المجلة الأردنية في العلوم التربوية، مجلد 21، عدد 2، 2025، 239-253

The effect of sample size and response grading on estimates of the scorer's general beta reliability coefficient for diverse types of transformed and raw scores

***Ahmed bin Saleh bin Ali Al-Saadi** ***, College of Education, Sultan Qaboos University, Sultanate of Oman

Rashid bin Saif Al-Mahrazi, College of Education, Sultan Qaboos University, Sultanate of Oman

Yousef Abdul Qader Abu Shandi, College of Education, Sultan Qaboos University, Sultanate of Oman

Received: 31/7/2024
Accepted: 20/10/2024
Published: 30/6/2025

***Corresponding author:**
ahmedsalehalsaadi8@gmail.com

Citation: Al-Saadi, A. B. S. B. A., Al-Mahrazi, R. B. S., & Abu Shindi, Y. A. Q. (2025). The effects of sample size and response rating scales on estimating Almehrizi generalized alpha reliability coefficient for different scaled scores. Jordan Journal of Educational Sciences, 21(2), 239–253. https://doi.org/10.47015/21.2.6

© 2025 Publishers / Yarmouk University. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license https://creativecommons.org/licenses/by-nc/4.0/

Abstract

Objectives: The study aimed to examine the effect of sample size and response grading on estimates of the scorer's general beta reliability coefficient for different types of transformed and raw scores using real data measuring student attitudes toward mathematics in TIMSS in 2019.
Methodology: The study sample consisted of 5,117 male and female fourth-grade students in the Sultanate of Oman. Three types of data were generated by modifying the number of response grading categories in the scale (four-point grading, three-point grading, and binary grading). For each grading type, 100 subsamples were randomly selected using three different subsample sizes (30, 200, and 1,000). Five types of transformed scores were calculated: raw score, percentile rank, standard score, first-TIMES score, and second-TIMES score, and the coefficient was calculated. The overall beta reliability of the scorers on the five transformed scores for all data was compared using descriptive statistics, bias ratios, and root mean square error (RMSE) values.
Results: The study reached a high degree of accuracy in the estimates of the general beta reliability coefficient for the scorers across all transformed grades, as the average values of the reliability coefficient estimates for all grades in the random samples were very close to the reliability coefficients of the main sample. The bias ratios and RMSE values were small and close to zero, including for the small sample size (30).
Conclusion: The study recommended expanding the use of the general beta reliability coefficient to calculate the stability of all types of transformed and raw scores in research studies and various assessment tools.

Keywords: Generalized Beta reliability coefficient, sample size, response rating scale, scale scores, TIMSS.

أثر حجم العينة وتدريج الاستجابة على تقديرات معامل ثبات بيتا العام للمحرزي لأنواع مختلفة من الدرجات المحولة والخام

أحمد بن صالح بن علي السعدي، كلية التربية، جامعة السلطان قابوس، سلطنة عمان

راشد بن سيف المحرزي، كلية التربية، جامعة السلطان قابوس، سلطنة عمان

يوسف عبد القادر أبو شندي، كلية التربية، جامعة السلطان قابوس، سلطنة عمان

الملخص

الأهداف: هدفت الدراسة إلى فحص تأثير حجم العينة وتدريج الاستجابة على تقديرات معامل ثبات بيتا العام للمحرزي لأنواع مختلفة من الدرجات المحولة والخام باستخدام بيانات حقيقية في مقياس اتجاهات الطلبة نحو الرياضيات فيTIMSS في عام 2019م.
المنهجية: تكونت عينة الدراسة من 5117 طالبًا وطالبة بالصف الرابع في سلطنة عمان، تم تكوين ثلاثة أنواع من البيانات من خلال تعديل عدد فئات تدريج الاستجابة في المقياس (تدريج رباعي –تدريج ثلاثي –تدريج ثنائي)، وفي كل تدريج تم اختيار عينات فرعية بعدد 100 عشوائيًا باستخدام ثلاثة أحجام مختلفة للعينة الفرعية (30– 200– 1000)، كما

© حقوق الطبع محفوظة لجامعة اليرموك، إربد، الأردن، 2025.

تم حساب خمسة أنواع من الدرجات المحولة: الدرجة الخام –الرتبة المئينيه –الدرجة المعيارية –درجة تيمز الأولى والثانية، وتم حساب معامل ثبات بيتا العام للمحرزي على الدرجات المحولة الخمس لجميع البيانات وتمت مقارنتها باستخدام الإحصاءات الوصفية، ونسب التحيز، وقيم جذر متوسطات مربع الخطأ (RMSE).

النتائج: توصلت نتائج الدراسة إلى درجة عالية في دقة تقديرات معامل ثبات بيتا العام للمحرزي في كل الدرجات المحولة، حيث كان متوسط قيم تقديرات معامل الثبات لجميع الدرجات في العينات العشوائية قريبة جداً من معاملات ثبات العينة الرئيسة، وكانت نسب التحيز وقيم RMSE صغيرة وقريبة من الصفر بما فيها حجم العينة الصغير (30).

الخلاصة: التوسع في استخدام معامل ثبات بيتا العام لحساب ثبات جميع أنواع الدرجات المحولة والخام في الدراسات البحثية ومختلف أدوات التقييم.

الكلمات المفتاحية: معامل ثبات بيتا العام، حجم العينة، تدريج الاستجابة، نوع الدرجة المحولة، TIMSS..

الكلمات المفتاحية: معامل ثبات بيتا العام، حجم العينة، تدريج الاستجابة، نوع الدرجة المحولة، TIMSS.

المقدمة

استقطبت الخصائص السيكومترية للاختبارات اهتمام العلماء، ومن بينهم علماء القياس النفسي والتربوي على مر العصور الماضية، ففي ضوء نتائجها يتم اعتماد نتائج الاختبارات في اتخاذ قرارات بشأن الأفراد أو الجماعات الذين يطبق عليهم تلك الاختبارات، ومن أهمها الصدق والثبات، فينبغي أن تتسم أدوات القياس بقدرٍ كبيرٍ من هاتين الخاصيتين، وذلك من أجل الوصول لنتائج سليمة ودقيقة وخالية من التحيز والتشويه.

والثبات من أهم الخصائص السيكومترية للاختبار، ويُعد سبيرمان مؤسس نظرية الثبات، فقد قدم عدة مفاهيم منها مفهوم الدرجة الحقيقية وخطأ القياس، فيُعرف الثبات على أنه مربع الارتباط بين الدرجة الحقيقية والدرجة المشاهدة، وكذلك يُعرف على أنه نسبة تباين الدرجة الحقيقية للاختبار إلى تباينه الإجمالي Streiner, (2003).

نماذج ثبات القياس وافتراضاتها

تعد نماذج ثبات القياس ذات أهمية بالغة في دقة معاملات ثبات درجات الاختبار، وهناك أربع نماذج للقياس، متمثلة في النموذج المتوازي، ونموذج تاو المتكافئ، ونموذج تاو المتكافئ في الأساس، والنموذج التقاربي ويتم دمج نموذج تاو المتكافئ ونموذج تاو المتكافئ في الأساس في نموذج واحد ليصبح عدد نماذج ثبات القياس ثلاثة نماذج (Dunn et al., 2014).

النموذج المتوازي

يعد هذا النموذج أكثر تقيداً عن بقية نماذج ثبات القياس وذلك بسبب افتراضاته، حيث يفترض أنّ بنود الاختبار كلها تقيس متغيراً كامنًا وحيداً (أحادية البعد)، ويقوم على افتراض أن الدرجات الحقيقية لكل بند من بنود الاختبار متساوية، وكذلك يفترض تساوي تباين درجات الخطأ لجميع البنود المكونة للاختبار (Graham, 2006).

نموذج تاو المتكافئ في الأساس

وتقوم افتراضات هذا النموذج على أنّ بنود المقياس كلها تقيس متغير كامنًا وحيداً (أحادية البعد)، ويفترض تساوي الدرجات الحقيقية، وعدم تساوي التباين في درجة الخطأ، ويسمح باختلاف قيم الدرجات الحقيقية للبنود بفارق قيم ثابتة (Raykov, 2001).

وتعد معادلة ألفا ضمن هذا النموذج، الذي يتحرر من شرط تساوي متوسطات وتباينات درجة الخطأ، ولكنه يتطلب أن يكون هناك تساوٍ في متوسطات وتباينات الدرجات الحقيقية في الأجزاء التي تكون الدرجة الكلية (Graham, 2006).

النموذج التقاربي

ويعد هذا النموذج أكثر النماذج مرونة، وذلك بسبب افتراضاته، حيث يفترض إمكانية تشبع بنود الاختبار على العامل الكامن بقيم تشبعات مختلفة، والتنوع في مستوى تباين الخطأ، ويعتبر معامل ثبات أوميجا مثالًا عليه (Graham, 2006).

