مجلة العلوم التربوية
جـامعـــة قطـــــر

تاريخ الاستلام: 2023/03/29 | تاريخ التحكيم: 2023/04/14 | تاريخ القبـول: 2023/05/26

OPEN ACCESS

تأثيـرُ حجـمِ العيِّنـة وتدريـجِ الاسـتجابةِ فـي تقديـراتِ مُعامـل ثبـاتِ ألفـا العـام للمحْرزي لأنواعِ الدَّرجات المحولة*

أحمد بن صالح بن علي السعدي
أخصائي إحصاء أول، وزارة التربية والتعليم–سلطنة عمان
ahmedsalehalsaadi8@gmail.com

يوسف عبد القادر أبو شندي[1]، راشد بن سيف المحرزي[2]

ملخص

هدفت الدراسة إلى فحص تأثير حجم العينة وتدريج الاستجابة في تقديرات معامل ثبات ألفا العام للمحرزي لأنواع مختلفة من الدرجات المحولة، باستخدام بيانات حقيقية في مقياس دولي لاتجاهات الطلبة نحو الرياضيات (TIMSS 2019)، وهو المطبَّق في عام 2019م. تكونت عينة الدراسة من 5117 طالبًا من طلبة الصف الرابع الأساسي في سلطنة عمان. ولتحقيق هدف الدراسة، تكوَّنت ثلاثة أنواع من البيانات من خلال تعديل عدد فئات تدريج الاستجابة في المقياس (تدريج رباعي، وهو التدريج الأصلي للمقياس – تدريج ثلاثي – تدريج ثنائي). وفي كل تدريج من تدريجات الاستجابة الثلاثة، جرى اختيار عينات فرعية بعدد 100 عشوائيا باستخدام ثلاثة أحجام مختلفة للعينة الفرعية (30 طالبًا، وتمثل الحجم الصغير للعينة – 200 طالب، وتمثل الحجم المتوسط للعينة – 1000 طالب، وتمثل الحجم الكبير للعينة)، وطُبِّق معامل ثبات ألفا العام للمحرزي على الدرجات المحولة الخمس لجميع البيانات السابقة، وقورِنت للإجابة عن أسئلة الدراسة باستخدام الإحصاءات الوصفية، وإيجاد نسب التحيز، وقيم جذر متوسطات مربع الخطأ (RMSE). توصلت نتائج الدراسة إلى درجة عالية في دقة تقديرات معامل ثبات ألفا العام للمحرزي في كل أنواع الدرجات المحولة؛ فقد كانت قيم تقديرات معامل الثبات لجميع لدرجات المحولة في العينات العشوائية قريبة جدا من معاملات ثبات العينة الرئيسة، وكانت نسب التحيز وقيم RMSE صغيرة وقريبة من الصفر، بما فيها حجم العينة الصغير (30). أوصت الدراسة بالتوسع في استخدام معامل ثبات ألفا العام لحساب ثبات جميع أنواع الدرجات المحولة في الدراسات البحثية ومختلِف أدوات التقييم.

الكلمات المفتاحية: معامل ثبات ألفا العام، حجم العينة، تدريج الاستجابة، نوع الدرجة المحولة، نسب التحيز

* البحث مستل من رسالة ماجستير بعنوان: «تأثير حجم العينة، وتدريج الاستجابة، ونوع الدرجة المحولة على تقديرات معاملي ثبات ألفا العام وبيتا العام للمحرزي»، للباحث أحمد بن صالح بن علي السعدي، في تخصص علم النفس (القياس والتقويم)، كلية التربية، جامعة السلطان قابوس، أجيزت في: 4 يناير 2023.

1 أستاذ مشارك، برنامج ماجستير القياس والتقويم، كلية التربية، جامعة السلطان قابوس–سلطنة عمان. yousefaaa@squ.edu.om
2 أستاذ، برنامج ماجستير القياس والتقويم، كلية التربية، جامعة السلطان قابوس–سلطنة عمان. mehrzi@squ.edu.om

للاقتباس: السعدي، أحمد وأبو شندي، يوسف والمحرزي، راشد. (2024). تأثيرُ حجمِ العيِّنة وتدريج الاستجابةِ في تقديراتِ مُعامل ثباتِ ألفا العام للمحْرزي لأنواعِ الدَّرجات المحولة، **مجلة العلوم التربوية، جامعة قطر، 24**(3)، ص101–130. https://doi.org/10.29117/jes.2024.0193

© 2024، السعدي وأبو شندي والمحرزي، الجهة المرخص لها: مجلة العلوم التربوية، دار نشر جامعة قطر. نُشرت هذه المقالة البحثية وفقًا لشروط Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0). تسمح هذه الرخصة بالاستخدام غير التجاري، وينبغي نسبة العمل إلى صاحبه، مع بيان أي تعديلات عليه. كما تتيح حرية نسخ، وتوزيع، ونقل العمل بأي شكل من الأشكال، أو بأية وسيلة، ومزجه وتحويله والبناء عليه، طالما يُنسب العمل الأصلي إلى المؤلف. https://creativecommons.org/licenses/by-nc/4.0

Journal of
Educational Sciences
Qatar University

Received: 29/03/2023 | Peer-reviewed: 14/04/2023 | Accepted: 26/05/2023

OPEN ACCESS

Effects of Sample Size and Response Rating Scales on Estimating Almehrizi Generalized Alpha Reliability Coefficient for Different Scaled Scores[1]

Ahmed bin Saleh bin Ali Al-Saadi
Senior Statistician, Ministry of Education–Sultanate of Oman
ahmedsalehalsaadi8@gmail.com

Youssef Abdel Qader Abu Shindi[2], Rashid bin Saif Al Mahrezi[3]

Abstract

The study aimed to examine the effect of sample size and response rating scales on estimating Almehrizi generalized alpha reliability coefficient for different scaled scores using real data for students' attitudes towards mathematics in TIMSS2019 for test administration in 2019. The sample consisted of 5117 male and female fourth-grade students in Sultanate of Oman. To achieve study objectives, three data sets were created through manipulating the response rating scales to obtain three types of response rating scales (four response categories which was the original scale – three response categories - two response categories). For each response rating scale, 100 subsamples were randomly selected using three sample sizes (30 for small sample size – 200 for medium sample size – 1000 for large sample size). In addition, five scaled scores were computed for the three types of rating scales: raw scores, standardized scores, percentile scores, and two TIMSS scaled scores. Almehrizi's generalized Alpha Coefficient was computed for each of these scaled scores for all data sets. Finally, these reliability coefficients were compared to answer research questions using descriptive statistics, bias and RMSE. Results revealed that the Almehrizi's Generalized Alpha Coefficients for all data sets showed high accuracy, as they were very close to the reliability coefficient values for the original data set for all scaled scores. Both bias and RMSE values were very close to zero indicating accurate estimation even for small sample size (30). The study recommended utilization of this coefficient to estimate reliability for different scaled scores in research and assessment tools.

Keywords: Generalized Alpha Reliability Coefficient, Sample size, Response rating scale, Bias ratio

1 The research is extracted from amaster's thesis entitled: "Effects of Sample Size, Response Scale, and Derived Scores on Estimating Almehrizi Generalized Alpha and Generalized Beta Reliability Coefficients," by Ahmed bin Saleh bin Ali Al-Saadi, in completion of the Program in Measurement and Evaluation, the College of Education, Sultan Qaboos University, awarded in 04/01/2023

2 Associate Professor, College of Education, Sultan Qaboos University–Sultanate of Oman. yousefaaa@squ.edu.om

3 Professor, College of Education, Sultan Qaboos University–Sultanate of Oman. mehrzi@squ.edu.om

Cite this article as: Al-Saadi, A.S.A.A., Abu Shindi, Y. A., & Al Mahrezi, R. S. (2024). Effects of Sample Size and Response Rating Scales on Estimating Almehrizi Generalized Alpha Reliability Coefficient for Different Scaled Scores. *Journal of Educational Sciences, Qatar University, 24*(3), pp. 101-130. https://doi.org/10.29117/jes.2024.0193

© 2024, Al-Saadi, A.S.A.A., Abu Shindi, Y. A., & Al Mahrezi, R. S., licensee, JES & QU Press. This article is published under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0), which permits non-commercial use of the material, appropriate credit, and indication if changes in the material were made. You can copy and redistribute the material in any medium or format as well as remix, transform, and build upon the material, provided the original work is properly cited. https://creativecommons.org/licenses/by-nc/4.0

مقدمة

يعدُّ الثبات من أهم الخصائص السيكومترية للاختبار، ويعدُّ سبيرمان براون Spearman-Brown مؤسس نظرية الثبات؛ فقد قدم عدة مفاهيمَ منها مفهوم الدرجة الحقيقية وخطأ القياس، ووفقًا للمفهوم الذي وضعه فإن الثبات وسيلة لمعرفة الارتباط الدقيق بين القيم الحقيقية، وذلك من خلال التحكم في التأثيرات التي تحدث للارتباط المشاهد، والناتجة عن الأخطاء العشوائية خلال عملية القياس، فُيعرف الثبات على أنه مربع الارتباط بين الدرجة الحقيقية والدرجة المشاهدة، وكذلك يُعرف على أنه نسبة تباين الدرجة الحقيقية للاختبار إلى تباينه الإجمالي (Streiner, 2003)

ويُعَّرف الثبات إحصائيًّا بأنه نسبة التباين الحقيقي إلى التباين الكلي؛ أي: كم من التباين الكلي في الدرجات يمكن أن يكون تباينًا حقيقيًّا، ويعبَّر عن هذه النسبة إحصائيًّا بمربع معامل الارتباط بين العلامة الحقيقية والعلامة الظاهرية. ويذكر أنه لا يمكن الاعتماد على هذه الطريقة في حساب الثبات؛ لأنها تقف عاجزة عن الوصول إلى العلامات الحقيقية؛ لذلك يؤكد على اللجوء إلى العلامة الظاهرية في تقدير الثبات؛ بمعنى الحصول على مؤشر إحصائي نستطيع من خلاله أن نحكم على دقة القياس، ويطلَق على هذا المؤشر اسم معامل الثبات Reliability Coefficient (عودة، 2010).

طُرق تقديرِ الثَّبات

هناك عدة طرق لتقدير معامل الثبات، وذلك بسبب ظهور بعض أخطاء القياس؛ لذلك وجب التنوع في معاملات الثبات؛ إذ إنَّ بعضها يظهر في بعض المعاملات والبعض الآخر لا يظهر فيها، ويمكن توضيح هذه المعاملات فيما يلي:

معاملُ ثبات الاستقرار Stability:

يُحسَب فيه الارتباط بين درجات المفحوصين على اختبار طُبِّق مرتين متتاليتين، ويطبَّق الاختبار في المرة الثانية بعد مرور فترة زمنية من تطبيقه في المرة الأولى، ويشترط أن تكون السمة المقاسة من هذا الاختبار صفة ثابتة لا تتغير بمرور الزمن؛ أي إنه لا يصلح استخدامه مثلًا في قياس القلق؛ فهي صفة ليست ثابتة، ويتأثر هذا المعامل بعدة عوامل كعامل التذكر، والتعلم إن طالت الفترة الزمنية، وظروف التطبيق (عودة، 2010).

معاملُ ثبات التكافؤ Equivalence:

تقوم فكرة هذا المعامل على إعداد صورتين متكافئتين للاختبار، ثم تطبيقه على المفحوصين في الفترة نفسِها أو بفاصل زمني قصير، ويشترط أن تكون الصورتان متكافئتين في الخصائص المتعلقة بعدد فقرات الاختبار، ومدى الصعوبة، والقدرة التميزية، ولهما المحتوى نفسُه الذي يقيسه السؤال، والتشابه في صياغة الأسئلة (أبو هلال وآخرون، 2020). ومن ثَم يصعب إعداد صورتين متكافئتين بهذه الشروط مما يسبب حدوث خطأ القياس (عودة، 2010).

معاملُ ثبات استقرار– تكافؤ Stability-Equivalence:

يحسَب من خلال إيجاد قيمة الارتباط بين درجات المفحوصين على صورتين متكافئتين لاختبار ما؛ إذ تُطبَّق الصورتان بشرط أن تكون مدة الفترة الزمنية بين التطبيقين طويلة نسبيًا، والهدف من هذا هو التقليل من أثر التذكر، وهذه الطريقة ينتج عنها قيمة ثبات أقل من الطريقتين السابقتين لأنها أكثر عرضة لخطأ القياس؛ فإنها تعدُّ طريقةً مركبةً تجمع بين طريقة حساب الثبات بمعامل الاستقرار ومعامل التكافؤ، ومن ثَم تتراكم بها الأخطاء (عودة، 2010).

