فاعلية القياس التكيفي متعدد المراحل باستخدام طريقة دلتا لتقدير الدرجات

أ.رضا أمين غالب مرجان
كلية التربية -جامعة الملك سعود- المملكة العربية السعودية
ameen.readh@gmail.com
أ.د. إسماعيل سلامة البرصان
كلية التربية – جامعة الملك سعود - المملكة العربية السعودية
ibursan@ksu.edu.sa

التقديم: 2023/11/1 القبول: 2023/1/22 النشر: 2023/6/15
Doi: https://doi.org/10.36473/ujhss.v62i2.2074

الملخص

هدفت هذه الدراسة إلى الكشف عن فاعلية القياس التكيفي متعدد المراحل باستخدام طريقة دلتا لتقدير الدرجات. وقد تم استخدام البيانات الناتجة عن تطبيق اختبار أوتيس لينون للقدرة العقلية العامة المستوى المقدم الصورة (J) على طلبة جامعة إب. وبلغ حجم عينة التدريج (1600) طالباً وطالبة، حيث تم تدريج البيانات باستخدام برنامج دلتا. وقد تكونت عينة القياس التكيفي من (130) مفحوصاً، استُخدمت بيانات استجاباتهم في القياس التكيفي الذي تكون من أربع مراحل. تكونت المرحلة الأولى من خمس مفردات، تلاها المراحل الثانية والثالثة والرابعة وتكونت كل منها من خمس مفردات أيضاً. ومن ثم تم حساب الدرجة لكل مفحوص في المراحل الأربع بشكل تراكمي، وتم مقارنتها بدرجة المفحوص المقدرة من الاختبار الخطي بمفرداته الثمانين. وأظهرت النتائج أن معامل الارتباط بين الدرجة المقدرة من الاختبار التكيفي والدرجة المقدرة من الاختبار الخطي بمفرداته الثمانين كان مقداره (0.878)، ومتوسط الفروق في الدرجة المقدرة كان بمقدار (5.728). كذلك بلغ الخطأ المعياري في المتوسط (0.096)، إضافة إلى ان الفرق بين الجذر التربيعي لمتوسط مربعات الخطأ في المتوسط كان (0.053)، الأمر الذي يفيد بفاعلية القياس التكيفي المبنى باستخدام طريقة دلتا.

الكلمات المفتاحية: القياس التكيفي متعدد المراحل، طريقة دلتا لتقدير الدرجات

The Effectiveness of a Multi-Stage Adaptive of Testing by using the Delta Scoring Method

Prof. Redha Ameen Ghaleb Morgan
College of Education - King Saud University- Suadia Arabia
ameen.readh@gmail.com

Prof. Dr. Ismail Salameh Al-Bursan
College of Education - King Saud University- Suadia Arabia
ibursan@ksu.edu.sa

Abstract:
This study aimed to reveal the effectiveness of multi-stage adaptive testing using the Delta- Scoring Method. The data resulting from the application of the Otis-Lennon Test of General Mental Ability (Figure J) was used on Ibb University students. The size of the staging sample was (1600) male and female students, and the data were graded using the Delta program. The adaptive testing sample consisted of (130) students, and the data of their responses were used in the adaptive testing, which consisted of four stages. The first stage consisted of five items, followed by the second, third and fourth stages, each of which consisted of five items as well. Then the score for each student in the four stages was cumulatively calculated, and it was compared to the student's estimated score from the written test with its eighty items. The results showed that the correlation coefficient between the estimated score from the adaptive test and the estimated score from the written test with its eighty items was (0.878), and the average difference in the estimated score was (5.728). The standard error of the mean was (0.096). In addition, the difference between the square root of the mean of the error squares in the mean was (0.053) which indicates the effectiveness of the adaptive testing built using the Delta method.

Keywords: multistage adaptive testing, Delta-Scoring Method.

المقدمة

شهد ميدان القياس النفسي والتربوي بعض التطورات والاتجاهات الحديثة والبارزة التي تستهدف بناء الاختبارات والمقاييس في العديد من المجالات النفسية والتربوية، ومن بين هذه الاتجاهات والتطورات الحديثة استخدام نظرية الاستجابة للمفردة Item Response Theory(IRT) والتي حققت طفرة علمية هائلة ومتطورة في مجال القياس النفسي والتربوي، فقد أكدت العديد من الدراسات العربية والأجنبية تفوق هذه النظرية في تحقيق الموضوعية المأمولة منها في هذا المجال إذا ما قورنت بالنظرية الكلاسيكية Classical Tests Theory (CTT) (De Champlin, 2010)، كما أن هذه النظرية تستند على افتراضات وشروط أقوى من الافتراضات والشروط ذات الصلة بالنظرية الكلاسيكية التي سيطرت على مجريات وإجراءات القياس النفسي والتربوي والتحليلات الخاصة بأدوات القياس لفترة طويلة من الزمن(Rotou & Elmore, 2002).

ومن أهم التطبيقات التي أصبحت ممكنة بعد ظهور نظرية الاستجابة للمفردة الاختبارات التكيفية، والتي كانت حلاً لمشكلة الاختبارات الكلاسيكية والمتمثلة في تعريض الأفراد للمفردات نفسها، بغض النظر عن ملائمة هذه المفردات لقدراتهم، حيث تشمل الاختبارات الكلاسيكية بعض المفردات السهلة، والتي يتعرض لها أفراد من ذوي القدرات العالية، وهذا بحد ذاته يمثل مضيعة للوقت والجهد، كما أن المفحوص قد يشعر بنوع من الملل جراء اختباره بمفردات لا تتحدى قدراته الأمر الذي يدفعُه للإجابة عنها بشكلٍ من عدم الاهتمام، كذلك عند اختبار الأفراد ذوي القدرات المتدنية بمفردات صعبة، مما يشعرهم بالإحباط، وبالتالي الإجابة عنها بشكلٍ عشوائي، وهذا يضيف خطأ جديداً للقياس(Sands et al., 2001).

وتقوم فكرة الاختبار التكيفي على اختيار المفردات اللاحقة من بنك الأسئلة اعتماداً على إجابات المفحوص على المفردات السابقة وبما يتناسب مع قدرة المفحوص، الأمر الذي يجعل الحصول على درجات متسقة لقدرة مفحوصين أمراً متعذراً في حال اعتماد النظرية الكلاسيكية في الاختبارات التكيفية، أما في نظرية الاستجابة للمفردة فإن ذلك ممكن بسبب خاصية تحرر قدرات الأفراد من خصائص المفردات التي يتعرض لها المفحوص Item Free؛ وكذلك بسبب تحرر معالم المفردات من خصائص المفحوص Person Free، بالإضافة إلى ذلك فإن نظرية الاستجابة للمفردة تقدم تقديراً موثوقاً به للقدرة تُقاس دقته بواسطة الخطأ المعياري في التقدير، وهو خاص بكل فرد أو مستوى كل قدرة على حدة، بعكس النظرية الكلاسيكية التي يعتبر الخطأ المعياري فيها موحداً لكل الأفراد(Warm, 1978).

وعلى الرغم من شيوع استخدام نظرية الاستجابة للمفردة، وانتشار استخدامها في بناء وإعداد الاختبارات النفسية والتربوية، وتحليل البيانات المستمدة منها، إلا أن الطبيعة المعقدة والشروط النظرية الصارمة لهذه النظرية تقف عائقاً عند استخدامها في بناء الاختبارات، وتفسير نتائجها، وخاصة عند استخدامها من قبل غير المتخصصين في القياس، بالإضافة إلى عدة مشكلات تواجه المستخدمين لهذه النظرية(Han et al.,2019)، لذلك اتجه عُلماء القياس في الآونة الأخيرة إلى مواجهة هذه الصعوبات مواجهة إيجابية، حيث أجريت العديد من الدراسات لتحويل هذه النماذج من الجانب النظري إلى الجانب العملي التطبيقي، وتبسيطها لتمكين المهتمين بالتطبيق من استعمالها دون عناء.

بالإضافة إلى تلك الجهود فهناك جهود مستمرة نظرية وتطبيقية تهدف إلى المقارنة بين المفاهيم والإجراءات المختلفة بين نظرية الاختبار الكلاسيكية ونظرية الاستجابة للمفردة، وربطهما ببعضهما البعض لتحقيق البساطة في تصحيح الاختبارات، واستخلاص نتائجها وتفسيرها، وتدريج المفردات، ومعادلة الصور الاختبارية المختلفة منها، وغير ذلك من التطبيقات النفسية والتربوية. وقد تبلورت تلك الجهود بظهور طرق جديدة في القياس، تتصف بنفس صفات وخصائص نظرية الاستجابة للمفردة من حيث الدقة، وتحقيق الموضوعية في القياس، حيث اقترح ديمتروف(Domingue & Dimitrov,2015) طريقة دلتا لتقدير الدرجات Delta- scoring method (DSM) وتقوم هذه الطريقة على تقدير معالم المفردات، وقدرات الأفراد على اختبار ما، من خلال تقدير الصعوبة المتوقعة للمفردة، والتي يتم فيها إيجاد الدرجة المتوقعة للمفحوص. وتتُطور هذه الطريقة نظرية الاختبار الكلاسيكيةCTT وذلك من خلال تنفيذ ميزات تشبه نظرية

الاستجابة للمفردة IRT على سبيل المثال الدرجة *D* تعتمد على نمط استجابة المفحوص، كما يمكن من خلال هذه الطريقة تقدير معالم المفردة والمفحوص بشكل مبسط ومباشر، دون اللجوء إلى تكرار عمليات التقدير لتحسين التقدير كما في أسلوب نيوتن رافسون أو التوقع البعدي، وتأخذ في الاعتبار هذه الطريقة صعوبة المفردة المتوقعة، كذلك تُشكل القدرات مقياساً للفاصل الزمني، وفيها يتم تمثيل قدرات الأفراد، ومعالم المفردة على نفس المقياس كما في النظرية الحديثة(Dimitrov, 2016؛ Dimitrov & Atanasov, 2020)، وتتميز طريقة دلتا لتقدير الدرجات بما تميّزت به أساليب النظرية الكلاسيكية لتقدير وتحليل درجات الاختبارات من البساطة والوضوح، إلا أنه يضيف عدداً من المميزات التي لم تتوفر في تلك الأساليب (Dimtrov,2018؛ Domingue & Dimitrov,2021) منها: أن تقدير درجة المفحوص يستند على متجه استجابته على مفردات الاختبار ويأخذ في الاعتبار الصعوبة المتوقعة للمفردة الاختبارية بالنسبة لأفراد المجتمع المستهدف بتطبيق الاختبار، وإمكانية تدريج درجات الأفراد وصعوبة المفردات على مقياس واحد يعّبِر عن السمة المقاسة. كما أن استخدام تقديرات درجات الأفراد (D) أفضل من تقدير (θ) للحصول على مقياس فئوي، حيث أنها تُنّتِج انتهاكات أقل لمسلمات القياس الموحد Additive Conjoint Measurement.

