Jordan Journal of Applied Science-Humanities Series

Volume 36
Issue 1 *36-1*

Article 8

2023

Estimating Reliability on Tests with dichotomous and Polychromous Items According to the classical Theory of Measurement and Generalizability Theory: A Comparative Study

Abdal hadi M.A Algharaibh
Yarmouk University, Irbid, Jordan., abed1973jor@yahoo.com

Follow this and additional works at: https://digitalcommons.aaru.edu.jo/jjoas-h

Part of the Arts and Humanities Commons, Business Commons, Education Commons, Law Commons, and the Social and Behavioral Sciences Commons

Recommended Citation

Algharaibh, Abdal hadi M.A (2023) "Estimating Reliability on Tests with dichotomous and Polychromous Items According to the classical Theory of Measurement and Generalizability Theory: A Comparative Study," *Jordan Journal of Applied Science-Humanities Series*: Vol. 36: Iss. 1, Article 8.
Available at: https://digitalcommons.aaru.edu.jo/jjoas-h/vol36/iss1/8

This Article is brought to you for free and open access by Arab Journals Platform. It has been accepted for inclusion in Jordan Journal of Applied Science-Humanities Series by an authorized editor. The journal is hosted on Digital Commons, an Elsevier platform. For more information, please contact rakan@aaru.edu.jo, marah@aaru.edu.jo, u.murad@aaru.edu.jo.

VOL 36 NO 1 ISSN: 2708-9126

Jordan Journal of Applied Science

Humanities Series

Applied science Private University

Estimating Reliability on Tests with dichotomous and Polytomous Items According to the classical Theory of Measurement and Generalizability Theory: A Comparative Study

تقدير الثبات في الاختبارات بفقرات ثنائية التدريج ومتعددة التدريج وفق النظرية الكلاسيكية في القياس ونظرية التعميم: دراسة مقارنة

Abdal hadi M.A Algharaibh[1]*, Ahmad Suliman Audeh bani Ahmad.[2]
Yarmouk University, Irbid, Jordan.[12]

ARTICLE INFO

Article history:

Received 25 Aug 2021

Accepted 05 Dec 2021

Published 01 Jul 2023

https://doi.org/10.35192/jjoas-h.v36i1.495

***Corresponding author at** Yarmouk University, Irbid, Jordan.

Abdal hadi M.A Algharaibh.

Email: abed1973jor@yahoo.com .

Keywords:

Classical-test theory

Generalization theory

Reliability, Raju equation

Signal-to-noise ratio

الكلمات المفتاحية :

نظرية الاختبار الكلاسيكية

نظرية التعميم

الثبات

معادلة راجو

نسبة اليقين إلى الشك

ABSTRACT

This current study aimed to compare reliability coefficients on tests with dichotomous and polytomous items according to classical theory of measurement and generalizability theory, to achieve this, a compound ability test was constructed in mathematics in two forms: Dichotomous and Polytomous by using five estimation methods, which as follows: (Alpha {α}, Theta{θ}, Raju, {G}, and {φ}), Furthermore, in order to achieve the study objectives, the test was applied on two forms, each test consists of 30 items to measure students mathematical ability on the field of numbers, Algebra, Geometries and Statistics, to 300 of the sample study. the reliability coefficients of the Dichotomous and estimated according to the Classical theory on the following manner: (α= 0.86), θ= 0.85, Raju= 0.86), and according to the Generalizability theory (G= 0.87, φ= 0.86), hence, the counter coefficients of Polytomous according to the Classical theory (α= 0.95, θ= 0.92, Raju= 0.93), and according to the Generalizability theory (G= 0.95, φ= 0.94). The results of the comparison between the coefficients according to the Cocron test and the signal/noise ratio test showed that there was a significant difference at (α =0.05) between the reliability coefficients values of the ability test on its two forms in favor of the polytomous, and the results also revealed a preference for generalizability theory in the estimation of reliability coefficients as extracted to classical coefficients.

هدفت هذه الدراسة إلى مقارنة معاملات الثبات في الاختبارات بفقرات ثنائية التدريج، ومتعدد التدريج وفق النظرية الكلاسيكية في القياس ونظرية التعميم، ولتحقيق ذلك، فقد تم بناء اختبار قدرة مركب في الرياضيات بصورتين: ثنائية التدريج، ومتعددة التدريج، باستخدام خمسة أساليب تقدير هي: (الفا(α)، ثيتا(θ)، راجو، G ، φ)، ولتحقيق أهداف الدراسة تم تطبيق الاختبار بصورتيه، وكل اختبار مكون من (٣٠) فقرة لقياس قدرة الطلبة الرياضية في مجال الأعداد والجبر والهندسة والإحصاء، على عينة من (٣٠٠) طالب وطالبة.وقد كانت معاملات الثبات للتدريج الثنائي والمقدرة وفق النظرية الكلاسيكية على النحو الآتي (α=٠,٨٧، θ=٠,٨٥، Raju=٠,٨٦)، ووفق نظرية التعميم (G=٠,٨٧، φ=٠,٨٦). وكانت المعاملات المقابلة للتدريج المتعدد وفق الكلاسيكية (α=٠,٩٥، θ=٠,٩٢، Raju=٠,٩٣)، ووفق نظرية التعميم (G=٠,٩٥، φ=٠,٩٤). وأظهرت نتائج المقارنة بين المعاملات وفق اختبار كوكرونCocron، واختبار نسبة اليقين إلى الشك signal/noise ratio إلى وجود فرق ذي دلالة إحصائية عند مستوى الدلالة (α=٠,٠٥) بين قيم معاملي ثبات اختبار القدرة بصورتيه لصالح التدريج المتعدد، كما كشفت النتائج عن أفضلية نظرية التعميم في تقدير معاملات الثبات مقارنة بالمعاملات المستخلصة وفق النظرية الكلاسيكية.

المقدمة

كثيرة هي الطرق التي تطرحها أدبيات القياس التربوي لتقدير الثبات (Reliability) بمعنى الدقة(precision) في القياس، وقد تعطي هذه الطرق تقديرات مختلفة لمعامل الثبات للمقياس الواحد؛ لأنها تقوم جميعها على تقدير الثبات، وليس حساب الثبات، ولذلك قد تفضي إلى قرارات مختلفة أو أراء أو تفسيرات مختلفة لنتائج القياس، كما ينظر إلى الثبات بمعنى تفسير نتائج القياس. وهناك مفاضلة بين المقاييس(فلا يوجد مقياس واحد لصفة تربوية أو نفسية واحدة)، بخلاف القياس الفيزيائي، الذي يقوم على وحدات قياس معروفة ومعيارية وقابلة للتحويل، ومن معادلات رياضية، بعكس القياس النفسي التربوي القائم على المنافسة بين المقاييس من خلال خصائصه السيكومترية، وتحديدا صدق المقياس وثباته. ومن هنا تأتي صعوبة تحديد الحد الأدنى المقبول لمعامل الثبات أو ما يطلق عليه بمستوى القطع للمعامل(cut point level)، وقد يخضع التحديد لأحكام شخصية، وقد يكون الاختلاف جوهريا بين هذه الأحكام، وقد تطرح الأدبيات بعض المقترحات يستند عليها الباحثون في الخروج من مآزق تحديد الحد الأدنى، وقد يكتفي البعض من الباحثين ومطوري أدوات القياس بعرض ما توصلوا إليه من خصائص سيكومترية تاركين للمستفيدين من هذه الأدوات أن يقرروا أو يحددوا خيارتهم. وقد تعددت أساليب تقدير الثبات إلى درجة أصبح يطلق عليها أنواع الثبات، وذلك لاختلاف الإجراءات والأساليب الإحصائية المعتمدة في هذا التقدير، ولكل من المختصين الذين ساهموا في تقويم هذه الأساليب مسوغات محددة، تاركين لمطوري الاختبارات، أو المستخدمين حرية اختيار الأسلوب المناسب، أو المعادلات الإحصائية المناسبة، حسب طبيعة أداة القياس ومصادر الأخطاء المحتملة، ومن أبرز الاختلافات بين هذه الأساليب هي تلك الاختلافات القائمة على النظرية، أو الأدب النظري في التعامل مع أخطاء القياس أو الافتراضات التي تقوم عليها نظرية القياس في التعامل مع الأخطاء.

وهناك عدة مفاهيم لمعنى الثبات، ويمكن أن نشير إلى إحداها في صورة مختصرة؛ بأنها ضمان الحصول على نفس النتائج تقريبا إذا أعيد تطبيق نفس الاختبار على نفس المجموعة من الأفراد، وهذا يعني قلة تأثير عوامل الصدفة العشوائية على نتائج الاختبار، وقد عرفت أنستازي (Anastasi, ١٩٨٢) الثبات، بأنه مدى دلالة تباين درجات الإفراد الذين اجري عليهم المقياس على الفروق الحقيقية في السمة، أي مدى الاتساق في الدرجات التي يحصل عليها من نفس الأفراد بتطبيق نفس الاختبار. وتلعب نظرية الاختبار الكلاسيكي دورا أساسيا، في فهم طرق تقدير ثبات درجات الاختبار، فالدرجة التي يحصل عليها الطالب في أي موقف اختباري تسمى بالدرجة الملاحظة(X) والتي تساوي الدرجة الحقيقية (T) + درجة الخطأ العشوائي (E)، وتمثل العلامة الملاحظة للفرد بالمعادلة التالية **X = T + E**. ويعرف الثبات إحصائيا بأنه النسبة بين تباين الدرجات الحقيقية وتباين الدرجات الملاحظة، وتتراوح قيمة معامل الثبات بين (٠) و (١) وكلما اقتربت قيمة معامل الثبات من (١) دل ذلك على درجة عالية من الثبات (علام ، ٢٠٠٢).

وتشير أدبيات القياس والتقويم التربوي والنفسي إلى طرق متعددة لتقدير قيم معامل الثبات بعضها تعتمد على تطبيق صيغتين متكافئتين للاختبار أو الاختبار نفسه مرتين بفاصل زمني، وبعضها الآخر يعتمد على تجزئة الاختبار إلى صيغتين متكافئتين. فنظرية الاختبار الكلاسيكي اعتمدت في تفسيرها للثبات على المقاييس المتوازية أو المتكافئة، فهي تفترض وجود مقياسين متكافئين أو أكثر لقياس سمة أو قدرة معينة، وذكر كروكر والجينا(٢٠٠٩) إن ضعف ومحدودية النظرية الكلاسيكية في معالجتها لمفهوم الثبات نظرا لأنها لا تميز بين أخطاء القياس، وإنما تعطي قيمة تقديريه كلية لمصادر أخطاء القياس المتعددة المتعلقة بإحدى صيغ الاختبار المستخدمة بواسطة فاحص أو مقيم معين في ظروف معينة. أن مثل هذه المحددات التي واجهت نظرية الاختبار الكلاسيكي في القياس، دعت كرونباخ وزملاءه إلى مراجعة مفهوم الثبات ومعالجته معالجة أكثر شمولية من منظور متعدد الأبعاد فاوجدوا ما يسمى بنظرية التعميم .

