مقارنة طرق التدريج العمودي المبنية على إجراءات النظرية الكلاسيكية في القياس بالطرق المبنية على نظرية الاستجابة للفقرة

*شريف عبدالرحمن السعودي، هلا محمد الشوا**

ملخص

هدفت هذه الدراسة إلى مقارنة طرق التدريج العمودي المبنية على إجراءات النظرية الكلاسيكية في القياس بالطرق المبنية على نظرية الاستجابة للفقرة، حيث تم استخدام طريقتي ثيرستون وهيرونيموس بناءً على النظرية الكلاسيكية، والنموذجين اللوجستيين الثنائي والثلاثي، كل منهما بطريقتي لتقدير القدرة هما: طريقة الأرجحية العظمى، وطريقة توقّع التوزيع البعدي، حيث تمت المقارنة على وفق أربعة محكّات هي: المتوسط الحسابي، الانحراف المعياري، حجم الأثر، المسافة الأفقية، وذلك باستخدام اختبار رياضيات تحصيلي متعدد المستويات للصفوف الرابع والخامس والسادس الأساسي في الأردن، وفق تصميم الفقرات المشتركة، على عينة بلغت (1630) طالباً وطالبة من الصفوف المستهدفة، وأشارت النتائج إلى اختلاف خط النمو الأكاديمي من صفٍ إلى آخر باختلاف الطريقة الاحصائية المستخدمة في التدريج العمودي.

الكلمات الدالة: التدريج العمودي، طريقة ثيرستون، طريقة هيرونيموس، نظرية الاستجابة للفقرة.

المقدمة

لا ريب أنّ تتبع تقدم ونمو الطلبة الأكاديمي عبر الصفوف الدراسية المتتالية، قد أصبح موضوعاً هاماً لبرامج التقييم المختلفة، إذ يحظى باهتمام التربويين القائمين على تقييم نمو الطلبة الأكاديمي مع مرور الوقت؛ نظراً لدوره في اتخاذ العديد من القرارات المتعلقة ببناء وتعديل المناهج وطرق التدريس والتخطيط لعمليات التعلم، مما يتطلب مقاييس خاصة تتمتع بالصدق والثبات لهذا الغرض، ويعدُ التدريج العمودي (Vertical Scaling) وسيلةً لمراقبة هذا التقدم والنمو الأكاديمي، إذ يتم استخدام اختبارات تقيس مجالات متماثلة لمستويات دراسية مختلفة ووضعها من خلال طرق إحصائية معينة على مقياس مشترك وحيد لجميع المستويات.

ويشير تونغ وكولين (Tong & Kolen, 2007) ولن (Linn, 2001) إلى أن التدريج العمودي طريقة لتعقب النمو الأكاديمي، وإن استخدام الدرجات الخام لا يقدم معلومات كافية لإجراء مثل هذه المقارنات، ومن ثم لا بدَ من التركيز على تقييم كم يتقدم الطالب مع الوقت وأن يكون التحصيل معتمداً على النمو الأكاديمي لا على تقديرات القدرة المختلفة، كما وترى هولاند ودورانس (Holland & Dorans, 2006) أن التدريج العمودي هو عملية تطوير مقياس مشترك لمستويات تعليمية متعددة، بهدف تكوين متصل تطوري يسمح بمقارنة درجات الاختبار على مستويات مختلفة، مما يتطلب اختبارات متعددة المستويات يطور كل منها؛ ليلائم المفحوصين في كل صف، والتي تجعل تقييم النمو الأكاديمي أكثر سهولة بالإضافة إلى إمكانية وصف أداء الطلبة لأي من مستويات الاختبار، حيث يشير باتز (Patz, 2007) إلى أنه من أكبر التحديات المعاصرة في ميدان التقييم التربوي هو معرفة كمية ما تعلم الطالب مع مرور الوقت من خلال التقدم في المستويات الدراسية.

وفي هذا الصدد، قدّم كولن وبرينان (Kolen & Brennan, 2004) تعريفين رئيسين للنمو هما: تعريف نمو المجال (Domain Growth) وتعريف النمو من صف دراسي لآخر، (Grade to grade Growth)، وفي تعريف نمو المجال يحدد نمو الطلبة الأكاديمي على كامل المجال الذي يتم قياسه، وعملياً يشير هذا التعريف إلى معدل النمو من سنة لأخرى على كامل مجال المحتوى، أما تعريف النمو من صف لآخر فيحدد النمو من خلال المحتوى الذي هو جزء من المنهاج في صف معين، إذ يركز هذا التعريف على الاختلافات في النمو بين الصفوف المتجاورة باستخدام مواضيع وثيقة الصلة بالمنهاج، مثل

* قسم المناهج والتدريس، الجامعة الأردنية. تاريخ استلام البحث 2014/12/1، وتاريخ قبوله 2015/2/25.

©2017 عمادة البحث العلمي وضمان الجودة/ الجامعة الأردنية. جميع الحقوق

الرياضيات الذي يعد من الموضوعات ذات الطبيعة التراكمية التطورية.

وأدى اختلاف تعريف النمو لوجود تصاميم مختلفة لجمع البيانات في القياس العمودي، إذ يوجد ثلاث تصاميم رئيسة لجمع البيانات لبناء المقياس العمودي، هي: تصميم الفقرات المشتركة (Common Item Design)، وتصميم اختبار التدريج (Design Scaling Test)، وتصميم المجموعات المتكافئة (Equivalent Groups Design)، حيث يتم في تصميم الفقرات المشتركة ربط الصفوف المتجاورة بفقرات مشتركة، وتقييم النمو الأكاديمي بناءً على الاختلافات في أداء المتعلمين على هذه الفقرات، أما في تصميم اختبار التدريج فيقدم اختبار واحد لجميع المفحوصين، من مختلف المستويات بالإضافة إلى اختبار آخر خاص بكل مستوى ويسمى اختبار المستوى، وفي تصميم المجموعات المتكافئة، يتم اختيار مجموعات عشوائية وتطبيق اختبارات معدّة حسب مستواهم، ويرتبط مع تعريف المحتوى، وترتبط خصائص المقياس الناتج حسب التصميم المستخدم (Kolen, 2011).

أما فيما يخص الطرق الإحصائية المستخدمة في التدريج العمودي، فقد ورد في الأدب النظري ثلاث طرق رئيسة للتدريج العمودي، هي: طريقة ثيرستون (Thurstone Method)، وطريقة هيرونيموس (Hieronymus Method)، وطرق نماذج نظرية الاستجابة للفقرة [Item Response Theory Models, (IRT)]، وتعدُ طريقة ثيرستون من أقدم الطرق الكلاسيكية، إذ تفترض أن توزيع درجات جميع المستويات يمثل توزيعاً طبيعياً، بالإضافة إلى أن الأداء على الفقرات لا يكون من خلال التخمين، أما طريقة هيرونيموس، فهي أيضاً من الطرق الكلاسيكية، ويتم ربط المستويات فيها من خلال طرق الربط المئينية، وفيما يخص نماذج نظرية الاستجابة للفقرة، فهي تضيف تعقيدات أكثر على التدريج العمودي من خلال افتراضاتها القوية، مثل: أحادية البعد والاستقلال الموضعي، ويمكن استخدام جميع نماذج نظرية الاستجابة للفقرة كإجراء للتوصل إلى المقياس العمودي (Williams, Pommerich & Thissen, 1998؛ Hoover, 1988 Kolen, 2006؛).

ويضيف كولين (Kolen, 2011) أن هنالك عدة عوامل تؤثر في نتائج التدريج العمودي، وخصائص المقياس الناتج، ومن اهم هذه العوامل تعريف النمو الذي يتم اعتماده مسبقاً، وتصميم جمع البيانات المستخدم، والطرق الإحصائية المستخدمة لبناء المقياس، بالإضافة لتأثير المنهاج موضع الاهتمام، في حين اقترح كل من تونغ وهاريس (Tong & Hariss, 2004) عوامل أخرى خاصة بنظرية الاستجابة للفقرة، وهي طرق تقدير القدرة والمعالم المستخدمة، وطريقة المعايرة بالإضافة لطريقة إعطاء الدرجات، وأضاف وليامز وآخرون (Williams at al, 1998) عوامل أخرى، وهي: خصائص فقرات الاختبار، والعلاقة بين قدرات الأفراد، وخصائص الفقرات، بالإضافة إلى البعدية ونوع الاستجابة على الفقرات (ثنائية، مبنية)، وأخيراً نوع التحويلات الخطية واللاخطية المستخدمة، وعليه فإن خصائص المقياس العمودي تعتمد على الطريقة والإجراءات التي يتم فيها بناء هذا المقياس.

مشكلة الدراسة وأسئلتها

بناءً على ما سبق يمكن القول أن عملية التدريج العمودي تتضمن العديد من المراحل لتطوير المقياس العمودي، وأن الطرق الكلاسيكية والحديثة تؤدي إلى مقاييس عمودية متباينة، وبالتالي اختلاف التفسيرات الناتجة لنمو الطلبة الأكاديمي، إذ يؤدي اختلاف الطريقة الإحصائية المستخدمة في التدريج العمودي إلى اختلاف خط النمو الأكاديمي الناتج، وعليه أوصت العديد من الدراسات في مجال التدريج العمودي إلى إجراء مقارنات تجمع طرق مختلفة لمعرفة أثر هذه الطرق في خط النمو الأكاديمي الذي ينتج من كلٍ منها، ومعرفة مدى ملائمة هذه النتائج وارتباطها بالواقع.