في ضوء هذا الأسلوب قدّم ديتمروف عدداً من النماذج الرياضية التي تُستخدم لنمذجة العلاقة القائمة بين استجابة المفحوصين، على مفردات مقياس ما وبين السمة أو الدرجة الكامنة وراء هذه الاستجابة كما في نماذج نظرية الاستجابة للمفردة، فطريقة دلتا في إطارها الكامن (DSM-L) مماثلة لـ (IRT) في العديد من الجوانب (ولكن ليس كلها)، مثل استخدام النمذجة الكامنة للتقدير المتزامن لمستويات قدرة المفحوصين (السمات)، ومعالم المفردات (مثل الصعوبة والتمييز)، ودالة الاستجابة للمفردة وغيرها &Dimtrov (Atanasov 2020).

مشكلة الدراسة

أن دقة تقدير معالم الأفراد والمفردات المستخلصة في ضوء نماذج نظرية الاستجابة للمفردة تتأثر بعدد من العوامل. وقد استهدفت الكثير من الدراسات التي أجرها الباحثون في القياس فحص مدى تأثير هذه العوامل على دقة تقدير المعالم، كطريقة التقدير، وحجم عينة المعايرة، وطول أداة القياس، وشكل توزيع المعالم، والنموذج المستخدم والعديد من العوامل الأخرى(Al- Sharifain, & Bani Atta,2017, AL-2017؛ Balawi ,2018؛Nasraween ؛ Mousavi,2015 ؛ Yurekli,2010).وقد أثبت نتائج الدراسات التي قام بها ديمتروف دقة تقديرات قدرة المفحوص باستخدام نماذج دلتا وفعالية استخدام تلك التقديرات في التطبيقات المختلفة لنظرية الاستجابة للمفردة كمعادلة الاختبارات وغيرها من التطبيقات، وبالتالي يمكن تطبيق طريقة دلتا DSM مع الاختبارات التكيفية متعددة المراحل Multistage Testing والتي يمكن أن تحسن من كفاءة القياس؛ وفي ظل تلك الجهود المبذولة في الوصول إلى طرق أكثر دقة وبساطة في القياس، واستكمالاً لتلك الجهود لابد من إجراء المزيد من الدراسات التي يمكن أن تسهم في تطوير

الاختبارات التكيفية من الناحية العملية، تأتي الدراسة الحالية في محاولة منها لدراسة فاعلية القياس التكيفي متعدد باستخدام طريقة دلتا في تقدير معلم الدراجات.

وتتلخص مشكلة الدراسة في محاولة الإجابة عن السؤال الآتي:

ما فاعلية القياس التكيفي متعدد المراحل باستخدام طريقة دلتا لتقدير معلم الدرجة؟ مقاسة بما يأتي:

ويمكن التعبير عن السؤال الرئيس في هذه الدراسة بالسؤالين الفرعيين الآتيين:

1. ما العلاقة بين تقديرات الدرجة المرجعية المقدرة من خلال الاختبار الخطي المكون (80) مفردة والدرجة المقدرة في كل مرحلة من المراحل الأربع المبنية؟

2. ما هو متوسط القيم المطلقة للفروق بين تقديرات الدرجة المرجعية المقدرة من خلال الاختبار الخطي المكون من (80) مفردة والدرجة المقدرة في كل مرحلة من المراحل الأربع المبنية؟

3. ما مقدار انخفاض الخطأ المعياري في التقدير أثناء التقدم في المراحل؟

4. ما مقدار انخفاض الجذر التربيعي لمتوسط مربعات الخطأ المعياري في التقدير أثناء التقدم في المراحل؟

أهمية الدراسة

تتمثل الأهمية النظرية للدراسة الحالية في أنها تسلط الضوء على طريقة دلتا لتقدير الدرجات، كطريقة معاصرة لتصحيح أدوات القياس وتفسير نتائجها ومعايرة مفرداتها في القياس النفسي والتربوي، ومدى فاعلية استخدامها في الاختبارات التكيفية.

أما الأهمية التطبيقية للدراسة الحالية فتتمثل في أنها توفر معلومات دقيقة يمكن أن يستفيد منها واضعي ومبرمجي الاختبارات التكيفية من خلال استخدامها في الحصول على قياس دقيق لقدرات الأفراد بأقل خطأ ممكن، وعلى وجه الخصوص في مجال الاختبارات التكيفية.

أهداف الدراسة

هدفت الدراسة الحالية إلى اختبار فاعلية القياس التكيفي متعدد المراحل باستخدام طريقة دلتا وذلك بالاعتماد على محكي معامل الارتباط، ومتوسط الفروق المطلقة بين الدرجة الحقيقة والدرجة المرجعية المأخوذة من الاختبار الخطي باستخدام أربع مراحل للاختبار التكيفي.

مصطلحات الدراسة

الاختبار التكيفي:The Adaptive Testing

يُعرفه هامبيلتون وسواميناثان (Hambleton & Swaminathan,1985) بأنه: اختبار يوائم بين قدرة المفحوص وصعوبة المفردات المقدمة له حيث لا يتقدم الأفراد لنفس المفردات، بل يجري تعريض كل فرد لمفردات تناسب قدرته، وهو التعريف النظري المعتمد في الدراسة الحالية.

ويُعرفه Al-Bursan (2012) بأنه اختبار يتكون من مفردات متتابعة أو مجموعات متتابعة من المفردات بحيث تتحدد المفردة أو مجموعة المفردات التالية اعتماداً على نتيجة المفردة السابقة أو نتائج مجموعة من المفردات السابقة، وذلك وصولاً إلى المستوى التقاربي (Asymptotic Level) للقدرة المقدرة.

ويعرَّف إجرائياً في الدراسة الحالية بأنه اختبار يحتوي على مفردات متتابعة أو مجموعات متتابعة من المفردات بحيث تتحدد المفردة أو مجموعة المفردات التالية اعتماداً على نتائج مجموعة من المفردات السابقة، وذلك وصولاً إلى المستوى التقاربي للقدرة المقدرة من اختبار أوتيس– لينون المستوى المتقدم الصورة (J) المطبق على عينة من طلبة جامعة إب.

طريقة دلتا Delta- Scoring Method: هي نموذج من النماذج الاحتمالية وضعها ديمتروف (Dimitrov, 2015)ويعرفها بأنها: إحدى طرق تقدير معالم المفردات وقدرات الأفراد من خلال إيجاد قيمة D على اختبار ما من خلال تقدير صعوبة المفردات المتوقع δ_j، وتعتمد في حسابها على متجه استجابة المفحوص. وتفسر درجة D على أنها نسبة من الدرجة المطلوبة للنجاح التام في الاختبار التي أظهرها المفحوص، والتي يمكن حسابها بشكل مباشر بالنظر إلى متجه استجابة المفحوص وتقديرات الصعوبة لمفردات الاختبار، وهذا التعريف النظري المعتمد في الدراسة الحالية.

وتعرَّف إجرائياً في الدراسة الحالية بأنها نموذج الدالة النسبية ثنائية المعلم بحيث يتم تقدير معالم الطريقة وفق الطريقة الكامنة لطريقة دلتا لتقدير الدرجة باستخدام اختبار أوتيس– لينون التكيفي الذي تم تطبيقه على عينة من طلبة جامعة إب.

معلم القدرة Ability Parameter: هو درجة المفحوص وتعبر عن مقدار ما يمتلكه من السمة (الدرجة) المقاسة بأداة القياس، ويعبر عن موقع المفحوص على متصل السمة المقاسة (Hambleton & Swaminathan,1985).

ويعرف إجرائياً بأنه مقدار سمة الدرجة العقلية للفرد (D -Score) (θ) والمقدرة باستخدام طريقتي دلتا والنموذج اللوغاريتمي ثنائي المعلم وذلك بعد تطبيقه على طلبة جامعة إب.

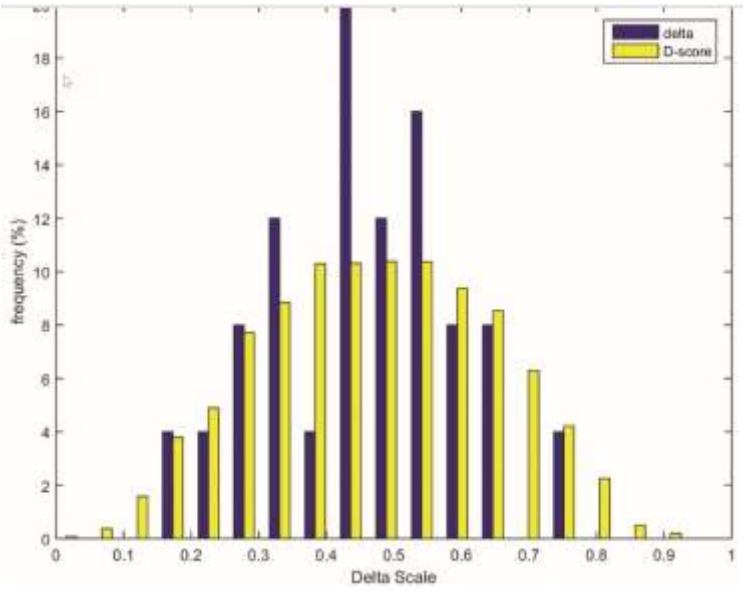
الخلفية النظرية

التدريج وفق طريقة دلتاD-Scale

طوّر ديمتروف تدريجاً فئوياً Interval scale يُعّبِر عن السمات المقاسة تتراوح درجاته من (0) إلى (1) بحيث يمكن استخدامه لتعيين درجات (قدرات) الأفراد (D_s) والصعوبات المتوقعة لمفردات الاختبار (δi) المقدرة وفق طريقة دلتا لتقدير الدرجات (DSM) وتحديد مواقعهم على متصل واحد للسمة، عُرِف بتدريج دلتا(D-Scale)، ويمكن تمثيل درجات الأفراد (D_s) والصعوبات المتوقعة للمفردات (δi) باستخدام التوزيعات التكرارية لكل منهما للحصول على خريطة للمفردة والمفحوص(Item person map (IPM، مما يساعد في مقارنة درجات الأفراد بخصائص المفردات(Dimtrov,2018). ويمثل الشكل (1) خريطة المفردة والمفحوص

لأحد اختبارات القدرات العامة والمكَّون من (25) مفردة والمطبق على (7,782) طالباً من طلبة المرحلة الثانوية.