وتعد نظرية التعميم امتدادًا لنظرية الاختبار (CTT) من خلال إجراء تحليل التباين (ANOVA) لقضايا القياس، فهي تقدم إطارا مفاهيميا وإحصائيا موسعا لنظرية الاختبار الكلاسيكية، وتعد نموذجا يمكنه التعامل مع مصادر متعددة للخطأ، مما يسمح بتقدير دقة القياسات، وتقديم طرق فعالة في تحسين إجراءات القياس المستقبلية (Brennan,٢٠٠١).

وتقدم نظرية التعميم عددا من المزايا مقارنة بالنظرية الكلاسيكية للاختبار في تصميم إجراءات القياس: حيث يمكنها دمج مصادر متعددة للخطأ، وتقّيم في نفس الوقت: ثبات الاختبار-أعادة الاختبار، ثبات الاتساق الداخلي والصدق ألتقاربي والثبات بين المقيمين، وتسمح بتقديرات لكل جوانب القياس الفردية والتفاعل بينهما، وتقدم معلومات عن ثبات التفسيرات النسبية وحتى التفسيرات المطلقة، وتسمح للباحثين بتحسين ثبات الأداء ضمن قيود التكلفة للمال والوقت(Yin&shavelson,٢٠٠٨).

وهناك عدة نظريات تستخدم لحساب الثبات، منها ما يندرج تحت نظرية الاختبار الكلاسيكي، ومنها ما يندرج تحت نظرية التعميم، وفي نظرية الاختبار الكلاسيكي، أشار عودة (٢٠١٠). إلى أن هناك أكثر من طريقة لتقدير معامل الثبات نظرا لتعدد مصادر أخطاء القياس، فكل طريقة تتعامل مع مصدر واحد من مصادر أخطاء الدرجات، بمعنى أن بعض أخطاء القياس تظهر في نوع معين من معاملات الثبات ولا تظهر في نوع أخر، فقد تعددت أساليب تقدير الثبات بدءا من ثبات الإعادة، ثم استخدام الصور المتكافئة، إلى فكرة الاستخدام طرق الاتساق الداخلي(Internal consistency method) والتي تقدم تقدير الثبات على نصفي الاختبار: (A , B) وفق المعادلة (١).

$$r_{AB} = \frac{2r}{1+r}, r \neq 1 \quad \text{----------(١)}$$

حيث: r_{AB} معامل ثبات التجزئة النصفية، وr معامل الارتباط بين نصفي الاختبار، لكن معادلة سبيرمان- براون تقوم على افتراض أن يتحقق شرط التوازي بين نصفي الاختبار

وكون افتراض التوازي صعب تحقيقه، فقد قدمت العديد من المعادلات الأخرى لحساب الثبات باستخدام التجزئة النصفية بتوفر شرط أقل تشددا من شرط التوازي، تم التفكير ببدائل لمفهوم التوازي، ومنها التكافؤ الحقيقي بالضرورة(essentially) tau-equivalent parts)، والمتمثل(بتساوي التباين الحقيقي لجزئي الاختبار بفرق ثابت ، وتباين الأخطاء مختلف) (Lord&Novick,1968)، ومن أشهر هذه المعادلات: معادلة جوتمان(Guttman ,١٩٤٥)، معادلة رولون (Rulon , 1939)، وطريقة كرونباخ والمعدلات المشتقة منها: كودر ريتشادسون(٢٠)، (٢١) وتقوم الفكرة على حساب الارتباط على جميع الفقرات الداخلة في الاختبار، ويتم حساب معامل كرونباخ ألفا وفق المعادلة (٢).

$$\alpha = \frac{n}{n-1}\left[1-\frac{\sum\sigma_{xi}^{2}}{\sigma_{x}^{2}}\right] \text{----------(2)}$$

ومع أن شرط التكافؤ بالضرورة أقل تشددا من شرط التوازي إلا أن تحقيقه يبقى أمرا صعبا، إلى أن قدم راجو (١٩٧٧): معادلة لتقدير ثبات التجانس الداخلي للعلامة الناتجة من قياسات متشاكلة (concentric forms)، بمعنى وجود علاقة خطية تامة بين الدرجات الحقيقية، وهي التي تسمح باختلاف التباينات الحقيقية وتباينات الخطأ على أجزاء الاختبار، والتي تشترط أن تكون الأوزان النسبية لهذه القياسات معلومة، وتعتبر معادلة راجو امتداد لطريقة كرونباخ ألفا، تتعامل مع التشاكل ولا تتعامل مع التباين المشترك في حساب الوزن النسبي للجزء، وهذه المعادلة تستخدم فيما كان الاختبار مكونا من ثلاثة أجزاء متشاكلة فأكثر معلومة الأوزان، ويتم حساب الثبات وفق معادلة (٣).

$$r_{xx} = \frac{1}{1-\sum\lambda_{i}^{2}}\left(1-\frac{\sum\sigma_{i}^{2}}{\sigma_{x}^{2}}\right) \text{----------(3)}$$

حيث تمثل r_{xx} معامل الثبات، λi الوزن النسبي للجزء i، σ_{i}^{2} تباين العلامات في الجزء i، σ_{x}^{2} تباين العلامات الكلي

ومن طرق الثبات الأخرى: معامل ثيتا(θ)، اقترح (armor, 1947) صيغة لحساب معامل الثبات، حيث يدل N على عدد الفقرات، λi على الجذر الكامن الأول المستخرج بتحليل طريقة المكونات الرئيسية ويتم تقديره وفق معادلة (٤).

$$\theta = \left(N/N-1\right)\left(1-1/\lambda_{i}\right) \text{----------(4)}$$

حيث N تمثل عدد الفقرات، λ_{i}: أكبر جذر كامن .

ومن ضمن طرق الثبات: معامل نسبة الإشارة إلى الضوضاء (signal /noise ratio). حيث تشير الإشارة (Signal) إلى الدرجة الحقيقية، وتشير قوة الإشارة إلى التباين الحقيقي، بينما تأتي الضوضاء من التخمين وعدم اختيار الفقرات عشوائيا، ومن التأثيرات الأخرى التي تسبب اختلاف الدرجة الملاحظة عن الدرجة الحقيقية والتي تعطى في المعادلة (٥)

$$S/N = \frac{r}{1-r} \text{----------(5)}$$

النسبة S /N تشير إلى نسبة اليقين إلى الشك، وتشير (r) إلى معامل ثبات الاختبار، وتفيد هذه النسبة في مقارنة إجرائيين للقياس فيما يتعلق بالكفاءة، وهي ذات قيمة خاصة عند المقارنة بين إجرائيين مختلفين لمعالجة نفس البيانات (مثل صيغتين للثبات)،(Cronbach & Gleser,١٩٦٤).

و تقدم نظرية التعميم مفهوما للثبات مختلفا عن مفهوم الثبات في نظرية الاختبار الكلاسيكي، إذ يأخذ مفهوم الثبات في النظرية الكلاسيكية بالاعتبار أن هناك مصدراً واحداً للخطأ في القياس ، لذلك لا يمكن استخدام نوع واحد من الثبات (على سبيل المثال ثبات أعادة الاختبار) لعمل استنتاجات حول الأنواع الأخرى من الثبات مثل ثبات الاتساق الداخلي(Web Rowley & Shavelson,١٩٨٨). بينما مفهوم الثبات في نظرية التعميم يعبر عن دقة التعميم من الدرجة الملاحظة للمفحوص على الاختبار مثلا أو أي أداة قياس أخرى إلى متوسط الدرجة التي سوف يحصل عليها المفحوص تحت كل الشروط الممكنة(Bernnan,٢٠٠١). ويتم تقييم الثبات من خلال معامل التعميم، وهو مؤشر مماثل لمعامل الثبات في النظرية الكلاسيكية، يتراوح من(٠ إلى ١)، وتشير القيم التي تقترب من ١ إلى أن درجات الاهتمام يمكن تمييزها بدرجة عالية من الدقة على الرغم من التقلبات العشوائية في ظروف القياس(Alla & Cardinet, ١٩٩٧; Shavelson &Webb, ١٩٩١; Strube, ٢٠٠٢) .

و في إطار نظرية التعميم، تعتبر الملاحظات(مثلا درجات الطلبة في الاختبار) التي يقوم الباحث بجمعها في موقف معين، بمثابة عينة من الملاحظات، التي يشتمل عليها نطاق شامل من الملاحظات يتضمن "شروط جمع الملاحظات"، أي المثيرات التي يتبع في ضوئها جمع الملاحظات، أو الحصول على القياسات، ومجموعة الشروط التي يتم إجراء الملاحظات لها تسمى (عالم الملاحظة المقبولة the universe of admissible observaton)، والذي يقوم بتحديده كل باحث، ويشير عالم الملاحظات المقبولة إلى كل الأبعاد التي يمكن اختيارها من أجل إجراء القياس حول موضوع القياس عادة ما يكون الفرد، ويمكن أن تتباين الدرجات التي تنتمي إلى نطاق شامل معين في أكثر من بعد، ويمكن أن تكون الأبعاد مهمات الاختبار أو المقيمين أو الفقرات التي يجري فيها الاختبار، وتسمى مصادر خطأ القياس أوجه(Facets)، ومستويات هذه الأوجه شروط(Conditions)، (علام،٢٠٠٢).

و يتضمن استخدام نظرية التعميم إجراء نوعين من الدراسات البحثية: دراسات التعميم(G) ودراسات القرار(D). فالباحث الذي يجري دراسة(G) يهتم بتحديد مدى جودة الدرجات في مواقف معينة، لذلك الاهتمام في دراسة (G) هو تعميم النتائج التي تم الحصول عليها من عينة القياس إلى نطاقه، وتتضمن دراسة (G) تقدير مكونات التباين التي ترجع إلى اختلاف مستويات الأوجه، وكذلك تقدير تفاعل هذه المكونات باستخدام طرق تحليل التباين (analysis of variance)، ومن خلال مكونات التباين يمكن تقدير ثبات أداة القياس عن طريق حساب معاملات التعميم (coefficients of generalizability)، والتي تعبر عن النسبة بين تباين الدرجة الشاملة (universe score) إلى تباين الدرجة الملاحظة (observed score). أما بالنسبة لدراسات(D) فإنها تهدف لتحديد الإجراء الأكثر كفاءة في تحديد ثبات الأداة بالاعتماد على المعلومات التي تقدمها دراسات التعميم. ونظرا لوجود نوعين من الخطأ: الخطأ النسبي والخطأ المطلق، يوجد في نظرية التعميم معاملان للثبات: معامل الثبات النسبي ومعامل الثبات المطلق(Bernnan,٢٠٠١;shvelson&Webb,١٩٩١).

