

Estimating Reliability of Tests with Dichotomous and Polytomous Items According to Classical Measurement Theory and Generalizability Theory: A Comparative Study

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

Abdal hadi Al-Gharaibh^{1*}, Ahmad bani Ahmad¹.

¹Yarmouk University, Irbid, Jordan.

ARTICLE INFO

Article history:

Received 25 Aug 2021

Accepted 05 Dec 2021

Published 01 Jul 2023

*Corresponding author:

Yarmouk University, Irbid, Jordan.

Email: abed1973jor@yahoo.com.

Abstract

This study aimed to compare the reliability coefficients of tests with dichotomous and polytomous items according to Classical Test Theory and Generalizability Theory. To achieve this, a compound mathematics ability test was developed in two formats: dichotomous and polytomous, using five estimation methods: Alpha (α), Theta (θ), Raju (G), and Phi (φ). Each test format comprised 30 items designed to assess students' abilities in numbers, algebra, geometry, and statistics, with a sample of 300 students.

Reliability coefficients for the dichotomous items, estimated using Classical Theory, were as follows: $\alpha = 0.87$, $\theta = 0.85$, and Raju = 0.86. For Generalizability Theory, the coefficients were: G = 0.87 and $\varphi = 0.86$. In contrast, reliability coefficients for the polytomous items, according to Classical Theory, were: $\alpha = 0.95$, $\theta = 0.92$, and Raju = 0.93; for Generalizability Theory, they were G = 0.95 and $\varphi = 0.94$.

Comparison results using the Cochran test, and the signal-to-noise ratio test indicated a significant difference at $\alpha = 0.05$, favoring the polytomous format for reliability. Additionally, the findings demonstrated a preference for Generalizability Theory over Classical Theory in estimating reliability coefficients.

Keywords: Classical Test Theory, Generalizability Theory, Reliability, Raju Equation, Signal-to-Noise Ratio.

الملخص

هدفت هذه الدراسة إلى مقارنة معاملات الثبات في الاختبارات بفقرات ثنائية التدرج، ومتعدد التدرج وفق النظرية الكلاسيكية في القياس ونظرية التعميم، ولتحقيق ذلك، فقد تم بناء اختبار قدرة مركب في الرياضيات بصورتين: ثنائية التدرج، ومتعددة التدرج، باستخدام خمسة أساليب تقدير هي: (الف) (α)، (ث) (θ)، راجو، (G)، (φ)، ولتحقيق أهداف الدراسة تم تطبيق الاختبار بصورتيه، وكل اختبار مكون من (30) فقرة لقياس قدرة الطلبة الرياضية في مجال الأعداد والجبر والهندسة والإحصاء، على عينة من (300) طالب وطالبة. وقد كانت معاملات الثبات للتدرج الثنائي والمقدرة وفق النظرية

الكلاسيكية على النحو الآتي ($\alpha=0.87$, $\theta=0.86$, $Raju=0.86$)، ووفق نظرية التعميم ($G=0.87$, $p=0.86$). وكانت المعاملات المقابلة للتدرج المتعدد وفق الكلاسيكية ($\alpha=0.95$, $\theta=0.92$, $Raju=0.93$)، ووفق نظرية التعميم ($G=0.95$, $p=0.94$). وأظهرت نتائج المقارنة بين المعاملات وفق اختبار Cochran واختبار نسبة اليقين إلى الشك signal/noise ratio إلى وجود فرق ذي دلالة إحصائية عند مستوى الدلالة ($\alpha=0.05$) بين قيم معاملي ثبات اختبار القدرة بصورتيه لصالح التدرج المتعدد، كما كشفت النتائج عن أفضلية نظرية التعميم في تقدير معاملات الثبات مقارنة بالمعاملات المستخلصة وفق النظرية الكلاسيكية.

الكلمات المفتاحية: نظرية الاختبار الكلاسيكية، نظرية التعميم، الثبات، معادلة راجو، نسبة اليقين إلى الشك.

