الأفضلية بين طرائق حساب ثبات فقرات اختبار القدرة العقلية

زهراء احمد عبد الحسين السراي
Yassin.h@ircoedu.uobaghdad.edu.iq
أ.د ياسين حميد عيال الربيعي
Zahraa.abd2302@ircoedu.uobaghdad.edu.iq
جامعة بغداد/ كلية التربية ابن رشد للعلوم الانسانية

يهدف البحث الحالي الى (الهدف الأول: اعداد اختبار هنمون نلسون للقدرة العقلية لطلبة المرحلة الإعدادية وفقاً لنظرية القياس التقليدية) و(الهدف الثاني: تأثير طرائق حساب ثبات الفقرات في الخصائص السيكومترية لاختبار القدرة العقلية). ومن اجل تحقيق اهداف البحث قاما الباحثان بتبني اختبار(هنمون نلسون للقدرة العقلية) الذي تم اعداده من قبل (السوداني،٢٠١٠) الذي يتكون من(٦٥) فقرة، حيث كانت عينة التحليل الاحصائي (٤٠٠) طالب وطالبة في مديريات تربية محافظة بغداد ، ليتم بعد ذلك اجراء التحليل الاحصائي للفقرات من اجل حساب المؤشرات الإحصائية لها وهي كل من(التحليل العاملي ، مؤشر الصعوبة، مؤشر التمييز، ارتباط الفقرة بالدرجة الكلية، وفعالية البدائل الخاطئة) واتضح من خلال هذه الإجراءات تمتع جميع فقرات الاختبار بالصدق وفقاً للمعايير المعتمدة في كل مؤشر من المؤشرات ، وتم حساب ثبات الاختبار اذ بلغت قيمته (٠.٩١٨) وهو معامل ثبات جيد. فقد استخدما الباحثان منهج البحث الوصفي المقارن لإجراء المفاضلة في الخصائص السيكومترية للاختبار بين طرائق حساب ثبات الفقرة. إذ قاما الباحثان بحساب ثبات فقرات اختبار القدرة العقلية بأربع طرائق مختلفة وهي (طريقة الاحتمال المنوالي، معامل الارتباط الثنائي الأصيل، طريقة تعتمد على تمييز الفقرة وانحرافها المعياري، معامل الفا)، ثم قاما باستبعاد الفقرات التي كانت معامل ثباتها منخفض (حسب معيار كل طريقة من الطرائق الأربع) واعتمدت فقط الفقرات التي حققت ثبات مقبول.

ولغرض اجراء المفاضلة بين طرائق حساب ثبات فقرات اختبار القدرة العقلية في خاصية الصدق قاما الباحثان بحساب الصدق التمييزي وفقاً لكل طريقة من الطرائق الأربع، وباستخدام الاختبار التائي لعينتين مستقلتين، ثم قاما باستخدام تحليل التباين للتعرف على دلالة الفروق في الصدق التمييزي بين الطرائق الأربع وبعد اجراء التحليل أظهرت النتائج ان القيمة الفائية المحسوبة البالغة (١.٠٧٧) هي اصغر من القيمة الفائية الجدولية البالغة (٢,٦٠) عند مستوى

دلالة (٠,٠٥) وبدرجتي حرية (٣، ١٩٦) مما تبين لم تظهر فروق ذات دلالة إحصائية بين طرائق حساب ثبات الفقرات في التحقق من صدق اختبار القدرة العقلية.

ولغرض اجراء المفاضلة بين طرائق حساب ثبات فقرات اختبار القدرة العقلية في خاصية الثبات قاما الباحثان بحساب معامل ثبات الاختبار باستخدام معادلة كيودر- ريتشاردسون (٢٠) لكل طريقة من الطرائق الأربع فكانت النتائج ان هناك فرق بين قيم معاملات ثبات الاختبار في الطرائق الأربع ومن اجل معرفة دلالة الفروق بين الطرائق الأربع في معاملات الثبات قاما باستخدام الاختبار الزائي لدلالة الفروق بين معاملي الارتباط وتبينت النتائج كما يأتي:

أ. لا توجد فروق ذات دلالة إحصائية بين طرائق حساب الثبات (الاحتمال المنوالي، الارتباط الثنائي الأصيل، تمييز الفقرة وانحرافها المعياري) إذا كانت القيم الزائية المحسوبة أصغر من القيمة الزائية الجدولية البالغة (١,٩٦) أي بمعنى لا يوجد افضلية لإحدى الطرائق الثلاثة في حساب ثبات الاختبار.

ب. توجد فروق ذات دلالة إحصائية بين طرائق حساب الثبات (الاحتمال المنوالي، الارتباط الثنائي الأصيل، تمييز الفقرة وانحرافها المعياري) وطريقة (الفا) في حساب ثبات الاختبار إذا كانت القيم الزائية المحسوبة بين هذه (الطرائق الثلاث) وطريقة (الفا) أكبر من القيمة الزائية الجدولية البالغة (١,٩٦) ولصالح طرائق الثلاث مما يدل على أفضلية هذه الطرائق في حساب معاملات ثبات الاختبار على طريقة (الفا). غى صد ؟ مئ ذ ك ة شك ك ئح بك و ل ع ب ل ئ إلإز ئ ج ئ ة ك ش ئ ة ك ق ح ئ ة.

The Superiority among Methods of Calculating Item Reliability in the Mental Ability Test

Zahraa Ahmed Abdul-Hussein Al-Saray

Prof. Dr. Yaseen Hameed Aayal Al-Rubaie

University of Baghdad /College of Education (Ibn Rushd)

Abstract

The current study aims to (The first objective: preparing the Henmon Nelson mental ability test for preparatory school students according to traditional measurement theory) and (The second objective: The effect of methods of calculating item reliability on the psychometric properties of the mental ability test. In order to achieve the objectives of the study, the researcher adopted (Henmon Nelson Mental Ability Test) which was prepared by (Al-Sudani, 2010) and which consists of (65) items. the

statistical analysis sample, which amounted to (400) male and female students in the Baghdad Governorate Education Directorates. Then, after which the statistical analysis of the items was conducted in order to calculate their statistical indicators, which are (factor analysis, difficulty index, discrimination index, item correlation (Total score, and effectiveness of false alternatives). These procedures revealed that all test items were valid according to the standards adopted for each indicator. Test reliability was calculated using Cronbach's alpha coefficient, which reached a value of 0.918, a good reliability coefficient. Thus, the first objective of developing Henmon-Nelson Mental Ability Test for preparatory school students was achieved, according to traditional measurement theory.

To achieve the second objective of the research, the researcher used the descriptive comparative research method to compare the psychometric properties of the test between methods for calculating item reliability.