مشكلة الدراسة وأسئلتها

إن بناء اختبارات تتمتع بخصائص سيكومترية موثوقة أمر بالغ الأهمية، فغالبية الدراسات التي اهتمت ببناء الاختبارات اعتمدت على حساب الثبات بالطرق التقليدية للقياس (الكحلوت،٢٠٠٢ ؛ الذنيبات،٢٠١٣ ؛ علاونة،٢٠١٦)، ويعتبر معامل كرونباخ ألفا الأكثر استخداما لتقدير ثبات درجات الاختبارات في الآداب التربوية والعلوم السلوكية، وعدد قليل من الدراسات التي اعتمدت على نظرية التعميم في حساب الثبات(فاروق، ٢٠١٦)، إن استخدام الاختبارات في البرامج الدراسية للكشف عن قدرات الطلاب وقياس مستواهم ألتحصيلي، وكما تستخدم الاختبارات في غايات تصنيف الطلاب، وقياس ذكائهم، وفي عمليات توجيههم وإرشادهم، لا يمكنها الاستغناء عن فحص جودة قياسات درجاتها، والتي تتطلب إجراءات ملائمة للتحقق من شروط جودتها الفنية، إن استخدام النظرية الكلاسيكية للاختيارات أصبحت غير كافية في تقدير ثبات الدرجات وذلك لأنها لا تأخذ بعين الاعتبارات جميع مصادر الأخطاء (مصححين وفترات التقييم وطرق التقييم والمهام وشكل الاختبار)، من هنا جاءت الحاجة إلى مثل هذه الدراسة لتسليط الضوء على نظرية التعميم بهدف مقارنة معاملات الثبات لاختبار قدرة مركب في الرياضيات بصورتين: ثنائية التدريج، ومتعددة التدريج لدى طلاب الصف التاسع الأساسي من خلال الإجابة على مجموعة من الفقرات التي تتعلق في مجال الأعداد والجبر والهندسة والإحصاء، وفقا لمصادر الخطأ في ضوء نظرية الاختبار الكلاسيكي ونظرية التعميم، وذلك من أجل الوقوف على أفضل إجراء يتمتع بالثبات في موقف معين.

ولتحقيق هدف الدراسة سوف يقوم الباحث ببناء اختبار قدرة في الرياضيات يتكون من نوعين من الفقرات احدهما اختبار اختيار من متعدد(ثنائي التدريج) ، والأخر يتكون من مهمات محددة الإجابة (متعدد التدريج) وبذلك يتوقع أن تجيب الدراسة عن الأسئلة التالية:

السؤال الأول: ما هي معاملات الثبات المقدرة لاختبار القدرة الرياضية حسب نظرية الاختبار الكلاسيكي؟

السؤال الثاني: ما معاملات الثبات المقدرة لاختبار القدرة الرياضية حسب نظرية التعميم؟

السؤال الثالث: هل تختلف معاملات الثبات باختلاف نوع أسئلة الاختبار؟

أهمية الدراسة

يمكن توضيح هذه الأهمية في جانبين من الناحية النظرية والعملية، فمن الناحية النظرية تقدم مقارنة بين معاملات الثبات لاختبار قدرة مركب في الرياضيات بصورتين: ثنائية التدريج، ومتعددة التدريج، باستخدام خمسة أساليب تقدير هي: (الفا (α)، ثيتا (θ)، راجو، معاملات التعميم(G)، (φ))، وتبين مدى التوافق بين معاملات الثبات المقدرة وفقا لهذه الطرق، وذلك من خلال بيانات حقيقية (Real Life Data) من الميدان من خلال وضع اختبار لتقييم القدرة الرياضية لدى الطلبة، للكشف عن المهارات والمعارف والقدرات التي يمتلكها الطالب، بحيث يمكننا من المقارنة بين الطلاب ومدى امتلاكهم لمهارات التفكير المعقدة ومن الناحية العملية الاستفادة من الاختبار في هذه الدراسة من اجل الوقوف على نطاق القوة والضعف التي يعاني منها الطلاب في مادة الرياضيات من اجل تنمية طرق التفكير والاستنتاج والقدرة على حل المشكلات المعقدة، ويتوقع الباحث أن تساهم هذه الدراسة في تحقيق الفهم العلمي لدى الباحثين في كيفية تقدير الخصائص السيكومترية من خلال حساب الثبات في النظرية الكلاسيكية في القياس ونظرية التعميم والمقارنة بين النظرتين، كما أنها ستفتح المجال أمام الباحثين في إجراء المزيد من الدراسات للمقارنة بين أشكال أخرى من الاختبارات ومعرفة أثرها على الخصائص السيكومترية للاختبار.

مصطلحات الدراسة:

معامل الثبات في التقليدية: هو نسبة التباين الحقيقي إلى التباين الكلي

تعريف الثبات في نظرية التعميم من الناحية الإجرائية: يشير الثبات إلى معامل الارتباط بين الدرجة الملاحظة للطالب في اختبار القدرة في الرياضيات من خلال أدائه على مهام الاختبار والدرجة الشاملة التي يحصل عليها ضمن شروط القياس الممكنة.

نظرية التعميم : أسلوب إحصائي يتم استخدامه لتحديد ثبات القياسات في ظروف محددة، وتهتم هذه النظرية بتحليل لمكونات التباين المتعددة لمجموعة من القياسات أو الملاحظات.

الاختبار الثنائي التدريج (Dichotomous): اختبار تكون الاستجابة لفقراته متغيرا ثنائيا وفيها تعطى الإجابة عن الفقرة العلامة (١) عندما تكون الإجابة صحيحة، وتعطى الإجابة العلامة (صفر) إذا كانت خطأ.

الاختبار المتعدد التدريج (Polytomous): اختبار تكون الاستجابة عن فقراته متدرجة، اذ تكون علامة الفرد متدرجة غير كاملة، فقد يحصل الفرد على علامة جزئية مثل العلامة ٢ من ٣ أو العلامة ١ من ٣، حيث يلجأ واضع السؤال عادة إلى تجزئة الإجابة إلى خطوات، وتعطى علامة جزئية على هذه الخطوات، لتشكل العلامة في مجموعها العلامة الكلية للسؤال وفي هذه الحالة يراعي هذا النوع من الأسئلة المعرفة الجزئية.

محددات الدراسة:

- اقتصرت عينة الدراسة على طلبة الصف التاسع الأساسي في المدارس الحكومية في محافظة عجلون في الفصل الدراسي الثاني لعام ٢٠٢٠ /٢٠٢١.
- اقتصرت الدراسة على استخدام اختبار قدرة في الرياضيات بنوعين من الفقرات: ثنائية التدريج، ومتعددة التدريج.

الدراسات السابقة:

الدراسات التي تناولت موضوع المقارنة بين أساليب تقدير الثبات قليلة في الاختبارات: بفقرات ثنائية التدريج ومتعدد التدريج وفق النظرية الكلاسيكية في القياس ونظرية التعميم لكن هناك دراسات تناولت المفاضلة بين النظريتين.

وتعد الدراسة التي أجراها أيسون (Eason, ١٩٩١) وهدفت للمقارنة بين نظرية الاختبار الكلاسيكية ونظرية التعميم، أكد فيها أن طرق تقدير معاملات الثبات في نظرية الاختبار الكلاسيكية لا تأخذ في الاعتبار جميع مصادر تباين الخطأ في آن واحد، وهذا يؤدي إلى اختلاف معاملات الثبات عند استخدام تلك الطرق، ومن أجل المقارنة بين النظريتين قام الباحث بحساب معامل الثبات لبعض البيانات الموجودة في الأبحاث السابقة مستخدما طرق نظرية الاختبار الكلاسيكية ونظرية التعميم، وتوصل الباحث إلى أن الدرجات تكون أكثر استقرارا وثباتا باستخدام الطرق التي تعتمد على نظرية التعميم في تقدير الثبات وذلك لان هذه النظرية تأخذ جميع المصادر لتباين الخطأ .

وكما قام ثومبسون وجرولي (Thompson&Growely,١٩٩٤) بدراسة هدفت للمقارنة بين النظرية الكلاسيكية للاختبار ونظرية التعميم، أكد فيها أنه لا يزال الاعتماد مستمرا على استخدام النظرية الكلاسيكية في حساب معامل الثبات، مما أدى إلى عدم تماسك النتائج، نظرا لما تعانيه الطرق التقليدية من قصور في مصادر تباين الخطأ، ولا زال التركيز في الجامعات على تدريس النظرية الكلاسيكية على الرغم من ظهور نظرية التعميم قبل أكثر من عقدين من الزمن، فنظرية التعميم تستخدم تحليل التباين والتي تقوم على تجزئة التباين حسب العوامل المؤثرة في درجات الاختبار مثل المفحوصين والمقيمين والفقرات والتفاعل بينهم.

وأجرى بردجمان(Brideman,١٩٩٤) دراسة هدفت إلى المقارنة بين درجات الطلاب على اختبار من متعدد، ودرجاتهم على اختبار الإجابة المصاغة، وكشفت نتائج الدراسة على أن متوسط درجات الطلاب على اختبار الإجابة المتعدد أقل من متوسط درجاتهم على اختبار الإجابة المصاغة، وأن ثبات اختبار الإجابة المصاغة أعلى من ثبات الاختيار من متعدد.

وفي دراسة الثبيتي(١٩٩٨) التي هدفت للمفاضلة بين نظرية الاختبار الكلاسيكية ونظرية التعميم في تقدير أدوات القياس في العلوم الإنسانية، تناولت الدراسة الطرق المختلفة لتقدير معامل ثبات أدوات جمع المعلومات في البحوث التربوية والنفسية المنشورة في بعض دوريات التربية وعلم النفس، وحققت مجموعة من الأهداف التي كان من أهمها، إظهار أهمية التحقق من ثبات أدوات القياس وإبراز العلاقة بين أخطاء القياس وثبات الدرجات وتوضيح المفاهيم الأساسية المتعلقة بالنظريتين وذلك من خلال اختيار (٥٠) دراسة كعينة عشوائية، توصلت الدراسة إلى مجموعة من النتائج كان أبرزها إن الأسباب وراء تناقض كثير من نتائج البحوث هو عدم إدراك الباحثين أهمية دراسة طرق حساب ثبات أدوات المعلومات وأن يدرك الباحث أن ثبات أدوات القياس مرتبط بالبيانات وليس بأداة القياس، وكان من ابرز نتائجها تفوق نظرية التعميم على نظرية الاختبار الكلاسيكية عند تقدير معامل الثبات.