معاملُ الاتِّساق الداخلي Internal Consistency:

وهو طريقة لحساب ثبات درجات الاختبار، ويتحقق ذلك من خلال حساب معامل الثبات بين نصفي الاختبار الذي طُبق مرةً واحدة، وذلك بعد تجزئته إحصائيًا إلى نصفين متكافئين، إما عشوائيًا أو تقسيم مفردات الاختبار على حسب الأعداد الفردية أو الزوجية، فالنصف الأول من المفردات يمثل الأعداد الفردية، والنصف الآخر يمثل الأعداد الزوجية، ومعامل الثبات بهذه الطريقة يسمى معامل الارتباط النصفي، ويحسَب بمعامل ارتباط بيرسون (Mehrens & Lehmann, 1978).

ويشير يانج وجرين (Yang & Green, 2011) إلى أن معامل ألفا دلالة على طول الأداة وتشبع البنود؛ فالقيمة المرتفعة لمعامل ألفا لا تعدُّ مؤشرًا للاتساق الداخلي أو التجانس؛ إذ إن طبيعة الاتساق الداخلي تعتمد على متوسط الارتباطات البينية للبنود وليس على عددها حسب ما يعتقد البعض.

معاملُ ثبات ألفا:

إنَّ المتتبع لتاريخ معادلة ألفا، سواءً من علماء أو باحثين في شتى مجالات البحث العلمي، يدرك أن معادلة ألفا ما هي إلا امتداد لمعادلات أخرى ومن بينها معادلة سبيرمان براون، التي كان من أهم افتراضاتها شرطُ التكافؤ بين نصفَي الاختبار وتساوي تباين الدرجة لكلا نصفَي الاختبار، ثم اقترح باحثون صيغًا أخرى لهذه المعادلة في حالة عدم تساوي تباينات الدرجة بين نصفي الاختبار، ومن بينهم (Flanagan, 1937; Mosier, 1994, Rulon, 1939)، ولكن جميع هذه الصيغ كان استخدامها صالحا لحساب الثبات بين نصفي الاختبار في حالة تجزئته.

ويذكر المحرزي (Almehrizi, 2022) أن معامل ألفا امتدادٌ لمعادلة كيودر وريتشاردسون (KR-20)؛ فهو أسلوب يُحسَب فيه تباين كل فقرة من فقرات الاختبار، ثم تُجمع هذه التباينات ويُحسب تباين الدرجة الكلية للاختبار، ويشار إليه بالصيغة 1-(مجموع تباين الفقرات/ تباين الدرجة الكلية) ومعادلته كما في المعادلة رقم (1):

$$\alpha = 1 - \frac{\sigma^2(\epsilon_x)}{\sigma^2(x)} = \frac{n}{n-1}\left(1 - \frac{\sum_i \sigma_i^2}{\sigma^2(x)}\right) \text{ ----------}$$

حيث إنَّ تباين الخطأ $\sigma^2(\epsilon_x)$ في نموذج تاو ويعطى كما في المعادلة (2):

$$\sigma^2(\epsilon_X) = \frac{n}{n-1}\left(\sum_i \sigma_i^2 - \frac{1}{n}\sigma^2(x)\right)$$ ----------

حيث σ_i^2 فيمثل تباين درجات السؤال الواحد.

وقارنت بعض الدراسات بين معامل ثبات ألفا كرونباخ ومعاملات ثبات أخرى، فتوصلت دراسة العمري (2018) إلى أن قيمة معامل ثبات ألفا كرونباخ أعلى من قيم معامل ثبات راجو، وبعض الدراسات توصلت إلى أن بعض معاملات الثبات أفضل من معامل ألفا كرونباخ تحت ظروف بحثية معينة، كدراسات كلٍّ من (Trizano-Hermosilla & Alvarado, 2016؛ كريش، 2018؛ زارع، 2021).

معاملُ ثبات ألفا العام للمحرزي:

طرح المحرزي (Almehrizi, 2021; 2022) معاملَي ألفا العام، وبيتا العام، ومن ثَم فحص دقة تقدير هذه المعاملات للثبات من خلال الدرجات المحولة، وتوصل إلى أن أعلى قيمة ثبات لكلا المعاملين في المقياسين لصالح الرتب المئينية مقارنة بالدرجات المحولة الأخرى، وأن أقل قيمة ثبات لصالح الدرجة المعيارية الاعتدالية والتساعيات. وبيَّنَ المحرزي (Almehrizi, 2022) من خلال بيانات حقيقية أن معامل ثبات ألفا العام $(\widehat{G\alpha})$ له نمط مشابه لمعامل ثبات بيتا العام $(\widehat{G\beta})$ بالنسبة إلى الدرجات المحولة، وأنه كذلك له قيم مختلفة في كل الدرجات المحولة، وأن هناك اختلافًا أيضًا في قيم معامل ثبات بيتا العام في كل درجة من الدرجات المحولة، وأكدت النتائج أن معامل ثبات ألفا العام $(\widehat{G\alpha})$ له قيم أعلى من قيم معامل ثبات بيتا العام $(\widehat{G\beta})$ في جميع الدرجات المحولة. ويُحسب معامل ثبات ألفا العام للدرجات المحولة (S) المقابلة للدرجات الخام (X) في الاختبار الذي يحتوي على عدد من الأسئلة (K)، المطبق على عينة بحجم (N) من خلال المعادلة الآتية:

$$\widehat{G\alpha} = 1 - \frac{\frac{N}{N-1}\frac{K}{K-1}\left[\hat{\epsilon}_p^2(S) + \hat{\epsilon}_i^2(S) - \hat{\epsilon}^2(S)\right]}{\hat{\sigma}^2(S)}$$ ----------

حيث تُحسب المقادير في المعادلة (3) من خلال:

$$\hat{\epsilon}_p^2(S) = \frac{1}{N}\sum_p\left[\sum_X S^2 f(X|\pi_p) - \left[\sum_X S f(X|\pi_p)\right]^2\right]$$ ----------

$$\hat{\epsilon}_i^2(S) = \sum_X S^2 f(X|\pi_i) - \left[\sum_X S f(X|\pi_i)\right]^2$$ ----------

$$\hat{\epsilon}^2(S) = \sum_X S^2 f(X|\pi) - \left[\sum_X S f(X|\pi)\right]^2$$ ----------

$$\hat{\sigma}^2(S) = \frac{N}{N-1}\left[\sum_X S^2 f(X) - \left[\sum_X S f(X)\right]^2\right]$$ ----------

ويمثل $f(X)$ التكرار النسبي للدرجات الخام في عينة الدراسة، ويُحسَب من خلال التوزيع التكراري للدرجات

الخام. وفي المقابل يمثل $f(X|\pi_p)$ التكرار النسبي المشروط للدرجات الخام باستخدام استجابات الطالب الواحد على جميع أسئلة الاختبار (π_p)، ويمثل $f(X|\pi_i)$ التكرار النسبي المشروط للدرجات الخام باستخدام استجابات جميع الطلبة على كل سؤال على حدة (π_i)، ويمثل $f(X|\pi)$ التكرار النسبي المشروط للدرجات الخام باستخدام استجابات جميع الطلبة على جميع أسئلة الاختبار مجتمعة (π). وتُحسَب التكرارات النسبية المشروطة الثلاثة من خلال تطبيق صيغة ارتدادية Recursion Formula قدَّمها لورد وونجيرسكي (Lord & Wingersky, 1984)، وتتطلب الصيغة الارتدادية حساب ثلاثة أنواع من التكرارات النسبية للاستجابات في أسئلة الاختبار: 1) التكرارات النسبية للاستجابات باستخدام استجابات الطالب الواحد على جميع الأسئلة (π_p)، 2) التكرارات النسبية للاستجابات باستخدام استجابات جميع الطلبة على كل سؤال على حدة (π_i)، 3) التكرارات النسبية للاستجابات باستخدام استجابات جميع الطلبة على جميع الأسئلة مجتمعة (π). ويمكن الرجوع الى الصيغة الارتدادية وطريقة تطبيقها في دراسة المحرزي (Almehrizi, 2022). ويجدر بالذكر أن معامل ثبات ألفا كرونباخ (Cronbach, 1951) حالة خاصة من معامل ثبات ألفا العام في حالة استخدام الدرجات الخام (X) محل الدرجات المحولة (S)، ويمكن إثباته رياضيا (Almehrizi, 2022) لتكون كالتالي:

$$\hat{\alpha} = 1 - \frac{\frac{N}{N-1}\frac{K}{K-1}\left[\hat{\epsilon}_p^2(X) + \hat{\epsilon}_i^2(X) - \hat{\epsilon}^2(X)\right]}{\hat{\sigma}^2(X)} = \frac{K}{K-1}\left[1 - \frac{\hat{\epsilon}_i^2(X)}{\hat{\sigma}^2(X)}\right] \text{----------}$$

حيث:

$$\hat{\epsilon}_i^2(X) = \frac{\sum_p \sum_i (X_{pi} - \bar{X}_i)^2}{N-1} \quad \text{----------}$$

$$\hat{\sigma}^2(X) = K^2 \frac{\sum_p (\bar{X}_p - \bar{X})^2}{N-1} \quad \text{----------}$$

العواملُ المؤثرة في معاملات الثبات

وتتأثر قيمة الثبات بعدة عوامل، منها ما يتعلق بأداة القياس، ومنها ما يتعلق بالمفحوصين. ومن العوامل التي تتعلق بأداة القياس طولُ الاختبار؛ فالاختبار الذي يتضمن بنودًا كثيرةً يحقق ثباتًا أعلى من الاختبار الذي يتضمن بنودًا قليلة، وكذلك أيضا طبيعةُ البنود؛ فالاختبارات الموضوعية تحقق ثباتًا أعلى من الاختبارات المقالية؛ فإنها لا تتأثر بذاتية المصحح (أبو زينة، 1992). ويذكر علام (2006) أن الثبات يتأثر كذلك بصعوبة مفردات الاختبار، فإذا كانت المفردات في غاية السهولة أو الصعوبة فإنه لا يمكن قياس الفروق الفردية.

وبالإضافة إلى هذه العوامل، فإن هناك عوامل أخرى تؤثر في تقدير معامل ثبات درجات الاختبار، وقد تناولتها العديد من الدراسات، كحجم العينة، وتدريج الاستجابة، ونوع الدرجة المحولة. فعلى سبيل الذكر، توجد دراسة علاونة (2016)، التي أشارت نتائجها إلى أن تقدير معامل الثبات يتناسب طرديًا مع حجم العينة،

ودراسةُ إسماعيل (2015)، التي أشارت نتائجها إلى أن معامل ألفا يزداد بتدريج الاستجابة، وأيضًا دراسة المحرزي(Almehrizi, 2022)، التي تناولت تقدير معامل الثبات للدرجات المحولة، ولقد تناولت هذه الدراسة الحالية هذه المتغيرات، وكان التطرق إليها على النحو الآتي:

حجمُ العينة:

يعدُّ حجم العينة من أهم العوامل التي تؤثر في الخصائص السيكومترية للاختبار، وخاصة التي تتعلق بالصدق والثبات؛ إذ يلجأ الباحثون في الدراسات الاستطلاعية إلى اختيار عينة صغيرة من المفحوصين، اعتقادًا منهم بأن ذلك يوفر لهم الوقت والجهد، من حيث جمع البيانات وتصحيحها ثم تحليلها للتحقق من الكفاءة السيكومترية للاختبار قيد التصميم، ولكن هذا اعتقاد خاطئ؛ فلا بد من أساس علمي يُرجَعُ إليه في تحديد حجم العينة المناسب والحصول على نتائج دقيقة (عبد السميع، 2017).

جذب تحديد حجم العينة أو عدد المفحوصين على الاختبار اهتمام العديد من الباحثين والمختصين في القياس النفسي والتربوي، ولم يكن وليد اليوم واختلفت الآراء فيه؛ فيتفق كلٌّ من (Kline & Bernstien, 1994; Segall, 1994) على ألَّا يقل حجم العينة عن 300 مفحوص، في حين أشار كارتر (Charter, 2003) إلى ألا يقل حجم العينة عن 400 مفحوص لحساب ثبات الاختبار، وبرر ذلك بأن هناك علاقة بين عدد المفحوصين وثبات الاختبار؛ فكلما قل عدد المفحوصين، أدى هذا إلى عدم استقرار معامل ثبات الاختبار، في حين أشار كروكر والجينا (Crocker & Algina, 1986) إلى ألا يقل عدد المفحوصين عن 200 مفحوص.

تدريجُ الاستجابة:

قدَّم ليكرت عام 1932 أسلوب التقدير الجمعي لقياس الاتجاهات والآراء، وهو يتضمن شطرين: الأول محتوى البند، والثاني تدريج الاستجابة التي من خلالها يختار المستجيب ما يراه معبرا عن رأيه، ورغم شيوعه واستخدامه فهناك عدم اتفاق على العدد المناسب لتدريج الاستجابة، مما يؤثر في الخصائص السيكومترية لأدوات القياس (Barnette, 2010)، ويؤكد ذلك دانر وآخرون (Danner et al., 2016) من أن تحديد تدريج الاستجابة من أكبر التحديات التي تعترض الباحثين المهتمين بتصميم أدوات القياس التربوية والنفسية، مما يستوجب المزيد من إجراء البحوث في هذا المجال.