The researcher calculated the reliability of the mental ability test items using four different methods: the modal probability method, the original bivariate correlation coefficient, a method based on item discrimination and standard deviation, and the alpha coefficient. She then excluded items with low reliability coefficients (according to the criteria for each of the four methods), and only items that achieved acceptable reliability were considered.

Keywords: Item Reliability, Mental Ability.

لـ كبطك

يعد الثبات من أبرز الخصائص السيكومترية الضرورية لضمان جودة الاختبارات والمقاييس النفسية والتربوية إذ يعبر عن مدى استقرار النتائج واتساقها عبر تطبيقات متكررة، على الرغم من أن معظم الدراسات تركز على تقدير ثبات درجة الاختبار ككل إلا أن هناك حاجة ملحة لدراسة ثبات كل فقرة على حدة، لاسيما في ظل الاستخدام الواسع لهذه الفقرات كمؤشرات فردية (Item Single – Measures) في مجالات متعددة (Zijlmans al et.2018a:554).

حيث ان الفقرات الفردية لا تمثل بالضرورة سمات قابلة للقياس بدقة إذا لم تكن ذات مستوى عالٍ من الثبات، إذ إن انخفاض ثبات الفقرة يؤدي الى تقليل ثبات النتائج ويضعف من صدق التفسيرات المستخلصة، وهذه المشكلة تزداد تعقيداً عند استخدام الفقرات الفردية كمؤشرات مستقلة في البحث العلمي والتطبيقات الميدانية حيث لا يكون بالإمكان الاعتماد على المعاملات التقليدية مثل الفا لكرونباخ الذي يعد تقديراً عاماً لثبات الاختبار ولا يصلح لتقدير ثبات الفقرات الفردية بسبب مشكلات تتعلق بالهوية (الصياغة) الرياضية التي تجعل التقديرات غير مستقرة او غير قابلة للتفسير على مستوى الفقرة الواحدة (Zijlmans al et.2018a:553-554)، على الرغم من أن درجة الاختبار الكلية تتكون في الأصل من درجات الفقرات فإن استقرار هذه الفقرات يعد اساساً في ضمان ثبات الدرجة الكلية وقد يؤدي ضعف ثبات فقرة واحدة أو أكثر الى تقويض الثبات الكلي للاختبار (Zijlmans et al.,2018b:999).

وان تقدير ثبات الفقرة الواحدة يطرح إشكالية جوهرية اذ لا تصلح الطرائق الكلاسيكية لقياس الثبات الكلي بسبب محدودية الفقرة الواحدة وقد طورت عدة طرائق بديلة لتقدير ثبات الفقرات أحادية الفقرة غير ان الدراسات التجريبية لم تحسم بعد أي من هذه الطرائق هي الأدق في تقدير ثبات الفقرة كما ان النتائج المتباينة بين الطرائق تثير إشكالية علمية ومنهجية (Zhang&colvin,2024:1).

آه بطك

وتلعب نظرية القياس التقليدية (Classical Test Theory) دوراً حيوياً في تقييم ثبات الفقرة حيث تقوم على فكرة ان الدرجة التي يحصل عليها الفرد في الاختبار تتكون من جزئيين الدرجة الحقيقية والخطأ القياسي تسعى هذه النظرية الى تقليل الخطأ القياسي من خلال تحسين ثبات الفقرات من خلال تحليل مكونات الفقرة والدرجة الكلية يمكن تحديد الفقرات التي قد تسهم في زيادة او تقليل الخطأ، مما يتيح للمطورين تحسين جودة الفقرات وبالتالي جودة الاختبار ككل (Lord & Novick,1968:327).

وتشكل الخصائص السايكومترية لفقرات الاختبارات النفسية أهمية كبيرة في تحديد هذه المؤشرات، وهي لا تقل أهميتها عن أهمية الخصائص السايكومترية للاختبار نفسه، لان الخصائص السايكومترية للاختبار تعتمد الى حد كبير على خصائص فقراته (Smith,1966:69-70).

ويرى المتخصصون في القياس النفسي ضرورة حساب ثبات الفقرات لان معامل ثبات الاختبار يعتمد كثيراً على معاملات ثبات فقراته، اذ ان الحصول على فقرات ذات ثبات عالي سوف يهيئ الفرصة لإعداد اختبار ثابت ودقيق، أي ان ثبات الاختبار يعتمد على ثبات فقراته،

مثلما يعتمد صدق الاختبار على صدق فقراته. لان يشير الى تجانس الفقرات مع بعضها في قياس ما يفترض ان تقيسه (عبد الرحمن،١٩٨٣: ٣٦٤).

ويعد حساب ثبات الفقرات خطوة أساسية في بناء الاختبارات والمقاييس النفسية والتربوية اذ يسمح للباحثين بتقدير مدى استقرار استجابات الافراد على كل فقرة بشكل مستقل، فاذا كانت الفقرة غير ثابتة فأنها قد تضيف قدراً كبيراً من الخطأ العشوائي مما يقلل من دقة الدرجة الكلية للاختبار ومن هنا يتيح فحص ثبات الفقرة إمكانية تحديد الفقرات الضعيفة او غير المناسبة التي تؤثر سلباً في صدق الاختبار وثباته الكلي كما ان الاهتمام بثبات الفقرات يساهم في تحسين عملية انتقاء الفقرات وبالتالي تعزيز قدرة الاختبار على قياس السمة او القدرة المستهدفة بدرجة اكبر من الموثوقية، وفي هذا السياق يؤكد الباحثون ان ثبات الفقرة قد يكون مهماً عند بناء الاختبار لان الفقرة ذات الثبات المنخفض قد لا تسهم كثيراً في ثبات الدرجة الكلية وبالتالي يمكن ان تكون مرشحة للحذف من الاختبار (Zijlmans et al.,2018b:999-1001).

وتبرز طرائق حساب ثبات الفقرة كأدوات تحليلية فعالة لتحسين جودة أدوات القياس وتساعد في اتخاذ قرارات دقيقة بشأن الاحتفاظ او حذف الفقرات وتوفر رؤية أعمق لتقييم الفقرات الفردية بدلاً من الاعتماد الحصري على الدرجة الكلية (Zijlmans et al. ,2018b:999).

ئ هئـــطك Research objectives:

ي ه ـــطك ط ك ىطك ع ـــطك و:

١- اعداد اختبار هنمون نلسون للقدرة العقلية لطلبة المرحلة الإعدادية وفقاً لنظرية القياس التقليدية.

٢- تأثير طرائق حساب ثبات الفقرات في الخصائص السيكومترية لاختبار القدرة العقلية.

