أثر بعض الطرق الوزنية على الخصائص السيكومترية لاختبار رياضيات في ضوء النظرية الكلاسيكية والنظرية الحديثة

The effect of some gravimetric methods on the psychometric Properties of a mathematics test In view of of classical theory and modern theory

إعداد

فيحان عبدالله سعود الدوسري

Faihan Abdullah Saud Al-Dosari

باحث دكتوراه - جامعة الملك سعود

أ.د/ عبد الرحمن سليمان الطريري

Prof. Dr. Abdul Rahman Suleiman Al-Tariri

أستاذ القياس والتقويم - جامعة الملك صعود

Doi: 10.21608/jasep.2022.247368

استلام البحث : ٢٦ / ٢ /٢٠٢٢

قبول النشر: ٨ / ٣ / ٢٠٢٢

الدوسري ، فيحان عبدالله سعود والطريري ، عبدالرحمن سليمان (٢٠٢٢). أثر بعض الطرق الوزنية على الخصائص السيكومترية لاختبار رياضيات في ضوء النظرية الكلاسيكية والنظرية الحديثة . ***المجلة العربية للعلوم التربوية والنفسية،*** المؤسسة العربية للتربية والعلوم والآداب، مصر، مج (٦)،ع(٢٨) يوليو ، ٥٣ – ١١٦.

http://jasep.journals.ekb.eg

أثر بعض الطرق الوزنية على الخصائص السيكومترية لاختبار رياضيات في ضوء النظرية الكلاسيكية والنظرية الحديثة

المستخلص:

هدفت هذه الدراسة إلى الكشف عن أثر الطرق الوزنية ودرجة القطع في الخصائص السيكومترية، ولتحقيق أهداف الدراسة تم بناء اختبار تحصيلي في مقرر الرياضيات للصف السادس الابتدائي، تكون في صورته النهائية من (٢٥) فقرة، واستخدم المنهج الوصفي المقارن، وتكون مجتمع الدراسة من (٢٦٧٢) طالباً، وتألفت العينة من (٥٢١) طالب، تم اختيارها بأسلوب المعاينة العشوائية البسيطة، وتم تقدير الدرجات من خلال أربع طرق وزنية، وهي: الطريقة العادية (الدرجة الخام)، وطريقة الدلتا سكور، وطريقة ألفا سكور، والنموذج الثنائي (طريقة التوقع البعدي)، وتم استخدام المتوسطات الحسابية، ومعاملات الصعوبة والتمييز، والتحليل العاملي الاستكشافي، للتحقق من الصدق العاملي وافتراض أحادية البعد، ومعاملات الثبات، ومعادلة فيلدت لمعرفة الفروق بين معاملات الثبات، ومعامل الصدق المرتبط بمحك، واختبار مربع كاي، ومعاملات الارتباط، واستخدام تحليل التباين للقياسات المتكررة، والخطأ المعياري بطريقة لورد، وأجرى تحليل البيانات باستخدام البرامج الإحصائية (SPSS)، و(Itemman)، و (Xcalibre)، و (BilogMG3). وأظهرت نتائج الدراسة الآتي: تفوق طريقة التوقع البعدي على الطرق الوزنية الأخرى حيث بلغت قيمة الخطأ المعياري بطريقة لورد (١.٦٣) وهو أقل قيمة تليها طريقة الدلتا (١.٧٨).

الكلمات المفتاحة: الطريقة الوزنية – الخصائص السكومترية – اختبار الرياضيات.

Abstract:

This study aimed to reveal the effect of the gravimetric methods and the degree of cut-off on the psychometric properties, and to achieve the objectives of the study, an achievement test was built in the mathematics course for the sixth grade of primary school.) students, and the sample consisted of (521) students, chosen by simple random sampling method, and grades were estimated through four weighted methods, namely: the normal method (raw score), the delta-score method, the alpha-score method, and the binary model (post-expectation method). arithmetic averages, difficulty and discrimination coefficients, and exploratory factor analysis were used to verify the factorial validity and the one-dimensional assumption, reliability coefficients, and Feldt's equation to find out the differences

between the reliability coefficients, the validity coefficient associated with a test, chi-square test, correlation coefficients, and the use of analysis of variance for measurements The data were analyzed using the statistical programs (SPSS), (Itemman), (Xcalibre), and (BilogMG3). The results of the study showed the following: The dimensional prediction method is superior to the other weighted methods, as the standard error value of the Lord's method is (1.63), which is the lowest value, followed by the delta method (1.78).

Key words: gravimetric method, scometric properties, mathematics test.

مقدمة:

نظريات القياس (الكلاسيكية، التعميم، الحديثة) تحاول الوصول إلى قياس ما يمتلكه المفحوص من السمة أو الخاصية المقاسة بدقة، لذا تعددت النماذج القياسية بهدف التقليل من أخطاء القياس والوثوق بالنتائج التي يتم الحصول عليها من أدوات القياس.

حيث إنَّ الهدف الأسمى والغاية الكبرى من تطبيق المقاييس والاختبارات، هو سيادة العدل ومحاولة تحقيقه في مجالات الحياة المتعددة. والعدل من الممكن أن يكون على مستوى الفرد ومن الممكن أن يكون على مستوى المجتمع برمته، فالقياس يمكن من خلاله وضع الفرد المناسب في المكان المناسب كما أنه بالقياس يمكن المحافظة على حقوق الأفراد المادية والمعنوية (الطريري، ٢٠١٤).

ومن زاوية أخرى فإن وزارة التعليم تولي اهتمام بالاختبارات الدولية (TIMSS)، ومن الأدوات المستخدمة في الاختبارات الدولية (TIMSS) اختبار في مادة الرياضيات، وهذا يشير إلى أن مادة الرياضيات مكون رئيس في الاختبارات الدولية.

ونظراً للاهتمام المستمر بتطوير أساليب القياس والتقويم يجب أن يكون هنالك إعادة نظر بصورة مستمرة في أدوات القياس وتطويرها ومراعاة الشروط والأُسس الفنية الواجب توافرها عند بنائها لكي يعتمد على نتائجها بدرجة أكبر من الثقة. وعادة تستخدم درجات الاختبار لاستخلاص استدلالات على سلوك المفحوصين فيما وراء العملية الاختبارية. وهذا يشير إلى أن الصدق والثبات من أهم شروط الاختبار الجيد.(كروكر و الجينا، ١٩٨٦/ ٢٠١٧).

وقد اهتمت كثير من الدراسات والبحوث السيكومترية بالتوصل إلى طرق وأساليب يمكن باستخدامها تحسين وزيادة صدق درجات الاختبار وثباته، حيث قام بعض المشتغلين بالقياس النفسي والتربوي بدراسة صدق وثبات درجات الاختبار في حالة استخدام طرق لتقدير درجات الاختبار تختلف عن طريقة التقدير التقليدية (الدرجة الخام) (عبدالوهاب، ٢٠٠١) .

حيث تطرقت دراسة غنيم (١٩٨٦) إلى مجموعة من الدراسات التي تناولت هذه الطرق (التقليدية، ديفز، ميشيل، اللوغاريتمية، التربيعية، جتمان)، في محاولة للبحث عن أفضل طريقة لتقدير درجات الاختبار، وبينت الدراسة أنَّ بعض هذه الطرق تحتاج إلى جهد ووقت وتكلفة أطول من الطريقة التقليدية ومع ذلك كان هنالك تباين في نتائج هذه الدراسات من حيث ارتفاع معاملات ثبات درجات الاختبار المقدر بالطرق المختلفة عن الطريقة التقليدية، والباحث المتتبع لطرق تقدير درجات الاختبار يجد أنها عديدة ومتنوعة.

مشكلة الدراسة:

مادة الرياضيات من المواد الأساسية والمهمة في جميع المراحل الدراسية، كما أن الاختبارات من أهم أدوات قياس مستوى تحصيل الطلاب، وفي الغالب عند تقدير درجات الاختبار يتم احتساب الإجابة الصحيحة بالقيمة(١)، والإجابة الخاطئة بالقيمة (٠) بغض النظر عن مضمون السؤال ومدى صعوبته وتمييزه بين الطلاب، وفي ذات السياق فإن الطرق الوزنية في النظرية التقليدية، وفي النظرية الحديثة من أهم المواضيع التي شغلت المهتمين بعملية القياس، للبحث عن أفضل طريقة تصف موقع المفحوص، وتضيف لدرجات الاختبار ثباتاً وصدقاً مرتفعاً، وتحدد ما يمتلكه المفحوص من معلومات بقيمة عددية.

حيث بينت الدراسات التي أُجريت لمقارنة هذه الطرق بعضها بالبعض الآخر، وجود اختلاف في دقة تقدير قدرات الأفراد يعزى للطريقة الوزنية المستخدمة حيث أن بعض هذه الطرق يؤدي إلى زيادة في ثبات وصدق درجات الاختبار ومنها دراسة: (Garee & Vermint, 2006؛ اللحياني، ١٤٣٠؛ الخرشة، ٢٠١٦). بينما أشارت دراسات أخرى إلى عدم وجود اختلاف في دقة تقدير قدرات الأفراد يعزى للطريقة الوزنية المستخدمة ومنها دراسة: (النعيمي، ٢٠١٥؛ اليامي،٢٠١٨).

أسئلة الدراسة:

تتلخص مشكلة الدراسة في الإجابة على السؤال الرئيس التالي:

هل تختلف الخصائص السيكومترية لدرجات الاختبار باختلاف الطريقة الوزنية (التقليدية، الدلتا سكور، ألفا سكور، التوقع البعدي)؟

أهداف الدراسة:

١- التعرف على الخصائص السيكومترية لدرجات الاختبار الموزون بالطرق الآتية (التقليدية، الدلتا سكور، ألفا سكور، التوقع البعدي).

٢- المقارنة بين الخصائص السيكومترية لدرجات الاختبار الموزون بالطرق الآتية (التقليدية، الدلتا سكور، ألفا سكور، التوقع البعدي).

أهمية الدراسة:

الأهمية النظرية:

تكمن أهمية الدراسة الحالية في الموضوع الذي تتناوله، فالحاجة إلى طريقة تقدير تجعل درجة الاختبار تتصف بصدق وثبات مرتفع وتجعل منها معنى أصبحت ضرورة ملحة لدى

العديد من المؤسسات التي تعد الاختبار أو تستخدم الاختبارات كأحد وسائل القياس، في ظل تنوع وتعدد طرق التقدير، سواءً باستخدام النظرية التقليدية، أو النظرية الحديثة. كما تكمن أهمية الدراسة الحالية في الموضوع الذي تتناوله من حيث استخدام طرق وزنية جديدة تعتمد على معامل التمييز (ألفا سكور) لتقدير درجات الاختبار

الأهمية التطبيقية:

تكمن الأهمية التطبيقية لهذه الدراسة في إمكانية الاستفادة من فقرات اختبار الرياضيات التحصيلي المعد في إعداد أو تحديث بنك اسئلة لمادة الرياضيات للصف السادس الابتدائي، وحث مستخدم الاختبار على التحقق من الخصائص السيكومترية لفقرات الاختبار.

حدود الدراسة:

الحدود الموضوعية: تتحدد الدراسة الحالية بموضوعها المتمثل في أثر الطرق الوزنية وتحديد درجة القطع في الخصائص السيكومترية.

الحدود المكانية: أُجريت الدراسة الحالية في محافظة الاحساء (منطقة الهفوف).

الحدود الزمانية: الفصل الدراسي الثاني لعام ١٤٤٢هـ.

الحدود البشرية: طلاب الصف السادس الابتدائي.

مصطلحات الدراسة:

- الطرق الوزنية weighting methods: نظم معينة لإعطاء الدرجات وتقديرها حيث تختلف فيها درجة كل مفردة، ولا يشترط أن تكون جميع درجات مفردات الاختبار متساوية (Jaradat & Tollefson, 1988).

التعريف الإجرائي للطرق الوزنية: الدرجة التي يحصل عليها المفحوص باستخدام الطرق المستخدمة بهذه الدراسة وهي (الطريقة التقليدية، وطريقة معامل السهولة (الدلتا سكور)، وطريقة معامل التمييز (ألفا سكور)، وطريقة التوقع البعدي). نجوف المعدلة.

- الخصائص السيكومترية Psychometric Properties:

١- الصدق Validity: عرفه كرونباخ بالعملية التي يجمع من خلالها مطور الاختبار الأدلة التي تدعم أنواع الاستدلالات التي سيتم استخلاصها من درجات الاختبار (Cronbach, 1970).

٢- الثبات Reliability: وتعرف انستازي الثبات بأنه الاتساق في الدرجات التي يحصل عليها من نفس الأفراد بتطبيق نفس الاختبار (Anastasi, 1982).

٣- صعوبة الفقرة Item Difficulty: نسبة الأفراد الذين أجابوا عن الفقرة إجابة صحيحة، عندما تكون الفقرات ثنائية التصحيح (كروكر والجينا،١٩٨٦/ ٢٠١٧).

٤- تمييز الفقرة Item Discrimination : النسبة التي تميز فيها الفقرة بشكل صحيح بين المفحوصين في السلوك الذي صمم الاختبار لقياسه (Anastasi, 1982).

التعريف الإجرائي للخصائص السيكومترية: القيم التي يتم الحصول عليها من معامل صعوبة الفقرات، ومعامل تمييز الفقرات، ومعاملات الصدق، ومعاملات الثبات.

الإطار النظري:

الطرق الوزنية

مفهوم القياس:

يختلف القياس في العلوم التربوية والنفسية عن القياس في العلوم الطبيعية لأسباب جوهرية أهمها أن طبيعة السلوك الإنساني وطبيعة الموضوعات التربوية كالتحصيل، هي موضوعات أكثر تعقيداً من الظواهر الطبيعية ولا يسهل قياسها كما في العلوم الطبيعة أما لعدم ثباتها أو لارتباطها بظواهر وعوامل يصعب عزلها بشكل مستقل لقياسها، لذلك تكون أغلب القياسات التربوية نسبية وليست مطلقة ومعرضة للخطأ أكثر مما في العلوم الطبيعية. ولا يكون القياس التربوي أو النفسي مباشراً كما في العلوم الطبيعية أنما هو استدلالي شأن أغلب الظواهر السلوكية. وهذا يتطلب أساليب متطورة ودقيقة للقياس تتوافر فيها صفات الصدق والثبات والقدرة على التمييز في المستويات الافتراضية. (مجيد، ٢٠١٤).

لذا تعدد نظريات القياس (الكلاسيكية- التعميم- الحديثة)، وسنتطرق إلى أشهر هذه النظريات بصورة مختصرة واهم افتراضاتها:

١- نظرية القياس الكلاسيكية (Classical Test Theory) :

ويطلق عليها نظرية الدرجة الحقيقية حيث ان الدرجة الملاحظة تساوي الدرجة الحقيقية مضافاً إليه درجة الخطأ، ومن افتراضاتها:

أ- الارتباط بين الدرجات الحقيقية ودرجات الخطأ لمجتمع من المفحوصين يساوي صفر.

ب- متوسط درجات الخطأ لمجتمع من المفحوصين يساوي صفر.

ت- عندما يتقدم مفحوصون لاختبارين منفصلين، وبافتراض أن درجات كل مفحوص على الاختبارين تم اختيارها عشوائياً من توزيعين مستقلين للدرجات الملاحظة الممكنة، فإن الارتباط بين درجات الخطأ للاختبارين يساوي صفر(رينولدز وليفنجستون، ٢٠١٣).

٢- نظرية الاستجابة للمفردة (Item Response Theory):

ومن مسمياتها نظرية السمات الكامنة، ونظرية القياس الحديثة، ونظرية المنحنى المميز للمفردة. ومن افتراضاتها:

أ- أحادية البعد (Unidimensnsioal):

تفترض النظرية وجود قدرة أو سمة واحدة فقط يمكن أن تفسر أداء المفحوص في الاختبار.

ب- الاستقلال الموضعي (Local Independence):

تقدير معالم المفردة لا يعتمد على تقديرات معالم المفردات الأخرى المكونة للاختبار. ولا على قدرات الأفراد الذين يجيبون عليها، ولا يعتمد تقدير قدرة الفرد على تقديرات قدرة أي مجموعة من الأفراد الذين يؤدون الاختبار أو على تقديرات صعوبة الفقرات.

جـ- المنحنى المميز للمفردة (Item Characteristic Curve) :

وجود دالة مميزة خاصة بكل مفردة على حدة يتخذ كل منها شكل منحنى الترجيح اللوغاريتمي الاحتمالي.

د- التحرر من السرعة (Speediness):
بمعنى أن الزمن ليس له دور في الإجابة على مفردات الاختبار.
وتعدد هذه النظريات يهدف إلى إعطاء المفحوص قيمة أو درجة تمثل ما يمتلكه من السمة المقاسة. وعندما يتم استخدام الاختبارات ذات الاختيار من متعدد، ومع تعدد نظريات القياس تعددت الطرق الوزنية التي تسعى إلى أفضل تقدير لقدرات المفحوصين، أو إعطاء قيمة أو درجة تمثل ما يمتلكه من السمة المقاسة (بيكر، ٢٠١٠؛ كروكر والجينا، ٢٠١٧).
الطرق الوزنية:
هناك العديد من الطرق الوزنية التي تختلف باختلاف النظرية، واختلاف إجراءاتها وأساليب معالجتها، حيث تعد الطرق الوزنية القاعدة التي يعطى في ضوئها قيماً كمية تعكس الدرجة المستحقة للطالب.
الطرق الوزنية في النظرية الكلاسيكية:
جدول (١): تصنف الطرق الوزنية في النظرية الكلاسيكية حسب طرق معالجتها

الأسلوب	الطريقة
التقليدي (العادي)	الدرجة الخام
تقدير الخبراء	طريقة ديفز، طريقة بوك
القيم الوزنية للخصائص السيكومترية للفقرات	طريقة الدلتا، طريقة ألفا
التغذية الراجعة	طريقة الدرجات الذاتية
الثقة و معالجة التخمين	طريقة جتمان ، طرق الثقة.

١- الدرجة الخام (Normal Method): تعتمد هذه الطريقة على إعطاء الإجابة الصحيحة واحد وبقية الإجابات الخاطئة صفر، وتستخدم هذه الطريقة عندما تكون إحدى الاختيارات صحيحة والباقي خطأ، وهذه الطريقة هي السائدة في تصحيح الاختبارات. وتم استخدامها في هذه الدراسة.
٢- طريقة ديفز وتسمى الدرجة الوزنية للاختيار (Choice weight Scoring):
تعتمد هذه الطريقة على صعوبة وسهولة الأسئلة بالإضافة إلى جاذبية المموهات، ويتم تحديد ذلك بواسطة المحكمين وواضع الاختبار بحيث لا تكون درجات الأسئلة متساوية، بل تحدد السؤال على حسب صعوبته ولا يشترط ان تأخذ الإجابة الصحيحة درجة موجبة وتأخذ الإجابة الخاطئة درجة سالبة.
٣- طريقة بوك (Bock method) : تعتمد هذه الطريقة على عرض فقرات الاختبار على مجموعة من الخبراء ويطلب منهم ترتيب الإجابات الخاطئة ترتيباً تنازلياً على أساس القرب من الإجابة الصحيحة، وذلك لكل مفردة من مفردات الاختبار. وبعد الانتهاء من ذلك الترتيب تجمع الآراء ويحسب وزن لكل بديل خاطئ في كل فقرة، بحيث يأخذ البديل الخاطئ القريب من الإجابة الصحيحة درجة أعلى من البدائل الأخرى (غنيم، ١٩٨٦).

٤- طريقة معامل السهولة (الدلتا سكور) Ease coefficient method (D- Scoring):
عندما تكون الفقرة ثنائية التصحيح فإن معامل صعوبة الفقرة هو نسبة المفحوصين الذين أجابوا على الفقرة إجابة صحيحة من أجمالي العينة.
D_s = ١ في حالة تقديم إجابات صحيحة لكل الفقرات.
من العرض السابق يتضح أنه إذا اجاب الطالب على الفقرة إجابة صحيحة فإنه يأخذ قيمة معامل السهولة لهذه الفقرة، وإذا أجاب الطالب على الفقرة إجابة خاطئة فإنه يحصل على الدرجة (٠). ويتم ايجاد مجموع قيم معاملات السهولة لجميع فقرات الاختبار ثم تقسم على اجمالي قيم معاملات السهولة، وتعبر عن الدرجة الكلية للطالب. ولتوضيح ذلك نستعرض المثال التالي:
اختبار يتكون من خمس فقرات، ومعاملات السهولة على الترتيب (٠,٢ ، ٠,٣٥ ، ٠,٥ ، ٠,٦٥، ٠,٨)، طبق على أربعة طلاب، وكانت إجاباتهم على النحو التالي:
جدول (٢): مثال توضيحي لتقدير درجات الطلاب وفق طريقة الدلتا سكور

الطالب	X_{s1} δ1=0.2	X_{s2} δ2=0.35	X_{s3} δ3=0.5	X_{s4} δ_4=0.65	X_{s5} δ_5=0.8	$\sum_{i=1}^{5} Xsi\delta i$	D_s
1	0	0	0	0	0	0	0
2	1	1	0	1	0	1.2	0.48
3	1	0	1	0	1	1.5	0.6
4	1	1	1	1	1	2.5	1

يتضح من الجدول السابق أن الطالب الأول حصل على الدرجة صفر بسبب أنه لم يجب على أي فقرة إجابة صحيحة، والطالب الرابع حصل على الدرجة الكاملة (١) لأنه أجاب على جميع الفقرات إجابة صحيحة، بينما أجاب الطالب الثاني والطالب الثالث على ثلاثة فقرات إجابة صحيحة، وعند استخدام الطريقة التقليدية فإنهما سيحصلان على نفس الدرجة (٣)، ولكنهما حصلا على درجات مختلفة عند استخدام طريقة الدلتا سكور (D_2=0.48, D_3=0.6). بسبب متجه الاستجابات المختلفة (Dimitrov, 2016). وتم استخدامها في هذه الدراسة.

٥- طريقة معامل التمييز (ألفا سكور) Discrimination coefficient method (α- Scoring): يتم تحديد المجموعة ذات الدرجات العالية من بين العينة والذين أجابوا إجابة صحيحة على الفقرة، وكذا تحديد المجموعة ذات الدرجات المنخفضة من بين العينة والذين أجابوا إجابة صحيحة على الفقرة، وقد درج القياسيون على استخدام نسب متعددة لتحديد أفراد المجموعتين، إلا أن الأكثر شيوعاً في الاستخدام نسبة ٢٧%. اختبار يتكون من خمس فقرات، ومعاملات التمييز على الترتيب (٠,٣ ، ٠,٣٥ ، ٠,٥ ، ٠,٦٥، ٠,٤٥)، طبق على أربعة طلاب، وكانت إجاباتهم على النحو التالي:

جدول (٣): مثال توضيحي لتقدير درجات الطلاب وفق طريقة ألفا سكور

الطالب	X_{s1} α1=0.3	X_{s2} α2=0.35	X_{s3} α3=0.5	X_{s4} $_4$=0.65 α	X_{s5} $_5$=0.45 α	$\sum_{i=1}^{5} Xsi\alpha i$	α_s
1	0	0	0	0	0	0	0
2	1	1	0	1	0	1.3	0.58
3	1	0	1	0	1	1.25	0.56
4	1	1	1	1	1	2.25	1

يتضح من الجدول السابق أن الطالب الأول حصل على الدرجة صفر بسبب أنه لم يجب على أي فقرة إجابة صحيحة، والطالب الرابع حصل على الدرجة الكاملة (١) لأنه أجاب على جميع الفقرات إجابة صحيحة، بينما أجاب الطالب الثاني والطالب الثالث على ثلاثة فقرات إجابة صحيحة، وعند استخدام الطريقة التقليدية فإنهما سيحصلان على نفس الدرجة (٣)، ولكنهما حصلا على درجات مختلفة عند استخدام طريقة ألفا سكور (α_2=0.58 , α_3=0.56). بسبب متجهات الإجابات المختلفة. وتم استخدامها في هذه الدراسة.

٦- طريقة الدرجات الذاتية (Self- Scoring Method):

تعود هذه الطريقة في الأصل إلى برسي (١٩٢٦) فهو الذي ابتكر هذه الطريقة وكانت تعرف حينذاك بأسلوب الدرجات الذاتية، وفي هذه الطريقة يمكن للمفحوص الاستمرار في الاستجابة على كل سؤال من أسئلة الاختبار بأسلوب أقرب ما يكون للحياة العادية، حيث تشير التغذية الراجعة في النهاية إلى النجاح.

في حالة إعطاء التغذية الراجعة المصاحبة لهذه الطريقة كان يطلب من المفحوص أن يستمر في الاستجابة على كل سؤال من أسئلة الاختبار حتى يصل إلى الإجابة الصحيحة، حيث يتم تعريفه بإجابته على كل سؤال ما إذا كانت صحيحة أم خاطئة، فإذا كانت اجابته خاطئة يطلب منه الاستمرار في الإجابة حتى يصل إلى الإجابة الصحيحة، أي يقال له : (استمر في الإجابة حتى تصل إلى الصواب) .