ويؤكد كلٌّ من (الضوي، 2011؛ علي وحسن، 2011) أن قيمة معامل ألفا كان عاليا في الأعداد الفردية لتدريج الاستجابة مقارنة بالأعداد الزوجية.

وعلى النقيض، أكدت دراساتُ كلٍّ من (Choudhury & Bhattacharjee, 2014; Preston & colman, 2000; Zumbo et al., 2007) أن قيمة معامل الثبات تزداد بزيادة تدريج الاستجابة، في حين ترى دراسة (Lozano et al., 2008) أن العدد المناسب لتدريج الاستجابة هو من (4-7) فئات، ونادرًا ما يزداد معامل الثبات في التدريجات الأعلى

من التدريج السباعي، ويؤكد هذا واكيتا وآخرون (Wakita et al., 2012) من أن الباحثين يفضلون استخدام تدريج الاستجابة الذي يتراوح من (4-7) فئات.

ويؤكد مايديو أوليفرز وآخرون (Maydeu-Olivares et al., 2009) أنه لا يوجد اتفاق بشأن العدد المناسب من تدريج الاستجابة الذي بدوره يعطي أفضل الخصائص السيكومترية لأدوات القياس؛ فعملية اختيار العدد المناسب لتدريج الاستجابة ليست عملية عشوائية تجري تبعًا لرغبة الباحثين وأهوائهم، وإنما هي عملية مبنية على أسس علمية.

الدرجاتُ المحوَّلة:

هي نوع من معايير الاختبارات مرجعية الجماعة، وتعدُّ من الأساسيات التي يرتكز عليها علم القياس والتقويم؛ فإن الدرجات الخام يجب تحويلها إلى درجات معيارية حتى يكون لها معنى. والدرجات المحولة من أهم الوسائل التي يعتمد عليها علم القياس والتقويم في تقييم نتائج المفحوصين الذين طبق عليهم الاختبار؛ فهي تحدد المركز النسبي للفرد في توزيع معين، من أجل وصف أدائه بالنسبة إلى أقرانه في اختبار يقيس صفة ما (علام، 2006).

وتناولت الدراسة ثلاث درجات محولة من الدرجات غير الخطية، وهي:

1. **الدرجةُ المعيارية الاعتدالية** (Normalized Standard Scores): وهي درجة معيارية تتبع التوزيع الطبيعي بمتوسط حسابي يساوي صفرا، وانحراف معياري يساوي واحدا، ويتحقق الحصولُ عليها من خلال تحويل الدرجة الخام إلى رتب مئينية، ثم استخراج الدرجة المعيارية الاعتدالية من خلال جدول التوزيع الطبيعي الاعتدالي والمقابلة لكل رتبة مئينية (عودة، 2010).
2. **الرتبُ المئينية** (Percentile Ranks): وهي درجات معيارية شاع استخدامها بكثرة من أجل تفسير درجات الاختبارات؛ فمن خلالها تتحقق معرفةُ النسب المئوية لمجموعة من الأفراد حصلوا على درجات تقل درجاتهم عن درجة ما في أحد الاختبارات، فعلى سبيل المثال، إذا كانت الرتبة المئينية المقابلة للدرجة الخام 70 هي 90 فهذا يدل على أن 90% من مجموعة الأفراد تقل درجاتهم أو تساوي الدرجة 70، فلا يمكن تفسير الرتب المئينية إلا ضمن مجموعة مرجعية طُبِّق الاختبار عليهم (علام، 2006).
3. **التِّساعيات** (Stanine): هي درجات معيارية تتوزع فيها الدرجات في تسع فئات، وكل فئة يطلق عليها (تساعي)، وتتضمن هذه الفئات مجموعة من الدرجات أو الأفراد، ولا يخضع تقسيم هذه الفئات إلى قاعدة رياضية معينة، وإنما لاعتبارات حقيقية تبنى على أساس ظروف المؤسسة أو الدائرة التعليمية، والهدف من الدرجات، والنظام المتبع في التقويم لهذه المؤسسة أو الدائرة التعليمية (عودة، 2010).

مشكلةُ الدراسةِ وأسئلتُها

تجاهل المستخدمون قضية فحص تحقيق البيانات لافتراضات معامل ثبات ألفا قبل استخدامه؛ لذا سارع الباحثون لتقديم مقترحات وصيغ معادلات أخرى لحساب قيمة الثبات، ومن بين هذه المعادلات معادلة ألفا العام،

فهي تساعد الباحثين والمختصين في مجال البحث العلمي في كيفية إعداد مقياس، الذي يحتاج إلى تحديد تدريج الاستجابة وحجم العينة المناسبة لحساب الثبات؛ إذ لا يوجد اتفاق في تحديد هذا العدد الذي يعطي تقديرًا عاليًا للثبات، وهذا يرجع إلى عدم وجود خطوط إرشادية متفق عليها لتحديد تدريج الاستجابة وحجم العينة، وهذا ما ذكره واجترز وآخرون (Weijters et el., 2011) من أنَّ بعض الباحثين يحددون تدريج الاستجابة وحجم العينة في أثناء تصميم المقياس، رغم عدم وجود إرشادات تعينهم في اختيار العدد الأمثل لهذين المتغيرَين للحصول على تقديرات ثبات تتسم بالدقة.

ولعل مؤشرَي التحيز BIAS وجذر متوسط مربع الخطأ Root Mean square Error (RMSE) من أكثر المؤشرات استخدامًا لفحص دقة التقديرات على نحو ما أشار هارول (Harwell, 2019)، وستوظَّف في هذه الدراسة للمقارنة بين دقة تقدير الثبات عبر مستويات متغيرات الدراسة وتقاطعاتها. وقد بيَّن هارول (Harwell, 2019) أن التحيز هو معدل الفرق بين القيم الحقيقية والقيم المقدرة عبر مرات تكرار اختيار العينات، أما عن متوسط مربع الخطأ فقد بيَّن أن تقديره يكون من خلال تباين الفرق بين القيم الحقيقية والقيم المقدرة، وقد توصل من خلال بحثه أن التحيز استخدم في 210 من أصل 677 دراسة منشورة في ست مجلات في الفترة 1985-2012 بنسبة (33.1%)، في حين أن جذر متوسط مربع الخطأ استخدم في 298 دراسة بنسبة (44.1%)، وهذا يدل على جودة هذه المؤشرات في فحص دقة التقدير.

تقتصر حدود إمكانيات معامل ألفا في تقدير ثبات الدرجات الخام، ولا يصلح استخدامه في تقدير ثبات الدرجات المحولة (Almehrizi, 2021)؛ لذلك جاءت هذه الدراسة لمعرفة تأثير حجم العينة، وتدريج الاستجابة، ونوع الدرجة المحولة بمعامل ألفا العام؛ لذا حاولت هذه الدراسة الإجابة عن الأسئلة الآتية:

1. ما الإحصاءاتُ الوصفية لمعامل ثبات ألفا العام لنتائج مقياس اتجاهات الطلبة نحو الرياضيات في TIMSS بسلطنة عُمان، تبعًا لنوع التدريج (رباعي، ثلاثي، ثنائي)؟
2. ما الإحصاءاتُ الوصفية لمعامل ثبات ألفا العام للمحرزي لنتائج العينات العشوائية المستخرجة من مقياس اتجاهات الطلبة نحو الرياضيات في TIMSS بسلطنة عُمان، وفقًا لكل من: تدريج الاستجابة، ونوع الدرجة المحولة، وحجم العينة؟
3. ما قيَمُ جذر متوسطات مربع الخطأ (RMSE) لمعامل ثبات ألفا العام للمحرزي لنتائج مقياس اتجاهات الطلبة نحو الرياضيات في TIMSS بسلطنة عُمان عبر العينات العشوائية، وفقًا لكل من: تدريج الاستجابة، ونوع الدرجة المحولة، وحجم العينة؟

أهميةُ الدِّراسة:

تنبثق الأهمية النظرية للدراسة من أنها توفر معلوماتٍ عن معامل ثبات ألفا العام، الذي يصلح لتقدير ثبات الدرجات الخام والمحولة، بخلاف معاملات الثبات المعتادة، كمعامل ألفا كرونباخ الذي يقتصر على تقدير ثبات الدرجات الخام فقط، وكذلك تفحص جودة هذه المعامل في ضوء عدة متغيرات (حجم العينة، وتدريج الاستجابة، ونوع الدرجة). كما ستوفر هذه الدراسة معلومات تساعد علماء النفس والقياس وأصحاب القرار والتربويين في الحصول على تقديرات ثبات عالية، وتوفير أدلة على أفضلية اختيار معامل الثبات المناسب لحجم العينة، وتدريج الاستجابة، ونوع الدرجة المحوَّلة.

مصطلحاتُ الدِّراسة

حجمُ العينة Sample size: عرَّف علام (2011) العينة بأنها مجموعة جزئية من المجتمع تُختار بطريقة احتمالية، بحيث تمثل خصائص المجتمع، وبيَّن أن هناك عدة عوامل يجب أن تؤخذ بعين الاعتبار عند تحديد حجم العينة، وهي: نوع البحث، وفروض البحث، وتكاليف البحث، وأهمية النتائج، وطرق جمع البيانات، والدقة المطلوبة. وفي هذا البحث اعتُمدت ثلاثة حجوم للعينة: 30، 200، 1000، واختيرت بطريقة عشوائية من مجتمع الدراسة.

تدريجُ الاستجابة Response scale: يتكون مقياس ليكرت من مجموعة من العبارات تقيس الاتجاهات نحو موضوع معين، ويُطلب من أفراد الدراسة الاستجابة لكل عبارة من خلال تدريج استجابة له من ثلاثة إلى سبعة تدريجات، وتعطى كل استجابة من هذه الاستجابات قيمة عددية، ومن ثم تُجمع هذه القيم العددية للحصول على الدرجة الكلية التي تعبر عن اتجاه الفرد نحو موضوع معين (علام، 2011).

الدرجةُ المحوَّلة Derived Scores: هي قيم إحصائية رقمية تستخدم في تحديد المركز النسبي للفرد في توزيع ما؛ لتسهيل وصف أدائه بالنسبة إلى أقرانه في اختبار يقيس سمة ما (علام، 2006).

ثباتُ الاتساق الداخلي Internal consistency: يعكس الاتساق الداخلي إلى أي مدى تقيس المفرداتُ المكوِّنة لأداة القياس الجوانبَ المختلفة التي صُممت الأداة لقياسها (Revicki, 2014).

معاملُ ثبات ألفا العام Generalized Alpha Reliability Coefficient: معادلة عامة لمعامل ثبات ألفا كرونباخ تقدم تقديرا لثبات مختلف الدرجات المحولة التي تُحسب للدرجات الخام في تفسير نتائج المستجيبين في أدوات القياس، مثل الرتب المئينية والدرجات المعيارية وغيرها، وقد طوَّره المحرزي (2013؛ 2022).

محدِّداتُ الدراسة

سيقتصر تعميم نتائج هذه الدراسة على طلاب الصف الرابع في بسلطنة عُمان، الذين طبق عليهم TIMSS عام 2019. وكذلك سيقتصر تعميم النتائج على دقة اختيار العينات، ودقة تقدير: BIAS, RMSE.

منهجيةُ الدراسةِ وإجراءاتُها

منهجُ الدراسة

استُخدم المنهج التجريبي في الدراسة الحالية والقائم على أسلوب المحاكاة؛ إذ إن النتائج التجريبية القائمة على هذا الأسلوب غالبًا ما تكون هي الأفضل (Zumbo et al., 2007)، وقد استُخدم هذا المنهج لمعرفة فاعلية أداء معامل ثبات ألفا العام وتأثره بحجم العينة، وتدريج الاستجابة، ونوع الدرجة المحولة.

مجتمعُ الدراسةِ وعينتُها

استخدمت الدراسة بيانات أرشيفية، وهي جميع استجابات طلبة الصف الرابع الأساسي في المدارس الحكومية والمدارس الخاصة والدولية بسلطنة عُمان للعام الدراسي 2020/2019م على مقياس اتجاهات الطلبة نحو الرياضيات، وقد حُصِل على البيانات في 3 مارس 2022، من الموقع الإلكتروني https://timss2019.org/international-database، وهو الموقع الرسمي للمنظمة الدولية لتقييم الإنجاز التعليمي IEA، وبلغ حجم المجتمع 6814 طالبًا وطالبة. ونظرًا لوجود بيانات مفقودة في البيانات، حُذف الطلبة الذين لم يستجيبوا على أي من عبارات المقياس لحساب معامل ثبات ألفا العام؛ وذلك لأن طبيعة معامل الثبات لا تتناسب مع البيانات المفقودة. وقد بلغ حجم العينة 5117 (ذكور 2510، وإناث 2607).