ح ططك : يتحدد البحث الحالي بـــ:

١. طلبة المرحلة الإعدادية التابعين لمديريات تربية محافظة بغداد للصفوف الدراسية (الرابع – الخامس – السادس) للدراسة الصباحية في المدارس الحكومية ولكلا الجنسين (ذكور- اناث) ولكلا التخصصين (العلمي والادبي)، للعام الدراسي (٢٠٢٤-٢٠٢٥).

٢. اختبار هنمون نلسون للقدرة العقلية لطلبة المرحلة الإعدادية.

٣. الطرق المستخدمة في حساب ثبات الفقرة من خلال:

- طريقة الاحتمال المنوالي.
- معامل الارتباط الثنائي الأصيل.
- تمييزها وانحرافها المعياري.
- معامل الفا.

ة ي ك ك ذ ة:

٠. ذ ق ئ ة (Item Reliability)

- كروكر والجينا **Crocker & Algina (2008)**

" تعد من أهم العوامل التي تؤثر في جودة الاختبارات النفسية، ويقصد بها مدى اتساق النتائج التي توفرها الفقرة عند تكرار استخدامها في ظروف متشابهة، ويعتبر الثبات عاملاً أساسياً لضمان أن تكون الفقرة مستقرة وتعكس القياس الصحيح للخاصية النفسية المراد قياسها. إذ أن الفقرات الثابتة تعزز من دقة التقييم وتقلل من احتمالية التحيز أو الخطأ في نتائج الاختبار"(Crocker & Algina,2008:423).

ك ع ك : تبنى الباحثان تعريف **(Crocker & Algina,2008)** تعريفاً نظرياً لثبات الفقرة.

ك ع ئ ى: قياس التباين الحقيقي لفقرات اختبار القدرة العقلية والذي يمثل مدى ثبات استجابة افرد عينة البحث من خلال استعمال طرائق حساب ثبات الفقرة المعتمدة في الدراسة الحالية.

٢. القدرة العقلية

- ز لـ م **Spearman (1904)** بانها: قدرة عامة تتضمن بشكل أساس استنباط العلاقات و المتعلقات (Gregory ,1996:153).

ك ع ك : لما كان البحث الحالي سيعتمد نظرية سبيرمان اطارا نظريا فقد اعتمد تعريفه للقدرة العقلية تعريفاً نظرياً.

ك ع ئ ى: الدرجة التي يحصل عليها المستجيب من خلال اجابته على فقرات اختبار القدرة العقلية المتبناة في الدراسة الحالية.

ئ ل ذ ك : ذ ق ب **(Item Reliability)**:

يعد ثبات الفقرات من المؤشرات الأساسية في تقدير ثبات الاختبار، إذ إن معاملات صدق الفقرات تعد مؤشراً لصدق الاختبار ككل، وبالتالي فإن توفر فقرات ذات ثبات مرتفع سيهيئ الاختبار على ان يكون ذات ثبات عال (عبد الرحمن ،٢٠٠٨: ٢٣٠). ويعد هولزنجر(Holzinger.1932) اول من قدم ابحاثاً بهذا المفهوم الجديد للفقرات الذي حاول ان يحسب هذا الثبات بطريقته التي سماها دالة الفروق (السيد، ١٩٧٩: ٥٤٤).

ويعد مؤشر ثبات الفقرة مقياساً لمدى اسهام الفقرة في التباين الكلي للاختبار الذي بدوره يؤدي الى ثباته، إذ تستند هذه العملية الى تحليل دقيق للفقرات، بهدف تحديد الفقرات التي تسهم فعلياً في زيادة التباين الكلي الذي بدوره يؤدي الى زيادة ثبات الاختبار وعلى هذا الأساس يتم

استبعاد الفقرة التي لا تساهم في زيادة الثبات الكلي والابقاء على الفقرة التي تزيد من نسبة التباين الكلي التي تؤدي الى ثبات مرتفع (Magnusson.1967:215).

إذ إن الأسلوب الاحصائي لثبات الاختبار ككل هو غير مناسب لحساب ثبات فقرات الاختبار لان الخصائص الإحصائية لدرجات الفقرات تختلف اختلافاً واضحاً عن الخصائص الإحصائية لدرجات الاختبار ككل لان درجات الاختبار مستمرة او متصلة بينما درجات الفقرات تكون ثنائية او متقطعة (الطيب، ١٩٩٩: ٢٦٢).

ئ ح ئ ا ت ئ ظ ق ب:

١. ق ب ئ إ ح ك ط ك ط ى **Model Probability** :

تستخدم هذه الطريقة بحساب ثبات الفقرات الاختبارات الموضوعية التي تتطلب من بين عدد من بدائل الإجابة اختيار بديل واحد صحيح، كما في اختبارات (الصواب والخطأ) او في اختبارات الاختيار من متعدد (الطيب،١٩٩٩: ٢٦٣).

ل ك ئ ا ت ئ ظ ق ئ ة ق ب ئ إ ح ك ط ك ط ى م ظ ك ئ ى ئ إ ة ب:

١. حساب تكرار الإجابة على كل بديل او احتمال لكل الفقرات.

٢. حساب التكرار النسبي لكل بديل في كل فقرة ويتم حسابه من خلال $\frac{\text{تكرار الاجابة}}{\text{المجموع}}$.

٣. نعيين اعلى تكرار نسبي من بين البدائل الفقرة الواحدة.

٤. ولحساب ثبات الفقرة نطبق المعادلة الاتية.

$$R_i = \frac{n}{n-1}\left(L - \frac{1}{n}\right)$$

حيث ان:

n = عدد البدائل. L = اعلى تكرار نسبي من بين البدائل (عبد الرحمن، ٢٠٠٨: ٢٣٠-٢٣١).