ومن الدراسات التي تناولت استخدام نظرية التعميم في تقدير معاملات الثبات الدراسة التي قام باجرائها مكبي وبارنس (Mcbee & Barnes,١٩٩٨) والتي هدفت لتقييم التحصيل في الرياضيات لدى عينة مكونة من (١٠١) من طلاب الصف الثامن، قام الباحثان بتطوير أربع مهام أداء، مهمتين أكثر تشابها إلى حد كبير لقياس مجال حل المشكلات المعقدة ، تم تقدير أدائهم بالاعتماد على نموذج تقيم المهام، اعتمد الباحثان على تصميم ملاحظة (فرد x مهمة x مقدر x فترة) لتحليل البيانات، أظهرت نتائج التحليلات أن معامل الثبات بين التقديرات (٠,٨٠) بينما كانت معاملات ثبات الإعادة اقل بكثير تتراوح بين(٠,٢٨ إلى ٠,٥٤)، أظهرت نتائج دراسة ارتفاعا في تباين طالب مع المهمة، وتفاعل طالب مع مهمة مع مقدر الممزوج بالخطأ العشوائي وانخفاضا في مصادر تباين الخطأ بين المهمات الأكثر تشابها ، وكشفت نتائج دراسات القرار أن عدد المهام المطلوبة للوصول إلى مستويات مقبولة من التعميم سيكون مرتفعًا بشكل محظور، حتى باستخدام مهام متشابهة للغاية.

وتعد دراسة لي وكانتور(٢٠٠٦ , Lee & Kantor) التي هدفت فحص تأثيرات المهمات والمقدرين على درجات الكتابة وتأثير عدد المهمات والمقدرين وتصميمات التقدير على ثبات الاختبار بواسطة نظرية التعميم، طبق الباحثان (٦) مهمات مدمجة في الكتابة (الاستماع والقراءة)، على عينة مكونة (٤٤٨) طالبا من طلبة اللغة الانجليزية في خمسة بلدان (أمريكيا ، كندا ، هونج كونغ ، المكسيك وتايوان)، وتم تقييم المهمات من قبل (٣) مقدرين، وتم تحليل البيانات من خلال تصميمين الأول ثنائي البعد متقاطع جزئيا " فرد: مقدرx مهمة "، والثاني ثنائي البعد متقاطع كليا " فرد x مهمة x مقدر"، وقد كشفت نتائج الدراسة أن معاملات التعميم كانت مقبولة، وان أكبر مصدر تباين للخطأ كان بسبب تفاعل الفرد مع المهمة مع المقدر الممزوج بالخطأ غير العشوائي.

وتفاعل الفرد مع المهمة، في حين أظهرت انخفاضا في مصادر التباين الأخرى، بالإضافة إلى إن زيادة عدد المهمات أكثر تأثيرا من زيادة عدد المقدرين في رفع معاملات الثبات بمعاملات تعميم نسبية ومطلقة بلغت (٠,٨٥) و(٠,٨١) على التوالي.

وبينت دراسة ني ويو ولو (Nie , Yeo & Lau ٢٠٠) التي تناولت فحص استخدام نظرية التعميم لتقييم جودة التقييم البديل (كتابة مجلة) في الرياضيات، شارك (٢٩) طالب من طلبة المرحلة الثانوية بكتابة مهام المجلات في الموضوعات المحددة، وقام اثنان (٢) من المقيمين بتقدير أداء الطلاب، باستخدام تصميم ثنائي البعد:(طالب x مهمة x مقدر) حيث توصلت الدراسة إلى أن معامل التعميم النسبي (٠,٧٦) ومعامل التعميم المطلق (٠,٧٢)، أظهرت النتائج أن زيادة عدد المهام له تأثير اكبر على معاملات التعميم أكثر من زيادة عدد المقيمين .

وأجرى باين (Bain, ٢٠٠٨) دراسة هدفت إلى إعداد اختبار مشترك لتقييم كفاءات الرياضيات وفق نموذج التعميم، طبق الباحث الاختبار المكون من (١٥) مهمة على عينة من (٤٣) طالبا من طلاب الصف التاسع بجنيف، وتم تحليل البيانات وفق تصميم متقاطع جزئيا (طالب x مهمة : مدرسة x مجموعة المستوى)، أظهرت نتائج الدراسة عن ارتفاعا في معاملات الثبات النسبية والمطلقة، حيث بلغت (٠,٩٠)، (٠,٨٩)، و أشارت النتائج إلى ارتفاع في مصدر تباين تفاعل طالب مع المهمة المتداخل مع المدرسة والمجموعة.

وتعد الدراسة التي قام بها جولر وجيلبال(Guler & gelbal , ٢٠١٠) التي هدفت إلى المقارنة بين نظرية الاختبار الكلاسيكية ونظرية التعميم لتحديد ثبات الدرجات التي تم الحصول عليها من أداة قياس النجاح في الرياضيات، قام الباحثان بتطبيق ٢٤ مسألة مفتوحة في الرياضيات في(TIMSS-١٩٩٩)على عينة (٢٠٣) طلاب في فصل الربيع ٢٠٠٧، وقد كشفت نتائج الدراسة أن معامل الاتساق الداخلي للدرجات بلغ (٠,٩٢)، وأن معامل التعميم لدرجات الرياضيات (٠,٩٢) وكان معامل فأي(٠,٩٠.)، استأثر مكون التباين للمقيمين بنسبة ٢,١٪ من التباين الكلي وفقًا لجميع النتائج، فقد لوحظ أن أداة قياس نجاح الرياضيات كانت تتمتع بثبات مرتفع، وعلى الرغم من وجود فارق بين أوساط درجات المقيمين الأربعة، فقد وجد أن هناك اتساقًا في نتائجهم.

وفي مجال الدراسات العربية لم يجد الباحث إي دراسة قارنت بين طرائق الثبات وفق نظرية الاختبار الكلاسيكي ونظرية التعميم، لكن وجد دراسات قارنت بين النظرية الحديثة والنظرية الكلاسيكية في القياس، وتعد دراسة الأسمر(٢٠٠٧) التي هدفت إلى تقدير الخصائص السيكومترية لصورتي اختبار في الرياضيات فقرات إحداهما ثنائية التدريج والأخرى متعددة التدريج وفق نظريتي القياس التقليدية والحديثة، وتمثلت عينة الدراسة من (٦٣٠) طالب وطالبة، حيث طبق الاختبار الثنائي التدريج على (٦٣٠) طالب وطالبة، بينما طبق الاختبار المتعدد التدريج على (٣٠٠)، توصلت الدراسة إلى أن معامل الثبات للاختبار الثنائي التدريج بطريقة كرونباخ ألفا (٠,٨٩)، وبمعادلة سبيرمان براون (٠,٩١)، بينما بلغ معامل الثبات للاختبار المتعدد التدريج بطريقة كرونباخ ألفا (٠,٩٣) وبمعادلة سبيرمان براون (٠,٩٥)، وبطريقة الصور المتكافئة فقد بلغ معامل الارتباط بين الصورتين (٠,٨٥)، أي أن معامل الثبات للمتعدد التدريج أعلى من معامل الثبات للثنائي التدريج.

وفي الدراسات العربية التي استخدم فيها معامل راجو لتقدير الثبات، فقد أجرت عليان (٢٠٠٤) دراسة لتقدير الثبات للعلامات المدرسية في مدينة اربد الكبرى، وقد أشارت نتائج الدراسة إلى تمتع العلامات في ٩٢,٣ % من الشعب المدرسية بدرجة عالية من الثبات، حيث يزيد معامل الثبات لها عن (٠,٨٠)، وأشارت النتائج إلى عدم وجود فروق ذات دلالة إحصائية بين توزيعات مستويات الثبات حسب متغيرات نوع المدرسة والمبحث ومتغير دراسة المعلم لمساقات القياس والتقويم أو عدم دراسته المختلفة .

كما أجرى الشايب (٢٠٠٧) دراسة هدفت إلى تقدير ثبات العلامات النهائية في جامعة آل البيت، وقد أظهرت النتائج باستخدام أسلوب راجو لتقدير الثبات عن تدني معاملات ثبات العلامات بشكل عام، حيث بلغ المتوسط العام لتقديرات الثبات (٠,٦٦)، كما أظهرت النتائج تمتع (٣١,٢٥%) فقط من المواد بمعاملات ثبات مقبولة.

كما أجرى العمري (٢٠١٨) دراسة هدفت إلى المقارنة بين ثلاث طرائق في تقدير ثبات الاختبارات المركبة التي تتضمن نوعين من الفقرات (ألفا، آلفا الطبقي، راجو) وذلك باستخدام ٨٤ اختبارا من الاختبارات النهائية بمشاركة عينة من (٣٠) طالب وطالبة للاستجابة على كل اختبار، بمتوسط (١٣) فقرة لكل اختبار وتكونت جميعها من نوعين من الفقرات الاختيار من متعدد والمقالية، أشارت نتائج الدراسة إلى إن الوسط الحسابي لتقديرات الثبات بلغ (٠,٧١)، (٠,٧٩)، (٠,٥٣) حسب معادلة ألفا وإلفا الطبقي وراجو على الترتيب ، وأشارت نتائج التحليل التباين واختبار كأي تربيع إلى وجود فروق دالة إحصائيا بين توزيعات معاملات الثبات وبمعامل توافق بينهما بلغ (٠,٣٦).

ومن الدراسات العربية التي استخدمت نظرية التعميم، فقد قام فاروق (٢٠١٦) بدراسة هدفت إلى تطبيق نظرية التعميم في تقدير ثبات اختبار تقييم كفاءة الطلاب في الرياضيات، للكشف عن مصادر التباين المؤثرة على ثبات درجات الطلاب، قام بتطبيق اختبار يتكون (٠٩) مهمات معقدة حول كفاءة الحساب والأعداد، على عينة مكونة من (٣٣١) في الرابع الابتدائي، وقد قام بتصحيح المهمات (٣) مصححين مدربين باستخدام شبكات تصحيح تحليلية، وتم تحليل البيانات بواسطة تصميم ثنائي البعد متقاطع كليا "طالب x مهمة x مقدر "باستخدام حزمة " EduG "، كشفت نتائج الدراسة وجود مصادر خطأ كبيرة راجعة إلى اثر تفاعل طالب مع مهمة والتأثير الرئيسي للمهمة، وكشفت الدراسة عن أن معامل التعميم النسبي بلغ (٠,٧٩) ومعامل التعميم المطلق (٠,٧٤) ومن اجل ضمان مستويات ثبات مقبولة يجب زيادة عدد المهمات وليس عدد المصححين. والدراسة التي أعدها الباحثان الحربي والحربي (٢٠١٧) حول مؤشرات الثبات باستخدام نظرية التعميم ومؤشرات صدق البناء لمقياس موهبة الإبداع، والتي هدفت إلى فحص جودة مقياس موهبة الإبداع من خلال دراسة مواصفات الاختبار الجيد: الثبات والصدق.