أداةُ الدِّراسة

استخدمت الدراسة مقياس اتجاهات الطلبة نحو الرياضيات ليكون أحد المقاييس المستخدمة مع اختبار TIMSS 2019 لطلبة الصف الرابع الأساسي، وهو اختصار لجملة Trends of International Mathematics and Science Studies، ويعني دراسات الاتجاهات الدولية في الرياضيات والعلوم، وتسهم الدول المشاركة في إعداد الإطار العام للاختبارات وصياغة محتواها تحت إشراف وإدارة الهيئة الدولية لتقويم التحصيل التربوي. كما تتولى كل دولة مشاركة مسؤولية ترجمة محتوى الأدوات ومواءمتها، وتنفيذ الاختبارات على مدارس العينة المرشَّحة والممثِّلة للدولة. ويمكن حصر أدوات الدراسة في: كتيبات الأسئلة (علوم ورياضيات للصفين الرابع والثامن)، والاستبانات الموجهة إلى (المدرسة، المعلم، الطالب، ولي الأمر). وتركز هذه الدراسة على استبانة الطالب، التي تقيس اتجاهات الطلبة نحو الرياضيات، ويتكون المقياس من ثلاثة محاور: يتضمن المحور الأول 9 عبارات، وتقيس اتجاهات الطلبة نحو تعلم الرياضيات؛ ويتضمن المحور الثاني 6 عبارات، وتقيس اتجاهات الطلبة نحو دروس الرياضيات؛ أما المحور الثالث فيتضمن 9 عبارات، وتقيس اتجاهات الطلبة نحو مادة الرياضيات. ولتحقيق هدف الدراسة اقتُصر على المحور الأول من استبانة الطالب، الذي يهتم باتجاهات الطلبة نحو تعلم الرياضيات للصف الرابع الأساسي، ويتحقق الحصولُ على استجابة الطالب باستخدام تدريج رباعي لدرجة الموافقة (أوافق كثيرًا = 4 درجات، أوافق قليلًا = 3 درجات، أرفض قليلًا = 2 درجتان، أرفض كثيرًا = 1 درجة)، وكانت عبارات المقياس كلها موجبة ما عدا العبارتيْن الثانية والثالثة؛ كانتا سالبتَين (TIMSS & PIRLS, 2019).

يتمتع المحور الأول – اتجاهات الطلبة نحو تعلم الرياضيات– للمقياس بثبات وصدق في جميع الدول التي طُبِّق عليها، وكان الحصول على مؤشرات الصدق والثبات من دليل المقياس المتوفر في الموقع الإلكتروني https:// timss2019.org/international-database/ التابع للمنظمة الدولية لتقييم الإنجاز التعليمي IEA. وبلغت قيمة معامل ثبات ألفا للمحور الأول في سلطنة عُمان 0.80. وبلغ معامل ارتباط بيرسون بين هذا المحور ونتائج التحصيل في اختبار TIMSS في سلطنة عُمان 0.28، في حين تراوح بين 0.12 و0.41 في مختلف الدول التي طبقت المقياس في عام 2019، وبلغ المتوسط الحسابي لمعامل ارتباط بيرسون بين هذا المحور ونتائج التحصيل في اختبار TIMSS 0.20. وأظهر تحليل المكونات الرئيسة (Principal component) للمقياس أن تشبعات عبارات هذا المحور في سلطنة عُمان تراوحت بين 0.27 و0.84، وتفسر ما نسبته 44% من التباين الكلي (TIMSS & PIRLS, 2019).

الإجراءات

المرحلةُ الأولى (تكوين العينات الأساسية الثلاث وحساب الدرجات المحولة): استُخرجت عينات أساسية من العينة الرئيسة، وذلك بنسخ العينة الرئيسة إلى ثلاث نسخ عُدِّل فيها عدد الاستجابات من خلال دمج الاستجابات للحصول على ثلاثة تدريجات. فالنسخة الأولى (التدريج الرباعي) تحتوي على أربع استجابات (أوافق كثيرًا= 4)، (أوافق قليلًا= 3)، (أرفض قليلًا= 2)، (أرفض كثيرًا= 1)، فتبلغ الدرجة الصغرى للمقياس 9، والدرجة العظمى 36. أما النسخة الثانية (التدريج الثلاثي)، فتحتوي على ثلاث استجابات من خلال دمج الاستجابتين الوسيطتين في استجابة واحدة: (موافق كثيرًا) = 3، (موافق قليلًا + أرفض قليلًا) =2، (أرفض كثيرًا) =1 فتبلغ الدرجة الصغرى للمقياس 9، والدرجة العظمى 27. في حين أن النسخة الثالثة تحتوي على استجابتين من خلال دمج كل استجابتين طرفيتين في استجابة واحدة: (أوافق كثيرًا + أوافق قليلًا) =2، (أرفض قليلًا + أرفض كثيرًا) =1، فتبلغ الدرجة الصغرى للمقياس 9، والدرجة العظمى 18.

المرحلةُ الثانية (توليد العينات العشوائية الفرعية): بهدف الحصول على ثلاثة أحجام للعينة، جرى الاختيار العشوائي لثلاث عينات فرعية من كل عينة أساسية بطريقة عشوائية بسيطة تختلف في أحجامها (30، 200، 1000)، وتختلف عن العينات الأساسية بأن حجمها يختلف عن حجم العينة الرئيسة، أما العينات الأساسية فلها حجم العينة الرئيسة نفسُه. وقد تحقق ذلك بطريقة (Bootstrap) في برنامج R، وصُنِّفت الأحجام واختيرت لتمثل ثلاثة مستويات لحجم العينة وفقًا لطبيعة البحث التجريبي القائم على أسلوب المحاكاة، وبناءً على الدراسات السابقة القائمة على هذا المنهج كدراستَي: (زارع، 2021؛ Geldhof et al., 2017).

المرحلةُ الثالثة (تحليل بيانات العينات: حساب معامل الثبات للعينات المختلفة): كتب المحرزي (Almehrizi, 2022) برمجةً خاصة باستخدام برنامج R لاستخراج معامل ثبات ألفا العام لأنواع الدرجات (الخام، والمحولة) في العينات الرئيسة والعينات الفرعية.

المرحلةُ الرابعة: تكررت المرحلتان الثانيةُ والثالثة 1000 مرة باستخراج عينات عشوائية بثلاثة أحجام مختلفة للتدريجات الثلاثة من مقياس اتجاهات الطلبة نحو الرياضيات، ومن ثَمَّ تحقق الحصولُ على إجمالي 9000 عينة (1000 تكرار × 3 تدريجات × 3 أحجام عينة). وقُدِّر معامل ثبات ألفا العام بعدد 42000 مرة لكل منهما (1000 تكرار × [1 تدريج × 3 أحجام عينة × 6 درجات + 2 تدريجات × 3 أحجام عينة × 48 درجات]).

وفيما يتعلق بحساب الدرجات المحولة، فقد استخدم برنامج SPSS في تقديرها؛ فقُدرت الدرجات المعيارية الاعتدالية من خلال تحويل الدرجات الخام إلى رتب مئينية، ومن ثم تستخرج الدرجات المعيارية الاعتدالية المقابلة لكل رتبة مئينية من جدول التوزيع الطبيعي، أما الرتب المئينية فتُعرَف فيها النسب المئوية لمجموعة الأفراد الذي حصلوا على درجات تقل عن درجة معينة، فهي تفسر من خلال مجموعة مرجعية طُبق الاختبار عليها. وفيما يتعلق بالتساعيات، فهي درجات معيارية تتوزع في تسع فئات يضم كل منها مجموعة من الدرجات، ويخضع هذا التوزيع لاعتبارات حقيقية تبنى على أساس ظروف المؤسسة، والهدف من الدرجات، ونظام التقويم المتبع (عودة، 2010).

وفيما يتعلق بدرجات تيمز، فقد استُخدمت درجتان معياريتان، وهما تيمز الأولى وتيمز الثانية، ودرجة تيمز الأولى تتم من خلال خطوتين: تحويل الدرجة الخام تحويلًا غير خطيٍّ إلى قيمة لوجيت باستخدام نظرية الاستجابة للمفردة، ومن ثم تحويل خطي لقيمة لوجيت باستخدام ميل وقيمته 1.234546 وثابت قيمته 8.402636. ونتيجة لهذا التحويل تراوحت درجة تيمز الأولى بين 3.85307 إلى 13.14434. أما درجة تيمز الثانية فتقسم درجة تيمز الأولى إلى ثلاث فترات باستخدام درجتي قطع، وهما 8.4 و10.2 وأعطيت الفترة الأصغر القيمة 1 تليها القيمة 2 ثم القيمة 3 للفترة الأكبر.

المعالجةُ الإحصائية:

للإجابة عن أسئلة الدراسة، استُخرجت الإحصاءات الوصفية ومعامل ثبات ألفا العام للدرجات الخام، والمحولة، وللعينات العشوائية في مقياس اتجاهات الطلبة نحو الرياضيات TIMSS في سلطنة عُمان عبر متغيرات: 1) تدريج المقياس و2) نوع الدرجة، وكذلك حُسبت نسب التحيز BIAS باستخدام المعادلة 9، وجذر متوسطات مربع الخطأ RMSE باستخدام المعادلة 11.

$$\text{Bias\%} = \frac{\widehat{G\alpha} - \overline{\widehat{G\alpha}_r}}{\widehat{G\alpha}} \text{ X } 100 \text{----------}$$

$\widehat{G\alpha}$: تقدير معامل ثبات ألفا العام في العينة الرئيسة لكل نوع من الدرجات المستخدمة الخام والمحوَّلة لكل تدريج باختلاف أحجامه الثلاثة. $\overline{\widehat{G\alpha}_r}$: متوسط تقديرات معامل ثبات ألفا العام في العينات العشوائية الفرعية لكل نوع من الدرجات المستخدمة الخام والمحولة لكل تدريج باختلاف أحجامه الثلاثة.

$$\text{RMSE} = \sqrt{\frac{\sum_{r=1}^{1000}(\widehat{G\alpha}_r - \widehat{G\alpha})^2}{1000}} \text{ ----------}$$

r: رقم التكرار (رقم العينة العشوائية الفرعية). $\widehat{G\alpha_r}$: تقديرات معامل ثبات ألفا العام في العينات العشوائية الفرعية لكل نوع من الدرجات المستخدمة الخام والمحولة.

ولفحص دقة التقديرات، قورِنت المتوسطات الحسابية لتقديرات معاملات الثبات (ألفا العام، وبيتا العام) لكل نوع من الدرجات المستخدمة الخام والمحولة في كل تدريج على حدة بأحجامه الثلاثة (30، 200، 1000)، بقيم معاملات ثبات (ألفا العام، وبيتا العام) للعينة الرئيسة، مع إيجاد نسب التحيز لمعرفة مدى قربها واقترابها من ثبات العينة الرئيسة بزيادة حجم العينة؛ فقد أشار تراكسلر (Traxler, 2017) إلى أن نسب التحيز تكون مقبولة إذا كانت أقل من أو يساوي 10%، وتدل الإشارة الموجبة أو السالبة في نسب التحيز على اتجاه التحيز من حيث التقليل أو التضخيم في تقديرات الثبات.

نتائجُ الدِّراسة

نتائجُ السؤال الأول، ونصُّه:

«ما معاملات ثبات ألفا العام لنتائج مقياس اتجاهات الطلبة نحو الرياضيات في TIMSS بسلطنة عُمان تبعًا لنوع التدريج (رباعي، ثلاثي، ثنائي)؟»

للإجابة عن هذا السؤال، حُسب معامل ثبات ألفا العام للمحرزي (2022) لأنواع الدرجات المختلفة في مقياس اتجاهات الطلبة نحو الرياضيات TIMSS في سلطنة عُمان في التدريجات الثلاثة (الرباعي، والثلاثي، والثنائي)؛ إذ يحتوي جدول (1) على تقديرات معامل ثبات ألفا العام للدرجات الخام والمعيارية عند الحالات المختلفة لتدريج المقياس (رباعي، ثلاثي، ثنائي) ويوضح جدول (1) هذه التقديرات.