شكل 1خريطة المفردة والمفحوص على تدريج دلتا (IPM)



(Dimtrov,2018: p.809)

وقد برهنت نتائج دراسة دومينقو وديتمروف(Domingue & Dimitrov,2015) أن استخدام قيم درجة الأفراد المقدّرة وفق طريقة دلتا لتقدير الدراجات(D_s) أفضل من نظيرتها (θ) المقدَّرة في ضوء قيم نظيرتها المقدَّرة في ضوء النموذج اللوغاريتمي ثلاثي المعلم(3PLM) للحصول على مقياس فئوي.

نماذج دالة الاستجابة للمفردة وفق طريقة دلتا

وضع ديتمتروف عدداً من النماذج الاحتمالية بحيث يتم من خلالها نمذجة العلاقة بين أداء المفحوص على مفردات الاختبار وبين السمة الكامنة خلف هذا الأداء. ويمكن التعبير عن هذه العلاقة بدالة تزايدية مطردة دالة الاستجابة للمفردة Item response function (IRF) وتُمثَّل هذه الدالة بيانياً بالمنحى المميز للمفردة (ICC) كما في نظرية الاستجابة للمفردة وتمتاز هذه النماذج بإمكانية استخدامها مع المفردات الثنائية أو متعددة التدريج.

1.نموذج الانحدار اللوغاريتمي(LRM) Logistic Regression Model

يتم في ضوء هذا النموذج تقدير احتمالية استجابة المفحوص استجابة صحيحة للمفردة وفقاً لدرجته D Score على تدريج دلتاD- Scale على أنها الدرجة المتنبأ بها للمفردة (Predicted Item Score $\widehat{xsi}$) باعتبار درجة المفحوص المقدرة(Ds) المتغير المستقل المنبئ Predictor، أي أن ($\widehat{xsi}$) هي الدرجة الحقيقة للفرد ذو الدرجة (الدرجة) (Ds) على المفردة i (Dimitrov,2018:p810).

ويُعبَّر عن الدالة الرياضية الممثلة لنموذج الانحدار اللوغاريتمي (LRM) بالشكل الآتي:

$$\widehat{xsi} = P\ (x_{si} = 1/\ D_s)\ 1 - \frac{1}{1+\left[\frac{Ds}{bs}\right]^{ai}} \qquad (1)$$

حين أن Ds درجة المفحوص، و $\widehat{xsi}$الدرجة الملاحظة للفردS على المفردة i، و b_i موقع المفردة i على متصل الدرجة (صعوبة المفردة)، و a_i ميل المنحنى المميز للمفردة i (تمييز المفردة).

وباستخدام هذا النموذج يمكن الحصول على تقديرات جيدة للدرجات الحقيقية للأفراد وعلى المفردة الاختبارية xsi في حال وقوع قدرات الافراد في المدى المتوسط أو الطرق الأدنى من تدريج دلتا للسمة المقاسة، بينما يقل جودة تقدير تلك الدرجات في حال كانت قدرات الأفراد في الطرف الأعلى من التدريج وخاصة المفردات الصعبة نسبياً، وهذا يقلل من فعالية استخدام هذا النموذج اللوغاريتمي ثنائي المعلم(2PLM) في نظرية الاستجابة للمفردة.

2.نماذج الدالة النسبية Rational Function Models (RFMs) لطريقة دلتا

لحل مشكلة عدم دقة تقديرات الدرجة الحقيقة $\widehat{xsi}$ في الطرف الأعلى من تدريج دلتا وضع ديتمتروف (Dimitrov,2020a)، ثلاثة نماذج للدالة النسبية للتعبير عن دالة الاستجابة للمفردة على تدريج دلتا، بحيث تختلف هذه النماذج في عدد المعالم لكل نموذج منها، وتتشابه هذه النماذج مع النماذج الثلاثة اللوغاريتمية لنظرية الاستجابة للمفردة (1PLM,2PLM, 3PLM) من حيث معالم النموذج مثلاً سُمّى نموذجا (2PLM)، و(RFM2) ثنائيي المعلم لاعتمادهما على تقديرات معلمي صعوبة وتمييز المفردة الاختبارية– إلا أن نماذج الدالة النسبية الثالثة تُضيف معلم وسيطاً وهو معلم مطابقة شكل المنحى(S) fit parameter for shape والذي يختلف عن معلم التمييز للمفردة. ويوضح الجدول التالي نماذج الدالة النسبية الثلاثة.

ويمكن استعراض هذه النماذج كما أوردها يدمتروف(Dimitrov,2020a):

1.نموذج الدالة النسبية أحادي المعلم (RFM1)

$$P = \frac{1}{1+\left[\frac{b(1-D)}{D(1-b)}\right]^{s}} \qquad (2)$$

علماً أن(b) معلم صعوبة المفردة، و(a) معلم تمييز المفردة، و(c) التخمين الزائف للمفردة، وs معلم مطابقة الشكل، وهي في هذا النموذج متساوية لجميع المفردات.

2.نموذج الدالة النسبية ثنائي المعلم (RFM2)

$$P = \frac{1}{1+\left[\frac{b(1-D)}{D(1-b)}\right]^{s}} \qquad (3)$$

علماً أن(b) معلم صعوبة المفردة، و(a) معلم تمييز المفردة، و(c) التخمين الزائف للمفردة، وs معلم مطابقة الشكل، حيث أن: $a = \frac{s}{4b(1-b)}$

3.نموذج الدالة النسبية ثلاثي المعلم(RFM3)

$$P = c + \frac{1-c}{1+\left[\frac{b(1-D)}{D(1-b)}\right]^{s}} \qquad (4)$$

علماً أن(b) معلم صعوبة المفردة، و(a) معلم تمييز المفردة، و(c) التخمين الزائف للمفردة، وs معلم مطابقة الشكل، حيث أن: $a = \frac{s(1-c)}{4b(1-b)}$

وتتشابه هذه النماذج مع النماذج اللوغاريتمية، في أنها تتوازي المنحنيات المميزة للمفردة في نموذجي (Basic RFM1) و(RFM1) كما في نموذجي (1PLM) و(Rasch). كذلك يُمكن أن تتقاطع المنحنيات المميزة للمفردات في نموذج (RFM2)، كما في نموذج(2PLM). كما يتضمن (RFM3) معلم التخمين الزائف (c)، كما في نموذج(3PLM).

وتختلف هذه النماذج عن النماذج اللوغاريتمية في عدة أمور منها: أنه يقدَّر معلم قدرة المفحوص (درجة المفحوص D) كمصدر أولي ثم يتم استخدامه كمتغير مستقل في أحد نماذج (RFMs)، لتقدير معالم المفردة في الإطار الكلاسيكي وكقيمة أولية في الإطار الكامن، بينما في النماذج اللوغاريتمية معالم المفحوص والمفردة غير معروفة مسبقاً، بل يتم تقديرها في وقت واحد باستخدام الأساليب القائمة على دالة الأرجحية أو النظرية البيزية. كما يُضمَّن معلم التمييز في الدالة الرياضية للنماذج اللوغاريتمية بشكل مباشر، على عكس نماذج (RFMs)، والتي يحسب فيها معلم التمييز كدالة لمعلمي موقع المفردة على متصل السمة (b) ومطابقة شكل المنحنى المميز لها (s). وفي هذه النماذج تتساوى قيمة معلم التمييز لجميع المفردات الاختبارية وفق نموذجي(1PLM) و(Rasch)، بينما تتساوي قيمة معلم التمييز للفقرتين ذات المواقع المتناظرة حول متوسط تدريج دلتا (0.5) في نموذجي Basic RFM1)، و(RFM1) فمثلاً إذا كانت قيم (b) لمفردتين (0.2، 0.8) فإن قيمة معلم التمييز لهاتين المفردتين متساوية.

أساليب تقدير معالم نماذج دالة الاستجابة للمفردة وفق طريقة دلتا لتقدير الدرجات

وضع ديمتروف إطارين مختلفين في ضوئهما يتم تقدير معالم نماذج دالة الاستجابة للمفردة باستخدام طريقة دلتا لتقدير الدرجات (DSM)، عُرف أحدهما بالإطار الكلاسيكي والآخر بالإطار الكامن وفيما يلي تفصيل لكل منها (Dimitrov,2020b).

أولاً: الإطار الكلاسيكي(DSM-C) DSM-Classical Framework

تعتمد درجة المفحوص المقدرة (D-score) والتي تُعبّر عن قدرته– في الإطار الكلاسيكي على متجه استجابته Person's response vector موزوناً بالصعوبة المتوقعة لمفردات الاختبار بالنسبة لأفراد المجتمع المستهدف بتطبيق الاختبار، ويُعبَّر عنها بالمعادلة التالية:

$$Ds = \frac{\sum_{i=1}^{n} Xsi\delta i}{\sum_{i=1}^{n} \delta i} \qquad (5)$$

حيث أن Ds قدرة المفحوص s، و Xs الدرجة الملاحظة للفرد s على المفردة i، و δi الصعوبة المتوقعة للمفردة i، و n عدد مفردات الاختبار.

