Journal of Educational and Psychological Sciences (JEPS) • Vol 8, Issue 9 (2024) • P: 115 - 98
https://journals.ajsrp.com/index.php/jeps
ISSN: 2522-3399 (Online) • ISSN: 2522-3399 (Print)

The impact of the Item Response Theory (IRT) model on the accuracy of test score equating

Dr. Easa Salamah Alhajory*[1], Dr. Mohammed Mahmoud Abdel-Wahab[2]

[1] Ministry of Education | KSA

[2] Umm Al-Qura University | KSA

Received: 21/05/2024

Revised: 03/06/2024

Accepted: 28/06/2024

Published: 30/09/2024

* Corresponding author: easa@live.ca

Citation: Alhajory, E. S., & AbdelWahab, M. M. (2024). The impact of the Item Response Theory (IRT) model on the accuracy of test score equating. *Journal of Educational and Psychological Sciences, 8*(9), 98 – 115. https://doi.org/10.26389/AJSRP.R210523

2024 © AISRP • Arab Institute of Sciences & Research Publishing (AISRP), Palestine, all rights reserved.

• **Open Access**

This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license

Abstract: The aim of the current study was to investigate the impact of the Item Response Theory (IRT) model on the accuracy of test score equating. The focus was on the effect of the two-parameter logistic (2PL) model and the three-parameter logistic (3PL) model, using the Loyd & Hoover mean/mean equating method (1980). To achieve the study's aim, simulated data were generated using the (Wingen3) software for 4 test forms. Two test forms were based on the 2PL model and the other two on the 3PL model, with 10 common items between each pair of test forms. All the data generation was repeated 30 times.

Data were analyzed using the R software through the Mirt and Sirt packages to verify the assumptions of the IRT model. Subsequently, the individual and item parameters were estimated using the PARSCLE software, and equating was performed using the IRTEQ software with the mean/mean method. To evaluate the accuracy of equating, the Root Mean Square Error (RMSE) was used. The results showed that the lowest errors occurred when equating was performed using the 2PL model, with an RMSE value of (0.591). When equating was performed according to the 3PL model, the RMSE value was 1.044, indicating that it is more accurate than the mean/mean method in equating test scores according to the 3PL model.

Through an independent two-sample t-test, it was revealed that the differences were statistically significant, with a significance level of (0.000**).**

Keywords: Item response theory, modern theory of measurement, test score equivalence, equivalence accuracy, mean/median method.

أثر نموذج نظرية الاستجابة للمفردة في دقة معادلة درجات الاختبارات

د/ عيسى بن سلامه الحجوري*[1]، د / محمد بن محمود عبد الوهاب[2]

[1] وزارة التعليم | المملكة العربية السعودية

[2] جامعة أم القرى | المملكة العربية السعودية

المستخلص: هدفت الدراسة الحالية إلى معرفة أثر نموذج نظرية الاستجابة للمفردة في دقة معادلة درجات الاختبارات. وركزت على معرفة أثر النموذج الثنائي البارامتر والنموذج الثلاثي البارامتر وذلك باستخدام طريقة المعادلة المتوسط/المتوسط لويد وهوفر(Loyd & Hoover, 1980) ، ولتحقيق هدف الدراسة تم توليد بيانات محاكاة باستخدام برمجية (Wingen3) لـ 4 صور اختبارية، حيث كانت صورتين اختباريتين وفق النموذج الثنائي البارامتر والأخرى وفق النموذج الثلاثي البارامتر، وكان هناك عدد (10) فقرات مشتركة بين كل صورتين اختباريتين، وتكرر توليد جميع البيانات السابقة (30) مرة. تم تحليل البيانات باستخدام برنامج R من خلال حزمتي Mirt وSirt لمعرفة تحقق افتراضات نظرية الاستجابة للمفردة، بعد ذلك تم تقدير بارامترات الأفراد والمفردات بواسطة برنامج PARASCLE، وتمت المعادلة بواسطة برنامج IRTEQ باستخدام طريقة المتوسط/المتوسط. ولتقييم دقة المعادلة استخدم جذر متوسط مربعات الخطأ RMSE. وأظهرت النتائج أن أقل الأخطاء كانت عندما تمت المعادلة باستخدام النموذج الثنائي البارامتر، حيث بلغت قيمة جذر متوسط مربعات الخطأ RMSE عند استخدام هذا النموذج (0.591) أما عندما تمت المعادلة وفق النموذج الثلاثي البارامتر فكانت قيمة جذر متوسط مربعات الخطأ RMSE تساوي (1.044) وبالتالي فهي أكثر دقة من طريقة المتوسط/المتوسط في معادلة درجات الاختبار وفق النموذج الثلاثي البارامتر، ومن خلال اختبار ت لعينتين مستقلتين اتضح أن الفروق دالة احصائياً حيث بلغت مستوى الدلالة (0.000).

الكلمات المفتاحية: نظرية الاستجابة للمفردة، النظرية الحديثة في القياس، معادلة درجات الاختبارات، دقة المعادلة، طريقة المتوسط/المتوسط.

1- مقدمة.

يسعى المتخصصون في التربية وعلم النفس بشكل عام، وفي مجال الاختبارات والمقاييس بشكل خاص، إلى بناء وتطوير الأدوات المستخدمة في عمليتي القياس والتقويم، والتحقق من كفاءتها وجودتها، وذلك من خلال محاولة تقليص الأخطاء إلى أقل ما يمكن؛ ذلك لأن تلك الأدوات تعطي نتائج تستخدم في صنع واتخاذ قرارات مهمة، ومدى دقة وصحة تلك النتائج، ينعكس على صحة القرارات المتخذة.

وتستخدم الاختبارات لأغراض عدة منها اختيار وتصنيف الطلبة وتحديد الحياة المهنية لهم، ومن هنا يتضح أهمية القرارات التي تعتمد على نتائج تلك الاختبارات (Inal & Anil, 2018).

إن الحاجة إلى صور مختلفة لنفس الاختبار من الأمور المهمة؛ وذلك للحفاظ على أمن وسرية الاختبار؛ حيث قد تختلف صور الاختبار المعدة لنفس الغرض في درجة صعوبتها، وقد يحصل الأفراد المتقدمين للاختبار الأسهل إلى درجات أعلى من نظيرهم المتقدمين للاختبار الأصعب، وهذا يؤدي إلى عدم العدل والمساواة بين المتقدمين للاختبار، وبذلك يفقد الاختبار عدالته بالرغم من تقديم نفس المحتوى، لذلك يجب أن تتسم صور الاختبارات التي تختلف في الصعوبة بالمساواة بين الأفراد وأن تكون متكافئة (Kolen & Brennan, 2004; von Davier, Holland, & Thayer, 2004)

عملية معادلة درجات الاختبار أحد التطبيقات المهمة التي ينادي بها المختصون في مجالات الاختبارات والمقاييس، والتي يتم فيها إزالة الاختلافات بين مستويات الصعوبة في صور الاختبارات، والتي تم إعدادها لتطبيقها على مجموعات من المفحوصين، للحفاظ على سرية الاختبار، والتي يراعى خلال بنائها بتماثل المحتوى والخصائص السيكومترية قدر الإمكان، حيث أنه في صور الاختبار من المتوقع أن يكون هناك اختلاف في مستويات الصعوبة وفي القدرات التي تتطلبها، وبالتالي يصعب المقارنة المباشرة بين درجات المفحوصين الذين أجابوا على تلك الصور المختلفة، مما يستوجب تحويلها لميزان مشترك تجعلها متكافئة (طومان وآخرون، 2006).

بدأ الانتباه لأهمية معادلة درجات الاختبار في أوائل ثمانينيات القرن الماضي (Inal & Anil, 2018)، وأدى ذلك إلى زيادة وتنوع البرامج المتخصصة في المعادلة من قبل المتخصصين (Kolen & Brennan, 2004).

ومعادلة درجات الاختبار إجراء تجريبي (Dorans, Moses, & Eignor, 2009) لأنها تتطلب القيام بعمل تصميم لجمع البيانات، وقاعدة لتحويل الدرجات من اختبار إلى آخر (Öztürk-Gübeş & Kelecioğlu, 2016)، كما أنها عبارة عن مجموعة من الإجراءات التي تحول تقديرات موقع الشخص على مقاييس مختلفة إلى مقياس مشترك، والتي تساعد في تسهيل مقارنة الأشخاص (De Ayala, 2022).

وهناك عدة طرق للمعادلة بعضها يكون في إطار نظرية القياس التقليدية، والآخر وفق نظرية الاستجابة للمفردة. وبسبب المزايا التي توفرها نظرية الاستجابة للمفردة، وقدرتها على معالجة بعض جوانب القصور التي كانت تعاني منها نظرية القياس التقليدية؛ أصبح مجال معادلة درجات الاختبارات وفق هذه النظرية محل اهتمام عديد من الباحثين والمختصين (الزعابي وآخرون،2020).

ومعادلة درجات الاختبارات القائمة على الدرجات الخام غير مرغوب فيها لأسباب تتعلق بالعدالة والتماثل والثبات، والمعادلة القائمة على أساس نظرية الاستجابة للمفردة تتغلب على تلك المشاكل إذا كان النموذج المستخدم يلائم البيانات (Kolen, 1981).

وقبل القيام بعملية معادلة الدرجات ينبغي معرفة التصميم المراد استخدامه في المعادلة (Akin-Arikan, & Gelbal, 2021)، فهناك عدة تصاميم للمعادلة، ومن أشهرها تصميم المجموعات المتكافئة ذات الجذع المشترك، والذي يتم فيه تقسيم مجموعة من الأفراد إلى مجموعتين متساويتين بشكل عشوائي، ويتم تطبيق الصورة الأولى من الاختبار على المجموعة الأولى من المفحوصين، وفي نفس الوقت يتم تقديم الصورة الثانية من الاختبار إلى المجموعة الثانية، بعد ذلك يتم تطبيق اختبار مشترك على كلتا المجموعتين وفي نفس الوقت (المدانات،2012).