١. المقدمة

١،١ الإطار النظري والدراسات السابقة

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 آراء 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 السيكميترية، وتحديدًا 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 جوهريًا بين هذه الأحكام، وقد تطرح الأدبيات بعض المقترحات يستند عليها الباحثون في الخروج من مأزق تحديد الحد الأدنى، وقد يكفي البعض من الباحثين ومطوري أدوات القياس بعرض ما توصلوا إليه من خصائص سيكميترية تاركين للمستفيدين من هذه الأدوات أن يقرروا أو يحددوا خياراتهم. 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 تقريبًا إذا أعيد تطبيق نفس الاختبار على نفس المجموعة من الأفراد، وهذا يعني قلة تأثير عوامل الصدفة العشوائية على نتائج الاختبار، وقد عرفت أنستازي (Anastasi, 1982) الثبات، بأنه مدى دلالة تباين درجات الأفراد 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 دورًا أساسيًا، 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 إحصائيًا 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 بفواصل 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 نظرًا لأنها لا تميز بين أخطاء القياس، وإنما تعطي قيمة تقديرية 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 فأوجدوا ما يسمى بنظرية التعميم.

وتعد نظرية التعميم امتداداً لنظرية الاختبار (CTT) من خلال إجراء تحليل التباين (ANOVA) لقضايا القياس، فهي تقدم إطاراً مفاهيمياً وإحصائياً موسعاً لنظرية الاختبار الكلاسيكية، وتعد نموذجاً يمكنه التعامل مع مصادر متعددة للخطأ، مما يسمح بتقدير دقة القياسات، وتقديم طرق فعالة في تحسين إجراءات القياس المستقبلية (Brennan, 2001).

وتقدم نظرية التعميم عدداً من المزايا مقارنة بالنظرية الكلاسيكية للاختبار في تصميم إجراءات القياس: حيث يمكنها دمج مصادر متعددة للخطأ، وتقييم في نفس الوقت: ثبات الاختبار-أعادة الاختبار، ثبات الاتساق الداخلي والصدق التقاربي 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, 2008).

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 نظراً 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 آخر، فقد تعددت أساليب تقدير الثبات بدءاً 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).

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

حيث: r_{AB} معامل ثبات التجزئة النصفية، و r معامل الارتباط بين نصفي الاختبار، لكن معادلة سييرمان- براون تقوم على افتراض أن يتحقق شرط التوازي بين نصفي الاختبار والمتمثل بافتراض تساوي التباينات الحقيقية وتساوي تباينات الأخطاء على جزأي الاختبار، وبالتالي تساوي التباينات الظاهرية.

وكون افتراض التوازي صعب تحقيقه، فقد قدمت العديد من المعادلات الأخرى لحساب الثبات باستخدام التجزئة النصفية بتوفر شرط أقل تشدداً 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, 1945)، 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).

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

ومع أن شرط التكافؤ بالضرورة أقل تشدداً من شرط التوازي إلا أن تحقيقه يبقى أمراً صعباً، إلى أن قدم راجو (١٩٧٧): معادلة لتقدير ثبات التجانس الداخلي للعلامة الناتجة من قياسات متشكلة (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 امتداداً لطريقة كرونباخ ألفا، تتعامل مع التشاكل ولا تتعامل مع التباين المشترك في حساب الوزن النسبي للجزء، وهذه المعادلة تستخدم فيما كان الاختبار مكوناً من ثلاثة أجزاء متشكلة فأكثر معلومة الأوزان، ويتم حساب الثبات وفق معادلة (٣).

$$r_{xx} = \frac{1}{1 - \sum \lambda_i^2} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right) \quad (3)$$

حيث تمثل r_{xx} معامل الثبات، λ_i الوزن النسبي للجزء i ، σ_i^2 تباين العلامات في الجزء i ، σ_x^2 تباين العلامات الكلي ومن طرق الثبات الأخرى: معامل ثبات (theta)، اقترح (armor, 1947) صيغة لحساب معامل الثبات، حيث يدل N على عدد الفقرات، λ_i على الجذر الكامن الأول المستخرج بتحليل طريقة المكونات الرئيسية ويتم تقديره وفق معادلة (٤).

$$\theta = (N / N - 1)(1 - 1 / \lambda_i) \text{-----} (4)$$

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 عشوائياً، 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, 1964).

وتقدم 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 إعادة الاختبار) لعمل استنتاجات حول الأنواع الأخرى من الثبات مثل ثبات الاتساق الداخلي (Web Rowley & Shavelson, 1988). بينما مفهوم الثبات في نظرية التعميم يعبر عن دقة التعميم من الدرجة الملاحظة للمفحوص على الاختبار مثلاً أو أي أداة قياس أخرى إلى متوسط الدرجة التي سوف يحصل عليها المفحوص تحت كل الشروط الممكنة (Bernnan, 2001). 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 (Ala & Cardinet, 1997; Shavelson & Webb, 1991; Strube, 2002).