واختصاراً يمكن أن ويُعبَّر عنها المعادلة التالية (Dimitrov & Atanaso,2021):

$$\boldsymbol{D_w} = \sum_{i=1}^{n} w_i X_{si} \qquad (6)$$

حيث $\boldsymbol{D_w}$ الدرجة الموزونة للفرد s (قدرة المفحوص)، و X_{si} الدرجة الملاحظة للفرد s على المفردة i، و w_i الوزن للصعوبة المتوقعة لمفردات الاختبار والمحسوب وفق المعدلة (Dimitrov & Atanaso,2020):

$$w_i = \frac{\delta_i}{\sum \delta_i} \qquad (7)$$

وتتراوح قدر المفحوص Ds بين $(0 \leq \boldsymbol{Ds} \leq 1)$، بحيث تكون $\boldsymbol{Ds} = 0$ عندما يجيب المفحوص إجابات خاطئة على جميع مفردات الاختبار أي عندما $(X_{si} = 0 \ , X_{s2} = 0 \ , \ldots\ldots X_{sn} = 0)$، بينما تكون $\boldsymbol{Ds} = 1$ عندما يجيب المفحوص بشكل صحيح عن جميع المفردات أي عندما $(X_{si} = 1 \ , X_{s2} = 1 \ , \ldots\ldots X_{sn} = 1)$. وتُفسَّر قدرة المفحوص $\boldsymbol{Ds}$ بأنها نسبة الدرجة المطلوبة للنجاح الكلي على الاختبار التي أظهرها ذلك المفحوص (Dimitrov & Atanaso,2021؛ Domingue & Dimitrov, 2015؛ Dimitrov,2018).

طرق حساب معامل صعوبة المفردات باستخدام طريقة دلتا

يتم تقدير الصعوبة المتوقعة للمفردة Expected item Difficulty والتي تُعرف بدلتا (Delta) ويرمز لها بالرمز δ_i بثلاث طرق هي:

الأولى: تعتمد على معالم المفردة المقدرة باستخدام أحد أساليب المستخدمة في نظرية الاستجابة للمفردة (IRT).

الثانية: تعتمد على استخدام أسلوب البوتستراب Bootstrap.

الثالثة: من خلال حساب قيمة مستوى الدلالة P-value ويتم الاعتماد على هذه الطريقة عندما يكون حجم العينة كبير جداً وممثل للمجتمع الأصلي.

الطريقة الأولى: باستخدام نظرية الاستجابة للمفردة (IRT).

يتم تقدير الصعوبة المتوقعة للمفردة بشكل مباشر من خلال بالاعتماد على الدرجة المتوقعة Expected Item Score التي يرمز لها بالرمز وفق المعادلة التالية (Dimitrov,2016a؛ Domingue & Dimitrov,2015).

$$\delta i = 1 - \pi i \qquad (8)$$

حيث أن δi الصعوبة المتوقعة للمفردة i، و πi الدرجة المتوقعة للمفردة i.

وتُعبِّر πi عن نسبة الأفراد المختبرين الذين يتوقع أن يجيبوا إجابة صحيحة عن المفردة i ووفقاً للنظرية الكلاسيكية (CTT) فإن πi يسمى صعوبة المفردة (بالرغم من أنه في الواقع يشير إلى سهولة المفردة Item (Easiness) Domingue & Dimitrov,2015)). ويتم إيجاد الدرجة المتوقعة للمفردة (πi) كدالة لمعالم المفردة(الصعوبة(b)، والتمييز(a)، والتخمين(c)) ووفقاً لنماذج نظرية الاستجابة للمفردة (IRT) بالمعادلة التالية (Domingue & Dimitrov,2015؛Dimitrov,2016).

$$\pi i = \frac{1 - \mathrm{erf}(xi)}{2} \qquad (1)$$

حيث أن erfهي دالة الخطأ Error Function، و xi وفق النموذج اللوغاريتمي ثنائي المعلم (2PLM) تحسب من المعادلة التالية:

$$\mathrm{X}i = \mathrm{a}i\mathrm{b}i\sqrt{2(1 + \mathrm{a}i2)} \qquad (10)$$

وعند استخدام النموذج اللوغاريتمي أحادي المعلم (1PLM) يتم استبدال معلم التمييز في المعادلة بـ($ai = 1$).

وفي حال استخدام النموذج اللوغاريتمي ثلاثي المعلم (3PLM) تحسب الدرجة المتوقعة للمفردة πi كما في النموذج (2PLM) ويضاف إليها عنصر التخمين وفق المعادلة التالية:

$$ci + (1\text{-}ci)\,\pi i \qquad (11)$$

الطريقة الثانية: باستخدام أسلوب البوتستراب Bootstrap.

يتم تقدير الصعوبة المتوقعة للمفردة (δi) بشكل مباشر بالاعتماد على أسلوب البوتستراب Bootstrap، وذلك بأخذ عدد كبيرة من العينات العشوائية (1000 عينة مثلاً) من استجابات الأفراد المختبرين وإيجاد نسبة الاستجابة الصحيحة(Proportion of Correct) للأفراد المختبرين على المفردة في كل عينة عشوائية(Dimitrov,2018)، حيث يُحدد مدى صعوبة المفردة في حالة المفردات ثنائية التدريج وفقاً للنظرية الكلاسيكية بنسبة الأفراد الذين يجيبون إجابة صحيحة عن المفردة التي تمثل معلم الصعوبة(Crocker & Algina, 1986,p90-301). أي أن هذه النسبة تعّبِر عن الدرجة المتوقعة للمفردة (πi) ومن ثم يتم حساب الصعوبة المتوقعة للمفردة من المعادلة التالية:

$$\delta i = 1 - \pi i \qquad (12)$$

وبحساب المتوسط لقيم δi للعينات العشوائية يتم تحديد قيمة الصعوبة المتوقعة لكل مفردة من مفردات الاختبار بالنسبة لأفراد المجتمع المستهدف بتطبيق الاختبار.

ومن أهم مميزات هذه الطريقة للتقدير أنها لا تعتمد على تقديرات مسبقة لمعالم المفردة وفق أي نموذج من نماذج نظرية الاستجابة للمفردة(IRT)، مما يجعلها لا تتطلَّب التحقق من افتراضات نظرية(IRT) – مثل مطابقة البيانات للنموذج، واعتدالية توزيع قدرات الأفراد، وغيرها– عند تقدير الصعوبة المتوقعة للمفردة (Dimitrov,2018,p807). ويتضمن الجدول التالي مثالاً تطبيقياً يُوضح كيفية حساب درجة المفحوص (D-Score) بعد حساب الصعوبة المتوقعة للمفردة بأحدي الطريقتين (Dimitrov,2020a).

جدول 1 حساب درجات أربعة مفحوصين وفقاً لمتجه استجاباتهم على خمس مفردات

D score	$\sum_{i=1}^{n} \delta i$	$\sum_{i=1}^{n} Xsi\delta i$	$\delta 5 = 0.80$	$\delta 4 = 0.65$	$\delta 3 = 0.50$	$\delta 2 = 0.35$	$\delta 1 = 0.20$	**Person**
			X_{s5}	X_{s4}	X_{s3}	X_{s2}	X_{s1}	
0	2.50	0	0	0	0	0	0	1
0.48	2.50	1.20	0	1	0	1	1	2
0.60	2.50	1.50	1	0	1	0	1	3
1	2.50	2.50	1	1	1	1	1	4

يتضح من الجدول السابق أن بالرغم من حصول المفحوصان الثاني والثالث على درجة كلية متساوية(X=3) فإن تقدير قدرتيهما اختلف؛ حيث كانت قيم التقدير للقدرة Ds_2 =0.48، Ds_3 =0.60 وذلك لاختلاف متجه استجابتهما على مفردات الاختبار، كما نلاحظ أن المفحوص الأول كان تقدير قدرته0= Ds_1، وذلك لأنه أجاب إجابات خاطئة على جميع المفردات، بينما كان تقدير قدرة المفحوص الرابع 1= Ds_4، لأنه أجاب إجابات صحيحة على جميع المفردات الخمس. وكمثال لتفسير درجات Ds تُشير Ds4 =1 أن المفحوص الثاني أظهر 100% من الدرجة المطلوبة للنجاح الكلي على الاختبار.

القيمة الحقيقة(المتوقعة) والخطأ المعياري في التقدير وفق الإطار الكلاسيكي (DCM-C)

تتحدَّد القيمة الحقيقة (المتوقعة) True (Expected) Value Of D_w Score لدرجة المفحوص المقدرة وفق طريقة دلتا لتقدير الدرجات في الإطار الكلاسيكي بالمعادلة للمفردة (Dimitrov,2018,p810-811):

$$E(D_W) = \left(\frac{1}{\sum_{i=1}^{n} \delta_i}\right) \sum_{i=1}^{n} \delta_i \, P_i(D_w) \qquad (13)$$

ويتم حساب الخطأ المعياري للتقدير وفق المعادلة التالية (Dimitrov,2018,p811):

$$SE(D_w) = \left(\frac{1}{\sum_{i=1}^{n} \delta_i}\right) \sqrt{\sum_{i=1}^{n} \delta_i^2 P_i(D_w)[1 - P_i(D_w)]} \qquad (14)$$

ثانياً: الإطار الكامن DSM- Latent Framework (DSM-L)

في الإطار الكامن لطريقة دلتا لتقدير الدرجات يتم التعامل مع نماذج الدالة النسبية (RFMs) كنماذج كامنة كما في نظرية الاستجابة للمفردة، حيث يتم تقدير معالم المفردة والمفحوص– غير المعروفة سابقاً باستخدام أحد أسلوبي الأرجحية العظمي(المشتركة)(JLM) أو الهامشية (MML). وبالرغم من استخدام أساليب تعتمد على تكرار عملية التقدير كما في نظرية الاستجابة للمفردة، إلا أنه يمكن تقليل عدد دورات التقدير في الإطار الكامن لطريقة دلتا لتقدير الدرجات باستخدام درجة المفحوص(D_w) المقدرة في ضوء الإطار الكلاسيكي كقيمة مبدئية لتقدير معلم قدرة المفحوص في الإطار الكامن (Dimitrov & Dimitrov,2020a ؛Dimitrov,2020b؛ Atanaso,2021).