وبالتالي جاءت هذه الدراسة لتحري الفروق في طرق التدريج العمودي الكلاسيكية (ثيرستون وهيرونيموس)، وطرق نماذج نظرية الاستجابة للفقرة باستخدام النماذج اللوجستية ثنائية الاستجابة الثنائي والثلاثي وبطريقتي تقدير للقدرة هما طريقة الأرجحية العظمى [(Maximum Likelihood Estimation, MLE) وطريقة توقّع التوزيع البعدي [(Expected A posteriori, EAP)، باستخدام اختبار رياضيات تحصيلي متعدد المستويات للصفوف الرابع والخامس والسادس الأساسي في الأردن، وفق تصميم الفقرات المشتركة، ومن خلال محكّات النمو والتغيّر عبر هذه الصفوف والمتمثلة في (المتوسط الحسابي، الانحراف المعياري، حجم الأثر، المسافة الأفقية)، وبشكل أكثر تحديداً تمثّلت مشكلة الدراسة بالسؤال البحثي الآتي:

"ما أثر طرق التدريج العمودي (ثيرستون، هيرونيموس، النموذج الثنائي باستخدام MLE، النموذج الثنائي باستخدام EAP، النموذج الثلاثي باستخدام MLE، النموذج الثلاثي باستخدام EAP) على المقياس العمودي الناتج (خط النمو) من حيث:

- معدل النمو بدلالة المتوسط الحسابي.
- التغيّر في النمو بدلالة الانحراف المعياري؟
- حجم الأثر
- المسافة الأفقية

أهمية الدراسة:

تكتسب الدراسة جانباً من أهميتها في أنها تناولت موضوع التدريج العمودي وليس المعادلة العمودية التي تناولتها الكثير من الدراسات العربية، ولا يوجد – في حدود علم الباحث – دراسات عربية تناولت هذا الموضوع، بالإضافة لاستخدامها ست طرق مختلفة من خلال بيانات حقيقية وليست افتراضية، واعتماد طرق التقدير المختلفة للمقارنة، ومحكّات حديثة للمقارنة.

كما ويمكن أن تسهم هذه الدراسة في توضيح معالم جديدة لمصممي الاختبارات في المجال التربوي، والمهتمين في مجال النمو الأكاديمي، وتسريع التعليم بأسلوب جديد في قياس تحصيل الطلبة، واتخاذ القرارات الخاصة بالنمو والتسريع، بالإضافة إلى أنها تعمل على توفير اختبار متعدد المستويات في الرياضيات للصفوف الرابع والخامس والسادس الأساسي باستخدام تصميم الفقرات المشتركة، يمكن استخدامه كمؤشر لتقييم نوعية التعليم من قبل المعلمين، والباحثين، ومتخذي القرار، مما يسهم في تحسين تعلم الطلبة.

حدود الدراسة ومحدداتها:

تحددت نتائج هذه الدراسة بما هو آتٍ:

– اقتصار طرق التدريج العمودي باستخدام نماذج نظرية الاستجابة للفقرة على النماذج أحادية البعد ثنائية الاستجابة الثنائي والثلاثي.

– اقتصار تصاميم جمع البيانات على تصميم الفقرات المشتركة.

– اقتصار مواضيع المناهج المستهدفة على محتوى الرياضيات للصفوف المستهدفة.

– درجة نجاح الباحث في بناء الأداة المستخدمة في هذه الدراسة.

تعريف المصطلحات

– التدريج العمودي (Vertical Scaling): طريقة إحصائية تستخدم للكشف عن النمو عبر مستويات دراسية مختلفة، وتتضمن إجراءات متعددة لوضع أداءات الأفراد على مقياس واحد مشترك، باستخدام اختبارات متعددة المستويات، (Kolen, 2011).

وتعرّف إجرائياً على أنها إجراءات رياضيّة، تم تطويرها لبناء المقياس العمودي، أو التطوري، ومنها ما ينبثق من النظرية الكلاسيكية، مثل: طريقة ثيرستون، وطريقة هيرونيموس، ومنها طرق حديثة باستخدام نماذج نظرية الاستجابة للفقرة.

– اختبار الفقرات المشتركة (Common Item Test): اختبارات تقيس محتوى معين لعدة مستويات دراسية بحيث يكون هناك فقرات مشتركة بين كل صفين متجاورين، تكون على الأقل (20%) من عدد الفقرات الكلي (Kolen, 2011).

– حجم الأثر (Effect Size): الفرق أو المتوسط المعياري بين توزيعي درجات صفين دراسيين متجاورين (Yen, 1986).

– المسافة الأفقية (Horizontal Distance): إحدى المحكّات المستخدمة في مقارنة طرق التدريج، ويعرّف على أنه الفرق بين الدرجات التي تقابل نفس الرتب المئينية لتوزيعين مختلفين (Holland, 2002).

– الاختبار متعدد المستويات (Multi-Level Test): اختبار يتكون من سلسلة من الاختبارات، بحيث يصمم كل اختبار لمستوى من مستويات السمة التي يغطيها الاختبار، ويعد أفضل أسلوب لقياس القدرات النامية مع التقدم في العمر أو المستوى الصفي، مثل: القدرة العقلية أو القدرة الرياضية او اللغوية (Thorndike, 1982).

– النمو الأكاديمي (Academic Growth): التغيّر في تحصيل الطلبة بين نقطتين زمنيتين أو أكثر، ويتضمن التقدم نحو تحقيق الأهداف الأكاديمية لبرنامج تعليمي معين (Jiao & Wang, 2007).

الدراسات ذات الصلة

تم الاطلاع على العديد من الدراسات التي تناولت مقارنة طرق التدريج العمودي، وكان هناك العديد من الدراسات والتي تتفاوت في طبيعة المتغيرات المتضمنة والمرتبطة بالتدريج العمودي، وفيما يلي عرض لهذه الدراسات التي تناولت متغيرات شبيهة بتلك التي تم تناولها في هذه الدراسة:

أجرى أساوكون (Asawakun, 1987) دراسة هدفت إلى تحري أثر تصميم جمع البيانات، وطريقة التدريج، ونوع المحتوى على التدريج العمودي لاختبارات التحصيل، واستخدم الباحث النموذج اللوجستي ثلاثي المعلمة المعدل، وطريقة هيرونيموس في التدريج العمودي، باستخدام تصميم الفقرات المشتركة وتصميم اختبار التدريج، وقد أشارت نتائج الدراسة المتعلقة بمعدل النمو إلى أن أكثر النتائج تشابهاً لاختبار اللغة كانت بين طريقة النموذج الثلاثي تحت تصميم الفقرات المشتركة، وطريقة هيرونيموس تحت تصميم اختبار التدريج وتصميم الفقرات المشتركة، أما في محتوى الرياضيات فكانت النتائج –تقريباً– متشابهة بجميع الطرق، والتصاميم، وفيما يتعلق بنمط التغيّر فقد أعطى النموذج الثلاثي المعدل تغيّراً متناقصاً في النمو، في حين كان التغيّر لذوي التحصيل المرتفع أكثر من التغيّر لذوي التحصيل المنخفض في طريقة هيرونيموس.

في حين أجرى بيكر وفورسث (Becker & Forsyth, 1992) دراسة هدفت إلى التحقق التجريبي من طرق ثيرستون، ونماذج نظرية الاستجابة للفقرة الأحادي، والثلاثي في تدريج اختبارات التحصيل، وأشارت نتائج هذه الدراسة إلى تزايد متوسط النمو الأكاديمي مع تقدم الصف الدراسي في جميع الطرق، وأن أكبر تزايد للمتوسط الحسابي عبر الصفوف كان عند المئين (90)، أي الطلبة ذوي التحصيل المرتفع أكثر نمواً من ذوي التحصيل المنخفض، وكذلك كان نمط التغيّر متزايداً مع تقدم الصف الدراسي بجميع الطرق أيضاً.

وقام كليمانز (Clemans, 1993) بدراسة هدفت إلى المقارنة بين طريقة ثيرستون والنموذج ثلاثي المعلمة باستخدام المعايرة المنفصلة ومن خلال طريقة الربط (Stocking & Lord)، وذلك من خلال استجابات افتراضية لـ (16000) طالب وطالبة من الصفوف من الأول الأساسي إلى الصف الثاني الثانوي بحيث تحاكي بيانات اختبار ولاية كاليفورونيا في القراءة، وأشارت النتائج المتعلقة بمعدل النمو إلى أن كلا الطريقتين أظهرت تزايداً في معدل النمو مع تقدم الصف الدراسي إلا أن طريقة ثيرستون أظهرت نمواً متزايداً أكثر من النموذج ثلاثي المعلمة الذي كان يميل النمو الأكاديمي فيها إلى الثبات، أما فيما يتعلق بالتغيّر في النمو، فقد كان متزايداً في طريقة ثيرستون للصفوف الدنيا، ثم اتجه إلى الثبات في الصفوف العليا، وفي النموذج ثلاثي المعلمة كان هناك تغير ثابت خلال الصفوف الدراسية المختلفة.

وأجرى ين وبوركت (Yen & Burket, 1997) دراسة هدفت إلى تحري الفروق بين طريقة ثيرستون، وطريقة النموذج اللوجستي ثلاثي المعلمة، باستخدام بيانات افتراضية ملائمة للنموذج ثلاثي المعلمة، حيث أشارت النتائج إلى أن معدل النمو كان متزايداً عبر الصفوف الدراسية، وذلك باستخدام الطريقتين ولكلا الاختبارين، إلا أن معدل النمو لاختبار القراءة كان متزايداً أكثر منه لاختبار الرياضيات، وكان للنموذج الثلاثي أكثر منه لطريقة ثيرستون، اما فيما يخص نمط التغيّر بدلالة الانحراف المعياري، فقد أظهرت كلا الطريقتين تزايد في الانحراف المعياري، مع تقدم الصف الدراسي في اختبار الرياضيات، في حين كان هناك تناقص في الانحراف المعياري، لاختبار القراءة مع تقدم الصف الدراسي.