و في إطار نظرية التعميم، تعتبر الملاحظات (مثلاً 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 ضوءها جمع الملاحظات، أو الحصول على القياسات، ومجموعة الشروط التي يتم إجراء الملاحظات لها تسمى (عالم الملاحظة المقبولة the universe of admissible observation)، والذي يقوم بتحديد كل باحث، ويشير عالم الملاحظات المقبولة إلى كل الأبعاد التي يمكن اختبارها 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 أوجهه (Facets)، ومستويات هذه الأوجه شروط (Conditions)، (علام، ٢٠٠٢).

ويتضمن 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. ونظراً لوجود نوعين من

الخطأ: الخطأ النسبي والخطأ المطلق، يوجد في نظرية التعميم معاملان للثبات: معامل الثبات النسبي ومعامل الثبات المطلق (Bernnan, 2001; shvelson & Webb, 1991).

١,١,١ الدراسات السابقة

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

وتعد الدراسة التي أجراها أيسون (Eason, 1991) 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 مستخدماً طرق نظرية الاختبار الكلاسيكية ونظرية التعميم، وتوصل الباحث إلى أن الدرجات تكون أكثر استقراراً وثباتاً باستخدام الطرق التي تعتمد على نظرية التعميم في تقدير الثبات وذلك لأن هذه النظرية تأخذ جميع المصادر لتباين الخطأ.

وكما قام ثومبسون وجرولي (Thompson & Growely, 1994) بدراسة هدفت للمقارنة بين النظرية الكلاسيكية للاختبار ونظرية التعميم، أكد فيها أنه لا يزال الاعتماد مستمراً على استخدام النظرية الكلاسيكية في حساب معامل الثبات، مما أدى إلى عدم تماسك النتائج، نظراً 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 برديمان (Brideman, 1994) 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 الاختبار من متعدد.

وفي دراسة الثبتي (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 أبرز 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 بإجرائها مكبي وبارنس (Mcbee & Barnes, 1998) والتي هدفت لتقييم التحصيل في الرياضيات لدى عينة مكونة من (١٠١) من طلاب الصف الثامن، قام الباحثان بتطوير أربع مهام أداء، مهمتين أكثر تشابهاً إلى حد كبير لقياس مجال حل المشكلات المعقدة، تم تقدير أدائهم بالاعتماد على نموذج تقييم المهام، اعتمد الباحثان على تصميم ملاحظة (فرد × مهمة × مقدار × فترة) لتحليل البيانات، أظهرت نتائج التحليلات أن معامل الثبات بين التقديرات (٠,٨٠) بينما كانت معاملات ثبات الإعادة أقل بكثير تتراوح بين (٠,٢٨ إلى ٠,٥٤)، أظهرت نتائج دراسة ارتفاعاً في تباين طالب مع المهمة، وتفاعل طالب مع مهمة مع مقدار الممزوج بالخطأ العشوائي وانخفاضاً في مصادر تباين الخطأ بين المهمات الأكثر تشابهاً، وكشفت نتائج دراسات القرار أن عدد المهام المطلوبة للوصول إلى مستويات مقبولة من التعميم سيكون مرتفعاً بشكل محظور، حتى باستخدام مهام متشابهة للغاية.

وتعد دراسة لي وكننور (Lee & Kantor, 2006) التي هدفت فحص تأثيرات المهمات والمقشرين على درجات الكتابة وتأثير عدد المهمات والمقشرين وتصميمات التقدير على ثبات الاختبار بواسطة نظرية التعميم، طبق الباحثان (٦) مهمات مدمجة في الكتابة (الاستماع والقراءة)، على عينة مكونة (٤٤٨) طالباً من طلبة اللغة الإنجليزية في خمسة بلدان (أمريكا، كندا، هونج كونغ، المكسيك وتايوان)، وتم تقييم المهمات من قبل (٣) مقدرين، وتم تحليل البيانات من خلال

تصميمين الأول ثنائي البعد متقاطع جزئياً " فرد: مقدار × مهمة "، والثاني ثنائي البعد متقاطع كلياً " فرد × مهمة × مقدار"، وقد كشفت نتائج الدراسة أن معاملات التعميم كانت مقبولة، وأن أكبر مصدر تباين للخطأ كان بسبب تفاعل الفرد مع المهمة مع المقدّر الممزوج بالخطأ غير العشوائي، وتفاعل الفرد مع المهمة، في حين أظهرت انخفاضاً في مصادر التباين الأخرى، بالإضافة إلى إن زيادة عدد المهمات أكثر تأثيراً من زيادة عدد المقدرين في رفع معاملات الثبات بمعاملات تعميم نسبية ومطلقة بلغت (٠,٨٥) و(٠,٨١) على التوالي.