جدول (1): معامل ثبات ألفا العام للدرجات الخام والمحولة في التدريجات الثلاثة (ن= 5117)

الدرجة	التدريج الرباعي	التدريج الثلاثي	التدريج الثنائي
الخام	0.8092	0.8099	0.7602
المعيارية الاعتدالية	0.8026	0.8044	0.7658
الرتبة المئينية	0.8763	0.7862	0.7534
التساعيات	0.7841	0.7772	0.6934
تيمز الأولى	0.7989	-	-
تيمز الثانية	0.6833	-	-

تظهر النتائج في جدول (1) أن قيم معاملات ثبات ألفا العام للدرجات الخام تراوحت بين 0.7062 و0.8099 للمقياس في التدريجات الثلاثة، كما يلاحَظ أن معامل ثبات ألفا العام كان أصغر في التدريج الثنائي. أيضًا تُظهر النتائج أن قيم معامل ثبات ألفا العام في كل أنواع الدرجات المحولة أقل من قيمته للدرجات الخام للمقياس في التدريجات الثلاثة، ما عدا الدرجة المعيارية الاعتدالية في التدريج الثنائي؛ إذ كان معامل ثباتها أكبر من معامل ثبات الدرجات الخام.

كما تُظهر النتائج أن قيم معامل ثبات ألفا العام للدرجة التساعية للمقياس أصغر منها للدرجات المحولة الأخرى في جميع التدريجات الثلاثة، وبالمقابل كانت الأكبر للدرجات المعيارية الاعتدالية. وهذا يرجع إلى نوعية الدرجات المحولة من حيث عدد الدرجات، وطبيعة التوزيع التكراري لها؛ فالدرجات التساعية تحتوي على 9 درجات، في حين أن الدرجات المعيارية تأخذ أي قيمة حقيقية.

وبغرض الوقوف بوضوح على الفروق في تقدير معامل الثبات عبر متغير تدريج المقياس، طُرحت التقديرات للتدريجات الثلاثية من تقديرات التدريج الرباعي، ويحتوي جدول (2) على هذه الفروق.

جدول (2): الفروق بين تقديرات معامل الثبات (رباعي– ثلاثي، رباعي – ثنائي) (ن= 5117)

نوع الدرجة	الرباعي– الثلاثي	الرباعي– الثنائي	المتوسط**
الخام	-0.0007	0.0490	0.0242
المعيارية الاعتدالية	-0.0018	0.0368	0.0175
الرتبة المئينية	0.0001	0.0329	0.0165
التساعيات	0.0069	0.0907	0.0488
المتوسط*	0.0110	0.0524	-

* متوسط الفروق لمعامل الثبات للدرجات في التدريج الواحد ** متوسط الفروق لمعامل الثبات للدرجة عبر التدريجات الثلاثة.

تُظهر النتائج في جدول (2) الفروق بين تقديرات معامل الثبات ألفا العام لكل نوع من درجات المقياس (الخام، والمحولة)، وقد حُسبت قيم الفروق بطرح قيمة معامل ثبات ألفا العام لكل نوع من الدرجات في التدريج الرباعي من نظيراتها في التدريجين الآخرين (الثلاثي، والثنائي)، وذلك باعتبار أن التدريج الرباعي يمثل التدريج الأصلي للمقياس. وانحصرت الفروقُ بين تقديرات معاملات الثبات ما بين -0.0018 و0.0907، وأتت قيم المتوسطات الحسابية للفروق في معامل ثبات ألفا العام لكل نوع من درجات المقياس (الخام، والمحولة) عبر التدريجات الثلاثة موجبةً (أكبر من الصفر)؛ أي إن معامل ثبات ألفا العام لكل نوع من الدرجات المستخدمة الخام والمحولة تعطي تقديراتِ ثباتٍ مختلفةً، وظهر هذا النمط في قيم الفروق في معامل ثبات ألفا العام بين التدريجين الرباعي والثنائي لكل أنواع درجات المقياس (الخام، والمحولة)، وكذلك ظهر هذا النمط بين التدريجين الرباعي والثلاثي لكل الدرجات ما عدا الخام، والدرجات المعيارية الاعتدالية؛ فقد أتت الفروق سالبة إذ قُدرت بـ -0.0007 و-0.0018 على التوالي،

وهو ما يشير إلى أن تقديرات معامل ثبات ألفا العام لهاتين الدرجتين كان أكبر في التدريج الثلاثي عن الرباعي. وجاءت قيم الفروق لتقديرات معامل ثبات ألفا العام بين التدريجين الرباعي والثلاثي أصغر عما كانت عليه بين التدريجين الرباعي والثنائي؛ فقد تراوحت بين -0.0018 و0.0069 بمتوسط 0.0110، وظهر هذا النمط في كل نوع من أنواع الدرجات؛ أي إن قيم معاملات ثبات ألفا العام كانت متقاربة بين التدريجين الرباعي والثلاثي، في حين كانت مختلفة أكثر بين التدريجين الرباعي والثنائي.

ولمعرفة الفروق في تقدير معامل الثبات عبر متغير نوع الدرجة، يحتوي جدول (3) على الفروق بين تقديرات معامل ثبات ألفا العام بين حالة الدرجة الخام، والدرجات المحولة (المحولة – الخام)

جدول (3): الفروق بين تقديرات معامل الثبات (محولة – الخام) (ن= 5117)

الدرجة	الرباعي	الثلاثي	الثنائي	المتوسط**
المعيارية الاعتدالية	0.0066-	0.0055-	0.0056	0.0020-
الرتبة المئينية	0.0229-	0.0273-	0.0068-	0.0178-
التساعيات	0.0251-	0.0327-	0.0668-	0.0145-
درجة تيمز الأولى	0.0103-	-	-	-
درجة تيمز الثانية	0.1259-	-	-	-
المتوسط*	0.0382-	0.0206-	0.0277-	-

* متوسط الفروق لمعامل ثبات الدرجات في التدريج الواحد ** متوسط الفروق لمعامل الثبات للدرجة عبر التدريجات الثلاثة.

تُظهر النتائج في جدول 3 قيم الفروق بين تقديرات معامل ثبات ألفا العام وبين الدرجات المحولة والدرجات الخام (ثبات الدرجة المحولة – ثبات الدرجة الخام) في كل تدريج من التدريجات الثلاثة (الرباعي، والثلاثي، والثنائي)؛ فقد انحصرت قيم الفروق بين تقديرات معامل ثبات ألفا العام بين -0.1259 و0.0056، وأتت قيم المتوسطات الحسابية للفروق في تقديرات معامل ثبات ألفا العام لكل درجة من الدرجات المحولة عبر التدريجات الثلاثة (الرباعي، والثلاثي، والثنائي) سالبة؛ أي إن تقديرات معامل ثبات ألفا العام في كل درجة من الدرجات المحولة أصغر من تقديراته للخام. وظهر هذا النمط في قيم الفروق في معامل ثبات ألفا العام في كل أنواع الدرجات المحولة في التدريجات الثلاثة، كلٌّ على حدة، فيما عدا معامل ثبات ألفا العام للدرجة المعيارية الاعتدالية في التدريج الثنائي؛ إذ جاء الفرق موجبًا (0.0056)، وهذا يشير إلى أن تقدير معامل ثبات ألفا العام للدرجة المعيارية الاعتدالية أكبر مما كان عليه عند الدرجة الخام.

وتتفق هذه النتيجة مع دراستَي المحرزي (Almehrizi, 2021; 2022)، في أنه لا يمكن التعميم بأن قيم معامل ثبات ألفا العام أكبر في الدرجات الخام عن الدرجات المحولة. وتختلف مع دراستَي كل من (Choudhury & Bhattacharjee, 2014; Preston & colman, 2000)، اللتين أشارتا إلى أنَّ الثبات يتناسب طرديًّا بزيادة تدريج الاستجابة.

نتائجُ السؤالِ الثاني، ونصُّه:

«ما الإحصاءاتُ الوصفية لمعامل ثبات ألفا العام للمحرزي لنتائج العينات العشوائية المستخرجة من مقياس اتجاهات الطلبة نحو الرياضيات في TIMSS بسلطنة عُمان، وفقًا لكل من: تدريج الاستجابة، ونوع الدرجة المحولة، وحجم العينة؟»

للإجابة عن هذا السؤال، استُخرجت الإحصاءات الوصفية لمعامل ثبات ألفا العام في العينات العشوائية عبر التدريجات؛ إذ يحتوي جدول (4) على الإحصاءات الوصفية لقيم معاملات الثبات للتدريج الرباعي عند المستويات المختلفة لمتغيرَي: نوع الدرجة، وحجم العينة.

جدول (4): الإحصاءات الوصفية لقيم معاملات الثبات للتدريج الرباعي
باستخدام تقديرات البوتستراب للأحجام الثلاثة للعينة (30، 200، 1000)

الدرجة	ثبات العينة الرئيسة	ن	نسبة التحيز%	المتوسط	الانحراف المعياري	الالتواء	التفلطح
الخام	0.8092	30	2.9	0.7860	0.0986	1.4858-	3.5664
		200	0.4	0.8063	0.0319	0.6574-	0.7078
		1000	0.1	0.8084	0.0124	0.1957-	0.0452-
المعيارية الاعتدالية	0.8026	30	0.5	0.7988	0.0711	1.1923-	3.2662
		200	0.1	0.8066	0.0255	0.4607-	0.3641
		1000	0.0	0.8026	0.0103	0.2673-	0.1205
الرتبة المئينية	0.8763	30	0.03-	0.7865	0.0711	1.0028-	2.2060
		200	0.10	0.7853	0.0262	0.3753-	0.1458
		1000	0.02-	0.7865	0.0107	0.2241-	0.0120
التساعيات	0.7841	30	0.20	0.7822	0.0706	1.0158-	2.0126
		200	0.010	0.7829	0.0254	0.3408-	0.1686
		1000	0.03-	0.7843	0.0105	0.2080-	0.0807-
تيمز الأولى	0.7989	30	0.50	0.7950	0.0728	1.0010-	2.3478
		200	0.20	0.7976	0.0266	0.4098-	0.2812
		1000	0.0	0.7989	0.0109	0.1830-	0.0988-
تيمز الثانية	0.6833	30	2.6	0.6654	0.1102	0.8585-	0.8290
		200	0.4	0.6806	0.0404	0.4062-	0.0623
		1000	0.1	0.6827	0.0162	0.2648-	0.2656

تُظهر النتائج في جدول (4) أن قيم المتوسطات الحسابية لمعامل ثبات ألفا العام لكل من الدرجات الخام، وتيمز الثانية جاءت أصغر من تقدير معامل ثبات ألفا العام في العينة الرئيسة في كل حجم من الأحجام الثلاثة (30، 200، 1000)، وتقترب منها بزيادة حجم العينة. فعلى سبيل المثال، أتت قيم المتوسطات الحسابية لمعامل ثبات ألفا العام في الدرجة الخام في الأحجام الثلاثة (30، 200، 1000) مساوية لـ 0.7860، 0.8063، 0.8084 على الترتيب، في حين كانت قيمُها قريبة من ثبات العينة الرئيسة في الرتب المئينية، على أنها كانت في التساعيات تقترب من ثبات العينة الرئيسة بزيادة حجم العينة عند الحجمين (30، 200)، وكانت قريبة من ثبات العينة الرئيسة عند الحجم 1000، في حين اقتربت قيم المتوسطات الحسابية لمعامل ثبات ألفا العام في كل من الدرجات (المعيارية الاعتدالية، وتيمز الأولى) من ثبات العينة الرئيسة بزيادة حجم العينة عند الحجمين (30، 200)، وكانت مساوية لثبات العينة الرئيسة عند الحجم 1000.

وتظهر نسب التحيز في التدريج الرباعي بوجهٍ عام أنها نسب حسب ما بيَّن تراكسلر (Traxler, 2017) لم تتعدَّ 10% في الدرجات الخام والدرجات المحولة، ويمكن تجاهلها؛ فإنها تراوحت بين (-0.03، 2.9) وهذا يُظهر استقرارَ تقديرات معامل الثبات في العينات العشوائية واقترابها من ثبات العينة الرئيسة، سواء أكان التحيز بالتقليل أم التضخيم، ولتوضيح ذلك أتت نسب التحيز في معامل ثبات ألفا العام في التساعيات للأحجام الثلاثة (30، 200، 1000) 0.2، 0.1، -0.03 على التوالي؛ فالنسبتان الأولى والثانية تدلان على التقليل في تقدير الثبات؛ بمعنى أن معامل ثبات ألفا العام للدرجة التساعية عند الحجمين (30، 200) كان أصغر من تقدير ثبات العينة الرئيسة، في حين أن النسبة الثالثة تدل على التضخيم في تقدير الثبات مقارنة بثبات العينة الرئيسة؛ أي إن معامل ثبات ألفا العام في التساعيات عند الحجم 1000 أكبر من ثبات العينة الرئيسة. ويوضح جدول (6) الإحصاءات الوصفية لقيم معامل الثبات بالتدريج الثلاثي لمقياس اتجاهات الطلبة نحو الرياضيات في الأحجام الثلاثة (30، 200، 1000).

وفيما يتعلق بالتدريج الثلاثي، فجدول (5) يحتوي الإحصاءات الوصفية لقيم معامل الثبات في حالة هذا التدريج عند المستويات المختلقة لمتغيرَي: نوع الدرجة، وحجم العينة.