وتقدر الدرجة للمفحوص في السؤال على أساس المعادلة الآتية:

درجة المفحوص في السؤال =عدد البدائل الموجودة في السؤال – عدد المحاولات للوصول إلى الإجابة الصحيحة.

فإذا وصل المفحوص في أحد الأسئلة المكونة من أربعة بدائل اختيارية إلى الإجابة الصحيحة من أول محاولة فإنه بذلك يحصل على الدرجة ٣ حيث أن : درجة المفحوص في السؤال = ٤ - ١ = ٣

وإذا وصل إلى الإجابة الصحيحة في المحاولة الثانية فإنه يحصل على الدرجة ٢، أما إذا وصل إليها في المحاولة الثالثة فإنه يحصل على الدرجة ١، وإذا وصل إليها في المحاولة الرابعة فإنه يحصل على الدرجة صفر.

وأثبت برسي (١٩٥٠) انه يحدث تقدم في تعلم الطلاب عندما يتم تطبيق الاختبارات ذات الاختيار من متعدد بواسطة الإجراءات المسماة الدرجات الذاتية، حيث أن التغذية الراجعة الفورية المصاحبة لهذه الطريقة تعمق التعلم وتزيد منه ويكون مصاحباً لها (Rippey, 1970).

٧- الطريقة الوزنية لجتمان (Getman's Weight Method): تستخدم هذه الطريقة استجابات المفحوصين، حيث يكون لكل بديل قيمة معينة مستنتجة من المفحوصين الذين اختاروا هذا البديل وذلك بمعرفة درجاتهم في بقية الأسئلة وتحديد وزن يمثل متوسط درجاتهم على بقية الأسئلة حيث يعتبر هذا المتوسط هو القيمة الوزنية للبديل المختار من قبل هؤلاء المفحوصين. ولإيضاح ذلك يمكن عرض هذه الطريقة في الخطوات التالية:

أ- يتم تحديد المفحوصين الذين وقع اختيارهم على بديل معين في فقرة من فقرات الاختبار. وتعزل أوراق إجابات هؤلاء المفحوصين.

ب-يتم التغاضي عن إجابة الفقرة سواء كانت صحيحة ام خاطئة، وتحسب درجة كل مفحوص في الاختبار عن طريقة المعادلة التالية :

درجة المفحوص = عدد الإجابات الصحيحة – (عدد الإجابات الخاطئة ÷ عدد البدائل -١).

جـ- يتم حساب المتوسط الحسابي للدرجات الناتجة من الخطوة السابقة، ويمثل هذا المتوسط قيمة وزنية للبديل (غنيم، ١٩٨٦).

٨- طرق الثقة : بحكم أن النظرية الكلاسيكية تعتمد على الدرجة الحقيقية ومحاولة الوصول إليها وتقديرها بتقليل نسبة الخطأ، حيث أن الدرجة الملاحظة = الدرجة الحقيقية + الخطأ، وبما أن التخمين يعتبر من مصادر الخطأ فهنالك عدة طرق تحاول التقليل من التخمين (الخطأ العشوائي)، ولكنها تتطلب إجراءات معينة أثناء تطبيق الاختبار ذات الاختيار من متعدد (Rippey, 1970).

حيث يطلب من المفحوصين وضع نسب مئوية أو درجات تعبر عن تأكدهم من صحة الإجابات التي يختارونها، ويختلف هذا الإجراء باختلاف حالة بدائل الإجابة المطروحة، فإذا كانت احدى هذه البدائل صحيحة والباقي خاطئ تماماً يُطلب من المفحوص أن يعطي نسبة مئوية أو درجة تعبر عن ثقته في الإجابة التي اختارها، اما إذا كانت هذه البدائل متدرجة في صحتها يطلب من المفحوص في هذه الحالة أن يوزع النسب المئوية على هذه البدائل، حيث تعبر النسبة المعطاة لكل بديل عن مقدار صحته من وجهة نظر المفحوص بشرط أن يكون مجموع هذه النسب يساوي ١٠٠%، أي أنه في حالة الانتقاء المطلق يُعطي المفحوص نسبة واحدة على الإجابة المختارة تعبر عن ثقته في صحتها، اما في حالة الانتقاء النسبي يوزع المفحوص النسب المئوية على البدائل كل حسب درجة قربها من الصحة من وجهة نظر المفحوص ويكون مجموع هذه النسب ١٠٠% (العدل، ١٩٨٦).

وتتناول طرق الثقة المختلفة بعد ذلك الإجراء معالجة نسب أو درجات الثقة معالجة رياضية سواء في حالة الانتقاء النسبي أو المطلق، ومن هذه الطرق:

١- الطرق اللوغاريتمية Logarithmic Methods :
تعتمد هذه الطرق على أساس معالجة نسب الثقة المعطاة على البدائل بواسطة اللوغاريتمات، إلا أن هذه المعالجة تختلف من طريقة إلى أخرى:
أ- طريقة ميشيل:
وضعها ميشيل Michael (1968) وهي تصلح للبدائل المتدرجة في درجة صحتها والتي فيها تعطى درجة الثقة على حسب درجة صحة البديل بحيث توزع درجات الثقة على جميع البدائل ويكون مجموعها مساوياً ١٠٠% فإذا كان هنالك سؤالاً ما مكون من أربع بدائل أ، ب، جـ، د. فإنه يطلب من المفحوص إعطاء درجة من الثقة تعبر عن مقدار صحة كل بديل من تلك البدائل الأربع بحيث يكون مجموع النسب مساوياً لمئة كأن يكون كالاتي:
٢٠% ل(أ) ، ٣٥ %لـ (ب) ، ١٥% لـ (جـ) ، ٣٠%لـ (د)
في هذه الحالة تستخدم المعادلة الآتية لحسابة درجة السؤال:
الدرجة = مجـ ن (١+ لو ح (س))
حيث : ح (س) تعبر عن النسبة المئوية المعطاة للاختيار (س)، ن عدد بدائل السؤال
بشرط أن : ٠.١ ≤ ح (س)≤ ١
فإذا أراد حساب درجة السؤال التي لها النسب المذكورة، فإن ذلك يكون كالاتي:
الدرجة = (١+ لو ٠.٢) + (١+ لو ٠.٣٥) + (١+ لو ٠.١٥) + (١+لو ٠.٣)=
١.٤٧ =٠.٤٧+٠.١٧+٠.٥٤+٠.٣
ويعاب على طريقة ميشيل أنها لم تضع أي اعتبار للإجابة الصحيحة، ولكن اكتفت بتوزيع درجات الثقة على البدائل المختلفة، حيث خلت المعادلة المستخدمة لحساب درجة السؤال من وجود أي تأثير لدرجة الثقة المحددة للإجابة الصحيحة على تلك المعادلة (Michael, 1968).
ب-الطريقة الثانية:
تعتمد هذه الطريقة على الإجابة الصحيحة فقط، وبالتالي فهي تصلح للانتقاء المطلق أفضل من صلاحيتها للانتقاء النسبي.
فإذا رمزنا للتوقع الاحتمالي المعطى للإجابة الصحيحة بالرمز ح (ق) فإنه في هذه الطريقة يتم تقدير درجة أي سؤال بواسطة الدالة الآتية:

$$\text{الدرجة اللوغاريتمية (ل)} = \frac{1}{2}(٢ + \text{لو}_{١٠}\ \text{ح(ق)} \quad \text{إذا كانت ح(ق)} \leq ٠.٠١$$

صفر إذا كانت ح (ق) > ٠.٠١
ففي المثال المعطى للطريقة السابقة إذا كانت الإجابة الصحيحة هي (د) فإن ح(ق) = ٠.٣
وبالتالي تكون الدرجة اللوغاريتمية = $\frac{1}{2}$ (٢+ لو$_{١٠}$ ٠.٣) = ٠.١٨ وبالمثل بالنسبة لبقية الأسئلة، وفي النهاية تجمع درجات الأسئلة جمعاً عادياً للحصول على الدرجة الكلية للاختبار.

ويعاب على هذه الطريقة اهمالها لنسب الإجابات الخاطئة في حالة الانتقاء النسبي، وعموماً فإن طريقة ميشيل تفضل إذا كان هنالك تدرج في صحة البدائل حيث يكون الانتقاء نسبيا، في حين تفضل الطريقة الثانية إذا كانت إحدى الإجابات صحيحة والباقي خاطئ تماماً حيث يكون الانتقاء مطلق (اللحياني، ١٤٣٠).

جـ- الطريقة الثالثة:

تعتمد هذه الطريقة على درجة الثقة في الإجابة الصحيحة المختارة وإذا كانت الإجابة المختارة غير صحيحة طرحت درجة الثقة من النهاية العظمى لهذه الدرجة (سواء كانت النهاية العظمى ١٠ أو ١٠٠)، فإذا رمزنا لتلك الدرجة في كل سؤال (سواء كانت هذه الدرجة مأخوذة مباشرة أو ناتجة من الطرح) بالرمز (ق) فإن درجة كل سؤال تساوي (لو ق)، أي يؤخذ لوغاريتم الدرجة المحددة للأساس ١٠، وتكون الدرجة الكلية للطالب في الاختبار تساوي (مجـ لو ق)، بمعنى أنها تجمع درجات الأسئلة المحددة بالصورة اللوغاريتمية للحصول على الدرجة الكلية للاختبار.

فعلى سبيل المثال إذا طلب من المفحوصين إعطاء درجة من ١٠ تعبر عن الثقة في صحة الإجابة المختارة، ووضع أحد المفحوصين الدرجة ٧ على إجابة سؤال ما، فإذا كانت هذه الإجابة صحيحة فإن الدرجة تساوي لو ٧ = ٠.٨٥ وإذا كانت الإجابة خاطئة فإن الدرجة تساوي (لو ١٠ – ٧) = لو ٣ = ٠.٤٨

د- الطريقة الرابعة :

تعتمد هذه الطريقة أيضاً على درجة الثقة في الإجابة المختارة سواء كانت صحيحة أم خاطئة، وبالتالي تعتبر درجة كل سؤال بهذه الطريقة تساوي (١- لو.١ (١٠ – ق)، حيث (ق) هي درجة الثقة المحددة كما سبق، ١٠ هي النهاية العظمى لدرجة الثقة كما في المثال السابق، وفي ذلك المثال إذا كانت الإجابة صحيحة فإن ق= ٧ وبالتالي تكون درجة كل سؤال بهذه الطريقة تساوي (١- لو (١٠-٧) = ١- لو ٣ = ٠.٥٢ أما إذا كانت الإجابة المختارة خاطئة فإن ق= ١٠-٧=٣ وتكون الدرجة بهذه الطريقة تساوي ١- لو (١٠-٣) = ١- لو ٧ = ٠.١٥

ويعاب على الطريقتين الثالثة والرابعة اعتمادها فقط على درجة الثقة في بديل واحد بصرف النظر عن البدائل الآخرى، ولكن يمكن استخدام الطريقتين المذكورتين في حالة الانتقاء المطلق فقط.

وعموماً فإن الطريقتين الأولى والثانية تستخدمان عندما يطلب من المفحوص توزيع درجات أو نسب الثقة على كل البدائل في الفقرة، في حين يفضل استخدام الطريقتين الثالثة والرابعة عندما يطلب من المفحوصين وضع درجة أو نسبة ثقة على الإجابة التي يختارونها (غنيم، ١٩٨٦).

٢- الطريقة الكروية (Spherical Method):
تستخدم هذه الطريقة عندما تكون درجات الثقة موزعة على البدائل كل حسب درجة صحته، حيث تحسب درجة الفقرة بواسطة الدالة التالية:

الدرجة الكروية (ك)=ح (ق) / (مجـ (ح(س))^{٢})$^{\frac{1}{2}}$

حيث يدل الرمز (ق) على التوقع الاحتمالي للإجابة الصحيحة، ويدل الرمز ح (س) على التوقع الاحتمالي لأي إجابة سواء كانت صحيحة أم خاطئة (ن) تدل على الرقم الثابت لعدد البدائل في المفردة، (س) تدل على الرقم المتغير لعدد البدائل في كل فقرة، فعلى سبيل المثال: إذا كان هنالك سؤال يحتوي على أربع بدائل هي : أ، ب، جـ، د وأعطى أحد المفحوصين التوقعات الاحتمالية (نسب الثقة) الاتية على الترتيب ٢٠%، ٥٠%، ١٠%، ٢٠% ، فإذا كانت الإجابة الصحيحة هي (ب) فإن : ح(ق) =٠.٥ ، ن = ٤ وبالتالي فإن :

الدرجة الكروية للسؤال=

٠.٥ ÷ ((٠.٢)^{٢} + (٠.٥)^{٢} + (٠.١)^{٢} +(٠.٢)^{٢})$^{٠.٥}$ = ٠.٨٦

وهكذا بالنسبة لبقية الفقرات ثم تجمع درجات الأسئلة جمعاً عادياً للحصول على الدرجة الكلية للاختبار(Rippey, 1970).

٣- طريقة الدالة التربيعية (Squar Function Method):
تقدر الدرجة وفقا لهذه الطريقة من الدالة الاتية:

الدرجة التربيعية = ٢ ح (ق) مجـ (ح(س))^{٢}

حيث ح (ق) تعبر عن النسبة المعطاة للإجابة الصحيحة ح(س) تعبر عن النسبة الاحتمالية المعطاة لأي إجابة سواء كانت صحيحة أم خاطئة.

فعلى سبيل المثال إذا السؤال يحتوي على أربع بدائل أ، ب، جـ، د. وإذا أعطى المفحوص النسب الاحتمالية التالية على الترتيب ٣٠%، ٢٥%، ٤٠%، ٥%، وكانت (ب) هي الإجابة الصحيحة.

فإن الدرجة التربيعية للسؤال= ٢× ٠.٢٥ – ((٠.٣)^{٢} + (٠.٢٥)^{٢} +(٠.٤)^{٢} +(٠.٠٥)^{٢})= ٠.١٨٥

٤- طريقة الاختيار الاستنتاجي (Inferred Choice Method):
في هذه الطريقة تعطى الدرجة من الدالة الاتية:

الدرجة = ١ إذا كانت ح(ق) > ح(س) لكل س≠ ق

صفر خلاف ذلك.

في المثال السابق، إذا كانت الإجابة الصحيحة (ب) فإن درجة السؤال تكون صفر، لإن البديل أ نسبته أكبر من نسبة (ب)، أما في حال كانت الإجابة الصحيحة (جـ) فإن درجة السؤال تكون ١ . وهكذا بالنسبة لبقية الأسئلة وللحصول على الدرجة الكلية للاختبار تجمع درجات الأسئلة (Koehler, 1971) .

٥- طريقة الاحتمال المقترح للإجابة الصحيحة (Probability Assigned to Correct Response):

تتطلب هذه الطريقة من المفحوص أن يعطى توقعات احتمالية أو درجات ثقة على البدائل، ثم يؤخذ التوقع الاحتمالي المعطى للبديل الصحيح على أنه درجة هذا السؤال، فإذا كانت التوقعات الاحتمالية للبدائل بالصورة التالية: أ (٢٠%)، ب (١٠%) ، جـ (٣٠%)، د (٤٠%)، والإجابة الصحيحة البديل (جـ) فإن درجة السؤال تقدر ٠.٣، اما إذا كانت الإجابة الصحيحة (أ) فإن درجة السؤال تساوي ٠.٢.

وهكذا بالنسبة لبقية البدائل، وفي النهاية يتم جمع درجات الأسئلة جمعاً عادياً للحصول على الدرجة الكلية للاختبار (عبدالوهاب، ٢٠٠١).

٦- طريقة وزن الثــقة :

في هذه الطريقة يطلب من المفحوصين ان يجيبوا على كل مفردة مع إعطاء أحد مستويات الثقة الثلاثة الآتية : تخمين – ثقة متوسطة – ثقة عالية . ويتم تعريف المفحوصين بالطريقة التي سيتم بها التصحيح والمتمثلة بالخطوتين الاتيتين:

– إذا كانت الإجابة صحيحة يتم إعطاء الدرجات ١،٢،٣ للمستويات الثلاث السابقة على الترتيب، أي أن المستوى تخمين يأخذ واحد، والمستوى (ثقة متوسطة) يأخذ ٢ ، والمستوى (ثقة عالية) يأخذ ٣.

– إذا كانت الإجابة خاطئة تأخذ مستويات الثقة الثلاثة السابقة الدرجات الآتية على التوالي: صفر، -٢، -٣ ، أي تطرح درجتين في حالة المستوى (ثقة متوسطة) كما تطرح ثلاث درجات في حالة المستوى (ثقة عالية)، ويمكن إعطاء مستويات الثقة في حالة الإجابة الخاطئة الدرجات الآتية على التوالي: صفر، ٤- ، -٦ على أن يعلم المفحوصين ذلك قبل بدء الإجابة، ثم تجمع الدرجات جمعاً عادياً والناتج يعبر عن الدرجة الكلية للاختبار(غنيم، ١٩٨٦).

٨- طريقة الدرجات التجريبية (Experimental Scoring Method):

هذه الطريقة تتطلب إجراءات معينة على شكل توجيهات تعطى للمفحوصين، وهذه التوجيهات هي: من فضلك أجب على مفردات الاختبار في ورقة الإجابة المخصصة لذلك والمعطاة لك، وسوف تجد انها مقسمة إلى خانات من أ إلى د ضع (١) في الخانة التي تعتقد انها تمثل الاختيار الصحيح بدرجة كبيرة، وممكن وضع (٢) في خانة أخرى تمثل اختيار تعتقد أيضاً أنه صحيح ولكن بدرجة أقل عن الاختيار الأول، ويمكنك الاستمرار في هذا العمل بالنسبة للخانة الثالثة والرابعة إذا رغبت في ذلك، وافعل ذلك لكل سؤال من أسئلة الاختبار.

إذا كنت واثق من إجابة معينة ثقة تامة ضع الرقم (١) في الخانة التي تمثله وهذا يكفي، وإذا كنت متأكد إلى حد ما من صحة اختيارين مختلفين فإنه يمكنك وضع (١) في الخانة التي تمثل الاختيار الأكثر تأكيداً، وتضع الرقم (٢) في الخانة التي تمثل الاختيار الأقل تأكيداً،

وإذا كنت لا تستطيع أن تقرر صحة ثلاثة اختيارات بثقة تامة ضع الأرقام ١،٢،٣ في الخانات التي تمثلهم بالترتيب التنازلي للثقة، أي الأكثر فالأقل ، واعلم ان التصحيح وإعطاء الدرجات لهذه الأسئلة سيكون كالآتي: إذا ملأت خانة واحدة وهذه الخانة تمثل الاختيار الصحيح فإنك ستحصل على ٣ درجات، وإذا ملأت خانتين تمثل اختيارين وواحد منهما صحيح فإنك ستحصل على درجتين، وإذا ملأت ثلاثة خانات وواحد منهما صحيح فإنك ستحصل على درجة واحدة، وإن ملأت الخانات الأربع ستحصل على صفر، ولا تضع في خانتين مختلفتين رقما واحد (غنيم، ١٩٨٦ ؛ عبدالوهاب، ٢٠٠١؛ اللحياني، ١٤٣٠).

١٠- طرق تصحيح أثر التخمين :

الطريقة الأولى: طريقة تصحيح الاختبار من نوع الاختيار من متعدد باستخدام معادلة التصحيح لأثر التخمين (العقاب) وتكون علامة المفحوص على الاختبار المصحح بهذه الطريقة تساوي:

درجة المفحوص = عدد الإجابات الصحيحة – (عدد الإجابات الخاطئة ÷ عدد البدائل -١)

يؤخذ على هذه الطريقة انها تعتبر أن جميع إجابات المفحوصين تأثرت بالتخمين لذا لابد أن يخصم جزء من الدرجة.

الطريقة الثانية: طريقة تصحيح الاختبار من نوع الاختيار من متعدد باستخدام معادلة التصحيح لأثر التخمين (المكافأة) وتكون علامة المفحوص على الاختبار المصحح بهذه الطريقة تساوي:

درجة المفحوص = عدد الإجابات الصحيحة + (عدد الفقرات المتروكة ÷ عدد البدائل)

وتقوم على تشجيع المفحوص على ألا يخمن وذلك بمكافأته إذا لم يخمن وتشير الدراسات النظرية إلى أن استخدام هذه المعادلة تعطي معامل ثبات عال إذا ما قورن بمعامل الثبات المستخرج عند استخدام معادلة عقاب التخمين السابقة. (الخراشة، ٢٠٠٤).

الطرق الوزنية وفق النظرية الحديثة:

تعد النظرية الحديثة نظرية اختبار احتمالية قائمة على النماذج، وهذا يحدد العلاقة بين الدرجة المشاهدة التي يحصل عليها الفرد عند إجابته على الفقرة والسمة التي يمتلكها. يمكن ربط هذه العلاقة بين السمة التي يملكها الفرد واستجابته على فقرات الاختبار من خلال منحنى خصائص الفقرة الذي يعد اللبنة الأساسية في النظرية الحديثة، حيث تعتمد عليه بقية مكونات النظرية (بيكر، ٢٠١٠).

ويمكن التعبير عن هذه العلاقة من خلال نماذج رياضية عدة ومن أشهرها:

١- النموذج الأحادي المعلم (One- parameter Model) :

ويعد أبسط نماذج النظرية الحديثة؛ وذلك لأنه يصف كل فقرة بدلالة معلم واحد بعد تحليل بيانات الفقرة التي تمت الاستجابة إليها من قبل الأفراد. ويفترض هذا النموذج أن جميع الفقرات لها قيمة تمييزية واحدة، وأن التخمين يكون في حده الأدنى، ومعادلة منحنى خصائص الفقرة أحادية المعلمة (1PLM) هي:

$$p(x_j = 1|\theta.a_j.\delta_j) = \frac{e^{a_j(\theta-\delta_j)}}{1+e^{a_j(\theta-\delta_j)}} \qquad p_g(\theta) = \frac{e^{D(\theta-b_g)}}{1+e^{D(\theta-b_g)}}$$

٢- النموذج الثنائي المعلمة (Two- Parameter Model):

يختلف هذا النموذج عن النموذج الأحادي في كونه يسمح للفقرات أن تختلف في معلم التمييز، وعندما تختلف الفقرات في تمييزها نجد أن دوال الاستجابة للفقرة تتقاطع فيما بينها، وذلك أن تمييز الفقرة يتناسب طردياً مع ميل الدالة، فكلما زادت قيمة التمييز يصبح ميل الدالة أكثر انحدارا؛ وبذلك تزداد قدرة الفقرة على التمييز بين الأفراد، وبدورها تزداد القيمة القصوى لمعلومات الفقرة عند موقع تلك الفقرة. ويوضح النموذج الآتي الصيغة الرياضية للنموذج ثنائي المعلمة

$$p(x_j = 1|\theta.a_j.\delta_j) = \frac{e^{a_j(\theta-\delta_j)}}{1+e^{a_j(\theta-\delta_j)}}$$

٣- النموذج الثلاثي المعلمة (Three- Parameter Model):

النموذج اللوجستي ذو الثلاثة معالم يتضمن الصعوبة، والتمييز، والتخمين، ويُعد النموذج العام للنماذج ذات الاستجابة الثنائية؛ إذ أنه يفترض تأثر الإجابات بعامل التخمين الذي تميز به، والذي يعد أحد العوامل المؤثره في أداء الاختبار، والصيغة العامة لهذا النموذج على النحو الآتي:

$$p(x_j = 1|\theta, a_j, \delta_j, x_j) = \delta x_j + (1-\delta x_j)\frac{e^{a_j(\theta-\delta_j)}}{1+e^{a_j(\theta-\delta_j)}}$$

كما يعد تقدير المعالم والقدرات من الخطوات الأساسية والمهمة ضمن النظرية الحديثة، فبمجرد اختيار النموذج المناسب يصبح من الضروري تقدير خصائص المفردات ومعالم الأفراد، وقد أشارت الأدبيات السابقة على وجود طرق عدة يمكن استخدامها في التقدير وفق النظرية الحديثة، وسنعرض فيما يلي أهم طرق التقدير التي يمكن استخدامها ضمن النظرية الحديثة.

١- التقدير باستخدام أسلوب الارجحية العظمى (Maximum liklehood):

قام بتطوير هذه الطريقة كل من لورد وبرنبوم (Lord, 1952; Birnbaum, 1968) وتتميز طريقة الأرجحية العظمى (Hambleton& Swaminthan, 1985) بالتالي:

- تعطي نتائج مثالية ومتناسقة في حال كان حجم العينة كبيراً، بمعنى أنها تقترب تدريجياً من معالم المجتمع الأصلي، وذلك كلما زاد حجم العينة.
- الإحصاءات الناتجة عن هذه الطريقة كافية، وتحتوي جميع المعلومات حول المعلمة.
- تعد من الطرق التي تحقق الكفاءة، حيث أن طريقة الأرجحية العظمى لديها أصغر تباين.
- التوزيعات الناتجة عن هذه الطريقة تتبع التوزيع الطبيعي.