وبيّنت دراسة ني و يو ولو (Nie et al., 2007) 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 × مهمة × مقدار) حيث توصلت الدراسة إلى أن معامل التعميم النسبي (٠,٧٦) ومعامل التعميم المطلق (٠,٧٢)، أظهرت النتائج أن زيادة عدد المهام له تأثير أكبر على معاملات التعميم أكثر من زيادة عدد المقيمين.

وأجرى باين (Bain, 2008) دراسة هدفت إلى إعداد اختبار مشترك لتقييم كفاءات الرياضيات وفق نموذج التعميم، طبق الباحث الاختبار المكون من (١٥) مهمة على عينة من (٤٣) طالباً من طلاب الصف التاسع بجنيف، وتم تحليل البيانات وفق تصميم متقاطع جزئياً (طالب × مهمة: مدرسة × مجموعة المستوى)، أظهرت نتائج الدراسة عن ارتفاعاً في معاملات الثبات النسبية والمطلقة، حيث بلغت (٠,٩٠)، (٠,٨٩)، وأشارت النتائج إلى ارتفاع في مصدر تباين تفاعل طالب مع المهمة المتداخل مع المدرسة والمجموعة.

وتعد الدراسة التي قام بها جولر وجيلبال (Guler & gelbal, 2010) 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-1999) على عينة (٢٠٣) طلاب في فصل الربيع ٢٠٠٧، وقد كشفت نتائج الدراسة أن معامل الاتساق الداخلي للدرجات بلغ (٠,٩٢)، وأن معامل التعميم للدرجات الرياضيات (٠,٩٢) وكان معامل فأي (٠,٩٠)، استأثر مكون التباين للمقيمين بنسبة ٢,١٪ من التباين الكلي وفقاً 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 اتساقاً في نتائجهم.

وفي مجال الدراسات العربية لم يجد الباحث أي 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 السيكمترية لصورتي اختبار في الرياضيات فقرات إحداها ثنائية التدرّج والأخرى متعددة التدرّج وفق نظريتي القياس التقليدية والحديثة، وتمثلت عينة الدراسة من (٦٣٠) طالب وطالبة، حيث طبق الاختبار الثنائي التدرّج على (٦٣٠) طالب وطالبة، بينما طبق الاختبار المتعدد التدرّج على (٣٠٠)، توصلت الدراسة إلى أن معامل الثبات للاختبار الثنائي التدرّج بطريقة كرونباخ ألفا (٠,٨٩)، وبمعادلة سييرمان براون (٠,٩١)، بينما بلغ معامل الثبات للاختبار المتعدد التدرّج بطريقة كرونباخ ألفا (٠,٩٣) وبمعادلة سييرمان براون (٠,٩٥)، وبطريقة الصور المتكافئة فقد بلغ معامل الارتباط بين الصورتين (٠,٨٥)، أي أن معامل الثبات للمتعدد التدرّج أعلى من معامل الثبات للثنائي التدرّج.

وفي الدراسات العربية التي استخدم فيها معامل راجو لتقدير الثبات، فقد أجرت عليان (٢٠٠٤) دراسة لتقدير الثبات للعلامات المدرسية في مدينة أريد 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، ألفا الطبقي، راجو) وذلك باستخدام ٨٤ اختباراً من الاختبارات النهائية بمشاركة عينة من (٣٠) طالب وطالبة للاستجابة على كل اختبار، بمتوسط (١٣) فقرة لكل اختبار وتكونت جميعها من نوعين من الفقرات الاختبار

من متعدد والمقالية، أشارت نتائج الدراسة إلى إن الوسط الحسابي لتقديرات الثبات بلغ (٠,٧١)، (٠,٧٩)، (٠,٥٣) حسب معادلة ألفا وإلغا الطبقى وراجو على الترتيب ، وأشارت نتائج التحليل التباين واختبار كأي تربيع إلى وجود فروق دالة إحصائية 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 العاملي التوكيدي لدراسة البناء التكويني للمقياس، وبعد تحليل نتائج الدراسة تبين أنَّ مقياس موهبة للإبداع يتمتع بجودة عالية، وتوصلت نتائج الدراسة إلى أن معاملات كرونباخ ألفا 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 2010) 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 التدرج ومتعددة التدرج وفق نظريتي الاختبار الكلاسيكية في القياس ونظرية التعميم .

