسة وعينتها:

تكون مجتمع الدراسة من جميع طلبة الصف السادس الأساسي في المدارس الحكومية التابعة لوزارة التربية والتعليم في المملكة الأردنية الهاشمية في تربية قصبة إربد للعام الدراسي 2023/2022، فقد بلغ عدد طلاب الصف السادس الأساسي (9856) طالباً وطالبة حسب إحصائيات قسم التخطيط التابع لوزارة التربية والتعليم الأردنية. وتكونت عينة الدراسة من طلاب وطالبات الصف السادس الأساسي في المدارس الحكومية في تربية قصبة إربد للعام الدراسي (2023/2022)، والتي تم اختيارهم باستخدام العينة العشوائية العنقودية، وبلغ عددهم (606) طالب وطالبة، جدول (١) يبين توزيع عينة الدراسة تبعاً لمتغير المدرسة والجنس.

جدول (١) توزيع عينة الدراسة تبعاً لمتغير المدرسة والجنس

المتغير	التصنيف	العدد	النسبة المئوية (%)
الجنس	أنثى	406	67.0
	ذكر	200	33.0
المدرسة	ابن زيدون	34	5.6
	الاندلس	104	17.2
	القصيلة	45	7.4
	المثنى	40	6.6
	أم سلمه	75	12.4

المتغير	التصنيف	العدد	النسبة المئوية (%)
	أمنه بنت وهب	31	5.1
	رفيدة الاسلامية	78	12.9
	ضاحية الحسين الثانوية للبنات	73	12.0
	عمر حلحول	126	20.8
	المجموع	**606**	**100.0**

أداة الدراسة:

تم استخدام احتبار اوتيس لينون للقدرة العقلية المستوى الثاني الصورة (K)، ويمثل هذا الاختبار أحد المستويات التي تتضمنها سلسلة اختبارات أوتيس لينون" للقدرة العقلية، والذي تم إعداده لقياس القدرات العقلية لطلاب الصف السادس الأساسي، ويقيس هذا الاختبار القدرة على التفكير والقدرة على حل المشكلات، والقدرة على التذكر.

إجراءات تطبيق الأداة:

تم التأكد من ملائمة الفقرات ووضوح فقرات الاختبار وتعليماته والتحقق من الزمن اللازم لتطبيق الاختبار وكذلك مناسبة الغرف الصفية من حيث التهوية والإضاءة وجاهزية الطلبة لتطبيق الاختبار عليهم، حيث طبق الاختبار على طلاب الصف السادس الأساسي في المدارس الحكومية (الإناث، والذكور) التابعة لوزارة التربية والتعليم في المملكة الأردنية الهاشمية في تربية قصبة إربد للعام الدراسي (2023/2022)، حيث تم تحديد الطلبة الذين طبق عليهم الاختبار بالطريقة العشوائية العنقودية ومن ثم وزعت عليهم أقلام الرصاص والقرطاسية المناسبة وكراسة الاختبار وتوضح لهم تعليمات الاختبار وقد تم طرح أمثلة توضيحية، والإجابة عن تساؤلات الطلاب والطالبات المفحوصين، وتم تسجيل وقت بدء الاختبار وتنبيه الطلاب بضرورة الالتزام بالوقت والإجابة على جميع الأسئلة، وعند الانتهاء من أداء الاختبار تم جمع كراسة الاختبار والقرطاسية منهم بمساعدة معلم الصف المتواجد في الصف.

تصحيح الأداة:

أولاً- الطريقة التقليدية (1،0) تعتمد هذه الطريقة على إعطاء المستجيب الدرجة (1) للإجابة الصحيحة، والدرجة (0) للإجابة الخاطئة.

ثانياً- التصحيح القائم على معامل الصعوبة حسب معادلة عودة: تعتمد هذه الطريقة على إعطاء المستجيب (قيمة معامل صعوبة الفقرة حسب معادلة عودة) للإجابة الصحيحة، والدرجة (0) للإجابة الخاطئة.

ثالثاً- التصحيح القائم على المموهات: تعتمد هذه الطريقة على إعطاء المستجيب (١) للإجابة الصحيحة، بينما الإجابة الخاطئة تأخذ وزن نسبي تبعاً للمجموعة التي حصلت على أعلى نسبة مئوية وتليها المجموعة الثانية ثم المجموعة الأخيرة.