طرق تقدير الثبات

يذكر ستريبر (Streiner, 2003) أن أكثر الطرق استخداماً لتقدير ثبات الاتساق الداخلي هي طريقة معامل ثبات كرونباخ ألفا الذي يعد امتدادا لمعامل ثبات كيودر ريتشاردسون 20، والذي تستخدم في الاختبارات ذات البنود المتجانسة، أي التي تقيس سمة واحدة، حيث يتم تقسيم الاختبار إلى عدد من المحاور أو الأجزاء مساوية لعدد مفرداته، لقد شاع استخدام معامل ألفا لدى الباحثين والمهتمين في أبحاث القياس النظرية والتطبيقية في تقدير ثبات الاتساق الداخلي لدرجات الاختبار منذ بداية ظهوره، وذلك لقلة

افتراضاته وإمكانية تعميمه وسهولة حسابه، حيث يفترض معامل ألفا أنّ نماذج الاختبار أو أسئلة الاختبار هي في الأساس مكافئ تاو الذي يقيس نفس السمة على نفس المقياس، ولكنها تختلف في تباين درجة الخطأ وتتشابه في تباين الدرجة الحقيقية، وهذا سوف يسمح إلى اختلاف الأسئلة في متوسطاتها الحسابية وانحرافاتها المعيارية (Graham, 2006)، وهذا المعادلة الشائعة لمعامل ثبات ألفا:

$$\hat{\alpha} = \frac{K}{K-1}\left(1 - \frac{\sum_i \hat{\sigma}_i^2}{\hat{\sigma}^2(X)}\right) = 1 - \frac{\hat{\sigma}_\alpha^2(E)}{\hat{\sigma}^2(X)}$$

حيث إنّ K يمثل عدد أسئلة الاختبار، ويمثل σ^_i^2 تباين درجات السؤال الواحد، ويمثل σ^2 (X) تباين الدرجة الكلية للاختبار، ويمثل σ^_α^2 (E) تباين الخطأ في نموذج تاو، ويحسب كالتالي:

$$\hat{\sigma}_\alpha^2(E) = \frac{K}{K-1}\left(\sum_i \hat{\sigma}_i^2 - \frac{1}{K}\hat{\sigma}^2(X)\right)$$

ويذكر المحرزي (Almehrizi, 2022) بأن معامل ثبات ألفا هو امتداد لمعادلة كيودر وريتشاردسون (KR-20)، ويستخدم لتقدير ثبات الدرجة الخام للاختبار والتي تمثل المجموع غير الموزون لدرجات أسئلة الاختبار، ولقد وضع المحرزي (Almehrizi, 2022) معامل ثبات ألفا في صيغة قابلة لتعميمه لتقدير ثبات الدرجات المحولة للدرجات الخام مثل الرتب المئينية والدرجات المعيارية وغيرها كالتالي:

$$\hat{\alpha} = 1 - \frac{\hat{\sigma}_\alpha^2(E)}{\hat{\sigma}^2(X)} = 1 - \frac{\frac{N}{(N-1)}\frac{K}{(K-1)}\left[\frac{\sum_p \sum_i (X_{pi}-\bar{X}_p)^2}{N} + \frac{\sum_i \sum_p (X_{pi}-\bar{X}_i)^2}{N} - \frac{\sum_i \sum_p \left(X_{pi}-\bar{X}\right)^2}{N}\right]}{\hat{\sigma}^2(X)}$$

ويمثل N حجم عينة الطلبة، ويمثل X_pi درجة كل طالب في كل سؤال، ويمثل X¯_i المتوسط الحسابي لدرجات الطلبة في كل سؤال، ويمثل X¯_p المتوسط الحسابي لدرجات كل طالب، ويمثل X¯ المتوسط الحسابي للاختبار.

معامل ثبات بيتا

في الفترة المنصرمة تم تطوير العديد من معاملات الثبات للاختبارات حسب افتراضاتها في النظرية الكلاسيكية، ولم تحْظَ معادلة (KR-21) باهتمام الباحثين مقارنة بمعادلة (KR-20) والتي نالت اهتماماً بالغًا من قبلهم، حيث يرجع السبب بأن معادلة (KR-21) تعطي ثباتاً أقل من (KR-20)، وكذلك تفترض تساوي فقرات الاختبار من ناحية الصعوبة وهذا يقلل من استخدامها (Almehrizi, 2021).

ولذلك قدم المحرزي (Almehrizi, 2021) معامل بيتا (β) فهي امتداد لمعادلة كيودر ريتشاردسون (KR-21)) في تقدير ثبات الدرجات الخام للاختبار، وبنفس افتراضات معامل (KR-21)، وممكن تطبيقه على الاختبارات متعددة الاستجابة، وليس فقط على الاختبارات ثنائية الاستجابة القائمة، ويتم حساب معامل ثبات بيتا للدرجات الخام (X) في الاختبار من خلال المعادلة الآتية:

$$\hat{\beta} = 1 - \frac{\hat{\sigma}_\beta^2(E)}{\hat{\sigma}^2(X)} = 1 - \frac{\frac{K}{K-1}\left[\frac{1}{N}\sum_p \sum_i (X_{pi}-\bar{X}_p)^2\right]}{\hat{\sigma}^2(X)}.$$

معامل ثبات بيتا العام

قدم المحرزي (Almehrizi,2021) معامل ثبات بيتا العام (Generalized Beta (Gβ)) فهو امتداد لمعادلة بيتا للدرجة الخام من أجل تقدير ثبات الدرجات المحولة للاختبار، ويتم حساب β^ للدرجات المحولة (S) المقابلة للدرجات الخام (X) في الاختبار من خلال المعادلة الآتية:

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

ويمثل $\widehat{\sigma}^2(S)$ تباين الدرجة المحولة، ويحسب بهذه المعادلة $\widehat{\sigma}^2(S) = \frac{\sum_p (S_p - \bar{S})^2}{N-1}$، في حين يمثل S_p الدرجات المحولة الملاحظة للممتحن، ويمثل σ^_p^2 (S) تباين الخطأ للدرجات المحولة لكل ممتحن، ويمثل E_p (S) الدرجة المتوقعة للدرجات الخام لكل ممتحن واللذان يحسبان كالتالي:

$$E_p(S) = \sum_X S_X\, f(X|\bar{X}_p)$$

$$\hat{\sigma}_p^2(S) = \sum_X \left(S_X - E_p(S)\right)^2 f(X|\bar{X}_p)$$

ويمثل $f(X|\bar{X}_p)$ التكرار النسبي المشروط للدرجات الخام باستخدام استجابات الطالب الواحد على جميع أسئلة الاختبار (X¯_p) ويتم حسابه من خلال تطبيق صيغة ارتدادية قدمها لورد وونجرسكي (Lord & Winger sky, 1984)، حيث تتطلب الصيغة الارتدادية حساب التكرارات النسبية للاستجابات باستخدام استجابات الطالب الواحد على جميع الأسئلة (X¯_p)، ويتم تطبيق هذه الصيغة لكل طالب على حده وتكرر بعدد الطلبة في عينة الدراسة (N)، ويمكن الرجوع إلى الصيغة الارتدادية وطريقة تطبيقها في دراسة (Almehrizi, 2022).

العوامل المؤثرة على معاملات الثبات في ضوء النظرية الحديثة للقياس

حجم العينة

يطرح العديد من الباحثين المهتمين في تطوير الاختبارات، تساؤلات عن الحجم المناسب للعينة، فليس من السهل الإجابة عن هذا التساؤل، وذلك بأن تحديد حجم العينة يرتبط بعدة عوامل كالتكلفة المادية، وتوفر أفراد العينة، والجهد، والوقت (Krish, 2018).

وجذب تحديد حجم العينة أو عدد المفحوصين على الاختبار اهتمام العديد من الباحثين والمختصين في القياس النفسي والتربوي ولم يكن وليد اليوم واختلفت الآراء فيه، حيث يتفق كلٌ من كلاين (Kline, 1986)؛ ونونالي وبرنستاين (Nunnally & Bernstein, 1994)؛ سيجال (Segall, 1994) بأن لا يقل حجم

العينة عن 300 مفحوص في حين أشار تشارتر (Charter, 2003) بأن لا يقل حجم العينة عن 400 مفحوص لحساب ثبات الاختبار، وبرر ذلك بأن هناك علاقة بين عدد المفحوصين وثبات الاختبار، فكلما قل عدد المفحوصين فإن هذا يؤدي إلى عدم استقرار معامل ثبات الاختبار، ويوصى جوهانسون وبروكس (Johanson & Brooks, 2010) بأن يكون أصغر حجم للعينة في الدراسة الاستطلاعية 30 مشاركًا.