٢. 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 التحصيلي، 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 للاختبارات أصبحت غير كافية في تقدير ثبات الدرجات وذلك لأنها لا تأخذ بعين الاعتبار 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 التدرج، ومتعددة التدرج 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 أحدهما اختبار اختيار من متعدد (ثنائي التدرج)، والآخر يتكون من مهمات محددة الإجابة (متعدد التدرج) وبذلك يتوقع أن تجيب الدراسة عن الأسئلة التالية.

٢,١ أسئلة الدراسة

١. ما هي معاملات الثبات المقدرة لاختبار القدرة الرياضية حسب نظرية الاختبار الكلاسيكي؟
٢. ما معاملات الثبات المقدرة لاختبار القدرة الرياضية حسب نظرية التعميم؟
٣. هل تختلف معاملات الثبات باختلاف نوع أسئلة الاختبار؟

٣. 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 التدرج، ومتعددة التدرج، باستخدام خمسة أساليب تقدير هي: الفا (α)، ثيتا (θ)، راجو، معاملات التعميم (G)، (φ)، وتبين مدى التوافق بين معاملات الثبات المقدرة وفقاً 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 أجل الوقوف على نطاق القوة والضعف التي يعاني منها الطلاب في مادة الرياضيات من أجل 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 النظريتين، كما أنها ستفتح المجال أمام الباحثين في إجراء المزيد من الدراسات للمقارنة بين أشكال أخرى من الاختبارات ومعرفة أثرها على الخصائص السيكومترية للاختبار.

٤. 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 للطلاب 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 التدرج (Dichotomous)

اختبار تكون الاستجابة لفقراته متغيراً ثنائياً وفيها تعطى الإجابة عن الفقرة العلامة (١) عندما تكون الإجابة صحيحة، وتعطى الإجابة العلامة (صفر) إذا كانت خطأ.

الاختبار المتعدد التدرج (Polytomous)

اختبار تكون الاستجابة عن فقراته متدرجة، إذ 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 يراعى هذا النوع من الأسئلة المعرفة الجزئية.

٥. محددات الدراسة

- اقتصر 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 التدرج، ومتعددة التدرج.

٦. الطريقة والإجراءات

٦,١ منهج الدراسة

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

٦,٢ 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).

٦,٣ عينة الدراسة

تم اختيار عينة الدراسة بالطريقة الطبقيّة 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.

٦,٤ 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 التدرج) لكل مفردة (٤) بدائل يتكون من (٣٠) فقرة والنوع الثاني يتكون من (٣٠) مهمة محددة الإجابة (متعدد التدرج) في الرياضيات وقد تم تحديد مادة الاختبار بناء على اختبارات التيمس TIMS 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، واقترح ما يروونه مناسبا من تعديلات، وذلك من خلال الإجابة على عن استبيان مرفق مع كل اختبار تم إعداده لتحقيق هذا الغرض، وفي ضوء آراء 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 بصورته النهائية والبالغ عددها (٣٠)، فبعد جلوس أفراد الدراسة للاختبار، وزع الاختبار على الطلاب بشكله حيث بدأ بأسئلة الاختبار الثنائي التدرج، وبعد الانتهاء من إجابة هذه الأسئلة أعطي الطلبة النوع الآخر، الاختبار المتعدد التدرج، وأشرف الباحث على إجراءات تطبيق الاختبار بمساعدة معلمو الرياضيات في المدارس التي طبق بها الاختبار.

❖ تصحيح الاختبار

قام الباحث بإعداد نموذج تصحيح للاختبار الثنائي التدرج، وذلك بإعطاء رقم (١) للإجابة الصحيحة وإعطاء رقم (٠) للإجابة الخاطئة، وإعداد نموذج تصحيح للاختبار المتعدد التدرج، 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، بمساعدة (٣) من المقيمين لتقدير أداء الطلاب على الفقرات المتعددة التدرج وتم استخراج العلامات وتفرغ إجابات المفحوصين حاسوبياً، وذلك من أجل عمل التحاليل الإحصائية المطلوبة واستخراج النتائج.