١-**تطبيق الاختبار على عينة استطلاعية:** تهدف هذه المرحلة للتحقق من الصياغة اللغوية للفقرات وبدائلها ووضوح الأسئلة وتعليمات الاختبار، ومدى كفاية الوقت اللازم للاستجابة على فقراته، وذلك قبل التطبيق النهائي على العينة الأساسية، حيث تم تطبيق الاختبار على عينة عشوائية من الأفراد مكونة من (215) طالباً وطالبة خلال الفصل الدراسي الأول من العام (٢٠٢٢ / ٢٠٢٣).

٢-الخصائص السيكومترية للعينة الاستطلاعية:

أولاً- الصدق:

١. **إعداد الاختبار بصورته النهائية:** بعد قيام الباحثة بالتجريب الأولي للفقرات على العينة الاستطلاعية والتحقق من الصياغة اللغوية للفقرات وبدائلها ووضوح الأسئلة وتعليمات الاختبار، ومدى كفاية الوقت اللازم للاستجابة على فقراته، بالإضافة إلى استخراج مؤشرات الصدق والثبات (معامل الصعوبة ومعامل التمييز) لفقرات اختبار اوتيس لينون للقدرات العقلية، وعملت الباحثة على حذف فقرات الاختبار بناءً على معاملي التمييز والصعوبة، للخروج بالصورة النهائية للاختبار والتي أثبت التجريب الأولي أنها جيدة بحيث تكون جاهزة للتطبيق النهائي على العينة الأساسية والتي يكون حجمها أكبر من حجم العينة الاستطلاعية. جدول (٢) يعرض نتائج التحليل على العينة الاستطلاعية متمثلة بمعامل الصعوبة ومعامل التمييز لفقرات اختبار اوتيس لينون للقدرات العقلية بعد حذف الفقرات غير المناسبة.

جدول (٢) معامل الصعوبة ومعامل التمييز لفقرات اختبار اوتيس لينون للقدرات العقلية

رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز	رقم الفقرة	معامل الصعوبة	معامل التمييز
١	0.89	0.24	13	0.71	0.20	25	0.70	0.37
٢	0.82	0.26	14	0.76	0.20	26	0.75	0.33
٣	0.86	0.29	15	0.81	0.43	27	0.36	0.31
4	0.35	0.59	16	0.68	0.44	28	0.59	0.21
5	0.89	0.25	17	0.91	0.20	29	0.66	0.38
6	0.80	0.20	18	0.70	0.42	30	0.53	0.50
7	0.86	0.36	19	0.24	0.50	31	0.26	0.25
8	0.72	0.22	20	0.50	0.32	32	0.26	0.43
9	0.40	0.26	21	0.41	0.46	33	0.59	0.36
10	0.93	0.34	22	0.33	0.4	34	0.49	0.38
11	0.37	0.49	23	0.39	0.49			
12	0.40	0.59	24	0.34	0.49			

يبين جدول (٢) معاملات الصعوبة للفقرات الواردة في الاختبار والتي تراوحت بين (0.24) و(0.93)، وأن قيم معاملات التمييز المحسوبة لفقرات الاختبار تراوحت بين (0.20) و(0.59).

ثانياً- الثبات:

للتحقق من خاصية الثبات لاختبار اوتيس لينون للقدرات العقلية، فقد تم حساب معامل كرونباخ الفا، حيث بلغ معامل الثبات (0.841) وتشير إلى ثبات الاختبار.

١- **تطبيق الصورة النهائية لفقرات الاختبار على عينة الدراسة وجمع الاستجابات وتصحيحها:** بعد كتابة الصورة النهائية لفقرات الاختبار قامت الباحثة بتحديد مكان وزمان تطبيق الاختبار، ومن ثم عملت على جمع البيانات من المدارس المشمولة بالدراسة، ومن ثم قامت بجمع أوراق الاختبار وفرزها وتصحيحها وتنظيمها وتهيئتها لإدخالها بياناتها على البرامج الحاسوبية الخاصة بتحليل النتائج.