ونالت تقديرات الأرجحية العظمى اهتمام خبراء القياس المعاصر، وهو ما أسهم في تطوير طرق عدة للتقدير وفق هذه الطريقة، ومنها (علام، ٢٠٠٥):

أ- التقدير بالأرجحية العظمى المشتركة (JMLE):

قام بتطوير هذه الطريقة بيرنبوم (Birnbaum, 1986) ، وتعمل على رفع الحد الأعلى لدالة الأرجحية لكل من المفحوصين والمفردات حتى يتم التقدير في آن معاً لمعالم كل من المفحوصين والمفردات، وذلك وفق الخطوات التالية (أيالا، ٢٠١٧):

- الخطوة الأولى: يتم تقدير مواقع المفردات باستخدام تقديرات مرحلية لمواقع المفحوصين، كما يتم التعامل مع هذه التقديرات المرحلية لمواقع المفحوصين بعدها " معروفة" لغايات تقدير مواقع المفردات.
- تجرى عملية تقدير مواقع المفردات أولاً؛ لأن المألوف أن يكون لدينا من المفحوصين أكثر بكثير مما يكون لدينا من المفردات، وبناء على ذلك تتوافر لدينا معلومات أكثر لتقدير مواقع المفردات، وبما أن تقدير معلم مفردة واحدة لا يعتمد على معالم المفردات الأخرى، لذلك يتم تقدير مفردة واحدة في كل مرة.
- الخطوة الثانية: تعد هذه التقديرات " معروفة"، وتستخدم في تقدير مواقع المفحوصين، ويتم تقدير موقع كل مفحوص بشكل مستقل عن مواقع الأخرين.
- نلاحظ أنه في تقدير مواقع المفردات في الخطوة الأولى استعملنا التقديرات المرحلية لمواقع المفحوصين، لكن بعد الخطوة الثانية تحسنت تقديرات مواقع المفحوصين، وهذه المواقع المحسنة يفترض أن تؤدي إلى تقديرات أكثر دقة لمواقع المفردات، وفي المرحلة التالية نكرر الخطوة الأولى باستعمال التقديرات المحسنة لمواقع المفحوصين في الخطوة الثانية لنحصل على تقديرات أفضل لمواقع المفردات، ومن خلال هذه التقديرات المحسنة لمواقع المفردات؛ وذلك لتقدير مواقع المفحوصين، وتستمر مراحل عملية تسمى : (لعبة كرة تنس الطاولة) في تحسينات متعاقبة في تقديرات مواقع المفحوصين والمفردات إلى أن تصبح غير متمايزة عن بعضها البعض.

ب- التقدير بالأرجحية القصوى الهامشية Maximum Marginal Likelihood Estimation(MMLE)

قام بتطوير هذا الطريقة كل من بوك وايتكن (Bock& Aitken, 1981)، وتعتمد طريقة التقدير بالأرجحية العظمى الهامشية على الفصل بين تقدير معالم المفردات وتقدير معالم الأفراد، وذلك وفق خطوتين (Baker& Kim, 2004):

الخطوة الأولى: الحصول على تقديرات معالم المفردات والتحقق بدرجة كافية من مطابقة البيانات للنموذج.

الخطوة الثانية: تقدير معالم الأفراد، وذلك إما بطريقة الأرجحية العظمى(MLE)، أو بطريقة بايز.

ويشير أيالا(٢٠١٧) إلى أن طريقة التقدير بالأرجحية العظمى الهامشية جاءت لتفادي مشكلات عدة تتضمنها طريقة التقدير بالأرجحية العظمى المشترك، منها:

- زيادة حجم العينة (لكي نحصل على أفضل تقدير لمعالم المفردات بالتزامن مع معالم الأفراد تجب علينا زيادة حجم العينة، غير أن هذه الزيادة تؤدي إلى الزيادة في عدد الأفراد المطلوب تقديرهم.
- إعادة التوزيع (عند تقدير معالم المفردات والأفراد بالتزامن معاً، فقد يحدث أن هنالك عدداً من المفردات غير المطابقة للنموذج، وفي حال تم استبعاد هذه المفردات، فإن هذا يتطلب إعادة التوزيع، وذلك للتخلص من الآثار السلبية للمفردات غير الجيدة.
- إن فصل تقدير معالم المفردات عن تقدير معالم الأفراد يمكن أن يحسن الدقة النظرية في التقديرات في بعض أدوات القياس.

جـ- طريقة الأرجحية العظمى الموزونة Weighted Maximum Likelihood (WML):

اقترح هذه الطريقة وورم (Warm, 1989)، بهدف تقليص تحيز المعالم الذي ينتج عن طريقة الأرجحية العظمى (MLE)، وقد أشار وورم إلى أن طريقته تعطي نتائج ذات تحيز أقل من طريقة الأرجحية العظمى(MLE)، وتعطي نتائج أقل حساسية من طريقة بايز في التقدير البعدي (EAP)، وذلك أنها تعتمد على أوزان معلم القدرة في دالة الأرجحية التي تعتمد بدورها على دالة المعلومات.

ويشير ديليدال وآخرون Deledalle et al (2009) أن هذه الطريقة تعمل على تقليص متوسط الخطأ المعياري عن طريق تقليل التباين في تقدير التحيز، وذلك من خلال تقدير أوزان مسبقة تعرف بـ (PPB)، وهي اختصار لـ: (Probabilistic Patch-Based).

٢- طريقة بايز في التقدير:

يعد جوهر طريقة بايز في التقدير تقوم على افتراض أن لدينا معلومات عن موقع الفرد بدلالة توزيع احتمالي قبل حصولنا على أي بيانات مشاهدة، ويشار إلى هذا التوزيع بالتوزيع القبلي، وبعد تطبيق الأداة نحصل على بيانات مشاهدة تدمج بالمعلومات القبلية للتوزيع، ومحصلة لهذا التكامل نحصل على توزيع يطلق عليه التوزيع البعدي، واستناداً على هذا التوزيع البعدي نحصل على تقدير موقع الفرد (أيالا، ٢٠١٧). ومن الطرق البايزية:

أ- طريقة التوقع البعدي Expected a posterior (EAP)

تتميز هذه الطريقة بأنها تعمل على حساب التقديرات في جميع الحالات سواء أكان حصل الفرد على الدرجة الكاملة من خلال إجابته على جميع المفردات بالشكل الصحيح أم حصل على الدرجة صفر في حال الإجابة على جميع المفردات بشكل خاطئ (Lord, 1986).

كما أن الانحراف المعياري الناتج عنها يعد منخفضاً مقارنة بباقي الطرق الأخرى وخاصة الطرق القائمة على الأرجحية العظمى (Zimowski et al, 2003).

وتعتمد هذه الطريقة على أسلوب مختلف في التقريب المتتابع، حيث اعتمدت على استخدام التوزيع الطبيعي المعياري وليس خوارزمية نيوتن – رافسون كما في الطرق الأخرى (Embretson & Reise, 2013) .

ب- الطريقة البايزية Bayesian Method Estimation (BME)

قدمت هذه الطريقة ساميجيما (Samejima, 1969) التي تعتمد على نظرية بايز، وتتطلب هذه الطريقة توافر معلومات أولية عن موقع الفرد بدلالة توزيع احتمالي قبل حصولنا على أي معلومات مشاهدة (أيالا، ٢٠١٧).

الفرق بين طريقتي (EAP) و (BME):

نظراً لعدم وجود توزيع قبلي معروف عن المجتمع فإننا نعتمد على التقديرات المشاهدة في الحصول على التوزيع القبلي، وذلك من خلال دالة الكثافة الاحتمالية، وتتميز (EAP) عن طريقة (BME) بأنها تعتمد على حساب متوسط معلم القدرة من خلال التوزيع القبلي، ولذلك فإن عدم ملاءمة التوزيع القبلي يؤثر بشكل مباشر في دقة التقدير، ويتضح ذلك في الاختبارات القصيرة (Mislevy& Stocking; Seong, 1990).

وتتميز (BME) عن طريقة (EAP) في كونها تعتمد على قيمة معلم القدرة من خلال المنوال، أي أنها تختار قيمة القدرة الأكثر تكراراً في التوزيع القبلي.

ثانياً- الخصائص السيكومترية Psychometric Properties

يهتم الباحثون في مجالات العلوم النفسية والتربوية والاجتماعية بالحصول على معلومات وبيانات دقيقة حول الظواهر السلوكية، وفي سبيل الحصول على تلك البيانات يستخدم الباحث عدداً من الأدوات كالاختبارات مثلاً.

وعندما نستخدم اختباراً من أجل الحصول على معلومات تساعدنا في اتخاذ قرار ما، فإننا نواجه مشكلة أساسية تتعلق باختيارنا للاختبار الذي يمكن أن يفيدنا حقيقة في اتخاذ القرار.

كما أن هنالك أمور كثيرة يمكن أن تؤخذ بعين الاعتبار عند تقويم جودة اختبار ما. منها:

- الشروط العامة للاختبار وتتضمن الشمول – التقنين- الموضوعية.
- الشروط التجريبية للاختبار وتتضمن الثبات- الصدق- المعايير.
- الاعتبارات العملية والقابلية لاستخدام الاختبار.

الخصائص السيكومترية لفقرات الاختبار:

تختلف طرق تقدير الخصائص السيكومترية باختلاف النظرية المستخدمة في تحليل البيانات معامل صعوبة الفقرة (Difficulty Item):

تعرف صعوبة الفقرة في ضوء النظرية التقليدية بأنها نسبة الأفراد الذين اجابوا عن الفقرة إجابة صحيحة، وعندما تكون الفقرات ثنائية الاستجابة فإن صعوبة الفقرة تساوي

متوسطها، الذي يساوي مجموع الدرجات مقسوما على عددها، وبالتالي فإن صعوبة الفقرة ثنائية الاستجابة تعطى وفق المعادلة الآتية:

معامل الصعوبة = عدد الأفراد الذين حصلوا على علامات صحيحة على الفقرة ÷ عدد المفحوصين.

وتتراوح قيمة معامل الصعوبة بين صفر و واحد (Anastasi, 1982).

وينصح بأن يكون معامل الصعوبة قريباً من (٠.٥) ويتراوح مؤشر الصعوبة بين (٠.١٥) و (٠.٨٥) ويتم حذف الفقرات خارج هذا النطاق إلا في ظل ظروف خاصة كأسئلة تقيس مهارات أساسية

ويمكن حساب معامل الصعوبة للفقرة في حال المجموعتين العليا والدنيا من خلال تطبيق المعادلة التالية:

معامل الصعوبة = (ص ع + ص د) ÷ $ن^{٢}$ ، حيث ص ع = عدد الأفراد الذين أجابوا على الفقرة إجابة صحيحة في المجموعة العليا، ص د = عدد الأفراد الذين أجابوا على الفقرة إجابة صحيحة في المجموعة الدنيا، ن = عدد الأفراد في إحدى المجموعتين (مجيد، ٢٠١٤).

أما معامل الصعوبة في النظرية الحديثة وتحديداً النموذج الأحادي والثنائي هو نقطة على متصل السمة تقابل احتمال قدره (٠.٥٠) للإجابة عن الفقرة إجابة صحيحة. (ايالا، ٢٠١٧؛ كروكر والجينا، ٢٠١٧)

ويتحدد المدى النظري لمعلمة الصعوبة بالفترة (∞+،∞-) بينما عملياً قد يكون في الفترة [٣+، ٣-]

أو ضمن المدى [٢+، ٢-] أو في الفترة [٢.٥+، ٢.٥-] وتشير القيمة (٢.٥-) إلى فقرة سهلة والقيمة (٢.٥+) إلى فقرة صعبة، بينما تعني القيمة صفر أن الصعوبة متوسطة (Warm, 1978؛ بيكر، ٢٠٠١).

معامل تمييز الفقرة (Item Discrimination):

عادة يكون الهدف من الاختبارات تقديم معلومات عن الفروق الفردية في السمة الذي يقيسه الاختبار، والمعلمة المهمة في انتقاء الفقرات هي معلمة التمييز، حيث أنه في النظرية التقليدية هنالك عدد من المعاملات تستخدم كمؤشرات لفعالية تمييز الفقرة

دليل التمييز ويحسب من المعادلة التالية

نسبة الأفراد الذين أجابوا عن الفقرة إجابة صحيحة من المجموعة العليا – نسبة الأفراد الذين أجابوا إجابة صحيحة عن الفقرة من المجموعة الدنيا، وتتراوح قيمته بين (١-، ١). ويستخدم دليل التمييز في حال الفقرات ثنائية الاستجابة ويتطلب ذلك تعيين نقطة أو نقطتين على توزيع الدرجات كدرجات قطع تفصل المفحوصين إلى مجموعات درجاتهم أعلى أو أدنى من درجات القطع هذه، فإذا كان المحك داخلياً كالدرجة الكلية للاختبار يمكن اعتماد (٥٠%)، وإذا كان عدد المفحوصين كبيراً فقد بين كيلي Kelley (١٩٣٩) أن هنالك

إمكانية للحصول على دليل تمييز أكثر حساسية واستقراراً عند استخدام أعلى (٢٧%) وأدنى (٢٧%) من المفحوصين (Anastasi, 1982).

ومعاملات التمييز الأخرى عبارة عن أنواع مختلفة من معاملات الارتباط تحت اشتراطات محددة وهي: معامل ارتباط بيرسون كمقياس لدرجة العلاقة بين درجة الفقرة والمحك، ومعامل ارتباط بوينت بايسيريال إذا كان الأداء على الفقرة ثنائية التصحيح ودرجات الاختبار الكلية كمتغير متصل مع مراعاة أن يكون عدد الفقرات كبير نسبياً (٢٥ أوأكثر)؛ أما عندما وجود عدد قليل من الفقرات حيث تحذف درجة الفقرة فإنه يمكن استخدام معادلة تصحيح لهذا المعامل، ومعامل ارتباط بايسيريال في حال كان توزيع المتغير الكامن توزيعاً اعتدالياً وأحد المتغيرين ثنائي مصطنع، ومعامل فاي عندما يكون المتغيرين ثنائيين حقيقيين، ومعامل الارتباط الرباعي (تتراكوريك) لمعرفة درجة الارتباط بين متغيرين ثنائيين كلاهما ناجم عن تحويل متغير متصل إلى متغير منفصل ثنائي.

وقدم ايبل (١٩٦٥) Ebel إرشادات لتفسير قيم معامل التمييز على النحو الآتي:

١- إذا كانت (٠.٤٠≤ معامل التمييز)، فإن الفقرة تلبي لغرض المقصود.

٢- إذا كانت (٠.٣٩≥معامل التمييز≥٠.٣٠)، فإن الفقرة تتطلب مراجعة بسيطة وقد لا تتطلب.

٣- إذا كانت (٠.٢٩ ≥ معامل التمييز ≥ ٠.٢٠)، فإن الفقرة تقع على الحد الفاصل وتحتاج إلى مراجعة.

٤- إذا كانت (٠.١٩≥معامل التمييز)، فإنه يجب حذف الفقرة أو إجراء مراجعة تامة لها.

ويذكر علام (٢٠١١) أن هنالك علاقة بين معامل صعوبة الفقرة وتمييزها، فكلما ابتعدت قيمة معامل الصعوبة عن (٠.٥٠) إلى أعلى أو إلى أدنى؛ فإن القيمة القصوى لمعامل التمييز تقل، لذا فإن الفقرات السهلة والفقرات الصعبة يكون تمييزها ضعيفاً، والمفردات التي تتراوح قيمة معامل صعوبتها بين (٠.٤٠) و (٠.٦٠) يكون تمييزها مرتفعاً.

وفي ضوء النظرية الحديثة فإن معلمة التمييز ترتبط بميل المنحنى المميز للمفردة؛ الذي يقابل النقطة التي تكون فيها علامة القدرة تساوي صعوبة المفردة وكلما زادت قيمة معامل التمييز يصبح ميل دالة الاستجابة أكثر انحداراً، وبذلك تزداد قدرة المفردة على التمييز بين الأفراد (ايالا، ٢٠١٧؛ كروكر والجينا، ٢٠١٧)

المدى النظري لمعلمة التمييز يرتبط بميل منحنى خصائص الفقرة ويتراوح بين (∞+، ∞) وقد يأخذ قيماً سالبة غير مرغوبة لأنها تعمل بطريقة مخالفة لما هو مقبول، إذ أن الفقرة ذات التمييز السالب تشير إلى احتمال الاستجابة الصحيحة للفقرة لدى افراد من ذوي القدرات المتدنية أكبر مما هو لدى الأفراد ذوي القدرات الأعلى.

لذا يكون المدى النظري في الفترة (∞+، ٠)، إلا أن المدى العملي المفضل يكون في الفترة [٢.٥، ٠.٥] (Warm, 1978)، أو في الفترة (٢.٥، ٠.٨) (دي إيالا، ٢٠١٧)، أو في الفترة [٢، ٠] (Hambleton& Swaminathan, 1985).

كما يجب الأشارة إلى أن انتقاء الفقرات في الاختبارات محكية المرجع لا يتم وفق المحكات التي تستخدم في الاختبارات معيارية المرجع، فقد تكون فقرة ما مناسبة لأغراض اختبارات التمكن أو الإتقان في حين أن معامل صعوبتها أكبر من (٠.٩٠)، أو يكون أقل من (٠.٤٠). فيما يكون معامل الصعوبة (٠.٥٠) أكثر ملائمة في الاختبارات معيارية المرجع.

فعالية المشتتات (المموهات):

المشتتات عبارة عن الخيارات الخاطئة للسؤال ذي الاختيار من متعدد، ويجب أن تكون جذابة بحيث يكون عدد المفحوصين الذين يختارونه من الفئة العليا أقل من عددهم في الفئة الدنيا، وكلما كانت قيمة المموه بالسالب كان هذا دليلاً على أن المشتت جيد، ويمكن حساب معامل فاعلية المشتت على النحو الآتي:

معامل فاعلية المشتت = (ن ع – ن د) ÷ ن ، حيث ن ع= عدد الأفراد الذين اختاروا المشتت في المجموعة العليا، ن د = عدد الأفراد الذين اختاروا المشتت في المجموعة الدنيا، ن = عدد الأفراد في إحدى المجموعتين (مجيد، ٢٠١٤).

ربط النظرية الحديثة بنظرية القياس التقليدية:

يمكن ربط النظرية الحديثة بالنظرية التقليدية من خلال معالم منحنى خصائص الفقرة وإحصائيات الفقرة بالنظرية التقليدية، ولتوضيح ذلك لتكن p_g تمثل ارتباط بايسيريال بين درجات الفقرة (g) ودرجات السمة الكامنة وهذا الارتباط مشابه لمعامل ارتباط بايسيريال بين الفقرة والدرجة الكلية، والذي يستخدم كمؤشر من مؤشرات التمييز عند تحليل الفقرات بنظرية القياس التقليدية. فإذا كان توزيع (ɵ) توزيعاً طبيعياً وسطه (٠) وانحرافه المعياري (١)، فإن قيمة المعلمة (a_g) لمنحنى خصائص الفقرة يكون:

$$a_g = \frac{pg}{\sqrt{1-(pg)\text{^}2}}$$

ويمكن ربط معلمة صعوبة منحنى خصائص الفقرة b_g بمعاملات تحليل الفقرة الكلاسيكي أيضاً، ووفق افتراضات التوزيع الطبيعي يمكن أن نبين أن:

$$b_g = \frac{-\varphi^{-1}(p_g)}{\rho_g}$$

وترمز p_g إلى مقياس نسبة الصحيح لمعامل صعوبة الفقرة g ، و($\varphi^{-1}(p_g)$) إلى الدرجة الزائية التي يقع على يمينها مساحة تساوي p_g في التوزيع الطبيعي المعياري، فإذا عرفنا p_g يمكننا الحصول على قيمة (($\varphi^{-1}(p_g)$) (كروكر والجينا، ٢٠١٧).

- **الثبات(Reliability):** الثبات يعني الاتساق أو الاستقرار الذي تكون عليه الصفة أو الخاصية المراد وصفها.

أساليب حساب الثبات:

١- طريقة إعادة الاختبار (Test Retest):

تعني أن يعطى الاختبار مرتين لنفس المفحوصين ويكون بين المرتين فاصل زمني قد يكون طويلاً أو قصيراً حسب الظروف التي تواجه عملية القياس أو تطلبها الحالة. وتكمن أهميته

في أنه يقيس مستوى الاستقرار. ويمكن التحقق من ذلك بحساب معامل الارتباط بين التطبيقين وقيمته هي مؤشر لمعامل الثبات، وكذلك معامل الارتباط بين الطبقات أو الداخلي Intraclass correlation coefficient (ICC) .

٢- الصور المتكافئة (Al ternate Forms): وتتمثل هذه الطريقة في إيجاد صورتين يفترض تكافؤهما في محتواهما أي أن تكون البنود في كلا الصورتين متكافئتين في السهولة والصعوبة وكذا في طبيعة المواضيع التي تتناولها هذه الصور. ومن أجل حساب الثبات باستخدام هذه الطريقة يتم استخراج معامل الارتباط بين درجات المفحوصين الذين تم تطبيق الصورتين عليهم.

٣- أساليب الاتساق الداخلي (Internal Consistency): عندما تطبق الأداة مرة واحدة، يتم حساب الثبات بطرق الاتساق الداخلي ويمكن تقسيم أساليب الاتساق الداخلي إلى نوعين فرعيين وهما أسلوب التجزئة النصفية (Split-half)، أما النوع الثاني فهو قائم على أسلوب تحليل التباين للاستجابات على الفقرات.

٤- الاتفاق أو التكافؤ(Equivalence): ويعني درجة الاتفاق أو الانسجام بين أثنين أو أكثر من المقيمين بشأن الدرجة على الأداة، ويمكن التحقق من ذلك باستخدام الملاحظين (كابا) Inter-observer reliability. (الطريري، ٢٠١٤؛ كروكر والجينا، ٢٠١٧، السيد، ٢٠٢٠).

جدول (٤): مقارنة طرق تقدير الثبات

تقدير الثبات	المعادلة	التوصيات للاستخدام
الصور البديلة Alternate Forms	معامل الارتباط بين الأداء في الصورتين	يستخدم عند توفر أكثر من نسخة للأداة.
إعادة الاختبار Test-retest	معامل الارتباط بين التطبيقين. معامل الارتباط بين الطبقات أو الداخلي	يستخدم عندما يكون الاستقرار الزمني للأداة مهم.
ألفا لكرونباخ Cronbach's α	$\alpha = \left(\frac{K}{K-1}\right)\left[1 - \frac{\sum_{k=1}^{p} S_k^2}{S_{Total}^2}\right]$	يستخدم عند الحاجة إلى تقدير الحد الأدنى للثبات.
التجزئة النصفية Spilt halves	معامل الارتباط بين النصفين مع التصحيح باستخدام معادلة سبيرمان بروان $SB = \frac{2r_{H1,H2}}{1 + r_{H1,H2}}$	يستخدم كتقدير عام لثبات الدرجة.
أوميجا Omega	$\omega = \frac{\left(\sum_{i=1}^{I} a_i\right)^2}{\left(\sum_{i=1}^{I} a_i\right)^2 + \sum_{i=1}^{I} \psi_i^2}$	يستخدم في حالة فهم البناء الكامن للمقياس، ونثق في نموذج العامل المناسب للبيانات.
ألفا الطبقية Stratified α	$\alpha_{Stratified} = 1 - \frac{\sum_{m=1}^{M} \sigma_m^2 (1 - \alpha_m)}{\sigma_X^2}$	يستخدم عند تقدير الثبات لمقياس يتكون من عدة مقاييس فرعية

ثبات الاختبارات محكية المرجع:

يعتبر ثبات الاختبارات محكية المرجع من القضايا الأساسية في القياس، وذلك أن الثبات بصورة عامة يعتمد على الفروق والاختلافات الموجودة بين المفحوصين عند تطبيق الاختبار عليهم، وفي الاختبارات محكية المرجع يكون الاهتمام منصباً على مقارنة درجات المفحوصين بمعيار محدد سلفاً مثل درجة النجاح على الاختبار التحصيلي والتي تمثل محكاً يتم مقارنة درجات المفحوصين به لمعرفة هل تجاوزوا هذه الدرجة أم لا. ومن هذا المنطلق فإن الاهتمام في الاختبارات محكية المرجع سيكون منصباً على الدقة في الدرجة بالإضافة إلى الاهتمام في أحيان أخرى على الدقة والصواب في القرار، مثل أن يصنف المفحوصين فيما أن كانوا قد بلغوا أو لم يبلغوا المحك، وبناء على ذلك فإن أساليب حساب الثبات بالطرق التقليدية لا يناسب استخدامها مع الاختبارات محكية المرجع، لكن ذلك لا يعني عدم أهمي الاتساق الداخلي للاختبار محكي المرجع، ولا عدم أهمية إعادة الاختبار فهي جميعاً مهمة ولكنها لا تغني عن حساب معامل الثبات للاختبارات محكية المرجع بالطرق المناسبة لها (الطريري، ٢٠١٤).