❖ تصميمات الدراسة لنظرية التعميم

تم تحليل البيانات باستخدام تصميم أحادي البعد (one-facet design) متقاطع كلياً (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 وتجدد الإشارة إلى أن الطلاب لا يمكن اعتبارهما بعداً للقياس، وإنما موضوعاً للقياس (Brannen, 1992)، والذي يشير إلى التباين الحقيقي الذي يعكس الفروق بين الطلاب في اختبار القدرة، وبالتالي فهم لا يمثلون مصدراً لخطأ القياس، ووصف التصميم بأنه متقاطع كلياً، لأن 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، اختبرت 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)	$\sigma^2 p$	MSp- MSpt
المهمة (t)	$\sigma^2 t$	MSt-MSpt
التفاعل (pt)	$\sigma^2 pt$	$\sigma^2 pt = MS Pt$

حيث يشير MS إلى متوسط المربعات (Mean of scores) المحصلة من جدول تحليل التباين، و n تشير إلى عدد مستويات عينة البعد، حيث يشير n_p إلى عدد الطلاب، و n_t إلى عدد المهمات، من أجل 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 ويشير كل رمز من رموز البسط في كل من الصيغتين إلى مكونات التباين، حيث ترمز t^2 إلى تباين المهمة، و σ^2_{pt} إلى تباين تفاعل طالب مع مهمة، و n_t 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).

٣. معامل الثبات النسبي:

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

٤. معامل الثبات المطلق

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

والتصميم ثنائي البعد (Tow-facet design): يعتبر من أكثر التصميمات ثنائية البعد استخداماً في القياسات النفسية والتربوية، وأكثر تحديداً لتأثيرات مصادر الخطأ في وضعيات القياس، فمثلاً التصميم (p x t x r) 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 أيضاً بمكونات التباين (variance of components) باستخدام متوسطات المربعات الناتجة عن تحليل التباين كما يوضحه جدول (٢)، (Bernman, 1992).

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)	$\sigma^2 pt$	$\frac{MSpt - MSptr}{nr}$
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 أجل تقدير معاملات التعميم يتم تقدير تباين خطأ القياس النسبي، وتباين خطأ القياس المطلق وذلك باستخدام الصيغ (٥)، (٦):

٥. تباين الخطأ النسبي:

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

٦. 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، وكل رمز من رموز البسط في كل من الصيغتين إلى مكونات التباين، حيث ترمز σ^2_t تباين المهمة، و σ^2_r 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، و σ^2_{ptr} تفاعل طالب مع مهمة 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).

٧. معامل الثبات النسبي:

$$\hat{\rho}_2 = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(\delta)}$$

٨. معامل الثبات المطلق:

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

٦,٥ المعالجة الإحصائية

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

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

ثانياً: إيجاد معاملات الثبات لاختبار القدرة الرياضية بصورتيه: ثنائية التدرج ومتعددة التدرج، وفق نظرية التعميم من خلال برمجية (EdeG) التي تعتمد على تحليل التباين وتحليلات التعميم، صممت من قبل (Cardinet)، بهدف إنجاز تحليلات التعميم (Cardinet, Sandra & Pini, 2010). حيث تمكنا من تحديد أكبر مصادر التباين تأثيراً على ملاحظات القياس، من خلال تحليل البيانات للتصميم الأحادي البعد لاختبار الثنائي التدرج $(p \times t)$ ، (حيث p ترمز إلى الأفراد، و t ترمز إلى مهمات الاختبار)، والتصميم الثنائي البعد لاختبار المتعدد التدرج $(p \times r \times t)$ ، (حيث p ترمز إلى الأفراد، t ترمز مهمات الاختبار، r مقيمين).

ثالثاً: استخدام اختبار Cochran بلغة R، واختبار signal / noise، للكشف عن دلالة الفروق في معاملات الثبات وفقاً لشكل الاختبار.

٧. نتائج الدراسة ومناقشتها

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 التدرج، ومتعددة التدرج، وفق نظرية الاختبار الكلاسيكي، وحساب 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 وبين الجدول (٣) ملخصاً لمعاملات الثبات.

جدول (٣): معاملات الثبات وفق نظرية الاختبار الكلاسيكي

معامل الثبات	معامل كرونباخ (α)	معامل (θ)	معامل ثبات راجو (Rjo)
الاختبار الثنائي التدرج	٠,٨٦٦	٠,٨٥٣	٠,٨٦٢
الاختبار المتعدد التدرج	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 ($p \times t$)، والجدولان (٤) و (٥) يلخصان النتائج.