ويجب تحري الدقة في إجراءات عملية معادلة درجات الاختبار، لما لها من أهمية بالغة للعدل بين الأفراد واتخاذ قرارات صحيحة بشأن مستقبلهم (Inal & Anil, 2018).

ولجعل المعادلة ذات موثوقية عالية، لابد من أخذ مجموعة من العوامل بعين الاعتبار كاختيار التصميم وطريقة المعادلة المناسبة، وكذلك اختيار النموذج الرياضي المناسب، وذلك من أجل الحصول على دقة معادلة جيدة، حيث أن اختيار النموذج الجيد يقود إلى مطابقة جيدة للمفردات، لأن لكل نموذج افتراضاته وخصائصه التي تختلف عن غيره، فقد تختلف مطابقة البيانات لنموذج عن الآخر، وقد يسبب ذلك معضلة للباحث، حيث يتبين له لاحقاً أن فقرات الاختبار قد لا تناسب النموذج، وبالتالي من هنا تأتي أهمية معرفة النموذج الرياضي المناسب الذي يطابق بيانات الدراسة (الشواورة، 2013).

وهناك مجموعة من النماذج الرياضية التي تندرج تحت إطار نظرية الاستجابة للمفردة والتي تستخدم في تقديرات بارامترات المفردات والقدرات للأفراد، والتي تعتمد أساساً على طبيعة البيانات، فهناك نماذج ثنائية الاستجابة ونماذج متعددة الاستجابة، ومن النماذج ثنائية الاستجابة، نموذج راش والذي يتم فيه تدريج فقرات الاختبار من خلال معلم صعوبة الفقرة فقط، والنموذج اللوجستي

ثنائي البارامتر والذي يحتوي على معلمتي الصعوبة والتمييز، والنموذج اللوجستي ثلاثي البارمتر والذي يمكن أن يقدر ثلاثة معالم هي معلمة الصعوبة والتمييز بالإضافة إلى معلمة التخمين، وهنالك النموذج اللوجستي رباعي المعلمة الذي يضيف معلمة رابعة تسمى بمعلمة الخط التقاربي الأعلى.

ومن طرق المعادلة في نظرية الاستجابة للمفردة، طريقة المتوسط/سيجما التي قدمها ماركو (Marco, 1977) والتي تستخدم المتوسطات والانحرافات المعيارية لبارامتر الصعوبة للمفردة المقدرة من الفقرات المشتركة لتحديد معاملات الربط، وطريقة المتوسط/المتوسط التي قدمها لويد وهوفر (Loyd& Hoover,1980) تستخدم متوسط بارامتر الصعوبة والميل المقدرة من الفقرات المشتركة، لتحديد معاملات الربط، وطريقة هايبرا (Haebara, 1980) وستوكنج لورد (Stocking & Lord, 1983) من الطرق التي تعتمد على المنحنى المميز، بحيث تأخذ بالاعتبار جميع البارامترات التي تقلل الاختلافات بين المنحنى المميز بين الاختبارات، وتستخدم طريقة هايبرا الفرق بين المنحنى المميز للفقرة، بحيث تأخذ مجموع مربعات الفروق بين المنحنيات المميزة للفقرة لكل فقرة لقدرة محددة للمفحوصين، أما في طريقة ستوكنج ولورد تستخدم مربع مجموع الفروق بين منحنيات خصائص الفقرات التي تخص كل مفردة، فيجب ملاحظة أنه يتم أخذ المجموع لكل المفردات لكل بارامتر مقدر قبل القيام بعملية التربيع (Kolen & Brennan, 2014)).

1-2-مشكلة الدراسة وتساؤلاتها:

تلجأ العديد من الهيئات والمؤسسات التعليمية والعاملين تحتها، لعمل صور متعددة للاختبارات التي تقدمها، ذلك من أجل الحفاظ على أمنها وسريتها، من أجل ألا تفقد الهدف التي أوجدت من أجله، وذلك بإعطاء نتائج مضللة في حالة تم تسريبها أو تكرارها، وبالتالي قد يبنى عليها قرارات خاطئة، علماً أن القرارات التي تتخذ غالباً ما تكون ذات أهمية كالمفاضلة في اختبارات القبول أو الاختبارات الخاصة بالترقية أو بالقدرات والتي يصنع فيها القرار بناءً على نتائج تلك الاختبارات، وذلك كقرارات المفاضلة بين الأشخاص أو التصنيف أو الاختيار.

وتعدد الصور للاختبار الواحد أمر جيد ومطلوب إذا روعي فيه تكافؤها، ولكن تكافؤ صور الاختبار يتطلب إجراءات لا بد القيام بها. حيث أنه من المتوقع أن يكون هناك اختلاف في مستويات الصعوبة وفي القدرات التي تتطلبها عند عمل صورتي أو أكثر للاختبار، وبالتالي يصعب المقارنة المباشرة بين درجات المفحوصين الذين أجابوا على تلك الصور المختلفة، مما يستوجب تحويلها لميزان مشترك تجعلها متكافئة (طومان وآخرون، 2006). ويتم ذلك من خلال عملية إحصائية تسمى بمعادلة الدرجات، والتي تستخدم لتحقيق العدالة والانصاف وبالتالي اتخاذ قرارات صحيحة. ويتم فيها إزالة الفروق بين مستويات الصعوبة في الاختبارات. وذلك لتسهيل مقارنة الأشخاص (De Ayala, 2022)

ويوصي جاد الله وآخرون (2021) باستخدام نظرية الاستجابة للمفردة وتطبيقاتها في بناء الاختبارات التحصيلية ومعادلة درجاتها. كما يوصي اللمع وعودة (2012) بإجراء مزيد من الدراسات في مجال المعادلة وفق نماذج النظرية الحديثة للحكم على دقة نتائج المعادلة.

كما أوصت دراسة حمادنة (2013) إجراء دراسة تأخذ بعين الاعتبار اختلاف طريقة المعادلة واختلاف النموذج المستخدم، فقد تتأثر دقة المعادلة بالنموذج المستخدم أو بالطريقة المستخدمة في المعادلة.

ومن خلال اطلاع الباحث على الدراسات السابقة، وجد بأن هناك حاجة للمزيد من البحث والتقصي في موضوع دقة طرق المعادلة وفق النظرية الحديثة للوصول لأفضل نماذج نظرية الاستجابة للمفردة وأفضل طرق معادلة درجات الاختبار في دقة المعادلة.

لذلك تبلورت مشكلة الدراسة الحالية في التحقق من أثر نموذج نظرية الاستجابة للمفردة في دقة معادلة الدرجات، وذلك من خلال توليد بيانات من خلال برمجية WinGen-3 للإجابة على أسئلة الدراسة.

ويمكن أن تلخص مشكلة الدراسة في السؤال البحثي التالي:

ما أثر نوع نموذج نظرية الاستجابة للمفردة في دقة المعادلة؟

ويتفرع منه السؤالان التاليان:

1- ما درجة دقة معادلة درجات صورتين اختباريتين عند اختلاف نموذج الاستجابة للمفردة في تقدير القدرة؟
2- هل توجد فروق ذات دلالة إحصائية بين متوسطي جذر متوسط مربعات الخطأ RMSE لطريقة المتوسط/المتوسط تعزى لنوع النموذج؟

1-3-أهداف الدراسة:

تهدف الدراسة الحالية إلى التعرف على:

1- درجة دقة معادلة درجات صورتين اختباريتين عند اختلاف نموذج الاستجابة للمفردة في تقدير القدرة.
2- الفروق ذات الدلالة الإحصائية بين متوسطي جذر متوسط مربعات الخطأ RMSE تعزى لنوع النموذج المستخدم.

1-4-أهمية الدراسة:

- معرفة مزيد حول طرق معادلة الدرجات وفق نظرية الاستجابة للمفردة والتي تعتبر من أهم تطبيقات هذه النظرية استخدام نماذج حديثة في القياس.
- إيجاد أفضل الحلول التي تجعل من صور الاختبار الواحد متعادلة.
- الاسهام في إلقاء الضوء على مدى دقة المعادلة عند اختلاف نوع النموذج وفق نظرية الاستجابة للمفردة.
- يساعد المهتمين بمجال القياس والتقويم في معرفة أفضل نماذج نظرية الاستجابة للمفردة في معادلة درجات الاختبار.
- تتكامل هذه الدراسة مع الدراسات والبحوث التي قامت بمقارنة طرق المعادلة وفق نظرية القياس التقليدية والحديثة والتي قد تساعد في الوصول إلى إطار معرفي متسق.
- تقدم صورة واضحة للمتخصصين في مراكز اعداد الاختبارات والمقاييس المختلفة والمعلمين حول مدى دقة طريقة المتوسط/المتوسط في ظل نموذج نظرية الاستجابة للمفردة المستخدم.

1-5-حدود الدراسة:

- اقتصرت الدراسة على استخدام طريقة المتوسط/المتوسط في معادلة درجات الاختبار.
- توليد بيانات وفق النموذج الثنائي والثلاثي البارمتر عن طريق برمجية WinGen3.
- تصميم المجموعات المتكافئة ذات الجذع المشترك.
- المحك المستخدم للحكم على دقة المعادلة في الدراسة متوسط مربع الأخطاء المعيارية RMSE.