٢- **تحليل نتائج الاختبار:** تعتبر هذه المرحلة من أهم مراحل بناء وتطوير الاختبار، حيث تم تحليل استجابات الطلبة على الفقرات إحصائياً وذلك باستخدام برامج حاسوبية خاصة مثل (Excel, SPSS)

متغيرات الدراسة:

اقتصرت الدراسة الحالية على متغير مستقل واحد هو نوع أثر التصحيح لفقرات الاختيار من متعدد وله ثلاثة مستويات بعدد أنواع التصحيح التي تم اختيارها في الدراسة وسيتم تحديد أوزانها بناء على رأي المحكمين.

أما المتغيرات التابعة فهي:

معامل صعوبة الفقرة Item Difficulty، معامل تمييز الفقرة Item Discrimination، معامل الثبات Test Reliability، أحادية البعد للاختبار Unidimensionality of the Test.

المعالجة الاحصائية:

تم بعد تطبيق الاختبار على عينة المفحوصين والحصول على استجاباتهم التي تمثل البيانات اللازمة لإجراءات التحليل للكشف عن أثر طريقة التصحيح لفقرات الاختبار من متعدد على الخصائص السيكومترية للاختبار وفقراته، وتم إدخال البيانات إلى الحاسوب وسيتم الإجابة عن أسئلة الدراسة باستخدام برنامج الرزمة الإحصائية للعلوم الاجتماعية SPSS وفق ما يلي:

- تم الإجابة عن السؤال الأول الذي ينص "هل يوجد أثر لطريقة التصحيح لفقرات الاختبار من متعدد على خصائص المقياس (معامل الصعوبة ومعامل التمييز)؟" من خلال استخراج معاملي الصعوبة والتمييز لطرق التصحيح.
- تم الإجابة عن السؤال الثاني الذي ينص: هل يوجد لطريقة التصحيح لفقرات الاختبار من متعدد على خصائص المقياس (الثبات، الصدق العاملي)؟ من خلال حساب معامل كرونباخ الفا للتحقق من خاصية الثبات لاختبار اوتيس لينون للقدرات العقلية لجميع طرق التصحيح وتم استخدام التحليل العاملي التوكيدي.

النتائج المتعلقة بالإجابة عن السؤال الأول الذي نصه "**هل يوجد أثر لطريقة التصحيح لفقرات الاختبار من متعدد على خصائص المقياس (معامل الصعوبة ومعامل التمييز)؟**

للإجابة عن هذا السؤال، تم حساب معامل الصعوبة ومعامل التمييز لفقرات اختبار أوتيس لينون للقدرات العقلية حسب طرق التصحيح كما يلي:

بالطريقة الأولى (التقليدية): تم تصحيح إجابات الطلبة على فقرات اختبار اوتيس لينون للقدرات العقلية بإعطاء الإجابة الصحيحة (1)، والإجابة الخاطئة (0)، **وبالطريقة الثانية** (معامل الصعوبة حسب معادلة عودة) تم تصحيح إجابات الطلبة على فقرات اختبار اوتيس لينون للقدرات العقلية بإعطاء الإجابة الصحيحة (قيمة معامل الصعوبة)، والإجابة الخاطئة (0).

بالطريقة الثالثة (الوزن النسبي للممووه): تم تصحيح إجابات الطلبة على فقرات اختبار اوتيس لينون للقدرات العقلية بإعطاء الإجابة الصحيحة (1)، والإجابة الخاطئة (الوزن النسبي للممووه)، وجدول (3) يعرض نتائج معاملات الصعوبة ومعاملات التمييز لفقرات اختبار اوتيس لينون للقدرات العقلية باستخدام طرق التصحيح الثلاثة.

جدول (٣) معاملات الصعوبة ومعاملات التمييز لفقرات اختبار اوتيس لينون للقدرات العقلية

رقم الفقرة	طريقة التصحيح								
	التقليدية			الثانية(0، الصعوبة)			الثالثة (0، الوزن النسبي للممووه)		
	B	a	T	b	a	t	b	a	t
١	0.83	0.38	55.33	0.681	0.400	55.329	0.83	0.38	55.33
٢	0.69	0.42	36.40	0.456	0.428	36.395	0.73	0.44	43.53
٣	0.71	0.29	38.76	0.489	0.298	38.757	0.74	0.30	43.25
4	0.80	0.28	49.76	0.629	0.300	49.759	0.80	0.29	49.76
5	0.78	0.44	45.73	0.591	0.448	45.725	0.78	0.44	45.73
6	0.48	0.37	23.49	0.208	0.365	23.485	0.57	0.37	33.50
7	0.54	0.39	26.63	0.271	0.385	26.629	0.60	0.38	33.93
8	0.53	0.41	26.02	0.259	0.396	26.018	0.62	0.43	36.33
9	0.43	0.20	20.15	0.171	0.169	15.398	0.61	0.23	36.91
10	0.83	0.46	53.42	0.673	0.478	53.421	0.83	0.46	53.42