جدول (٤): تحليل التباين للتصميم المتقاطع كليا ($p \times t$)

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 x مهمة (pt)	١٧٤٠,٢٤٠٠	٨٦٧١	٠,٢٠٠٧	٠,٢٠٠٧		٠,٢٠٠٧	٨٠,٢	٠,٠٠٣٠
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، وإن مصدر تباين الخطأ المطلق الذي يأتي بعد تفاعل طالب مع مهمة هو مهمات الاختبار بنسبة (٣,١٪) من تباين الخطأ المطلق. وتم تقدير مصادر تباين الخطأ ومعاملات التعميم النسبية والمطلقة للتصميم ثنائي البعد ($p \times t \times r$)، كما موضحة في الجدول (٦)، والذي يتكون من سبع 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)	١٣٩٠,٦٣٤٢	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
مهمة (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
مقدر - مهمة (rt)				٠,٠٠٠١	٠,٣
طالب - مقدر - مهمة (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	٠,٠١٦٧٨	الخطأ المعياري المطلق	٠,١٨٠٣
معامل التعميم النسبي				٠,٩٥	
معامل التعميم المطلق				٠,٩٤	

يبين الجدول (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 أكبر مصدر تباين للخطأ النسبي والمطلق راجع لتفاعل طالب ومهمة بنسبة (٩٨٪) من تباين الخطأ النسبي، و (٨٤,٩٪) من تباين الخطأ المطلق، وأن ثاني أكبر 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 لصورتى اختبار القدرة، وفقا لنظرية الاختبار الكلاسيكي ونظرية التعميم باستخدام (اختبار Cochran) واستخدام Signal / noise.

جدول (٨): مقارنة بين معاملات الثبات باستخدام اختبار Cochran

معامل الثبات	ثنائي التدرج	متعدد التدرج	عدد أفراد العينة	عدد الفقرات	قيمة Chsiq	معامل الارتباط البيني	درجات الحرية	فترة الثقة للتدرج	فترة الثقة للمتعدد التدرج	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
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	٠,٩٥	٠,٩٤	نسبة اليقين إلى الشك (للمحددة)
١٧,١٨	١١,٥	١٣,٢٩	١٩	١٥,٦٧	

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

٧,١ 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 صورتى اختبار القدرة الرياضية، وقد أظهرت النتائج اختلافا في قيم معاملات الثبات بالطرائق الثلاثة، وتوصلت النتائج إلى أن معاملات الثبات للاختبار المتعدد التدرج كانت أعلى من معاملات الثبات المقدر للاختبار الثنائي التدرج، وقد اتفقت هذه النتيجة مع دراسات (الأسمر، ٢٠٠٧؛ Brideman, 1994). التي أشارت في نتائجها إلى أن معاملات الثبات للفقرات المتعددة التدرج أعلى من الثنائية التدرج، وربما يعود ذلك إلى مصادر الخطأ في مفردات الثنائي التدرج أكثر منها في مفردات المتعدد التدرج، 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 (Bain, 2008; guler & gelbal, 2010; Mcbee & Barners, 1998)؛ (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, 2007) والتي تشير إلى أن أكبر 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 التدرج كانت أعلى من معاملات الثبات للاختبار الثنائي التدرج، وفق التصميم المستخدمة في الدراسة، ويمكن تفسير قيم الثبات المرتفعة إلى أن عدد فقرات الاختبار كانت كبيرة، وهذا يتفق مع دراسات (Bain, 2008; Guler)

(Gelbal, 2010 ; Lee & Kantor, 2007) 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 (Nie et al., 2007 ; McBee & Barens, 1998) التي كشفت عن ضعف في ثبات الاختبارات، والذي يعود إلى قلة عدد الفقرات في الاختبار.

النتائج المتعلقة بالسؤال الثالث: أشارت نتائج اختبار Cochran، ونتائج اختبار signal/ratio إلى وجود فروق ذات دلالة إحصائية بين معاملات الثبات لصالح الاختبار المتعدد التدرج، وقد تعزى هذه النتيجة إلى متغير الدراسة (صورة الاختبار) وإلى طبيعة الاختبار، وما يتمتع من خصائص سيكومترية جيدة، وهذا يتفق مع دراسات (الأسمر، ٢٠٠٧؛ Bridgeman, 1994).

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 لأن 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؛ Eason, 1991؛ guler & gelbal, 2010)، التي بينت أفضلية نظرية التعميم.

٨. التوصيات

بناء على نتائج الدراسة يوصي الباحث:

١. توعية معلمين الرياضيات باستخدام الفقرات متعددة التدرج على نطاق واسع في اختبارات القدرة الرياضية وذلك بسبب الفعالية العالية للفقرات متعددة التدرج 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 اقتصرَت هذه الدراسة على الاختبارات التي تتضمن نوعين فقط من الفقرات.
٣. إجراء دراسة مماثلة حول تقدير الثبات ومقارنة معاملات الثبات للاختبارات بفقرات ثنائية ومتعددة التدرج وفق نظريات القياس: الكلاسيكية ونظرية التعميم والنظرية الحديثة في القياس.