جدول (5): الإحصاءات الوصفية لقيم معامل الثبات للتدريج الثلاثي للعينات (30، 200، 1000)

الدرجة	ثبات العينة الرئيسة	ن	نسبة التحيز %	المتوسط	الانحراف المعياري	الالتواء	التفلطح
الخام	0.8099	30	3.6	0.7806	0.0963	1.4559-	3.6292
		200	0.3	0.8073	0.0302	0.6181-	0.6697
		1000	0.2	0.8079	0.0114	0.0688-	0.2224-
المعيارية الاعتدالية	0.8044	30	1.5	0.7925	0.0744	1.1678-	2.3558
		200	0.1	0.8035	0.0259	0.5384-	0.5118
		1000	0.01-	0.8045	0.0100	0.0238-	0.1252-

الدرجة	ثبات العينة الرئيسة	ن	نسبة التحيز %	المتوسط	الانحراف المعياري	الالتواء	التفلطح
الرتبة المئينية	0.7862	30	0.8	0.7803	0.0741	1.0504-	2.2844
		200	0.03	0.7860	0.0267	0.4420-	0.3020
		1000	0.03-	0.7864	0.0106	0.0851-	0.1093-
التساعيات	0.7772	30	0.8	0.7711	0.0750	1.2230-	4.0932
		200	0.03	0.7770	0.0267	0.4115-	0.2255
		1000	0.03-	0.7774	0.0105	0.0597-	0.1389-

يتضح من جدول (5) أن قيم المتوسطات الحسابية لمعامل ثبات ألفا العام في الدرجات الخام تقترب من قيمة معامل ثبات ألفا العام في العينة الرئيسة كلما زاد حجم العينة؛ فعلى سبيل المثال، جاءت قيم المتوسطات الحسابية لمعامل ثبات ألفا العام في الدرجة الخام في الأحجام الثلاثة (30، 200، 1000) 0.7806، 0.8073، 0.8079 على التوالي، وتظهر أيضًا قيم المتوسطات لمعامل ثبات ألفا العام عند كل من الدرجات (الرتب المئينية، والتساعيات) أنها تقترب من ثبات العينة الرئيسة كلما زاد الحجم عند الحجمين (30، 200)، في حين تكون قريبةً من ثبات العينة الرئيسة عند الحجم 1000.

وتقترب قيم الانحرافات المعيارية لمعامل ثبات ألفا العام لجميع أنواع درجات المقياس من الصفر كلما زاد حجم العينة، وهذا يدل على استقرار تقديرات معامل الثبات بزيادة حجم العينة، وكذلك تشير مؤشراتُ التوزيع التكراري للمعامل في العينات العشوائية إلى أنها قريبة من مؤشرات التوزيع الطبيعي، وهذا ما تؤكده قيم الالتواء والتفلطح في جميع أنواع الدرجات المستخدمة الخام والمحولة؛ فإنها كانت تقترب من الصفر كلما زاد حجم العينة.

وتُظهر النتائج أن نسب التحيز لمعامل ثبات ألفا العام في جميع الدرجات الخام والدرجات المحولة لم تتعدَّ 10%، وهي نسب مقبولة؛ فقد تراوحت بين -0.03 و3.6، وهذا يدل على استقرار تقديرات معامل ثبات ألفا العام في كل الدرجات الخام والدرجات المحولة؛ فعلى سبيل الذكر، أتت قيم نسب التحيز لمعامل ثبات ألفا العام في الرتب المئينية للأحجام الثلاثة (30، 200، 1000) 0.8، 0.03، -0.03 على التوالي، فالنسبتان الأولى والثانية تدلان على التقليل في الثبات، في حين أن النسبة الثالثة تدل على التضخيم في تقدير الثبات مقارنة بثبات العينة الرئيسة، وتدل على قرب تقديرات معامل ثبات ألفا العام في العينات العشوائية واقترابها من ثبات العينة الرئيسة كلما زاد حجم العينة. ويوضح جدول (7) الإحصاءات الوصفية لقيم معامل الثبات بالتدريج الثنائي لمقياس اتجاهات الطلبة نحو الرياضيات في الأحجام الثلاثة (30، 200، 1000).

وفيما يتعلق بالتدريج الثنائي، فجدول (6) يحتوي الإحصاءات الوصفية لقيم معامل الثبات في حالة هذا التدريج عند المستويات المختلفة لمتغيرَي: نوع الدرجة، وحجم العينة.

جدول (6): الإحصاءات الوصفية لقيم معاملات الثبات للتدريج الثنائي
باستخدام تقديرات البوتستراب لأحجام العينات (30، 200، 1000)

الدرجة	ثبات العينة الرئيسة	ن	نسبة التحيز %	المتوسط	الانحراف المعياري	الالتواء	التفلطح
		30	6.3	0.7118	0.1477	1.3064-	1.9884
الخام	0.7602	200	0.7	0.7548	0.0431	0.7240-	1.0032
		1000	0.1	0.7591	0.0165	0.2886-	0.0615
		30	0.9	0.7586	0.0849	0.8486-	1.0987
المعيارية الاعتدالية	0.7658	200	0	0.7658	0.0297	0.3158-	0.1112
		1000	0.01-	0.7659	0.0119	0.0379-	0.0477-
		30	0.1-	0.7543	0.0798	0.7146-	0.9672
الرتبة المئينية	0.7534	200	0.2-	0.7550	0.0286	0.2298-	0.1697
		1000	0.1-	0.7540	0.0115	0.0229-	0.0383-
		30	0.9	0.6872	0.1015	0.8595-	1.2449
التساعيات	0.6934	200	0.1-	0.6941	0.0348	0.3059-	0.2138
		1000	0.1-	0.6939	0.0141	0.0229	0.0518-

تُظهر النتائج في جدول (6) أن قيم المتوسطات الحسابية لمعامل ثبات ألفا العام في الدرجات الخام في العينات العشوائية تقترب من معامل ثبات ألفا العام في العينة الرئيسة كلما زاد حجم العينة؛ فعلى سبيل الذكر، أتت قيم المتوسطات الحسابية لمعامل ثبات ألفا العام في الدرجة الخام في الأحجام الثلاثة (30، 200، 1000) 0.7118، 0.7548، 0.7591 على التوالي، في حين تظهر قيم المتوسطات الحسابية لمعامل ثبات ألفا العام في الدرجات المعيارية الاعتدالية أنها قريبة من ثبات العينة الرئيسة عند الحجمين (30، 1000)، وتكون مساويةً لثبات العينة الرئيسة عند الحجم 200، في حين جاءت في الرتب المئينة والتساعيات قريبةً من ثبات العينة الرئيسة. وتؤكد النتائج في جدول (7) استقرارَ تقديرات معامل ثبات ألفا العام في كل أنواع الدرجات كلما زاد حجم العينة، وهذا ما تُظهره قيم الانحرافات المعيارية لمعامل الثبات؛ إذ إنها تقترب من الصفر كلما زاد حجم العينة، وكذلك تقترب قيمُ الالتواء والتفلطح من الصفر بزيادة حجم العينة، وهذا يؤكد أن مؤشرات التوزيع التكراري لمعامل الثبات تقترب من مؤشرات التوزيع الطبيعي.

وتظهر نسب التحيز لمعامل ثبات ألفا العام بوجهٍ عام أنها نسب مقبولة لم تتعدَّ 10% في الدرجات الخام المحولة حسب ما بيَّن Traxler (2017)، وهذا يُظهر استقرارَ تقديرات معامل الثبات في العينات العشوائية واقترابها من ثبات

الرئيسة، سواء أكان التحيز بالتقليل أم المبالغة؛ فعلى سبيل المثال، أتت نسب التحيز في الدرجة الخام لمعامل ثبات ألفا العام للأحجام الثلاثة 6.3، 0.7، 0.1 على التوالي، وهي نسب تدل على التقليل في تقدير الثبات مقارنة بثبات الرئيسة؛ بمعنى أن الثبات في العينات العشوائية أقل من ثبات الرئيسة ولكن بنسب لم تتعدَّ 10%، وتقل هذه النسب بزيادة حجم العينة، وتقترب بذلك من ثبات الرئيسة.

وتتفق هذه النتائج مع دراسة تريزانو هيرموسيلا وألفارادو (Trizano-Hermosilla & Alvarado, 2016) من ناحية إجراءاتها التي تتمثل في اختبار فاعلية معاملات الثبات، وذلك من خلال أخذ عينات عشوائية من العينة الرئيسة بأحجام مختلفة بطريقة البوتستراب، ثم مقارنتها بمعاملات ثبات العينة الرئيسة، وإيجاد نسب التحيز في تقديرات معاملات الثبات، التي أثبتت نتائجُها استقرارَ تقديرات معاملات الثبات بزيادة حجم العينة. وكذلك تتفق مع دراسة زارع (2021)، التي أشارت إلى أفضلية أربع معاملات ثبات عن معامل ألفا من أصل 13 معامل ثبات، وهي على الترتيب: معامل ثبات أوميجا الكلية (ω_t)، ومعامل ثبات أكبر حد أدنى (glb)، ومعامل ثبات الحد الأدنى لجتمان (λ_4)، ومعامل ثبات ألفا للأقسام المتعددة، بعد مقارنة تقديرات كل معامل ثبات في العينات العشوائية، التي أخذت بأحجام مختلفة، ومدى اقترابها من ثبات العينة الرئيسة بزيادة حجم العينة، والنظر إلى نسب التحيز، للحكم على استقرار تقديرات معاملات الثبات. وتتفق أيضًا مع دراسة زارع (2022) في منهجها التجريبي؛ إذ عُقدت مقارنةٌ بين ثبات ألفا على نوعين من البيانات: (الرتبية، والفترية) على مقياس ليكرت ذي 5، 11 بديل، واعتمدت الدراسة على الحكم على فاعلية أداء المعاملين على نسب التحيز، وقد أظهرت النتائج أن معامل ثبات ألفا في البيانات الرتبية كان يعطي تقديرات ثبات مسقرة، ودون تحيز عن معامل ثبات ألفا في البيانات الفترية. واختلفت هذه النتيجة مع دراسة زومبو وآخرين (Zumbo et al., 2007) القائمة على المنهج التجريبي بأسلوب المحاكاة؛ إذ قورِن فيها بين معاملات الثبات (ألفا، وألفا الرتبي، وثيتا الرتبي) للحكم على فاعلية أداء المعاملات، واستخدمت لذلك بيانات من نوع ليكرت، وأشارت النتائج إلى أن معامل ثبات ألفا يعطي تحيزًا سالبًا في جميع فئات الاستجابة، وقد تحقق ذلك من خلال المقارنة بين معاملات ثبات العينات العشوائية ومدى قربها من معامل ثبات الرئيسة، في حين كان معاملا ثبات ألفا وثيتا الرتبي يعطيان تقديرات مستقرة للثبات.

نتائجُ السؤالِ الثالث، ونصُّه:

«ما قيمُ جذر متوسطات مربع الخطأ (RMSE) لمعامل ثبات ألفا العام للمحرزي لنتائج مقياس اتجاهات الطلبة نحو الرياضيات في TIMSS بسلطنة عُمان عبر العينات العشوائية، وفقًا لكل من: تدريج الاستجابة، ونوع الدرجة المحولة، وحجم العينة؟»

للإجابة عن هذا السؤال، استُخرجت قيَم جذر متوسطات مربع الخطأ (RMSE) لمعامل ثبات أنواع الدرجات الخام والمحولة في التدريجات الثلاثة (الرباعي، والثلاثي، والثنائي)، في مقياس اتجاهات الطلبة نحو الرياضيات لأحجام العينات (30، 200، 1000)؛ لمعرفة مدى فاعلية أداء معامل ثبات ألفا العام، كما يوضحها جدول (7)؛ إذ

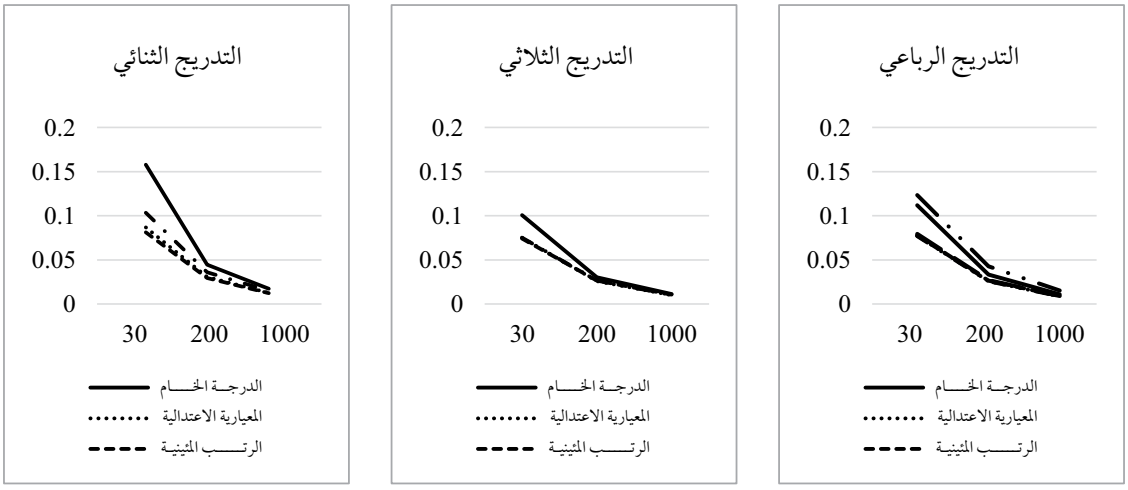
يحتوي على قيم جذر متوسطات مربع الخطأ لمعامل الثبات عند تقاطعات مستويات متغيرات: نوع التدريج، ونوع الدرجة، وحجم العينة.