1-6-مصطلحات الدراسة:

- **النموذج الثنائي البارامتر Two Parameter Logistic Model :** "وهو تعميم للنموذج اللوجستي أحادي المعلمة حيث يفترض هذا النموذج أن المفردة تختلف في صعوبتها وتمييزها، وعدم تأثرها بالتخمين" (Han & Hambleton, 2007).
- **النموذج الثلاثي البارامتر Three Parameter Logistic Model:** "أحد نماذج نظرية الاستجابة للمفردة الاختبارية ثنائية التدريج، والذي يتم فيه تقدير ثلاثة معالم للمفردة: معلم الصعوبة، معلم التمييز، ومعلم التخمين" (سليمان والصالح، 2016).
- **معادلة درجات الاختبار Test Equating:** "هي عملية إحصائية تحول الاختلافات في الصعوبة بين نماذج الاختبارات التي تم بناؤها على نفس المواصفات لتصبح الدرجات في النماذج قابلة التبادل" .(Kolen & Brennan, 2014).
- **دقة المعادلة Equating accuracy:** "امكانية الحصول على أقل قدر من مقدار الخطأ عندما يتم معادلة درجات الاختبارات، سواء أكان ذلك الخطأ منتظماً أو عشوائياً" (عبدالوهاب، 2015).

2- الإطار النظري والدراسات السابقة

2-1-الإطار النظري:

2-1-1-نماذج نظرية الاستجابة للمفردة

يندرج تحت نظرية الاستجابة للمفردة العديد من النماذج الرياضية والتي بدورها تحقق القياس الموضوعي، وتعالج كثير من المشكلات التي واجهت نظرية القياس التقليدية، ومن أكثر هذه النماذج شيوعاً واستخداماً في مجال القياس والتقويم النفسي والتربوي، النموذج اللوغاريتمي الأحادي المعلمة، النموذج اللوغاريتمي الثنائي المعلمة، والنموذج اللوغاريتمي الثلاثي المعلمة (الناغي، 2011).

وبسبب اختلاف الافتراضات التي تتعلق بالبيانات، تشتمل هذه النظرية على عديد من النماذج، ففي حال أن الأداء في الاختبار يقيس سمة واحدة، يمكن استخدام أحد النماذج التابعة لنظرية الاستجابة للمفردة أحادية البعد، بحيث إذا كانت الاستجابات ثنائية، تستخدم نماذج نظرية الاستجابة للمفردة أحادية البعد ثنائية الاستجابة، وفي حال كانت الاستجابات متعددة، ننتقل إلى النماذج متعددة الاستجابة، أما في حال كان هناك أكثر من بعد أو سمه فيتم استخدام النماذج التابعة لنظرية الاستجابة للمفردة متعددة الأبعاد، ولكل نموذج من النماذج المختلفة خصائصه وافتراضاته.

ويوضح القضاة وآخرون (2020) أنه يتم الانتقال إلى النماذج أحادية البعد عندما يكون هناك سمة أو قدرة واحدة تكمن خلف استجابات الأشخاص على فقرات الاختبار، بحيث أن التباين بين الاستجابات يرجع إلى الفروق في هذه السمة، أما النماذج متعددة الابعاد فهي تعتبر امتداد إلى النماذج السابقة، ولكنها تفترض أن هناك أكثر من بعد يكمن خلف استجابات الأشخاص على فقرات الاختبار.

2-1-2-نماذج نظرية الاستجابة للمفردة أحادية البعد ثنائية الاستجابة:

- النموذج اللوغاريتمي أحادي المعلمة.
- النموذج اللوغاريتمي ثنائي المعلمة.
- النموذج اللوغاريتمي ثلاثي المعلمة.
- النموذج اللوغاريتمي رباعي المعلمة.
- النموذج اللوغاريتمي خماسي المعلمة.

وأكثر هذه النماذج انتشاراً واستخداماً هي النماذج اللوجستية الأحادية المعلم، الثنائية المعلم، والثلاثية المعلم (Bock & Gibbons, 2021).

وفيما يلي بعض التفاصيل:

2-1-3-النموذج اللوجستي أحادي المعلم (نموذج راش) One Parameter Logistic Model (Rasch Model)

يعد هذا النموذج من أكثر نماذج نظرية الاستجابة للمفردة استخداماً، نظراً لبساطته مقارنة بغيره، ويسمى بنموذج المعلمة الواحدة، لاحتوائه على معلمة الصعوبة فقط لتمثيل الفروق بين المفردات، ويفترض أن تتشابه المفردات في التمييز، ولا تتأثر بعملية التخمين (Hableton & Swaminthan, 1985)، وغالباً ما يطلق على النموذج اللوجستي الأحادي البارامتر اسم نموذج راش، مع إن شكل نموذج راش يختلف عن هذا النموذج لكنه يعادله رياضياً (Hambleton et al., 1991)، ويعتبر هذا النموذج حالة خاصة من النموذج الثلاثي البارامتر، وذلك عندما يكون التمييز متساوي والتخمين مساوي للصفر (Lord, 1980).

والوحدة التي تستخدم في قياس القدرة الكامنة وصعوبة المفردة في النماذج اللوجستية والنموذج الأحادي المعلمة بشكل خاص هي اللوجت، والتي يمكن حسابها من خلال أخذ ناتج قسمة احتمال الإجابة عن المفردة، على احتمال عدم الإجابة عنها وإيجاد اللوغاريتم الطبيعي لهذا الناتج (التقي، 2013).

ويمكن التعبير عن هذا النموذج رياضياً من خلال الصيغة التالية:

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$$

حيث:

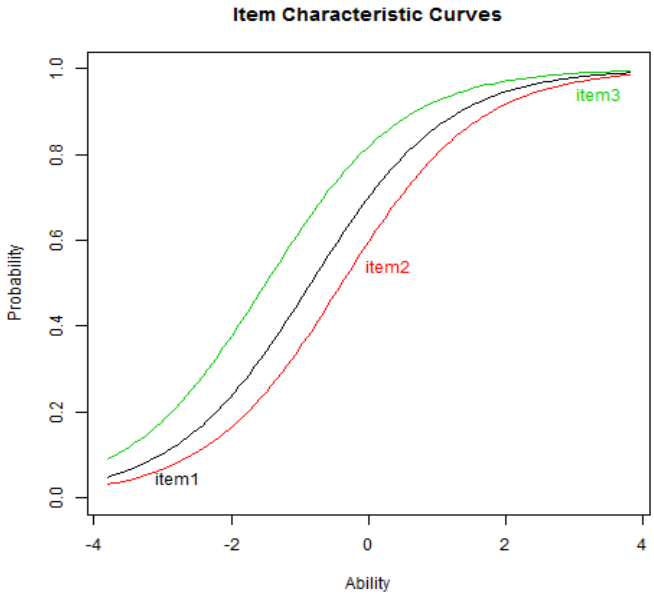
$P_i(\theta)$= احتمالية إجابة المفحوص ذو القدرة (θ) إجابة صحيحة على الفقرة (i).

θ = قدرة الفرد.

b_i= معلم صعوبة الفقرة.

D= 1.7

والشكل التالي يمثل منحنيات خصائص (3) مفردات وفق النموذج أحادي المعلم حيث يفترض بأن التمييز موحد، ولا يوجد تخمين، ويمكن ملاحظة تشابه التمييز حيث المنحنيات متوازية عند نقطة الانقلاب ولا تتقاطع، كما يفترض هذا النموذج عدم وجود تخمين من قبل منخفضي القدرة، ويمكن ملاحظة ذلك من خلال خروج المنحنيات للمفردات الثلاثة من النقطة التي احداثياتها صفر، ويظهر أن الاختلاف فقط في معلم الصعوبة، وهي النقطة التي يحدث عندها انقلاب في المنحنى، ويتضح بأن المفردة رقم (2) تبدو أكثر صعوبة من المفردتين (1) و(3) حيث انقلب المنحنى عند قيمة أكبر من انقلاب المنحنيات للمفردتين (1) و(3).



شكل (1) منحنى خصائص المفردة لثلاث مفردات وفق نموذج راش

2-1-4-النموذج اللوجستي ثنائي المعلمة Two Parameter Logistic Model

وهذا النموذج يفترض أن المفردة تختلف في صعوبتها وتمييزها، كما أنه يفترض عدم تأثرها بالتخمين. ويعتبر تعميماً للنموذج اللوجستي أحادي المعلمة، فبدلاً من ثبات التمييز والذي يأخذ القيمة (1) عبر كل المفردات في النموذج أحادي المعلمة، في هذا النموذج لكل فقرة تمييزها الخاصة بها (Han & Hambleton, 2007).

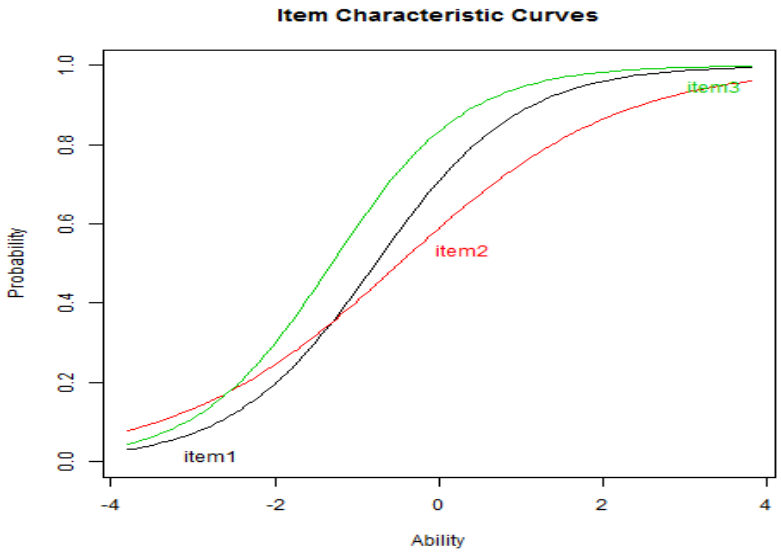
ويعبر عنه بالصيغة الرياضية التالية:

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

حيث:

a_i= يمثل معلم تمييز الفقرة.