رقم الفقرة	طريقة التصحيح								
	التقليدية			الثانية(0، الصعوبة)			الثالثة (0، الوزن النسبي للمموه)		
	B	a	T	b	a	t	b	a	t
11	0.77	0.35	45.08	0.579	0.373	45.085	0.78	0.36	48.83
12	0.74	0.41	41.07	0.528	0.417	41.066	0.76	0.42	45.88
13	0.56	0.45	27.72	0.293	0.460	27.715	0.61	0.47	33.34
14	0.56	0.37	27.53	0.291	0.363	27.531	0.64	0.38	37.98
15	0.63	0.32	31.89	0.376	0.317	31.894	0.67	0.33	38.90
16	0.49	0.52	24.19	0.220	0.487	24.194	0.59	0.52	35.22
17	0.71	0.61	38.60	0.488	0.637	38.601	0.73	0.62	41.85
18	0.44	0.33	21.68	0.172	0.311	21.683	0.54	0.33	32.80
19	0.44	0.29	21.98	0.174	0.259	21.976	0.54	0.29	31.83
20	0.22	0.30	13.04	0.034	0.276	13.043	0.42	0.29	33.32
21	0.51	0.38	25.34	0.241	0.364	25.339	0.59	0.36	35.10
22	0.39	0.34	19.58	0.131	0.343	19.576	0.46	0.39	26.41
23	0.51	0.38	25.26	0.241	0.367	25.255	0.59	0.39	33.72
24	0.72	0.37	39.39	0.501	0.386	39.391	0.73	0.37	42.12
25	0.55	0.30	27.26	0.283	0.284	27.256	0.62	0.31	35.47
26	0.67	0.37	34.91	0.430	0.366	34.915	0.68	0.38	37.46
27	0.70	0.57	37.99	0.476	0.588	37.990	0.72	0.58	40.37
28	0.43	0.34	21.18	0.163	0.320	21.179	0.52	0.38	29.80
29	0.58	0.40	28.66	0.314	0.394	28.663	0.64	0.39	37.79
30	0.67	0.48	34.79	0.429	0.503	34.785	0.69	0.49	38.99
31	0.50	0.48	24.43	0.227	0.457	24.435	0.58	0.47	34.21
32	0.74	0.29	41.78	0.533	0.310	41.775	0.74	0.28	41.07
33	0.49	0.37	24.11	0.220	0.377	24.114	0.59	0.38	35.40
34	0.50	0.47	24.43	0.227	0.445	24.435	0.59	0.48	36.02

يظهر من البيانات في جدول (3) النتائج التالية:

وفقاً للطريقة الأولى نلاحظ أن معاملات الصعوبة لفقرات اختبار اوتيس لينون للقدرات العقلية تراوحت بين (0.22) و(0.83)، وأن قيم معاملات التمييز المحسوبة لفقرات الاختبار تراوحت بين (0.20) و(0.61). وللطريقة الثانية نلاحظ أن معاملات الصعوبة لفقرات اختبار اوتيس لينون للقدرات العقلية تراوحت بين (0.034) و(0.681)، وأن قيم معاملات التمييز المحسوبة لفقرات الاختبار تراوحت بين (0.169) و(0.637). وللطريقة الثالثة نلاحظ أن معاملات الصعوبة لفقرات اختبار اوتيس لينون للقدرات العقلية تراوحت بين (0.42) و(0.83)، وأن قيم معاملات التمييز المحسوبة لفقرات الاختبار تراوحت بين (0.23) و(0.62).