استخدم الباحث مُعامل كرونباخ ألفا لدراسة اتساق العلامات الناتجة من المقياس ونظرية التعميم للحصول على جودة تقويم المقيمين لإجابات الطلاب على مقياس الإبداع، واستخدم الباحث التحليل ألعاملي ألتوكيدي لدراسة البناء التكويني للمقياس، وبعد تحليل نتائج الدراسة تبين أنَّ مقياس موهبة للإبداع يتمتع بجودة عالية، وتوصلت نتائج الدراسة إلى أن معاملات كرونباخ الفا تراوحت بين (٠,٧٤) و(٠,٨٨) وبلغت (٠,٩٠) للمقياس ككل، وأنَّ مؤشرات نظرية التعميم لدراسة ثبات المقيمين قدرت بـ (٠,٨١) لمقيم واحد و(٠,٨٦) لثلاثة مقيمين.

وعليه فان الدراسات السابقة قد تناولت استخدام هذه المعادلات على انفراد في تقدير الثبات، ولم أجد دراسة تناولت موضوع المقارنة بين طرق تقدير الثبات في النظرية التقليدية ونظرية التعميم لاختبار قدرة في الرياضيات باستخدام نوعين من الفقرات في ضوء بيانات حقيقية، إلا أنها تتشابه مع بعض الدراسات في بعض الجوانب فمثلا دراسة (Guler & Gelbal ٢٠١٠)تقارن بين معاملات الثبات وفق النظرية الكلاسيكية ونظرية التعميم، وركزت معظم دراسات نظرية التعميم على استخدام مهمات في الرياضيات، وهو مجال اهتمام الدراسة الحالية، واختلفت مع دراسات نظرية التعميم من حيث عدد المهمات التي كانت قليلة في معظم الدراسات وتتشابه مع بعض الدراسات من حيث استخدام تصاميم أحادية وثنائية البعد متقاطع كليا ، لذلك ستنفرد هذه الدراسة بتقديم مقارنة بين خمسة أساليب لتقدير الثبات في ضوء بيانات حقيقية باستخدام اختبار قدرة مركب بفقرات: ثنائية التدريج ومتعددة التدريج وفق نظريتي الاختبار الكلاسيكية في القياس ونظرية التعميم .

الطريقة والإجراءات:

منهج الدراسة : بما أن الدراسة ذات طبيعة سيكومترية اهتمت بتقدير ثبات اختبار القدرة الرياضية، وفق نظريتي القياس الكلاسيكية والتعميم، فان اغلب هذا النوع من الدراسات وصفية استكشافية، لذلك فان المنهج المستخدم في هذه الدراسة وصفي يهدف إلى تقدير معاملات الثبات.

مجتمع الدراسة:

يتكون مجتمع الدراسة من جميع طلاب وطالبات الصف التاسع الأساسي للعام الدراسي ٢٠٢٠\ ٢٠٢١ التابعين لمدارس مديرية التربية والتعليم لمحافظة عجلون في وزارة التربية والتعليم الأردنية، ويبلغ عددهم (٣٠٦٨) طالبا وطالبة، منهم (١٥١٩) طالبا (١٥٤٩) .

عينة الدراسة:

تم اختيار عينة الدراسة بالطريقة الطبقية العنقودية العشوائية، حيث تم اختيار المدارس التابعة لمديرية التربية والتعليم لمحافظة عجلون ، ثم اختيار الصف المشمول بالدراسة (التاسع الأساسي)، حيث تكونت عينة الدراسة من (٣٠٠) طالب وطالبة من طلبة الصف التاسع الأساسي، موزعين على (٧) مدارس.

أداة الدراسة:

لتحقيق أهداف الدراسة، تم إعداد اختبار قدرة مركب في الرياضيات، للصف التاسع الأساسي في مجال الأعداد والجبر والهندسة والإحصاء يتكون من نوعين من الفقرات: النوع الأول هو اختبار من متعدد(ثنائي التدريج) لكل مفردة (٤) بدائل يتكون من (٣٠) فقرة والنوع الثاني يتكون من (٣٠) مهمة محددة الإجابة (متعدد التدريج) في الرياضيات وقد تم تحديد مادة الاختبار بناء على اختبارات التميسTIMS الدولية، وقد تم إعداد الاختبار بناء على الخطوات المتبعة في إعداد مثل هذه الاختبارات،

صدق المحتوى للاختبار(Content Validity):

وللتأكد من صدق المحتوى تم عرض الصورة الأولية للاختبار على مجموعة من المختصين في القياس والتقويم والرياضيات، ومناهج وطرق تدريس الرياضيات، وذلك لأخذ وجهات نظرهم في مدى صدق الفقرة الاختبارية في قياس الهدف السلوكي المحدد، والتحقق من سلامة الصياغة اللغوية ووضوح المعنى، واقتراح ما يرونه مناسبا من تعديلات، وذلك من خلال الإجابة على عن استبيان مرفق مع كل اختبار تم إعداده لتحقيق هذا الغرض، وفي ضوء أراء المحكمين، تم تعديل بعض أسئلة الاختبار إضافة إلى بعض البدائل الغامضة، بعد التحقق من صدق محتوى الاختبار، قام الباحث بتطبيق الاختبار بصورته الأولية على عينة استطلاعية تألفت من (٥٠) طالباً وطالبة من خارج عينة الدراسة، بهدف التحقق من وضوح التعليمات ومناسبة الوقت، والتي من خلالها تم التأكد من وضوح محتوى الفقرات، وتعليماتها وملائمة زمن الإجابة عليها.

إجراءات تطبيق الدراسة:

طبق الباحث الاختبار على عينة الدراسة خلال الفصل الدراسي الثاني لعام ٢٠٢٠ /٢٠٢١ م، وبفقرات الاختبار بصورتيه النهائية والبالغ عددها (٣٠)، فبعد جلوس أفراد الدراسة للاختبار، وزع الاختبار على الطلاب بشكليه حيث بدأ بأسئلة الاختبار الثنائي التدريج، وبعد الانتهاء من إجابة هذه الأسئلة أعطي الطلبة النوع الأخر، الاختبار المتعدد التدريج، واشرف الباحث على إجراءات تطبيق الاختبار بمساعدة معلمو الرياضيات في المدارس التي طبق بها الاختبار.

تصحيح الاختبار:

قام الباحث بإعداد نموذج تصحيح للاختبار الثنائي التدريج، وذلك باعطاء رقم (١) للإجابة الصحيحة وإعطاء رقم (٠) للإجابة الخطأ، و إعداد نموذج تصحيح للاختبار المتعدد التدريج، تتراوح درجات الطالب على كل فقرة بين (٠ إلى ٣)، بحيث تكون الدرجة صفر عند عدم الإجابة، والدرجة ٣ في حالة الإجابة الكاملة، تم تصحيح أوراق الاختبار بصورتيه، بمساعد(٣) من المقيمين لتقدير أداء الطلاب على الفقرات المتعددة التدريج وتم استخراج العلامات وتفريغ إجابات المفحوصين حاسوبيا، وذلك من اجل عمل التحاليل الإحصائية المطلوبة واستخراج النتائج.

تصميمات الدراسة لنظرية التعميم

تم تحليل البيانات باستخدام تصميم أحادي البعد (one-facet design) متقاطع كليا (fully Crossed): اشتمل على بعدين عشوائيين متقاطعين هما الأفراد (p) والمهمات (t) ، ويرمز لهذا التصميم بالرمز (pxt) ويتصف هذا التصميم بأنه أحادي البعد لأنه يشتمل على بعد واحد محتمل لتباين الخطأ، وهو المهمات وتجدر الإشارة إلى إن الطلاب لا يمكن اعتبارهما بعدا للقياس، وإنما موضوعا للقياس(Brnnen,١٩٩٢)، والذي يشير إلى التباين الحقيقي الذي يعكس الفروق بين الطلاب في اختبار القدرة، وبالتالي فهم لا يمثلون مصدرا لخطأ القياس، ووصف التصميم بأنه متقاطع كليا، لان كل الطلاب يجيبون على كل مهمة من مهمات الاختبار، ويشتمل بعد الطلاب على (٣٠٠) مستوى وبعد المهمات على (٣٠) مستوى، واعتبرت عينة الطلاب والمهمات عشوائية غير محدودة، اختيرت من نطاق واسع من الطلاب، والمهمات التي يمكن اختيارها، أو إعدادها من طرف الباحث، كما يوضح جدول(١)، مصادر تباين الخطأ الراجعة إلى أبعاد القياس التي تعرف بمكونات التباين.

جدول (١) صيغ تقدير مكونات التباين للتصميم أحادي البعد (pxt)

مصدر التباين	مكون التباين	تقديرات مكونات التباين
الطالب (p)	$\sigma^2 p$	$\frac{MSp - MSpt}{nt}$
المهمة (t)	$\sigma^2 t$	$\frac{MSt - MSpt}{np}$
التفاعل (pt)	$\sigma^2 pt$	$\sigma^2 pt = MS_{Pt}$

حيث يشير MS إلى متوسط المربعات(Mean of scores) المحصلة من جدول تحليل التباين، وn تشير إلى عدد مستويات عينة البعد، حيث يشير n_p إلى عدد الطلاب، و n_t إلى عدد المهمات، من اجل تقدير معاملات التعميم يتم تقدير تباين خطأ القياس النسبي، وتباين خطأ القياس المطلق وذلك باستخدام الصيغ (١)، (٢):

١) تباين الخطأ النسبي:

$$\sigma^2(\delta) = \frac{\sigma^2(pt)}{nt}$$

٢) تباين الخطأ المطلق:

$$\sigma^2(\Delta) = \frac{\sigma^2(t)}{nt} + \frac{\sigma^2(pt)}{nt}$$

وتشير $\sigma 2(\delta)$ إلى تباين الخطأ النسبي و $\sigma 2(\Delta)$ إلى تباين الخطأ المطلق و ويشير كل رمز من رموز البسط في كل من الصيغتين إلى مكونات التباين،حيث ترمز σ^2_t إلى تباين المهمة، و σ^2_{pt}إلى تباين تفاعل طالب مع مهمة، وnt إلى عدد المهمات، وبناء على قيم تباين الخطأ النسبي، وقيم تباين الخطأ المطلق، يمكن تقدير معاملات الثبات النسبي، والمطلق من خلال المعادلات: (٣)، (٤). (Brennan, 2001; Shavelson &Webb, 1991).

معامل الثبات النسبي :

$$\hat{\rho}2 = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(\delta)} \text{ ----------(3)}$$

معامل الثبات المطلق:

$$\phi = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(\Delta)} \text{ ----------(4)}$$

والتصميم ثنائي البعد(Tow-facet design): يعتبر من أكثر التصميمات ثنائية البعد استخداما في القياسات النفسية والتربوية، وأكثر تحديدا لتأثيرات مصادر الخطأ في وضعيات القياس، فمثلا التصميم (pxtxr) يفترض ثلاثة أبعاد متقاطعة الطلاب (p) مع المهام (t) ومع المقدرين (r)، ومن الناحية الإحصائية يتم تقدير أحجام مصادر تباين الخطأ الراجعة إلى أبعاد القياس التي تعرف أيضا بمكونات التباين (variance of copponents) باستخدام متوسطات المربعات الناتجة عن تحليل التباين كما يوضحه جدول (٢)،(Bernnan , ١٩٩٢).