٢. ل ع ئ ل ئ إ ة ئ ك ئ ى ئ آ ش **Point – Biserial Correlation Coefficient**:

ولتقدير معامل ثبات الفقرة في هذا الأسلوب نتبع نفس الخطوات المستخدمة في تقدير صدق الفقرة وذلك بالاعتماد على الدرجة الكلية للمقاييس أو للاختبار او على الدرجة الكلية للمقاييس الفرعية بدلاً من استخدام الدرجة الكلية على مقياس المحك الخارجي، وفي هذه الحالة تستخدم الدرجة الكلية على المقياس (او الدرجة الكلية على أحد المقاييس الفرعية للاختبار) كمحك لتقسيم العينة الى عدد من المجموعات الفرعية، وبالتالي تحديد مجموعات المقارنة الطرفية ثم يتم حساب معامل الارتباط الثنائي الأصيل بين درجات الافراد على كل فقرة على حدة ودرجاتهم على الاختبار نفسه بصورة كلية. وعند الحصول على معامل ارتباط دال احصائياً بين درجات الافراد على الفقرات والدرجة الكلية للاختبار يعد ذلك مؤشر لثبات الفقرات وكما يمثل دليلاً على الاتساق الداخلي لفقرات الاختبار ومن ثم وجب على معد الاختبار الاحتفاظ

بتلك الفقرات لأنه يترتب على استخدامها الحصول على اختبار يتصف بدرجة عالية من الاتساق الداخلي فيه تميز كل فقرة بين مجموعات المقارنة في نفس الاتجاه الذي تمييز فيه الدرجة الكلية بين هذه المجموعات (خطاب ،٢٠٠٣: ٣٤٩).

لك ئ ا ت ئ ةطفقوب اهنطك ق بمـ ظطك ئ ةئإلإة بـ:

٠. نحسب الانحراف المعياري لدرجات الاختبار ككل.

١. نعين متوسط درجات الافراد ككل للإجابات الصحيحة على الفقرة.

٣. نعين متوسط درجات الافراد ككل لإجابات الخاطئة على الفقرة

٤. نعين معامل السهولة للفقرة وصعوبتها.

٥. نطبق المعادلة التالية لحساب ثبات الفقرة.

$$R_i = \frac{\bar{x}1 - \bar{x}2}{s} \times \sqrt{p \times q}$$

حيث ان:

$\bar{x}1$= متوسط الإجابات الصحيحة. $\bar{x}2$ = متوسط الإجابات الخاطئة.

S= الانحراف المعياري. **p**= نسبة الإجابات الصحيحة.

q = نسبة الإجابات الخاطئة (خطاب،٢٠٠٣: ٣٥٠).

٣. ح ئ ا ت ئ ةطفقوب لـ خلإك ة ةئ ةهـ ئغهـئطك ئذ (**Standarddivision& Discrimination)**

يحدد معامل ثبات كل فقرة بمعلومية تمييزها وانحرافها المعياري (Thayn,2011:30)،

لك ئ ا ت ئ ةطفقوب اهنطك ق بمـ ظطك ئ ةطك بـ:

١. حساب معامل التمييز للفقرة.

٢. حساب معامل الصعوبة.

٣. حساب معامل السهولة من خلال (١_ معامل الصعوبة)

٤. حساب الانحراف المعياري وذلك من خلال الجذر التربيعي لضرب معامل الصعوبة بمعامل السهولة.

٥. نطبق المعادلة الاتية لحساب ثبات الفقرة.

$$ir = r_{it}\sqrt{pq}$$

حيث ان:

ir = ثبات الفقرة r_{it}= تمييز الفقرة

q= معامل السهولة (p−1). p = معامل الصعوبة (صعوبة الفقرة)

٤. لعئل طفئ Coefficient alpha:

يعرف معامل ثبات الفقرة (ir) بأنه حاصل ضرب الانحراف المعياري لدرجات الممتحنين على الفقرة بمعامل ارتباط الدرجة على الفقرة بالدرجة الكلية على المحك الداخلي (x) وهو علامات الطلبة الكلية على الاختبار وهذا ما أشار اليه كروكر والجينا (Crocker&Algina,1986).

وبالرموز فإن:

$ir = s_i r_{ix} = \sqrt{p\ i\ qi}\ r_{ix}$

حيث ان:

s_i : الانحراف المعياري للدرجات على الفقرة.

r_{ix} : معامل الارتباط بين الدرجات على تلك الفقرة من ناحية ودرجات المختبرين الكلية على الاختبار. ومعامل الارتباط هو من نوع بوينت بايسيريال (النبهان،٢٠١٣: ٢٥٦).

و أعتمد معامل الفا في تقدير معامل ثبات وفقاً لخصائص العينة التي تم توليدها الفقرات أحادية البعد (ثنائية الاستجابة) حيث يساوي هذا المعامل حاصل ضرب الانحراف المعياري لدرجات الفقرات وارتباط هذه الفقرات بالدرجة الكلية، وبناء على ذلك فان معامل الفا هو دالة لمعاملات ثبات الفقرات بحيث يرتفع معامل الفا كلما كانت معاملات الفقرات مرتفعة (الوسيدي،٢٠١٧: ٣٩).

م بلآعئك ك لئم (Two Factor Theory)

تعد نظرية العاملين او العامل العام من النظريات الرائدة في تفسير القدرة العقلية للعالم الانجليزي سبيرمان (Spearman.1904) , حيث تعد تاريخياً من اول النظريات المتكاملة لتفسير الذكاء والتي بنيت على نتائج التحليل الاحصائي العاملي للاختبارات (Sternberg, 1990: 983). فقد اعتمدت نظرية سبيرمان بتفسيرها للعامل العام الذي يقفى خلف مختلف النشاطات العقلية على أساس أسلوب التحليل العاملي , وتمحورت بشكل أساسي على ان مختلف النشاطات العقلية تشترك في عامل واحد يدعى بالعامل العام , وافترضت ايضاً وجود عدة عوامل طائفية أخرى اطلق عليها اسم العوامل الخاصة (Specific Factors) يتخصص كل منها بنشاط عقلي وحيد , وان أي ارتباط موجب بين أي مهمتين معينتين يرجع الى العامل العام , (Anastasi, 1997: 380).

خ ئزئ ة زئق ب:

خ ئز ب ز م ئمع ه ك ى لف (Sijun Zhang& Kimberly colvin,2024)

ط ع ئ م: مقارنة بين الطرائق المختلفة في تقدير ثبات الفقرة: دراسة محاكاة

Comparison of different reliability estimation methods for single-item assessment: a simulation study

ه ـطيخ يخز ب: مقارنة خمس طرائق لتقدير ثبات الفقرة عبر دراسة محاكاة لتحديد أي من هذه الطرائق تقدم تقديرات أكثر دقة وثبات تحت ظروف محاكاة مختلفة من حيث حجم العينة، نوع المقياس، وطول الاختبار.

طيخع ب: لم تعتمد على عينة بشرية فعلية، بل استخدم بيانات محاكاة حاسوبية

ئلآخيئ: هي بيانات محاكاة حاسوبية لمقاييس أحادية الفقرة ومتعددة الفقرات، باستخدام فقرات متعددة الفئات على نمط مقاييس ليكرت، دون استخدام أداة واقعية مطبقة على أفراد فعليين.