دالة المعلومات (Information Function):

تعد دالة المعلومات للاختبار سمة ذات أهمية بالغة في نظرية الاستجابة للفقرة، وتدل على مدى دقة فقرات الاختبار على تقدير قدرة الفراد عبر متصل السمة المقاسة، وكلما كانت هناك فقرات أكثر في الاختبار كلما كان مقدار المعلومات أكبر (بيكر، ٢٠١٠). ويشير مفهوم دالة المعلومات للفقرة في النموذج الثنائي المعلم إلى كمية المعلومات التي تقدمها الفقرة لتقدير مستوى قدرة الفرد، وهي تكافئ مفهوم الثبات في النظرية الكلاسيكية، ولدالة المعلومات للفقرة تعريف رياضي دقيق من خلال الصيغة الرياضية الآتية:

$$I_g(\theta) = \frac{[p'_g(\theta)]^2}{[p_g(\theta)][Q_g(\theta)]}$$

حيث Ig(ɵ) : دالة معلومات الفقرة g، و (P`g(ɵ)): المشتقة الأولى لمنحنى خصائص الفقرة؛ والمشتقة الأولى محسوبة عند أي قدرة ثيتا (ɵ) تساوي ميل مماس منحنى خصائص الفقرة المرسوم عند النقطة (P(ɵ)).

وترمز (g): للفقرة، و $P_g(\theta)$: نسبة احتمال إجابة المفحوص إجابة صحيحة عن الفقرة(g)، و ($Q_g(\theta)= 1- P_g(\theta)$): نسبة احتمال إجابة المفحوص إجابة خاطئة عن الفقرة (g)، ولذا تميل المعلومات لأن تكون أكبر عندما تكون قيمة المشتقة أكبر (كروكر، والجينا، ٢٠١٧).

ويذكر ايالا (٢٠١٧) أن دالة معلومات الفقرة للنموذج الأحادي المعلم هي:

$$I_j(\theta) = a^2 p_j(1 - p_j)$$

حيث $I_j(\theta)$ يشير إلى معلومات الفقرة (j) عند مستوى القدرة (ɵ)، و (α) معلمة التمييز، و(p_j): احتمالية الإجابة على الفقرة إجابة صحيحة، وبما أن التمييز يساوي واحداً صحيحاً؛

وناتج ضرب $p_j*(1-p_j)$ تصل قيمته القصوى عندما تكون $p_j = (1-p_j)$ فإن المعلومات القصوى في النموذج الأحادي المعلم تساوي (٠.٢٥).

ودالة معلومات الفقرة للنموذج الثنائي المعلمة تأخذ قيمتها القصوى عند $I_j(\theta)=0.25a_j^2$، بينما في النموذج الثلاثي المعلمة تكون دالة المعلومات على الصورة التالية:

$$I_j(\theta) = a_j^2 \left[\frac{p_j - c_j}{1 - c_j}\right]^2 \left[\frac{(1 - p_j)}{p_j}\right]$$

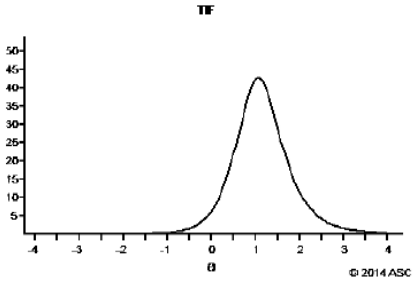
حيث يشير (c_j) إلى معلمة التخمين للفقرة (j).

كما أن دالة المعلومات الكلية للاختبار تساوي مجموع دوال معلومات فقرات الاختبار جميعها، وذلك وفق الصيغة الآتية:

$$I(\theta) = \frac{1}{\sigma_e^2(\theta)} = \sum_{j=1}^{L} I_{j(\theta)}$$

حيث يشير $\sigma_e^2(\theta)$ يشير إلى تباين الخطأ، ويكون المنحنى البياني لدالة المعلومات للاختبار كما في الشكل (٢-٢) أدناه.

شكل (١) دالة معلومات الاختبار (Test Information Function)



يوضح الشكل (٢-٢) متصل السمة أفقياً لقدرة الفرد؛ والمحور الرأسي الأيسر معلومات الاختبار (أيالا، ٢٠١٧).

الخطأ المعياري للقياس (Standard Error of Measurement)

تعرف كروكر والجينا (٢٠١٧) الخطأ المعياري في القياس على أنه متوسط الانحراف المعياري لتوزيعات الخطأ لعدد كبير من المواقف الاختبارية المتكررة، وتتعدد مصادر أخطاء القياس، وطريقة تأثير هذه الأخطاء تحدد مقدار الخطأ ونوعه؛ فقد يكون التأثير عشوائياً ويسمى الخطأ في هذه الحالة الخطأ العشوائي (Random error) وقد يكون التأثير في اتجاه واحد (بالزيادة أو النقصان) ويسمى الخطأ في هذه الحالة الخطأ المنتظم (Systematic error).

ويذكر علام (٢٠١١) أنه من المفيد الاهتمام بمصادر أخطاء القياس من أجل التحكم فيها أثناء إعداد الاختبار وعند تطبيقه وتصحيح فقراته، وذلك لجعل تأثيرها في الدرجة الحقيقة

أقل ما يمكن وبذلك تقترب الدرجات الملاحظة للأفراد في الاختبار من درجاتهم الحقيقة ويزيد تباين الدرجات الحقيقية إلى أقصى حد ممكن، وبالتالي تكون قيم معامل الثبات أكبر ما يمكن.

ونظراً لأن أخطاء القياس العشوائية التي تؤثر في موقف اختباري تختلف باختلاف مجموعة الأفراد المختبرين ونوع الاختبار وعينة فقراته، وظروف تطبيقه وتصحيح فقراته فإن مصادر هذه الأخطاء تكون كثيرة ومتعددة. ومن أهم هذه المصادر: أخطاء تتعلق بأداة القياس، وأخطاء ترتبط بإجراءات تطبيق الاختبار وتصحيحه، وأخطاء ذات علاقة بالأفراد المختبرين.

ويشير عودة (٢٠٠٤) أن القياس لا يمكن أن يكون مثالياً؛ وهذا يعني أن دقة تقدير الدرجة الحقيقية يمكن أن تقع ضمن مدى معين للدرجة الظاهرية يسمى فترة الثقة، ويتطلب تحديدها معرفة الخطأ المعياري في القياس في ضوء النظرية الكلاسيكية من خلال معامل الثبات والانحراف المعياري للدرجات الظاهرية وفق العلاقة التالية:

$$\sigma_E = \sigma_X \sqrt{(1 - P_{XT})}$$

حيث تعني σ_E الخطأ المعياري للقياس، و σ_X الانحراف المعياري للدرجة المشاهدة، و P_{XT} معامل الارتباط (الثبات) بين الدرجتين المشاهدة والحقيقية (آنستازي وأورينا، ٢٠١٥).

ويلاحظ أن فترة الثقة تعتمد على الخطأ المعياري الذي بدوره يعتمد على معامل الثبات فإذا كان معامل الثبات = (١.٠٠) نظرياً فإن الخطأ المعياري = (٠) وبالتالي تكون الدرجة الحقيقية مساوية للدرجة المشاهدة، أي عدم وجود خطأ في تقدير الدرجة الحقيقية ومدى الثقة في هذه الحالة = (٠)، أما إذا كان معامل الثبات = (٠) فإن الخطأ المعياري = الانحراف المعياري للدرجات بمعنى أن الفروق في الدرجات كلها ناتجة عن أخطاء القياس وليست فروقاً حقيقية مرتبطة بالسمة أو الخاصية التي يقيسها الاختبار، ويمكن حساب مدى الثقة للدرجة الحقيقية من خلال العلاقة الآتية:

مدى الثقة للدرجة الحقيقية= الدرجة الملاحظة ± الخطأ المعياري للقياس × الانحراف المعياري للمنحنى الاعتدالي عند مستوى الثقة المحدد (٠.٩٥ أو ٠.٩٩).

وهذا يعني أن مدى الثقة للدرجة الحقيقية عند مستوى ٠.٩٥= الدرجة الملاحظة ± الخطأ المعياري للقياس ×١.٩٦، وأن مدى الثقة للدرجة الحقيقية عند مستوى ٠.٩٩= الدرجة الملاحظة ± الخطأ المعياري للقياس ×٢.٥٨.

وبهذا تتضح أهمية الخطأ المعياري في القياس عند حساب فترة الثقة التي يمكن أن تقع ضمنها الدرجة الحقيقية.

ولتكوين فترة ثقة لدرجة المفحوص الحقيقية يجب أن تتحقق الافتراضات التالية:

١- الافتراضات الأساسية للنظرية الكلاسيكية.

٢- خطأ القياس لكل مفحوص يتبع التوزيع الطبيعي.

٣- الخطأ المعياري في القياس متساو لجميع المفحوصين.

وفي الواقع العملي فإن الافتراضات السابقة لتكوين فترة ثقة لتقدير درجة المفحوص الحقيقية امر يصعب تحقيقه أو خاصة فيما يتعلق بالتحقق من افتراضات تساوي الخطأ المعياري في القياس لجميع المفحوصين، وأن خطأ القياس لكل مفحوص يتبع التوزيع الطبيعي. حيث أن توزيع الدرجات المشاهدة قد يكون سالب الالتواء، وقد يكون توزيع الدرجات المشاهدة موجب الالتواء.

ولذلك إذا لم تتحقق الافتراضات السابقة فإن فترات الثقة لتقدير موقع القدرة الحقيقية ستكون غير دقيقة.

وللتخلص من الافتراضات الثلاثة السابقة والمرتبطة بتكوين فترة الثقة، يتم استخدام الخطأ المعياري في القياس عند مستويات محددة لعلامات المفحوصين بدلاً من استخدام قيمة واحدة للخطأ المعياري لجميع المفحوصين.

ويتم ذلك بحساب قيمة الخطأ المعياري في القياس لكل مفحوص بناء على درجته الكلية، مما يمكن من تقدير دقيق للدرجة الحقيقية لكل مفحوص، وكذلك تقدير دقيق لفترات الثقة التي يتوقع أن تقع ضمنها الدرجة الحقيقية للمفحوص (عباس، ١٩٩٣).

وقد تم اقتراح عدد من الطرق التي يمكن من خلالها إيجاد الخطأ المعياري في القياس عند مستويات محددة لدرجات المفحوصين. بدلاً من استخدام قيمة واحدة للخطأ المعياري في القياس لجميع المفحوصين، وفيما يلي عرض لطرقتين من هذه الطرق:

١- طريقة لورد (Lord's Method):

يفترض لورد أن الاختبار عبارة عن اختيار عينات عشوائية للفقرات من مجتمع الفقرات، وأن الاختبارات المختارة بهذه الطريقة والتي تحتوي على العدد نفسه من الفقرات هي اختيارات عشوائية متوازية، وأن الدرجة التي يحصل عليها فرد ما (عدد الإجابات الصحيحة) من خلال إجابته على عدد كبير من فقرات تتبع توزيع ذات الحدين وتنص معادلته:

$$SEM_{CONDLord} = \left[\frac{X(k-X)}{(k-1)}\right]^{1/2}$$

٢- طريقة كيتس (Keats Method):

لقد اقترح كيتس تحسيناً على طريقة لورد بحيث يصبح بالإمكان الحصول على قيم للخطأ المعياري في القياس ليست أعلى مما هو متوقع لها، ويكمن اقتراح كيتس بضرب الناتج التي نحصل عليها للخطأ المعياري في القياس بطريقة لورد بعدد ثابت، وبالتالي فإن معادلة كيتس لإيجاد الخطأ المعياري في القياس عند مستويات محددة من درجات المفحوصين هي:

$$SEM_{CONDKeats} = \left[\left(\frac{X(k-X)}{k-1}\right)\left(\frac{1-r_{xx'}}{1-r_{21}}\right)\right]^{1/2}$$

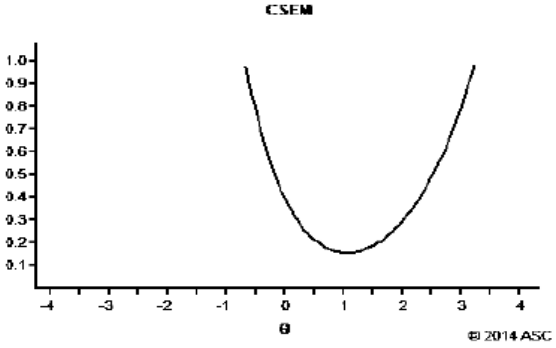
(Lord, 1959؛ طاهات، ٢٠١٦).

ويعرف الخطأ المعياري للتقدير في نظرية الاستجابة للفقرة بأنه مقلوب دالة المعلومات؛ إذ ترتبط الزيادة في المعلومات المتوافرة عند مستوى قدرة محدد بنقص متزامن في الخطأ المعياري لموقع الفرد، ويعرف من خلال العلاقة:

$$SEM(\theta) = {}^{1}/_{\sqrt{I(\theta)}}$$

حيث I(θ) ترمز إلى دالة معلومات الاختبار عند مستوى القدرة (θ) (Hambleton, & Swaminathan, 2010).

والشكل (٢) لتوضيح المنحنى البياني لدالة الخطأ المعياري للتقدير.



يوضح الشكل (٢) متصل السمة أفقياً لقدرة الفرد، والمحور الرأسي الأيسر الخطأ المعياري للتقدير، كما يلاحظ أن المنحنيين في الشكلين (٢-٢) و (٢-٤) متعاكسان.

الصدق(Validity):

الصدق: عرف الصدق بأنه الدرجة التي يتمتع بها الاختبار في تحقيق أهداف محددة، ويؤكد على أهداف عامة:

١ – التنبؤ في أمور ذات علاقة بالمفحوص.

٢ – وصف المفحوص من جانب أو مجموعة من الجوانب.

ويعتبر الاختبار صادق إذا كانت الاستنتاجات المبنية عليه مناسبة وذات دلالة وفائدة.

الدليل على صدق الاختبار ربط بتصنيفات الصدق الرئيسة والتي هي بمثابة الدليل على صدق المحتوى، الدليل على صدق المحك، والدليل على صدق المفهوم. وتجدر الإشارة إلى أن هذا التقسيم لا يعني الانفصال، بل من الممكن أن تكون ادلة صدق المحتوى وصدق المحك مفيدة لصدق المفهوم.

حقائق مرتبطة بصدق الاختبارات:

- الصدق يعتبر أهم خاصية يفترض أن تتحقق في الاختبار.

- الصدق يعني فعالية الاختبار في قياس ما قصد قياسه.
- الصدق يمكن تعريفه بعدة تعريفات كما أنه ذو أنواع وصيغ مختلفة.
- يمكن التحقق من صدق الاختبار من خلال استخدام الاختبار بصورة عملية بعد أن يتم تطويره.
- للتحقق من صدق الاختبار لابد من محكات حقيقية أو افتراضية أو مصطنعة.
- الاختبار الصادق لابد من أن تكون مواصفات محتواه واضحة ومحددة بالإضافة إلى إعطاء نتائج ثابتة.
- يمكن تعريف ما يقيسه الاختبار من خلال إعطاء وصف دقيق ومحدد للمهمات والصفات والخصائص وليس مجرد تسميتها فقط.
- صدق الاختبار يتمثل في استخدامات الدرجات وليس في الاختبار بحد ذاته إذ أن الاختبار بمعزل عن الاستخدامات المترتبة عليه لا قيمة له.
- الأساليب والطرائق المنطقية أفضل من الأساليب الإحصائية في تأكيد صدق الاستخدامات المترتبة على درجات الاختبار. الاختبار يفقد مصداقيته إذا لم يتمكن مستخدم الاختبار من الخروج ببعض الاستنتاجات المبنية على درجات الاختبار، إن صدق الاختبار لا يمكن حصره في نتائج إحصائية بحته بل لابد من أدلة عملية وواقعية.
- الصدق صدق نسبي، أي أنه لا يمكن أن يكون الاختبار صادقاً تمام الصدق في كل الظروف والحالات.
- الدقة في التقدير يعتبر علامة على مصداقية الاختبار في التنبؤ، لذا فإن الاعتماد على متغير وحيد يتنبأ من خلاله بالظاهرة قد يكون مضللاً.

صدق الاختبار واتخاذ القرار: من أجل ضمان أن يكون الاختبار سليماً يتم تحديد ما يسمى بدرجة القطع على الاختبار والتي تمثل الحد الأدنى، مما يجب توفره في الفرد الذي قد يقع عليه الاختيار بناءً على نتيجته في الاختبار.

أساليب تحقيق الصدق:

وقد يتم استخدام أساليب إحصائية وهو ما نطلق عليه معامل الصدق، أو قد يكون أسلوب (مؤشراً) عملياً أو سلوكياً نستشهد به على فعالية الاختبار وصدقه، ومن الممكن أن نصنف أساليب تحقيق الصدق للأنواع التالية:

١- صدق المحتوى: معرفة إلى أي مدى تمثل أسئلة الاختبار المجال السلوكي والمعرفي المراد قياسه من خلال الاختبار. ويتحقق صدق المحتوى إذا كانت هذه الأسئلة تتطلب من المفحوصين أن يبرزوا المهارات والكفاءات المتضمنة في الأهداف التي يتعهد المحتوى الدراسي تثبيتها، ويتحقق صدق المحتوى عندما يكون السلوك الذي تتناوله الفقرات والأسئلة

تتوافق مع السلوك أو الموضوع المحدد في الأهداف محل العناية. ومن المجالات الاكثر حاجة لصدق المحتوى اختبارات التحصيل.

خطوات تحقيق صدق المحتوى:

أ- تعريف وتحديد مجال الأداء المراد قياسه بصورة واضحة.

ب- تحديد الأهداف السلوكية المعني بها من خلال المحتوى الدراسي. ومعرفة أهمية وأوزان هذه الأهداف بالنسبة للمجال أو المحتوى. ويتم عن طريق المحكمين.

ت- مقارنة الأسئلة بالأهداف، ويتم عن طريق المحكمين.

ث- من أجل أن تتم عملية التناظر بين الأسئلة والأهداف لابد من إعطاء المحكمين معلومات كافية حول الأسئلة. جميع هذه الإجراءات كيفية، وقد يتطلع كثير من مستخدمي الاختبارات إلى صيغ كمية. ومنها معامل التطابق.

معامل التطابق = (ن ÷ ٢ن-٣) (س ك – س)

٢- صدق المحك: يعني الحكم الذي يصل له مستخدم الاختبار إزاء إمكانية الاستنتاج حول المفحوص بالاعتماد على نتيجته في الاختبار. ويمكن اعتباره الارتباط الذي يوجد بين الدرجات على الاختبار ومحك خارجي.

وأنواعه :

أ- الصدق التلازمي: ويعني الارتباط الذي يوجد بين الاختبار وبين المحك على أن يتم جمع البيانات باستخدام الاختبار والمحك في نفس الوقت أو في أوقات متقاربة.

ب- الصدق التنبؤي: القدرة التي يتمتع بها الاختبار من أجل التنبؤ بما سيكون عليه الفرد في صفة أو خاصية من الخصائص في المستقبل.

خطوات تحقيق صدق المحك:

١- تحديد واختيار محك مناسب يكون على شكل مؤشرات سلوكية يمكن الاستدلال من خلالها.

٢- تحديد عينة من المفحوصين يمثلون المجتمع.

٣- يتم تطبيق الاختبار عليهم ومن ثم يحتفظ بنتائجه.

٤- يتم جمع البيانات حول المحك ووضع الأفراد ومستواهم في المحك.

٥- حساب معامل الارتباط بين الدرجات على الاختبار وأداء المفحوصين على المحك.

٣- صدق المفهوم : هو الحكم والقناعة العلمية التي تتولد لدى مطور المقياس حول مناسبة الاستنتاجات التي يتوصل إليها من خلال الدرجات على الاختبار.

أساليب صدق المفهوم:

١- الاتساق الداخلي: يتمثل في التأكد من أن الاختبار يقيس شيئاً أو مفهوماً واحداً. ويتحقق من خلال حساب معامل الارتباط بين المقاييس الفرعية المكونة للاختبار والدرجة الكلية للاختبار.

٢- التحليل العاملي: عبارة عن إجراءات إحصائية يقوم بها مطور الاختبار. وينتج من ذلك تحديد العامل الرئيس أو مجموعة العوامل المؤثرة.
٣- التمييز بين المجموعات المختلفة: حيث يطبق المقياس على أفراد المجموعة ويتم مقارنته بمحك خارجي أو اختبار سابق فإذا تساوت درجات الأفراد في كلا المجموعتين فهذا دليل على الإخفاق في تحقيق صدق المفهوم.
٤- أسلوب التشابه والاختلاف مع خصائص أخرى: التشابه يعني أن المقياس الذي طوره الباحث له معاملات ارتباط مع مقاييس أخرى تقيس نفس السمة. أما الاختلاف فيعني أن المقياس الذي طوره الباحث ليس له معاملات ارتباط دالة مع مقاييس يفترض أنها تقيس سمات وخصائص تختلف عن السمة التي يقيسها المقياس.
٥- صدق المفهوم من خلال التغيرات النمائية: يتضح من خلال الفروق التي توجد في درجات المفحوصين المختلفين في الأعمار.
٦- مؤشرات صدق المفهوم من خلال الاختبار القبلي والبعدي وفي حالة تشابه النتائج بين الاختبار القبلي والاختبار البعدي وقوة العلاقة فهذا دليل على أن الاختبار يقيس نفس السمة.
٤- الصدق الظاهري: وهو من أنواع الصدق أيضاً، ويعني ما يبدو أن الاختبار يقيسه ولا يعني بالضرورة ما يقيسه الاختبار وهذا جوهر الاختلاف بينه وبين صدق المحتوى، وتكمن أهمية الصدق الظاهري في الكيفية التي يبدو بها الاختبار لمن سيطبق عليهم، وكذا لمن سيقوم بتطبيقه. وهذا يؤدي إلى زيادة دافعية المفحوصين وثقتهم بالفاحص. أما عن الكيفية التي يتم بها الكشف عن الصدق الظاهري من خلال اسم الاختبار واستعراض بنوده.
٥- الصدق التعارضي: دراسة ما إذا كان أداء المجموعات التي تتمايز في الواقع في السمة المقاسة، على الاختبار متمايزاً، فإذا كشف الاختبار عن هذا التفاوت في الأداء بين المجموعتين فإن ذلك يدل على أن الاختبار يعكس السمة التي تميز بين المجموعتين في الواقع.

صدق الاختبارات محكية المرجع:

تتشابه المبادئ الأساسية في صدق الاختبارات سواء كان الاختبار معياري المرجع أو محكي المرجع، إذ يشير مصطلح صدق الاختبار إلى ما إذا كان الاختبار يقيس ما صمم أصلاً لقياسه، وبالتالي فإن الصدق عادة ما يرتبط بغرض الاختبار وبهدفه وبالقرارات التي يتم اتخاذها بناءً على درجات الطلبة على الاختبار، ولأن الاختبار يستخدم لأغراض مختلفة فإن دلائل وجود صدق الاختبار تتنوع بتنوع أغراضه، وعليه فإن أدلة صدق الاختبار عادة ما تعتمد على أحكام الخبراء ونتائج التحقق الإمبريقي، والتحليل الإحصائي لدرجات الاختبار، ولقد استخدم المختصون بالقياس والتقويم عدة مصطلحات تتعلق بصدق الاختبار منها صدق المفهوم والصدق التجريبي والصدق التمييزي، إلا أن أشهر المسميات المستخدمة في حالة الاختبارات معيارية المرجع هي صدق المحتوى والصدق المرتبط بالمحك والصدق البنائي (عبابنة ، ٢٠٠٩). وتوجد ثلاث أدلة رئيسة لتقدير صدق الاختبارات محكية المرجع

تماثل إلى حد ما الطرق المتبعة في التحقق من صدق درجات الاختبارات معيارية المرجع وهي: الصدق الوصفي، والصدق الوظيفي، وصدق انتقاء النطاق السلوكي، وهذه الطرق تناظر صدق المحتوى والصدق التجريبي وصدق المفهوم في الاختبارات معيارية المرجع (علام، ٢٠٠١).

العوامل المؤثرة على الصدق:

١- الفترة الفاصلة بين الاختبار والمحك: إذا زادت الفترة تسببت في حدوث عوامل الصدفة التي تقلل من الصدق، أما إذا قلت الفترة الزمنية لأمكن التقليل من عوامل الصدفة، وبالذات في عملية التنبؤ.

٢- تفاوت أفراد العينة في الصفة: كلما زاد التباين زاد معامل الصدق.

٣- معامل ثبات المحك: كلما زاد معامل ثبات المحك زاد معامل الصدق.

٤- عدد الفقرات في الاختبار: زيادة الفقرات كما نعرف يزيد من معامل الثبات والصدق، ولكن زيادة الفقرات ليس هدفاً بحد ذاته وإنما يؤدي إلى تناول الظاهرة من كل مكوناتها.

٥- تحيز الاختبار: وهو من أهم العوامل المؤثرة في الصدق لأنه من الأخطاء المنتظمة.

الدراسات السابقة:

قارن كلاً من اليجي، واوناه، وابانوبي (Eleje, Onah & Abanobi, 2018) بين النظرية الكلاسيكية والنموذج ثلاثي المعلم لنظرية استجابة الفقرة في نتائج تحليل فقرات اختبار مهارات التحليل الكمي الاقتصادي والمكون من (٥٠) فقرة من نوع الاختيار من متعدد تم تطبيقه على عينة مكونة من (٥١٧) طالب وطالبه من طلبة الاقتصاد في السنة الثالثة. كما أظهرت النتائج أن معامل الثبات في النظرية الكلاسيكية قد بلغ (٠.٦٦) بينما بلغ في النموذج ثلاثي المعلم (٠.٨٦). وقد أوصت هذه الدراسة باستخدام النموذج ثلاثي المعلم ضمن نظرية استجابة الفقرة كنموذج أكثر فاعلية في الكشف عن الخصائص السيكومترية.