تدريج الاستجابة

يعد تدريج الاستجابة عاملاً مهمًا في تحديد الخصائص السيكومترية للاختبار، ومؤشرًا رئيسًا في تقدير معاملات الثبات في الدراسات التي تناولته، وهذا ما يؤكده شافتيل وآخرون (Shaftel et al., 2012) بأن اختيار تدريج الاستجابة من أهم الاعتبارات في تصميم الاختبار، وخاصة التي تعتمد على أسلوب ليكرت، وقد أشارت عدد من الدراسات كدراسة: إسماعيل (Ismail, 2015)؛ وحسين (Hussein, 2007)؛ ولي بايك (Lee & Paek, 2014) إلى مدى تأثير هذا المتغير على تقدير معامل الثبات، وأن معامل الثبات يزداد بزيادة تدريج الاستجابة.

ويذكر دانر وآخرون (Danner et al., 2016) بأن تحديد تدريج الاستجابة من أكبر التحديات التي تعترض الباحثين المهتمين في تصميم أدوات القياس التربوية والنفسية، مما يستوجب المزيد من إجراء البحوث في هذا المجال.

ويعتقد بعض الباحثين بأن زيادة تدريج الاستجابة قد يربك المستجيبين ويشعرهم بالحيرة أثناء الاستجابة، ولذلك يتجه بعض الباحثين إلى التقليل من تدريج الاستجابة؛ لسهولة التصحيح وإدخال البيانات ثم تحليلها (Birkett, 1986).

ويؤكد كل من أبو حطب وصادق (Abu Hatab & Sadiq, 1991) أن يكون تدريج الاستجابة فرديًا بدلًا من أن يكون زوجيًا للحصول على قيمة ثبات عالية لاحتوائه على النقطة الوسيطة. وأكد الضوي (Al-Dawy, 2011) وعلي وحسن (Ali & Hassan, 2011) بأن قيمة معامل ألفا كان عاليًا في الأعداد الفردية لتدريج الاستجابة مقارنة بالأعداد الزوجية. وعلى النقيض أكدت دراسة كل من: شودري وبهاتاشارجي (Choudhury & Bhattacharjee, 2014)، بريستون وكولمان (Preston & colman, 2000)، زومبو وأخرون (Zumbo et al., 2007) بأن قيمة معامل الثبات تزداد بزيادة تدريج الاستجابة، في حين ترى دراسة لوزانو وأخرون (Lozano et al., 2008) إلى أن العدد المناسب لتدريج الاستجابة هو من (4-7) فئات، ونادرًا ما يزداد معامل الثبات في التدريجات الأعلى من التدريج السباعي، ويؤكد هذا واكيتا وأخرون (Wakita et al., 2012) على أن الباحثين يفضلون استخدام تدريج الاستجابة الذي يتراوح من (4-7) فئات.

الدرجات المحولة

تستخدم هذه الدرجات المحولة في العديد من الاختبارات كمعايير، فمثلًا تستخدم في اختبارات القبول بالكليات في الولايات المتحدة الأمريكية كاختبار الاستعداد الأكاديمي، وكذلك أيضا في اختبارات الذكاء كاختبار ويكسلر، واختبارات التصنيف العام للأفراد الملتحقين في السلك العسكري فهي ذات أهمية بالغة (Allam, 2006).

وتكمن أهمية الدرجات المحولة في أنها تقوم بتوحيد وحدات القياس، في حين وحدات القياس في الدرجة الخام كانت متعددة أي مختلفة، وتعمل على تحديد موقع الفرد من مجتمع العينة، وهذا يعطي تفسيرًا للدرجة الذي حصل عليها الفرد (Odeh, 2021).

وتعمل الميادين العلمية والاجتماعية وغيرها من العلوم على تحويل الدرجة الخام إلى درجة محولة، وذلك لتفسير درجة الاختبار من أجل اتخاذ القرار المناسب المبني على هذه التفسيرات، فلابد من التأكد من موثوقية هذه الدرجات المحولة، ولقد أكدت ذلك الجمعية الأمريكية للبحوث عام 2014 بأنه من الضروري التأكد من موثوقية هذه الدرجات المحولة، فعدم فحصها يسبب إعطاء قيم ثبات أكبر أو أصغر من الدرجات الخام (Almehrizi, 2013)، وأشارت دراستا المحرزي (Almehrizi, 2021; 2022) إلى تأثير نوع الدرجة المحولة على تقدير معاملات الثبات، من أجل الحصول على ثبات دقيق.

مشكلة الدراسة

يغفل الباحثون عن تفحص الافتراضات اللازمة لمعامل ألفا في البيانات قبل تطبيقه، مما يؤدي إلى الحصول على قيمة تبخس قيمته الحقيقية وتمثل الحد الأدنى له أو تضخم قيمته، وهذا ما أكده جرين وهيرشبرجر (Green & Hershberger, 2000) بأن معامل ألفا يعطي قيمة منخفضة في تقدير ثبات الدرجات في حالة انتهاك افتراضاته المتعلقة بتكافؤ تاو، أما إذا كانت الأخطاء مرتبطة فإن ذلك يؤدي إلى زيادة قيمته.

وأجريت العديد من الدراسات في فحص موثوقية معامل ألفا في تقدير الثبات، وذلك بمقارنته بمعاملات ثبات أخرى وهذا ما أشارت إليه عدد من الدراسات كدراسة كل من: العمري (AL-Omari, 2018)، وكريش (Krish, 2018)، وتريزانو- هيرموسيلا وألفارادو (Trizano-Hermosilla & Alvarado, 2016).

إن التعدي على هذه الافتراضات جعل الباحثين يتسارعون في تقديم مقترحات وصيغ معادلات أخرى لحساب قيمة الثبات، ومن بين هذه المعادلات معادلة بيتا العام، فهي تساعد الباحثين والمختصين في مجال البحث العلمي في كيفية إعداد اختبار، فمن أهم خطواته تحديد تدريج الاستجابة، وحجم العينة المناسبة لحساب الثبات، حيث إنه لا يوجد هناك اتفاق في تحديد هذا العدد الذي يعطي تقديرًا عاليًا للثبات، فبعض منهم يختار عددًا قليلاً من تدريج الاستجابة

وحجم العينة، اعتقاداً منهم بأنَّ هذا يوفر الوقت والجهد خلال فترة تصميم الاختبار، وهذا يرجع إلى عدم وجود خطوط إرشادية متفق عليها لتحديد تدريج الاستجابة وحجم العينة.

وهذا ما ذكره فايترز وأخرون (Weijters at el., 2011) بأنَّ بعض الباحثين يقومون بتحديد تدريج الاستجابة وحجم العينة أثناء تصميم الاختبار بالرغم من عدم وجود إرشادات تعينهم في اختيار العدد الأمثل لهذين المتغيرين للحصول على تقديرات ثبات تتسم بالدقة.

وتقتصر حدود إمكانيات معامل بيتا 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 على بيتا العام، لذا حاولت هذه الدراسة الإجابة عن الأسئلة الآتية:

1- ما الإحصاءات الوصفية للدرجات الخام والدرجات المحولة، ومعامل ثبات بيتا العام للمحرزي لنتائج مقياس اتجاهات الطلبة نحو الرياضيات في TIMSS بسلطنة عُمان تبعًا لنوع التدريج (رباعي، ثلاثي، ثنائي)؟

2- ما الإحصاءات الوصفية لمعامل ثبات بيتا 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 قيم جذر متوسطات مربع الخطأ لمعامل ثبات بيتا 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؟

متغيرات الدراسة

المتغيرات التابعة

معامل ثبات بيتا ومعامل ثبات بيتا العام (Almehrizi, 2021).

المتغيرات المستقلة

أ) الدرجات المحولة: الدرجة الخام –الدرجة المعيارية الاعتدالية – الرتب المئينية –التساعيات –درجة تيمز الأولى -درجة تيمز الثانية، ولم يتم استخدام هاتين الدرجتين حيث إنهما درجات تتواجد في التدريج الرباعي (العينة الرئيسة) فقط.

ب) تدريج الاستجابة: التدريج الرباعي -التدريج الثلاثي -التدريج الثنائي.

ج) حجم العينة: حجم صغير (30 طالب) -حجم متوسط (200 طالب) -حجم كبير (1000 طالب)

أهمية الدراسة

الأهمية النظرية

تنبثق الأهمية النظرية للدراسة من أنها توفر معلومات عن معامل جديد لتقدير الثبات وهو معامل ثبات بيتا العام، والذي يصلح لتقدير ثبات الدرجات الخام والدرجات المحولة، بخلاف معاملات الثبات المعتادة كمعامل ألفا كرونباخ، ومعامل بيتا اللذين يقتصران على تقدير ثبات الدرجات الخام فقط، وكذلك تتفحص جودة هذا المعامل في ضوء عدة متغيرات (حجم العينة، وتدريج الاستجابة، ونوع الدرجة المحولة).