تعتبر نظرية الاستجابة للمفردة IRT أكثر تعقيداً من الناحية النظرية على سبيل المثال خاصية "الثبات"، وبالرغم من عدم تمتُّع الإطار الكامن (DSM-L) بخاصية اللاتغاير Invariance كما في نظرية بعدد من المميزات العملية تتمثل في أن التقديرات للمعالم في ضوئه تتطلَّب عمليات حسابية أقل تعقيداً من العمليات التي تتطلبها التقديرات المستخرجة بنماذج نظرية الاستجابة للمفردة، وهذا يساعد في تجنب المشاكل التي قد تحدث عند التقدير باستخدام أسلوب الأرجحية العظمى كمشكلة عدم التقارب التي قد تحدث في إطار(IRT)؛ كما أنها توفر مستويات أعلى من التمييز بين المفحوص ين في نهاية مقياس (D-Scale) (أي المفحوص ين ذوي الدرجة المنخفضة والمفحوص ين ذوي القدرات العالية)؛ إضافة إلى سهولة تفسير الدرجات على تدريج الدلتا(من0 إلى1) (Dimitrov,2020b).

دالة المعلومات الاختبار والخطأ المعياري في التقدير وفق الإطار الكامن (DSM-L)

تتحدد دالة المعلومات للمفردة وفق الإطار الكامن بالمعادلة التالية:

(Dimitrov & Atanaso,2021):

$$l_i(D) = \frac{s_i^2 P_i(1-P_i)}{[D(1-D)]} \quad (15)$$

أما دالة المعلومات الاختبار فتحسب وفق المعادلة التالية:

$$l_i(D) = \sum_{i=1}^{n} l_i(D) \quad (16)$$

ويتم حساب الخطأ المعياري في التقدير بالمعادلة التالية:

$$SE(\widehat{D}) = \frac{1}{\sqrt{1-(D)}} \quad (17)$$

وتُشير معادلتا كل من دالة المعلومات الاختبار والخطأ المعياري للتقدير إلى مدى التشابه بين الإطار الكامن(DSM-L) ونظرية الاستجابة للمفردة، حيث تُعرف دالة معلومات الاختبار كمجموعة مباشر لدوال معلومات مفردات الاختبار، ويُعَّرف الخطأ المعياري للتقدير بمقلوب الجذر التربيعي لدالة معلومات الاختبار، كذلك الخطأ المعياري في التقدير لدرجة D الكامنة، يرتبط عكسيًا بدالة معلومات الاختبار كما في نظرية الاستجابة للمفردة (Dimitrov at al.,2021).

الدراسات السابقة

أجريت العديد من الدراسات في مجال الاختبار التكيفية حيث أجرت Al-*Shdifat*(2008) دراسة هدفت إلى بناء اختبار تكيف لقياس القدرة الرياضية وفق الاستراتيجية ثنائية المرحلة في نظرية الاستجابة للمفردة، وتكونت أداة الدراسة في صورتها النهائية من (60) مفردة وتكونت عينة القياس التكيفي (100) طالب وطالبة، وقد تكون تصميم الاختبار التكيفي المحوسب متعدد المراحل من مرحلتين اختباريتين، وقد أوضحت نتائج الدراسة أن للاختبار التكيفي المحوسب متعدد المراحل دقة في تقدير قدرة الأفراد وذلك مقارنة مع الاختبار الخطي.

كما أجرت Al-Bayada(2011) دراسة كان الهدف منها بناء تصميم تكيفي محوسب متعدد المراحل (خمس مراحل) تضمنت المرحلة الاختبارية الأولى صورة اختبارية مصغرة متوسطة الصعوبة، أما المرحلة الاختبارية الثانية فتضمنت (2) صورة اختبارية مصغرة(سهلة – صعبة) والمرحلة الاختبارية الثالثة تتضمن (3) صور اختبارية مصغرة، والمرحلة الاختبارية الرابعة تتضمن(4) صور اختبارية مصغرة، أما المرحلة الاختبارية الخامسة فتتضمن(5) صور اختبارية مصغرة، وقد تكونت عينة التدريج من (1200) طالباً وطالبة، وقد أوضحت أهم نتائج الدراسة إلى فاعلية الاختبار التكيفي المحوسب متعدد المراحل في تقدير القدرة الرياضية بدقة وأقل وقت وجهد.

وقدم Al-Bursan(2012) بحثا هدف فيه إلى الكشف عن فاعلية القياس التكيفي المبني من مفردات ذات إجابة منتقاة ثنائية التدريج ومفردات ذات إجابة منشأة متعددة التدريج في نفس الاختبار باستخدام بيانات مولدة من الحاسوب، واختيرت عينة للقياس التكيفي بلغت (100) مفحوص جرى استخدام بيانات استجاباتهم في القياس التكيفي الذي تكون من ثلاث مراحل تكونت المرحلة الاستطلاعية الأولى من خمس مفردات ذات إجابة منتقاة ثنائية التدريج، والمرحلة الثانية من خمس مفردات ذات إجابة منتقاة ثنائية التدريج، في حين تكونت المرحلة الثالثة والأخيرة من فقرتين ذاتي إجابة منشأة، وأظهرت النتائج معامل ارتباط بين قدرة المقدرة من الاختبار التكيفي والقدرة المقدرة من الاختبار الخطي مقداره 0.93، ومتوسط فرق في القدرة مقدار 0.087 لوجيت، إضافة إلى فرق الخطأ المعياري مقداره في المتوسط 0.278 الأمر الذي يفيد بفاعلية عالية للقياس التكيفي المبني باستخدام مفردات ثنائية ومتعددة التدريج معاً.

وقدم ديلز (2014) Dallas دراسة هدفت إلى معرفة أثر محك البداية وطريقة حساب الدرجة على تصميم الاختبار التكيفي المحوسب متعدد المراحل، حيث تم إجراء الدراسة على بيانات محاكاة على أربعة مستويات تضمن المستوى الأول توليد البيانات وتدريجها، تم المستوى الثاني ويتمثل في تجميع مفردات الصور الاختبارية الدينامية وكانت(10) صور اختبارية دينامية، أما المستوى الثالث والذي قد تمثل في توليد البيانات بمعالم

المفردات المشتقة من المرحلة الأولى تم المستوى الرابع فتمثل في توليد بيانات الاختبار التكيفي المحوسب متعدد المراحل وذلك لعينة تتكون من (200000) طالب، وتمثلت أهم نتائج الدراسة في أن الطرق القائمة على نظرية الاستجابة للمفردة تكون أفضل من طريقة الإجابات الصحيحة في بدء الاختبار، بالإضافة إلى أن زيادة عدد المراحل الاختبارية يعمل أيضاً على زيادة دقة القياس.

وهدفت دراسة ساري ومانلي (Sari & Manley,2017) إلي المقارنة بين الاختبار التكيفي المحوسب لكل مفردة والاختبار التكيفي المحوسب متعدد المراحل، وذلك في حالة تغيير عدد من مجالات المحتوى المختلفة وتم أيضاً تغيير أطوال الاختبار؛ لذلك قام الباحثان بالمقارنة بين تصميم الاختبار التكيفي المحوسب لكل مفردة وتصميمين للاختبار التكيفي المحوسب متعدد المراحل، وهما 1–3 وتصميم 1–3–3 وذلك بأطوال اختبار مختلفة هي (24) مفردة، و(48) مفردة، وقد تم التحكم في قيود مجالات المحتوى إلى (8) مجالات للمحتوى، وقد أوضحت أهم نتائج الدراسة أن طول الاختبار ونوع التصميم لها تأثير كبير على نتائج وخصائص الاختبار بدرجة أكبر من تأثير مجالات المحتوى وعددها، كما وجد أن التصميم التكيفي المحوسب لكل مفردة يعطي نتائج أفضل قليلاً من تصميمات الاختبار التكيفي متعدد المراحل وذلك من ناحية متوسط التحيز في تقدير القدرة.

وقام هان وآخرون(Han et al., 2019) دراسة لتطوير ومقارنة تصميمات الاختبار التكيفي متعدد المراحل باستخدام طريقة دلتا(DSM) حيث أجريت الدراسة على بيانات محاكاة، وقد استخدمت تصميم الاختبار التكيفي المحوسب لكل مفردة أجريت الدراسة على جزأين الأول مع تصميم 1–3 بواقع (15) مفردة لكل مرحلة فقد بلغ العدد الكلي للاختبار(30) مفردة، والجزء الثاني مع تصميم 1 – 2–3 بواقع (15) مفردة لكل مرحلة فقد بلغ العدد الكلي للاختبار(60) مفردة. وقد أشارت نتائج الدراسة إلى فاعلية الاختبار التكيفي متعدد المراحل باستخدام طريقة دلتا، حيث يمكن مقارنة الدرجات بين المسارات المختلفة للاختبار التكيفي. كما أشارت النتائج أن الاختبار التكيفي متعدد المراحل (MST) باستخدام طريقة دلتا يمكن أن تحقق تحسينات في دقة القياس وكفاءته مقارنة بالاختبار الخطي المبنى على طريقة دلتا.

ودراسة أوزتورك (Öztürk , 2019) التي هدفت الدراسة إلى معرفة أثر كل من طول الصورة الاختبارية المصغرة المبدئية وخصائصها في تصميم الاختبار التكيفي المحوسب متعدد المراحل على دقة تقدير القدرة، وقد تمثلت تصميمات الاختبار التكيفي المحوسب متعدد المراحل في تصميمين التصميم الأول ثنائي المرحلة (1–3)، والتصميم الثاني ثلاثي المرحلة(1–3–3) وقد تراوح عدد مفردات الصورة الاختبارية المصغرة المبدئية (الأولى) ما بين(5) مفردات إلى(30) مفردة، وتراوح العدد الكلي لمفردات الاختبار ما بين (25) مفردة إلى (50) مفردة، وذلك لكل تصميم على حدة، وأوضحت أهم نتائج الدراسة أنه كلما زاد طول الصورة الاختبارية المصغرة المبدئية كلما كان أفضل لتحقيق دقة في القياس، كما أن التصميم التكيفي المحوسب ثلاث المرحلة(1–3–3) يحقق دقة أفضل في القياس من التصميم ثنائي المرحلة (1–3).