جدول (٢)صيغ تقدير مكونات التباين في دراسة التعميم للتصميم ثنائي البعد (p x t x r)

مصدر التباين	مكون التباين	تقديرات التباين
الطالب (p)	$\sigma^2 p$	$\frac{MSp - MSpt - MSpr + MSptr}{nt\,nr}$
المهمة (t)	$\sigma^2 t$	$\frac{MSt - MStr - MSpt + MSptr}{np\,nt}$
المقدر(r)	$\sigma^2 r$	$\frac{MSr - MStr - MSpr + MSptr}{np\,nt}$
تفاعل طالب-مهم (pt)	σ^2_{pt}	$\frac{MSpt - MSptr}{nr}$
تفاعل طالب- مقدر(pr)	$\sigma^2 pr$	$\frac{MSpr - MSptr}{nt}$
تفاعل مهمة –مقدر(tr)	$\sigma^2 tr$	$\frac{MStr - MSptr}{np}$
تفاعل طالب-مهمة – مقدر (ptr)	$\sigma^2 ptr$	$\sigma^2 ptr = MSptr$

حيث يشير MS إلى متوسط المربعات (Mean of squares) التي يتم الحصول عليها من جدول تحليل التباين، وتشير n إلى عدد مستويات البعد، n_p تشير عدد الطلاب، nt تشير إلى عدد المهمات، nr تشير إلى عدد المقدرين. ومن اجل تقدير معاملات التعميم يتم تقدير تباين خطأ القياس النسبي، وتباين خطأ القياس المطلق وذلك باستخدام الصيغ (٥)، (٦):

تباين الخطأ النسبي

$$\hat{\sigma}2(\delta) = \frac{\hat{\sigma}^2 ptr}{nt\,nr} + \frac{\hat{\sigma}2\,pr}{nr} + \frac{\hat{\sigma}2\,pt}{nt}$$

تباين الخطأ المطلق

$$\hat{\sigma}^2(\Delta) = \frac{\hat{\sigma}^2 t}{nt} + \frac{\hat{\sigma}^2 r}{nr} + \frac{\hat{\sigma}^2 tr}{nt\,nr} + \frac{\hat{\sigma}^2 pt}{nt} + \frac{\hat{\sigma}^2 pr}{nr} + \frac{\hat{\sigma}^2 ptr}{nt\,nr}$$

وتشير($\sigma^2(\delta)$ إلى تباين الخطأ وتشير($\sigma^2(\Delta)$ إلى تباين الخطأ المطلق، و كل رمز من رموز البسط في كل من الصيغتين إلى مكونات التباين،حيث ترمز(σ^2_t) تباين المهمة، و(σ_r^2) تباين المقدر، و(σ^2_{tr}) تباين تفاعل مهمة مع مقد، و(σ^2_{pt}) تباين تفاعل طالب مع مهمة (σ^2_{pr}) تفاعل طالب مع مقد، ($\sigma^2 ptr$) تفاعل طالب ومهمة مع المقدر، وأما رموز المقام في الصيغتين تشير إلى عدد مستويات الأبعاد حيث (n_t) إلى عدد المهمات، و(n_r) إلى عدد المقدرين. وبناء على قيمة التباين الخطأ النسبي، وقيمة تباين الخطأ المطلق، تم تقدير معامل التعميم النسبي $\hat{\rho}2$ ، ومعامل التعميم المطلق Φ من خلال المعادلات(٧)، (٨).

معامل الثبات النسبي:

$$\hat{\rho}2 = \frac{\sigma^2(p)}{\sigma^2(p)+\sigma^2(\delta)} \text{ ----------(7)}$$

معامل الثبات المطلق

$$\phi = \frac{\sigma^2(p)}{\sigma^2(p)+\sigma^2(\Delta)} \text{ ----------(8)}$$

المعالجة الإحصائية

من اجل الإجابة على أسئلة الدراسة تم استخدام المعالجات الإحصائية التالية:

أولا: إيجاد معاملات الثبات لاختبار القدرة الرياضية بصورتيه: ثنائية التدريج ومتعددة التدريج، وفق النظرية الكلاسيكية في القياس بالطرق التالية (ألفا، راجو، ثيتا(θ))، باستخدام برنامج(spss) وبرنامج (microsoft Excel) والمعادلات اللازمة.

ثانيا: إيجاد معاملات الثبات لاختبار القدرة الرياضية بصورتيه: ثنائية التدريج ومتعددة التدريج، وفق نظرية التعميم من خلال برمجية (EdeG) التي تعتمد على تحليل التباين وتحليلات التعميم، صممت من قبل(Cardinet)، بهدف انجاز تحليلات التعميم(Cardinet,Sandra&Pini,٢٠١٠). حيث تمكننا من تحديد اكبر مصادر التباين تأثيرا على ملاحظات القياس، من خلال تحليل البيانات للتصميم الأحادي البعد لاختبار الثنائي التدريج (p x t)، (حيث p ترمز إلى الأفراد، و t ترمز إلى مهمات الاختبار)، والتصميم الثنائي البعد لاختبار المتعدد التدريج (p x r x t)، (حيث p ترمز إلى الأفراد، t ترمز مهمات الاختبار، r مقيمين).

ثالثا: استخدام اختبار كوكرون cocron بلغة R، واختبار signal / noise، للكشف عن دلالة الفروق في معاملات الثبات وفقا لشكل الاختبار.

نتائج الدراسة :

الإجابة عن السؤال الأول: معاملات الثبات المقدرة لاختبار القدرة في الرياضيات حسب نظرية الاختبار الكلاسيكي .

تم تحليل البيانات باستخدام برنامج (spss) ، وبرنامج (Microsoft excel)، والمعادلات المتعلقة بحساب معاملات الثبات لاختبار القدرة في الرياضيات بصورتيه: ثنائية التدريج، ومتعددة التدريج، وفق نظرية الاختبار الكلاسيكي، و حساب معامل الثبات بالطرق التالية (ألفا، راجو، ثيتا(θ)) ولحساب معامل الثبات (θ)، باستخدام برنامج (spss) من خلال استخدام أسلوب تحليل المكونات الأساسية) principal components analysis) ويبين الجدول٣ ملخصا لمعاملات الثبات:

جدول (٣) معاملات الثبات وفق نظرية الاختبار الكلاسيكي

معامل الثبات	معامل كرونباخ (α)	معامل (θ)	معامل ثبات راجو(Rjo)
الاختبار الثنائي التدريج	٠,٨٦٦	٠,٨٥٣	٠,٨٦٢
الاختبار المتعدد التدريج	٠,٩٤٥	٠,٩٢	٠,٩٣

الإجابة عن السؤال الثاني: معاملات الثبات المقدرة لاختبار القدرة في الرياضيات نظرية التعميم .

تم تقدير مصادر تباين الخطأ ومعاملات التعميم النسبية، والمطلقة للتصميم أحادي البعد (p x t)، والجدولان (٤) و(٥) يلخصان النتائج.

جدول (٤) تحليل التباين للتصميم المتقاطع كليا (p x t)

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	العشوائية	المختلطة	المصححة	النسبة%	الخطأ المعياري
طالب (p)	٤٤٧,٤٦٦٧	٢٩٩	١,٤٩٦٥	٠,٠٤٣٢	٠,٠٤٣٢	٠,٠٤٣٢	١٧,٣	٠,٠٠٤١
مهمة(t)	٦١,٨٩٣٣	٢٩	٢,١٣٤٣	٠,٠٠٦٤	٠,٠٠٦٤	٠,٠٠٦٤	٢,٦	٠,٠٠١٨
الباقي(طالب Xمهمة)(pt)	١٧٤٠,٢٤٠٠	٨٦٧١	٠,٢٠٠٧	٠,٢٠٠٧	٠,٢٠٠٧	٠,٢٠٠٧	٨٠,٢	٠٠٣٠.
المجموع	٢٢٤٩,٦٠٠	٨٩٩					١٠٠%	

في هذا التصميم تم الاعتماد على تصميم قياس يعتبر الطلاب(p) كبعد لتباين الدرجة الشاملة والمهمة(t) كبعد لتباين خطأ القياس، والجدول (٥) يوضح نسب تباين الخطأ النسبي ونسب تباين الخطأ المطلق لكل بعد من أبعاد القياس والتفاعل فيما بينهم .

جدول (٥) تحليل التعميم لتصميم القياس (P/T)

مصدر التباين	تباين التمييز	تباين الخطأ النسبي	نسبة تباين الخطأ النسبي %	تباين الخطأ المطلق	نسبة تباين الخطأ المطلق %
طالب(p)	٠,٠٤٣٢	---------	---------	---------	---------
مهمة (t)	---------	---------	---------	٠,٠٠٠٢	٣,١
طالب-مهمة(pt)	---------	٠,٠٠٦٧	١٠٠	٠,٠٠٦٧	٩٦,٩
مجموع التباينات	٠,٠٤٣٢	٠,٠٠٦٧	١٠٠%	٠,٠٠٦٩	١٠٠%
الانحراف المعياري	٠,٢٠٧٨	الخطأ المعياري النسبي	٠,٠٨١٨	الخطأ المعياري المطلق	٠,٠٨٣١
معامل التعميم النسبي		٠,٨٧			
معامل التعميم المطلق		٠,٨٦			

يبين الجدول رقم (٥) أن معامل التعميم لنسبي (٠,٨٧)، ومعامل التعميم المطلق(٠,٨٦)، وتعتبر قيم هذه المعاملات من الثبات مقبولة لاتخاذ قرارات نسبية لمقارنة أداء الطلاب مع زملائهم، أو قرارات مطلقة مقارنة أداء الطالب بمحك خارجي، ويبين جدول(٥) أن أكبر مصدر تباين للخطأ النسبي والمطلق راجع لتفاعل طالب ومهمة بنسبة (١٠٠%) من تباين الخطأ النسبي، و(٩٦,٩%) من تباين الخطأ المطلق، وان مصدر تباين الخطأ المطلق الذي يأتي بعد تفاعل طالب مع مهمة هو مهمات الاختبار بنسبة (٣,١ %) من تباين الخطأ المطلق. و تم تقدير مصادر تباين الخطأ ومعاملات التعميم النسبية والمطلقة للتصميم ثنائي البعد(p x t x r)، كما موضحة في الجدول (٦)، والذي يتكون من سبعة مكونات للتباين تعود إلى الطلاب: تباين الدرجة الشاملة(p)، والمقدرين(r)، والمهمات(t)، وتفاعل طلب مع المهمة(pt)، وتفاعل طالب مع المقدر(pr)، وتفاعل مقدر مع المهمة (rt)، والباقي الذي يعكس تفاعل طالب مع المهمة مع المقدر(ptr).