الأهمية التطبيقية

يتوقع من هذه الدراسة توفير 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 المحولة، فهي سوف تتناول متغير حجم العينة بثلاثة أحجام مختلفة، وتدريج بثلاثة مستويات وخمسة درجات محولة.

مصطلحات الدراسة

حجم العينة (Sample size)

عرَّف علام (2011) العينة بأنها مجموعة جزئية من المجتمع يتم اختيارها بطريقة احتمالية، بحيث تمثل خصائص المجتمع.

تدريج الاستجابة (Response scale)

يتكون مقياس ليكرت من مجموعة من العبارات تقيس الاتجاهات نحو موضوع معين، ويطلب 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 تجمع هذه القيم العددية للحصول على الدرجة الكلية التي تعبر عن اتجاه الفرد نحو موضوع معين (علام، 2011).

الدرجة المحولة (Derived Scores)

هي قيم إحصائية رقمية تستخدم في تحديد المركز النسبي للفرد في توزيع ما؛ لتسهيل وصف أدائه بالنسبة لأقرانه في اختبار يقيس سمة ما (علام، 2006).

ثبات الاتساق الداخلي Internal consistency

يقوم على مدى اتساق استجابات مجموعة من الأفراد على مجموعة من الأسئلة (العبارات)، أي أنَّ العبارات كلها تقيس المفهوم أو المتغير ذاته، وتسمى هذه الحالة بأحادية البعد، بمعنى أن العبارات زادت أو نقصت تقيس مفهوماً واحداً (أبو هلال وآخرون، 2020).

ويتم حسابه إجرائيا في هذا البحث باستخدام معامل ثبات ألفا العام.

معامل ثبات بيتا العام Generalized Beta reliability coefficient:

معادلة عامة لمعامل ثبات بيتا تقدم تقديرا لثبات مختلف الدرجات المحولة التي يتم حسابها للدرجات الخام في تفسير نتائج المستجيبين في أدوات القياس مثل الرتب المئينيه والدرجات المعيارية وغيرها، وتم تطويره من قبل المحرزي (2021).

محددات الدراسة

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

منهجية الدراسة وإجراءاتها

مجتمع الدراسة وعينتها

استخدمت الدراسة بيانات أرشيفية من استجابات طلبة الصف الرابع الأساسي في المدارس الحكومية، والمدارس الخاصة والدولية بسلطنة عُمان في العام الدراسي 2020/2019م على مقياس اتجاهات الطلبة نحو الرياضيات، وتم الحصول على البيانات من الموقع الالكتروني https://timss2019.org/international-database/ ، وهو الموقع الرسمي للمنظمة الدولية لتقييم الإنجاز التعليمي IEA، وتم الحصول على البيانات في 3 مارس 2022، وبلغ حجم العينة 6814 طالبًا وطالبة، ونظراً لوجود بيانات قيم مفقودة في البيانات تم حذف الطلبة الذين لم يستجيبوا على أي من عبارات المقياس لحساب معاملات ثبات بيتا العام، وذلك لأن طبيعة معامل الثبات بيتا العام لا تتناسب مع أي بيانات مفقودة، وقد تم الاقتصار على عينة بحجم 5117 طالبًا وطالبة وقد بلغ عدد الذكور 2510 وعدد الإناث 2607.

أداة الدراسة

استخدمت الدراسة مقياس اتجاهات الطلبة نحو الرياضيات كأحد المقاييس المستخدمة مع اختبار TIMSS2019 لطلبة الصف الرابع الأساسي، حيث هدف المقياس إلى توفير مصادر غنية من المعلومات لصانعي القرار ومطوري المناهج والعاملين من أجل وضع الخطط التي تسهم في تحسين تعليم الرياضيات، ويتكون المقياس من ثلاثة محاور، فالمحور الأول يتضمن 9 عبارات وتقيس اتجاهات الطلبة نحو تعلم الرياضيات، والمحور الثاني يتضمن 6 عبارات وتقيس اتجاهات الطلبة نحو دروس الرياضيات، أما المحور الثالث، فيتضمن 9 عبارات وتقيس اتجاهات الطلبة نحو مادة الرياضيات، ولتحقيق هدف الدراسة تم الاقتصار على المحور الأول من استبانة الطالب، والذي يهتم باتجاهات الطلبة نحو تعلم الرياضيات للصف الرابع الأساسي، ويتم الحصول على استجابة الطالب باستخدام تدريج رباعي لدرجة الموافقة (أوافق كثيراً = 4 درجات، أوافق قليلاً = 3 درجات، أرفض قليلًا = 2 درجتان، أرفض كثيراً = 1 درجة)، وكانت عبارات المقياس كلها موجبة ما عدا كل من العبارة الثانية والثالثة كانت سالبة .(TIMSS & PIRLS, 2019)

ويتمتع المحور الأول (اتجاهات الطلبة نحو تعلم الرياضات) للمقياس بثبات وصدق في جميع الدول التي تم تطبيقه عليها، وتم الحصول على مؤشرات الصدق والثبات من دليل المقياس المتوفر في الموقع الإلكتروني https://timss2019.org/international-database/ التابع للمنظمة الدولية لتقييم الإنجاز التعليمي IEA، حيث بلغت قيمة معامل ثبات ألفا للمحور الأول في سلطنة عُمان 0.80، وأظهر تحليل المكونات الرئيسية (Principal component) للمقياس أن تشبعات عبارات هذا المحور في سلطنة عُمان تراوحت بين 0.27 و0.84، وتفسر ما نسبته 44% من التباين الكلي.

الإجراءات

تم تنفيذ إجراءات الدراسة من خلال تطبيق أربع مراحل:

المرحلة الأولى

للحصول على المتغير الأول في الدراسة والمتمثل في تدريج الاستجابة، تم استخراج ثلاث عينات أساسية من العينة الرئيسة، وذلك بنسخ العينة الرئيسة إلى ثلاث نسخ تم فيها تعديل عدد الاستجابات من خلال دمج الاستجابات، والعينة الرئيسة (التدريج الرباعي) تحتوي على أربع استجابات (أوافق كثيراً= 4)، (أوافق قليلاً= 3)، (أرفض قليلاً= 2)، (أرفض كثيراً= 1)، فتبلغ الدرجة الصغرى للمقياس 9، والدرجة العظمى 36، أما العينة الثانية (التدريج الثلاثي) فتحتوي على ثلاث استجابات من خلال دمج الاستجابتين الوسيطتين في استجابة واحده: (موافق كثيراً= 3، موافق قليلًا + أرفض قليلاً= ٢، أرفض كثيراً =١) فتبلغ الدرجة الصغرى للمقياس 9، والدرجة العظمى 27. بينما العينة الثالثة تحتوي على استجابتين من خلال دمج كل استجابتين طرفيتين في استجابة واحده: (أوافق كثيراً + أوافق قليلاً= ٢، أرفض قليلاً + أرفض كثيراً =١)، فتبلغ الدرجة الصغرى للمقياس 9، والدرجة العظمى 18.

وللحصول على المتغير الثاني في الدراسة والمتمثل في الدرجات المحولة، تم حساب ٣ درجات محولة باستخدام العينات الرئيسة في التدريجات الثلاثة وهي: الدرجة المعيارية الاعتدالية (تحويل غير خطي للدرجات الخام بحيث يتم تحويل الدرجات الخام إلى درجات زائية باستخدام التوزيع التكراري المعياري الاعتدالي بمتوسط صفر وانحراف معياري ١) والرتب المئينية (الموقع النسبي للفرد أو النسبة المئوية من خلال درجات أقرانه والتي تقل عن درجته، وهي درجة تتراوح ما بين ١-٩٩ فمتوسطها الحسابي 50 وانحرافها المعياري 28) والتساعيات (درجات معيارية تتوزع فيها الدرجات في تسع فئات باستخدام الرتب المئينيه).

وتم استخدام درجتين محولتين أخيرتين في التدريج الرباعي وهي درجتي تيمز، حيث إن درجة تيمز الأولى تتم من خلال خطوتين:

1) تحويل الدرجة الخام تحويلاً غير خطيّ إلى قيمة لوجيت باستخدام نظرية الاستجابة للمفردة.
2) 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.

وأما 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 (TIMSS & PIRLS, 2019).

وتم استخدام جداول التحويل من الدرجات الخام إلي الدرجات المحولة الخمس في جميع العينات العشوائية المستخرجة.