والشكل التالي يوضح ثلاث مفردات مختلفة في تمييزها، فيوجد هناك تقاطع بين المنحنيات، وكذلك تختلف في صعوبتها، حيث تختلف نقاط انقلاب المنحنيات، ولكن يمكن ملاحظة عدم تأثرها بالتخمين وهو ما يفترضه هذا النموذج.



شكل (2) منحنى خصائص المفردة لثلاث مفردات وفق النموذج الثنائي البارامتر.

2-1-5-النموذج اللوجستي ثلاثي المعلمة Three Parameter Logistic Model

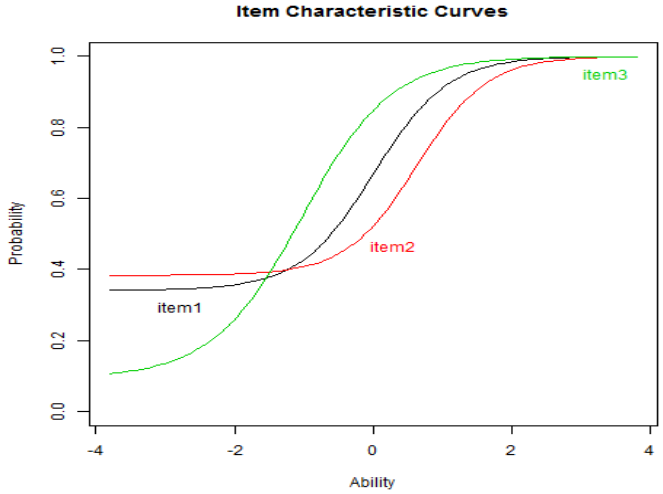
في عام (1968) قام بيرنباوم Birnbaum بتعديل النموذج اللوجستي ثنائي المعلمة بإضافة معلمة ثالثة وهي معلمة التخمين، حيث افترض أن التخمين يساهم في الإجابة على المفردة بشكل صحيح، وأصبح النموذج الناتج معروف باسم النموذج الثلاثي المعلمة (Baker, 2001). وهو أحد نماذج نظرية الاستجابة للمفردة الاختبارية ثنائية التدريج، والذي يتم فيه تقدير ثلاثة معالم للمفردة: معلم الصعوبة، معلم التمييز، ومعلم التخمين (الصالح وسليمان، 2016).

ويأخذ هذا النموذج الصيغة الرياضية التالية:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

ويمثل c_i خط التقارب السفلي لمنحنى خصائص الفقرة.

ونلاحظ بالشكل أدناه ثلاث مفردات مختلفة في صعوبتها وتمييزها وتأثرها بعامل التخمين، حيث اختلاف نقطة الانقلاب في كل منحنى والذي يدل على اختلاف مستوى الصعوبة، والتقاطع بين المنحنيات يشير إلى اختلاف التمييز، وكذلك نلاحظ عدم خروج المنحنيات من النقطة التي احداثياتها صفر، وهذا يدل أن هناك أثر للتخمين.



شكل (3) منحنى خصائص المفردة لثلاث مفردات وفق النموذج الثلاثي البارامتر

2-1-6-معادلة درجات الاختبار Test Scores Equating

غالبًا ما يتم استخدام نماذج بديلة من فترة إلى أخرى في الاختبارات عامة والمصيرية خاصة، من أجل عدم الإلمام بالأسئلة والحفاظ على محتواها، وهذه النماذج قد تعطي نتائج مضللة يبنى عليها قرارات خاطئة، ومعادلة درجات الاختبار تكون مهمة لهذا النوع من الاختبارات (Bock & Gibbons, 2021)، ولذلك يكون الغرض من المعادلة هو تسهيل مقارنة الأشخاص (De Ayalaa, 2022).

وتشكل معادلة درجات الاختبار جزءاً هاماً من مهام المختصين في القياس والتقويم، حيث قاموا في السنوات الأخيرة بعمل العديد من الأبحاث وبشكل واسع اعترافاً بأهمية المعادلة في بناء الاختبارات، والتي يمكن أن تجرى وفق نظرية القياس التقليدية أو نظرية الاستجابة للمفردة التي تختلف اختلافاً كبيراً في الافتراضات والإجراءات الحسابية عن النظرية التقليدية (الزعابي وآخرون، 2020).

ومعادلة درجات الاختبار هي عملية إحصائية تستخدم لضبط درجات الاختبار في الصور المختلفة، بحيث يمكن استخدام الدرجات في الصورتين بشكل تبادلي، فهي تقوم بضبط الاختلافات في الصعوبة بين الصور، لجعلها متشابهة في الصعوبة والمحتوى (Kolen & Brennan, 2014).

ومعادلة درجات الاختبار وفق نظرية الاستجابة للمفردة عندما يكون النموذج المستخدم مطابق للبيانات المعدة لعملية المعادلة، يمكنه حل الكثير من المشاكل التي عجزت نظرية القياس التقليدية عن حلها (المحروق، 2016).

وقد تكون المصطلحات المستخدمة في وصف المعادلة مربكة نوعا ما، حيث في كثير من الحالات نجد من يعتبر ربط الدرجات مرادف لمعادلة الدرجات، ولكن معادلة الدرجات نوع من أنواع الربط وهو أقوى صور الربط (Bock & Gibbons, 2021).

شروط المعادلة Conditions of Equivalency (Lord,1980; Angoff,1984; Kolen & Bernann,2004; Yang& Houang,1996)

- يجب أن يقيس كلا الاختبارين نفس السمة أو المهارة، حيث لا يمكن معادلة الاختبارات التي تقيس سمات مختلفة.
- يجب تساوي الثبات فلا يمكن معادلة درجات الاختبارات مختلفة الثبات.
- تحقق العدالة ويعني ذلك أن كل مجموعة من الأفراد لهم نفس القدرة، بحيث لا يهم أي اختبار طبق عليهم، بحيث يحصل الافراد على نفس الدرجات على النموذجين، سواء طبق عليهم النموذج الأول أو الثاني، وبذلك يكون التوزيع التكراري المشروط على النموذج الأول هو نفسه على النموذج الأخر.
- يجب أن يكون تحويل الدرجات من نموذج إلى نموذج قابل للعكس من أجل تحقيق التماثل، وهذا ما يطلق عليه التماثل.
- ينبغي أن تكون نتائج المعادلة مستقلة عن خصائص العينة المستخدمة في عملية المعادلة، أي ينبغي ألا تتغير نتائج المعادلة عند استخدام مجموعة أخرى من المفحوصين، ويمكن التأكد من هذا الشرط من خلال فحص علاقة التكافؤ عبر المجموعات الفرعية ويسمى ذلك اللاتغاير في مجتمع الدراسة.
- إذا كانت المعادلة وفق نظرية الاستجابة للمفردة يتطلب تحقق أحادية البعد في الدرجات، وقد لا يعتبر هذا الشرط شرطاً أساسياً من شروط المعادلة بشكل عام، لكنه من الافتراضات الأساسية لنظرية الاستجابة للمفردة، ولذلك فهو مطلوب عند المعادلة وفق نظرية الاستجابة للمفردة.

2-1-7-خطوات تنفيذ المعادلة Steps of Implementing Equating

يمكن اتباع الخطوات التالية عند تنفيذ المعادلة والتي من الممكن أن يختلف الترتيب عند التطبيق (Kolen & Bernnan, 2014):

- اتخاذ القرار حول الهدف من المعادلة.
- بناء النماذج البديلة بحيث تكون بنفس المواصفات والمحتوى.
- اختيار تصميم جمع البيانات المناسب.
- القيام بجمع البيانات وفق التصميم المختار.
- اختيار التعريف الاجرائي للمعادلة.
- اختيار طريقة التقدير المناسبة مثل طريقة تكر أو طريقة ليفين.
- تقييم نتائج المعادلة.

طريقة المتوسط - المتوسط Mean-Mean Method

قدم لويد وهوفر (Loyd & Hoover, 1980) طريقة مختلفة قليلاً عن الطريقة السابقة، سميت بطريقة المتوسط - المتوسط، وذلك لتقدير كل من المعاملات A وB حيث استخدما متوسط بارامتر التمييز لتحديد منحنى المعادلة (A) ومتوسط بارامتر الصعوبة لتحديد ثابت المعادلة (B):

$$A = \frac{\mu(\alpha_i)}{\mu(\alpha_j)}$$

$$B = \mu(b_j) - A\mu(b_i)$$

حيث:

$\mu(\alpha_i)$ =متوسط بارامترات التمييز للمقياس (i).

$\mu(b_i)$ = متوسط بارامترات الصعوبة للمقياس (i).

$\mu(\alpha_j)$ =متوسط بارامترات التمييز للمقياس (j).

$\mu(b_j)$ = متوسط بارامترات الصعوبة للمقياس (j).

A = الميل.

B = القاطع.