وسعت دراسة الصيخان و المومني (٢٠٢١) إلى المقارنة بين النظرية الكلاسيكية والنموذج ثلاثي المعلم لنظرية استجابة الفقرة في انتقاء فقرات اختبار تحصيلي لمادة اللغة الإنجليزية. ولتحقيق أهداف الدراسة تم بناء اختبار تحصيلي لمادة اللغة الإنجليزية للصف الأول ثانوي مكون من (٤٢) فقرة من نوع الاختيار من متعدد لأربعة بدائل. وشملت عينة الدراسة على (١٠٧٦) طالبة تم اختيارهم بالطريقة العشوائية العنقودية متعددة المراحل من (١١) مدرسة للبنات. حيث أظهرت النتائج وجود فروق ذات دلالة إحصائية بين معاملي الثبات بين النظرية الكلاسيكية والنموذج ثلاثي المعلم لصالح النظرية الكلاسيكية. وفروق ذات دلالة إحصائية بين معاملي الصدق المحكي بين النظرية الكلاسيكية والنموذج ثلاثي المعلم لصالح النموذج ثلاثي المعلم.

وكشفت دراسة روبتسيش (Robitzsch, 2021) عن مدى التكافؤ بين طريقة الدلتا سكور الكامن و النموذج ثنائي المعالم، حيث تم استخدام المنهج التحليلي لبيانات اختبار PISA عام (٢٠٠٦) وكانت عينة الدراسة ١١٠٢٣٦ طالباً وكان اختبار القراءة يتضمن ٢٨

فقرة. وأشارت النتائج أن طريقة الدلتا سكور الكامنة لديها مزايا عند تفسير النتائج يمكنها أن تستخدم كنموذج بديل للنموذج الثنائي.

وقارنت دراسة دومنجو و ديمتروف (Domingue, Dimitrov, 2021) بين تقديرات النظرية الحديثة والدلتا سكور من منظور القياس الجمعي، حيث استخدمت الدراسة المنهج التجريبي وكانت أداة الدراسة اختبار كفايات المعلمين الذي يعده المركز الوطني للقياس والمكون من ٧٩ فقرة ،وكانت عينة الدراسة ٤٥٦١٢، وأظهرت النتائج أن الدرجات المقدرة بالدلتا تنتج عدد أقل من المخالفات لمسلمات ترتيب القياس المشترك الجمعي مقارنة بتقديرات النظرية الحديثة؛ ومن الملاحظ أن جميع الطرق السابقة تحاول إيجاد طريقة تقدير تضيف لدرجات الاختبار صدق وثبات مرتفع، أيضاً في هذه الدراسة نريد الحصول على طريقة تقدير تضيف لدرجات الاختبار صدق وثبات مرتفع، ولكن عن طريق استخدام قيم المعالم الأساسية للفقرات (معلم السهولة، ومعلم التمييز) كقيم وزنية لها.

يتضح من العرض السابق للدراسات أن الهدف الرئيس لمعظم هذه الدراسات هو البحث عن أفضل طريقة لتقدير الدرجات والمقارنة بين الطرق المستخدمة في تقدير الدرجات، حيث تتفق الدراسة الحالية مع دراسة ودراسة الصيخان والمومني (٢٠٢١)، ودراسة ديمتروف (٢٠١٦)، ودراسة روبتسيش (٢٠٢١)، ودومنجو وديمتروف (٢٠٢١) في المتغير المستقل (طرق التقدير)، وتتفق دراسة أبو فودة (٢٠١٦)، ودراسة الخضر والبرصان (٢٠١٧) في النموذج الثنائي كأحد نماذج النظرية الحديثة، كما تتفق مع دراسة الصيخان والمومني (٢٠٢١) في الطريقة التقليدية كأحد طرق تقدير الدرجات. واتفقت مع دراسة ديمتروف (٢٠١٦)، ومع دراسة روبتسيش (٢٠٢١)، ومع دراسة دومنجو وديمتروف (٢٠٢١) في طرية الدلتا كأحد الطرق الوزنية المستخدمة في الدراسة. واتفقت دراسة أبوفودة (٢٠١٦)، بخصوص أداة الدراسة اختبار مادة الرياضيات مع اختلاف المرحلة العمرية. ومن ناحية عينة الدراسة فقد اتفقت مع دراسة الصيخان والمومني (٢٠٢١) تناولت المرحلة الثانوية. بينما دراسة دومنجو وديمتروف (٢٠٢١) تناولت عينة المعلمين. وتختلف في المتغير التابع مع دراسة ديمتروف التي تناولت أثر طرق التقدير على معادلة الاختبارات.

منهج الدراسة وإجراءاتها:

يتناول هذا الفصل وصفاً لمنهج وإجراءات الدراسة، ويسلط الضوء على مجتمع الدراسة وعينتها، والأساليب الإحصائية المستخدمة، وأدلة الثبات والصدق، والتحقق من افتراضات النموذج المستخدم.

منهج الدراسة:

في هذه الدراسة سيتم استخدم المنهج الوصفي (الأسلوب المقارن) لملائمته لطبيعة هذه الدراسة. يعتمد الأسلوب الوصفي على دراسة الظاهرة من خلال الواقع، مع الاهتمام بوصفها بصورة دقيقة والتعبير عنها كماً وكيفاً، حيث إن التعبير الكيفي يركز على

خصائصها، بينما التعبير الكمي يهتم بوصف خصائص الظاهرة رقمياً من حيث مقدارها أو حجمها ودرجة ارتباطها ببقية الظواهر الأخرى (عبيدات وآخرون، ٢٠١٣).

مجتمع الدراسة:

تألف مجتمع الدراسة من طلاب الصف السادس الابتدائي بالمدارس الحكومية بمدينة الهفوف للفصل الدراسي الثاني من العام الدراسي ١٤٤٢هـ، والبالغ عددهم (٢٦٧٢) طالب، في (٤٤) مدرسة حسب بيان إدارة التخطيط والتطوير وحدة المعلومات للفصل الدراسي الأول ١٤٤٢هـ. والجدول التالي يوضح المجتمع الأصلي للدراسة.

جدول (٥) مجتمع الدراسة

م	المدرسة	عدد الطلاب	م	المدرسة	عدد الطلاب	م	المدرسة	عدد الطلاب
١	ابن المشرف	٥٢	٢	الشاطبي	٥١	٣	الأمير سعود بن جلوي	٦٤
٤	ابن النفيس	٥٤	٥	الشافعي	٤٥	٦	الأمير محمد بن فهد	١٢٥
٧	ابن الهيثم	٦٩	٨	النسفي	٤٥	٩	الأمير نايف بن عبدالعزيز	٦٧
١٠	ابن خلدون	٢٥	١٠	حمزة الزيات	٥٧	١٢	البحتري	١٣٧
١٣	ابي دجانة	٦٨	١٤	الإمام عاصم	٧٤	١٥	الجارود	٢٠
م	اسم المدرسة	عدد الطلاب	م	اسم المدرسة	عدد الطلاب	م	اسم المدرسة	عدد الطلاب
١٦	احمد بن حنبل	٣٥	١٧	علي بن أبي طالب	٦٧	١٨	الخالدية	٨٣
١٩	الرافعي	٦١	٢٠	الإمام مسلم	١٣٥	٢١	السلمانية	٦٤
٢٢	عبدالرحمن السعدي	٧٠	٢٣	الفضل بن العباس	٢٦	٢٤	المباركية	٣٢
٢٥	المثنى بن حارثة	٤٧	٢٦	المروة	١٣٠	٢٧	الملك فيصل	٥٤
٢٨	النجاح	٤٠	٢٩	النجاشي	٢٥	٣٠	أنس بن مالك	٥٧
٣١	بلال بن رباح	٤٨	٣٢	حي الملك فهد	٩٩	٣٣	حي جنوب الهفوف	٣٣
٣٤	زيد بن ثابت	٧١	٣٥	سراقة بن مالك	٢٦	٣٦	سيبويه	٨٧
٣٧	عبادة بن الصامت	٤٢	٣٨	عثمان بن عفان	٥٧	٣٩	عقبة بن نافع	٤٢
٤٠	عمار بن ياسر	١١٠	٤١	فلسطين	٣٥	٤٢	مالك بن أنس	٢٥
٤٣	معاوية بن أبي سفيان	٢٤	٤٤	وادي طوى	٩٤			

عينة الدراسة:

اختيرت ١٠ مدارس بأسلوب العينة العشوائية البسيطة، حيث تألفت عينة الدراسة من (٥٢١) طالب أي ما نسبته (١٩%) من مجتمع الدراسة، اختيرت بأسلوب العينة العشوائية

البسيطة من مدارس الهفوف حيث تم ترقيم المدارس واختيار ١٠ مدراس وهي تمثل مانسبته (٢٣%) من إجمال المدارس، والجدول التالي يوضح عينة الدراسة :

جدول (٦) عينة الدراسة

م	المدرسة	عدد الطلاب	م	المدرسة	عدد الطلاب	م	المدرسة	عدد الطلاب
١	ابن خلدون	٢٥	٢	النسفي	٤٥	٣	البحتري	١٣٧
٤	عبدالرحمن السعدي	٧٠	٥	النجاح	٤٠	٦	النجاشي	٢٥
٧	حي جنوب الهفوف	٣٣	٨	سراقة بن مالك	٢٦	٩	عبادة بن الصامت	٤٢
١٠	عقبة بن نافع	٤٢						

أداة الدراسة:

اختبار تحصيلي من إعداد الباحث في مادة الرياضيات للصف السادس الابتدائي الفصل الدراسي الثاني يشتمل على الوحدات التالية (العمليات على الكسور الاعتيادية، النسبة والتناسب، النسبة المئوية والاحتمالات)، مفرداته على شكل اختيار من متعدد عددها ٢٥ فقرة.

إجراءات الدراسة:

تكونت إجراءات الدراسة من الخطوات التالية:

أ- بناء اختبار تحصيلي في مادة الرياضيات وفق خطوات بناء الاختبار التحصيلي الجيد:

١- تحديد الغرض من الاختبار:

تم بناء اختبار تحصيلي محكي المرجع في مادة الرياضيات الصف السادس الفصل الدراسي الثاني لعام ١٤٤٢هـ لقياس تحصيل الطلاب في الوحدات التالية (العمليات على الكسور الاعتيادية، النسبة والتناسب، النسبة المئوية والاحتمالات).

تحليل محتوى المادة:

تم تحليل محتوى مادة الرياضيات للفصل الدراسي الثاني في الوحدة السادسة والسابعة والثامنة من المقرر، واشتقاق الأهداف التعليمية حيث تم صياغة (٤٨) هدفاً تعليمياً اشتقت من (١٣) موضوعاً موزعة على الفصول الثلاثة من موضوعات الكتاب، وصنفت الأهداف التعليمية وفقاً لتصنيف بلوم، وصيغت بطريقة مدمجة بين جهد الباحث وفي ضوء الأهداف التعليمية الوزارية لكل درس؛ وبالاستعانة بإرشادات الكتاب، وتمت الصياغة وفق مواصفات الهدف الجيد القابل للملاحظة والقياس بحيث تقيس الفقرة الاختبارية ناتج تعليمي واحد بسيط وغير مركب،

جدول المواصفات :

تم إعداد جدول مواصفات بناء الاختبار بتوزيع الأهمية النسبية لكل وحدة اعتماداً على عدد الحصص التي تلزم المعلم في شرح كل وحدة، والأهمية النسبية لكل مستوى معرفي من

مستويات بلوم. وتعتبر هذه الخطوة والخطوة السابقة من مؤشرات الصدق(صدق المحتوى).

٢- صياغة الفقرات:

بعد الرجوع إلى مقرر مادة الرياضيات للصف السادس الابتدائي الفصل الدراسي الثاني، وتحديداً الوحدات التالية (العلميات على الكسور الاعتيادية، النسبة والتناسب، النسبة المئوية والاحتمالات)، تم صياغة ٢٥ فقرة من نوع الاختيار من متعدد، بأربع بدائل، منها بديل واحد فقط صحيح. رُعيت فيها الإرشادات والقواعد الفنية للصياغة، مع استقلالية المفردات بحيث لا تعتمد إجابة الفقرة على إجابة فقرة أخرى، وكانت الفقرات شاملة مواضيع الاختبار، تم عرض الاختبار على محكمان (مشرفة/ معلم) لا تقل خبرتهما عن ١٥ سنة، ملحق (٥) وطلب منهما تحديد ما إذا كان السؤال (الفقرة) مناسبة جداً أو مناسبة أو مناسبة إلى حد ما أو غير مناسبة بحيث أن اختيار مناسبة جداً أو مناسبة تصنف إلى قوي الانتماء، ومناسبة إلى حد ما أو غير مناسبة تصنف إلى ضعيف الانتماء. والجدول التالي يوضح ذلك

جدول (٧) معامل صدق المحتوى

المحكمان		المحكم الأول		المجموع
		قوي الانتماء	ضعيف الانتماء	
المحكم الثاني	قوي الانتماء	١٨ (ج)	٤ (د)	٢٢
	ضعيف الانتماء	٣ (ب)	٠ (أ)	٣
المجموع		٢١	٤	٢٥

ولحساب معامل صدق المحتوى = ج ÷ (أ+ب+ج+د) (الطريري، ٢٠١٤)

معامل صدق المحتوى= ١٨ ÷ (١٨+٤+٣+٠) = ٠.٧٢

ولايوجد اتفاق على أي سؤال بأنه ضعيف الانتماء.

كما عُرض الاختبار بصورته الأولية – المؤلف من (٢٥) فقرة – على مجموعة من المحكمين من ذوي الاختصاص للتحقق من صدق المحتوى للاختبار بحيث يتم الإبقاء على الفقرة التي تصل نسبة الاتفاق عليها (٨٠%) (علام، ٢٠١١؛ مجيد، ٢٠١٤). وقد تم عرض الفقرات والأهداف وتحديد معايير الحكم على الاختبار (ارتباط الفقرة بالهدف التي تقيسه، وضوح الأسئلة، سلامة الصياغة، مناسبة البدائل) على (٢) من معلمي الرياضيات، و (٣) من المشرفين التربويين المتخصصين في الرياضيات، و (٢) من المتخصصين في طرق تدريس الرياضيات، و(٢) من المتخصصين في القياس والتقويم، وتم التعديل في ضوء مرئياتهم ملحق رقم (٦). كما تم الأخذ بآراء المحكمين، حيث قدموا مجموعة من الاقتراحات، ساعدت الباحث في إعادة النظر لعدد من الفقرات التي كانت تحتاج إلى إعادة صياغة، ويعتبر هذا الإجراء من مؤشرات الصدق (صدق المحكمين).

٣- التجريب والتطبيق الأولى للاختبار:

تم تطبيق الاختبار بصورته الأولية (٢٥) فقرة على عينة استطلاعية من خارج عينة الدراسة بلغ عددهم (١٤٧) طالب وفق نموذج الأسئلة والإجابة المرفقة في الملحق رقم (٧

)، إذ تم تحديد حجم العينة الاستطلاعية لإمكانية إجراء التحليل العاملي الاستكشافي، وكان التطبيق يوم الاثنين ١٤٤٢/٨/٢هـ حيث انحصر الغرض من التطبيق على: التأكيد على مناسبة الزمن الذي يستغرقه الطلاب في الإجابة عن فقرات الاختبار، والتعرف إلى مستوى صعوبة الفقرات وقدرتها التمييزية، ومدى وضوح الصياغة اللغوية للفقرات، ومناسبة البدائل، وكذلك حساب معاملات الصدق والثبات، حيث تم الاستعانة بالبرامج التالية SPSS, ITMAN، وقد استجاب على الاختبار (١٣٤) طالب وظهرت المخرجات والنتائج التالية:

جدول (٨) معاملات الصعوبة والتمييز

رقم الفقرة	معامل الصعوبة	معامل البوينت بايسيريال	رقم الفقرة	معامل الصعوبة	معامل البوينت بايسيريال
١	٠.٤٥٥	٠.٤٨٣	١٤	٠.٥٠٧	٠.٥٦٧
٢	٠.٨٨٨	٠.٣١٧	١٥	٠.٦٠٤	٠.٤٦٦
٣	٠.٩١	٠.١٧٨	١٦	٠.٧٦١	٠.٣٧٣
٤	٠.٤٠٣	٠.٣٠٦	١٧	٠.٧٤٦	٠.٥٥٤
٥	٠.٩٠٣	٠.٣٤٨	١٨	٠.٦٥٧	٠.٢٩٨
٦	٠.٥٩٧	٠.٤٣٧	١٩	٠.٤٠٣	٠.٢٧٩
٧	٠.٦٢٧	٠.٢٩٥	٢٠	٠.٤٨٥	٠.٤٩٨
٨	٠.٨٨١	٠.٢٢٢	٢١	٠.٣٨٨	٠.٤٥٨
٩	٠.٥٧٥	٠.٤٩	٢٢	٠.٥٥٢	٠.٤٤٩
١٠	٠.٤٧٨	٠.٤٦٣	٢٣	٠.٣٩٦	٠.٥٧٤
١١	٠.٧٦٩	٠.٤٨٥	٢٤	٠.٥٩٧	٠.٥٣٧
١٢	٠.٧٥٤	٠.٤٥٢	٢٥	٠.٧٦١	٠.٤٦٤
١٣	٠.١٦٤	٠.٢٦٠			

كما يتضح من الجدول (٨) قيمة كل من معامل الصعوبة (نسبة الطلاب الذين أجابوا إجابة صحيحة عن الفقرة)، ومعامل التمييز (معامل البونت بايسيريال) وذلك لكل فقرة من فقرات الاختبار بصورته المطبقة على عينة التجريب، حيث تراوحت قيم معاملات الصعوبة بين (٠.١٦- ٠.٩١) بمتوسط حسابي (٠.٦١)، وهذا يشير إلى أن الاختبار متوسط الصعوبة كما تراوحت قيم معاملات التمييز بين (٠.١٨ - ٠.٥٧) بمتوسط حسابي (٠.٤١) وهو أعلى من ٠.٢ الذي يشير إلى الحد الأدنى لقبول معاملات التمييز وفق محك آيبل (Ebel) الذي يصنف قيم هذا المؤشر وفق الجدول (٥) (كروكر والجينا، ٢٠١٧)

جدول (٩) دلالات قبول المفردات وفق معامل تمييز بوينت بايسيريال

دلالة القيمة	قيمة المؤشر
المفردة تلبي الغرض المقصود	أكبر من ٠.٤
المفردة تتطلب مراجعة بسيطة وقد لا تتطلب	من ٠.٣ إلى ٠.٣٩
المفردة تقع على الحد الفاصل وتحتاج إلى مراجعة	٠.٢ إلى ٠.٢٩
تحذف المفردة أو تحتاج إلى مراجعة تامة لها	أقل من ٠.١٩

جدول رقم (١٠) معاملات الثبات.

معاملات ثبات درجات الاختبار	ألفا لكرونباخ	أوميجا	معامل التجزئة النصفية المصحح بالطريقة		
			العشوائية	الأول- الأخير	الفردي -الزوجي
	٠.٨٦٢	٠.٨٦٥	٠.٨٢٩	٠.٨٢٢	٠.٨٦

يتضح من الجدول السابق أن معاملات الثبات تراوحت بين ٠.٨٢ – ٠.٨٧ وهي قيم مقبولة.

وتم تطبيق حساب الثبات بإعادة الاختبار على عينة متاحة عددها (٥٦) من أصل ١٣٤ وكان معامل الارتباط بين التطبيقين ٠.٧١ وهو دال عند ٠.٠١ . كما أن معامل الارتباط بين الفئات ٠.٨٢ وكانت فترة التطبيق الأولى بتاريخ (١٤٤٢/٨/٢هـ) يوم الاثنين وفترة التطبيق الثاني يوم الاثنين ١٤٤٢/٨/١٦هـ ، وفاصل زمني يقدر بـــ ١٤ يوم لتقليل من إمكانية تذكر الإجابات يكون الاعتماد على القدرة التحصيلية للمفحوص مع ملاحظة أنه في الاختبار الأول لم يقدم للمفحوصين أي تغذية راجعة.

الصدق البنائي (صدق المفهوم):

تم استخدام التحليل العاملي الاستكشافي لتحقق من الصدق البنائي للاختبار، وأشارت نتائج التحليل العاملي إلى تنوع قيم مصفوفة الارتباط، وبلغت قيمة المحدد ٠.٠٠١ أكبر من صفر مما يشير إلى أن مصفوفة الارتباط لا تحتوي على معاملات ارتباط عالية وقيمة كايز ٠.٧٩ وهي مؤشر جيد لكفاية العينة ومعامل بارتليت دال وهذا مؤشر لكفاية العينة.

جدول (١١) كفاية العينة

قيمة اختبار كايزر- ماير- وأولكن لكفاية العينة ٠.٧٩٧	
اختبار بارتليت للكروية	٩٠٠.٨٧
درجة الحرية	٣٠٠
الدلالة	٠.٠٠

بإيجاد المصفوفة الصورية (Anti-image Matrices) للاختبار كانت الارتباطات القطرية في مقياس كفاية المعاينة وملاءمتها لكل فقرة (MSA) Measures of Sampling Adequacy أكبر من(٠.٥٠) وهذا يدل على أن العينة مناسبة وكافية لكل فقرة والمصفوفة مناسبة لإجراء التحليل العاملي (تيغزة، ٢٠١١؛ الشافعي، ٢٠١٤). وجميع القيم تشير إلى إمكانية استخدام التحليل العاملي.

وقد استخدم التحليل العاملي الاستكشافي بطريقة المكونات الأساسية أو الرئيسة (PCA) Principal Components Analysis حيث تتميز هذه الطريقة بدقتها وقدرتها على الوصول إلى حل يتفق مع محك أدنى مربعات للمصفوفة الارتباطية؛ وهو أحد المحكات الرياضية التي تلاقي قبولاً واضحاً في مجال الأساليب التلخيصية للعلاقات بين المتغيرات (حسن، ٢٠٠٤)، كما يشير تيغزة (٢٠١٢) إلى أن هذه الطريقة تستعمل التباين الكلي بما في ذلك التباين الخاص وتباين الخطأ.

وتم التحقق من أحادية البعد باعتماد خمسة محكات، المحك الأول: قيم معاملات الثبات للاتساق الداخلي، والمحك الثاني: الجذر الكامن (Eigen Value)، والمحك الثالث: نسبة التباين المفسر (Percent of Variance Extracted Criterion)، والمحك الرابع: اختبار رسم المنحدر لكاتيل (Kattell Scree Test)، والمحك الخامس: التشبعات العاملية (Loadings) وقيم الشيوع أو الاشتراكيات (Communalities)، وبيانها كالتالي:

المحك الأول: قيم معاملات الثبات للاتساق الداخلي حيث أشار دينك وتفكول (٢٠١١) Dennick and Tavakol أن قيم معاملات الثبات التي تزيد عن (٠.٧٠) تُعد مقبولة لأغراض تحقق افتراض أحادية البعد؛ وأن القيم التي تقل عن ذلك تشير إلى ارتباطات ضعيفة بين مفردات الاختبار مما يدل على عدم تجانسها. وبالنظر إلى معامل الثبات (٠.٨٦) فإن تحقق أحادية البعد للاختبار وفق هذا المحك متحقق.

أما المحك الثاني: الجذر الكامن حيث حدد جلورفيلد (١٩٩٥) Glorfeld أن تكون نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني أكبر من أو تساوي (٢) للدلالة على أحادية البعد (Hatti, 1984; Lord, 1980).

والمحك الثالث: نسبة التباين المُفَسر، وقد كانت نتائج المحكين الثاني والثالث على الاختبار موضحة في الجدول التالي: وقد تم الحصول على ٧ عوامل حيث بلغت الجذور الكامنة أكبر من ١ وفق لمقياس كايزر

جدول (١٢) الجذور الكامنة

العامل	الجذور الكامنة		
	المجموع	نسبة التباين	النسبة التراكمية
١	٥.٩٥٧	٢٣.٨٢٨	٢٣.٨٢٨
٢	٢.٠٣٦	٨.١٤٥	٣١.٩٧٤
٣	١.٥٤٩	٦.١٩٦	٣٨.١٦٩
٤	١.٣٩٦	٥.٥٨٤	٤٣.٧٥٣
٥	١.٣٣٥	٥.٣٤	٤٩.٠٩٣
٦	١.١١٦	٤.٤٦٤	٥٣.٥٥٧
٧	١.٠٣١	٤.١٢٥	٥٧.٦٨١

من الجدول السابق يوجد (٧) عوامل زادت قيم جذورها الكامنة عن الواحد الصحيح مسؤولة عن تفسير التباين بين المفحوصين؛ إذ فسرت العوامل مجتمعة نسبة قدرها (٥٧.٦٨%)، فقد كانت قيمة الجذر الكامن للعامل الأول (٥.٩٦)؛ وفسر ما نسبته (٢٣.٨٣%) من تباين الأداء بين الأفراد على نموذج الاختبار، وهذه النسبة تشكل ما مقداره (٤١.١٣%) من التباين الكلي الذي تفسره العوامل الـــ(٧) مجتمعة.