لـ ه ثطيخ يخز ب: منهج تجريبي باستخدام المحاكاة الحاسوبية لتوليد بيانات افتراضية وتحليلها باستخدام الطرق الخمس لتقدير ثبات الفقرة تحت ظروف إحصائية وتجريبية مختلفة.

طيخ زئ ئلأح ئ ب: تم تنفيذ جميع التحليلات باستخدام لغة البرمجة R (الإصدار ٤.١.٠).

طيخ ئ ث: أظهرت نتائج اختلافاً واضحاً في أداء الطرائق الخمس لتقدير ثبات التقييمات أحادية الفقرة تبعاً لظروف المحاكاة المختلفة. (Zhang& colvin,2024:1-16).

لـ ه بطيخ Research Methodology

اعتمد في البحث الحالي المنهج الوصفي المقارن للأسباب (الدراسات المقارنة السببية) لملاءمته لطبيعة وأهداف البحث الحالي، فالمنهج الوصفي يعد من أكثر مناهج البحث استخداما خصوصا في بحوث العلوم التربوية والنفسية، فهو يهتم بوصف الظاهرة المدروسة بشكل دقيق ويعمل على إيجاد العلاقات بينها وبين ظواهر اخرى وكذلك المقارنة بينهما.

لـ ظطيخ Society of the Research

يتألف مجتمع البحث الحالي من طلبة المرحلة الإعدادية، المنتظمين في المدارس الحكومية التابعة لمديريات التربية العامة في محافظة بغداد (رصافة – كرخ) للعام الدراسي (٢٠٢٤-٢٠٢٥) للدراسة الصباحية ذكوراً واناث، حيث بلغ المجتمع الإحصائي الكلي (٣٨٧٤٢٠) طالباً وطالبة، موزعين بحسب المديرية والتخصص والجنس والصف.

ع ئ طيخ Samples of the Research

يتطلب تحقيق أهداف البحث الحالي اختيار مجموعة من العينات منها (عينة وضوح التعليمات والفقرات) ، عينة التحليل الاحصائي ، عينة النتائج أو التطبيق النهائي)، سيتم توضيح طرق اختيار كل منها وحجم كل منها، كل بحسب الاجراء في حينه.

آخيئطيخ Instrument of the Research:

بعد إطلاع الباحثة على مجموعة من الاختبارات التي تقيس الجوانب العقلية والأدبيات ووجهات نظر علماء القياس العقلي والنفسي، وجدت الباحثة ان اختبار القدرة العقلية لـ (هنمون – نلسون Henmon – Nelson) مناسب للبحث الحالي لا سيما انه يلاءم عينة البحث، فقامت الباحثة بتبني النسخة النهائية من اختبار (هنمون – نلسون Henmon – Nelson) للقدرة العقلية الذي

قامت (السوداني ٢٠١٠) بأعداده للبيئة العراقية وفقاً لإنموذج راش، ويتكون الاختبار من (٧٠) فقرة اختبارية من توع الاختيار من متعدد وبواقع خمس بدائل لكل فقرة، ويكون بذلك تصحيحه بمنح درجة واحدة للإجابة الصحيحة و (صفر) للإجابة الخاطئة .

ك ك ق ك إ ذ

استند الباحث كخطوة أولية في التحقق من صدق الفقرات وصلاحيتها وذلك من خلال عرض الاختبار بصورته الأولية على مجموعة من الخبراء المتخصصين في العلوم التربوية والنفسية والذي بلغ عددهم (١٥) خبيراً وقد اعدت جميع فقرات اختبار التفكير التقني مقبولة بنسبة (١٠٠%).

ع ك ئ إ

وتشير انستازي "Anastasia" ان حجم عينة التحليل الإحصائي يجب ألا يقل عن (٤٠٠) فرد يتم اختيارهم بدقة من مجتمع الأصلي للبحث (محمود ،٢٠١٥: ٧٢٥). وعلى هذا الأساس تكونت عينة التحليل الاحصائي من (٤٠٠) طالباً وطالبة من طلبة المرحلة الإعدادية (لمديريات الست لتربية محافظة بغداد) اذ يتم اختيارها بإسلوب العينة العشوائية الطبقية المتناسبة.

ك ك ئ إ ف ئ إ ذ:

ان تحليل الفقرات احصائياً يعد جزء لا يتجزأ من البحث بحد ذاته ويترتب عليه التوصل الى نتائج صادقة ويمكن الوثوق بها.(علام , ٢٠٠٩: ٣١). وبإستخراج مجموعة من المؤشرات الاحصائية التي تدلل على الخصائص السيكومترية الجيدة للاختبار، وهي كل من (التحليل العاملي، معاملات صعوبة الفقرة، معاملات تمييز الفقرة، معامل ارتباط الفقرة بالدرجة الكلية للاختبار، فاعلية البدائل الخاطئ) واتضح من خلال هذه الإجراءات تمتع جميع فقرات الاختبار بالصدق وفقاً للمعايير المعتمدة في كل مؤشر من المؤشرات، وتم حساب ثبات الاختبار باستخدام معامل الفا كرونباخ اذ بلغت قيمته (٠.٩١٨) وهو معامل ثبات جيد.

ئ ح ئ ا ت ئ ف ئ ة:

قاما الباحثان بحساب ثبات فقرات اختبار القدرة العقلية لهنمون نلسون بأربع طرائق مختلفة وهي (طريقة الاحتمال المنوالي، طريقة الارتباط الثنائي الأصيل، طريقة تعتمد على تمييز الفقرة وانحرافها المعياري، طريقة الفا).

ئ ل: ق ب ئ إ ك ك ى: تستخدم هذه الطريقة لحساب ثبات الفقرات الاختبارية الموضوعية التي تتطلب من بين عدد من بدائل الإجابة اختيار بديل واحد صحيح، والمعيار المعتمد لقبول معامل ثبات الفقرة في هذه الطريقة هو ان تكون قيمة معامل الثبات أكبر او يساوي (٠.٥٠) بهذه الحالة تعد الفقرة مقبولة وأنها تتمتع بمعامل ثبات مقبول، اما إذا كانت

قيمة معامل الثبات أقل من (٠.٥٠) بهذه الحالة تعد الفقرة غير مقبولة وأنها لا تتمتع بمعامل ثبات مقبول (معمريه،٢٠٠٢: ١٥٠).

تئميؤ: لعئل ىئلإلة ئ ظك ئئىىئلآ ش :

نستعين بهذه الطريقة لحساب الاقتران القائم بين درجات الاختبار ودرجات اي فقرة من فقراته وأن المعيار المعتمد لقبول معامل ثبات الفقرة في هذه طريقة هو أن تكون قيمة معامل الارتباط دالة احصائيا، بهذه الحالة تعد الفقرة مقبولة وأنها تتمتع بمعامل ثبات مقبول، أما أذا كانت قيمة معامل الارتباط غير دالة احصائيا، بهذه الحالة تعد الفقرة غير مقبولة وأنها لا تتمتع بمعامل ثبات مقبول (خطاب، ٢٠٠٣: ٣٦١).