بعد ذلك يمكن استبدال القيم التي تم الحصول عليها من (A) و(B) في المعادلة التالية:

$$\theta_{J_j} = A^*\theta_{I_j} + B^*$$

تقييم نتائج المعادلة Evaluating Equating Results

يتبقى خطوة أخيرة بعد الانتهاء من معادلة درجات الاختبار، وهي معرفة ما إذا تمت المعادلة بشكل جيد من عدمه، ويمكن معرفة ذلك من خلال مجموعة من المحكات أو المؤشرات التي توضح مدى دقة المعادلة. وقد استعرض كل من هاريس وكروس Harris & Crouse,1993)) مجموعة من المحكات والمؤشرات التي قد تساعد في تقييم نتائج المعادلة والتي تحدد إلى أي مدى كانت دقة المعادلة.

ولا يوجد محك متفق عليه وحاسم من أجل الحكم على دقة معادلة درجات الاختبار (Kolen, 1981)، وسوف نستعرض بعض من المحكات والمؤشرات التي يمكن استخدامها في الحكم على دقة المعادلة:

2-1-8-الجذر التربيعي لمتوسط مربعات الأخطاء Root-mean-squared of equating

ويختصر ب (RMSE) أو (RMSD) والتي هي اختصار لجذر معدل مربعات الانحرافات، ويعتمد على بواقي الدرجات المعادلة من قياسات محك الدقة للمجتمع (العلي، 2017)، ويعكس هذا المحك مقدار التحيز للمفردة ودقة نتائج المعادلة مقارنة بمعيار المعادلة الذي تم استخدامه (المحروق وعليان، 2011).

ويمكن حسابه من خلال المعادلة التالية:

$$RMSE = \left[\frac{1}{n_s}\sum\frac{1}{n_e}\sum_j (y_i - E_j)^2\right]^{\frac{1}{2}}$$

حيث أن:

n= عدد الأفراد.

j =رمز الفرد.

n_s = عدد العينات المولدة.

y_i = الدرجة على الصورة الاولى للمفحوص *j* في ضوء درجته على الصورة الثانية.

n_e= عدد المفحوصين في العينة الواحدة.

E_j = الدرجة الحقيقية المتوقعة للمفحوص على الصورة الأولى.

2-2-الدراسات السابقة:

تعددت الدراسات التي تناولت استخدام نظرية الاستجابة للمفردة في بناء الأسئلة والكشف عن خصائصها السيكومترية، وفيما يلي سيتم عرض بعض منها مرتبة ترتيبًا تصاعديًا بحسب تاريخها:

- **دراسة جاد الله وآخرون (2021)** هدفت إلى التعرف على أثر اختلاف طريقة المعادلة في معادلة صورتي اختبار تحصيلي وفق نظرية الاستجابة للفقرة، حيث أعد الباحث صورتي اختبار تحصيلي لمادة العلوم، وبلغ حجم العينة (1032) من طلبة الثاني الاعدادي بمنطقة القاهرة الأزهرية، واعتمد في جمع البيانات على تصميم المجموعات غير المتكافئة ذات الجذع المشترك، ووجد الباحث أن بيانات الدراسة تلائم النموذج الثلاثي البارامتر، وتوصل إلى وجود فروق ذات دلالة إحصائية عند مستوى (0.01) بين متوسطات القيم الناتجة من عملية المعادلة وفقًا لطرق المعادلة التابعة لنظرية الاستجابة للمفردة، وكانت طريقة المتوسط/الانحراف المعياري أعلى القيم مقارنة بالطرق الأخرى، كذلك أظهرت النتائج وجود فروق ذات دلالة إحصائية عند مستوى (0.01) بين متوسطات قيم دقة المعادلة لصورتي الاختبار وفقاً لتلك الطرق، كانت طريقة المتوسط/الانحراف المعياري هي أيضاً الأكثر دقة وفق النموذج الثلاثي البارامتر المستخدم.
- **وسعت دراسة أكين أريكان وجيلبال (2021) AKIN-ARIKAN & GELBAL** إلى مقارنة أداء طريقة معادلة نظرية الاستجابة للفقرة وطريقة كيرنيل بناء على أخطاء المعادلة RMSD والأخطاء المعيارية للمعادلة، باستخدام تصميم المجموعات غير المتكافئة ذات الجذع المشترك. تحت بعض الشروط التي شملت، توزيع القدرة، نوع الربط في الجذع المشترك (داخلي أو خارجي)، ونسبة فقرات الجذع المشترك، وأظهرت النتائج أن توزيع القدرة، ونوع الجذع المشترك (داخلي، خارجي)، ونسبة فقرات الجذع المشترك، أثرت على أداء طرق المعادلة، وعندما يكون متوسط الاختلافات في القدرة مرتفعة بين المجموعات، يكون هناك ارتفاع في الأخطاء المعيارية في طرق كيرنيل، بينما تكون الأخطاء المعيارية أقل في طرق معادلة نظرية الاستجابة للمفردة، وعند استخدام الربط الخارجي للجذع المشترك تكون الأخطاء المعيارية وRMSD أقل منها عند استخدام الربط الداخلي، وأيضاً تقل تلك الأخطاء عند ارتفاع نسبة فقرات

الجذع المشترك، وبشكل عام كان لطريقة معادلة نظرية الاستجابة للفقرة معدل أخطاء أقل، لذلك يمكن استخدام طرق كيرنيل عندما لا يتم استيفاء افتراضات نظرية الاستجابة للمفردة.

- **وتهدف دراسة أويسال وكيلمن (2016) Uysal & Kilmen** إلى التحقق من أداء طرق معادلة الاختبار في اطار نظرية الاستجابة للمفردة، وتحقيقا لهذا الهدف، أجريت دراسة محاكاة لمقارنة أخطاء المعادلة لطريقة (المتوسط / المتوسط، المتوسط / سيجما، المتوسط سيجما القوي للين Linn وزملاءه، هيبرا، وستوكنج لورد) تحت ظروف مختلفة، تم استخدام 40 مفردة، وتأثيرات طول الجذع المشترك التالية (10%، 20%، و30%)، وتوزيع القدرة (طبيعي، التواء سالب، التواء موجب) وطبق ذلك على عينه حجمها (1000) مشارك، وتم استخدام تصميم المجموعة غير المتكافئة ذات الجذع المشترك، وكانت الاختبارات وفق النموذج الثلاثي المعلمة لبيانات مولدة ثنائية الاستجابة، والنموذج الجزئي المعمم لبيانات مولدة متعددة الاستجابة، وأظهرت النتائج عن أن طريق المتوسط / سيجما القوية للين وزملاءه احتوت على أعلى أخطاء معادلة، وكان أقل خطاء معادلة حدث عندما استخدمت طريقة ستوكنج لورد في حالة المجموعات ذات التواء موجب وطول اختبار الجذع المشترك بالنسبة للاختبار (30%)، وأشارت النتائج إلى أن المجموعات التي لها نفس توزيع القدرة (طبيعي / طبيعي، التواء سالب / التواء سالب،...) تعطي أخطاء معادلة أقل من تلك التي تختلف في توزيع القدرة.
- **وفي دراسة أجراها المحروق (2016)** هدفت إلى معرفة أثر صعوبة الفقرة وحجم العينة في دقة معادلة الدرجات في إطار نظرية الاستجابة للمفردة، حيث تم استخدام ثلاثة أحجام للعينة: (200، 600، 1000) ومستويين للصعوبة، الأول التشابه في متوسط معدل الصعوبة، والثاني الاختلاف في متوسط معدل الصعوبة، واستخدم تصميم المجموعات العشوائية، وتم توليد بيانات تجريبية باستخدام برمجية WinGen2 وتم معادلة درجات الاختبارات باستخدام طريقة الدرجات الملاحظة في نظرية استجابة المفردة. وأسفرت النتائج عن أن حجم العينات الكبير يقلل من الخطأ المعياري للمعادلة وكذلك يقلل من البواقي المعيارية. كما أوضحت النتائج أن النماذج التي تختلف في صعوبتها تميل قيم الخطأ المعياري وقيم متوسط مربع الأخطاء المعيارية إلى الانخفاض، والنماذج التي تتشابه في صعوبتها تميل قيم الخطأ المعياري وقيم متوسط مربع الأخطاء المعيارية إلى الارتفاع.
- **وهدفت دراسة (اللمع، 2012)** إلى تقصي الأثر المشترك لمستوى الصعوبة والتمييز لمفردات الاختبار المستهدف والمشترك في دقة المعادلة وذلك عند استخدام تصميم المجموعات المتكافئة ذات الفقرات المشتركة، وتم استخدام برنامج Wingen3 لتوليد بيانات الدراسة وفق النموذج الثلاثي البارامتر، واستخدمت طريقة المعادلة الخطية الأفقية للمعادلة، وأظهرت النتائج أن هناك أثر ذو دلالة إحصائية لتقديرات الأخطاء المعيارية للقيم المعادلة يعزى لمتغيري الدراسة، وأن مستوى صعوبة وتمييز المفردات يؤثر على دقة المعادلة.

3-منهجية الدراسة وإجراءاتها.

3-1-منهج الدراسة:

نظراً لطبيعة الدراسة وبياناتها استخدم الباحث المنهج التجريبي للإجابة على تساؤلات الدراسة والمتعلقة بدراسة أثر نوع نموذج نظرية الاستجابة للمفردة في دقة معادلة الاختبارات، ولضبط تأثير عدد كبير من العوامل التي قد تؤثر في معادلة درجات صورتين اختباريتين مثل شكل توزيع القدرة ومدى بارامترات الفقرات وطريقة تقدير القدرة ووجود بيانات مفقودة تم استخدام بيانات مولدة. ويتم استخدام دراسات المحاكاة للمقارنة بين أثر نماذج نظرية الاستجابة في دقة المعادلة تحت عوامل محددة، لأن الحالات في البيانات الحقيقية تكون معقدة وتحتوي على عدد عوامل أكبر (Harris & Crouse, 1993). وعندما تصمم دراسات المحاكاة بشكل حذر وبتأني، من الممكن أن تمثل بيانات حقيقية، لذلك هي توفر نموذج قبلي للبيانات الحقيقية (Uysal & Kilmen, 2016).