وأجرى وليامز وبومرش وثيسين (Williams, Pommerich & Thissen, 1998) دراسة هدفت إلى مقارنة المقاييس التطورية المبنية استناداً إلى طريقة ثيرستون وطرق نظرية الاستجابة للفقرة، وكانت الطرق المستخدمة هي طريقة النموذج ثلاثي المعلمة وثلاث طرق لثيرستون هي (ثيرستون 1925، ثيرستون1938 لكل البيانات، ثيرستون 1938 لبعض البيانات)، وأشارت النتائج إلى تزايد المتوسط الحسابي مع تقدم الصف الدراسي لجميع الطرق المستخدمة، إلا أن طريقتي نظرية الاستجابة للفقرة أعطت معدلاً أعلى للنمو، تليها طريقة ثيرستون 1938 لبعض البيانات، في حين أظهرت طريقة ثيرستون 1925، وطريقة ثيرستون 1938 لكل البيانات معدلاً أقل للنمو، وفيما يخص نمط التغيّر بدلالة الانحراف المعياري، فقد كان – تقريباً– ثابتاً ومتشابهاً لطريقتي نظرية الاستجابة للفقرة، في حين كان متناقصاً في طريقة ثيرستون 1925، ومتزايداً في طريقة ثيرستون 1938 لكل البيانات، أما طريقة ثيرستون 1938 لجزء من البيانات فاظهرت تزايداُ قليلاً وثابتاً نسبياً، ومن ثم أظهرت النتائج بشكل عام أن طريقتي نظرية الاستجابة للفقرة، وطريقة ثيرستون 1938 لجزء من البيانات لهما –تقريباً– نتائج متشابهة.

وفي دراسة لهندرسون وكولين وتونغ (Hendrickson, Kolen & Tong 2004) هدفت للمقارنة بين تصاميم جمع البيانات، وطرق تقدير القدرة، ونوع البرنامج الحاسوبي المستخدم في التحليل، على نتائج التدريج العمودي باستخدام نماذج نظرية الاستجابة للفقرة، حيث تم استخدام طريقتين لتقدير القدرة هما طريقة الأرجحية العظمى (MLE)، وطريقة توقّع التوزيع البعدي (EAP)، وذلك باستخدام النموذج اللوجستي ثلاثي المعلمة، وأشارت النتائج المتعلقة بطرق التقدير إلى أن أنماط النمو والتغيّر متباينة عبر طرق تقدير القدرة، حيث تزايد متوسط النمو الأكاديمي في كلا طريقتي التقدير ، في حين أظهرت طريقة (EAP) تغيّر أقل من طريقة (MLE) حسب معيار التباين، من خلال تناقص التغيّر في النمو عبر الصفوف لدى الطلبة من مستوى القدرة نفسه، وفيما يخص حجم الأثر فقد كان متناقصاً مع تقدم الصف الدراسي، وكان اتجهاه لطرق التقدير عكس اتجاه نمط التغيّر بدلالة التباين.

وأجرى تونغ (Tong, 2005) دراسة هدفت إلى مقارنة المنهجيات والنتائج في التدريج العمودي لاختبارات التحصيل التربوية، وتم استخدم طريقة ثيرستون تحت تصميم اختبار التدريج، وطريقة النموذج اللوجستي ثلاثي المعلمة تحت تصميم اختبار التدريج، وتصميم الفقرات المشتركة، وفي النموذج ثلاثي المعلمة تم استخدام طريقتين لإعطاء الدرجات، هما الدرجة الكلية [Summed Scoring (SC)] من خلال طريقة (EAP) في تقدير القدرة، وطريقة نمط الإجابة [Pattern Scoring (PC)] من خلال ثلاث طرق لتقدير القدرة هي (EAP, MLE, MAP)، وأشارت نتائج معدل النمو إلى زيادة الوسط الحسابي مع تقدم الصف بجميع الطرق، وكان لطريقة ثيرستون وطريقة (IRT-SC) نتائج متماثلة في معدل النمو وبشكلٍ عام كانت أعلى من طريقة (IRT-PS)، وكان لجميع طرق التقدير نتائج متماثلة، وكان لنمط التغيّر اتجاهاً متزايداً بطريقة ثيرستون،

وثابتاً بطريقة (IRT-SC) ومتناقصاً بحدة في طريقة (IRT-PS)، وبالنسبة لطرق التقدير، كانت قيم الانحراف المعياري تأخذ الاتجاه التالي (MAP < EAP-SC < EAP < MLE)، أما حجم الأثر فكان متناقصاً في جميع الطرق، وبشكلٍ عام أعطت طرق (IRT) حجم أثر أعلى من طريقة ثيرستون وبخاصةً في الصفوف العليا، وكان اتجاه حجم الأثر بطرق التقدير عكس اتجاه نمط التغيّر، وفيما يتعلق بالمسافة الأفقية فكانت متزايدة من المئينات المنخفضة إلى الأعلى في طريقة ثيرستون، وفي طرق (IRT) كانت تتناقص في الصفوف الدنيا وترتفع قليلاً في العليا، ولم تختلف قيم المسافة الأفقية عبر طرق التقدير المختلفة.

وأجرى مينغ وكولين ولومان (Meng, Kolen & Lohman, 2006) دراسة هدفت إلى معرفة اختلاف نتائج التدريج العمودي باختلاف نماذج نظرية الاستجابة للفقرة، وطرق المعايرة، وطرق تقدير القدرة والمحتوى، حيث تم استخدام طريقتين لتقدير القدرة (MLE ,EAP)، من خلال النموذجين اللوجستيين الثلاثي والأحادي المعلمة، وأشارت النتائج المتعلقة بالمتوسط الحسابي، وحجم الأثر إلى أن نتائجها كانت متشابهة لطرق المعايرة، إلا أن طريقة (EAP) أظهرت حجم أثر أكبر من طريقة (MLE)، وفيما يتعلق بنمط التغيّر أظهرت طريقة (EAP) تغيراً أكبر ضمن المستوى، وبخصوص المسافة الأفقية بين المئينات المختلفة فقد كانت متماثلة لطريقة تقدير القدرة.

وفي دراسة قامت بها كيم (Kim, 2007) هدفت إلى مقارنة طرق تقدير القدرة، وطرق المعايرة في بناء مقاييس (IRT) العمودية، حيث تم استخدام النموذج ثلاثي المعلمة وأجريت المقارنات بين طرق تقدير القدرة (MLE ,EAP) وطريقتي المعايرة المتزامنة والمنفصلة، وأشارت النتائج المتعلقة بمعدل النمو إلى زيادة معدل النمو بزيادة الصف الدراسي، وإلى أن طرق التقدير المختلفة كان لها نتائج متشابهة وفيما يتعلق بنمط التغيّر فكان بشكل عام في طريقة (MLE) أكبر من طريقة (EAP)، أما النتائج الخاصة بحجم الأثر، فقد أشارت إلى نقصان حجم الأثر مع تقدم الصف الدراسي في طريقتي التقدير، وأشارت نتائج المسافة الأفقية إلى نقصان متوسط هذه المسافة مع تقدم الصف الدراسي، وكانت نتائج متوسط المسافة متشابهة عبر طرق التقدير، وأعطت طريقة (MLE) أعلى متوسط مسافة أفقية، أما المسافة الأفقية عند نقاط معينة، فكانت المسافة الأفقية بين المئينات المختلفة متشابهة عبر طرق التقدير.

وأخيراً أجرى ثوماس (Thomas, 2008) دراسة للتحقق من أثر طريقة التدريج على المقياس العمودي، إذ تم استخدام طريقة هيرونيموس، والنموذج ثلاثي المعلمة، وأشارت النتائج إلى أن معدل النمو في طريقتي التدريج يتزايد بتقدم الصف الدراسي، وتقل هذه الزيادة بشكل غير منتظم مع زيادة الصف الدراسي، وفيما يتعلق بنمط التغيّر أفادت النتائج بأنه لا يوجد نمط ثابت من التغيّر سواءً بالزيادة أم النقصان مع زيادة الصف الدراسي في طريقة هيرنيموس، ونمط متناقص من التغيّر في طريقة النموذج ثلاثي المعلمة، وفيما يتعلق بحجم الأثر، فكان يتناقص بشكل ثابت مع تقدم الصف في النموذج ثلاثي المعلمة، في حين أظهرت طريقة هيرنيموس حالات من عدم الانتظام في قيم حجم الأثر، ولمحك المسافة الأفقية، أشارت النتائج أنه لا يوجد نمط ثابت من المسافة الأفقية حيث كان يقل ويزداد بزيادة الصف، ومن خلال نسبة النمو أظهر النموذج ثلاثي المعلمة أن نمو الطلبة ذوي التحصيل المنخفض أعلى من ذوي التحصيل المرتفع، أما طريقة هيرونيموس، فكانت نسبة معدل النمو متساوية بين الطلبة من التحصيل المرتفع، والمنخفض.