السؤال الثاني: هل يوجد لطريقة التصحيح لفقرات الاختبار من متعدد على خصائص المقياس (الثبات، الصدق العاملي)؟

للإجابة عن هذا السؤال تم حساب معامل كرونباخ الفا للتحقق من خاصية الثبات لاختبار اوتيس لينون للقدرات العقلية لجميع طرق التصحيح، حيث كانت النتائج كما في جدول (٤):

جدول (٤) معامل كرونباخ الفا لطرق التصحيح الثلاثة

رقم الطريقة	طريقة التصحيح	معامل كرونباخ الفا
الأولى	التقليدية	0.874
الثانية	(0، معامل الصعوبة)	0.872
الثالثة	(الوزن النسبي للممووه، 1)	0.880

يظهر من النتائج الواردة في جدول 4 أن معامل الثبات بلغ (0.874) باستخدام طريقة التصحيح الأولى. وفيما يتعلق بالطريقة الثانية فقد بلغ معامل الثبات (0.872). وفيما يتعلق بالطريقة الثالثة فقد بلغ معامل الثبات (0.880) وتشير جميعها إلى ثبات الاختبار.

للتحقق من الصدق العاملي لجميع طرق التصحيح تم استخدام التحليل العاملي التوكيدي من خلال برنامج التحليل الاحصائي (AMOS) وكانت نتائج مؤشرات حسن المطابقة والمواءمة **المقارن** ومؤشر البواقي (RMSEA) واختبار (Chi^2) كما في جدول (٥):

جدول (٥)

مؤشرات حسن المطابقة والمواءمة المقارن ومؤشر البواقي (RMSE) واختبار (Chi^2)

طريقة التصحيح	مؤشرات المطابقة والمواءمة المقارن								
	Chi^2	DF	CMIN/DF	GFI	NFI	TLI	CFI	RMSEA	RMR
التقليدية	7149.40	527	13.566	.632	.292	.260	.305	.144	.023
(0، معامل الصعوبة حسب معادلة عودة)	7142.76	527	13.554	.633	.291	.260	.305	.144	.007
(الوزن النسبي للممووه، 1)	5733.45	527	10.879	.667	.346	.325	.366	.156	.016

يوضح جدول (٥) نتائج مؤشرات المطابقة والمواءمة المقارن لنموذج القياس لاختبار اوتيس لينون للقدرات العقلية وفيما يلي النتائج لكل طريقة من طرق التصحيح:

الطريقة الأولى: التقليدية (1،0): بلغت قيمة (Chi^2/DF) المحسوبة (13.566)، وهي ذات دلالة عند مستوى (0.05) ودرجات حرية (527) . وبلغت قيمة Goodness of Fit (GFI) Index وهو مؤشر ملاءمة الجودة ما قيمته (.305) وهو مقارب إلى قيمة الواحد صحيح (الملاءمة التامة). وبنفس السياق بلغ Comparative fit index (CFI) مؤشر المواءمة

(.632)، وأما مؤشر جذر متوسط مربع الخطأ التقريبي Root Mean Square Error of Approximation (RMSEA) بلغت قيمة (.144) وهو مساوي لقيمة الصفر.

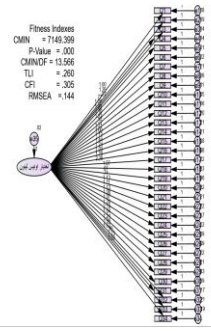
الطريقة الثانية: (0، معامل الصعوبة حسب معادلة عودة): بلغت قيمة (Chi2/DF) المحسوبة (13.554)، وهي ذات دلالة عند مستوى (0.05) ودرجات حرية (527) . وبلغت قيمة Goodness of Fit Index (GFI) وهو مؤشر ملاءمة الجودة ما قيمته (0.633) وهو مقارب إلى قيمة الواحد صحيح (الملاءمة التامة). وبنفس السياق بلغ Comparative (CFI) fit index مؤشر المواءمة (0.305)، وأما مؤشر جذر متوسط مربع الخطأ التقريبي Root Mean Square Error of Approximation (RMSEA) بلغت قيمة (0.144) وهو مساوي لقيمة الصفر.

الطريقة الثالثة: (الوزن النسبي للمموه، 0): بلغت قيمة (Chi^2/DF) المحسوبة (10.879)، وهي ذات دلالة عند مستوى (0.05) ودرجات حرية (527). وبلغت قيمة Goodness (GFI) of Fit Index وهو مؤشر ملاءمة الجودة ما قيمته (.667) وهو مقارب إلى قيمة الواحد صحيح (الملاءمة التامة). وبنفس السياق بلغ Comparative fit index (CFI) مؤشر المواءمة (.366)، وأما مؤشر جذر متوسط مربع الخطأ التقريبي Root Mean Square Error of Approximation (RMSEA) بلغت قيمة (.156) وهو مساوي لقيمة الصفر.