ئكيئ: ح ئ ا ت ئ ظظقوب ل خلإك ة ظظقوب ىظ ىئغىئك ئذ :

يحدد معامل ثبات كل فقرة بمعلومية تمييزها وانحرافها المعياري. وأن المعيار المعتمد لقبول معامل ثبات الفقرة في هذه طريقة هو أن تكون قيمة معامل الثبات أكبر أو يساوي (٠.١٠) بهذه الحالة تعد الفقرة مقبولة وأنها تتمتع بمعامل ثبات مقبول، أما أذا كانت قيمة معامل الثبات أقل من (٠.١٠) بهذه الحالة تعد الفقرة غير مقبولة وأنها لا تتمتع بمعامل ثبات مقبول (Thayn, 2011: 30).

ئ ىيئ: لعئل ظظقئ:

يستخدم معامل الفا في تقدير معامل ثبات وفقاً لخصائص العينة التي تم توليدها الفقرات أحادية البعد (ثنائية الاستجابة) حيث يساوي هذا المعامل حاصل ضرب الانحراف المعياري لدرجات الفقرات وارتباط هذه الفقرات بالدرجة الكلية، وأن المعيار المعتمد لقبول معامل ثبات الفقرة في هذه طريقة هو أن تكون قيمة معامل الثبات أكبر أو يساوي (٠.٣٠) بهذه الحالة تعد الفقرة مقبولة وأنها تتمتع بمعامل ثبات مقبول، أما أذا كانت قيمة معامل الثبات أقل من (٠.٣٠) بهذه الحالة تعد الفقرة غير مقبولة وأنها لا تتمتع بمعامل ثبات مقبول (Zijlmans et al., 2018 :1007).

واستخلاصاً لما تقدم لقد تم حساب ثبات الفقرة لاختبار القدرة العقلية لهنمون نلسون بالطرائق الأربع: الطريقة الأولى هي طريقة الاحتمال المنوالي وأظهرت النتائج ان نسبة الفقرات التي كانت ثباتها أعلى من المعيار المحدد بهذه الطريقة بلغ عددها (٤٠ فقرة) بنسبة (٦٢%) من عدد فقرات الاختبار، وأن الفقرات التي لم تحقق الحد الادنى للثبات مقبول وفق هذه الطريقة بلغ عددها (٢٥ فقرة) بنسبة (٣٨%) من عدد فقرات الاختبار. حيث ان المعيار المعتمد لقبول معامل ثبات الفقرة في هذه طريقة هو أن تكون قيمة معامل الثبات أكبر أو يساوي (٠.٥٠) بهذه الحالة تعد الفقرة مقبولة. واما الطريقة الثانية طريقة الارتباط الثنائي الأصيل تبين ان جميع الفقرات حققت معامل ثبات مقبول، أي ان الفقرات التي تتمتع بثبات مقبول وفقاً لهذه الطريقة بلغ

عدد (٦٥ فقرة) بنسبة (١٠٠%) من عدد فقرات الاختبار، إذ ان المعيار المعتمد لقبول معامل ثبات الفقرة في هذه طريقة هو أن تكون قيمة معامل الارتباط دالة احصائيا، بهذه الحالة تعد الفقرة مقبولة وأنها تتمتع بمعامل ثبات مقبول. واما الطريقة الثالثة فهي طريقة حساب ثبات الفقرة من خلال تمييزها وانحرافها المعياري تبينت ان جميع الفقرات تتمتع بثبات مقبول بلغ عددها (٦٥ فقرة) بنسبة (١٠٠%) من عدد فقرات الاختبار. إذ أن المعيار المعتمد لقبول معامل ثبات الفقرة في هذه طريقة هو أن تكون قيمة معامل الثبات أكبر أو يساوي (٠.١٠) بهذه الحالة تعد الفقرة مقبولة وأنها تتمتع بمعامل ثبات مقبول. اما الطريقة الرابعة فهي طريقة معامل الفا ان الفقرات التي كانت قيم ثباتها أعلى من المعيار المحدد بهذه الطريقة بلغ عددها (٣٠ فقرة) بنسبة (٤٦%) من عدد فقرات الاختبار، وأن الفقرات التي لم تحقق الحد الادنى للثبات مقبول وفق هذه الطريقة بلغ عددها (٣٥ فقرة) بنسبة (٥٤%) من عدد فقرات الاختبار حيث ان المعيار المعتمد لقبول معامل ثبات الفقرة في هذه طريقة هو أن تكون قيمة معامل الثبات أكبر أو يساوي (٠.٣٠) بهذه الحالة تعد الفقرة مقبولة وأنها تتمتع بمعامل ثبات مقبول.

ثم بعد ذلك قاما الباحثان باستبعاد الفقرات التي كانت معامل ثباتها منخفض (حسب معايير كل طريقة من الطرائق الأربع) واعتمدت الفقرات التي حققت معامل ثبات مقبول وفق معايير كل طريقة من طرائق الثبات الأربع والتي عددها الكلي بعد استبعاد الفقرات الغير ملائمة للمعايير (٢٠٠) فقرة في الطرائق الأربع، من أجل التحقق من الفرضيات الآتية:

آ. لإ ة ج غ ف ئ ة خ لإ ك ب وء ح ئ ا ئ ح ئ ا ة ئ ط ف ق ئ ة غ ى ك ق ل ش ف ئ خ ة ذ ط ق ب ن ك ق ك ب.

للتأكد والتحقق من هذه الفرضية قاما الباحثان بحساب ثبات الفقرات لاختبار القدرة العقلية لهنمون نلسون بالطرائق الأربع المذكورة انفاً، ثم قاما باستبعاد الفقرات التي كانت معامل ثباتها منخفض (حسب معايير كل طريقة من الطرائق الأربع) واعتمدت فقط الفقرات التي حققت ثبات مقبول ثم قاما بحساب معامل صدق الاختبار (الصدق التمييزي) لكل طريقة من الطرائق الاربع، وذلك اعتمادا على (الحقيبة الاحصائية للعلوم الاجتماعية SPSS) وباستخدام (الاختبار التائي لعينتين مستقلتين T-Test) بين الدرجات الكلية للمجموعة العليا والدرجات الكلية للمجموعة الدنيا فكانت النتائج ان كل من طريقة (الارتباط الثنائي الأصيل) وطريقة (تمييز الفقرة وانحرافها المعياري) كانت قيم معاملات صدقها التمييزي اكبر من قيم الصدق التمييزي في الطريقتين الآخرتين، ومن أجل معرفة دلالة الفروق بين الطرائق الاربع في الصدق التمييزي قاما الباحثان باستخدام تحليل التباين الاحادي (One Way ANOVA) وذلك بالاعتماد على (الحقيبة الاحصائية للعلوم الاجتماعية SPSS) وبعد إجراء التحليل أظهرت النتائج ان القيمة الفائية

المحسوبة البالغة (١.٠٧٧) هي أصغر من القيمة الفائية الجدولية البالغة (٢.٦٠) عند مستوى دلالة (٠.٠٥) ودرجتي حرية (٣ ، ١٩٦).