جدول (7): جذر متوسطات مربع الخطأ لمعامل ثبات الدرجات الخام والمحولة في التدريجات الثلاثة في مقياس اتجاهات الطلبة نحو الرياضيات لأحجام العينات (30، 200، 1000)

نوع التدريج/ حجم العينة		30	200	1000
الرباعي	الخام	0.1013	0.0320	0.0124
	المعيارية الاعتدالية	0.0712	0.0255	0.0103
	الرتبة المئينية	0.0710	0.0262	0.0107
	التساعيات	0.0706	0.0255	0.0105
	تيمز الأولى	0.0729	0.0267	0.0108
	تيمز الثانية	0.1116	0.0404	0.0162
الثلاثي	الخام	0.1006	0.0303	0.0114
	المعيارية الاعتدالية	0.0753	0.0259	0.01
	الرتبة المئينية	0.0743	0.0267	0.0106
	التساعيات	0.0752	0.0267	0.0105
الثنائي	الخام	0.1554	0.0435	0.0166
	المعيارية الاعتدالية	0.0852	0.0297	0.0119
	الرتبة المئينية	0.0797	0.0286	0.0115
	التساعيات	0.1017	0.0348	0.0141

يتضح من جدول (7) أن قيم جذر متوسطات مربع الخطأ RMSE لمعامل ثبات ألفا العام لكل أنواع الدرجات (الخام، والمحولة) في كل تدريج من التدريجات الثلاثة (الرباعي، والثلاثي، والثنائي) بأحجامه الثلاثة المختلفة (30، 200، 1000) تقل بزيادة حجم العينة، وتقترب بذلك من الصفر؛ فعلى سبيل المثال، جاءت RMSE لمعامل ثبات ألفا العام في الدرجة الخام في التدريج الرباعي في كل حجم من الأحجام الثلاثة (30، 200، 1000) 0.1013، 0.032، 0.0124 على التوالي، وظهر هذا النمط في كل أنواع درجات المقياس بتدريجاته الثلاثة، وهذا يدل على استقرار تقديرات معامل ثبات ألفا العام في العينات العشوائية لكل نوع من الدرجات المستخدمة، واقترابها من ثبات الرئيسة. وتتفق كذلك مع دراستَي كلٍّ من زارع (2021) وتريزانو هيرموسيلا وألفارادو (Trizano-Hermosilla & Alvarado, 2016)، اللتين أثبتت نتائجهما أن قيم معاملات الثبات في العينات العشوائية تقترب من ثبات الرئيسة؛ إذ تقلُّ قيم جذر متوسطات

مربع الخطأ بزيادة حجم العينة وتقترب من الصفر. وتُظهر النتائج تقارب سلوك بعض الدرجات المحولة في تقديرات معامل ثبات ألفا العام في كل تدريج من التدريجات الثلاثة (الرباعي، والثلاثي، والثنائي)، وهذا يدل على استقرار تقديرات معامل الثبات في العينات العشوائية؛ فمثلًا، جاءت قيم جذر متوسطات مربع الخطأ (RMSE) لمعامل ثبات ألفا العام في كل من الدرجات (تيمز الأولى، والتساعيات، والرتب المئينية) في التدريجات الثلاثة متقاربة جدًا في تقدير الثبات، وهذا يدل على استقرار تقديرات معامل ثبات ألفا العام كما يوضحه الشكل التالي:



شكل (1): قيم RMSE لمعامل ثبات ألفا للدرجات الخام المحولة في التدريجات الثلاثة وفي الأحجام المختلفة (30، 200، 1000)

الخلاصةُ والمقترَحات

خلصت الدراسة إلى أن معامل ثبات ألفا العام للمحرزي يعطي تقديرات دقيقة للدرجات، بغض النظر عن كونها خامًا أو محولةً، وهذا ما يميز هذا المعامل عما سبقه من المعاملات، وذلك من خلال مؤشرات التحيز وجذر متوسط مربع الخطأ واقتراب ثبات العينات العشوائية من ثبات العينة الرئيسة. وفيما يتعلق بنتائج الدراسة المتعلقة بمتغيراتها: حجم العينة (30، 200، 1000)، وعدد مستويات التدريج (رباعي، ثلاثي، ثنائي)، ونوع الدرجة (خام، معيارية اعتدالية، رتبة مئينية، تساعيات)، فقد خلصت النتائج إلى أن دقة تقدير معامل الثبات تزداد بزيادة حجم العينة، وتوضح ذلك من خلال مؤشرَي: نسبة التحيز، وجذر متوسط مربع الخطأ، وذلك عند جميع حالات تقاطعات متغيرَي عدد فئات التدريج، ونوع الدرجة. وفيما يتعلق بمتغير عدد مستويات التدريج، فقد خلصت النتائج إلى أن ثبات البيانات ذات التدريج الرباعي كان بوجهٍ عام أعلى الثلاثي والثنائي، إلا في حالة الدرجتين الخام والمعيارية الاعتدالية؛ فقد كان الثلاثي أعلى. أما عن متغير نوع الدرجة المحولة، فقد أظهرت النتائج بوجهٍ عام أن ثبات الدرجات المحولة أعلى من معامل ثبات الدرجات الخام، باستثناء حالة التدريج الثنائي؛ فقد كان ثبات الدرجات الخام أعلى من ثبات الدرجات المعيارية الاعتدالية.

ومع ضرورة الأخذ بعين الاعتبار أن صحة تعميم هذه النتائج يقتصر على استخدامها بياناتٍ أرشيفيةً من موقع TIMSS، وكذلك في عدم تناوله لمعاملات تقدير ثبات أخرى، وفي ضوء ذلك توصي الدراسة بما يأتي:

1. استخدامِ معامل ثبات ألفا العام في تقدير ثبات مختلف الدرجات المحولة غير الخطية للمقاييس والاختبارات.
2. استفادةِ الباحثين من الدراسة الحالية، من خلال ما وفرته من أدلة من اختيار الحجم المناسب للعينة، وتدريج الاستجابة، ونوع الدرجة المناسبة لتحقيق أفضل ثبات.
3. استخدامِ مطوري الاختبارات التحصيلية في وزارة التربية والتعليم لمعامل ثبات ألفا العام؛ للحكم على ثبات هذه الاختبارات.
4. حثِّ الباحثين على استخدام معامل ثبات ألفا العام للمحرزي في بحوثهم، وذلك من خلال عمل ورش تدريبية للتعريف بمعامل الثبات، وكيفية استخدامه في برنامج R.

المَراجع

أولًا: العربية

أبو هاشم، سيد. (2006). **الخصائص السيكومترية لأدوات القياس في البحوث النفسية والتربوية باستخدام برنامج SPSS**. مركز البحوث التربوية، كلية التربية، جامعة الملك سعود، السعودية.

أبو هلال، ماهر والعابد، عدنان والمطيري، معصومة. (2020). **البحوث والرسائل الجامعية في العلوم السلوكية**.. مكتبة الفلاح للنشر والتوزيع، الكويت.

إسماعيل، محمد. (2015). أثر عدد بدائل الاستجابة في مقياس ليكرت على الخصائص السيكومترية للمقياس وقياس الاتجاهات: دراسة تطبيقية على متدربي معهد الإدارة العامة، المملكة العربية السعودية. **الإدارة العامة، 55**(4)، 833-875.

تيغزة، أمحمد. (2017). توجيهات حديثة في تقدير صدق وثبات درجات أدوات القياس: تحليل نظري تقويمي وتطبيقي. **مجلة العلوم النفسية والتربوية، 3**(1)، 7-29.

الجراح، خلدون. (2015). **أثر حجم العينة وصعوبة الفقرة وتمييزها على دالة المعلومات والخطأ حسب نظرية الاستجابة للفقرة: النموذج ثنائي المعلمة**]رسالة ماجستير غير منشورة[. كلية العلوم التربوية، جامعة مؤتة، الأردن.

حسين، محمد. (2007). أثر اتجاه المفردة وعدد بدائل الاستجابة على ثبات أدوات القياس من نوع «ليكرت». **المجلة المصرية للدراسات النفسية، 17**(56)، 353-392.

زارع، نسرين. (2021). مقارنة معاملات ثبات درجات الاختبار في ظل مجموعة من الاشتراطات: دراسة محاكاة مونت كارلو. **مجلة كلية التربية بسوهاج، 88**(88)، 1107-1174.

زارع، نسرين. (2022). تأثير اختلاف طريقة التحليل الإحصائي تبعا لنوع البيانات الفترية والرتبية على الخصائص السيكومترية لمقياس ليكرت: دراسة محاكاة. **مجلة العلوم التربوية والدراسات الإنسانية**، (22)، 238- 266.

الزبون، حابس. (2013). أثر حجم العينة على تقدير دالة المعلومات للاختبار والخطأ المعياري في تقديرها باستخدام النظرية الحديثة في القياس. **مجلة جامعة النجاح للأبحاث-العلوم الإنسانية، 27**(6)، 1313-1334.

زكري، علي. (2014). أثر اختلاف حجم العينة على فحص الثباتية لاختبار (أوتيس- لينون) للقدرة العقلية المستوى المتوسط الصورة (j) باستخدام نموذج راش. **رسالة الخليج العربي، 35**(131)، 37-66.

الضوي، محسوب. (2011). تحري تأثير الدرجات المتطرفة وعدد فئات الاستجابة على تقدير معامل ألفا كرونباخ. **مجلة كلية التربية، 27**(1)، 117-175.

الطراونة، نادية. (2017). **أثر حجم العينة في تقدير معامل الثبات في ضوء نظرية التعميم** [رسالة ماجستير غير منشورة]. كلية العلوم التربوية، جامعة مؤتة، الأردن.

عبد السميع، محمد. (2017). تأثير عدد فئات الاستجابة وعدد المشاركين على دقة قيم معاملي ألفا وأوميجا في تقدير ثبات درجات المقياس النفسي. **المجلة المصرية للدراسات النفسية، 27**(96)، 317-384.

علام، رجاء. (2011). **مناهج البحث العلمي في العلوم النفسية والتربوية** ط7. دار النشر للجامعات، القاهرة، مصر.

علام، صلاح الدين. (2006). **القياس والتقويم التربوي والنفسي: أساسياته وتطبيقاته وتوجهاته المعاصرة**. دار الفكر العربي، القاهرة، مصر.

علاونة، معزوز. (2016). أثر حجم العينة ونمط الفقرة وجنس الطلبة على الخصائص السيكومترية لاختبار تحصيلي في الرياضيات لدى طلبة الصف الحادي عشر في المدارس الحكومية الفلسطينية. **مجلة مجمع**، (17)، 426-473.

علي، حجاج وحسن، ياسر. (2011). تأثير عدد بدائل ليكرت على الخصائص السيكومترية للمقياس النفسي وافتراضات التصميم العاملي ثنائي الاتجاه. **مجلة الإرشاد النفسي**، (29)، 56-131.

العمري، حسان. (2018). المقارنة بين ثلاث طرائق في تقدير ثبات الاختبارات المركبة التي تتضمن نوعية من الفقرات (ألفا، ألفا الطبقي، راجو). **مجلة اتحاد الجامعات العربية للتربية وعلم النفس، 16**(2)، 85-102.

عودة، أحمد. (2010). **القياس والتقويم في العملية التدريسية**. دار الأمل، إربد، الأردن.

كريش، أحمد. (2018). معامل ألفا الرتبي: تقدير معامل ثبات درجات الاختبار باستخدام البيانات الرتبية. **مجلة العلوم النفسية والتربوية، 4**(1)، 10-23.

اللصاصمة، شريفة. (2018). **أثر حجم العينة وطريقة التعامل مع القيم المفقودة على ثبات الاختبار ومعاملات تمييز وصعوبة الفقرات**]رسالة ماجستير غير منشورة[. كلية العلوم التربوية، جامعة مؤتة، الأردن.