نموذج القياس لاختبار اوتيس لينون للقدرات العقلية

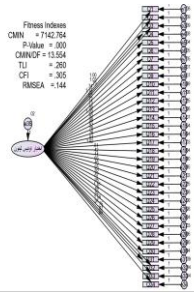
شكل 1

الطريقة الأولى (التقليدية ((0،1))



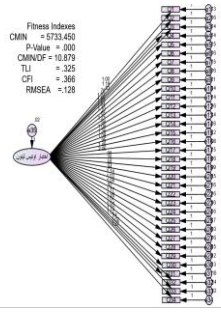
شكل 2

الطريقة الثانية (0، معامل الصعوبة حسب معادلة عودة)



شكل 3

الطريقة الثالثة (الوزن النسبي للمموه، 0):



بالاعتماد على نتائج السؤال الأول أظهرت النتائج أن أثر طريقة التصحيح لفقرات الاختبار من متعدد على خصائص المقياس (معامل الصعوبة ومعامل التمييز) باختلاف طرق التصحيح

وفقا للطرق (التقليدية) و (قيمة معامل الصعوبة) و (الوزن النسبي للمموه)، وذلك من خلال قراءة النتائج وفقاً للطريقة (التقليدية) التي تم تصحيح إجابات الطلبة بإعطاء الإجابة الصحيحة (١)، والإجابة الخاطئة (٠)، نلاحظ أن معاملات الصعوبة لفقرات اختبار اوتيس لينون للقدرات العقلية تراوحت بين (0.22) و(0.83) حيث ظهرت هناك فقرتين رقم (١. ١٠) سهله حيث بلغ معامل صعوبتها (٠.٨٣)، أما باقي الفقرات وقعت ضمن المدى (٠.٢٢) و (٠.٨٠) التي كان معامل صعوبتها متوسطة، وأن قيم معاملات التمييز المحسوبة لفقرات الاختبار تراوحت بين (0.20) و(0.61). وفقاً للطريقة الثانية تم تصحيح إجابات الطلبة بإعطاء الإجابة الصحيحة (قيمة معامل الصعوبة) المصححص باستخدام معادلة عودة (٢٠١٤)، والإجابة الخاطئة (٠). نلاحظ أن معاملات الصعوبة لفقرات اختبار اوتيس لينون للقدرات العقلية تراوحت بين (0.034) و(0.681) حيث لم تظهر أي فقرات سهلة أو صعبة وجميها وقع ضمن المدى، وأن قيم معاملات التمييز المحسوبة لفقرات الاختبار تراوحت بين (0.169) و(0.637). وفقاً للطريقة الثالثة تم تصحيح إجابات الطلبة بإعطاء الإجابة الصحيحة (١)، والإجابة الخاطئة (الوزن النسبي للمموه)، نلاحظ أن معاملات الصعوبة لفقرات اختبار اوتيس لينون للقدرات العقلية تراوحت بين (0.42)و(0.83) حيث ظهرت هناك فقرتين رقم (١. ١٠) سهله حيث بلغ معامل صعوبتها (٠.٨٣)، أما باقي الفقرات وقعت ضمن المدى (٠.٤٢) و (٠.٨٠) التي كان معامل صعوبتها متوسطة، وأن قيم معاملات التمييز المحسوبة لفقرات الاختبار تراوحت بين (0.23) و(0.62). وتدل هذه النتيجة على أن طرق التصحيح (التقليدية) و(الوزن النسبي للمموه) أظهرت وجود فقرات سهله بأن الطلبة لم تكن طريقة الإجابة تعود لأثر التخمين باختيار الإجابة الصحيحة، بينما طريقة التصحيح (قيمة معامل الصعوبة) المصححص باستخدام معادلة عودة (٢٠١٤) لم تظهر وجود فقرات سهله أو صعبة. وقد تشابهت هذه النتائج مع دراسة عودة (٢٠١٤) وقد أشارت نتائج الدراسة أن هناك توافقاً عالياً بين نتائج التقديرات لمعاملات الصعوبة المشتقة الصورة المقترحة للمعادلة والأساس النظري الذي ترتكز عليه، وهي معادلة توفيقية تجمع بين نظرية القرار ونظرية القياس.