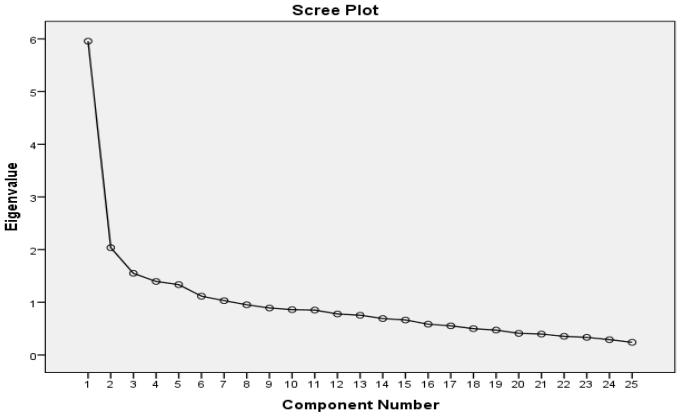
وكانت قيمة الجذر الكامن للعامل الثاني (٢.٠٤)؛ وفسر ما نسبته (٨.١٥%) من تباين الأداء، كما يلاحظ أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني بلغت

(٢.٩٢) ضعفاً، وهي نسبة قريبة جداً من الثلاثة أضعاف الأمر الذي يُعد مؤشراً على تحقق سمة أحادية البعد في الأداء على الاختبار.

والمحك الرابع: اختبار رسم المنحدر لكاتيل، وهي طريقة تقوم على الجذور الكامنة وذلك برسم المحور السيني الأفقي الذي يدل على العوامل في مقابل الجذور الكامنة التي ترسم نقاطها على المحور الصادي العمودي، وينتج عن ذلك منحنى ينطلق من أعلى جذر عند العامل الأول ثم يأخذ في التناقص إلى أن يصل نقطة ما تقابل عاملاً معيناً تتباطأ درجة انحداره عندها كاسراً وتيره الانخفاض (تيغزة، ٢٠١٢).

وكانت نتائج المحك الرابع اختبار رسم المنحدر للاختبار كما في الشكل التالي:

شكل (٣) اختبار رسم المنحدر



من الشكل السابق يظهر عامل واحد فقط قبل نقطة القطع الأولى التي تمثل تباطؤ انحدار المنحنى ليتخذ منحنى مستوي نسبياً، بمعنى أن قيمة الجذر الكامن للعامل الأول ينفرد بقيمة مرتفعة نسبياً مقارنة ببقية العوامل الأخرى التي تبدو قيمها صغيرة نسبياً ومتقاربة، مما يشير إلى تحقق سمة أحادية البعد في هذا الاختبار. ويشير الرسم البياني إلى وجود فرق كبير بين العامل الأول وبقية العوامل وهذا يدل على أن هذا الاختبار يقيس سمة واحدة .

المحك الخامس: التشبعات العاملية وقيم الشيوع، أشار علام (٢٠٠٣) أنه يفضل الاقتصار على التشبعات التي تمثل نسبة لا تقل عن (٢٥%). كما ذكر تيغزة (٢٠١٢) أن الكثير من البحوث استعملت القاعدة المألوفة التي تعتبر أن التشبع (٠.٣) يدل على القيمة الفاصلة بين التشبع الهام وغير الهام. وتفضل بعض المراجع ألا يقل التشبع عن (٠.٤٠) (Stevens,2002). والجدول التالي يوضح تلك القيم: تشبعات الفقرات على العوامل لنموذج الاختبار

جدول رقم (١٣) : تشبعات الفقرات على العوامل وقيم الشيوع للاختبار

فقرة	1	٢	3	5	6	الشيوع
1	0.546					0.443
2		0.393				0.587
3				-0.393		0.476
4			-0.514			0.597
5	0.422					0.566
6	0.513					0.629
7				0.58		0.579
8		0.573				0.702
9	0.566					0.529
10	0.532					0.610
11	0.557					0.59
12	0.53					0.669
13			0.363			0.465
14	0.627					0.617
15	0.523					0.518
16	0.446					0.591
17	0.629					0.623
18			-0.608			0.580
19					0.476	0.581
20	0.571					0.583
21	0.52					0.541
22	0.512					0.489
23	0.636					0.616
24	0.609					0.605
25	0.531					0.631

وبالنظر إلى الجدول السابق نجد أن أقل تشبع (٠.٣٦٣) للفقرة (١٣) على العامل الثالث، وأعلى تشبع (٠.٦٣٦) للفقرة (٢٣)، ويلاحظ أن أغلب تشبعات الفقرات على العوامل أعلى من (٠.٤) وهي تشبعات مرتفعة جداً وفق المحك المفضل، كما يبين الجدول أن هنالك (١٧) فقرة تشبعت على العامل الأول وتمثل ما نسبته (٦٨%) من فقرات الاختبار، و فقرتين تشبعت على العامل الثاني، و(٣) فقرات تشبعت على العامل الثالث، و فقرتين على العامل الخامس، وفقرة على العامل السادس، وحيث أن هذه العوامل تشبع عليها عدد قليل من الفقرات فإنها لا تؤخذ بعين الاعتبار، إضافة إلى أن العوامل الرابع والسابع لم تتشبع عليهم أي فقرة. مما يؤكد البعدية لقياس القدرة الرياضية المتمثلة في التحصيل الدراسي. أما قيم الشيوع فقد تراوحت بين (٠.٤٤٣) و (٠.٦٦٩) للفقرتين (١) و (١٢) على الترتيب، وأغلب

القيم أكبر من (٠.٥٠) الحد الأدنى المقبول لقيم الشيوع ما عدا (١، ٣، ١٣ ، ٢١) فقد كانت قيم الشيوع الخاصة بها (٠.٤٤ ، ٠.٤٨ ، ٠.٤٧ ، ٠.٤٩) على الترتيب وهي قيم شيوع قريبة من المحك المحدد.

ومما سبق يتضح أن هنالك عاملاً رئيساً منفرداً بتاثيره عن بقية العوامل الأخرى، ويدل ذلك على أن الاختبار يقيس مفهوماً واحداً.

صدق المحك التنبؤي :

تم حساب الصدق التنبؤي عن طريق معامل ارتباط بيرسون بين الاختبار التحصيلي ودرجات الطلاب في الفصل الدراسي الثاني حيث بلغت قيمة معامل الارتباط (٠.٤) ومستوى الدلالة المرتبط به أقل من (٠.٠١). وتم تصحيح معامل الصدق عن طريق المعادلة التالية:

معامل الصدق المصحح = $\frac{\text{معامل الصدق المحسوب}}{\sqrt{\text{معامل ثبات المحك} \times \text{معامل ثبات الاختبار}}}$ (الطريري، ٢٠١٤).

معامل الصدق المصحح= ٠.٤ ÷ (٠.٥٤×٠.٨٦) = ٠.٤÷٠.٦٣= ٠.٦٣

الصدق التمييزي :

تم ترتيب العينة تنازلياً وفقاً لدرجة المحك (درجات الطلاب في الفصل الدراسي الثاني)، واستخدم اختبار (ت) لمعرفة الفروق بين المجموعة العليا والمجموعة الدنيا. حيث تم اختيار عدد ٣٦ طالب من ذوي الدرجات العليا على المحك و٣٦ طالب من ذوي الدرجات الدنيا على المحك، ومن خلال نتائج اختبار (ت) يتضح أن متوسط درجات المجموعة العليا يساوي ١٧.٣٩ بانحراف معياري ٥.٣٥ ومتوسط درجات المجموعة الدنيا يساوي ١٢.٧٥ بانحراف معياري ٥.٢٧ ، وبالنظر إلى قيمة (ف) التي تحدد مدى تجانس العينتين نجد انها تساوي ٠.٣٣٣ وبمستوى دلالة ٠.٥٦ وهذا يشير إلى أن العينتين متجانستين (تساوي التباين) لذا فإن قيمة (ت) في حالة تساوي التباين تقدر ٣.٧ وهي دالة إحصائياً حيث بلغت قيمة مستوى الدلالة ٠.٠ وهذا يشير إلى فرق دال بين المجموعتين العليا والدنيا لصالح المجموعة ذات المتوسط الأعلى. (يتضح من خلال نتائج اختبار (ت) وجود فروق دالة إحصائياً في الأداء بين المجموعة العليا والدنيا لصالح المجموعة العيا، وهذا يدل على قدرة الاختبار على التمييز بين الطلاب.

الاتساق الداخلي :

من النظر إلى معاملات الارتباط بين الفقرات نجد أن متوسط الارتباط بين الفقرات يساوي (٠.٢) تقريباً وأعلى قيمة بلغت (٠.٥٢) وأقل قيمة (٠.٠٩-) وبانحراف معياري يساوي ٠.١١.

كما أن الجدول التالي يوضح معاملات الارتباط بين الفقرات والدرجة الكلية

جدول (١٤) معاملات الارتباط بين الفقرات والدرجة الكلية

الفقرة	الارتباط	الفقرة	الارتباط	الفقرة	الارتباط
١	**٠.٥٥	٢	**٠.٣٦٨	٣	**٠.٢٢٨
٤	**٠.٣٨٦	٥	**٠.٣٩٤	٦	**٠.٥٠٨
٧	**٠.٣٧٥	٨	**٠.٢٧٨	٩	**٠.٥٥٦
١٠	**٠.٥٣٣	١١	**٠.٥٤٢	١٢	**٠.٥١٣
١٣	**٠.٣٢٢	١٤	**٠.٦٢٦	١٥	**٠.٥٣٤
١٦	**٠.٤٣٩	١٧	**٠.٦٠٧	١٨	**٠.٣٧٦
١٩	**٠.٣٦١	٢٠	**٠.٥٦٥	٢١	**٠.٥٢٦
٢٢	**٠.٥١٩	٢٣	**٠.٦٣١	٢٤	**٠.٥٩٨
٢٥	**٠.٥٢٤				

(**: مستوى الدلالة الإحصائية ٠.٠١= a ، العينة = ١٣٤)

يتضح من الجدول السابق أن معاملات الارتباط بين الفقرات والدرجة الكلية تراوحت من(٠.٢٣- ٠.٦٣) للفقرتين (٣ – ١٤)على الترتيب. وجميعها دالة عند مستوى ٠.٠١ . بمتوسط (٠.٤٧) ، وبلغت نسبة التباين المشترك من خلال معامل التحديد (R^2) لأقل وأعلى ارتباط (٥%) ، (٤٠%).

التخمين:

تم التحقق من شرط عدم وجود التخمين تبعاً لهامبيلتون وسوامينثال (١٩٨٥) يتم التحقق من شرط عدم وجود التخمين من خلال فحص صعوبة الفقرات، وتحديد الفقرات الصعبة، وفحص أداء أصحاب القدرة المنخفضة على هذه الفقرات، فإن كانت نسبة إجاباتهم بشكل صحيح عنها أقل عن نسبة التخمين العشوائي ، كان ذلك دليلاً على تحقق شرط عدم وجود التخمين (Hambleton, Swaminathan, 1985). وفي هذه الدراسة اختيرت أصعب الفقرات التي تتراوح معاملات صعوبتها من(٠.١٦ – ٠.٣٩) وعددها ثلاثة فقرات (١٣، ٢١، ٢٣) من فقرات الاختبار (٢٥) فقرة ، وتم فحص أداء أصحاب أقل (٢٧%) درجة كلية على الاختبار وعددهم (٣٦) طالب على هذه الفقرات الصعبة، حيث يوضح جدول (٥) أن نسبة منخفضي القدرة الذين أجابوا إجابة صحيحة عن الفقرات الصعبة في الاختبار نسبة منخفضة، مقارنة بنسبة الإجابة العشوائية عن هذه الفقرات (٠.٢٥) وهي نسبة التخمين العشوائي على فقرة الاختيار من متعدد ذات البدائل الأربعة، لذلك يمكن القول بتحقق شرط عدم وجود التخمين.

جدول (١٥) التحقق من عدم وجود التخمين

رقم الفقرة	عدد الطلاب	نسبة الطلاب الذين اجابوا إجابة صحيحة
١٣	٣٦	٠.٠٩
٢١	٣٦	٠.١١
٢٣	٣٦	٠.١٤

الزمن المحدد للاختبار :

عرض الاختبار على مجموعة من المتخصصين عددهم ٦ لتحديد الزمن الكافي للاختبار وتراوحت تقديراتهم لزمن الكافي بين (١١ - ٦٠) دقيقة ، وعند حساب المتوسط ٢٦.٥ دقيقة، ونظراً لوجود قيم متطرفة فتم حساب الوسيط المقدر ٢٢.٥دقيقة وبعد تطبيق الاختبار على العينة الاستطلاعية وحساب متوسط الزمن المستغرق للانتهاء من الاختبار ٢٣ دقيقة ، وفي ضوء ذلك تم تحديد زمن الاختبار ٣٠ دقيقة لمناسبته زمن الحصة الافتراضية.

تم إعداد الاختبار في صورته النهائية (٢٥) فقرة، الكترونياً تضمنت تعليمات الاختبار وطريقة الإجابة على فقراته. وتم التواصل مع مدراء ومعلمي المدارس العشرة المختارة بهدف ضبط عملية التطبيق وتوحيد ظروفه وإجراءاته، وكذلك الوصول لأقصى أداء لدى الطلاب عن طريق تحفيز معلمي المادة لطلابهم وحثهم على أداء الاختبار حيث يعتبر مراجعة للاختبار النهائي.

تطبيق الصورة النهائية من الاختبار الكترونياً على أفراد عينة الدراسة (٥٢١) طالب، يوم الاثنين ١٤٤٢/٨/٢٥هـ . وتمت برمجة بداية الاختبار ونهايته إلكترونياً، وكانت مدة الاختبار (٣٠) دقيقة فقط (٣:٣٥ – ٤:٠٥) مساءً .

- تصحيح الاختبار، ومراجعة وفحص استجابات الطلاب، وتحليل البيانات باستخدام البرامج الإحصائية المناسبة. حيث وصل عدد الاستجابات على الاختبار (٤٦٦) استجابة.
- فحص الافتراضات التي يتطلبها النموذج اللوجستي ثنائي المعلمة (2PLM) وهي: أحادية البعد، والاستقلال الموضعي، والمنحنى المميز للفقرة، والتحرر من السرعة.
- فحص مطابقة البيانات (الأفراد والفقرات) للنموذج اللوجستي الثنائي المعلمة.

ثبات أداة الدراسة (Reliability):

للتحقق من ثبات درجات الاختبار وحيث أن درجات مفردات الاختبار ثنائية الاستجابة (٠) أو (١) فإن معامل الثبات المناسب للاتساق الداخلى يتم بإستخدام معادلة كيودروريتشاردسون ٢٠ (KR-20) التي تعد حالة خاصة من معادلة ألفا لكرونباخ، فمعامل ألفا لكرونباخ تمثل الصيغة العامة لـ (KR-20) (آنستازي وأورينا، ٢٠١٥ ؛ وليفينجسون، ٢٠١٣) ، وفي حالة الاستجابات الثنائية يكون معامل ثبات درجات الاختبار ألفا لكرونباخ يكافئ (KR-20) (Thonmpson, 2003). وبمراجعة دليل البرنامج الإحصائي (SPSS) يتضح أن معامل ألفا بالنسبة للبيانات الثنائية يكافئ معامل كيودر وريتشاردسون (٢٠) (حسن، ٢٠٢٠)، وعليه تم حساب معاملات الثبات باستخدام معامل ألفا لكرونباخ بافتراض تحقق اشتراطاته، حيث بلغ معامل الثبات ٠.٨٩ وهو معامل ثبات جيد مقارنة بالنسب المقبولة في مثل هذه الدراسات. وباستخدام معامل أوميقا ٠.٨٩ أي أن أداة الدراسة على درجة عالية من الدقة والاتساق (الدوسري، ٢٠١٨)، ويرى أبو علام (٢٠٠٦) أن يكون

معامل الثبات أعلى ما يمكن، فإذا زادت عن قيمة (٠.٨) كان أفضل في مقاييس الميول والاتجاهات، أما معامل الثبات في الاختبارات المعرفية كالتحصيل الدراسي والاستعداد فيجب أن تكون قريبة من (٠.٩) ويفضل أن تزيد عن هذه القيمة، ويرجع ذلك إلى طبيعة الأسئلة التي تتعلق بهذا النوع من الاختبارات.

صدق الأداة (Validity):

لأهمية الخصائص السيكومترية للأداة فقد تم التأكد من أدلة الصدق وفق المنظور التكاملي لمفهوم الصدق في وثيقة المعايير الحديثة كما وردت في (رينولدز وليفينجستون، ٢٠١٣)، بالتحقق من صدق نموذج الاختبار بثلاث طرق؛ فكلما تعددت الأدلة على صدق الاختبار ومصادرها كان ذلك أفضل.

فالطريقة الأولى: صدق المحتوى (Content Vaildity) من خلال تطبيق جدول المواصفات والمحكمين؛ ويستخدم للتأكد من صدق اختبارات التحصيل (مجيد، ٢٠١٤)، والطريقتان الثانية والثالثة: صدق المفهوم (Construct Validity) أو صدق التكوين الفرضي باستخدام الصدق العاملي (Factorial Validit) وإجراء التحليل العاملي الاستكشافي، وصدق الاتساق الداخلي (Internal Validity) بإيجاد معاملات الارتباط بين المفردات والدرجة الكلية، وصدق المحك (حسن، ٢٠٠٤؛ تيغزة، ٢٠١١؛ الطريري، ٢٠١٤؛ مجيد، ٢٠١٤).

التحقق من افتراضات النظريتين: النظرية الكلاسيكية والنظرية الحديثة:

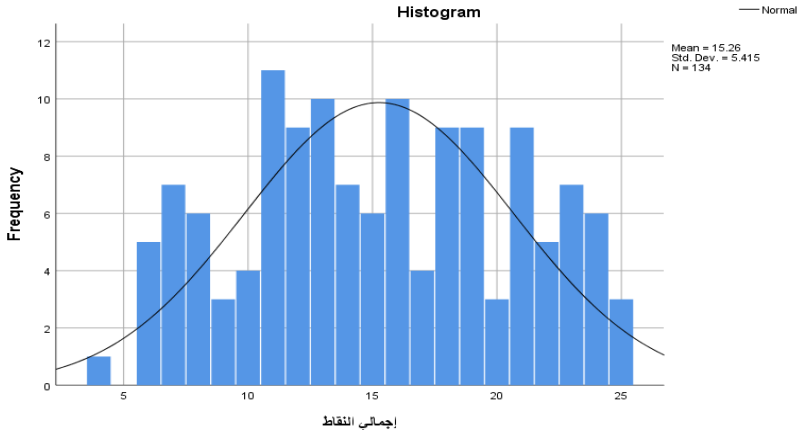
أحد الافتراضات العملية في النظرية الكلاسيكية (CTT) التوزيع الطبيعي، وحيث إن حجم عينة الدراسة أكبر من (٣٠) طالب، حيث كانت (٤٦٦) طالب، واستناداً إلى نظرية النهاية المركزية (Central Limit Theorem) (دودين، ٢٠١٣). فإن افتراض التوزيع الطبيعي لعينة الدراسة في النظرية التقليدية للقياس متحقق بقدر مناسب وكاف. كما توجد بعض المؤشرات الإحصائية والبيانية التي تدل على اعتدالية التوزيع، والجدول التالي يوضح الإحصاءات الوصفية لدرجات الاختبار:

جدول (١٦) الإحصاءات الوصفية

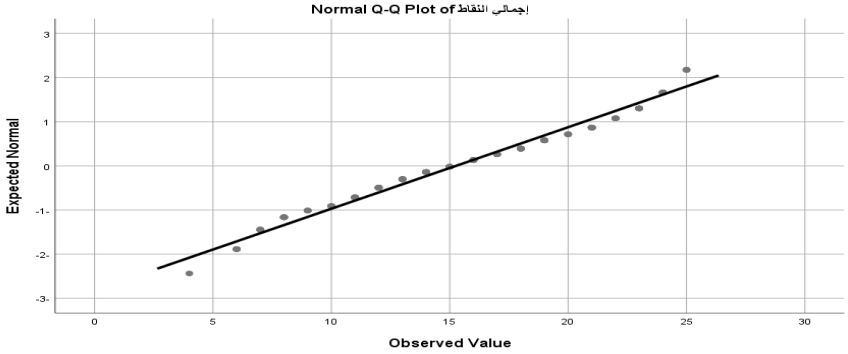
أقل درجة	أعلى درجة	المدى المطلق	الرباعي الأول	الوسيط	الرباعي الثالث	المدى الربيعي	المتوسط
٤	٢٥	٢١	١١	١٥	١٩.٢٥	٨.٢٥	١٥.٢٦

يلاحظ من الجدول السابق أن قيمة المتوسط قريبة جداً من الوسيط وهذا مؤشر لقرب توزيع الدرجات من التوزيع الطبيعي. وبفحص توزيع الدرجات بيانياً تم إيجاد المدرج التكراري، والرسم البياني للنقاط الاحتمالية لمدى ملاءمة البيانات والأشكال التالية توضح ذلك:

شكل (٤) : المدرج التكراري لتوزيع درجات الاختبار



شكل (٥) : ملاءمة الدرجات لنموذج الاختبار



المدرج التكراري لتوزيع درجات الاختبار يشير إلى أن الدرجات توزعت بشكل طبيعي بينما الرسم البياني يوضح مدى ملائمة النقاط الاحتمالية الطبيعية، وان القيمة الملاحظة لكل درجة قد تم وضعها في ضوء القيمة المتوقعة من التوزيع الطبيعي، ويدل الخط المستقيم على اعتدالية التوزيع.

افتراضات النظرية الحديثة:

هنالك أربعة افتراضات أساسية في النظرية الحديثة تتحدد في أحادية البعد، والاستقلال الموضعي، وخصائص المنحنى المميز للفقرة، والتحرر من السرعة، وبعد تحليل بيانات العينة الأساسية، تم التحقق من افتراضات النظرية على النحو الآتي:

١- أحادية البعد (Unidimensnsioa;):

تفترض نظرية (IRT) وجود قدرة أو سمة واحدة فقط يمكن أن تفسر أداء الفرد في اختبار ما، وتم تحقق هذا الافتراض من خلال الخمسة محكات السابقة، المحك الأول: قيم معاملات الثبات للاتساق الداخلي، والمحكات الأربعة المستخرجة من التحليل العاملي الاستكشافي المحك الثاني: الجذر الكامن، والمحك الثالث: نسبة التباين المفسر، والمحك الرابع: اختبار رسم المنحدر لكاتيل، والمحك الخامس: التشبعات العاملية وقيم الشيوع، ولفحص أحادية البعد كانت النتائج على النحو الآتي:

المحك الأول: قيم معاملات الثبات للاتساق الداخلي، حيث تم استخراج معامل ثبات درجات الاختبار ألفا لكرونباخ إذ بلغ (٠.٨٩) وهو معامل ثبات مرتفع لأنها أكبر من (٠.٧٠) وتدل على أحادية البعد.

ولتحقق المحكات الأربعة الآخرى تم فحص مدى مناسبة البيانات للتحليل العاملي وتوفر افتراضاته في مصوفة الارتباطات، والجدول التالي يوضح المحكات المستخدمة لتحقق افتراضات التحليل العاملي للاختبار.

جدول (١٧): محكات تحقق افتراضات التحليل العاملي

المحك		
KMO> ٠.٥	BarTlett's	MSA>٠.٥
٠.٩١٩	٣٢٦٤.٧١٥Chi- Sguare= DF= 300 Sig=0.00	أقل قيمة = ٠.٨٧ أعلى قيمة = ٠.٩٥ المتوسط= ٠.٩٢

يوضح الجدول (١٧): أن قيمة محك كايزر- ماير- أولكين (KMO) Kaiser- Meyer- Olkin Measure of Sampling Adequacyهي أكبر من (٠.٥) مما يدل على أن عينة الدراسة الاستطلاعية كافية وملائمة؛ وأن مصفوفة معاملات الارتباط مناسبة لإجراء التحليل العاملي، كما أن اختبار بارتليت للكروية (Bartlett's Test of Sphericity) دال إحصائياً (Chi- Sguare= 3264.715، DF= 300، Sig= 0.000) أي أن مصفوفة معاملات الارتباط في الاختبار ليست مصفوفة الوحدة، وبإيجاد المصفوفة الصورية (Anti-image Matrices) للاختبار كانت الارتباطات القطرية في مقياس كفاية المعاينة وملاءمتها لكل متغير Measures of sampling Adequacy (MSA) أكبر من (٠.٥٠) وهذا يدل على العينة مناسبة وكافية لكل متغير والمصفوفة مناسبة لإجراء التحليل العاملي (حسن، ٢٠٠٤؛ تيغزة، ٢٠١١؛ الشافعي، ٢٠١٤).