المرحلة الثانية

للحصول على المتغير الثالث في الدراسة والمتمثل في حجم العينة، تم تكوين ثلاث عينات فرعية من كل عينة أساسية بطريقة عشوائية بسيطة تختلف في أحجامها (30، 200، 1000) وتم ذلك بطريقة البوتستراب (Bootstrap) في برنامج R، وتم تصنيف واختيار الأحجام لتمثل ثلاثة مستويات لحجم العينة وفقًا لطبيعة البحث التجريبي القائم على أسلوب المحاكاة، وبناءً على الدراسات السابقة القائمة على هذا المنهج كدراسة كل من: زراع (Zarea, 2021)؛ جولدهوف واخرون (Geldhof et al., 2014)؛ تراكسلر (Traxler, 2017)؛ يورداجول (Yurdugül, 2008).

المرحلة الثالثة

تم استخدام برمجية خاصة باستخدام برنامج (R) من إعداد المحرزي (Almehrizi2022,) لاستخراج معامل ثبات بيتا العام لأنواع الدرجات (الخام، والمحولة) في العينات الرئيسة والعينات الفرعية.

المرحلة الرابعة

تم تكرار المرحلتين الثانية والثالثة 1000 مرة باستخراج عينات عشوائية بثلاثة أحجام مختلفة للتدريجات الثلاث من مقياس اتجاهات الطلبة نحو الرياضيات، وبالتالي تم الحصول على إجمالي 9000 عينة (1000×3×3)، وتم تقدير معامل ثبات ألفا العام ومعامل ثبات بيتا العام بعدد 42000 مرة لكل منهما (1000×[4×3×2 + 6×3×1]).

المعالجة الإحصائية

للإجابة عن أسئلة الدراسة، تم استخراج الإحصاءات الوصفية ومعامل ثبات بيتا العام للدرجات الخام والمحولة، وللعينات العشوائية في مقياس اتجاهات الطلبة نحو الرياضيات TIMSS في سلطنة عُمان عبر متغيرات: تدريج المقياس ونوع الدرجة، وكذلك تم حساب نسب التحيز، وجذر متوسطات مربع الخطأ باستخدام المعادلتين:

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

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

حيث

r: رقم التكرار (رقم العينة العشوائية الفرعية).

$\widehat{G\beta}$: تقدير معامل ثبات بيتا العام في العينة الرئيسة لكل نوع من الدرجات المستخدمة الخام والمحولة لكل تدريج باختلاف أحجامه الثلاث.

$\widehat{G\beta}_r$: تقديرات معامل ثبات بيتا العام في العينات العشوائية الفرعية لكل نوع من الدرجات المستخدمة الخام والمحولة.

$\overline{\widehat{G\beta}_r}$: متوسط تقديرات معامل ثبات بيتا العام في العينات العشوائية الفرعية لكل نوع من الدرجات المستخدمة الخام والمحولة لكل تدريج باختلاف أحجامه الثلاث.

وأشار تراكسلر (Traxler, 2017) بأن 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.

نتائج الدراسة

للإجابة عن السؤال الأول تم حساب كل من:

أ- الإحصاءات الوصفية للدرجات الخام، والدرجات المحولة:

تم استخراج الإحصاءات الوصفية للدرجات الخام والدرجات المحولة في مقياس اتجاهات الطلبة نحو الرياضيات TIMSS في سلطنة عُمان، لمعرفة درجة الاتجاهات لدى الطلبة نحو الرياضيات، ويوضح جدول رقم (1) قيم الإحصاءات الوصفية لكل نوع من الدرجات المستخدمة الخام والمحولة.

جدول 1: الإحصاءات الوصفية للدرجات الخام والدرجات المحولة في مقياس اتجاهات الطلبة نحو الرياضيات في سلطنة عُمان في التدريجات الثلاثة.

نوع التدريج	الدرجة	المتوسط	الانحراف المعياري	الالتواء	التفلطح
الرباعي	الخام*	31.81	5.10	-1.69	**3.09**
	المعيارية الاعتدالية	-0.04	0.90	-0.52	**0.49**
	الرتبة المئينية	50.01	28.26	-0.14	**1.38**
	التساعيات	4.08	1.65	-0.13	**1.39**
	تيمز الأولىx	11	1.83	-0.33	**0.41**
	تيمز الثانيةxx	2.58	0.61	-1.16	**0.30**
الثلاثي	الخام**	24	3.48	-1.49	**2.28**
	المعيارية الاعتدالية	0.04	0.90	-0.52	**0.48**
	الرتبة المئينية	50.01	28.22	-0.14	**1.38**
	التساعيات	4.05	1.68	-0.13	**1.43**
الثنائي	الخام***	16.81	1.74	-2.02	**4.28**
	المعيارية الاعتدالية	0.06	0.84	-0.81	**0.26**
	الرتبة المئينية	50.01	26.81	-0.38	**1.46**
	التساعيات	3.86	1.38	-0.72	**1.13**

ملاحظة.*من 9 إلى 36 ، ** من 9 إلى 27، *** من 9 إلى 18، x من 3.85307 إلى 13.14434 وتم استخدام معادلة y= Logit Scale Score *1.234546+8.402636، xx من 1 إلى 3.

من النتائج في جدول رقم (1) تُظهر المتوسطات الحسابية للدرجات الخام في جميع التدريجات الثلاث أن اتجاهات الطلبة نحو الرياضيات بالصف الرابع بسلطنة عُمان إيجابية بدرجة مرتفعة (31.81/36 للتدريج الرباعي، حيث تعتبر 36 الدرجة الكلية للتدريج الرباعي (4×9 = 36)، و24/27 للتدريج الثلاثي (3×9 = 27)، و16.81/18 للتدريج الثنائي (2×9 = 18))، ويؤكد ذلك القيمة السالبة للالتواء التوزيع التكراري للتدرجات الثلاث.

وتشير النتائج في جدول رقم (1) أن قيم الإحصاءات الوصفية لكل نوع من أنواع الدرجات المحولة جاءت متوقعة في التدريجات الثلاث حسب طبيعة هذه الدرجات، فعلى سبيل المثال، كانت المتوسطات الحسابية والانحرافات المعيارية للدرجات المعيارية الاعتدالية مقاربة للصفر والواحد الصحيح على التوالي، وكانت المتوسطات الحسابية للرتب المئينية قريبة من 50 وانحرافاتها المعيارية قريبة من 28.

ب- معامل الثبات للدرجات الخام، والدرجات المحولة:

تم حساب معامل ثبات بيتا العام للمحرزي (2021) لأنواع الدرجات المستخدمة الخام والمحولة في مقياس اتجاهات الطلبة نحو الرياضيات TIMSS في سلطنة عُمان في التدريجات الثلاثة، ويوضح جدول رقم (2) تقديرات معامل ثبات بيتا العام لكل نوع من الدرجات المستخدمة الخام والمحولة.

جدول 2: معامل ثبات بيتا العام للدرجات الخام والدرجات المحولة في التدريجات الثلاثة لمقياس اتجاهات الطلبة نحو الرياضيات في سلطنة عُمان.

الدرجة	التدريج الرباعي	التدريج الثلاثي	التدريج الثنائي
الخام	0.7948	0.7984	0.7407
المعيارية الاعتدالية	0.7891	0.7931	0.7435
الرتبة المئينيه	0.7726	0.7744	0.7303
التساعيات	0.7721	0.7665	0.6743
تيمز الاولى	0.7868	-	-
تيمز الثانية	0.6801	-	-

تظهر النتائج في جدول رقم (2) أن قيم معامل ثبات بيتا العام للدرجات الخام بين 0.7407 و0.7984 في التدريجات الثلاث، وتظهر النتائج أيضاً أن قيم معامل ثبات بيتا العام في كل أنواع الدرجات المحولة أقل من قيم معامل ثبات بيتا العام للدرجات الخام للمقياس في التدريجات الثلاث ما عدا الدرجة المعيارية الاعتدالية في التدريج الثنائي، حيث كانت معاملات ثباتها أكبر من معامل ثبات الدرجات الخام، ويجدر الإشارة إلى أن الدراسة لم تستخدم فروق الدلالة الإحصائية حيث إنها مستندة على طريقة المحاكة لمونت كارلو.

كما تظهر النتائج أن قيم معامل ثبات بيتا العام للدرجة التساعية للمقياس أصغر من معاملات الثبات للدرجات المحولة الأخرى في جميع التدريجات الثلاث، وبالمقابل كانت القيمة الأكبر للدرجات المعيارية الاعتدالية، وهذا يرجع لنوعية الدرجات المحولة من حيث عدد الدرجات وطبيعة التوزيع التكراري لها، فالدرجات التساعية تحتوي على 9 درجات فقط بينما الدرجات المعيارية تأخذ أي قيمة حقيقية.

وللتعرف بشكل تفصيلي على أنماط معامل ثبات بيتا العام لمقياس اتجاهات الطلبة نحو الرياضيات وفقًا للمتغيرات تدريج المقياس ونوع الدرجة، ويعرض الجدول رقم (3) الفرق بين تقديرات معامل ثبات المقياس وفقًا لهذه المتغيرات.