التعقيب على الدراسات ذات الصلة

من خلال استعراض الدراسات ذات الصلة يتضّح بأنه لا يوجد اتفاق في الدراسات ذات الصلة حول أي طرق التدريج الكلاسيكية والحديثة أكثر فاعلية، وتعطي أنماط واضحة للنمو، ومن ثم تفسيرات منطقية لطبيعة النمو الأكاديمي لدى المتعلمين، وكذلك بعدم وجود دراسات تناولت النموذج ثنائي المعلمة في بناء التدريج العمودي، وأن هناك دراسات قليلة جداً حاولت الجمع بين أكثر من طريقتين في التدريج العمودي، وفي هذه الدراسة سيتم الجمع بين الطرق الكلاسيكية ونموذجين من نماذج نظرية الاستجابة للفقرة أحادية البعد، وكذلك دراسات قليلة تطرقت إلى مقارنة طرق التقدير في نماذج نظرية الاستجابة للفقرة، وحتى الدراسات التي استخدمت هذه الطرق تناولت نموذج واحد فقط من نماذج نظرية الاستجابة للفقرة، وأن الدراسات الحديثة فقط، هي التي تناولت المسافة الأفقية كمحك لمقارنة طرق التدريج العمودي، وبناءً على ما تم استعراضه من الدراسات السابقة، فإن هذه الدراسة تحاول الجمع بين متغيرات لم يتم التطرق لها سابقاً متمثلة بالطرق الاحصائية، ومحكات المقارنة.

أفراد الدراسة

لأغراض هذه الدراسة تم اختيار عينة من الطلبة بالطريقة العشوائية العنقودية الطبقية على مستوى المدارس، والشعب من مديريتي التربية والتعليم لمنطقة الطفيلة، ولواء بصيرا، وكان عدد أفراد الدراسة (1630) طالباً وطالبة من الصفوف الرابع، الخامس، والسادس الأساسي، للعام الدراسي 2013–2014 والجدول (1) يوضح عدد أفراد الدراسة، وتوزيعهم على الصفوف والجنس.

الجدول (1) عدد أفراد عينة الدراسة وجنسهم

الصف	ذكور	إناث	المجموع
الرابع	246	286	532
الخامس	254	308	562
السادس	255	281	536
المجموع	755	875	1630

أداة الدراسة

لأغراض هذه الدراسة، قام الباحث ببناء اختبار تحصيلي في مبحث الرياضيات متعدد المستويات معياري المرجع للصفوف الرابع، الخامس والسادس الأساسي في الأردن ثنائي التدريج، يقيس قدرات رياضية محددة، وهي: الكسور العادية والكسور العشرية، وذلك باستخدام تصميم الفقرات المشتركة، وذلك على وفق خطوات بناء الاختبار التحصيلي العلمية، وقواعد كتابة الفقرات الصحيحية، حيث كان العدد الأولي للفقرات يمثل (1.5) ضعف العدد النهائي المطلوب لكل صف، لضمان كفاية عدد الفقرات إذا تم حذف، أو دمج الفقرات اعتماداً على التحكيم، أو المحكات الإحصائية للفقرات.

وتم عرض الصورة الأولية من الفقرات على محكمين متخصصين في مبحث الرياضيات (معلّمي الصفوف المستهدفة، ومشرفي المبحث)، للتحقق من الصدق الظاهري لصور الاختبار، وقد طلب منهم إبداء الرأي في مدى وضوح الفقرات وارتباطها مع الأهداف التي تنطوي تحتها، ومدى ملائمتها للفئة العمرية المستهدفة، كل منها على مقياس رباعي التدريج، وكانت الخطوة التالية هي التجريب الأولي للفقرات، على عينة تجريبية بلغت (128) طالباً وطالبةً من الصفوف الثلاثة، من خارج عينة الدراسة ومن داخل مجتمع الدراسة، وذلك للتحقق من وضوح الفقرات والتعليمات، وتحديد الزمن اللازم للإجابة، واشتقاق خصائص أولية للفقرات، حيث استغرق زمن الإجابة للصفوف الثلاثة بين (32 –43) دقيقة، أي ما يعادل حصة دراسية واحدة، ويوضّح الجدول (2) مدى صعوبة وتمييز الفقرات ضمن العينة التجريبية الأولى.

الجدول (2) مدى صعوبة وتمييز الفقرات ضمن العينة التجريبية الأولى

الصف	الصعوبة	التمييز	حجم العينة
الرابع الأساسي	0.87 – 0.21	0.73 – 0.23	44
الخامس الأساسي	0.83 – 0.26	0.76 – 0.18	41
السادس الأساسي	0.89 – 0.22	0.69 – 0.24	43

واعتماداً على هذه النتائج – بالإضافة إلى نتائج التحكيم – تم اختيار فقرات الصورة النهائية للاختبار، وذلك من خلال اختيار فقرات ذات صعوبة ضمن المدى (0.30 – 0.70)، على أن يكون لها معاملات تمييز مرتفعة، والملحق (1) يبين الصورة النهائية للاختبار، حيث بلغ عدد الفقرات النهائي (46) فقرة لجميع المستويات ثنائية التدريج رباعية البدائل، منها (12) فقرة مشتركة بين كل صفين (مستويين) متجاورين، حيث تم وضع الفقرات المشتركة في نهاية المستوى الأدنى، وبداية المستوى الأعلى، وعليه تم تطبيق الصورة النهائية على عينة تجريبية ثانية مكونة من (484) طالباً وطالبة من طلبة الصفوف المستهدفة، لاستخراج خصائص الفقرات والاختبار كما هو آتٍ:

صدق الاختبار (Test validation)

1– صدق البناء (Construction validation)

للتحقق من صدق البناء، تمّ استخدام طريقة الصدق العاملي للاختبارات، من خلال إجراء التحليل العاملي، بطريقة المكونات الرئيسة، لاختبارات المستويات الثلاثة، وذلك لبحث العوامل المسؤولة عن الأداء في كل مستوى من مستويات الاختبارات الثلاثة، والتأكد من أن الأداء على الاختبار يعود لعامل القدرة الرياضيّة، ويبيّن الجدول (3) قيم الجذور الكامنة (Eigen Value) ونسبة التباين المفسر، ونسبة التباين التراكمي للعاملين الأول والثاني، من كل اختبار على جميع الفقرات.

جدول (3) قيم الجذور الكامنة ونسبة التباين المفسر للعوامل في الصفوف الثلاثة

الصف	العامل	الجذر الكامن	نسبة التباين المفسر%	نسبة التباين المفسر التراكمي%
الرابع	1	3.446	17.232	17.232
	2	1.254	6.270	23.503
الخامس	1	4.117	17.154	17.154
	2	1.208	5.035	22.189
السادس	1	4.324	16.630	16.630
	2	1.328	5.106	21.736

ويتضح من خلال الجدول، أنه يوجد فرق في نسبة التباين المفسر، وقيم الجذر الكامن بين العامل الأول، والعامل الذي يليه في كل اختبار وقيمة هذا الفرق أكثر من ثلاثة اضعاف، مما يدل على وجود عامل واحد مسيطر، على أداء الأفراد على الاختبارات، وهو عامل القدرة الرياضية.

2- الصدق المرتبط بالمحك (Criterion-related validation)

للتحقق من صدق الاختبار المرتبط بالمحك، تم إيجاد معامل ارتباط (بيرسون) بين الدرجات التي حصل عليها الطلبة على الاختبار، ودرجات الرياضيات المدرسية، حيث بلغت قيم الصدق المرتبط بالمحك (0.84, 0.86, 0.87) للصفوف الثلاثة على التوالي، وهي قيم موجبة ومرتفعة، ومن ثم فإن للاختبارات الثلاث صدق محكي مرتفع.

ثبات الاختبار (Test Reliability)

بهدف التحقق من ثبات الاختبار للصفوف الثلاثة، تم استخدام الطرق الآتية:

1- ثبات الإتساق الداخلي: حيث حساب معاملات الثبات باستخدام معادلة (كيودر- ريتشادسون) (KR20)، وكذلك باستخدام معادلة التجزئة النصفية.

2- ثبات الاستقرار: من خلال إعادة تطبيق الاختبار بمستوياته الثلاثة على أفراد العينة التجريبية الثانية، وذلك بعد فاصل زمني مقداره عشرة أيام، وحساب معامل ارتباط بيرسون بين الدرجات الكلية للتطبيق الأول، والتطبيق الثاني، ولجميع الصفوف.

3- الثبات في ضوء نظرية الاستجابة للفقرة: حيث تم استخدام استجابات أفراد العينة التجريبية من الصفوف الثلاثة، وباستخدام برنامج (Bilog-MG) تم تقدير معالم الفقرات، ومعلمة القدرة حسب النموذج اللوجستي ثلاثي المعلمة، ومن ثم تم حساب مؤشرات ثبات تباعد الأفراد، وثبات تباعد الفقرات، ويوضح الجدول (4) قيم معاملات الثبات التي تم اشتقاقها للاختبارات الثلاثة وبجميع الطرق.

جدول (4) قيم معاملات ثبات الاختبارات بجميع الطرق

الصف	كيودر- ريتشادسون	التجزئة النصفية	الاستقرار	تباعد الأفراد	تباعد الفقرات
الرابع	0.85	40.8	0.84	0.79	60.8
الخامس	0.82	0.81	0.88	0.77	40.8
السادس	0.86	80.8	0.91	0.81	30.8

ويتضح من خلال الجدول أن جميع قيم معاملات الثبات تراوحت بين (0.77-0.91)، ومن ثم فان الاختبار بمستوياته الثلاث، يتمتع بمقدار جيد من الثبات.