3-2-توليد البيانات:

تم استخدام برمجية WinGen3 في توليد بيانات ثنائية الاستجابة ل (6000) فرد على (200) مفردة في 4 صور اختبارية كما في الجدول (1). صورتين اختباريتين وفق النموذج الثنائي البارامتر والأخرى وفق النموذج الثلاثي البارامتر، وكررت هذه العملية (30) مرة، واستخدمت الفقرات المشتركة لمعرفة العلاقة بين صور الاختبارات.

جدول (1) صورتين اختباريتين وفق النموذج الثنائي البارامتر والأخرى وفق النموذج الثلاثي البارامتر

نوع النموذج	النموذج الثنائي البارامتر		النموذج الثلاثي البارامتر	
الصورة	الأولى	الثانية	الأولى	الثانية
عدد المفحوصين	1500	1500	1500	1500
عدد الفقرات الفريدة	40	40	40	40
الفقرات المشتركة	10	10	10	10
عدد الفقرات الكلي	50	50	50	50
عدد العينات المولدة	30	30	30	30

واتبعت معالم القدرة التوزيع الطبيعي بمتوسط قدره (0) وانحراف معياري (1)، أما بارامترات التمييز، والصعوبة، والتخمين اتبعت التوزيع المنتظم بحيث تراوحت قيم التمييز بين (0.01 إلى 2.0) وقيم الصعوبة بين (3 إلى 3-)، أما قيم التخمين فتراوحت بين (0.02 إلى 0.25)، وكما هو واضح في الجدول (1) تكون كل اختبار من عينة حجمها (1500) فرد ومن فقرات عددها (50) فقرة، كانت بينها (10) فقرات مشتركة بين كل صورتي اختبار لكل نموذج. حيث يؤكد كولن وبيرنان (Kolen and Brennan, 2004) أنه عند المعادلة وفق النموذج الثلاثي البارامتر، فإن ذلك يتطلب عينة حجمها (1500) وذلك عند استخدم تصميم المجموعات غير المتكافئة ذات الجذع المشترك، كما أكدا على أن طول الاختبار عند معادلة درجات الاختبار يجب أن يكون على الأقل من 30-40 فقرة.

وروعي في عدد الفقرات المشتركة اتباع ما أشار إليه (Cook & Eignor, 1991; Kolen & Brennan, 2014)، أنه يجب أن تكون الفقرات المشتركة لا تقل عن (20%) من مجموع الفقرات الكلي.

تم التحقق من افتراضات نظرية الاستجابة للمفردة في بيانات الدراسة من خلال حزمة Sirt وبعد ذلك تم تقدير بارامترات الأفراد والفقرات من خلال حزمة Mirt بواسطة برنامج R وذلك لمعرفة مدى ملائمة الأفراد والفقرات للنموذج المستخدم، حيث تم التحقق من ملائمة الأفراد للنموذج والتحقق من ملائمة المفردات للنموذج المستخدم من خلال حساب مربع كاي (χ^2) والذي يتطلب أن يكون غير دال احصائياً، أكبر من (0.05) حتى تكون المفردة مطابقة للنموذج، ووجد بأن هناك مجموعة من الأفراد والمفردات الغير مطابقة للنموذج المستخدم وتم حذفها، ويوضح الجدول (8) أول ثلاثة صور اختبارية للعدد النهائي للأفراد والمفردات المطابقة، حيث كان العدد الأصلي (1500) فرد و(50) مفردة لكل صورة اختبار.

جدول (2) الأعداد المطابقة لكل صورة من الأفراد والمفردات

النموذج الثنائي البارامتر					
العينة	الصورة الأولى (أ)		العينة	الصورة الثانية (ب)	
	المفردات	الأفراد		المفردات	الأفراد
1	48	1445	1	49	1455
2	48	1450	2	49	1463
3	48	1450	3	48	1455
النموذج الثلاثي البارامتر					
العينة	الصورة الأولى (أ)		العينة	الصورة الثانية (ب)	
	المفردات	الأفراد		المفردات	الأفراد
1	47	1455	1	47	1458
2	47	1470	2	46	1459
3	46	1472	3	47	1457

وتم تقدير بارامترات الأفراد والمفردات باستخدام طريقة الاحتمال الأقصى من خلال برنامج PARSCALE، وذلك لتوفير الامتداد المطلوب لبرنامج IRTEQ للبارامترات، وتم القيام بعملية معادلة الدرجات من خلال برنامج IRTEQ باستخدام طريقة المتوسط /المتوسط Mean/Mean وفق النموذج الثنائي البارامتر والنموذج الثلاثي البارامتر.

3-3-المعالجة الإحصائية:

للإجابة عن أسئلة الدراسة استخدم الباحث البرامج التالية:

1- برنامج R. وهو برنامج مفتوح المصدر يسمح بإجراء التحليل الاحصائي من خلال عدة حزم إحصائية واستخدم في هذه الدراسة:
- للتحقق من مطابقة الأفراد والمفردات للنموذج.
- للتحقق من افتراضات نظرية الاستجابة للمفردة.

2- برنامج PARSCALE. برنامج تم تطويره في أوائل التسعينيات من قبل Eiji Muraki بعد ذلك طوره Bock ويستخدم في تقدير معالم المفردات الثنائية والمتعددة الاستجابة معتمداً على أسلوب الأرجحية العظمى الهامشية (Du Toit, 2003). واستخدم في هذه الدراسة لتقدير بارامترات الأفراد والمفردات.

3- برنامج IRTEQ. تطبيق مجاني يعمل على نظام الويندوز تم تطويره بواسطة هان T. Han يقوم بالتدريج والمعادلة وفق نظرية الاستجابة للمفردة ويستخدم عدة طرق للمعادلة كالمتوسط- المتوسط والمتوسط – سيجما والمتوسط - سيجما القوي، وطريقة هيبرا وستكونج لورد، كما يدعم النماذج الشائعة من نماذج نظرية الاستجابة للمفردة الأحادية البعد سواء الثنائية أو المتعددة الاستجابة لإجراء معادلة درجات الاختبار (Han, 2009).

4- برنامجSPSS. للكشف عن وجود فروق ذو دلالة إحصائية.

4-نتائج البحث ومناقشتها.

للإجابة عن أسئلة الدراسة، تم إعادة تقدير بارامترات الأفراد والفقرات بعد حذف الأفراد والفقرات الغير مطابقة للنموذج المستخدم، بواسطة برنامج PARSCALE لكي تتوافق مخرجاته مع متطلبات برنامج IRTEQ، حيث تم اجراء معادلة صور الاختبار باستخدام برنامج IRTEQ بطريق (المتوسط - المتوسط Mean-Mean) وفق النموذج الثنائي البارامتر والنموذج الثلاثي البارامتر، وكان ذلك لجميع الصور الاختبارية، وتم حساب معاملي المعادلة A وB Equating coefficients بواسطة البرنامج وكانت النتائج على النحو التالي:

4-1-نتائج معاملي المعادلة A وB وفق النموذج الثنائي البارامتر والثلاثي البارامتر

جدول (3) النموذج الثنائي البارامتر

الطريقة / الصورة،العينة	طريقة المتوسط/المتوسط	
	A	B
أ1،ب1	0.935	-0.119
أ2،ب2	0.893	0.203
أ3،ب3	1.093	0.251
أ4،ب4	0.891	-0.212
أ5،ب5	1.059	-0.142
أ6،ب6	1.135	-0.199
أ7،ب7	0.968	-0.29
أ8،ب8	1.013	-0.08
أ9،ب9	1.019	0.028
أ10،ب10	0.89	-0.064
أ11،ب11	1.124	-0.305
أ12،ب12	1.003	0.132
أ13،ب13	0.85	-0.129
أ14،ب14	1.102	-0.212
أ15،ب15	1.032	-0.313
أ16،ب16	1.084	0.14

الطريقة / الصورة،العينة	طريقة المتوسط/المتوسط	
	A	B
أ17،ب17	1.233	-0.483
أ18،ب18	1.084	-0.11
أ19،ب19	0.983	0.087
أ20،ب20	1.217	0.403
أ21،ب21	1.003	-0.11
أ22،ب22	1.034	0.091
أ23،ب23	1.043	-0.269
أ24،ب24	1.096	-0.18
أ25،ب25	1.166	-0.283
أ26،ب26	0.989	0.06
أ27،ب27	1.025	0.21
أ28،ب28	0.873	-0.458
أ29،ب29	1.109	0.031
أ30،ب30	0.901	0.322
المتوسط	1.03	-0.064

جدول (4) النموذج الثلاثي البارامتر

الطريقة / الصورة،العينة	طريقة المتوسط/المتوسط	
	A	B
أ1،ب1	0.948	-0.199
أ2،ب2	0.945	0.008
أ3،ب3	0.971	0.124
أ4،ب4	0.984	-0.031
أ5،ب5	0.979	-0.229
أ6،ب6	0.932	-0.003
أ7،ب7	0.954	0.045
أ8،ب8	1.005	-0.11
أ9،ب9	0.978	0.023
أ10،ب10	1.034	0.121
أ11،ب11	1.023	-0.059
أ12،ب12	0.957	0.057
أ13،ب13	0.981	0.108
أ14،ب14	0.899	-0.003
أ15،ب15	1	0.164
أ16،ب16	0.974	0.056

الطريقة / الصورة.العينة	طريقة المتوسط/المتوسط	
	A	B
أ17،ب17	0.946	0.12
أ18،ب18	0.973	-0.046
أ19،ب19	0.975	0.176
أ20،ب20	1.052	0.084
أ21،ب21	0.946	-0.066
أ22،ب22	1.064	-0.05
أ23،ب23	1.083	0.249
أ24،ب24	1.04	0.018
أ25،ب25	1.034	0.133
أ26،ب26	0.974	-0.054
أ27،ب27	0.982	-0.029
أ28،ب28	0.985	-0.075
أ29،ب29	0.984	-0.065
أ30،ب30	0.895	-0.056
المتوسط	0.983	0.014

وللحكم على دقة نتائج المعادلة اعتمد الباحث على جذر متوسط مربعات الخطأ RMSE والذي يتطلب استخراج الدرجة الحقيقية المتوقعة لكل مفحوص، وتم تقدير الدرجة الحقيقية وفقاً للمعادلة التالية:

$$E_j = \sum_i P_i(\theta)$$

حيث:

E_j= الدرجة الحقيقية المتوقعة، والتي تساوي مجموع احتمال الإجابة الصحيحة للمفحوص j على جميع فقرات الاختبار.