جدول 3: الفروق بين تقديرات معامل الثبات لمقياس اتجاهات الطلبة نحو الرياضيات في سلطنة عُمان وفقًا لنوع الدرجة والتدريج.

نوع الدرجة/ التدريج	الرباعي- الثلاثي	الرباعي- الثنائي	المتوسط**
الخام	0.0036-	0.0541	0.0253
المعيارية الاعتدالية	0.004-	0.0456	0.0208
الرتبة المئينية	0.0018-	0.0423	0.0203
التساعيات	0.0056	0.0978	0.0517
المتوسط*	0.0010-	0.0600	-

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

تظهر النتائج في جدول رقم (3) الفروق بين تقديرات معامل الثبات بيتا العام لكل نوع من درجات المقياس (الخام، والمحولة)، وتم حساب قيم الفروق بطرح قيمة معامل ثبات بيتا العام لكل نوع من الدرجات المستخدمة الخام والمحولة في التدريج الرباعي من نظيراتها في التدريجين الآخرين (الثلاثي، والثنائي)، وذلك باعتبار أن التدريج الرباعي يمثل التدريج الأصلي للمقياس.

وانحصرت الفروق بين تقديرات معامل الثبات ما بين -0.0036 و0.0978، وجاءت قيم المتوسطات الحسابية للفروق في معامل ثبات بيتا العام لأنواع الدرجات الخام والدرجات المحولة للمقياس عبر التدريجات الثلاث موجبة، أي أن معامل ثبات بيتا العام لكل نوع من الدرجات المستخدمة الخام والمحولة تعطي تقديرات ثبات مختلفة، وظهر هذا النمط في قيم الفروق في معامل ثبات بيتا العام بين التدريجين الرباعي والثنائي في كل أنواع درجات المقياس الخام والمحولة، وكذلك ظهر هذا النمط بين التدريجين الرباعي والثلاثي في درجة التساعيات فقط، في حين جاءت فيه قيم الفروق لمعامل ثبات بيتا العام في كل من الدرجات (الخام، والمعيارية الاعتدالية، والرتب المئينية) سالبة حيث قدرت ب -0.0036 و -0.004 و -0.0018 على التوالي، وتشير هذه القيم السالبة أن معامل ثبات بيتا العام في هذه الدرجات كان أكبر في التدريج الثلاثي عن التدريج الرباعي، وأتت قيم الفروق لتقديرات معامل ثبات بيتا العام بين التدريجين الرباعي والثلاثي أصغر من نظيراتها من الفروق بين التدريجين الرباعي والثنائي، حيث تراوحت بين -0.0036 و 0.0056 بمتوسط حسابي -0.0010، وظهر هذا النمط في كل أنواع الدرجات المستخدمة الخام والمحولة، وهذا يشير إلى تقارب معاملات ثبات بيتا العام بين التدريجين الرباعي والثلاثي، بينما كانت مختلفة أكثر بين التدريجين الرباعي والثنائي.

تظهر النتائج في الجدول رقم (4) قيم الفروق بين تقديرات معامل ثبات بيتا العام بين الدرجات المحولة والدرجات الخام (ثبات الدرجة المحولة – ثبات الدرجة الخام) في كل تدريج من التدريجات الثلاث، حيث انحصرت قيم الفروق بين تقديرات معامل ثبات بيتا العام بين -0.0664 و 0.0028، وأتت قيم المتوسطات الحسابية للفروق في تقديرات معامل ثبات بيتا العام لكل نوع من أنواع الدرجات المحولة لجميع التدريجات الثلاث سالبة، وهذا ما تشير إليه القيم المتواجدة في العمود الأخير من الجدول رقم (4)، وهذا يدل على أن تقديرات معامل ثبات بيتا العام في كل أنواع الدرجات المحولة أصغر من تقديرات معامل ثبات بيتا العام للدرجة الخام.

وبرز هذا النمط أيضاً في قيم الفروق في معامل ثبات بيتا العام في كل أنواع الدرجات المحولة في كل تدريج من التدريجات الثلاث، فيما عدا معامل ثبات بيتا العام للدرجة المعيارية الاعتدالية في التدريج الثنائي، حيث أتى الفرق موجباً (0.0028) مما يدل على أن تقدير معامل ثبات بيتا العام كان فيها أكبر عما كان عليه في الدرجة الخام. ويعرض الجدول رقم (4) الفرق بين تقديرات معامل ثبات المقياس وفقًا لهذه المتغيرات.

جدول 4: الفروق بين تقديرات معامل الثبات (محولة - الخام) لمقياس اتجاهات الطلبة نحو الرياضيات في سلطنة عُمان وفقًا لنوع الدرجة والتدريج.

الدرجة	نوع التدريج			المتوسط**
	الرباعي	الثلاثي	الثنائي	
المعيارية الاعتدالية	0.0057	0.0053	0.0028	0.0027
الرتبة المئينية	0.0222	0.0240	0.0104	0.0189
التساعيات	0.0227	0.0319	0.0664	0.0403
درجة تيمز الأولى	0.0080	-	-	-
درجة تيمز الثانية	0.1147	-	-	-
المتوسط *	0.0347	0.0204	0.0247	-

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

وتتفق هذه النتيجة مع دراستي المحرزي (Almehrizi, 2021; 2022)، بأنه لا يمكن التعميم بأن قيم معامل ثبات بيتا العام أكبر في الدرجات الخام عن الدرجات المحولة، وكذلك تتفق مع دراسة كلٍ من: الضوي (Al-Dawy, 2011)؛ وعلي وحسن (Ali & Hassan, 2011) لاحتوائها على النقطة الوسيطة، وهذا ما أكده أبو حطب وصادق (Abu Hatab & Sadiq, 1991) أن يكون تدريج الاستجابة فرديًا للحصول على قيمة ثبات عالية، لاحتوائها على النقطة الوسيطة، وتختلف مع دراسة كل من: شودوري وبهاتاشارجي (Choudhury & Bhattacharjee, 2014)؛ وبريستون وكولمان (Preston & colman, 2000) التي أشارت إلى أنّ الثبات يتناسب طرديا بزيادة تدريج الاستجابة.

وللإجابة عن السؤال الثاني تم استخراج الإحصاءات الوصفية لمعامل ثبات بيتا العام في العينات العشوائية المستخرجة (1000 عينة) بطريقة تقديرات البوتستراب (Bootstrap) لمقياس اتجاهات الطلبة نحو الرياضيات بسلطنة عُمان في التدريجات الثلاث باستخدام ثلاثة أحجام للعينة، ويوضح الجدول رقم (5) الإحصاءات الوصفية لقيم معامل الثبات للتدريج الرباعي لمقياس اتجاهات الطلبة نحو الرياضيات في الاحجام الثلاث.

جدول 5: الإحصاءات الوصفية لقيم معامل الثبات للتدريج الرباعي لمقياس اتجاهات الطلبة نحو الرياضيات في سلطنة عُمان باستخدام تقديرات البوتستراب للأحجام الثلاثة للعينة (30، 200، 1000).

الدرجة	ثبات العينة الرئيسية	ن	نسبة التحيز%	المتوسط	الانحراف المعياري
الخام	0.7948	30	4.3	0.7606	0.1153
		200	0.6	0.7904	0.0359
		1000	0.1	0.7938	0.0139
المعيارية الاعتدالية	0.7891	30	1.6	0.7763	0.0819
		200	0.3	0.7866	0.0287
		1000	0.04	0.7888	0.0115
الرتب المئينية	0.7726	30	1.1	0.7641	0.0799
		200	0.3	0.7702	0.0289
		1000	0.01	0.7725	0.0118
التساعيات	0.7721	30	1.3	0.7624	0.0781
		200	0.3	0.7697	0.0278
		1000	0	0.7721	0.0114
تيمز الأولى	0.7868	30	1.6	0.7746	0.0815
		200	0.3	0.7842	0.0292
		1000	0.02	0.7866	0.0118
تيمز الثانية	0.6801	30	3.3	0.6577	0.1122
		200	0.5	0.6768	0.0405
		1000	0.1	0.6794	0.0162

تظهر النتائج في الجدول رقم (5) أن قيم المتوسطات الحسابية لمعامل ثبات بيتا العام في كل أنواع درجات المقياس تقترب من ثبات العينة الرئيسة كلما زاد حجم العينة.

وكما أن قيم الانحرافات المعيارية لمعامل ثبات بيتا العام للدرجات الخام والدرجات المحولة تقترب من الصفر كلما زاد حجم العينة، وهذا يؤكد استقرار تقديرات معامل الثبات بزيادة حجم العينة، وأظهرت قيم الالتواء والتفلطح لمعامل ثبات بيتا العام لكل نوع من الدرجات المستخدمة الخام والمحولة اقترابها من الصفر بزيادة حجم العينة، وهذا يشير إلى أن مؤشرات التوزيع التكراري لمعامل الثبات

تقترب من مؤشرات التوزيع الطبيعي.