خصائص الفقرات

للتحقق من خصائص الفقرات، تم استخدام برنامج (Bilog-MG) لحساب خصائص الفقرات في ضوء كل من النظرية الكلاسيكية، ونظرية الاستجابة للفقرة، حيث تراوحت معاملات الصعوبة الكلاسيكية بين (0.28-0.77) لجميع الصفوف، وهذا يتفق مع تصميم الاختبار كونه معياري المرجع، وكان متوسط قيم الصعوبة قريباً من (0.50) لجميع الصفوف، وأن قيم الصعوبة كانت متزايدة مع التقدم في الصف الدراسي، وهو الأمر المرغوب في تصميم اختبارات التدريج العمودي، ويلاحظ أيضاً أن الفقرات المشتركة بين الصفوف المتجاورة، كانت أسهل للصف الأعلى، وأصعب للصف الأدنى، أما فيما يخص معاملات تمييز الفقرات الكلاسيكية فكانت لجميع الاختبارات ضمن مدى متقارب حيث تراوحت بين (0.38-0.67) وهي معاملات تمييز مرتفعة حسب ما أشار إليه ايبل (Ebel, 1973).

وفيما يتعلق بخصائص الفقرات وفق نظرية الاستجابة للفقرة، بلغ متوسط معلمة تمييز الفقرات للصفوف الرابع والخامس والسادس (1.34، 1.35، 1.36)، على التوالي، وهي معاملات تمييز مرتفعة، وفيما يخص معلمة صعوبة الفقرات، بلغت متوسطاتها أيضاً للصفوف الثلاثة (0.75، 0.62، 0.67) على الترتيب وهي معلمة صعوبة متوسطة، وهذا يتفق مع تصميم الاختبار، أما معلمة التخمين، فكانت متوسطاتها للصفوف الثلاثة على الترتيب هي (0.24، 0.22، 0.22)، وهي متقاربة، ومتدنية نسبياً.

اجراءات الدراسة

بهدف التوصل إلى نتائج الدراسة تم تطبيق الاختبار بصورته النهائية على عينة التطبيق الواردة في الجدول (1)، ثم تم تصحيح أوراق الاختبار من خلال أعطاء الدرجة واحد (1) للإجابة الصحيحة، والدرجة صفر (0) للإجابة الخاطئة، كما وتم إدخال نتائج التصحيح على برنامج (Ecxel) حتى يصار إلى

نقلها لبرامج التحليل الإحصائي المستخدمة في الدراسة وهي: (SPSS, Bilog-MG, LazStsts, CIPE, IRTPRO)، بهدف تطبيق خطوات التوصل إلى المقياس المشترك، باعتبار الصف الرابع صفاً مرجعياً، وفيما يأتي تفصيل خطوات التدريج العمودي لكل طريقة من هذه الطرق:

طريقة ثيرستون

تم التوصل إلى المقياس المشترك بطريقة ثيرستون، وبتصميم الفقرات المشتركة من خلال استخدام برمجية (SPSS) وضمن الخطوات التالية:

1. الحصول على توزيع الدرجات الخام لجميع الصفوف على الاختبار ككل، وكذلك على الفقرات المشتركة.
2. إيجاد الرتب المئينية المقابلة للدرجات الخام على الفقرات المشتركة، ثم إيجاد الدرجات المعيارية (Normalized Z Scores) المقابلة للرتب المئينية.
3. استخدام المعادلتين (1،2) بهدف ربط الصفوف مع بعضها، مع مراعاة ان يكون الصف الأدنى في كل عملية ربط من ربط السلسلة بوسط حسابي صفر وانحراف معياري واحد صحيح.

$$\sigma_2(\mathrm{SC}) = \frac{\sigma[\mathrm{Z}_1^*(\mathrm{y})]}{\sigma[\mathrm{Z}_2^*(\mathrm{y})]}\ \sigma_1(\mathrm{SC}) \dots\dots\dots [1]$$

$$\mu_2(\mathrm{SC}) = \sigma_1(\mathrm{SC})\left[\mu\left(\sigma \mathrm{Z}_1^*(\mathrm{y})\right) - \frac{\sigma[\mathrm{Z}_1^*(\mathrm{y})]}{\sigma[\mathrm{Z}_2^*(\mathrm{y})]}\ \mu[\mathrm{Z}_2^*(\mathrm{y})]\right] + \mu_1(\mathrm{SC}) \dots\dots [2]$$

حيث
$\mu[z_1^*(y)]$ المتوسط الحسابي للدرجات المعيارية في المجموعة الأولى
$\mu[z_2^*(y)]$ المتوسط الحسابي للدرجات المعيارية في المجموعة الثانية
$\sigma[z_1^*(y)]$ الانحراف المعياري للدرجات المعيارية في المجموعة الأولى
$\sigma[z_2^*(y)]$ الانحراف المعياري للدرجات المعيارية في المجموعة الثانية

4. الوصول للمقياس المشترك بحيث يكون للصف الأدنى (الرابع) بوسط حسابي صفر وانحراف معياري واحد، وللصف الخامس بوسط حسابي ناتج من مجموع الوسط الحسابي للصفين الرابع، والخامس في الخطوة (3)، وكذلك الانحراف المعياري ناتج من حاصل ضرب الانحراف المعياري للصفين الرابع والخامس، ويكون الوسط الحسابي للصف السادس هو مجموع الوسط الحسابي للصفوف الرابع والخامس والسادس من الخطوة (3)، والانحراف المعياري هو ناتج حاصل ضرب الانحراف المعياري للصفوف الثلاث الناتجة من الخطوة (4).
5. استخدام المعادلة (3) بهدف إعادة المقياس (Rescale) بوسط حسابي (200) وانحراف معياري (15).

$$sc = z_1^*(y)\sigma_1(\mathrm{SC}) + \mu_1(\mathrm{SC}) \dots\dots\dots [3]$$

6. إعادة الخطوات (1−2) على كامل الاختبار لكل صف، والذي يعدُ بمثابة اختبار مستوى خاص بكل صف دراسي بهدف الحصول على المقياس النهائي.

طريقة هيرونيموس

في هذه الطريقة تم القيام بالتدريج العمودي، والتوصل للمقياس المشترك تحت تصميم الفقرات المشتركة من خلال الخطوات التالية:

1. الحصول على توزيع الدرجات الخام لجميع الصفوف على الاختبار ككل، وكذلك على الفقرات المشتركة.
2. إيجاد التوزيع التكراري، والتكراري النسبي والنسبي التراكمي للدرجات الخام على الفقرات المشتركة، وعلى الاختبار ككل.
3. عمل تمهيد (Smoothing) للتوزيع التكراري للدرجات الكلية والدرجات على الفقرات المشتركة، وذلك باستخدام برنامج [Common Item Percentile Equating, (CIPE)] والذي تم

تطويره من قبل كولين (Kolen, 2004).

4. القيام بعملية ربط كل صفين متجاورين بطريقة ربط المئينات المتساوية (Eqiupercentile Linking)، وذلك باستخدام نفس البرنامج الوارد في الخطوة (3)، وهو معد لغاية الربط، والمعادلة تحت تصميم الفقرات المشتركة، حيث يتم ربط الصف الخامس إلى الصف الرابع، والصف السادس إلى الصف الخامس.

5. الحصول على المتوسطات الحسابية والانحرافات المعيارية والمتوسطات للتوزيعات الجديدة للدرجات بعد عملية الربط.

6. استخدام المعادلة (4) لربط الصفوف جميعاً مع بعضها، وذلك من خلال تثبيت وسط الصف الرابع (المرجعي) عند (200)، والصف السادس عند (225).

$$SC(\hat{\tau}) = \left[\frac{md_h - md_l}{Mdn_{gh} - Mdn_{gl}}\right]\hat{\tau} + \left(md_l - \left[\frac{md_h - md_l}{Mdn_{gh} - Mdn_{gl}}\right]\right) Mdn_{gl} \ldots\ldots [4]$$

حيث $\hat{\tau}$ الدرجات الحقيقية

$SC(\hat{\tau})$ درجات المقياس للدرجات الحقيقية

Mdn_{gh} وسيط درجات الصف الأعلى

Mdn_{gl} وسيط درجات الصف الأدنى

md_h قيمة الوسيط المراد تثبيت الأعلى عندها

md_l قيمة الوسيط المراد تثبيت الصف عندها

طريقة نماذج نظرية الاستجابة للفقرة

بهدف الحصول على المقياس المشترك بواسطة التدريج العمودي، وباستخدام نماذج نظرية الاستجابة للفقرة تم اتباع الخطوات التالية:

1- التحقق من افتراضات نظرية الاستجابة للفقرة، وهي افتراض أحادية البعد من خلال إجراء التحليل العاملي الوارد في باب صدق الاختبار، وافتراض الاستقلال الموضعي باستخدام برمجية (IRTPRO) الذي أفادت نتائجه أن جميع الفقرات تتمتع بافتراض الاستقلال الموضعي، إذ تراوحت هذه القيم ضمن المدى (3.9-0.7)، وقد أشار شن وثيسن (Chen & Thissen, 1997) إلى أن هذه القيم إذا كانت ضمن المدى (−2,+2) أو (−3,+3) فان افتراض الاستقلال الموضعي يتحقق بشكل كبير وموثوق، والقيم ضمن المدى (5−10) فهي مقبولة إلى حدٍ ما، أما القيم التي تزيد عن (10) فتدل على عدم تحقق افتراض الاستقلال الموضعي.

2- ادخال بيانات الاستجابة على شكل مصفوفة غير كاملة بهدف القيام بإجراءات التدريج بطريقة المعايرة المتزامنة على برنامج (Bilog-MG).