النعيمي، ضرغام. (2012). تأثير حجم ومستوى الفئة العمرية على معامل الثبات لبعض الاختبارات البدنية ذات الأداء الأقصى. **مجلة جامعة كركوك للدراسات الإنسانية، 7**(3)، 1458-1473.

يحيى، جهاد. (2009). **أثر بعض المتغيرات السياقية على المعرفة الرياضية لدى معلمي الصف الثامن وتحصيل طلابهم في الرياضيات في محافظة قلقيلية**]رسالة ماجستير غير منشورة[. جامعة النجاح الوطنية كلية الدراسات العليا، فلسطين.

ثانيًا:

References:

Abu Hashim, S. (2006). *Psychometric properties of measurement tools in psychological and educational research using the SPSS program. Educational Research Center* (in Arabic), College of Education, King Saud University.

Abu Hilal, M., Al-Abed, A., & Al-Mutairi, M. (2020). R*esearch and theses in behavioral sciences* (in Arabic). Al-Falah Library for publication and distribution. Kuwait.

Abdel Samie, M. (2017). The effect of the number of response categories and the number of participants on the accuracy of the Alpha and Omega Coefficient values in estimating the stability of psychological scale scores. *Egyptian Journal of Psychological Studies*, *27*(96), 317-384.

Al Nuaimi, D. (2012). The effect of the size and level of age group and type of data on the stability coefficient of some maximum performance physical tests (in Arabic). *Carcook University Journal of the Humanities, 7*(3), 1458-1473.

Alawneh, M. (2016). The effect of sample size, paragraph style, and students' gender on the psychometric characteristics of an achievement test in mathematics for eleventh grade students in Palestinian public schools (in Arabic). *Majma Magazine*, (17), 426-473.

Al-Dawy, M. (2011). An investigation of the effect of extreme scores and the number of response categories on the estimation of the Alpha Cronbach coefficient (in Arabic). *Journal of the College of Education, 27*(1), 117-175.

Ali, H. & Hassan, Y. (2011). Effect of Likert number of substitutions on the psychometric properties of the psychometric scale and assumptions of two-way factorial design (in Arabic). *Journal of Psychological Counseling*, (29), 56-131.

Al-Jarrah, K. (2015). *The effect of sample size, paragraph difficulty and its distinction on the information function and standard error according to the paragraph response theory: the two-parameter model [Unpublished master's thesis]* (in Arabic). Mutah University, Jordan.

Allam, R. (2011). *Scientific research methods in psychological and educational sciences*. Universities Publishing House, Cairo, Egypt.

Allam, S. (2006). *Educational and psychological measurement and evaluation: Its basics, applications and contemporary trends* (in Arabic). Dar El Fikr El Arabi, Cairo, Egypt.

Al-Lasasmeh, S. (2018). *The effect of sample size and the method of dealing with missing values on test stability, discrimination coefficients, and difficulty of paragraphs [Unpublished master's thesis]* (in Arabic). Mutah University, Jordan.

Almehrizi, R. S. (2013). Coefficient alpha and reliability of scale scores. *Applied Psychological Measurement*, *37*(6), 438-459. https://doi.org/10.1177/0146621613484983

——. (2021). Coefficient β As Extension of KR-21 Reliability for Summed and Scaled Scores for Polytomously-scored Tests. *Applied Measurement in Education*, *34*(2), 139-149. https://doi.org/10.1080/08957347.2021.1890740

——. (2022). Reconceptualization of Coefficient Alpha reliability for test summed and scaled scores. *Educational Measurement: Issues and Practice*, *41*(3), 38-47.

Al-Omari, H. (2018). Comparison of three methods in estimating the reliability of composite tests that include the quality of the paragraphs (alpha, alpha stratification, Raju) (in Arabic). *Journal of Association of Arab Universities for Education and Psychology, 16*(2), 85-102.

Al-Zaboun, H. (2013). The effect of sample size on estimating the information function of the test and the standard error in estimating it using the modern theory of measurement. *An-Najah University Journal of Research-Human Sciences*, *27*(6), 1313-1334.

Barnette, J. J. (2010). Likert Scaling In: *Encyclopedia of Research Design*. Thousand Oaks, CA: SAGE Publications, Inc.

Bonett, D. G. (2002). Sample size requirements for testing and estimating Coefficient Alpha. *Journal of Educational and Behavioral Statistics*, *27*(4), 335-340.

Cho, E. (2016). Making reliability reliable: A systematic approach to reliability coefficients. *Organizational Research Methods*, *19*(4), 651-682. https://doi.org/10.1177/1094428116656239

Choudhury, S. & Bhattacharjee, D.(2014). Optimal number of scale points in Likert scales for quantifying compulsive buying behavior. *Asian Journal of Management Research*, *4*(3), 431-440.

Cronbach, L. J. (1951). Coefficient alpha and internal structure of tests. *Psychometrika, 16,* 297-334

Cronbach, L. J., Schönemann, P., & McKie, D. (1965). Alpha coefficients for stratified-parallel tests. *Educational and Psychological Measurement*, *25*(2), 291-312. https://doi.org/10.1177/001316446502500201

Customer, T. (2013). The effect of the sample size on the estimation of the information function of the test and the standard error in its estimation using the modern theory of measurement (in Arabic). An-Najah University Journal for Research-Human Sciences, 27(6), 1313-1334.

Danner, D., Blasius, J., Breyer, B., Eifler, S., Menold, N., Paulhus, D. L., ... Ziegler, M. (2016). Current Challenges, New Developments, and Future Directions in Scale Construction. *European Journal of Psychological*

Assessment, 32(3), 175-180. https://doi.org/10.1027/1015-5759/a00037

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72-91. https://doi.org/10.1037/a0032138

Green, S. B., & Hershberger, S. L. (2000). Correlated errors in true score models and their effect on coefficient alpha. *Structural Equation Modeling*, *7*(2), 251-270. https://doi.org/10.1207/S15328007SEM0702_6

Harwell, M. (2019). A strategy for using bias and RMSE as outcomes in Monte Carlo studies in statistics. *Journal of Modern Applied Statistical Methods*, *17*(2), 2-16. https://doi.org/10.22237/jmasm/1551907966

Hussein, M. (2007). The effect of item direction and the number of response alternatives on the stability of Likert-type measuring instruments (in Arabic). *Egyptian Journal of Psychological Studies, 17*(56), 353-392.

Ismail, M. (2015). The effect of the number of response alternatives in the Likert scale on the psychometric properties of the scale and measuring attitudes: An applied study on the trainees of the Institute of Public Administration, Saudi Arabia (in Arabic). *Public Administration, 55*(4), 833-875.

Johanson, G. A., & Brooks, G. P. (2010). Initial scale development: sample size for pilot studies. *Educational and psychological measurement*, *70*(3), 394-400. https://doi.org/10.1177/0013164409355692

Kline, P. (1986). *A handbook of test construction (psychology revivals): Introduction to psychometric design*. Routledge.

Krish, A. (2018). Ordinal Alpha Coefficient: Estimating the stability coefficient of test scores using ordinal data (in Arabic). *Journal of Psychological and Educational Sciences, 4*(1), 10-23.

Lee, J. & Paek, I. (2014). In search of the optimal number of response categories in a rating scale. *Journal of Psychoeducational Assessment*, *32*(7), 663-673. https://doi.org/10.1177/0734282914522200

Lord, F. M. & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score equatings. *Applied Psychological Measurement, 8*(4), 453-461. https://doi.org/10.1177/014662168400800409

Lozano, L. M., García-Cueto, E., & Muñiz, J. (2008). Effect of the number of response categories on the reliability and validity of rating scales. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, *4*(2), 73. https://doi.org/10.1027/1614-2241.4.2.73

Maydeu-Olivares, A., Kramp, U., García-Forero, C., Gallardo-Pujol, D., & Coffman, D. (2009). The effect of varying the number of response alternatives in rating scales: Experimental evidence from intra-individual effects. *Behavior Research Methods*, *41*(2), 295-308. https://doi.org/10.3758/BRM.41.2.295

McComas, W. F. (2014). *Trends in international mathematics and science study (TIMSS)*. The Language of Science Education: An Expanded Glossary of Key Terms and Concepts in Science Teaching and Learning.

Mehrens, W. A. & Lehmann, I. J. (1978). Measurement and evaluation in education and psychology: A Qualitative approach.

Morera, O. F. & Stokes, S. M. (2016). Coefficient α as a measure of test score reliability: Review of 3 popular misconceptions. *AmericanJournalofPublicHealth*, *106*(3),458-461.https://doi.org/10.2105/AJPH.2015.302993

Novick, M. R. & Lewis, C. (1967). Coefficient Alpha and the reliability of composite measurements. *Psychometrika*, *32*(1),1-13.https://doi.org/10.1002/j.2333-8504.1966.tb00356.x

Odeh, A. (2021). *Measurement and evaluation in the teaching process* (in Arabic). Dar alamal, Irbid, Jordan.

Preston, C. C. & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, *104*(1), 1-15. https://doi.org/10.1016/S0001-6918(99)00050-5

Revicki, D. (2014). Internal consistency reliability. In *Encyclopedia of Quality of Life and Well-Being Research,* pp. 3305–3306 [Michalos, AC, editor]. Dordrecht, The Netherlands: Springer Netherlands.

Shaftel, J., Nash, B. L., & Gillmor, S. C. (2012, April). Effects of the number of response categories on rating scales. In *Proceedings of the annual conference of the American Educational Research Association* (pp.1-24).

Streiner, D. L. (2003). Starting at the beginning: An introduction to Coefficient Alpha and internal consistency. *Journal of Personality Assessment*, *80*(1), 99-103.

Tarawneh, N. (2017). *The effect of sample size in estimating the stability coefficient in the light of generalization theory [Unpublished master's thesis]* (in Arabic). Mutah University, Jordan.

Tighza, M. (2017). Recent directions in estimating the validity and reliability of the scores of measurement tools: A theoretical, evaluative and applied analysis (in Arabic). *Journal of Psychological and Educational Sciences, 3*(1), 7-29.

TIMSS & PIRLS (2019). *International Study Center*. Boston College. Retrieved from: https://timss2019.org/international-database/

Traxler, K. (2017). *Estimating Bias in Multilevel Reliability Coefficients: A Monte Carlo Simulation*. University of Northern Colorado.

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's Alpha reliability in realistic conditions: Congeneric and asymmetrical measurements. *Frontiers in Psychology*, *7*,1-8. https://doi.org/10.3389/fpsyg.2016.00769

Wakita, T., Ueshima, N., & Noguchi, H. (2012). Psychological distance between categories in the Likert scale: Comparing different numbers of options. *Educational and Psychological Measurement*, *72*(4), 533-546. https://doi.org/10.1177/0013164411431162

Weijters, A. J. M. M., & Ribeiro, J. T. S. (2011, April). Flexible heuristics miner (FHM). In *2011 IEEE symposium on computational intelligence and data mining* (pp. 310-317).

Yahya, J. (2009). *The effect of some contextual variables on the mathematical knowledge of eighth grade teachers and their students' achievement in mathematics in Qalqilya Governorate [Unpublished master's thesis]* (in Arabic). Palestine.

Yang, Y. & Green, S. B. (2011). Coefficient Alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, *29*(4), 377-392. https://doi.org/10.1177/0734282911406668

Zakry, A. (2014). The effect of the difference in sample size on the reliability test (Otis-Lennon) for the intermediate level mental ability form (j), using the Rasch model (in Arabic). *Arabian Gulf Message, 35* (131), 37-66.

Zaree, N. (2021). Comparison of test scores reliability coefficients under a set of variables: A Monte Carlo simula-

tion study (in Arabic). *Journal of the College of Education in Sohag, 88* (88), 1107-1174.

——. (2022). The effect of different statistical analysis methods depending on the type of interval and ordinal data on the psychometric properties of the Likert scale: A simulation study (in Arabic). *Journal of Educational Sciences and Human Studies*, (22), 238-266.

Zumbo, B. D. (2006). Validity: Foundational issues and statistical methodology. *Handbook of Statistics*, *26*, 45-79. https://doi.org/10.1016/S0169-7161(06)26003-6

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of Coefficients Alpha and Theta for Likert rating scales. *Journal of Modern Applied Statistical Methods*, *6*(1), 21-29. https://doi.org/10.22237/jmasm/1177992180

تصريحات ختامية:

- يصرح المؤلف/ المؤلفون بالحصول على موافقة الأشخاص المتطوعين للمشاركة في الدراسة وعلى الموافقات المؤسسية اللازمة.
- تتوفر البيانات الناتجة و/ أو المحلَّلة المتصلة بهذه الدراسة من المؤلف المراسل عند الطلب.

Final declarations:

- The authors declare that they got the required voluntary human participants consent to participate in the study as well as the necessary institutional approvals.
- The datasets generated and/or analyzed during the current study are available from the corresponding author upon reasonable request.