وقد استخدم التحليل العاملي الاستكشافي بطريقة المكونات الأساسية أو الرئيسة (PCA) Principal Components Analysis، حيث تتميز هذه الطريقة بدقتها وقدرتها على الوصول إلى حل يتفق مع محك أدنى مربعات للمصفوفة الارتباطية؛ وهو احد المحكات الرياضية التي تلاقي قبولاً واضحاً في مجال الأساليب التلخيصية للعلاقات بين المتغيرات (حسن، ٢٠٠٤)

يقول باهي وآخرون (٢٠٠٢): إن طريقة المكونات الأساسية تتميز بكونها تؤدي إلى تشبعات دقيقة، إذ يستخلص كل عامل بأقصى تباين ممكن؛ وتتلخص المصفوفة الارتباط في أقل عدد من العوامل المتعامدة، ويشير تيغزة (٢٠١٢) إلى أن هذه الطريقة تستعمل التباين الكلي بما في ذلك التباين الخاص وتباين الخطأ، ولذلك استخدمت لتحديد البنية العاملية.

وتم التحقق من أحادية البعد باعتماد خمسة محكات، المحك الأول: قيمة معامل الثبات للاتساق الداخلي، والمحك الثاني: الجذر الكامن (Eigen Value)، والمحك الثالث: نسبة التباين المفسر (Percent of Variance Extracted Criterion)، والمحك الرابع: اختبار رسم المنحدر لكاتيل (Kattell Scree Test)، والمحك الخامس: التشبعات العاملية (Loadings) وقيم الشيوع أو الاشتراكيات (Communalities)، وبيانها على النحو الآتي:

المحك الأول: أشار دينك وتفكول (٢٠١١) Dennick and Tavakol أن قيم معاملات الثبات التي تزيد عن (٠.٧٠) تُعد مقبولة لأغراض تحقق افتراض أحادية البعد؛ وأن القيم التي تقل عن ذلك تشير إلى ارتباطات ضعيفة بين مفردات الاختبار مما يدل على عدم تجانسها. حيث بلغ معامل ثبات درجات الاختبار ٠.٨٩. وبالتالي تحقق أحادية البعد للاختبار وفق هذا المحك.

المحك الثاني: الجذر الكامن (Eigen Value)، إذ حدد جلورفيلد (١٩٩٥) Glorfeld أن تكون نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن الثاني أكبر من أو تساوي (٢) للدلالة على أحادية البعد (Hatti, 1984; Lord,1980).

والمحك الثالث: نسبة التباين المفسر (Precent of Variance Extacted Criterion)

وقد كانت نتائج المحكين الثاني والثالث للاختبار على النحو الآتي:

جدول (١٨) الجذور الكامنة والتباين المفسر

العوامل	الجذور الكامنة	التباين المفسر%	التباين المفسر التراكمي%
١	٧.٠١٢	٢٨.٠٤٨	٢٨.٠٤٨
٢	٢.٢٦٤	٩.٠٥٥	٣٧.١٠٣
٣	١.٠٩٩	٤.٣٩٦	٤١.٤٩٩
٤	١.٠٦٧	٤.٢٦٩	٤٥.٧٦٨
٥	١.٠٠٨	٤.٠٣٣	٤٩.٨٠٢

يوضح الجدول (١٨) وجود (٥) عوامل زادت قيم جذورها الكامنة عن الواحد الصحيح مسؤولة عن تفسير التباين بين المفحوصين؛ إذ فسرت العوامل مجتمعة نسبة قدرها (٤٩.٠٨%)، فقد كانت قيمة الجذر الكامن للعامل الأول (٧.٠١)؛ وفسر ما نسبته (٢٨.٠٥%) من تباين الأداء بين المفحوصين على الاختبار، وهذه النسبة تشكل ما مقداره (٥٦.٣٣%) من التباين الكلي الذي تفسره العوامل الـ(٥) مجتمعة.

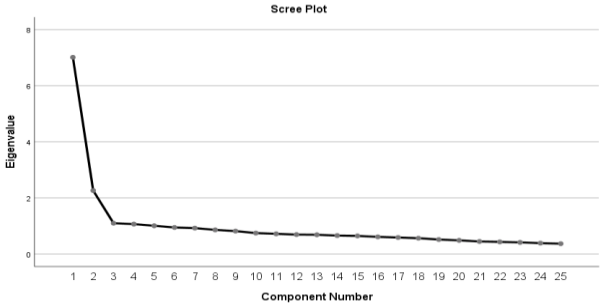
وكانت قيمة الجذر الكامن للعامل الثاني (٢.٢٦)؛ وفسر ما نسبته (٩.٠٦%) من تباين الأداء، كما يُلاحظ أن نسبة الجذر الكامن للعامل الأول إلى الجذر الكامن للعامل الثاني بلغت

(٣.١) ضعفاً، وهي نسبة تزيد عن الثلاثة أضعاف، كما يشير لورد (Lord,1980) إلى أنه إذا كانت نسبة الجذر الكامن للعامل الأول كبيرة مقارنة بالعامل الثاني، ولم يكن الجذر الثاني أكبر بقدر كبير من أي جذور أخرى؛ فإن هذا يُعد مؤشراً على تحقق سمة أحادية البعد في الأداء على هذا الاختبار.

والمحك الرابع: اختبار رسم المنحدر لكاتيل (Kattell Scree Test)، وهي طريقة تقوم على الجذور الكامنة وذلك برسم المحور السيني الأفقي الذي يدل على العوامل في مقابل الجذور الكامنة التي ترسم نقاطها على المحور الصادي العمودي، وينتج عن ذلك منحنى ينطلق من أعلى جذر عند العامل الأول ثم يأخذ في التناقص إلى أن يصل نقطة ما تقابل عاملاً معيناً تتباطأ درجة انحداره عندها كاسراً وتيرة الانخفاض (تيغزة، ٢٠١٢).

وكانت نتائج المحك الرابع اختبار رسم المنحدر للاختبار كما في الشكل التالي:

شكل (٦) اختبار رسم المنحدر



اختبار منحنى المنحدر (Scree Plot Test) لنموذج الاختبار يظهر عامل واحد فقط قبل نقطة القطع الأولى التي تمثل تباطؤ انحدار المنحنى ليتخذ منحنى مستوي نسبياً، بمعنى أن قيمة الجذر الكامن للعامل الأول ينفرد بقيمة مرتفعة نسبياً مقارنة ببقية العوامل الأخرى التي تبدو قيمها صغيرة نسبياً ومتقاربة، مما يشير إلى تحقق سمة أحادية البعد في هذا النموذج. ويشير ستيفنز (٢٠٠٩) Stevensإلى أن اختبار منحنى المنحدر (Scree Plot Test) دقيق عندما يكون حجم العينة أكبر من (٢٥٠) ومتوسط قيم الشيوع أكبر من أو يساوي (٠.٦٠).

والمحك الخامس: التشبعات العاملية (Loadings) وقيم الشيوع أو الاشتراكيات (Communalities)، وقد أشار علام (٢٠٠٣) أنه يفضل الاقتصار على التشبعات التي تمثل نسبة لا تقل عن (٢٥%)، كما ذكر تيغزة (٢٠١٢) أن الكثير من البحوث استعملت القاعدة المألوفة التي تعتبر أن التشبع (٠.٣٠) يدل على القيمة الفاصلة بين التشبع الهام وغير الهام، كما أن القيمة (٠.٥٠) هي الحد الأدنى لقيم الشيوع المقبولة؛ وتفضل بعض المراجع ألا يقل التشبع عن (٠.٤٠) (Stevens, 2002).

ويلاحظ أن جميع تشبعات الفقرات على العوامل أعلى من (٠.٤٠) كما يبين الجدول أن هنالك (١٩) فقرة تشبعت على العامل الأول وتمثل ما نسبته (٧٦%) من فقرات الاختبار، و فقرة تشبعت على العامل الثاني، وفقرة تشبعت على العامل الثالث، و(٢) فقرة تشبعت على العامل الرابع، و(١) فقرة تشبعت على العامل الخامس، وحيث أن هذه العوامل تشبع عليها عدد قليل من الفقرات فإنها لا تؤخذ بعين الاعتبار، مما يؤكد البعدية لقياس القدرة الرياضية المتمثلة في التحصيل الدراسي، أما قيم الاشتراكيات فقد تراوحت بين (٠.٦١) و(٠.٣٦) للفقرتين (١١) و(٦) على الترتيب. حيث أن جميع القيم أعلى من المحك (٠.٥٠) عدا الفقرات (١) و(٣) و(١٣) و(٢٢) التي بلغت قيم شيوعها (٠.٤٤)، (٠.٤٨)، (٠.٤٧)، (٠.٤٩) على الترتيب وهي جميعاً مرتفعة وقريبة من المحك المفضل.

وعليه فإن التشبعات العاملية (Loadings)، وقيم الشيوع (Communalities) الناتجة من التحليل العاملي الاستكشافي مناسبة وجيدة وتدل على أحادية البعد وفق هذا المحك.

الاستقلال الموضعي (Local Independence):

تم في هذه الدراسة الاعتماد على محك Q3 في الكشف عن تحقق افتراض الاستقلال الموضعي، وذلك من خلال مصفوفة الارتباط بين البواقي، وتم الاعتماد على القيم المطلقة لـــQ3 حيث بلغت أعلى قيمة (٠.٣٣) وكانت بين الفقرة (٥ و ٨) وعند النظر إلى هذه الفقرتين اتضح أن الفقرة الخامسة تناولت النسبة المئوية والفقرة الثامنة تناولت العمليات على الكسور مما يؤكد أن هذه الفقرتان مستقلتان. وأقل قيمة (٠) وكانت بين الفقرات (١٧ و٢٢)، وأغلب القيم تقع ضمن المحك الذي اقترحه ين (٠.٢) (إيالا، ٢٠١٧). وباستخدام برنامج الإكسل. والجدول التالي يبين أعلى القيم في هذه المصفوفة، حيث تم أخذ أعلى قيمة مطلقة من كل عمود من مصفوفة الارتباط بين البواقي وتم ترتيبها تنازلياً والجدول التالي يوضح ذلك .

جدول (١٩) القيم المطلقة لـــQ3

معامل الارتباط	معامل الارتباط
٠.٣٣	٠.١٤
٠.٢٤	٠.١٤
٠.٢٢	٠.١٣
٠.٢٢	٠.١٢
٠.٢١	٠.١١
٠.١٩	٠.١١
٠.١٩	٠.١٠
٠.١٨	٠.٠٩
٠.١٨	٠.٠٩
٠.١٦	٠.٠٨
٠.١٦	٠.٠٧
٠.١٥	٠.٠٦

أما من ناحية المحك التي اقترحه ين فإنه يوجد (٥) قيم تجاوزت هذا المحك وهي نسبة قليلة جداً من أصل ٣٠٠ معامل ارتباط حيث تمثل (٢%) بافتراض أنها نسبة خطأ. وعند تقريب هذه القيم نجد أنها جميعها ضمن حدود المحك المقترح من ين باستثناء القيمة (٠.٣) تعتبر قريبة من المحك.

المنحنى المميز للفقرة (Item Characteristic Curve):

تفترض النظرية وجود دالة مميزة خاصة بكل فقرة على حدة يتخذ كل منها شكل منحنى الترجيح اللوغاريتمي الاحتمالي. والذي يسمى بمنحنى خصائص الفقرة، وقد تم إيجاد المنحنى المميز لكل فقرة من فقرات الاختبار باستخدام برنامج اكسكليبر (Xcalibre).

التحرر من السرعة (Speediness):

تم التحقق من هذا الافتراض من خلال فحص نسبة الأفراد الذين أكملوا الإجابة عن الفقرات. وللتحقق من هذا الافتراض يتم حساب الزمن المحدد للاختبار الذي يمكن جميع الطلاب من الإجابة على جميع فقرا ت الاختبار. يمكن التأكد من أن الاختبار ليس اختبار سرعة من خلال نسبة الطلاب الذين أكملوا الاختبار. فوفقاً لما ذكره هامبلتون وسوامينثان وروجرز (١٩٩١) Hambleton & Swaminthan and Rogers أنه إذا كان (٧٥%) من الأفراد أكملوا الإجابة عن الاختبار، وكان (٨٠%) من مفردات الاختبار قد تمت الإجابة عنها فإن السرعة لا تعتبر في هذه الحالة عاملاً مهماً في الأداء على الاختبار، وبحكم أن الاختبار رابط الكتروني ولا يستطيع الطالب ارسال النموذج حتي يجيب عن جميع الأسئلة. مما يؤكد تحقق افتراض التحرر من السرعة. ممكن إضافة (أن ١٠١ من أصل ١٣٤ اجابوا عن الاختبار في أقل من ٣٠ دقيقة وهي تمثل النسبة المطلوبة كما أن (١٠٠%) من مفردات الاختبار تمت الإجابة عنها لإن الاختبار رابط الكتروني ولا يستطيع الطالب ارسال النموذج حتى يجيب عن جميع الأسئلة.

مطابقة البيانات للنموذج الثنائي المعلمة (2PL):

يوجد (٤) فقرات أظهرت مؤشراتها عدم مطابقته وفق برنامج اكسكليبر (Xcalibre) وهي (١٠، ١٤، ١٦، ٢٣) وقيم كاي تربيع الخاصة بها على الترتيب (٣٠.٤، ٢٦.٢٨ ، ٣١.٢٢ ، ٧٧.٢١) وجميعها دالة عند مستوى دلالة أقل من ٠.٠٥ وقيم Z Resid (٢.١، ٢.٣، ٢.٣، ٢.٥)، وبالنظر إلى مضمون هذه الفقرات اتضح أنها تتضمن المهارات التالية على التوالي (جمع الأعداد الكسرية، إيجاد النسبة المئوية للكسر الاعتيادي، إيجاد المعدل، جمع الكسور غير المتشابهة) وجميع هذه المهارات مهارات أساسية، كما أنه في دراسة (المركز القومي للامتحانات والتقويم التربوي، ٢٠١٣) تم الاعتماد على محكات الملاءمة الإحصائية (٢.٥ + ، ٢.٥-) لذا تم الإبقاء عليها.

كما تم فحص مطابقة الأفراد (person Fit) للنموذج الثنائي المعالم من خلال تقييم اللاتغاير وذلك بتقسيم العينة إلى عينتين فرعيتين مستقلتين وايجاد معامل ارتباط بيرسون بين تقديرات قدرات الأفراد (أيالا، ٢٠١٧).

حيث تم تقسيم العينة حسب المستوى التحصيلي وبلغ معامل الارتباط (٠.٩٧٤) وهو معامل ارتباط مرتفع ودال عند مستوى (٠.٠١)، وتدل على علاقة قوية موجبة شبه تامة، مما يؤكد مطابقة بيانات الأفراد للنموذج الثنائي المعلمة.

الإجابة عن السؤال الأول:

- هل تختلف الخصائص السيكومترية لدرجات الاختبار باختلاف الطريقة الوزنية (التقليدية، الدلتا سكور، ألفا سكور، التوقع البعدي)؟

تم إيجاد معاملات الصعوبة وفق النظرية التقليدية والنظرية الحديثة، حيث تشتمل النظرية التقليدية على الطرق الوزنية التالية (الطريقة التقليدية، طريقة الدلتا سكور، طريقة ألفا سكور)، والنظرية الحديثة تشتمل على طريقة التوقع البعدي.

والجدول التالي يحتوي على معاملات الصعوبة في كلا النظريتين مرتبة من السؤال السهل إلى السؤال الأصعب

جدول (٢٠): معاملات الصعوبة للفقرات وفق النظرية الكلاسيكية والحديثة

النظرية التقليدية						النظرية الحديثة النموذج الثنائي					
ترتيب السؤال	رقم السؤال	معامل الصعوبة	ترتيب السؤال	رقم السؤال	معامل الصعوبة	ترتيب السؤال	رقم السؤال	معامل الصعوبة	ترتيب السؤال	رقم السؤال	معامل الصعوبة
١	٣	٠.٨٩	١٤	٦	٠.٥٩	١	٣	١.٥٧-	١٤	٦	٠.٤-
٢	٥	٠.٨٩	١٥	٩	٠.٥٨	٢	٥	١.٥٥-	١٥	٩	٠.٣٧-
٣	٨	٠.٨٦	١٦	١٤	٠.٥٨	٣	٢	١.٥-	١٦	١٤	٠.٣٥-
٤	٢	٠.٨٥	١٧	١٠	٠.٥٨	٤	٨	١.٤٣-	١٧	١٠	٠.٣٤-
٥	٢٥	٠.٨٢	١٨	٢٢	٠.٥٧	٥	٢٥	١.٢٥-	١٨	٢٢	٠.٣٤-
٦	١٢	٠.٧٧	١٩	٢٠	٠.٥٥	٦	١٢	١.٢١-	١٩	٢٠	٠.٢٤-
٧	١٧	٠.٧٥	٢٠	١	٠.٥١	٧	١١	١.١٦-	٢٠	١	٠.١١-
٨	١٦	٠.٧٥	٢١	٢٣	٠.٤٩	٨	١٦	١.٠٨-	٢١	٢٣	٠.٠٢-
٩	١١	٠.٧٣	٢٢	٤	٠.٤٦	٩	١٧	٠.٩٤-	٢٢	٤	٠.١٩
١٠	٧	٠.٦٩	٢٣	٢١	٠.٤٦	١٠	٧	٠.٨٣-	٢٣	٢١	٠.٢٣
١١	١٥	٠.٦٥	٢٤	١٩	٠.٤	١١	١٥	٠.٦٦-	٢٤	١٩	٠.٥
١٢	٢٤	٠.٦٥	٢٥	١٣	٠.٣٢	١٢	٢٤	٠.٥٩-	٢٥	١٣	٠.٧١
١٣	١٨	٠.٦١				١٣	١٨	٠.٥٣-			

يلاحظ من الجدول السابق أن النظرية التقليدية والحديثة اتفقت على ترتيب ٢١ سؤال من أصل ٢٥ سؤال بنسبة ٨٤% كما يلاحظ أنه كلما زاد صعوبة السؤال كان هنالك اتفاق بين النظرية التقليدية والحديثة في ترتيب الأسئلة من حيث صعوبتها. فالعشر الأسئلة الأولى لم يكن هنالك اتفاق تام بين النظريتين في مواقع الأسئلة، بينما في ١٥ سؤال التالية هنالك اتفاق تام بين النظريتين.

وبلغ متوسط معامل الصعوبة في النظرية التقليدية (٠.٦٤) وبانحراف معياري (٠.١٦). وأكبر قيمة معامل صعوبة في النظرية التقليدية (٠.٨٩) للسؤال (٣) (٥) وأقل قيمة

(٠.٣٢) للسؤال (١٣) ونظراً لأن مفهوم معامل الصعوبة في النظرية التقليدية يمثل نسبة الطلاب الذين أجابوا إجابة صحيحة على السؤال فإن القيمة الأكبر تعني سؤال أسهل (سهل). بينما في النظرية الحديثة بلغ متوسط معامل الصعوبة (-٠.٥٩) وبانحراف معياري (٠.٦٥) وبلغت أعلى قيمة (٠.٧١) للسؤال (١٣) وأقل قيمة (-١.٥٧) للسؤال (٣).
كما بلغ معامل الارتباط بين معاملات الصعوبة في النظرية الحديثة والنظرية التقليدية (٠.٩٩٤) وهي قيمة مرتفعة جداً تقترب من (١).
أما بخصوص معاملات التمييز، فسيتم عرضها في الجدول التالي:

جدول (٢١): معاملات التمييز وفق النظرية الكلاسيكية والحديث

النظرية التقليدية				النظرية الحديثة النموذج الثنائي			
ترتيب السؤال	رقم السؤال	معامل التمييز Total Rpbis بوينت بايسيريال	ترتيب السؤال	رقم السؤال	معامل التمييز Total Rpbis بوينت بايسيريال	ترتيب السؤال	رقم السؤال	معامل التمييز
	١	٠.٥٦		١٤	٠.٥٦		١	٠.٩٩
	٢	٠.٣٩		١٥	٠.٤٧		٢	١.٠٤
	٣	٠.٤١		١٦	٠.٤٥		٣	١.٣٢
	٤	٠.٣٩		١٧	٠.٥٨		٤	٠.٥٩
	٥	٠.٤٢		١٨	٠.٤٢		٥	١.٤٠
	٦	٠.٤٧		١٩	٠.٣٣		٦	٠.٨
	٧	٠.٤٥		٢٠	٠.٦٤		٧	٠.٧٩
	٨	٠.٤٤		٢١	٠.٣٢		٨	١.٢١
	٩	٠.٥٦		٢٢	٠.٤٤		٩	٠.٩٨
	١٠	٠.٥٥		٢٣	٠.٥١		١٠	٠.٩٧
	١١	٠.٣٤		٢٤	٠.٥٧		١١	٠.٦٣
	١٢	٠.٤١		٢٥	٠.٤٧		١٢	٠.٨٢
	١٣	٠.٤٨					١٣	٠.٨٢

من الجدول السابق يلاحظ أن متوسط معامل التمييز (البوينت بايسيريال) في النظرية التقليدية بلغ (٠.٤٧) وبانحراف معياري (٠.٠٨). كما بلغت أعلى قيمة (٠.٦٤) للفقرة (٢٠)، وأقل قيمة (٠.٣٢) للفقرة (٢١). كما بلغ متوسط معامل التمييز في النظرية الحديثة (٠.٩٤) وبانحراف معياري (٠.٢٧). كما بلغت أعلى قيمة (١.٤) للفقرة رقم (٥)، وأقل قيمة (٠.٥٠) للفقرة رقم (٢١).
كما بلغ معامل ارتباط بيرسون بين معامل التمييز وفق الطريقة التقليدية ووفق الحديثة (٠.٥٥) وهو دال حيث بلغ مستوى الدلالة المرتبط به (٠.٠٠٤).

استخرجت معاملات الثبات (ألفا لكرونباخ، أوميجا، التجزئة النصفية بمعادلة سبيرمان – براون) لكل مجموعة من المجموعات الثلاث (التقليدية، الدلتا سكور، ألفا سكور)، بينما تم استخراج معامل الثبات التجريبي لطريقة التوقع البعدي، ثم اختبرت الفروق في معامل الثبات باستخدام معادلة فيلدت والجدول رقم (٥-٣) يبين معاملات الثبات لدرجات الاختبار المقدرة بالطرق الوزنية:

جدول (٢٢): معاملات الثبات وفق النظرية الكلاسيكية والحديثة

الطرق الوزنية / معاملات الثبات	التقليدية	معامل السهولة	معامل التمييز	التوقع البعدي	
ألفا لكرونباخ	٠.٨٨٩	٠.٨٦٣	٠.٨٧٨	معامل الثبات التجريبي	٠.٨٧٨
أوميجا	٠.٨٩١	٠.٨٦٩	٠.٨٨٧		
التجزئة النصفية	٠.٨٧٧	٠.٨٦٧	٠.٨٨١		

ومن الملاحظ وجود فروق ظاهرية في معامل ألفا وعند استخدام معادلة فيلدت لمعرفة هل الفروق جوهرية اتضح أن هناك فروقاً جوهرية بين الطريقة التقليدية والدلتا حيث بلغت قيمة ت المحسوبة بواسطة معادلة فيلدت ٢.٤٧ وهي أكبر من القيمة الجدولية عند ٠.٠٥ أما بخصوص الفروق بين الطريقة التقليدية وألفا سكور فهذه الفروق غير جوهرية وكذلك بين ألفا سكور والدلتا سكور.

وعند استخدام معاملات أوميجا وجدت أن هناك فروقاً جوهرية بين الطريقة التقليدية والدلتا وبين الدلتا والفا ولصالح التقليدية ولصالح ألفا بينما لايوجد فروق بين التقليدية وألفا. كما توجد فروق بين التقليدية والتوقع ولصالح التقليدية، كما لا توجد فروق بين التوقع والدلتا وألفا.

وعند استخدام التجزئة النصفية لا يوجد هنالك فروق جوهرية بين الطرق الثلاثة وبحكم أن لا يوجد اتفاق تام على جوهرية الفروق بين معاملات الثبات في الثلاث طرق المستخدمة. لذا سيتم الاعتماد على معامل أوميجا حيث يعتبر أنه أقل افتراضات من معامل ألفا.

تم حساب معاملات الصدق لدرجات الاختبار المقدرة بالطرق الأربع عن طريق معاملات ارتباط درجات الاختبار بدرجات الفصل الدراسي الثاني.

جدول (٢٤): معاملات الصدق

طرق تقدير الدرجات	معامل الصدق	مستوى الدلالة
الطريقة التقليدية	٠.٣٣٣	٠.٠٠
طريقة الدلتا سكور	٠.٣٣١	٠.٠٠
طريقة ألفا سكور	٠.٣٤٨	٠.٠٠
التوقع البعدي	٠.٣٤٨	٠.٠٠

توجد فروق ظاهرية بين معاملات الصدق ولمعرفة هل هذه الفروق جوهرية تم استخدام اختبار مربع كاي للعينات المترابطة للتحقق هل هذه الفروق جوهرية.

جدول (٢٥): اختبار مربع كاي

قيمة اختبار مربع كاي	مستوى الدلالة
١.٢٩٣	٠.٧٣١

وحيث أن مستوى الدلالة المرتبط بقيمة مربع كاي ببلغ (٠.٧٣١) وهو أكبر من (٠.٠٥) وهذا يشير إلى أنه لايوجد فروق أو أن هذه الفروق فروق ظاهرية فقط حيث أن جميع هذه الطرق كفاءتها متساوية في الارتباط بالمحك.