3- التحقق من مدى مطابقة بيانات الاختبار للنموذجين الثنائي، والثلاثي، حيث أشارت نتائج المطابقة إلى أن جميع فقرات الاختبار، والبالغ عددها (46) فقرة، كانت مطابقة للنموذج اللوجستي ثلاثي المعلمة باستثناء فقرة واحدة هي الفقرة (22)، وكذلك للنموذج ثنائي المعلمة باستثناء فقرتين هما (34,43)، حيث كان لهذه الفقرات الثلاث قيم احتمالية أقل من مستوى الدلالة (P <0.05).

4- بهدف المقارنة مع نتائج ثيرستون، تم عمل المعايرة المتزامنة، ثم عمل إعادة للمقياس بحيث يكون وسط الصف الرابع (المرجعي) هو (0) وبانحراف معياري (1)، ثم يتم أيضاً عمل (Rescaling) للمقياس بوسط حسابي (200)، وانحراف (15)، للنموذجين الثنائي والثلاثي تحت توزيعي (MLE, EAP).

5- بهدف المقارنة مع طريقة هيرونيموس، يتم استخدام المعادلة (4) على قيم القدرة (θ)، دون عمل (Rescaling)، ويتم أيضاً تثبيت وسيط الصف الرابع عند (200)، والسادس عند (225)، وذلك للنموذجين الثنائي والثلاثي تحت توزيعي (MLE, EAP).

وأخيراً تم استخراج المحكّات المستخدمة في مقارنة طرق التدريج العمودي وهي:

- المتوسط الحسابي للدلالة على معدل النمو.
- الانحراف المعياري ونصف المدى الربيعي للدلالة على نمط التغيّر ضمن الصف.
- حجم الأثر: من خلال المعادلة التالية:

$$E.F = \frac{\bar{X}_u - \bar{X}_l}{\sqrt{\frac{\sigma_u^2 + \sigma_l^2}{2}}} \ldots\ldots\ldots [5]$$

حيث $\bar{X}_u$ ، σ_u^2 متوسط وتباين الصف الأعلى

$\bar{X}_l$ ، σ_l^2 متوسط وتباين الصف الأدنى

- المسافة الأفقية: من خلال حساب المئينات (10,25,50,75,90)، ثم تطبيق المعادلات (6،7،8) لحساب المسافة الأفقية، ومتوسط المسافة الأفقية، ونسبة النمو.

$$HD(f) = Y(p) - X(p) \dots\dots\dots\dots\dots [6]$$

حيث Y(P) مئين معين لصف معين، X(P) نفس المئين لصف آخر.

$$\overline{HD} = \frac{\sum_{p=1}^{99} HD_P}{99} \dots\dots\dots\dots\dots [7]$$

$$Ratio\ of\ Growth_{75,25} = \frac{\sum HD_{P75}}{\sum HD_{P25}} \dots\dots\dots\dots\dots [8]$$

وتكون لهذه النسبة ثلاث قيم هي :

R.G < 1	نسبة نمو الطلبة ذوي التحصيل المنخفض أكبر منها للطلبة ذوي التحصيل المرتفع
R.G > 1	نسبة نمو الطلبة ذوي التحصيل المرتفع أكبر منها للطلبة ذوي التحصيل المنخفض
R.G = 1	نسبة نمو الطلبة ذوي التحصيل المنخفض نفسها للطلبة ذوي التحصيل المرتفع.

النتائج

بداية تم حساب مجموعة من الإحصاءات الوصفية لأداء أفراد العينة على الاختبار متعدد المستويات، ويوضح الجدول (5) الإحصاءات الوصفية لأداء الطلبة على الاختبار من الصفوف الثلاثة باستخدام الدرجات الخام الكلية، في حين يبين الجدول (6) الأداء على الفقرات المشتركة بين الصفوف المتجاورة، البالغ عددها (12) فقرة، بين كل صفين متجاورين.

جدول (5) الإحصاءات الوصفية للدرجات الكلية الخام على الاختبار من الصفوف الثلاثة

الإحصائي / الصف	المتوسط الحسابي	الانحراف المعياري	الالتواء	التفلطح	عدد الفقرات	حجم العينة
الصف الرابع	8.63	3.62	420.	410.	20	533
الصف الخامس	11.00	3.91	10.2	310.	24	563
الصف السادس	11.48	4.24	0.29	110.	26	537

ويلاحظ من خلال الجدول (5) زيادة متوسط الأداء والانحراف المعياري نظراً لزيادة عدد الفقرات بزيادة الصف الدراسي، أما بخصوص قيم الالتواء والتفلطح والتي تعطي فكرة عن طبيعة التوزيع للدرجات، كانت قيمها قريبة من الصفر، وبالتالي فإن توزيع الدرجات الخام الكلية يقترب من التوزيع الطبيعي للصفوف الثلاثة.

جدول (6) الإحصاءات الوصفية للدرجات الخام على الفقرات المشتركة

الفقرات / الإحصائي	الرابع/الخامس		الخامس/السادس	
	الرابع	الخامس	الخامس	السادس
المتوسط الحسابي	4.97	5.38	5.60	66.0
الانحراف المعياري	2.26	2.25	2.28	2.32

ويتضح من خلال الجدول (6) أن المتوسط الحسابي لأداء الطلبة على الفقرات المشتركة يزداد بزيادة الصف الدراسي، وذلك نظراً لزيادة القدرة.

أما بخصوص النتائج المتعلقة بطرق التدريج العمودي المستخدمة في هذه الدراسة، فقد تم التوصل إلى المقياس النهائي المشترك، من خلال جميع الطرق، ويوضّح الجدول (7) نتائج طرق التدريج العمودي، عند مقارنة طريقة ثيرستون، مع نماذج نظرية الاستجابة للفقرة، في حين يوضّح الجدول (8) نتائج طرق التدريج العمودي، عند مقارنة طريقة هيرونيموس، مع نماذج نظرية الاستجابة للفقرة.

جدول (7) نتائج مقياس ثيرستون مع مقاييس نماذج (IRT)

الإحصائي	الطريقة / الصف	ثيرستون	2PLM-EAP	2PLM-MLE	3PLM-EAP	3PLM-MLE
المتوسط الحسابي	الرابع الأساسي	200.03	200.00	200.62	200.00	199.24
	الخامس الأساسي	202.57	204.29	204.44	204.71	204.47
	السادس الأساسي	204.52	208.03	208.13	207.03	206.30
الانحراف المعياري	الرابع الأساسي	14.80	15.01	16.01	15.01	15.98
	الخامس الأساسي	14.69	14.17	14.16	14.02	12.49
	السادس الأساسي	15.28	15.18	15.86	14.14	12.87
نصف المدى الربيعي	الرابع الأساسي	10.84	9.24	8.58	9.80	10.21
	الخامس الأساسي	9.97	9.06	7.88	9.57	7.27
	السادس الأساسي	11.56	9.44	8.36	10.97	6.48
حجم الأثر	الرابع/الخامس	0.17	290.	0.25	0.32	0.36
	الخامس/السادس	0.13	0.25	0.24	0.17	0.14
	الكلي	0.30	0.54	0.49	0.49	0.50
متوسط المسافة الأفقية	الرابع/الخامس	2.55	4.30	3.83	4.74	5.21
	الخامس/السادس	1.95	3.75	3.67	2.31	1.83
نسبة النمو	جميع الصفوف	1.46	1.05	0.94	1.27	0.36

جدول (8) نتائج مقياس هيرونيموس مع مقاييس نماذج (IRT)

الإحصائي	الطريقة / الصف	هيرونيموس	2PLM-EAP	2PLM-MLE	3PLM-EAP	3PLM-MLE
المتوسط الحسابي	الرابع الأساسي	204.65	205.81	208.24	203.59	191.44
	الخامس الأساسي	221.47	217.47	220.38	218.88	208.43
	السادس الأساسي	228.11	227.62	232.01	226.43	214.40
الانحراف المعياري	الرابع الأساسي	27.74	40.77	50.77	48.76	51.95
	الخامس الأساسي	31.30	38.48	44.89	45.52	40.59
	السادس الأساسي	7631.	41.21	50.14	45.91	41.85
نصف المدى الربيعي	الرابع الأساسي	19.23	25.10	27.21	31.83	33.19
	الخامس الأساسي	19.84	24.60	24.95	31.08	23.65
	السادس الأساسي	22.19	25.64	26.49	35.62	21.07
حجم الأثر	الرابع/الخامس	0.57	290.	0.25	0.32	0.36
	الخامس/السادس	0.32	0.25	0.24	0.17	0.14
	الكلي	0.89	0.54	0.49	0.49	0.50
متوسط المسافة الأفقية	الرابع/الخامس	16.67	11.68	12.14	15.38	16.95
	الخامس/السادس	6.78	10.18	11.58	7.51	5.95
نسبة النمو	جميع الصفوف	1.33	1.05	0.94	1.27	0.36

ملخص نتائج معدل النمو (المتوسط الحسابي)

- تزايد المتوسط الحسابي للمقياس النهائي المشترك، بجميع الطرق مع تقدم الصف الدراسي، إلا ان هذه الزيادة كانت تتناقص في طريقة هيرونيموس والنموذج ثلاثي المعلمة، مع تقدم الصف الدراسي، في حين كانت تقريباً ثابتة في طريقتي ثيرستون والنموذج ثنائي المعلمة.
- لا يوجد فروق تذكر بين طريقتي تقدير القدرة EAP, (MLE) وللنموذجين الثنائي والثلاثي من حيث معدل النمو

بدلالة المتوسط الحسابي، إلا أن طريقة (MLE) كانت تعطي متوسط (معدل نمو) أعلى من طريقة (EAP) في النموذج ثلاثي المعلمة، أما في النموذج ثنائي المعلمة كانت الفروق بينهما غير ثابتة.