وتظهر نسب التحيز لمعامل ثبات بيتا العام في التدريج الرباعي بشكل عام بأنها نسب مقبولة لم تتعد 10% في الدرجات الخام والدرجات المحولة ويمكن تجاهلها، حيث تراوحت بين (0، 4.3) وهذا يُظهر استقرار تقديرات معامل الثبات في العينات العشوائية واقترابها من ثبات العينة الرئيسة سواء كان التحيز بالتقليل أم التضخيم، ولتوضيح ذلك أتت نسب التحيز في معامل ثبات بيتا العام في الرتب المئينية في كل حجم من الاحجام الثلاثة (30، 200، 1000) 1.1، 0.3، 0.01 على التوالي، وهي نسب تدل على التقليل في تقدير الثبات مقارنة بثبات العينة الرئيسة، ولكن بنسب قليلة لم تتعدَ 10%، بمعنى أن تقديرات معامل ثبات بيتا العام لهذه الدرجة في العينات العشوائية أقل من ثبات العينة الرئيسة، وتقل هذه النسب كلما زاد حجم العينة وتقترب بذلك من الصفر.

ويوضح الجدول رقم (6) الإحصاءات الوصفية لقيم معامل الثبات بالتدريج الثلاثي لمقياس اتجاهات الطلبة نحو الرياضيات في الأحجام الثلاث.

جدول (6)

الإحصاءات الوصفية لقيم معامل الثبات للتدريج الثلاثي لمقياس اتجاهات الطلبة نحو الرياضيات في سلطنة عُمان باستخدام تقديرات البوتستراب لأحجام العينات.

الدرجة	ثبات العينة الرئيسية	ن	نسبة التحيز%	المتوسط	الانحراف المعياري
الخام	0.7984	30	5.1	0.7580	0.1105
		200	0.5	0.7944	0.0335
		1000	0.1	0.7979	0.0126
المعيارية الاعتدالية	0.7931	30	2.7	0.7716	0.0841
		200	0.3	0.7909	0.0286
		1000	0.03	0.7929	0.0110
الرتب المئينية	0.7744	30	1.9	0.7594	0.0821
		200	0.2	0.7729	0.0292
		1000	0.01	0.7743	0.0115
التساعيات	0.7665	30	1.9	0.7520	0.0821
		200	0.2	0.7651	0.0288
		1000	0.01	0.7664	0.0133

ويتضح من الجدول رقم (6) أن قيم المتوسطات الحسابية لمعامل ثبات بيتا العام في كل أنواع الدرجات الخام والدرجات المحولة تقترب من معامل ثبات بيتا العام في العينة الرئيسة كلما زاد حجم العينة، فعلى سبيل المثال جاءت قيم المتوسطات الحسابية لمعامل ثبات بيتا العام للدرجة الخام في الأحجام الثلاث (30، 200، 1000) 0.7580، 0.7944، 0.7979 على التوالي.

وتظهر النتائج أن نسب التحيز لمعامل ثبات بيتا العام في جميع الدرجات الخام والدرجات المحولة لم تتعدَ 10%، وهي نسب مقبولة حيث تراوحت بين 0.01 و 5.1 وهذا يدل على استقرار تقديرات معامل ثبات بيتا العام في كل الدرجات الخام والدرجات المحولة، فعلى سبيل الذكر أتت قيم نسب التحيز لمعامل ثبات بيتا العام في الدرجة المعيارية الاعتدالية للأحجام الثلاثة (30، 200، 1000) 2.7، 0.3، 0.03 على التوالي، وتقل هذه النسب كلما زاد حجم العينة وتقترب بذلك من الصفر، وهذا يدل على استقرار تقديرات معامل ثبات بيتا العام في العينات العشوائية واقترابها من ثبات العينة الرئيسة كلما زاد حجم العينة.

ويوضح الجدول رقم (7) الإحصاءات الوصفية لقيم معامل الثبات بالتدريج الثنائي لمقياس اتجاهات الطلبة نحو الرياضيات في الأحجام الثلاث.

جدول (7): الإحصاءات الوصفية لقيم معامل الثبات للتدريج الثنائي لمقياس اتجاهات الطلبة نحو الرياضيات في سلطنة عُمان باستخدام تقديرات البوتستراب لأحجام العينات.

الدرجة	ثبات العينة الرئيسية	ن	نسبة التحيز%	المتوسط	الانحراف المعياري
الخام	0.7407	30	8.9	0.6748	0.1740
		200	1.01	0.7332	0.0488
		1000	0.2	0.7393	0.0187
المعيارية الاعتدالية	0.7435	30	2.6	0.7244	0.0992
		200	0.2	0.7419	0.0343
		1000	0.01	0.7434	0.0138
الرتب المئينية	0.7303	30	1.3	0.7208	0.0898
		200	0.01-	0.7304	0.0320
		1000	0.1-	0.7307	0.0129
التساعيات	0.6743	30	2.8	0.6553	0.1172
		200	0.2	0.6731	0.0394
		1000	0.03-	0.6745	0.0160

تظهر النتائج في الجدول رقم (7) أن قيم المتوسطات الحسابية لمعامل ثبات بيتا العام لكل من الدرجات (الخام، والمعيارية الاعتدالية) بأنها تقترب من معامل ثبات بيتا العام للعينة الرئيسة كلما زاد حجم العينة، فعلى سبيل المثال جاءت قيم المتوسطات الحسابية لمعامل ثبات بيتا العام للدرجة المعيارية الاعتدالية في الأحجام الثلاث (30، 200، 1000) 0.7244، 0.7419، 0.7434 على التوالي، في حين أتت قيم المتوسطات الحسابية في الرتب المئينية قريبة من ثبات العينة الرئيسة في الأحجام الثلاث، في حين اقتربت قيم المتوسطات الحسابية لمعامل ثبات بيتا العام في التساعيات من ثبات العينة الرئيسة بزيادة حجم العينة عند الحجمين (30، 200)، في حين كانت قريبة عند الحجم 1000 من ثبات العينة الرئيسة.

وتؤكد النتائج استقرار تقديرات معامل ثبات بيتا العام في كل أنواع الدرجات الخام والمحولة كلما زاد حجم العينة، وهذا ما تظهره قيم الانحرافات المعيارية لمعامل الثبات حيث إنها تقترب من الصفر كلما زاد حجم العينة، وكذلك أظهرت قيم الالتواء والتفلطح اقترابها من الصفر بزيادة حجم العينة، وهذا يؤكد أن مؤشرات التوزيع التكراري لمعامل الثبات تقترب من مؤشرات التوزيع الطبيعي.

وتظهر نسب التحيز لمعامل ثبات بيتا العام بشكل عام بأنها نسب مقبولة لم تتعدَ 10% في الدرجات الخام والدرجات المحولة، وهذا يظهر استقرار تقديرات معامل الثبات بيتا العام في العينات العشوائية واقترابها من ثبات العينة الرئيسة سواء أكان التحيز بالتقليل أم المبالغة، فعلى سبيل الذكر أتت نسب التحيز في معامل ثبات بيتا العام في الرتب المئينية للأحجام الثلاث (30، 200، 1000) 1.3، -0.01، -0.1 على التوالي، فالنسبة الأولى تدل على التقليل في تقدير الثبات مقارنة بثبات العينة الرئيسة، أي أن معامل ثبات بيتا العام في الرتب المئينية عند الحجم 30 أقل من تقدير معامل ثبات بيتا العام في العينة الرئيسة، في حين أن النسبة الثانية والثالثة تدل على المبالغة في تقدير الثبات مقارنة بثبات العينة الرئيسة، وهذا يدل على أن تقديرات معامل ثبات بيتا العام في العينات العشوائية للرتب المئينية عند الحجمين (200، 1000) أكبر من معامل ثبات بيتا العام في العينة الرئيسة.

وتتفق هذه النتائج مع دراسة تريزانو-هيرموسيلا و ألفارادو (Trizano-Hermosilla & Alvarado, 2016) من ناحية إجراءاتها التي تتمثل في اختبار فعالية معامل الثبات، وذلك من خلال أخذ عينات عشوائية من العينة الرئيسة بأحجام مختلفة بطريقة البوتستراب، ثم مقارنتها بمعامل ثبات العينة الرئيسة، وإيجاد نسب التحيز في تقديرات معامل الثبات، والتي أثبتت نتائجها استقرار تقديرات معامل الثبات بزيادة حجم العينة.

للإجابة على السؤال الثالث تم استخراج قيم جذر متوسطات مربع الخطأ لمعامل ثبات أنواع الدرجات الخام والمحولة في التدريجات الثلاث في مقياس اتجاهات الطلبة نحو الرياضيات لأحجام العينات، لمعرفة مدى فعالية أداء معامل ثبات بيتا العام، كما يوضحها الجدول رقم (8).