بيان تضارب المصالح

يقر جميع المؤلفين أنه ليس لديهم أي تضارب في المصالح.

المراجع

- الأسمر، مصطفى. (٢٠٠٧). الخصائص السيكمترية لصورتي اختبار في الرياضيات فقرات أحدهما ثنائية التدرج والأخرى متعددة التدرج وفق نظريتي القياس التقليدية والحديثة [أطروحة دكتوراه غير منشورة]. جامعة عمان العربية.
- الثبيتي، علي. (١٩٩٨). ثبات أدوات القياس في العلوم الإنسانية بين النظرية التقليدية للقياس ونظرية التعميم. المركز العربي للتعليم والتنمية.
- الحري، خليل، والحري، عيد. (٢٠١٧). مؤشرات الثبات باستخدام نظرية التعميم ومؤشرات صدق البناء لمقياس موهبة الإبداع. مجلة جامعة طيبة للعلوم التربوية، ١٢ (٣)، ٤٢٥-٤٤١.
- الذنيبات، سجي أحمد. (٢٠١٣). أثر اختلاف شكل الاختبار على الخصائص السيكمترية للاختبار [رسالة ماجستير غير منشورة]. جامعة مؤتة، الأردن.
- الشايب، عبد الحافظ. (٢٠٠٧). تقدير ثبات علامات عينة من المواد في جامعة آل البيت. مجلة جامعة دمشق، ٢٣ (٢)، ٢٧١-٢٥٥.
- علام، صلاح الدين محمود. (٢٠٠٢). القياس والتقويم التربوي والنفسي أساسياته وتطبيقاته وتوجهاته المعاصرة. القاهرة، مصر: دار الفكر العربي.

علاونة، معزوز جابر. (٢٠١٦). أثر 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*، ١٧، ٤٧٣-٤٢٦.

عليان، ريم محمد. (٢٠٠٤). تقدير الثبات للعلامات المدرسية في مدينة إربد الكبرى في الفصل الدراسي الأول ٢٠٠٣/٢٠٠٢ م [رسالة ماجستير غير منشورة]. جامعة اليرموك، الأردن.

العمرى، حسان. (٢٠١٨). المقارنة بين ثلاث طرائق في تقدير ثبات الاختبارات المركبة التي تتضمن نوعين من الفقرات (ألف، ألف الطبعي، راجو). *مجلة اتحاد الجامعات العربية للتربية وعلم النفس*، ٦ (٢)، ٨٥-١٠٢.

عودة، أحمد. (٢٠١٠). *القياس والتقويم في العملية التدريسية*. إربد، الأردن: دار الأمل للنشر والتوزيع.

فاروق، طباع. (٢٠١٦). تقييم نموذج إمكانية التعميم لاختبار كفاءات وفق الوضعيات المركبة [أطروحة دكتوراه غير منشورة]. جامعة محمد ليمن دباغين، سطيف ٢، الجزائر.

الكحلوت، أحمد إسماعيل. (٢٠٠٢). مقارنة الخصائص السيكومترية لكل من اختبارات الاختيار من متعدد واختبارات التكميل. *مجلة مركز البحوث*، ٢٢، ١٢٧-١٥٣.

كروكر، ليندا، والجينا، جيمز. (٢٠٠٩). *مدخل إلى نظرية القياس التقليدية والمعاصرة* (زينات دعنا، مترجم). عمان، الأردن: دار الفكر.

References

- Allal, L., & Cardinet, J. (1997). Generalizability theory. In J. P. Keeres (Ed.), *Educational research, methodology, and measurement: An international handbook* (2nd ed., pp. 737–741). Cambridge, United Kingdom: Cambridge University Press.
- Allen, M., & Yen, W. (1979). *Introduction to measurement theory*. California: Cole Publishing Company.
- Anastasia, A. (1988). *Psychology testing* (6th ed.). New York: Macmillan.
- Brennan, R. L. (2001). *Generalizability theory*. New York: Springer-Verlag.
- Bridgeman, B. (1994). A comparison of quantitative questions in open-ended and multiple-choice format. *Journal of Educational Measurement*, 29(3), 253–271.
- 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 modern test 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: A primer* (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. 22–66). Washington, DC: American Psychological Association.
- Thompson, B., & Growley, G. (1994, April 30). When classical measurement theory is insufficient, and generalizability theory is essential [Presented paper]. The annual meeting of the Western Psychological Association, Kailua-Kona. (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. *Measurement and Evaluation in Counseling and Development*, 21, 81–90.