ملخص نتائج نمط التغيّر (الانحراف المعياري ونصف المدى الربيعي)

- كان نمط التغيّر في طريقة هيرونيموس متزايد مع التقدم في الصف الدراسي، من خلال معياري الانحراف المعياري ونصف المدى الربيعي، ومتناقص باستخدام طريقة النموذج اللوجستي ثلاثي المعلمة تحت طريقة (MLE) في التقدير، من خلال معيار نصف المدى الربيعي، وتقريباً متناقص باستخدام معيار الانحراف المعياري.
- تميل طريقة ثيرستون لإعطاء نمط تغير في النمو تقريباً ثابت باستخدام معيار الانحراف المعياري، ولا يوجد نمط محدد للتغير (متذبذب) من خلال معيار نصف المدى الربيعي، أما النموذج الثنائي تحت طريقة التقدير (EAP) أظهرت نمط تغير ثابت من خلال معيار نصف المدى الربيعي، ونمط متذبذب من خلال الانحراف المعياري، وكذلك النموذج ثنائي المعلمة وباستخدام طريقة (MLE) والنموذج ثلاثي المعلمة باستخدام طريقة (EAP) أظهرت تغير متذبذب في النمو الأكاديمي عبر الصفوف.
- وفيما يخص طرق التقدير وبشكل عام أعطت طريقة الأرجحية العظمى (MLE) تغيّراً أكبر من طريقة توقع التوزيع (EAP)، في النموذج ثنائي المعلمة، والعكس في النموذج ثلاثي المعلمة.

ملخص نتائج حجم الأثر

يتناقص مع تقدم الصف الدراسي باستخدام جميع الطرق، إلا إن طريقة هيرونيموس أظهرت أعلى حجم أثر وبالتالي أعلى نمواً، يليها النموذج ثلاثي المعلمة والنموذج ثنائي المعلمة، وأخيراً كانت طريقة ثيرستون الأقل نمواً، وفيما يتعلق بطرق تقدير القدرة فكانت نتائج حجم الأثر لها على عكس نتائج نمط التغيّر بدلالة الانحراف المعياري ونصف المدى الربيعي، حيث كانت قيمة حجم الأثر في طريقة الأرجحية العظمى (MLE) أكبر منه في طريقة توقع التوزيع (EAP) في النموذج ثلاثي المعلمة، والعكس في النموذج ثنائي المعلمة.

ملخص نتائج المسافة الأفقية

يلاحظ من النتائج التي توصلت إليها هذه الدراسة أن متوسط المسافة الأفقية يتناقص مع تقدم الصف الدراسي، أي أن نتائجه جاءت مشابهة لنتائج حجم الأثر، بالإضافة إلى أن الفروق بين طرق تقدير القدرة في النموذج الواحد كانت قليلة، أي إن نتائج طرق التقدير كانت متشابهة لنفس النموذج، وعموماً أظهر النموذج اللوجستي ثلاثي المعلمة تحت طريقة الأرجحية العظمى في التقدير أعلى متوسط للمسافة الأفقية بين الصفوف المتجاورة، وكذلك أظهرت طرق ثيرستون والنموذج الثنائي بكلا طريقتي التقدير تناقصاً قليلاً في متوسط المسافة الأفقية، أي إن المسافة تقريباً تبقى ثابتة مع تقدم الصف الدراسي في طريقتي ثيرستون والنموذج ثنائي المعلمة، في حين كان التناقص حاداً وملحوظاً في طريقة هيرونيموس والنموذج ثلاثي المعلمة.

أن الطرق الكلاسيكية (ثيرستون وهيرونيموس) والنموذج ثلاثي المعلمة باستخدام طريقة التقدير (EAP)، كان لها نسبة نمو أكبر من الواحد الصحيح، وبالتالي فإن نمو الطلبة ذوي التحصيل المرتفع أكبر من نمو الطلبة ذوي التحصيل المنخفض، وكانت هذه النسبة قريبة من الواحد الصحيح في النموذج ثنائي المعلمة بطريقتي تقدير القدرة، أي ان النمو متماثل للطلبة ذوي التحصيل المرتفع والمنخفض، وفي النموذج ثلاثي المعلمة باستخدام طريقة التقدير (MLE) كانت نسبة النمو أقل من الواحد الصحيح وعليه فإن نمو الطلبة ذوي التحصيل المنخفض أعلى منه لذوي التحصيل المرتفع.

مناقشة النتائج

أظهرت النتائج المتعلقة بمعدل النمو، تزايد المتوسط الحسابي للمقياس النهائي المشترك بجميع الطرق مع تقدم الصف الدراسي، مما يدل على تحقق النمو الأكاديمي من صف لآخر، وتعتبر هذه نتيجة منطقية وواقعية، نظراً لزيادة النضج المعرفي، واكتساب كم أكثر من المعرفة والمهارات مع تقدم الصف أو المستوى الدراسي، حيث اتفقت هذه النتيجة مع جميع الدراسات السابقة ذات الصلة.

وكان معدل النمو من الصف الرابع إلى السادس الأساسي متساوياً تقريباً بجميع الطرق عدا طريقة ثيرستون كان لها معدل نمو أقل من الطرق الأخرى، ويتفق ذلك مع ما جاء في نتائج دراسة كل من ين وبوركت (Yen & Burket, 1997) ودراسة وليامز وآخرون (Williams et. al, 1998)، ودراسة أساوكون (Asawakun, 1987)، التي أشارت نتائجها إلى أن طريقة هيرونيموس والنموذج ثلاثي المعلمة تعطيان معدل للنمو أكبر من طريقة ثيرستون، وقد يعود السبب في ذلك إلى بساطة وقلة عدد التحويلات التي تجرى على الدرجات الخام، ضمن طريقة ثيرستون، التي تعتمد على خطوات وإجراءات بسيطة في التوصّل إلى المقياس العمودي النهائي المشترك.

اختلفت الطرق في معدل النمو بين الصفوف المتجاورة، حيث

كان معدل النمو في طريقة هيرونيموس والنموذج الثلاثي يتناقص مع تقدم الصف الدراسي، في حين كانت نسبة التزايد ثابتة تقريباً في طريقتي ثيرستون والنموذج ثنائي المعلمة، وجاءت هذه النتيجة متفقة مع نتائج دراسة ثوماس (Thomas, 2008)، وبالتالي تكون طريقة ثيرستون، وطريقة النموذج ثنائي المعلمة، هما الأقرب إلى المقياس الأصلي، من خلال الدرجات الخام على الفقرات المشتركة، والتي كان فيها معدل النمو بدلالة المتوسط الحسابي، متزايداً عبر الصفوف، وكانت هذه الزيادة شبه ثابتة مع تقدم الصف الدراسي.

أظهرت طريقتي تقدير القدرة وهما طريقة الأرجحية العظمى (MLE) وطريقة توقع التوزيع البعدي (EAP)، نتائج متماثلة فيما يتعلق بمعدل النمو العام، وكذلك معدل النمو بين الصفوف المتجاورة، وجاءت هذه النتيجة متفقة مع دراسة كل من تونغ (Tong, 2005) وكيم (Kim, 2007)، والتي أشارت نتائجها إلى أن طرق التقدير تعطي نتائج متماثلة فيما يتعلق بمحك المتوسط الحسابي.

أما النتائج المتعلقة بنمط التغير، فأظهرت أنه كان متزايداً مع التقدم في الصف في طريقة هيرونيموس، واختلفت هذه النتيجة مع دراسة ثوماس (Thomas, 2008) التي أشارت لعدم وجود نمط ثابت من التغيّر في طريقة هيرونيموس عزاه الباحث إلى الفروق في حجم العينة في الصفوف المختلفة، وتميل طريقة ثيرستون لإعطاء نمط تغيّر في النمو ثابت تقريباً، وتتفق هذه النتيجة مع دراسة ين وبوركت (Yen & Burket, 1997) ودراسة وليامز وآخرون (Williams et. al, 1998)، في حين اختلفت مع دراسة بيكر وفورسث (Becker & Forsyth, 1992) ودراسة تونغ (Tong, 2005) التي أشارت إلى تزايد نمط التغيّر وفق طريقة ثيرستون، أما النموذج اللوجستي ثنائي المعلمة فقد أظهر نمطاً ثابتاً من التغيّر تحت طريقة التقدير (EAP)، في حين كان نمط التغيّر متذبذباً تحت طريقة (MLE) لنفس النموذج، علماً بأنه لا توجد دراسات تناولت هذا النموذج مما يحد من إمكانية مقارنة هذه النتيجة مع دراسات سابقة.

وأخيراً كان نمط التغيّر متناقصاً في النموذج اللوجستي ثلاثي المعلمة تحت طريقة (MLE) في التقدير، في حين أظهرت طريقة (EAP) لنفس النموذج تغيّراً متذبذباً في نمط النمو الأكاديمي عبر الصفوف، وتتفق هذه النتيجة مع دراسة تونغ (Tong, 2005) ودراسة ثوماس (Thomas, 2008) ودراسة كيم (Kim, 2007) ودراسة أساوكون (Asawakun, 1987) حيث أفادت جميعها بتناقص نمط التغيّر في النموذج الثلاثي ، في حين اختلفت مع نتائج ين وبوركت (Yen & Burket, 1997) ودراسة كليمانز (Clemans, 1993) ودراسة وليامز وآخرون (Williams et. al, 1998) التي أشارت جميعها إلى أن نمط التغيّر ثابتاً في النموذج ثلاثي المعلمة، وفيما يخص طرق التقدير وبشكل عام أظهرت طريقة الأرجحية العظمى (MLE) تغيّراً أكبر من طريقة توقع التوزيع (EAP) في النموذج ثنائي المعلمة، والعكس في النموذج ثلاثي المعلمة، وجاءت هذه النتيجة متفقة مع جميع الدراسات التي تناولت مقارنة طرق تقدير القدرة.