منهجية الدراسة وإجراءاتها

منهج الدراسة: اعتمدت الدراسة الحالية على تطبيق المنهج الوصفي المقارن، وهو ما يحقق أهداف الدراسة الحالية في ضوء طبيعة مشكلة الدراسة متغيراتها.

مجتمع وعينة الدراسة: اعتمدت الدراسة الحالية على تطبيق المنهج الوصفي المقارن شمل المجتمع الإحصائي للدراسة الحالية جميع طلاب وطالبات جامعة إب ومن كل المستويات للعام الجامعي (2019 – 2020م) والبالغ عددهم (14519) طالب وطالبة، وقد تم اختيار عينة بالطريقة العشوائية الطبقية؛ وقد بلغت التدريج النهائية (1600) طالباً وطالبة.

أداة الدراسة:

استخدمت الدراسة الحالية اختبار أوتيس– لينون للقدرة العقلية المستوى المتقدم الصورة(J) وهذا الاختبار عبارة عن سلسلة من اختبارات أوتيس ولينون واسعة الاستخدام، ويتكون الاختبار من (80) مفردة متنوعة لقياس مختلف مظاهر الدرجة العقلية. وقد تم التحقق من الخصائص السيكومترية للاختبار. من خلال حساب معامل الاتساق الداخلي للاختبار أوتيس– لينون للقدرة العقلية، حيث تم حساب معاملات ارتباط المفردة مع البعد الذي تنتمي إليه، وقيم معاملات ارتباط المفردة مع الدرجة الكلية للاختبار، حيث بلغت قيم معاملات ارتباط المفردة مع البُعد الذي تنتمي إليه جاءت مرتفعة، وتراوحت بين (0.42–0.83)، كما تراوحت قيم معاملات الارتباط بين المفردات والاختبار ككل بين (0.36– 0.76). وتجدر الإشارة إلى أن الباحثان اعتمدا معياراً لقبول، أو حذف المفردة بأن لا يقل معامل ارتباطها بالمجال الذي تنتمي إليه، وبالاختبار ككل عن (0.25)، وبناءً على ذلك فقد تم قبول جميع المفردات باعتبار أن جميع المفردات كانت دالة إحصائياً عند مستوى دلالة (0.01)، و(0.05). كما تم استخراج قيم معاملات الارتباط البينية لأبعاد لاختبار أوتيس لينون، وبين الابعاد والاختبار ككل، فقد بلغت قيم معاملات الارتباط البينية لأبعاد اختبار أوتيس– لينون كانت متوسطة، وتراوحت بين (0.455– 0.706)، كما تراوحت قيم معاملات الارتباط بين المجالات والقائمة ككل بين (0.69–0.90). وللتحقق من ثبات درجات اختبار أوتيس لينون، تم حساب معامل الثبات باستخدام معامل الفا كرونباخ، وقد بلغت قيمة ثبات معامل الفا لاختبار أوتيس لينون للقدرة العقلية (0.887) وهو معامل ثبات مرتفع وجيد. كما تم استخراج مؤشر الثبات التجريبي Reliability Index والذي يعد مؤشر على دقة التقدير وقد بلغ(0.899)

إجراءات القياس التكيفي:

1.مطابقة البيانات: تم التحقق من مطابقة البيانات لطريقة دلتا حيث تم إدخال البيانات لبرنامج دلتا للتحقق من مطابقة استجابات الأفراد للطريقة دلتا، أعتمد البحث الحالي على أحد مؤشرات مطابقة الأفراد المستخرجة باستخدام برنامج دلتا والذي يتمثل في إحصائي (Ud Statistic) والذي يعتمد على المؤشر الإحصائي(U3) U3 Person Fit- Statistic، حيث تنحصر قيمة المؤشر (Ud) بين (0–1) حيث يشير

الصفر إلى مطابقة استجابة المفحوص للنموذج بينما يشير الواحد إلى عدم المطابقة التامة Dimitrov (et,al.,2021)، وقد أظهرت نتائج التحليل وجود (30) استجابة غير مطابقة، حيث تم استبعاد الأفراد غير المطابقين. وبذلك أصبح العدد النهائي لعينة التدريج (1370)، وهي الاستجابات التي تم إدخالها من أجل تدريج كل من المفردات والأفراد وفقاً لطريقة دلتا.

وللكشف عن مطابقة المفردات لطريقة دلتا اعتمدت على المؤشرات المدرجة في برنامج (DALTA). فقد استخدم مؤشر متوسط الاختلاف المطلق(MAD) Mean Absolute Difference المشار إليه من قبل يمتروف وآخرون (Dimitrov et,al.,2021) في التعرف على مطابقة المفردة للنموذج وفق المحكات الآتية: تدل قيمة المؤشر($MAD \leq 0.07$) على المطابقة الجيدة للمفردة؛ بينما يدل وقوع قيمة المؤشر بين($0.07< MAD <0.01$) على المطابقة المقبولة للمفردة، وتدل القيمة($MAD \geq 0.01$)على المطابقة الضعيفة للمفردة. ومن خلال فحص قيم مؤشر متوسط الاختلاف المطلق فقد أظهرت جميع المفردات وقوعها في مدى المطابقة الجيد والمقبول لطريقة دلتا مما يدل على ملائمة جيدة لجميع المفردات (Dimitrov & Atanaso,2021).

2.عينة القياس التكيفي: تكونت عينة القياس التكيفي من (130) مفحوص، وقد روعي خلال ذلك أن ينتشر المفحوصين على كافة فئات الدرجة للمفحوصين.

3.تدريج المفردات: تم ترتيب المفردات تصاعدياً حسب قيم معلم الصعوبة المفردات.

4.خطوات إجراء القياس التكيفي:

1. بعد ترتيب المفردات تصاعدياً حسب معلم الصعوبة لكل مفردة، جري اختيار خمس مفردات تغطي مستويات صعوبة المفردات جميعها لتكوّن مفردات الاختبار الاستطلاعي الذي يجري بواسطته تحديد قدرة مبدئية للمفحوص يجري على أساسها اختيار مفردات المرحلة اللاحقة التي ستقدم للمفحوصين، وأرقام هذه المفردات هي (26، 36، 73، 32، 13)، وقد تراوحت قيم معلم الصعوبة لهذه المفردات بين 0.823 لوجيت 0.094 لوجيت.
2. قسمت المفردات الثمانون لتكوّن خمس اختبارات فرعية اعتماداً على معلم الصعوبة لكل مفردة يجري اختيار مفردات المراحل الثانية والثالثة والرابعة منها بناءً على الدرجة التي تُعطي للمفحوص إثر إجابته على الاختبار الاستطلاعي أولاً ثم المراحل التي تسبق المرحلة المعنية.
3. جري تحديد المفردات التي ستعطى للمفحوص بعد مرحلة الاختبار الاستطلاعي في المرحلة الثانية والثالثة والرابعة اعتماداً على الدرجة المتحققة من الاختبار الاستطلاعي.
4. جرى تقديم مفردات الاختبار الاستطلاعي لكل المفحوصين في عينة القياس التكيفي وحساب درجة مبدئية لكل منهم باستخدام برنامج Delta وذلك باستخدام البيانات السابقة التي تحتوي على استجاباتهم على الاختبار الخطي، بحيث تم حذف الاستجابات جميعها باستثناء الاستجابات على مفردات الاختبار الاستطلاعي الخمس.

5. بناء على الدرجة المحسوبة لكل مفحوص جرى تقديم خمس مفردات أخرى تتناسب مع درجة المفحوص المبدئية من المجموعات الفرعية الخمسة، وحساب درجة المفحوص الجديدة باستخدام المفردات الجديدة والمفردات السابقة التي كونت الاختبار الاستطلاعي.
6. بناء على الدرجة المحسوبة من المرحلة الثانية جرى تعريض المفحوصين لخمس مفردات جديدة تتناسب مع قدرة المفحوص الجديدة وحساب درجة المفحوص الجديدة باستخدام المفردات الجديدة والمفردات السابقة في المرحلة الأولى والمرحلة الثانية.
7. بناء على الدرجة المحسوبة من المرحلة الثالثة جرى تعريض المفحوصين لخمس مفردات جديدة تتناسب مع درجة المفحوص الجديدة وحساب درجة المفحوص النهائية باستخدام المفردات الجديدة والمفردات السابقة في المراحل الثلاث السابقة معاً، وبالتالي أصبح لكل مفحوص أربع درجات(DL_1) متحصلة من الاستجابات على مفردات الاختبار الاستطلاعي،(DL_2) متحصلة من الاستجابات على مفردات المرحلة الثانية ومفردات المرحلة الأولى(الاستطلاعية)، و(DL_3) متحصلة من الاستجابات على مفردات المراحل الثلاث المرحلة الأولى(الاستطلاعية)، والمرحلة الثانية والمرحلة الثالثة، و(DL_4) متحصلة من الاستجابات على الأربع المراحل.

النتائج

تم حسب تقديرات الدرجات المتحصلة لأفراد عينة القياس التكيفي من تطبيق القياس التكيفي في المرحلة الأولى(DL_1) وهي مرحلة الاختبار الاستطلاعي، و(DL_2) وهي الدرجة المقدرة من المرحلة الثانية والمضاف إليها مفردات المرحلة الأولى، و(DL_3) وهي الدرجة المقدرة المرحلة الثالثة والمكونة من خمس مفردات والمضاف إليها مفردات المرحلتين الأولى والثانية، (DL_4) وهي الدرجة المتحصلة من المرحلة الرابعة مضافاً إليها مفردات المراحل الثلاث السابقة، إضافة إلى حساب الخطأ المعياري في التقدير لجميع المراحل و الدرجة المقدرة من الاختبار الخطي(DL_L) لجميع أفراد عينة القياس التكيفي.