بالاعتماد على نتائج السؤال الأول أظهرت النتائج بينت النتائج أن معاملات الثبات باستخدام معامل كرونباخ الفا للطرق الثلاثة للتصحيح؛ حيث أن معامل الثبات باستخدام طريقة التصحيح الأولى التقليدية فقد بلغ معامل الثبات (٠.٨٧٤). وفيما يتعلق بالطريقة الثانية معامل الصعوبة فقد بلغ معامل الثبات (٠.٨٧٢). وفيما يتعلق بالطريقة الثالثة فقد بلغ معامل الثبات (٠.٨٨٠) وتشير النتائج إلى أن قيم معاملات الثبات متقاربة بين طرق التصحيح الثلاثة، يمكن أن تعزى إلى التجانس بين مجموعات المفحوصين، حيث أنهم من نفس المستوى الدراسي، وتم

اختيارهم عشوائياً مما أدى إلى ارتفاع الفرق بين عينة المفحوصين، وبالتالي إلى تباين درجاتهم، مما انعكس على ارتفاع معامل الثبات في طرق التصحيح الثلاثة، وقد يعزى السبب إلى وضوح صياغة الأسئلة، مما ساعد الطلبة على فهم المقصود من السؤال. وتشير النتائج إلى تحقق الصدق العاملي لطرق التصحيح الثلاثة باستخدام التحليل العاملي، وتدل هذه النتيجة على مطابقة البيانات لمحتوى اختبار اوتيس لينون للقدرات العقلية، وهذا دليل على صحة البناء العاملي للاختبار على البيئة الأردنية. وقد تشابهت هذه النتائج مع دراسة أجرى البركات (٢٠٢٢) والتي أشارت إلى عدم وجود فرق ذات دلالة إحصائية بين معاملي ثبات الاتساق الداخلي (كرونباخ ألفا) لفقرات الاختبار يعزى لطريقة التصحيح.

التوصيات:

في ضوء ما توصلت إليه الدراسة من نتائج توصي بما يلي:

- إجراء المزيد من الدراسات لمقارنة أثر طرق التصحيح المختلفة بخصائص سيكومترية وحجم العينة.
- إجراء المزيد من الدراسات واستخدام التحليل العاملي التوكيدي للتأكد من استقرار البنية العاملية.
- إجراء المزيد من الدراسات باعتماد طرق تصحيح مختلفة عن الطريقة التقليدية (التقليدية (٠،١)، والتمييز (a، ٠).

المراجع

ألكين، لويس. (٢٠٠٧). الاختبارات والامتحانات "قياس القدرات والأداء"، (فرح السراج، ترجمة؛ ط١). شركة العبيكان للأبحاث والتطوير.

البركات، غيث علي. (٢٠٢٢). *ثبات الصدق العاملي في اختبارات الاختيار من متعدد بتغير طرق التصحيح المرتكزة على التخمين والمعرفة الجزئية*. اطروحة دكتوراه. جامعة اليرموك.

خصاونة، ابتسام عيسى. (٢٠١٢). *أثر اختلاف الأوزان النسبية لقواعد انتهاك صياغة فقرات الاختيار من متعدد في الاختبارات على خصائصها السيكومترية* [رسالة دكتوراه غير منشورة]. جامعة اليرموك.

ربيع، هادي، وأحمد، مشعان. (٢٠٠٨). *القياس والتقويم في التربية والتعليم*. دار زهران للنشر والتوزيع.

رينولدز، سيسيل وليفقنستون، رونالد (٢٠١٣). إتقان القياس النفسي الحديث. ترجمة صلاح الدين محمود علام (ط١). دار الفكر.

الشايب، عبد الحافظ. (٢٠٠٧). تقدير ثبات علامات عينة من المواد في جامعة آل البيت. *مجلة جامعة دمشق، ٢٣*(٢).

الظاهر، زكريا محمد؛ وتمرجيان، جاكلين؛ عبد الهادي، جودت عزت؛ والمنيزل، عبد الله. (٢٠٠٢). مبادئ القياس والتقويم في التربية (ط٢). الدار العلمية الدولية للنشر والتوزيع ودار الثقافة للنشر والتوزيع.