جدول (٦) تحليل التباين للتصميم المتقاطع كليا "طالب X مهمة X مقدر "(p x t x r)

مصدر التباين	مجموع المربعات	درجات الحرية	متوسط المربعات	العشوائية	المختلطة	المصححة	النسبة%	الخطأ المعياري
طالب (p)	١٣٩٠٦,٦٣٤٢	٢٩٩	٤٦,٥١٠٥	٠,٤٨٨٦	٠,٤٨٨٦	٠,٤٨٨٦	٣٢,٩	٠,٠٤٢١
مقدر (r)	١٤,٧٧٦٩	٢	٧,٣٨٨٤	٠,٠٠٠٥	٠,٠٠٠٥	٠,٠٠٠٥	٠,٠	٠,٠٠٠٦
مهمة (t)	٣٣٤٨,٣٧٤٢	٢٩	١١٥,٤٦١٢	٠,١٢٢٧	٠,١٢٢٧	٠,١٢٢٧	٨,٢	٠,٠٣٢٦
طالب -مقدر) (pr)	٣٠,٤٠٠٩	٥٩٨	٠,٠٥٠٨	٠,٠٠٠٤	٠,٠٠٠٤	٠,٠٠٠٤	٠,٠	٠,٠٠٠١
((طالب- مهمة)(pt)	٢١٨٧٢,٩٣٦٩	٨٦٧١	٢,٥٢٢٥	٠,٨٢٧٩	٠,٨٢٧٩	٠,٨٢٧٩	٥٥,٧	٠,٠١٢٨
(مقدر – مهمة) (rt)	١٤٧,٧٧٤٢	٥٨	٢,٥٤٧٨	٠,٠٠٨٤	٠,٠٠٨٤	٠,٠٠٨٤	٠,٦	٠,٠٠١٦
(طالب-مقدر –مهمة)(prt)	٦٧٢,٣٨١٣	١٧٣٤٢	٠,٠٣٨٨	٠,٠٣٨٨	٠,٠٣٨٨	٠,٠٣٨٨	٢,٦	٠,٠٠٠٤
المجموع	٣٩٩٩٣,٢٧٨٧	٢٦٩٩٩					%١٠٠	

في هذا التصميم تم الاعتماد على تصميم قياس يعتبر الطلاب(p) كبعد لتباين الدرجة الشاملة، والمهمة (t)، والمقدر(r) كأبعاد لتباين خطا القياس، والجدول (٧) يوضح نسب تباين الخطأ النسبي، ونسب تباين الخطأ المطلق لكل بعد من أبعاد القياس، والتفاعل فيما بينهم .

جدول (٧) تحليل التعميم لتصميم القياس (P/RT)

مصدر التباين	تباين التمييز	تباين الخطأ النسبي	نسبة تباين الخطأ النسبي %	تباين الخطأ المطلق	نسبة تباين الخطأ المطلق %
طالب(p)	٠,٤٨٨٦	.....	.....	.....	.....
مقدر (r)	----------	.....	.....	٠,٠٠٠٢	٠,٥
مقدر(t)	----------	.....	.....	٠,٠٠٤١	١٢,٦
طالب-مقدر(pr)	----------	٠,٠٠٠١	٠,٥	٠,٠٠٠١	٠,٤
طالب-مهمة(pt)	----------	٠,٠٢٧٦	٩٨,٠	٠,٠٢٧٦	٨٤,٩
مهمة- مقدر(rt)	----------	.....	.....	٠,٠٠٠١	٠,٣
طالب-مهمة-مقدر(prt)	----------	٠,٠٠٠٤	١,٥	٠,٠٠٠٤	١,٣
مجموع التباينات	٠,٤٨٨٦	٠,٠٢٨١	١٠٠%	٠,٠٣٢٥	١٠٠%
الانحراف المعياري	٠,٦٩٩٠	الخطأ المعياري النسبي	٠,٠,١٦٧٨	الخطأ المعياري المطلق	٠,١٨٠٣
معامل التعميم النسبي			٠,٩٥		
معامل التعميم المطلق			٠,٩٤		

يبين الجدول رقم (٧) أن معامل التعميم لنسبي (٠,٩٥)، ومعامل التعميم المطلق(٠,٩٤) وتعتبر قيم هذه المعاملات من الثبات مقبولة، ومرتفعة لاتخاذ قرارات نسبية لمقارنة أداء الطلاب مع زملائهم، أو قرارات مطلقة مقارنة أداء الطالب بمحك خارجي، ويبين جدول(٧) أن اكبر مصدر تباين للخطأ النسبي والمطلق راجع لتفاعل طالب ومهمة بنسبة (٩٨%) من تباين الخطأ النسبي، و (٨٤,٩%) من تباين الخطأ المطلق، وأن ثاني اكبر مصدر تباين الخطأ المطلق الذي يأتي بعد تفاعل طالب مع مهمة هو درجة صعوبة مهمات الاختبار بنسبة (١٢,٦ %) من تباين الخطأ المطلق.

ثالثا -النتائج المتعلقة بالسؤال الثالث: المقارنة بين معاملات الثبات باختلاف نوع الفقرات

الجدولان (٨) و(٩) يلخصان نتائج المقارنة بين معاملات الثبات المقدرة لصورتي اختبار القدرة، وفقا لنظرية الاختبار الكلاسيكي ونظرية التعميم باستخدام (Cocron) واستخدام Signal / noise .

جدول (٨) مقارنة بين معاملات الثبات باستخدامcocron

معامل الثبات	ثنائي التدريج	متعدد التدريج	عدد أفراد العينة	عدد الفقرات	قيمة Chsiq	معامل الارتباط البيني	درجات الحرية	فترة الثقة للثنائي التدريج	فترة الثقة للمتعدد التدريج	مستوى الدلالة	P - value	القرار
α	٠,٨٦٦	٠,٩٤٥	٣٠٠	٣٠	٧٩,٨٩٣٧	٠,٧٥	١	٠,٨٦٦٦ - ٠,٩٠٣٨	٠,٩٣٥٧- ٠,٩٥٣٦	٠,٠٥	٠,٠٠٠	رفض
θ	٠,٨٥٣	٠,٩٢	٣٠٠	٣٠	٥٦,٦٧٩٩	٠,٧٥	١	٠,٨٢٨٠ – ٠,٨٧٥٩	٠,٩٠٦٤ – ٠,٩٣٢٥	٠,٠٥	٠,٠٠٠	رفض
Raju	٠,٨٦٢	٠,٩٣	٣٠٠	٣٠	٦٩,٨٣٢٤	٠,٧٥	١	٠,٨٣٨٥ – ٠,٨٨٣٥	٠,٩١٨١ – ٠,٩٤٠٩	٠,٠٥	٠,٠٠٠	رفض

معامل الثبات	ثنائي التدريج	متعدد التدريج	عدد أفراد العينة	عدد الفقرات	قيمة Chsiq	معامل الارتباط البيني	درجات الحرية	فترة الثقة للثنائي التدريج	فترة الثقة للمتعدد التدريج	مستوى الدلالة	P - value	القرار
G	٠,٨٧	٠,٩٥	٣٠٠	٣٠	١٣١,٧٣٠٨	٠,٧٥	١	٠,٨٤٧٩ – ٠,٨٩٠٣	٠,٩٤١٥ – ٠,٩٥٧٨	٠,٠٥	٠,٠٠٠	رفض
φ	٠,٨٦	٠,٩٤	٣٠٠	٣٠	١٠٥,٧٧٧٩	٠,٧٥	١	٠,٨٣٦٢ – ٠,٨٨١٨	٠,٩٢٩٨ – ٠,٩٤٩٤	٠,٠٥	٠,٠٠٠	رفض

جدول (٩) يوضح المقارنة من خلال نسبة اليقين إلى الشك

	α	θ	Raju	G	φ
معامل ثبات المتعدد	٠,٨٦٦	٠,٨٥٣	٠,٨٦٢	٠,٨٧	٠,٨٦
نسبة اليقين إلى الشك (للمتعدد)	٦,٤٦	٥,٨	٦,٢٥	٦,٦٩	٦,١٤
معامل ثبات لإجابة المحددة	٠,٩٤٥	٠,٩٢	٠,٩٣	,٩٥	٠,٩٤
نسبة اليقين إلى الشك(للمحددة)	١٧,١٨	١١,٥	١٣,٢٩	١٩	١٥,٦٧

يوضح الجدول (٨) نتائج اختبار الفروق بين معاملي ثبات مرتبطين، لاختبار القدرة الرياضية وفقا لصورتي الاختبار: ثنائية التدريج ومتعددة التدريج، حيث تشير النتائج إلى وجود فروق ذات دلالة إحصائية بين معاملات الثبات المقدرة، باختلاف طريقة التقدير لصالح الاختبار المتعدد التدريج، كما وبين الجدول (٩) نتائج المقارنة من خلال نسبة اليقين إلى الشك Signal/noise، فكلما زادت نسبة القين زاد معامل الثبات وقلت نسبة الشك، إي كلما زادت نسبة اليقين إلى الشك كلما زاد معامل الثبات، ونلاحظ أن النسبة للفقرات المتعددة التدريج أعلى منها لثنائية التدريج، وتشير النتائج إلى فروق بين معاملات الثبات لاختبار القدرة في الرياضيات وفقا لصورتي الاختبار لصالح الفقرات المتعددة التدريج.

مناقشة النتائج :

النتائج المتعلقة بالسؤال الأول: أشارت النتائج المتعلقة بتقدير الثبات لاختبار القدرة الرياضية بصورتيه، وفق نظرية الاختبار الكلاسيكي إلى قيما مرتفعة لمعاملات الثبات، وهي مؤشر على أن هناك اتساقا داخليا عاليا، وتوافقا في علامات الطلبة في صورتي اختبار القدرة الرياضية، وقد أظهرت النتائج اختلافا في قيم معاملات الثبات بالطرائق الثلاثة، وتوصلت النتائج إلى أن معاملات الثبات للاختبار المتعدد التدريج كانت أعلى من معاملات الثبات المقدرة للاختبار الثنائي التدريج، وقد اتفقت هذه النتيجة مع دراسات(الأسمر،٢٠٠٧؛ (Brideman,١٩٩٤).التي أشارت في نتائجها إلى أن معاملات الثبات للفقرات المتعددة التدريج أعلى من الثنائية التدريج، وربما يعود ذلك إلى مصادر الخطأ في مفردات الثنائي التدريج أكثر منها في مفردات المتعدد التدريج، وعلى رأس تلك المصادر التخمين العشوائي .