حيث تبينت النتائج بان لا توجد فروق ذات دلالة احصائية بين الطرائق الاربع في الصدق التمييزي للاختبار، لذلك قامت الباحثة بقبول الفرضية الصفرية التي تنص على (لا توجد فروق ذات دلالة إحصائية بين طرائق حساب ثبات الفقرات في التحقق من صدق اختبار القدرة العقلية).

ا. لإ ة ج غ ف ئ ة خ لإ ك ب وء ح ئ ب ا ئ ح ئ ا ت ئ ة ط ق ئ ة غ ي ك ق ل ت ئ ة ئ خ ئ ذ ط ق ب ذ ك ق ب ك ب.

للتأكد والتحقق من هذه الفرضية قاما الباحثان بحساب ثبات فقرات اختبار القدرة العقلية لهنمون نلسون بالطرائق الأربع المذكورة انفاً، ثم قاما باستبعاد الفقرات التي كانت معامل ثباتها منخفض (حسب معياري كل طريقة من الطرائق الأربع) واعتمدت فقط الفقرات التي حققت ثبات مقبول ثم قاما بحساب معامل ثبات الاختبار (طريقة تحليل التباين باستخدام معادلة كيودر- ريتشاردسون 20) لكل طريقة من الطرائق الاربع، وذلك اعتمادا على (برنامج ميكروسوفت إيكسل Microsoft Excel) وباستخدام (معادلة كيودر- ريتشاردسون 20) فكانت النتائج ان هناك فرق بين قيم معاملات ثبات الاختبار في الطرائق الأربع، ومن اجل معرفة دلالة الفروق بين الطرائق الاربع في معاملات الثبات، قاما الباحثان باستعمال الاختبار الزائي لدلالة الفرق بين معاملي الارتباط، للمقارنات الثنائية لمعرفة دلالة الفرق بين معاملي ثبات كيودر-ريتشاردسون ٢٠ بين من الطرائق الأربع لثبات الفقرات.

وتبينت بان لا توجد فروق ذات دلالة إحصائية بين طرائق حساب الثبات (الاحتمال المنوالي، الارتباط الثنائي الأصيل، تمييز الفقرة وانحرافها المعياري) إذا كانت القيم الزائية المحسوبة أصغر من القيمة الزائية الجدولية والبالغة (١,٩٦) مما يدل على تقارب حساب قيم معاملات الثبات وفقاً لهذه الطرائق أي بمعنى لا يوجد افضلية لإحدى الطرائق الثلاث في حساب ثبات الاختبار.

وتبينت بان توجد فروق ذات دلالة إحصائية بين طرائق حساب الثبات (الاحتمال المنوالي، الارتباط الثنائي الأصيل، وتمييز الفقرة وانحرافها المعياري) وطريقة (الفا) في حساب ثبات الاختبار، إذا كانت القيم الزائية المحسوبة بين هذه (الطرائق الثلاث) وطريقة (الفا) اكبر من القيمة الزائية الجدولية والبالغة (١,٩٦) ولصالح الطرائق (الاحتمال المنوالي، الارتباط الثنائي الأصيل، وتمييز الفقرة وانحرافها المعياري) مما يدل على افضلية هذه الطرائق في حساب معاملات ثبات الاختبار على طريقة (الفا).

ك زئ ئلأح ئ ب:

آيلإبك ق بئلإح ئ بكعكلئلإج ئ ب (Spss)ئزَ ل لإز ئث ل :

١. التحليل العاملي بطريق المكونات الأساسية أُستعمل للتأكد من الصدق العاملي.

٢. معادلة الفاكرونباخ لحساب معامل ثبات الاختبار.

٣. الاختبار التائي لعينتين مستقلتين (T–Test) لحساب الصدق التمييزي.

٤. تحليل التباين الاحادي (One Way ANOVA) لحساب دلالة الفرق في تمييز الفقرات بين طرق ثبات الفقرات الاربعة.

٥. طرائق حساب ثبات الفقرة.

آ. الاحتمال المنوالي $\mathbf{R = \frac{n}{n-1}\left(L - \frac{1}{n}\right)}$

حيث ان: n = عدد البدائل. L = اعلى تكرار نسبي من بين البدائل.

ا. حساب ثبات الفقرة باستخدام الارتباط الثنائي الأصيل.

$$\boldsymbol{R = \frac{\bar{x}1 - \bar{x}2}{S} \times \sqrt{p \times q}}$$

حيث ان: $\bar{x}1$= متوسط الإجابات الصحيحة. $\bar{x}2$ = متوسط الإجابات الخاطئة.

S= الانحراف المعياري. p = نسبة الإجابات الصحيحة(معامل الصعوبة)

q = نسبة الإجابات الخاطئة (معامل السهولة).

ث. حساب ثبات الفقرة من خلال تمييزها وانحرافها المعياري $\mathbf{LRi = rit \times \sqrt{pq}}$

حيث ان: rit = تمييز الفقرة. P = نسبة الإجابات الصحيحة (معامل الصعوبة).

q = نسبة الإجابات الخاطئة (معامل السهولة).

خ. حساب ثبات الفقرة باستخدام معامل الفا.

$\mathbf{Ir = s_i\, r_{ix} = \sqrt{p\ i\ qi}\ r_{ix}}$

حيث ان: S_i = الانحراف المعياري للفقرة. r_{ix} = معامل ارتباط الفقرة بالدرجة الكلية.

٦. معادلة كيودر– ريتشاردسون 20 لحساب ثبات الاختبار للطرق الاربعة.

$$\boldsymbol{R_{kr20} = \left[\frac{K}{k-1}\right]\left[1 - \frac{\sum pq}{\sigma^2}\right]}$$

حيث ان: $\boldsymbol{K}$= عدد فقرات المقياس.