وللإجابة عن السؤال الأول من أسئلة الدراسة يبين الجدول(2) المتوسطات الحسابية للدرجات المقدرة من المراحل الأربع ومعاملات الارتباط بيرسون بين كل من الدرجات المقدرة في المراحل الأربع والدرجة المقدرة من الاختبار الخطي بمفرداته الثمانين، ومتوسط القيم المطلقة للفروق في تقديرات الدرجة في المراحل الأربع مقارنة بالدرجة المقدرة من الاختبار الخطي بمفرداته الثمانين، والمتوسط الحسابي للخطأ المعياري في التقدير للمفحوصين، و الجذر التربيعي لمتوسط مربعات الخطأ المعياري في التقدير(RMSE) لكل مرحلة منسوبة إلى الخطأ المعياري في تقدير الدرجة المتأتية من الاختبار الخطي بمفرداته الثمانين.

جدول (2) يوضح القيم الإحصائية المتعلقة بتقديرات درجات دي (DSM-L) في المراحل المختلفة للقياس التكيفي مقارنة بالدرجة الحقيقة المقدرة من الاختبار الخطي لطريقة دلتا الكامنة

	المرحلة الأولى (DL_1)	المرحلة الثانية (DL_2)	المرحلة الثالثة (DL_3)	المرحلة الرابعة (DL_4)	الاختبار الخطي (DL_L)
المتوسط الحسابي للقدرة	47.676	45.219	46.810	47.128	46.246
معامل الارتباط	0.710	0.750	0.854	0.878	1
متوسط القيم المطلقة للفروق	17.855	10.371	7.315	5.728	-
المتوسط الحسابي للخطأ المعياري	0.151	0.134	0.113	0.096	0.044
متوسط الفروق في الخطأ المعياري	0.119	0.091	0.069	0.053	-

ويتضح من الجدول رقم (2) أن المتوسط الحسابي لقدرات المفحوصين في الاختبار الاستطلاعي كانت (47.676) درجة وهي قريبة جداً من المتوسط الحسابي للقدرة الحقيقية المتأتية من الاختبار الخامس بمفرداته السبعين وهي(46.246) أي بفارق(1.433) درجة، وفي المرحلة الثانية كان متوسط تقديرات الدرجة يساوي(45.219) وهنا يظهر أن المرحلة الثانية لم تحسن تقدير الدرجة للاقتراب من الدرجة الحقيقة إلا بمقدار بسيط وهو (1.027) درجة، وفي المرحلة الثالثة كان متوسط تقديرات الدرجة يساوي(46.810) وهنا يظهر أن المرحلة الثالثة حسنت تقدير الدرجة بدرجة عالية جداً واقتربت من الدرجة الحقيقة، أما المرحلة الرابعة كانت متوسط تقديرات الدرجة يساوي(47.128) هنا يظهر أن المرحلة الرابعة زادت عن تقدير الدرجة الحقيقة بمقدار ضئيل وهو(0.882) والسبب هنا أن المرحلة الرابعة تضمنت عدد مفردات أكثر من المراحل السابقة وهو (20) مفردة، وهذا يعنى انه عند استخدام طريقة دلتا يمكن الاكتفاء بثلاث مراحل وعدد مفردات (15) مفردة. وهذه النتائج تتفق مع دراسة هان وآخرون(Han et.al., 2019) أن استخدام طريقة دلتا يمكن أن تحقق تحسينات في دقة القياس وكفاءته مقارنة بالاختبار الخطي المبنى على طريقة دلتا.

أما بالنسبة للمتوسط الحسابي للفروق المطلقة أظهر أن تقديرات المرحلة الأولى الاستطلاعية تبتعد في المتوسط عن الدرجة الحقيقة بمقدار (-1.43) درجة، وفي المرحلة الثانية أصبحت تبتعد بمقدار(1.02)

درجة، وفي المرحلة الثالثة اقتربت من الدرجة الحقيقة بمقدار(-0.564) وهي تعد قريبة من الدرجة الحقيقة، لاسيما لو أننا أعدنا تطبيق الاختبار الخطي بمفرداته الثمانين فسنحصل على فروقات ربما تساوي الفروق الحالية مع القياس التكيفي الذي استخدم فقط (15) مفردة من أصل (80)، أما المرحلة الرابعة أصبحت تبتعد بمقدار(-0.882) وتتفق هذه النتيجة مع نتائج دراسة هان وآخرون(Han et al., 2019) أن استخدام طريقة دلتا يمكن أن تحقق تحسينات في دقة القياس وكفاءته مقارنة بالاختبار الخطي المبنى على طريقة دلتا.

ومن حيث المؤشر الثاني على فاعلية القياس التكيفي باستخدام طريقة دلتا لتقدير الدرجات وهو معامل الارتباط فيظهر أن معامل الارتباط في المرحلة الاستطلاعية الأولى مع الدرجة الحقيقة كان (0.710) بتباين مفسر قدره(0.504) أما المرحلة الثانية فقد أصبح معامل الارتباط (0.750) ويلاحظ هنا مدى التحسن في التقدير باتجاه الدرجة الحقيقية، أما المرحلة الثالثة فقد أصبح معامل الارتباط (0.854) ويلاحظ هنا مدى التحسن الكبيرة في التقدير باتجاه الدرجة الحقيقية، لكن في المرحلة الأخيرة أصبح معامل الارتباط(0.878) وبتباين مفسر مقداره (0.770) الأمر الذي يبن مدى اقتراب تقدير الدرجة باستخدام الاختبار الذي احتوى فقد(20) مفردة من تقدير الدرجة باستخدام الاختبار الخطي بمفرداته الثمانين، علماً بأن معاملات الارتباط المذكورة دالة عند مستوى(α=0.01)، وتتفق نتيجة الدراسة الحالية جزئياً مع نتائج دراسة كل بتسلو (patsula,1999)، أنه كلما زاد عدد المراحل الاختبارية كلما كان هناك دقة أكبر لتكييف الاختبار ليلائم قدرة الطالب، وعندما تمت زيادة عدد المراحل الاختبارية من ثلاثة مراحل إلى خمس زادت دقة القياس على متصل الدرجة. وتتفق النتيجة السابقة مع نتائج دراسة Al-Shdifat (2008)، دارسة Al-Bayaida (2011)، دراسة Al-Bursan (2012)، ودراسة أوزتورك (Öztürk , 2019) التي أوضحت أن للاختبار التكيفي متعدد المراحل دقة في تقدير قدرة الأفراد وذلك مقارنة مع الاختبار الخطي، كما تتفق النتيجة السابقة مع ما أشار إليه ديلز (Dallas, 2014) أن زيادة عدد المراحل من مرحلتين إلى ثلاثة يزيد من دقة القياس.

وبالنسبة للخطأ المعياري في التقدير يُظهر الجدول أن متوسط الخطأ المعياري في التقدير كان في المرحلة الأولى والاستطلاعية(0.151) ثم أصبح في المرحلة الثانية(0.134) ويلاحظ هنا مقدار الانخفاض في الخطأ المعياري في التقدير الذي اعتمد على (10) مفردات ، لكن في المرحلة الثالثة التي تضمنت (15) مفردة أصبح (0.113) وهو يعتمد على (15) مفردة، أما المرحلة الرابعة أصبح(0.096) وهو يبتعد بمقدار(0.052) عن متوسط الخطأ المعياري في تقديرات الدرجات في الاختبار الخطي والذي يبلغ(0.044)، ويلاحظ أن الخطأ المعياري بدأ بمقدار كبير في المرحلة الأولى(الاستطلاعية) إلا أنه وصل لمقدار معقول في المرحلة الرابعة حيث كانت نسبة متوسط الخطأ المعياري في المرحلة الرابعة إلى متوسط الخطأ المعياري في الاختبار الخطي تساوي(1: 2) تقريباً، بينما كانت نسبة عدد مفردات الاختبار الخطي إلى عدد مفردات الاختبار التكيفي(1: 4) وهذا يبين مقدار الخفض الكبير في الكلفة والجهد للفاحص والمفحوص في الاختبار

التكيفي مقارنة بالاختبار الخطي في مقابل التضحية بمقدار قليل جداً من الدقة في القياس باستخدام طريقة دلتا لتقدير الدرجات.

أما بالنسبة للجذر التربيعي لمتوسط مربعات الخطأ المعياري في التقدير كان في المرحلة الأولى والاستطلاعية(0.119) ثم أصبح في المرحلة الثانية(0.091) ويلاحظ هنا مقدار الانخفاض في الخطأ المعياري في التقدير الذي اعتمد على (10) مفردات ، لكن في المرحلة الثالثة التي تضمنت (15) مفردة أصبح (0.069) وهو يعتمد على (15) مفردة، أما المرحلة الرابعة أصبح(0.053) وهذا يدل على زيادة في دقة القياس حيث أن قيمة الجذر التربيعي لمتوسط مربعات الخطأ المعياري في التقدير(RMSE) كلما انخفضت القيمة واقتربت من الصفر دل هذا دقة أعلى (Vaispoel, 1993؛ vispoel et al,1997 ؛ Weiss, 1982؛wang et al.,1999).

وتتفق النتيجة السابقة مع ما ورد في الأدب في مجال القياس التكيفي من حيث أن دقة القياس تكون أفضل في المرحلتين الثانية والثالثة من مراحل التصميمات المختلفة للاختبار التكيفي متعدد المراحل، وبالتالي يُقترح ألا يقل عدد مفردات وحدة التوجيه عن (10) مفردات، كما تتفق النتيجة السابقة جزئياً مع نتيجة دراسة ساري ومانلي (Sari & Manley,2017) حيث أشارت أن طول الاختبار ونوع التصميم لها تأثير كبير على نتائج وخصائص الاختبار بدرجة أكبر من تأثير مجالات المحتوى وعددها.

التوصيات والمقترحات

في ضوء نتائج الدراسة الحالية فإن الباحثان توصي بالآتي: -

1- اعتماد القياس التكيفي متعدد المراحل باستخدام طريقة دلتا لتقدير الدرجات مكان القياس الخطي القائم على طريقة دلتا لتقدير الدرجات مع مراعاة عدد المفردات في كل مرحلة وعدد المراحل الاختبارية حيث يمكن الاكتفاء ب (15) مفردة وثلاث مراحل.