النتائج المتعلقة بالسؤال الثاني: أشارت النتائج المتعلقة بتقدير الثبات لاختبار القدرة في الرياضيات بصورتيه، وفق نظرية التعميم من خلال نتائج تحليل التصميم الأحادي والتصميم الثنائي إلى أن أكثر مصادر تباين الخطأ تأثيرا على معاملات ثبات درجات الطلاب في اختبار القدرة الرياضية، راجع إلى تفاعل طالب ومهمة، وتباين مهمة، وتتفق هذه النتائج مع دراسات (١٩٩٨ , Mcbee& Barners ; ٢٠١٠ , guler& gelbal ; ٢٠٠٨ , Bain; فاروق، ٢٠١٦)، والتي بينت أن أكثر مصادر تباين الخطأ تأثيرا على معاملات ثبات درجات الطلاب، راجع إلى تفاعل طالب ومهمة، وتباين مهمة، واختلفت مع دراسة (Lee & kantor,٢٠٠٧) والتي تشير إلى أن اكبر مصدر لتباين الخطأ راجع إلى تفاعل طالب مع مهمة ومع المقدر الممزوج بالأخطاء العشوائية، وسبب ارتفاع مكون تباين طالب مع مهمة إلى التباين في متوسط أداء الطلاب من مهمة إلى أخرى، ويرجع ارتفاع مكون تباين المهمة إلى الاختلاف في درجة صعوبة المهمات، وأشارت النتائج إلى أن معاملات الثبات النسبية، والمطلقة للاختبار المتعدد التدريج كانت أعلى من معاملات الثبات للاختبار الثنائي التدريج، وفق التصاميم المستخدمة في الدراسة، ويمكن تفسير قيم الثبات المرتفعة إلى أن عدد فقرات الاختبار كانت كبيرة، وهذا يتفق مع دراسات(٢٠٠٧ , Lee & kantor ; ٢٠١٠ ,Guler & Gelbal ; ٢٠٠٨ ,Bain) من حيث بلوغ الاختبارات مستويات مقبولة من الثبات، والتي بينت أن زيادة عدد الفقرات له الأثر الأكبر على قيم الثبات، بدلا من زيادة عدد المقدرين، وتختلف النتائج مع بعض الدراسات (١٩٩٨ , Mcbee & barens ; ٢٠٠٧, Nie et al) التي كشفت عن ضعف في ثبات الاختبارات، والذي يعود إلى قلة عدد الفقرات في الاختبار.

النتائج المتعلقة بالسؤال الثالث : أشارت نتائج اختبار كوكرون cocron، و نتائج اختبار signal/ratio إلى وجود فروق ذات دلالة إحصائية بين معاملات الثبات لصالح الاختبار المتعدد التدريج، وقد تعزى هذه النتيجة إلى متغير الدراسة (صورة الاختبار) والى طبيعة الاختبار، وما يتمتع من خصائص سيكومترية جيدة، وهذا يتفق مع دراسات (الأسمر،٢٠٠٧; Bridgeman,١٩٩٤)، وبينت نتائج الدراسة على الرغم من تقارب معاملات ثبات ألفا مع معاملات التعميم لصورتي الاختبار، اختلاف قيم معاملات الثبات المقدرة وفق النظرية الكلاسيكية في القياس، حيث كان معامل ألفا الأعلى، ثم يليه معامل راجو، ثم معامل(θ)، وهذا يدعم أفضلية نظرية التعميم في تقدير معامل الثبات، وذلك لان نظرية التعميم تتعامل مع جميع مصادر تباين الخطأ، بينما نظرية الاختبار الكلاسيكي تتعامل مع مصدر تباين خطأ واحد وفقا لطريقة تقدير الثبات، وقد اتفقت الدراسة مع دراسات (الثبيتي، ١٩٩٨;١٩٩١ ,Eason; ٢٠١٠ , guler& gelbal)، التي بينت أفضلية نظرية التعميم.

التوصيات :

بناء على نتائج الدراسة يوصي الباحث:

١. توعية معلمين الرياضيات باستخدام الفقرات متعددة التدريج على نطاق واسع في اختبارات القدرة الرياضية وذلك بسبب الفعالية العالية للفقرات المتعددة التدريج في قياس القدرة الرياضية للطلبة.

٢. إجراء دراسات مماثلة لتقدير الثبات في حالة اختبارات القدرة التي تتضمن أكثر من نوعين من الفقرات وفق نظريتي القياس الكلاسيكية والتعميم، حيث اقتصرت هذه الدراسة على الاختبارات التي تتضمن نوعين فقط من الفقرات.

٣. إجراء دراسة مماثلة حول تقدير الثبات ومقارنة معاملات الثبات للاختبارات بفقرات ثنائية ومتعددة التدريج وفق نظريات القياس: الكلاسيكية ونظرية التعميم والنظرية الحديثة في القياس.

قائمة المراجع العربية :

- الأسمر، مصطفى .(٢٠٠٧) . الخصائص السيكومترية لصورتي اختبار في الرياضيات فقرات أحداهما ثنائية التدريج والأخرى متعددة التدريج وفق نظريتي القياس التقليدية والحديثة . أطروحة دكتوراه ، جامعة عمان العربية .
- الثبيتي، علي. (١٩٩٨). ثبات أدوات القياس في العلوم الإنسانية بين النظرية التقليدية للقياس ونظرية التعميم. المركز العربي للتعليم والتنمية.
- الحربي، خليل والحربي، عيد. (٢٠١٧). مؤشرات الثبات باستخدام نظرية التعميم ومؤشرات صدق البناء لمقياس موهبة الإبداع. مجلة جامعة طيبة للعلوم التربوية. ١٢ (٣)، ٤٢٥-٤٤١.
- الذنيبات، سجى احمد. (٢٠١٣). اثر اختلاف شكل الاختبار على الخصائص السيكومترية للاختبار، رسالة ماجستير منشورة، جامعة مؤته، الاردن.
- الشايب، عبد الحافظ . (٢٠٠٧). تقدير ثبات علامات عينة من المواد في جامعة آل البيت ، مجلة جامعة دمشق، ٢٣(٢)، ٢٥٥-٢٧١.
- علام ، صلاح الدين محمود. (٢٠٠٢). القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة. القاهرة: دار الفكر العربي.
- علاونة، معزوز جابر. (٢٠١٦). اثر حجم العينة ونمط الفقرة وجنس الطلبة على الخصائص السيكومترية لاختبار تحصيلي في الرياضيات لدى طلبة الحادي عشر في المدارس الحكومية الفلسطينية. مجلة جامعة المدينة العالمية ماليزيا. ع(١٧). ٤٢٦-٤٧٣
- عليان ، ريم محمد . (٢٠٠٤). تقدير الثبات للعلامات المدرسية في مدينة اربد الكبرى في الفصل الدراسي الأول ٢٠٠٢/٢٠٠٣ م . رسالة ماجستير، بجامعة اليرموك
- العمري ، حسان .(٢٠١٨). المقارنة بين ثلاث طرائق في تقدير ثبات الاختبارات المركبة التي تتضمن نوعين من الفقرات
- (ألفا، ألف الطبقي، راجو). مجلة اتحاد الجامعات العربية للتربية وعلم النفس.٦(٢).٨٥-١٠٢
- عودة، احمد . (٢٠١٠). القياس و التقويم في العملية التدريسية. إربد، الأردن: دار الأمل للنشر و التوزيع .
- فاروق، طباع. (٢٠١٦) . تقييم نموذج إمكانية التعميم لاحتيار كفاءات وفق الوضعيات المركبة. أطروحة دكتوراه، جامعة محمد ليمن دباغين ، سطيف٢ ،الجزائر
- الكحلوت، احمد اسماعيل.(٢٠٠٢) . مقارنة الخصائص السيكومترية لكل من اختبارات الاختيار من متعدد واختبارات التكميل. مجلة مركز البحوث. جامعة قطر. (٢٢). ١٢٧-١٥٣
- كروكر ، ليندا ، والجينا ، جيمز. (٢٠٠٩) . مدخل إلى نظرية القياس التقليدية والمعاصرة (زينات دعنا ، مترجم). دار الفكر : عمان.
- Allal, L., &Cardinet, J. (1997). Generalizability theory. In J.P. Keeres (Ed.), Educational research, methodology, and measurement: Aninternational handbook (2nd, pp. 737- 741). Cambridge, United Kindom: Cambridge University.
- Allen , M. ,& Yen , W.(1979). Introduction to measurement theory .Brooks/Colepulishing company : California .
- Anastasia. A. (1988). *Psychology testing* (6th ed). New York: Macmillan
- Brennan, R.L. (2001). Generalizability theory. New York: Springer-verlag .
- Bridgeman, B.(1994). Acomparison of quantitative questions in open-ended and multiple choice format. *Journal of educational measurement*,٢٩(٣)، ٢٥٣-٢٧١.
- Cardinet, J., Johnson, S., &Pini, G. (2010). Applying generalizability theory using EduG. New York,NY: Routledge .
- Crocker, L., &Algina, J. (1986). Introduction to classical and moderntest theory. Belmont, CA: Wadsworth Group/ Thomson Learning.
- Cronbach, L. J., &Gleser, G. C. (1964). The signal/noise ratio in the comparison of reliability coefficients. *Educational and Psychological Measurement*, 24(3), 467-480.
- *Cronbach, L. J., Gleser, G. C., Nanda, H. &Rajaratnam, N. (1972). The Dependability of Behavioral Measurements. New York: Wiley.*
- Eason, S. (1991). Why Generalizability Theory yields better results than classical test theory: A primer with concrete examples. *Advances in educational research*, 1. 83 –98.
- Guler, N., &Gelbal, S. (2010). Studying reliability of open-ended mathematics items according to the Generalizability Theory. *Educational Sciences: Theory & Practice*, 10 (2), 1011-1019.

- Lee, Y. W. & Kantor, R. (2007). Evaluating prototype tasks and alternative rating schemes for a new ESL writing test through G-theory. *International Journal of Testing*, 7(4), 353-385.
- Rajaratnam, N. (1960). Reliability formulas for independent decision data when reliability data are matched. *Psychometrika*, 25, 261-271
- *Shavelson, R. J. & Webb, N. M. (1991). Generalizability theory: Aprimer (Vol. 1). Sage.*
- Shavelson, R. J., & Webb, N. M. (1981). Generalizability theory: 1973–1980. *British Journal of Mathematical and Statistical Psychology*, 34(2), 133-166.
- Strube, M. J. (2002). Reliability and generalizability theory. In L.G. grimm& P.R. Yarnold (Eds.), Reading and understanding more multivariate statistics (pp. 23-66). Washington, DC: American Psychological Association.
- Thompson, B &Growley, G. (1994). When classical measurement theory is insufficient and Generalizability Theory is essential. Paper presented at the annual meeting of the western psychological association. Kailue-Kone, April 30(ERIC No. ED 377218).
- Webb, M., Schlackman, J. &Sugrue, B. (2000). The dependability and interchangeability of assessment methods in science. *Applied Measurement in Education*, 13(3), 277-301.
- *Webb, N.M., Rowley, G. L., &Shavelson, R. J. (1988). Using generalizability* theory in counseling and development. *Measurementand Evaluation in Counseling and Development*, 21, 81- 90.