$\boldsymbol{P}$= نسبة الاجابات الصحيحة. $\boldsymbol{q}$= نسبة الاجابات الخاطئة. $\boldsymbol{\sigma^2}$= تباين المقياس الكلي.

٧. الاختبار الزائي لعينتين لدلالة الفروق بين معاملي الارتباط: استعمل في المقارنات الثنائية لمعرفة دلالة الفرق بين معاملي ثبات كيودر–ريتشاردسون ٢٠ بين من الطرق الأربعة لثبات الفقرات.

$$Z = \frac{Dr1 - Dr2}{\sqrt{\frac{1}{n1-3} + \frac{1}{n2-3}}}$$

حيث ان: Dr = القيمة الزائية المعيارية لمعامل الارتباط.

ئلإز ئجئ ة:

١. إمكانية استعمال طريقة الاحتمال المنوالي وطريقة الارتباط الثنائي الأصيل وطريقة تمييز الفقرة مضروبا بانحرافها في حساب ثبات الفقرات والابتعاد عن استعمال طريقة الفا.

٢. يعد ثبات درجة الفقرة مؤشراً فعالاً ومكملاً لمؤشرات الفقرة التقليدية ويمكن استخدامها لتحسين جودة الاختبارات النفسية والتربوية، ويعد ثبات الفقرة (Item Reliability) مؤشراً دقيقاً يعبر عن مدى اتساق كل فقرة داخل الأداة مع بقية الفقرات وقدرته على التنبؤ بالدرجة الكلية للاختبار، فالفقرات هي الوحدات البنائية التي تقوم عليها أدوات القياس وتؤثر بدرجة مباشرة على خصائص الأداة ومصداقيتها.

٣. اختيار فقرات بناء على ثباتها يمكن أن ينتج اختبارات ذات خصائص سيكومترية قوية دون الحاجة إلى الاعتماد فقط على معامل التمييز.

طك شدئ ة:

١. حث طلبة الدراسات العليا على حساب ثبات الفقرة بعده خاصة من خصائص السيكومترية للفقرات التي سترتبط بصدق البناء وباعتبارهِ كأداة تحليل إضافية الى جانب مؤشرات تحليل الفقرات الاخرى، وذلك عند بناء أو تعديل أدوات القياس النفسية والتربوية.

٢. على الباحثين والممارسين دمج هذا المؤشر ضمن قائمة مؤشرات جودة الفقرات خاصة في حالات تقويم الاختبارات القصيرة او عند الحاجة الى أدوات ذات استقرار مرتفع.

٣. يجب على الباحثين اختيار الفقرات الأحادية بعناية لضمان وضوحها وتوافقها مع البنية المستهدفة وتقليل احتمالية الخطأ العشوائي.

طكق حئ ة:

١. اجراء دراسة لحساب ثبات الفقرة في مقاييس الشخصية والاتجاهات والميول وأثرها في الخصائص السيكومترية للمقاييس.

٢. اجراء دراسة في حساب ثبات الفقرة باستعمال الطرائق الأخرى غير المستعملة في الدراسة الحالية وأثرها في الخصائص السيكومترية.

٣. اجراء دراسة مقارنة بين الطرائق التي يتم فيها حساب ثبات الفقرة من خلال إعادة التطبيق مثل (معامل ارتباط الرباعي الحقيقي فاي، معامل ارتباط الرباعي تتراشورك، معامل ارتباط بيرسون).

ك ئخذ:

- عبد الرحمن، سعد (١٩٨٣):ك ئ زكئف ى، الكويت، مكتبة الفلاح.
- ..... (٢٠٠٨):ك ئ زكئف ىك ب ك ، ط٥، هبة النيل العربية للنشر والتوزيع.
- السيد، فؤاد البهي (١٩٧٩): عك كئف ئإح ئئى ه ئ زكئق ك ، ط ٣، دار الفكر العربي للطباعة والنشر.
- الطيب، احمد محمد (١٩٩٩):كئق ك ئ ز كئف ى ك ، الإسكندرية، المكتب الجامعي الحديث.
- خطاب، علي ماهر (٢٠٠٣):ك ئ ز كئق غىكئعكلكئف ب ك ب ئإج ئ ب، ط٣، القاهرة، مكتبة الأنجلو المصرية.
- الوسيدي، أسامة بن عبد العزيز (٢٠١٧): فئذهب خ ئ كئف خئ ة غ ك ب كئقك ب م ب ئإز ئ بكئكف خ كك ئ ز ئز ىل ك ئقئ. (رسالة ماجستير منشورة). جامعة ام القرى. المملكة العربية السعودية.

النبهان، موسى محمد (٢٠١٣):ئزئز ئ ةك ئ ز كئق غىكئعكلك ك ب. ط٢، عمان – الأردن، دار الشروق للنشر والتوزيع

- معمرية، بشير (٢٠٠٢):ك ئ زكئف ى ة ئإخ ئئ ةكئف ب، منشورات شركة بانتيت، الجزائر.

– Anastasi, A & Urbina, S. (1997). **Psychological testing**. 7th ed; New York; prentice–Hall.

– Crocker, L.,& Algina,J (2008).**introduction to classical and modern test theory**. cengage learning.

– Gregory.R.L.(1996): **Psychological testing**: Hoistory, principles, and applications. 2nded Boston, Allyan and Bacon

– Lord, F. M & Novick, M. B. (1986). **Statistical theories of mental Test Scores**. **Addison– Wesley**.

– Magnusson, D (1967): **Testing theory**. London, Addison–wesley publishing company.

– Smith, m. (1966)." **The re lation ship between item validity**" Psychometrick " Vol.1, no 3. 69–76.

- Thayn, S.(2011). **An evaluation of multiple choice test questions deliberately designed to include multiple correct answers**, Ph. D, dissertation, the Faculty of Brigham Young university, from Proquest digtal.

- Zhang, s., & Colvin, K. (2024). ***Comparison of different reliability estimation methods for single- item assessment: a simulation study***. Frontiers in Psychology, https://doi:10.3389/fpsyg.2024.1482016

- Zijlmans, E. A. O., Tijmstra, J., van der Ark, L.A., & Sijtsma, K. (2018a) .**Methods for Estimating Item –Score Reliability**. Applied psychological measurement, Vol. 42(7),553-570. https://doi.org/10.1177/0146621618758290

- Zijlmans, E. A. O., Tijmstra, J., Van der Ark ,L.A., & Sijtsma, K.(2018b).**Item_Score reliability in empirical - data sets and its relationship with other item indices**. Educational and Pychological measurement, vol. 78(6) ,998-1020. https://doi.org/10.1177/0013164417728358.