في النموذج الثنائي البارامتر تم حساب احتمال الإجابة الصحيحة لكل مفحوص من خلال معادلة النموذج التالية:

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

حيث:

a_i= يمثل معلم تمييز الفقرة.

b_i =معلم صعوبة الفقرة.

D= 1.7

أما في النموذج الثلاثي البارامتر فتم حساب احتمال الإجابة الصحيحة لكل مفحوص من خلال المعادلة التالية:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

c_i= خط التقارب السفلي لمنحنى خصائص الفقرة.

وبمساعدة برنامج EXCEL تم حساب احتمالية إجابة كل مفحوص على جميع الفقرات التي أجاب عنها، والتي تمثل الدرجة الحقيقية المتوقعة للمفحوص.

وعندما أجريت المعادلة بين صور الاختبار تكونت لدينا لكل مفحوص درجة معادلة ملاحظة، وذلك لكل طريقة من طرق المعادلة، وتم حساب قيمة جذر متوسط مربعات الخطأ RMSE من خلال المعادلة التالية:

$$RMSE = \left[\frac{1}{n_s}\sum\frac{1}{n_e}\sum_j (y_i - E_j)^2\right]^{\frac{1}{2}}$$

حيث:

n: عدد الأفراد.

j: رمز الفرد.

n_s:عدد العينات المولدة.

y_i : الدرجة على الصورة الأولى للمفحوص j في ضوء درجته على الصورة الثانية.

n_e :عدد المفحوصين في العينة الواحدة.

E_j: الدرجة الحقيقية المتوقعة للمفحوص على الصورة الأولى.

وأعطت قيم جذر متوسط مربعات الخطأ RMSE لطريقة المعادلة المستخدمة نتائج مختلفة ظاهريا، فعندما كانت البيانات ملائمة للنموذج الثنائي البارامتر كانت نتائجRMSE على النحو الموضح بالجدول (5):

جدول (5) قيم RMSE عندما تمت المعادلة للبيانات المطابقة للنموذج الثنائي البارامتر

الصورة_العينة	قيم RMSE	الصورة_العينة	قيم RMSE
أ1،ب1	1.026	أ16،ب16	0.64
أ2،ب2	0.237	أ17،ب17	0.479
أ3،ب3	0.7	أ18،ب18	0.271
أ4،ب4	0.233	أ19،ب19	0.847
أ5،ب5	0.813	أ20،ب20	0.466
أ6،ب6	0.319	أ21،ب21	0.473
أ7،ب7	0.251	أ22،ب22	0.321
أ8،ب8	1.12	أ23،ب23	1.289
أ9،ب9	1.36	أ24،ب24	0.087
أ10،ب10	0.56	أ25،ب25	1.014
أ11،ب11	0.277	أ26،ب26	0.304
أ12،ب12	0.362	أ27،ب27	0.137
أ13،ب13	0.447	أ28،ب28	0.457
أ14،ب14	0.452	أ29،ب29	0.341
أ15،ب15	0.897	أ30،ب30	0.627
		المتوسط	0.56

أما عندما تمت معادلة صور الاختبار وفق النموذج الثلاثي البارامتر كان متوسط قيم جذر متوسط مربعات الخطأ RMSE كما هو مبين بالجدول (6):

جدول (6) قيم RMSE لطرق المعادلة وفق النموذج الثلاثي البارامتر

الصورة_العينة	قيم RMSE	الصورة_العينة	قيم RMSE
أ1،ب1	1.758	أ16،ب16	1.342
أ2،ب2	0.37	أ17،ب17	0.828
أ3،ب3	2.494	أ18،ب18	1.003
أ4،ب4	0.949	أ19،ب19	0.543
أ5،ب5	0.3	أ20،ب20	1.167
أ6،ب6	1.199	أ21،ب21	0.505
أ7،ب7	0.842	أ22،ب22	0.192

الصورة_العينة	قيم RMSE	الصورة_العينة	قيم RMSE
أ8،ب8	1.113	أ23،ب23	0.351
أ9،ب9	2.155	أ24،ب24	1.317
أ10،ب10	0.464	أ25،ب25	0.866
أ11،ب11	1.715	أ26،ب26	0.37
أ12،ب12	0.809	أ27،ب27	1.045
أ13،ب13	0.645	أ28،ب28	1.22
أ14،ب14	2.067	أ29،ب29	1.274
أ15،ب15	1.869	أ30،ب30	0.562
		المتوسط	1.044

وفيما يتعلق بالإجابة عن السؤال الأول، والذي ينص على "ما درجة دقة معادلة درجات صورتين اختباريتين عند اختلاف نموذج الاستجابة للمفردة في تقدير القدرة؟"

فكما هو مبين بالجدولين (5، 6) تم حساب جذر متوسط مربعات الخطأ RMSE بعد أن تمت المعادلة بطريقة المتوسط - المتوسط Mean-Mean وفق النموذج الثنائي البارامتر وكذلك وفق النموذج الثلاثي البارامتر، وكانت قيمة متوسط RMSE عند المعادلة بطريقة المتوسط - المتوسط باستخدام النموذج الثنائي البارامتر تساوي (0.56) وكانت قيمة متوسط RMSE عندما تمت المعادلة بطريقة المتوسط - المتوسط باستخدام النموذج الثلاثي البارامتر (1.044) وكما هو ملاحظ نجد أن قيمة RMSE عند استخدام النموذج الثنائي البارامتر أقل مقارنة باستخدام النموذج الثلاثي البارامتر، وكلما اقتربت قيمة RMSE من الصفر دل على دقة المعادلة (Kim et al., 2019)، لذلك يمكن القول أن طريقة المتوسط-المتوسط تكون أدق عندما تكون البيانات ملائمة للنموذج الثنائي البارامتر مقارنة بالنموذج الثلاثي البارامتر.

4-2-الإجابة على سؤال الدراسة الثاني: "هل توجد فروق ذات دلالة إحصائية بين متوسطي جذر متوسط مربعات الخطأ RMSE لطريقة المتوسط/المتوسط تعزى لنوع النموذج؟

وللإجابة على هذا السؤال تم استخدام اختبار ت لعينتين مستقلتين للتعرف على الفروق بين متوسطي قيم جذر متوسط مربعات الفروق RMSE الناتجة من عملية المعادلة باستخدام طريقة المتوسط/المتوسط Mean/Mean وفق النموذج الثنائي البارامتر والثلاثي البارامتر، ويوضح الجدولين (7، 8) نتائج الاختبار:

جدول (7) نتائج اختبار ت لعينتين مستقلتين للتعرف على الفروق بين متوسطي قيم جذر متوسط مربعات الفروق RMSE

طريقة المتوسط/المتوسط Mean/Mean	النموذج	العدد	المتوسط	الانحراف المعياري
	النموذج الثنائي البارامتر	30	.5602	.34229
	النموذج الثنائي البارامتر	30	1.0445	.59994

جدول (8) قيمة "ت" والدلالة الإحصائية لمعرفة اتجاه الفرق بين متوسطي RMSE الناتجة من عملية المعادلة

F	Sig.	t	df	Sig. (2-tailed)	Mean Difference
7.014	.010	-3.840	58	0.000	-.48423

عند النظر في الجدول (8) يمكن ملاحظة أن الفرق بين المتوسطين (-0.484) وبلغ مستوى الدلالة (0.000) وهو أقل من (0.05) وهذا يعني أن هناك فروق جوهرية عند استخدام طريقة المعادلة المتوسط/المتوسط تعزى لنوع النموذج، وكانت الفروق باتجاه النموذج الثلاثي البارامتر، حيث كانت قيم جذر متوسط مربعات الخطأ RMSE الناتجة عن هذه الطريقة وفق هذا النموذج أكبر بشكل دال احصائياً منها عند استخدام النموذج الثنائي البارامتر، وذلك يعني أن طريقة المتوسط/ المتوسط وفق النموذج الثلاثي البارامتر أقل دقة، حيث القيمة الأقل لجذر متوسط مربعات الأخطاء RMSE تدل على دقة أعلى حسب ما أشار إليه (Kim et al., 2019).

وقد يعزي الباحث السبب في أن المعادلة أعطت دقة أعلى عندما كانت البيانات ملائمة للنموذج الثنائي البارامتر منها عندما كانت ملائمة للنموذج الثلاثي البارامتر؛ بأن النموذج الثلاثي البارامتر يحمل معلمة ثالثة وهي التخمين وقد تكون السبب في وجود أخطاء أدت إلى تدني الدقة في المعادلة، ولقلة الدراسات التي تناولت هذا الموضوع يوصي الباحث بإجراء دراسات تستخدم طرق معادلة مختلفة مع النموذجين ومقارنة نتائجها مع هذه الدراسة.