ولمعرفة هل توجد فروق بين متوسطات ارتباط الفقرة بالدرجة الكلية لكل طريقة من الطرق الوزنية المستخدمة في هذا البحث تم استخدام اختبار القياسات المتكررة، حيث تم التحقق من اعتيدالية التوزيع باختبار كولمجروف و سميرنوف، والجدول التالي يبن الإحصاءات الوصفية بالطرق الأربع .

جدول (٢٦): الإحصاءات الوصفية لارتباط الفقرة بالدرجة الكلية واختبار الاعتدالية

الطريقة	ارتباط الفقرة بالدرجة	الانحراف المعياري	قيمة اختبار كولمجرو ف سميرنوف	درجة الحرية	مستوى الدلالة	اختبار ماكولي		قرين - جيسر	
						قيمة اختبار ماكولي	مستوى الدلالة	قيمة ف	مستوى الدلالة
التقليدية	٠.٥٢	٠.٠٨	٠.١٤٣	٢٥	٠.٢				
الدلتا	٠.٥	٠.١١	٠.١٠٤	٢٥	٠.٢	٠.٠٣	٠.٠٠	٠.٤٢٣	٠.٦
ألفا	٠.٥١	٠.١	٠.١٧٣	٢٥	٠.٠٥٣				
التوقع	٠.٥١	٠.٠٩	٠.١٢٥	٢٥	٠.٢				

ظاهرياً توجد فروق ولمعرفة هل هذه الفروق داله ، يجب التحقق من شرط الكروية وذلك عن طريقة قيمة اختبار ماكولي وبالنظر إلى الجدول التالي نجد أن قيمة اختبار ماكولي بلغت ٠.٠٣ وهي قيمة دالة حيث بلغ مستوى الدلالة المقترن بها أقل من (٠.٠٥)، مما يشير إلى عدم تحقق الكروية (تجانس القياسات الأربع)، وبالتالي نختار اختبار قرين هاوس و جيسر.

يتضح من الجدول السابق باختبار تباين درجات القياسات الأربع لارتباط الفقرات بالدرجة الكلية أن قيمة (ف) بلغت (٠.٤٢) وهي قيمة غير دالة حيث أن قيمة مستوى الدلالة المقترن بها يساوي (٠.٦) وهو أكبر من (٠.٠٥) مما يشير إلى عدم وجود فروق جوهرية بين متوسطات ارتباط الفقرات بالدرجة الكلية للطرق الأربع.

مناقشة نتائج الدراسة

١- السؤال الأول:

أسفرت نتائج الإجابة عن السؤال الأول: هل تختلف الخصائص السيكومترية لدرجات الاختبار باختلاف الطريقة الوزنية (التقليدية، معامل السهولة، معامل التمييز، النموذج الثنائي (التوقع البعدي) ؟

عدم وجود فروق بين معاملات الصدق المرتبط بالمحك للطرق الأربع كذلك اتضح أنه لا يوجد فروق عندما تم استخدام اختبار القياسات المتكررة لمتوسطات ارتباط الفقرات بالدرجة الكلية لطرق الأربع وقد يرجع ذلك إلى أنه تم التحقق من صدق درجات الاختبار بعدة أدلة و مؤشرات (صدق المحتوى وصدق المحك وصدق المفهوم) مستخدماً النظرة الحديثة لمفهوم الصدق وقد اختلفت نتائج هذه الدراسة مع نتائج دراسة عبدالوهاب (٢٠٠١) ، ودراسة أبو فودة (٢٠١٦)، ودراسة الصيخان والمومني (٢٠٢١)، واتفقت نتائج هذه الدراسة مع نتائج دراسة روبتسيش (٢٠٢١).

وعندما تم مقارنة معاملات الثبات بين الطرق الأربع وتحديداً عندما تم استخدام معامل أوميجا اتضح أن هناك فروق بين معاملات الثبات ولصالح النظرية التقليدية ممثلة في الطريقة العادية وطريقة ألفا وهذه النتيجة تتفق مع دراسة الصيخان والمومني (٢٠٢١)، بينما أشارت دراسة اليجي واوناه وابانوبي (٢٠١٨)، ودراسة أبو فودة (٢٠١٦) إلى عكس ذلك بأن معامل الثبات في النظرية الحديثة أعلى من معامل الثبات في النظرية التقليدية.

يمكن تلخيص نتائج الدراسة في النقاط الاتية:

جدول (٢٧): ملخص نتائج الفروق بين الطرق الوزنية

وجه المقارنة	الطريقة التقليدية / طريقة الدلتا	الطريقة التقليدية / طريقة ألفا	الطريقة التقليدية/ التوقع البعدي	طريقة الدلتا/ طريقة ألفا	طريقة الدلتا/ طريقة التوقع	طريقة ألفا / طريقة التوقع
معامل الثبات	التقليدية أعلى	متساوية	التقليدية أعلى	ألفا أعلى	لا يوجد فرق	ألفا أعلى
معامل الصدق	لا يوجد فرق	لا يوجد فرق	لا يوجد فرق	-	-	-
متوسط ارتباط الفقرة بالدرجة الكلية	-	-	-	-	-	-

الخطأ المعياري وفق لورد :

الخطأ المعياري للقياس في النظرية التقليدية له قيمة واحدة لجميع المفحوصين بينما الخطأ المعياري في النظرية الحديثة لكل تقدير قدرة يوجد خطأ، لذا وجدت طرق في النظرية التقليدية لايجاد خطأ معياري لكل درجة، ومن هذه الطرق طريقة لورد لتحديد الخطأ المعياري لدرجات المفحوصين والتي تنص معادلته على التالي :

(Lord, 1959) $SE_{(lord)}= \sqrt{X * (n - X)/(n - 1)}$

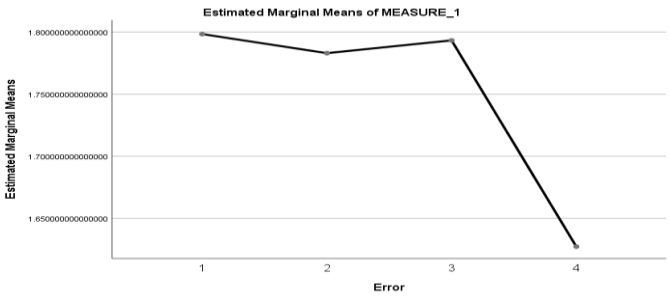
حيث أن : X= درجة المفحوص ، N= عدد فقرات الاختبار

والجدول التالي يمثل الإحصاءات الوصفية للخطأ المعياري للورد

جدول (٢٨): الإحصاءات الوصفية للخطأ المعياري للورد

االطريقة	متوسط الخطأ المعياري لطريقة لورد	الانحراف للخطأ المعياري لطريقة لورد
التقليدية (١)	١.٨٠	.٠٤٩
دلتا سكور (٢)	١.٧٨	.٠٥٤
ألفا سكور (٣)	١.٧٩	.٠٥٢
التوقع البعدي (٤)	١.٦٣	.٢٤٢

شكل (٥): متوسطات الأخطاء المعيارية للطرق الوزنية



من الجدول و الرسم البياني يتضح لنا أن قيم متوسطات الأخطاء المعيارية قريبة من بعض وأن متوسط الخطأ المعياري لطريقة لورد للطريقة التقليدية يساوي (١.٨) ولطريقة الدلتا سكور (١.٧٨) ولطريقة ألفا (١.٧٩) ولطريقة التوقع البعدي (١.٦٣). ولمعرفة هل هذه الفروق فروق جوهرية تم استخدام اختبار القياسات المتكررة.

جدول (٢٩): اختبار قيمة ماكولي واختبار قرين وجيسر

قيمة ماكولي	مستوى الدلالة	اختبار قرين - جيسر		
		متوسط المربعات	قيمة (ف)	مستوى الدلالة
٠.٠٠	٠.٠٠٠	٩.٤٤٩	٣٤٣.٣٩	٠.٠٠١

لجدول السابق يتضح أن قيمة ماكولي بلغت (٠٠.) ومستوى الدلالة المقترن بها أقل من (٠.٠٥) وهذا يشير إلى أن شرط الكرويه غير متحقق لذا يتم الحكم على جوهرية الفروق من خلال اختبار قرين هاوس جيسر حيث أن قيمة (ف) بلغت (٣٤٣.٣٩) ومستوى الدلالة المقترن بها أقل من (٠.٠٥) وهذا يشير إلى أن هذه الفروق جوهرية ولمعرفة اتجاه هذه الفروق تم إجراء المقارنات الثنائية. الجدول التالي يوضح تلك المقارنات:

جدول (٣٠): المقارنات الثنائية

الطرق الوزنية		الدلتا سكور	ألفا سكور	التوقع البعدي
الطريقة التقليدية	فرق المتوسطات	.٠١٥*	.٠٠٥*	.١٧١*
	مستوى الدلالة	.٠٠١	.٠٥٥	.٠٠١
طريقة الدلتا سكور	فرق المتوسطات		-.٠١٥*	.١٥٦*
	مستوى الدلالة		.٠٠١	.٠٠١
طريقة ألفا سكور	فرق المتوسطات			.١٦٦*
	مستوى الدلالة			.٠٠١

وحيث أن المقارنة بين متوسطات الإخطاء فإن قيمة الخطأ المنخفضة أفضل لذا فإن أفضل طريقة وفق هذا المحك هي طريقة التوقع البعدي.

كما أن الفروق بين متوسطات الأخطاء المعيارية للطرق الوزنية فروق جوهرية؛ وحيث أن الخطأ المعياري يحدد فترة الثقة للدرجات بمعنى آخر يحدد دقة تقدير الدرجات، فإن أفضل طريقة وفق محك الخطأ المعياري هي طريقة التوقع البعدي تليها طريقة الدلتا سكور وتليها طريقة ألفا سكور ثم الطريقة التقليدية.

توصيات الدراسة:

تأتي هذه الدراسة في محاولة لمعرفة أثر الطرق الوزنية على الخصائص السيكومترية، وتحديد الطريقة الأفضل من الطرق الوزنية (الطريقة التقليدية، طريقة الدلتا، طريقة ألفا، طريقة التوقع البعدي)

١- طريقة التوقع البعدي تعتبر أفضل طريقة وزنية في تقدير الدرجات، في ضوء محك الخطأ المعياري، وكذلك وفق محك معامل الصدق.

٢- الاهتمام ببناء الاختبار وفق خصائص الاختبار الجيد، واستثمار نتائج تحليل الخصائص السيكومترية للفقرات في تقدير درجات هذه الفقرات.

٣- عند يتم الاهتمام ببناء الاختبار وفق خصائص الاختبار الجيد، فإن ذلك قد يؤدي إلى تكافئ الطرق الوزنية في بعض مؤشرات الحكم على أفضلية طريقة على أخرى.

المراجع العربية:

أبو فودة، باسل خميس (٢٠١٦). التوافق بين النظرية التقليدية في القياس ونظرية استجابة الفقرة في مطابقة فقرات اختبار محكي المرجع في وحدة الهندسة التحليلية. رابطة التربويين العرب، ع(٧٣)، ص ص ١٨٩-٢١٥.

آنستازي، آني وأورينا، سوسن (٢٠١٥). القياس النفسي (ترجمة صلاح الدين علام). دار الفكر.

أيالا، أر (٢٠١٧). النظرية والتطبيق في نظرية الاستجابة للفقرة (ترجمة عبدالله الكيلاني وإسماعيل البرصان). الرياض: دار جامعة الملك سعود للنشر والتوزيع.

بيكر، فرانك (٢٠١٠). أسس نظرية الاستجابة للمفردة (ترجمة عبدالرحمن الطريري والسيد أبوهاشم وسوسن شلبي). الرياض: دار جامعة الملك سعود للنشر والتوزيع.

تيغزة، أمحمد بوزيان (٢٠١١). اختبار صحة البنية العاملية للمتغيرات الكامنة في البحوث: (منحنى التحليل والتحقق) مركز بحوث كلية التربية، جامعة الملك سعود.

تيغزة، أمحمد بوزيان (٢٠١٢). التحليل العاملي الاستكشافي والتوكيدي مفاهيمهما ومنهجيتهما بتوظيف حزمة SPSS وليزرل LISREL. دار المسيرة للنشر والتوزيع والطباعة.

جمحاوي، إيناس محمود (٢٠٠٠). مقارنة خصائص الفقرات وفق النظرية التقليدية ونظرية اتجابة الفقرة في مقياس للقدرة الرياضية (رسالة ماجستير غير منشورة). جامعة اليرموك.

حسن، السيد محمد أبوهاشم (٢٠٠٤). الدليل الإحصائي في تحليل البيانات باستخدام SPSS الرياض: مكتبة الرشد.

حسن، السيد محمد أبوهاشم (٢٠٢٠). معامل ألفا للتحقق من درجات أدوات القياس بين الحقائق والمعتقدات الخاطئة لدى الباحثين. مجلة البحث العلمي في التربية، ع(٢١)، ١٧٩-٢١٠.

الحكماني، رحاب سعيد (٢٠٠٧). مقارنة بين النظرية الكلاسيكية للاختبار ونظرية الاستجابة للمفردة في تقدير قدرات الأفراد ومدى استقرار مؤشرات المفردات الاختبارية (رسالة ماجستير غير منشورة). جامعة السلطان قابوس.

الخرشة، طه (٢٠١٦). أثر بعض طرائق تصحيح اختبارات الاختيار من متعدد في دقة تقدير صعوبة الفقرات وقدرات الأفراد وفق نموذج راش في نظرية الاستجابة للفقرة. العلوم التربوية، ٢٤ (١)، ٦٠٧- ٦٢٧.

الخضر، محمد وإسماعيل، البرصان (٢٠١٧). دقة القياس بين النظرية التقليدية في القياس والنظرية الحديثة في الاختبارات" دراسة تجريبية". المجلة الدولية للدراسات النفسية، ٩ (١)، ٧١-٨١.

الدوسري، راشد (٢٠١٢). تحديد درجة القطع ومستوى الأداء في اللغة العربية والرياضيات لطلبة الحلقة الأولى من التعليم الأساسي. رسالة الخليج لعربي، ع(١٢٤)، ص ص ١٠٥- ١٥١.

الدوسري، راشد (٢٠١٨). القياس والتقييم التطبيقي للمعلمين. الكويت: دار المسيلة.

الرحيل، راتب، والدرابسة، رياض (٢٠١٤). أثر طريقة تقدير القدرة وطريقة التعامل مع القيم المفقودة على دقة تقدير معالم الفقرات والأفراد. المجلة الدولية التربوية المتخصصة، ٣ (٦)، ٢٣- ٤٧.

الشريم، أحمد (٢٠٠٣). دراسة مقارنة لنموذج إنجوف ونموذج نيدلسكي في تحديد درجة القطع لاختبار محكي المرجع في الرياضيات (رسالة ماجستير غير منشورة). جامعة اليرموك، الأردن.

الشريم، أحمد (٢٠٠٨). تطوير أسلوب لتحديد القطع يرتكز على الأهداف ومقارنته مع أسلوب أنجوف (أطروحة دكتوراه غير منشورة). جامعة اليرموك، الأردن.

صهوان، أكرم (٢٠١٧). أثر استخدام بيانات التغذية الراجعة وخبرة المحكم في تقدير درجة لقطع وممارسات المحكمين وفق طريقة إنجوف المعدلة. المجلة المصرية للدراساتالنفسية، ٢٧ (٩٧)، ٦١- ١٤٧.

الصيخان، رحاب و المومني، رنا (٢٠٢١). مقارنة بين النظرية الكلاسيكية والنموذج ثلاثي المعلم في انتقاء فقرات اختبار تحصيلي للغة الإنجليزية. المجلة الدولية للدراسات التربوية والنفسية، ع (١)، ص ص ١٣٥-١٥٦.

طاهات، خالد (٢٠١٦). المقارنة بين طرق تقدير الخطأ المعياري الشرطي في القياس تبعاً لطريقة التصحيح لأثر التخمين وعدد البدائل (رسالة ماجستير غير منشورة). جامعة مؤتة، الأردن.

الطريري، عبدالرحمن (٢٠١٤). القياس النفسي والتربوي: نظريته-أسسه- تطبيقاته. الرياض: مكتبة الرشد.

عبابنة، عماد (٢٠٠٩). الإختبارات محكية المرجع: فلسفتها وأسس تطويرها. عمان: دار المسيرة.

عباس، محمد (١٩٩٣). المقارنة بين خمس طرق لتقدير الخطأ المعياري الشرطي في القياس عند مستويات محددة لعلامات الاختبار (أطروحة دكتوراه غير منشورة). الجامعة الأردنية، الأردن.

عبدالهادي، نبيل. (٢٠٠١). القياس والتقويم التربوي، واستخدامه في مجال التدريس. عمان: وائل للنشر.

عبدالوهاب، صلاح (٢٠٠١). أثر بعض الطرق الوزنية لتقدير الدرجات على صدق الاختبارات مرجعية المحك ذات الاختيار من متعدد. مجلة كلية التربية، جامعة بنها، ١٢(٤٩)،٢٠٢ – ٢٥٥.

العدل، عادل (١٩٨٦). أثر بعض طرق تقدير الدرجات للمفردات على ثبات الاختبار ذات الاختيار من متعدد (رسالة ماجستير غير منشورة). جامعة الزقازيق، مصر.
علام، صلاح الدين (٢٠٠٥). نماذج الاستجابة للمفردة الاختبارية أحادية البعد ومتعددة الأبعاد وتطبيقاتها في القياس النفسي والتربوي. القاهرة: دار الفكر العربي.
علام، صلاح الدين (٢٠٠٧). الاختبارات التشخيصية مرجعية المحك في المجالات التربوية والنفسية والتدريبية. القاهرة: دار الفكر العربي.
علام، صلاح الدين (٢٠١١). القياس والتقويم التربوي والنفسي- أساسياته وتطبيقاته وتوجهاته المعاصرة. دار الفكر العربي.
العنبكي، حيدر (٢٠٠٩). المفاضلة في تحديد درجات القطع لاختبار محكي المرجع (أطروحةدكتوراه غير منشورة). جامعة بغداد، العراق.
عودة، أحمد (٢٠٠٤). القياس والتقويم في العملية التدريسية. ط٣، دار الأمل.
الغانمي، وحيد (٢٠٠٩). المقارنة بين طرق مختلفة لتحديد علامة القطع لمعدل امتحان شهادة الدراسة الثانوية العامة للقبول في التخصصات الجامعية المختلفة (رسالة ماجستير غير منشورة). جامعة مؤته، الأردن.
غنيم، أحمد (١٩٨٦). بعض طرق تقدير الدرجات للمفردات ذات الاختيار من متعدد. مجلة كلية التربية بالمنصورة، ٥ (٧)، ٧٣- ١٠٧.
كروكر، ألـ ، والجينا، جي. (٢٠١٧). مدخل إلى نظرية القياس التقليدية والمعاصرة. (ترجمة هند الحموري وزينات دعنا). عمان: دار الفكر. (العمل الأصلي نشر في ١٩٨٦).
اللحياني، عفاف (١٤٣٠). أثر بعض طرق تقدير الدرجات للمفردات على ثبات وصدق درجات اختبار تحصيلي في الرياضيات ذي الاختيار من متعدد لدى طالبات الصف الأول الثانوي بمكة المكرمة(رسالة ماجستير غير منشورة). جامعة أم القرى، مكة المكرمة.
مجيد، سوسن (٢٠١٤). أسس بناء الاختبارات والمقاييس النفسية والتربوية. عمان: مركز ديبونو لتعليم التفكير.
المركز القومي للامتحانات والتقويم التربوي (٢٠١٣). تطوير بنك أسئلة لقياس الاستعداد للدراسة بالمرحلة الثانوية (المرحلة الثانية). القاهرة.
مصلح، موفق (٢٠٠٩). فاعلية أربع طرق في تحديد درجة القطع لاختبار محكي المرجع في العلوم لطلبة المرحلة الأساسية في دولة الامارات العربية المتحدة(رسالة ماجستير غير منشورة). جامعة عمان العربية، الأردن.
النعيمي، عز الدين (٢٠١٥). معالم الفقرات والأفراد وخاصية اللاتغير في الاختبارات الوطنية لضبط جودة التعليم في الأردن مقارنة بين النظرية الكلاسيكية والنظرية

الحديثة في القياس. مجلة اتحاد الجامعات العربية للتربية وعلم النفس، ١٣ (١)، ١٣٦ – ١٥٥.

نيتكو، أ. ، وبروخات، س. (٢٠١٢). التقييم التربوي للطلبة (ترجمة علي القرني الدوسري وراشد المحرزي، وحسين الخروصي). مكتب التربية العربي لدول الخليج.

اليامي، محمد (٢٠١٨). مقارنة طرق تصحيح اختبار الاختيار من متعدد من حيث أثرها على الخصائص السيكومترية. المجلة العربية للعلوم الاجتماعية، ٣ (١٣)، ١ - ٣٤.

المراجع الأجنبية:

Anastasi, A. (1982). Psychological tsting. (5th,Ed). New York: Macmillan

Baker, F., & Kim,S. (2004). *Item response theory: Parameter estimation techniques*. CRC press.

Buckendahl, C., Smith, R., Impara, J., & Plake, B. (2002). ACompariso of the Angoff and bookmark standard setting methods. Journal of Educational Measurement, 39 (3), 253- 263 .

Cetin, S., Gelbal, S. (2013). A Comparison of Bookmark and Angoff standard setting methods. Educational Sciences: Theory & Practice, 13 (4), 2169 – 2175.

Cizek,G. (2001). Setting performance standards concepts methods and perspectives. 15(2). 20-31.

Cronbach, L. (1970). Essentials of psychological testing. (3th, Ed). California: Stanford

Deledalle, C., Denis, L., & Tupin, F. (2009). Iterative weighted maximum likelihood denoising with probabilistic patch-based weights. IEEE Transactions on Image Processing , 18(12), 2661-2672.

Domingue, B., Dimitrov, D. (2021). A Comparison of IRT Theta estimates and Delta scores from the perspective of Additive conjoint measurement. Saudi Arabia National Center for Assessment.

Dimitrov, D. (2016). An Approach to Scoring and Equating Tests With Binary Items: Piloting With Large-scale Assessments. Educational and Psychological Measurement, 76 (6), 954- 975.

Eleje, L., Onah, F., & Abanobi,C. (2018). Comparative study of classical test theory and item response theory using diagnostic quantitative economics skill test item analysis results. European Journal of Education and Social Sciences, 3(1), 57-75.

Embretson, S., & Reise, S. (2013). Item response theory. Psychology Press.

Garre, G., Vermunt, K. (2006). Avoiding boundary estimation in latent class analysis by bayesian posterior estimation. Behaviormentrika, 33 (1), 43 - 59.

Hambleton, R. (1978). On the use of cutoff scores with criterion-referenced tests in instructional settings. Journal of Educational Measurement, 15 (4), 277-290 .

Hambleton, R., & Swaminathan, H. (1985). Item response Theory: Principles and applications. Boston: Kluwer Nijhoff Publishing.

Jaradat, D., Tollefson, M. (1988). The impact of altemative scoring procedures for multiple choice items on test reliability, validity, and grading. Educational and psychological measurement, 48 (3), 627- 635 .

Koehler, R. (1971). A Comparison of the Validities of Conventional Choice Testing and Various Confidence Marking Procedures. Journal of Educational Measurement, 8(4). 297-303.

Lord, F (1959). Tests of the same length do Have the same standard error of measurement. Educational and psychological measurem, 19(2). 233-239.

Lord, F. (1986). Maximum likelihood and bayesian parameter estimation in item response theory. Journal of Educational Measurement , 23 (2), 157- 162 .

Michael, J. (1968). The Reliability of A multiple-choice Examination Under Various test Taking Instruction, Journal of Educational Measurement, 5(4), 307-314.

Mislevy, R., Stocking, M. (1989). A consumer's guide to and bilog. *Applied psychological measurement*, 13(1). 57-75

Peterson, C., Schulz, E., & Engelhard, G. (2011). Reliability and Validity of Bookmark-Based methods for standard setting: Comparisons to Angoff-Based methods in the National Assessment of Educational Progress. Educational Measurement, 30 (2), 3- 14 .

Rippey, R. (1970). A comparison of five different scoring functions for confidence tests. Journal of educational measurement, 7(3), 165-170.

Robitzsch, A. (2021). About the equivalence of the latent D-Scoring model and the Tow-Parameter logistic item response model. MDPI.

Shulruf, B., Wilkinson, T., Weller, J., Jones, P., & Poole, P. (2016). Insights into the Angoff method: Results from simulation study. BMC Medical Education, 16 (134), 1- 10 .

Warm, T. (1978). A primer of Item Response Theory. Oklahomn: U.S Coast Guard Institute.

khairani, A., Razak, N. & Shamsuddin, H. (2014). Application of the Rasch model and the Bookmark method in setting cut scores in Mathematics. International journal of information and education technology, 4(2). 198-202.

Kim, J., & Yang, J. (2020). How to improve reliability of cut-off scores in dental competency exam: A comparison of rating methods in standard setting. Eur J Dent Educ, 24, 734-740.