جدول 8: جذر متوسطات مربع الخطأ لمعامل ثبات الدرجات الخام والدرجات المحولة في التدريجات الثلاثة في مقياس اتجاهات الطلبة نحو الرياضيات لأحجام العينات.

نوع التدريج	نوع الدرجة	حجم العينة		
		30	200	1000
الرباعي	الخام	0.1202	0.0361	0.0139
	المعيارية الاعتدالية	0.0829	0.0287	0.0115
	الرتبة المئينية	0.0803	0.0290	0.0118
	التساعيات	0.0787	0.0279	0.0114
	تيمز الاولى	0.0824	0.0293	0.0118
	تيمز الثانية	0.1143	0.0406	0.0163
الثلاثي	الخام	0.1176	0.0337	0.0126
	المعيارية الاعتدالية	0.0867	0.0287	0.0110
	الرتبة المئينية	0.0835	0.0292	0.0115
	التساعيات	0.0834	0.0289	0.0113
الثنائي	الخام	0.1860	0.0494	0.0188
	المعيارية الاعتدالية	0.1010	0.0343	0.0138
	الرتبة المئينية	0.0903	0.0319	0.0129
	التساعيات	0.1187	0.0394	0.0160

يتضح من الجدول رقم (8) أن قيم جذر متوسطات مربع الخطأ لمعامل ثبات بيتا العام لكل أنواع الدرجات (الخام، والمحولة) في كل تدريج من التدريجات الثلاث بأحجامه الثلاث المختلفة تقل بزيادة حجم العينة وتقترب بذلك من الصفر، فعلى سبيل المثال: جاءت قيم جذر متوسطات مربع الخطأ لمعامل ثبات بيتا العام في الدرجة الخام في التدريج الرباعي في كل حجم من الأحجام الثلاث (30، 200، 1000) 0.1202، 0.0361، 0.0139 على التوالي، وظهر هذا النمط في جميع درجات المقياس بتدريجاته الثلاث، وهذا يؤكد استقرار تقديرات معامل ثبات بيتا العام في العينات العشوائية لكل نوع من الدرجات المستخدمة الخام والمحولة ومدى اقترابها من ثبات العينة الرئيسة وتتفق كذلك مع دراسة كل من: (Zarea, 2021)؛ تريزانو-هيرموسيا وألفارادو (Trizano-Hermosilla & Alvarado, 2016) والتي أثبتت نتائجها بأن قيم معاملات الثبات في العينات العشوائية تقترب من ثبات العينة الرئيسة، حيث تقل قيم جذر متوسطات مربع الخطأ بزيادة حجم العينة وتقترب من الصفر.

وتتفق أيضًا هذه النتيجة من ناحية إجراءاتها مع دراسة يورداجول (Yurdugül, 2008) والتي أثبتت أن معامل الثبات في العينات العشوائية يقترب من ثبات العينة الرئيسة، وذلك من خلال اقتراب قيم جذر متوسطات مربع الخطأ من الصفر بزيادة حجم العينة.

وتظهر النتائج تقارب سلوك بعض الدرجات المحولة في تقديرات معامل ثبات بيتا العام في كل تدريج من التدريجات الثلاث، وهذا يدل على استقرار تقديرات معامل ثبات بيتا العام في العينات العشوائية، فعلى سبيل الذكر جاءت قيم جذر متوسطات مربع الخطأ لمعامل ثبات بيتا العام في كل من الدرجات (تيمز الأولى، والتساعيات، والرتب المئينية) في التدريجات الثلاثة متقاربة جداً في تقدير الثبات، وهذا يدل على استقرار تقديرات معامل ثبات بيتا العام.

الخلاصة والمقترحات

توصلت الدراسة الحالية إلى أن معامل ثبات بيتا العام للمحرزي يعطي تقديرات دقيقة للدرجات سواء كانت خام أم محولة، وذلك من خلال مؤشرات التحيز وجذر متوسط مربع الخطأ واقتراب ثبات العينات العشوائية من ثبات العينة الرئيسة، وفيما يرتبط بنتائج الدراسة المرتبطة بمتغيراتها: حجم العينة (30، 200، 1000)، وعدد مستويات التدريج (رباعي، ثلاثي، ثنائي)، ونوع الدرجة (خام، معيارية اعتدالية، رتب مئينيه، تساعيات)، فقد توصلت النتائج إلى أن دقة تقدير معامل ثبات بيتا العام تزداد بزيادة حجم العينة، وتبين ذلك من خلال مؤشري: نسب التحيز، وجذر متوسط مربع الخطأ، وذلك عند جميع حالات تقاطعات متغيري عدد فئات التدريج، ونوع الدرجة.

وفيما يتعلق بمتغير عدد فئات التدريج فقد توصلت الدراسة إلى أن ثبات البيانات ذات التدريج الرباعي كان بشكل عام أعلى عن التدريجين الثلاثي والثنائي، إلا في حالة الدرجتين الخام، والمعيارية الاعتدالية، فقد كان التدريج الثلاثي أعلى.

وفيما يرتبط بمتغير نوع الدرجة المحولة، فقد بينت النتائج بشكل 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 بيانات أرشيفية من موقع TIMSS، وكذلك في عدم تناوله لمعاملات تقدير ثبات أخرى، وفي ضوء ذلك توصي الدراسة ما يأتي:

- استخدام معامل ثبات بيتا العام في تقدير ثبات مختلف الدرجات

المحولة غير الخطية للمقاييس والاختبارات.

- الاستفادة من الدراسة الحالية من قبل الباحثين من خلال ما وفرته من أدلة من اختيار الحجم المناسب للعينة، وتدريج الاستجابة، ونوع الدرجة المناسبة لتحقيق أفضل ثبات.
- استخدام معامل ثبات بيتا العام من قبل مطوري الاختبارات التحصيلية في وزارة التربية والتعليم للحكم على ثبات هذه الاختبارات.
- حث الباحثين على استخدام معامل ثبات بيتا للمحرزي في بحوثهم، وذلك من خلال عمل ورش تدريبية للتعريف بمعامل الثبات، وكيفية استخدامه في برنامج R.

References

Abu Hatab, F., & Sadiq, A. (1991). Research methods and statistical analysis methods in psychological, educational, and social sciences (In Arabic). Anglo Egyptian Library.

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

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 the 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.

Allam, R. (2011). Scientific research methods in psychological and educational sciences. Universities Publishing Dar.

Allam, S. (2006). Educational and psychological measurement and evaluation: its basics, applications, and contemporary trends (In Arabic). Arab Thought House.

Almehrizi, R. S. (2013). Coefficient alpha and reliability of scale scores. Applied Psychological Measurement, 37(6), 438-459.

Almehrizi, R. S. (2016). Normalization of mean squared differences to measure agreement for continuous data. Statistical methods in medical research, 25(5), 1955-1974.

Almehrizi, R. S. (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.

Almehrizi, R. S. (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.

Birkett, N. J. (1986). Selecting the number of response categories for a Likert-type scale. In Proceedings of the American statistical association, 1, 488-492.

Charter, R. A. (2003). Study samples are too small to produce sufficiently precise reliability coefficients. The Journal of Generalized Psychology, 130(2), 117-129.

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

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.

Dunn, T. J., Baguley, T., & Bruns den, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. British journal of psychology, 105(3), 399-412.

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.

Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. Educational and psychological measurement, 66(6), 930-944.

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 .DOI: 10.1207/S15328007SEM0702_6

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.

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.

Lord, F. M., & Winger sky, M. S. (1984). Comparison of IRT true-score and equipe centile observed-score" equating's. 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.

Nunnally, J. C., & Bernstein, C. (1994). Psychometric theory. New York: McGraw-Hill. Accessed.

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

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.

Raykov, T. (2001). Bias of coefficient α for fixed congeneric measures with correlated errors. Applied psychological measurement, 25(1), 69-76.

Segall, D. O. (1994). The reliability of linearly equated tests. Psychometrika, 59(3), 361-375.

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.

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., U .Eshima, 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.

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 (CIDM) (pp. 310-317). IEEE.

Yurdugül, H. (2008). Minimum sample size for Cronbach's coefficient alpha: a Monte-Carlo study. Journal of Education: Haccet tape University, 35(35), 1-9.

Zarea, Nisreen. (2021). Comparison of Test Scores Reliability Coefficients under a Set of Variables: A Monte Carlo Simulation Study (In Arabic). Journal of education Sohag UNV, 88(88), 1107- 1174.

Zarea, Nisreen. (2022). The Effect of Different Statistical Analysis Method According to Interval and Ordinal Data Nature on the Psychometric Properties of Likert Scale: A Simulation Study (In Arabic), Humanities and Educational Sciences Journal, 22, 238- 266.

Zumbo, B. D., Gaderman, 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.