2- استخدام تصميمات الاختبار التكيفي متعدد المراحل قائمة على النماذج المختلفة لطريقة دلتا وذلك للكشف عن العوامل مثل (طريقة اختيار المفردات، طريقة تقدير القدرة، حجم العينة، البيانات المفقودة، وتوزيع معالم المفردات والأفراد...) التي تؤثر على دقة تقدير معالم الأفراد والمفردات في ضوء نماذج دلتا.

3- استخدام محكات مختلفة لانتقال المفحوص بين المراحل الاختبارية وبيان أثر أكثر محكات الانتقال ملائمة للتصميم متعدد المراحل المستخدم، وبيان مدى تأثير تغير محكان الانتقال بين المراحل الاختبارية على دقة تقدير قدرة المفحوص النهائية، وعدم الاقتصار على محك الانتقال تبعاً لمستوى صعوبة المفردة، في ضوء مستوى قدرة المفحوص الحالية.

4- استخدام طرق مختلفة لتقدير قدرة المفحوص في كل مرحلة اختبارية، والمقارنة بين تلك الطرق لبيان مدى دقة كل منها في تقدير قدرة المفحوص، وعدم الاقتصار على طريقة الأرجحية العظمي لتقدير القدرة وذلك باستخدام طريقة دلتا.

المقترحات

- إجراء دراسات أخرى للكشف عن أثر طريقة اختيار المفردة في الاختبار التكيفي باستخدام طريقة دلتا لتقدير الدرجات يحتوي على مزيجاً من المفردات ثنائية ومتعددة التدريج، حيث اعتمدت الدراسة الحالية على فقرة ثنائية الاستجابة.
- إجراء مزيد من الدراسات للمقارنة بين الاختبارات التكيفية، والاختبارات التكيفية متعددة المراحل باستخدام تصاميم مختلفة، حيث اعتمدت الدراسة الحالية على تصميم واحد.

المصادر

أولاً: المراجع العربية

- البرصان، إسماعيل سلامة (2012). فاعلية القياس التكيفي باستخدام فقرات ذات إجابة منتقاة وفقرات ذات إجابة منشأة. *مجلة جامعة الملك سعود، العلوم التربوية والدراسات الاسلامية، 24* (4)، 1487-1518.
- البلوي، خضراء عبدالله (2019). أثر اختلاف قدرة الأفراد على دقة تقدير معالم المفردات والأفراد في أسئلة الاختيار من متعدد وفق النموذج اللوجستي الثلاثي المعلم، *المجلة الدولية المتخصصة 8*(5)، 78-91.
- البيايضة، آلاء محمد معزي (2011). *بناء اختبار تكيفي للقدرة الرياضية للصف السابع الأساسي وفق الإستراتيجية الهرمية باستخدام نظرية الاستجابة للفقرة*. (رسالة ماجستير غير منشورة). جامعة مؤتة، مؤتة. مسترجع من http://search.mandumah.com/Record/785139
- الشديفات، صباح جميل فدعوس (2008) *بناء اختبار تكيفي لقياس القدرة الرياضية وفق الاستراتيجية ثنائية المرحلة في نظرية الاستجابة للمفردة* (رسالة ماجستير غير منشورة). جامعة اليرموك، إربد. مسترجع من http://search.mandumah.com/Record/742259
- الشريفين، نضال كمال؛ وبنى عطا، زايد صالح (2013). تقصى أثر عدد خطوات الأسئلة متعددة التدريج وشكل التوزيع لصعوبتها على تقديرات القدرة للأفراد والصعوبة للأسئلة ودالة المعلومات للاختبار وفق نموذج التقدير الجزئي. *المجلة التربوية،28*(109)، 213-275.
- نصراوين، معين سلمان (2018). دقة تقدير معالم الفقرات عند استخدام أربعة نماذج لوجستية في إطار نظرية الاستجابة للفقرة. *مجلة العلوم التربوية، 45*(4)، 179-205.

Refernces

- Al Sharifain, Nidal Kamal; Bani Atta, Zayed Saleh (2013). Tracing the effect of a number of steps of the multiple-graded questions and the shape of the distribution of their difficulty on the ability ratings of individuals, the difficulty of the questions and the information function of the test according to the partial estimation model. *Journal Educational, 28*(109), 213-275.
- Al-Balawi, Khadra Abdullah (2019). The effect of the difference in the ability of individuals on the accuracy of estimating vocabulary and individuals' parameters in multiple-choice items according to the three-way logistic model of the teacher. *Journal of Specialized International 8*(5), 78-91.
- Al-Bayaida, Alaa Muhammad Moazi (2011). Building an adaptive test of mathematical ability for the seventh grade according to the hierarchical strategy using item response theory. (Unpublished Master dissertation). Mutah University, Mutah. Retrieved from http://search.mandumah.com/Record/785139.
- Al-Bursan, Ismail Salameh (2012). The effectiveness of adaptive testing using dichotomous and polytomous items. *Journal of King Saud University, Educational Sciences, and Islamic Studies, 24* (4), 1487-1518.
- Al-Shdifat, Sabah Jamil Fadoos (2008). Building an adaptive test to measure mathematical ability according to the two-stage strategy in the item response theory (unpublished master's dissertation). Yarmouk University, Irbid. Retrieved from http://search.mandumah.com/Record/742259.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. Harcourt Brace Jovanovich.
- Dallas, A. (2014). *The effects of routing and scoring within a computer adaptive multi-stage framework* (Doctoral dissertation, University of North Carolina at Greensboro).
- De Champlain, A. F. (2010). A primer on classical test theory and item response theory for assessments in medical education. *Medical education*, *44*(1), 109-117.
- Dimitrov, D. (2020b). *Using the D-Scoring Method in Large-Scale Assessments* [Workshop]. International Conference on Education & Training Evaluation, Education & Training Evaluation Commission (ETEC).
- Dimitrov, D. M. (2016). An Approach to Scoring and Equating Tests with Binary Items: Piloting with Large-Scale Assessments. *Educational and Psychological Measurement*, *76*(6), 954–975. https://doi.org/10.1177/0013164416631100
- Dimitrov, D. M. (2018). The Delta-Scoring Method of Tests with Binary Items: A Note on True Score Estimation and Equating. *Educational and Psychological Measurement*, *78*(5), 805-825.
- Dimitrov, D. M. (2020a). Modeling of item response functions under the D-scoring method. *Educational and Psychological Measurement*, *80*(1), 126-144.

- Dimitrov, D. M., Atanasov, D. V., & Luo, Y. (2021). Person-fit assessment under the D-scoring method. *Measurement: Interdisciplinary Research and Perspectives*, *18*(3), 111-123.
- Dimitrov, D.M., & Atanasov, D.V. (2020). Latent D-scoring modeling: Estimation of item and person parameters. *Educational and Psychological Measurement*, *81*(2), 388-404.
- Domingue, B. W., & Dimitrov, D. M. (2015). A comparison of IRT theta estimates and delta scores from the perspective of additive conjoint measurement (Research Rep., RR-4-2015). *Riyadh, Saudi Arabia: National Center for Assessment.*
- Domingue, B., & Dimitrov, D. (2021). A comparison of IRT theta estimates and delta scores from the perspective of additive conjoint measurement.
- Hambleton, R.H., & Swaminathan, H. (1985). Item Response Theory. Principles and Application. Boston: k lower-nigh of tests and *Item. Applied Psychological Meassurment*,9(2), (136- 164).
- Han, K. T., Dimitrov, D. M., & Al-Mashari, F. (2019). Developing multistage tests using D-scoring method. Educational and Psychological Measurement, 79(5), 988-1008. https:// doi.org/10.1177/0013164419841428
- Mousavi, S. A. (2015). *The effect of person misfit on item parameter estimation A simulation study*. (Unpublished doctoral dissertation). University of Alberta.
- Nasraween, Mueen Salman (2018). The accuracy of estimating the parameters of the paragraphs when using four logistic models within the framework of the paragraph response theory. *Journal of Educational Sciences*, *45*(4), 179-205.
- Öztürk, N. B. (2019). How the Length and Characteristics of Routing Module Affect Ability Estimation in ca-MST? Universal Journal of Educational Research, 7(1), 164-170.
- Patsula, L. N. (1999). A comparison of computerized adaptive testing and multistage testing. University of Massachusetts Amherst.
- Rotou, O., Headrick, T. C., & Elmore, P. B. (2002). A proposed number correct scoring procedure based on classical true-score theory and multidimensional item response theory. *International Journal* of Testing, 2(2), 131-141.
- Sands, W., Walters, B., & Bride, J. (2001). *Computerized adaptive testing: From Inquiry to operation. American psychological association*. Washington DC: American Psychological Association.
- Sari, H. İ., & Huggins-Manley, A. C. (2017). Examining content control in adaptive tests: Computerized adaptive testing vs. computerized adaptive multistage testing. Educational Sciences: *Theory & Practice*, *17*(5).
- Vispoel, W. P. (1993). The Development and Evaluation of a Computerized Adaptive Test of Tonal Memory. *Journal of Research in Music Education*, 41(2), 111–136. https://doi.org/10.2307/3345403.

- Vispoel, W. P., Wang, T., & Bleiler, T. (1997). Computerized adaptive and fixed-item testing of music listening skill: A comparison of efficiency, precision, and concurrent validity. *Journal of Educational Measurement*, *34*(1), 43-63.
- Wang, T., Hanson, B. A., & Lau, C. M. A. (1999). Reducing bias in CAT trait estimation: A comparison of approaches. *Applied Psychological Measurement*, *23*(3), 263-278.
- Warm, A. (1978). *A Primer of Item Response Theory*: Us Cost Guard Institute Oklahoma 73/69.
- Weiss, D. J. (1982). Improving Measurement Quality and Efficiency with Adaptive Testing. *Applied Psychological Measurement*, 6(4), 473–492. https://doi.org/10.1177/014662168200600408.
- Yurekli, H. (2010). *The Relationship Between Parameters from Some Polytomous Item Response Theory Models*. (Unpublished master's thesis). Florida State University Gainesville.