التوصيات والمقترحات.

1. أعطت طريقة المعادلة المستخدمة (المتوسط/المتوسط) خطأ أقل عندما استخدم النموذج الثنائي البارامتر مقارنة بالنموذج الثلاثي البارامتر، لذلك عندما تكون البيانات ملائمة للنموذجين يوصي الباحث باستخدام طرق المعادلة وفق النموذج الثنائي البارامتر للحصول على دقة أعلى.
2. يقترح الباحث إجراء دراسات مماثلة وفقا للعناوين التالية:
   أ. دراسة مماثلة على نماذج أخرى من نماذج نظرية الاستجابة للمفردة غير المستخدمة في هذه الدراسة.
   ب. دراسة لمعرفة أثر طرق المعادلة الأخرى على نماذج نظرية الاستجابة للمفردة.
   ج. دراسة مماثلة وفق بيانات حقيقية ومقارنتها مع ما توصلت له هذه الدراسة.
   د. دراسة لمعرفة أثر اختلاف طول الاختبار وعدد الفقرات المشتركة وطريقة معادلة درجات الاختبار على دقة المعادلة.

قائمة المراجع.

أولاً- المراجع بالعربية:

- التقي، أحمد محمد. (2009). *النظرية الحديثة في القياس*، دار المسيرة للنشر والتوزيع.
- جاد الله، أشرف سعد محمد، عبد الواحد، إبراهيم سيد أحمد، وعلي، محمد سعد محمد. (2021). أثر اختلاف طريقة المعادلة في معادلة صورتي اختبار تحصيلي وقف نظرية الاستجابة للمفردة. *مجلة التربية، 2*(192)، 599 - 640.
- حمادنة، راتب صايل الخضر الرحيل. (2013). أثر وجود أداء تفاضلي في الفقرات المرساوية على دقة المعادلة العمودية لاختبار اوتيس لينون للقدرة العقلية. *المجلة التربوية الدولية، 2*(8)، 754-771.
- الزعابي، أمل، المحرزي، راشد، حسن، عبد الحميد. (2020). معادلة كتيبات اختبار الدراسة الدولية لقياس مهارات القراءة (PRILS 2011) بسلطنة عمان باستخدام نظرية الاستجابة للمفردة. *مجلة العلوم التربوية، 15* (15)، 68 -93.
- سليمان، شاهر خالد، والصالح، علي محمد. (2016). بناء اختبار تحصيلي محكي المرجع في القياس والتقويم باستخدام النموذج اللوقاريتمي ثلاثي المعلم لنظرية الاستجابة للمفردة الاختبارية. *مجلة التربية، 1* (168)، 175-201.
- الشواورة، ياسين. (2013). *دراسة مقارنة بين نموذج التقدير الجزئي والاستجابة في معادلة درجات الاختبارات* [دكتوراة غير منشورة]. جامعة القاهرة.
- طومان، منار أحمد، شريف، نادية محمود، وأبو علام، رجاء محمود.(2006) . *فاعلية استخدام نظرية الاستجابة للمفردة (IRT) في عملية معادلة درجات الاختبارات متعددة الأبعاد والمتغيرات المؤثرة فيها* [رسالة دكتوراه غير منشورة]. جامعة القاهرة.
- عبدالوهاب، محمد محمود. (2015). أثر اختلاف معامل ثبات صورتي اختبار في دقة معادلة درجاتها باستخدام الطرق القائمة على نظرية الاستجابة للمفردة .*مجلة كلية التربية،25*(3)، 23-54.
- العلي، ديالا، ودرويش، رمضان محمد. (2017). مقارنة فاعلية طريقتي معادلة براون - هولند والمئينيات باستخدام الجذع المشترك بين صورتي اختبار تحصيلي في الرياضيات. *مجلة جامعة البعث للعلوم الإنسانية 39*(43)، 119 - 148.
- القضاة، بسمة عواد (2010) بناء اختبار تحصيلي في الرياضيات للصف التاسع الأساسي وتقدير خصائصه السيكومترية وفق النظرية الكلاسيكية ونموذج المعلمتين للنظرية الحديثة [رسالة ماجستير غير منشورة]. جامعة مؤتة.
- اللمع، انور نعيم يعقوب، وعودة، أحمد سليمان. (2012). *أثر مستوى الصعوبة والتمييز لفقرات الاختبار المستهدف والاختبار المشترك على دقة المعادلة باستخدام تصميم المجموعات العشوائية* [رسالة دكتوراه غير منشورة]. جامعة اليرموك.
- المحروق، يوسف عبد العاطي محمد. (2016). أثر حجم العينة وطول الاختبار في دقة معادلة درجات الاختبارات متعددة الحدود باستخدام طريقة كيرنيل. *مجلة العلوم التربوية والنفسية، 17* (3)، 201 - 228.
- المحروق، يوسف عبد العاطي، وعليان، خليل محمد. (2011). *مقارنة طرق كيرنيل والمئينات وطرق نظرية استجابة الفقرة عند استخدام تصميم الفقرات المشتركة في دقة معادلة درجات الاختبارات متعددة الحدود* [رسالة دكتوراه غير منشورة]. الجامعة الاردنية.
- المدانات، رائد فايز. (2012). مقارنة فاعلية طريقتي معادلة العلامات الحقيقية والمشاهدة في معادلة الاختبارات باستخدام جذع مشترك ومجموعات غير متكافئة. *مجلة العلوم التربوية والنفسية، 13* (2)، 365 - 394.
- الناغي، هبة إبراهيم.(2011). أثر عدد مفردات الاختبار على تقدير قدرات الأفراد ودالة المعلومات لاختبار تحصيلي مرجعي المحك في الرياضيات باستخدام نماذج نظرية الاستجابة للمفردة، *مجلة كلية التربية*، 10، 676-721.

ثانياً- المراجع بالإنجليزية:

- Akin-Arikan, Ç., & Gelbal, S. (2021). A Comparison of Kernel Equating and Item Response Theory Equating Methods. *Eurasian Journal of Educational Research*, 21(93), 179–198.
- Angoff, W. H. (1984). Scales, norms, and equivalent scores. Princeton, NJ: Educational Testing Service.
- Baker, F.B.. (2001). *The Basics of Item Response Theory*. ERIC Clearinghouse on Assessment and Evaluation, College Park, MD.
- Bock, D. R., & Gibbons, R. D. (2021). *Item Response Theory*. Wiley.
- De Ayala, R. J. (2022). *The Theory and Practice of Item Response Theory, Second Edition*. Guilford Publications.
- De Ayala, R. J., Smith, B. D., & Dvorak, R. N. (2018). A Comparative Evaluation of Kernel Equating and Test Characteristic Curve Equating. *Applied Psychological Measurement, 42*(2), 155–168.
- Dorans, N. J., Moses, T., & Eignor, D. R. (2009). Equating Test Scores: Toward Best Practices. In *Springer eBooks* (pp. 21–42).
- Du Toit, M. (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Scientific Software International.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and Applications*. MA: Kluwer-Nijhoff.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of Item Response Theory. SAGE Publications.
- Han, K. S. (2007). WinGen: Windows Software That Generates Item Response Theory Parameters and Item Responses. *Applied Psychological Measurement, 31*(5), 457–459.
- Han, K. T. (2009). IRTEQ: Windows application that implements IRT scaling and equating [computer program]. *Applied Psychological Measurement*, 33(6), 491-493.
- Han, K. T., & Hambleton, R. K.,.(2007). *User's Manual for WinGen: Windows Software that Generates IRT Model Parameters and Item Responses*, Holt Rinehart and Winston.
- Harris, D. J. & Crouse, J. D. (1993). A study of criteria used in equating. *Applied Measurement in Education*, 6(3), 195-240.
- Inal, H., & Anil, D. (2018). Investigation of group invariance in test equating under different simulation conditions. *Eurasian Journal of Educational Research*, 2018(78), 67–86.
- Kim, K. H., Lim, E., & Lee, W. (2019b). A Comparison of the Relative Performance of Four IRT Models on Equating Passage-Based Tests. *International Journal of Testing, 19*(3), 248–269.
- Kolen, M. J. (1981). Comparison of Traditional and Item Response Theory Methods For Equating Tests. *Journal of Educational Measurement*, *18*(1), 1–11
- Kolen, M. J., & Brennan, R. W. (2004*). Test Equating, Scaling, and Linking: Methods and Practices*. Springer.
- Kolen, M. J., & Brennan, R. W. (2014). Test Equating, Scaling, and Linking. In *Springer eBooks*.
- Lord, F. M. (1980). Applications of Item Response Theory to Practical Testing Problems. Routledge.
- Ozturk-Gubes, N., & Kelecioglu, H. (2016). The impact of test dimensionality, common-item set format, and scale linking methods on mixed-format test equating. *Educational Sciences: Theory and practice*, *16*(3), 715-734.
- Uysal, İ. Ö., & Kilmen, S. (2016). Comparison of Item Response Theory Test Equating Methods for Mixed Format Tests. *International Online Journal of Educational Sciences, 8*(2).
- Von Davier, A. A., Holland, P. R., & Thayer, D. T. (2004). The Kernel Method of Test Equating. In *Springer eBooks*.
- Yang, W., & Houang, R.T. (1996). The Effect of Anchor Length and Equating Method on the Accuracy of Test Equating: Comparisons of Linear and IRT-Based Equating Using an Anchor-Item Design. American Educational Research