عبد الله، صهيب حمدي. (٢٠١٩). قياس أثر التخمين في اختبارات الاختيار من متعدد في مادة الرياضيات للصف التاسع الأساسي في ضوء النظرية الكلاسيكية في القياس. مجلة جامعة عمان العربية للبحوث، ١ (٢) ١١٧-١٣٠.

العجلوني، جهاد. (2016). *فحص الأداء التفاضلي لمموهات فقرات اختبار الرياضيات في التقييم الوطني الأردني لاقتصاد المعرفة وفقا لجنس الطالب وموقع المدرسة* [رسالة دكتوراه غير منشورة]. جامعة اليرموك

عودة، أحمد سليمان. (٢٠١٤). تصحيح معاملات صعوبة الفقرات لأثر التخمين في أسئلة الاختيار من متعدد: صورة معدلة لمعادلة جيلفورد. مجلة الدراسات التربوية والنفسية جامعة السلطان قابوس، ٨ (٢)، ٢٤٨-٢٥٧.

عودة، أحمد. (2022). *القياس والتقويم في العملية التدريسية*. دار الأمل للنشر والتوزيع.

عودة، أحمد. (2022). التصحيح لأثر التخمين في اختبارات الاختيار من متعدد بين المعرفة الجزئية والمخاطرة: نموذج مقترح لتدريج ثنائي الاتجاه. *مجلة اتحاد الجامعات العربية للبحوث في التعليم العالي، ٤٢ (١).*

عيلبوني، سمير. (٢٠٠٧). *أثر عدد البدائل وتمييز المموهات في فقرات الاختيار من متعدد على التوافق مع نموذج ثلاثي المعلم* [رسالة دكتوراه غير منشورة]. جامعة اليرموك.

Aiken, L. (2003). *Psychological testing and assessment*. Boston: Pearson Education Group.

Crocker. L. & Algina, T. (1986). *Introduction to classical and modern test theory*. New York: university of florida, U.S.A. Holt. Rinehart and Winston.

Dodeen, H. (2005). Using Distractors in Correcting for Guessing in Multiple-Choice Tests. *Dirasat, Educational Sciences*, *2*(1), 2005.

Frey B, Bruce., Petersen, S., Edwards, M., Lisa, Pedrotti T. Jennifer and Peyton V. (2005). Item – Writing Rules: Collective Wisdom Department of Psychology and Research in education. *School of Education, University of Kansas, Lawrence, KS.*

Frisbie, D.A. (1973). Multiple – choice versus true -false: A Comparison of reliability and Concurrent validities. *Journal of Education Measurement, 10*(4), 297-304.

Gay, L. R. (1980). *Educational Evaluation and Measurement*, Ohio Chales, E. Merrill.

Gronlund, N. & Linn, R. (1990). *Measurement and evaluation in teaching.* New York.

Hopkins, C.D. and Antes, *R. L.* (1978). *Classroom Measurement and Evaluation*. Itasca: F.E. Peacock Publishers.

Howe, K. (1995). Validity, Bias, and Justice in Educational Testing: *The Limits of the consequetialist conception*. Retrieved January 14, 2009, form http://www.ed.uiuc.edu/eps/pes-year book //as-does / howe -html.

Lau P. N. K., Lau S. H., Hong K. S. & Usop H. (2011). Guessing, Partial Knowledge, and Misconceptions in Multiple-Choice Tests. Educational Technology & Society, 14 (4), 99–110.

Marshall, J. C. and Hales, L.w. (1971). *Classroom test Construction Reading. Massachusetts*: Addison – Wesley.

Murphy, K.R and Hofer, c.o. (1994). *Psychological Testing Principles and Application*. Third edition. New jersey.

Osterlind, S. J. (2006). *Modern measurement. Upper Saddle River. NJ*: Pearson Education. Inc

Otis, A.S. & Lennon R.T. (1969). Otis - *Lennon mental ability test: Technical handbook.* New York.

Penfield, R. D. (2008). An odds ratio approach for assessing differential distractor functioning effects under the nominal response model. *Journal of Educational Measurement, 45(3), 247-269.*

Yaşar, M., Kula Kartal, S., & Aybek, E. C. (2021). Scoring methods for multiple choice tests: How does the item difficulty weighted scoring change student's test results? *Bartın University Journal of Faculty of Education, 10*(2), 309-324. https://doi.org/10.1016/buefad.878504