وتدلّل جميع هذه النتائج، إلى أثر الطريقة الإحصائية المستخدمة في التدريج العمودي على المقياس العمودي الناتج، وعموماً أظهرت طريقتي ثيرستون، والنموذج ثنائي المعلمة، نمطاً ثابتاً من التغيّر تقريباً، وهو النمط الأقرب إلى المقياس الأصلي، بدلالة الانحراف المعياري للدرجات الخام على الفقرات المشتركة بين الصفوف المتجاورة.

أظهرت النتائج المتعلقة بحجم الأثر، أنه يتناقص مع تقدم الصف الدراسي باستخدام جميع الطرق، ويعدّ هذا مؤشراً على تباطؤ النمو مع تقدم الصف الدراسي، إلا ان طريقة هيرونيموس أظهرت أعلى حجم أثر وبالتالي أعلى نمواً، يليها النموذج ثلاثي المعلمة والنموذج ثنائي المعلمة، وأخيراً أظهرت طريقة ثيرستون أقل قيمة لحجم الأثر، وبالتالي كانت الأقل نمواً، وتتفق هذه النتيجة مع دراسة تونغ (Tong, 2005) ودراسة ثوماس (Thomas, 2008)، ودراسة كيم (Kim, 2007)، ودراسة هندرسون وآخرون (Hendrickson et. al, 2004)، في حين اختلفت مع نتيجة دراسة ثوماس (Thomas, 2008)، التي أظهرت عدم انتظام في حجم الأثر في طريقة هيرونيموس، وفيما يتعلق بطرق تقدير القدرة فكانت نتائج حجم الأثر لها على عكس نتائج نمط التغيّر، حيث كانت قيمة حجم الأثر في طريقة الأرجحية العظمى (MLE)، أكبر منه في طريقة توقع التوزيع (EAP) في النموذج ثلاثي المعلمة، والعكس في النموذج ثنائي المعلمة، وتتفق هذه النتيجة مع الدراسات التي تناولت حجم الأثر عبر طرق التقدير، والتي أشارت إلى أن اتجاه حجم الأثر لطرق التقدير هو عكس اتجاه نمط التغيّر، وقد يعود السبب في ذلك إلى اعتماد قيمة حجم الأثر على قيمة التباين لكل صفين متجاورين، وبما أنه في طريقة (MLE) يتم حذف استجابات الطلبة ذوي الدرجات الكاملة، وكذلك الطلبة الذين لم يحصلوا على أي درجة، وبالتالي يتم حذف القيم المتطرفة، مما يؤدي إلى نقصان قيمة التباين، ومن ثم زيادة قيمة حجم الأثر.

وفيما يتعلق بمتوسط المسافة الأفقية يتناقص مع تقدم الصف الدراسي، أي نتائجه جاءت مشابهة لنتائج حجم الأثر، بالإضافة إلى أن الفروق بين طرق تقدير القدرة في النموذج الواحد كانت قليلة، أي إن نتائج طرق التقدير كانت متشابهة لنفس النموذج، وعموماً أظهر النموذج اللوجستي ثلاثي المعلمة تحت طريقة

الأرجحية العظمى في التقدير، أعلى متوسط للمسافة الأفقية بين الصفوف المتجاورة، وكذلك أظهرت طرق ثيرستون والنموذج الثنائي بكلا طريقتي التقدير تناقصاً قليلاً في متوسط المسافة الأفقية، في حين كان التناقص حاداً وملحوظاً في طريقة هيرونيموس والنموذج ثلاثي المعلمة، وهذا يتفق مع نتائج دراسة كيم (Kim, 2007) وهي الدراسة الوحيدة التي استخدمت هذا المحك، حيث أشارت نتائجها إلى تناقص متوسط المسافة الأفقية مع تقدم الصف الدراسي، وأن متوسط المسافة لم يختلف باختلاف طرق تقدير القدرة.

وأخيراً كانت نتائج نسبة النمو، تشير إلى أن الطرق الكلاسيكية (ثيرستون وهيرونيموس) والنموذج ثلاثي المعلمة باستخدام طريقة التقدير (EAP)، كان لها نسبة نمو أكبر من الواحد الصحيح، وبالتالي فإن نمو الطلبة ذوي التحصيل المرتفع أكبر من نمو الطلبة ذوي التحصيل المنخفض، وكانت هذه النسبة قريبة من الواحد الصحيح في النموذج ثنائي المعلمة بطريقتي تقدير القدرة، أي إن النمو متماثل للطلبة ذوي التحصيل المرتفع والمنخفض، وفي النموذج ثلاثي المعلمة باستخدام طريقة التقدير (MLE) كانت نسبة النمو أقل من الواحد الصحيح وعليه فان نمو الطلبة ذوي التحصيل المنخفض أعلى منه لذوي التحصيل المرتفع، ويتفق هذا مع دراسة ثوماس (Thomas, 2008) التي أشارت إلى أن نسبة النمو لذوي التحصيل المنخفض أعلى من الطلبة ذوي التحصيل المرتفع باستخدام النموذج ثلاثي المعلمة باستخدام طريقة التقدير (MLE)، في حين اختلفت هذه النتيجة مع نفس الدراسة فيما يخص طريقة هيرونيموس والتي كانت نسبة النمو فيها متماثلة للطلبة بغض النظر عن مستوى التحصيل أو القدرة.

وهنا لا بد من الإشارة إلى ما ذكره هوفر (Hoover, 1984)، بأن التزايد عبر المئين دليل على أن سرعة النمو الأكاديمي، وأورد فرضيته التي يدعي فيها أن تطوّر الطلبة الموهوبين يكون على منحنى نمو عقلي متسارع، في حين يكون للطلبة ذوي التحصيل المنخفض على منحنى نمو عقلي ثابت نسبياً، وبناءً على ما تقدم فان طريقة ثيرستون، وطريقة هيرونيموس، والنموذج ثلاثي المعلمة ، تعد نتائجها المتعلقة بنسبة النمو متسقة مع هذه الفرضية، في حين اختلفت نتائج النموذج ثنائي المعلمة، مع هذه الفرضية، وقد يعود السبب في ذلك إلى طبيعة الافتراضات الخاصة بهذا النموذج، والتي من ضمنها بأن استجابة الطلبة على الفقرات، لا تكون من خلال التخمين، وأن العاملين الوحيدين اللذان يحكمان استجابة الفرد هما صعوبة وتمييز الفقرة، وهذا ما يصعب تحققه، لا سيما في الفقرات ذات الاستجابة المنتقاة (الموضوعية)، وخاصة لدى الطلبة ذوي التحصيل المنخفض.

الاستنتاجات

إن عملية مقارنة نتائج هذه الدراسة مع نتائج الدراسات ذات الصلة التي تم الإشارة إليها في الفصل الثاني، فيه شيء من الصعوبة، كون نتائج التدريج العمودي تعتمد على طبيعة الاختبارات المستخدمة، والمحتوى الدراسي، ونوعية الفقرات، ومستويات الصفوف الدراسية، وتشكّل كل هذه العوامل عائقاً لمقارنة النتائج التي تم التوصل إليها مع نتائج الدراسات الأخرى ذات الصلة، إلا إنه يمكن التوصل إلى الاستنتاجات التالية:

1- إن المقاييس المطوّرة باستخدام طريقة ثيرستون، تعد إجراءات سهلة نسبياً، وافتراضاتها بسيطة، خاصة إذا كان للدرجات توزيعاً طبيعياً، وإن لم تتوفر خاصية أحادية البعد في البيانات، فإن طريقة ثيرستون تكون اختياراً جيداً، حيث أظهرت هذه الطريقة نتائج متسقة مع نتائج المقياس الأصلي بدلالة الأداء على الفقرات المشتركة.

2- تتطلب طريقة هيرونيموس إجراءات أكثر وأصعب مقارنة مع طريقة ثيرستون، وكذلك تتطلب برنامج خاص للربط، عدا عن تأثرها بنوع التحويلات اللاخطية المستخدمة، وبالتالي في حال اللجوء للطرق الكلاسيكية في التدريج العمودي، يفضّل استخدام طريقة ثيرستون على هذه الطريقة، نظراً لسهولة اجراءاتها وواقعية نتائجها.

3- الطرق المبنية على نماذج نظرية الاستجابة للفقرة، تتطلب افتراضات أكثر، وعمليات أعقد، لبناء المقياس العمودي، حيث كانت نتائج النموذج ثنائي المعلمة أقرب لنتائج طريقة ثيرستون، وفيها كفاية أكثر من خلال المحافظة على خصائص المقياس الأصلي وكذلك المقياس المعياري، في حين كانت نتائج النموذج ثلاثي المعلمة أقرب لنتائج طريقة هيرونيموس، وبالتالي يكون النموذج ثنائي المعلمة، هو النموذج الأكثر واقعية ومنطقية، من نماذج نظرية الاستجابة للفقرة، في حال تحققت افتراضات نظرية الاستجابة للفقرة، المتمثلة في أحادية البعد والاستقلال الموضعي، واذ لم تتوفر هذه الافتراضات تكون طريقة ثيرستون هي الطريقة الكلاسيكية الأكثر ملاءمة وواقعية.

4- فيما يتعلق بطرق تقدير القدرة كانت نتائجها متشابهة عبر النموذج الواحد، عدا محك المسافة الأفقية من خلال نسبة النمو والمسافة الأفقية عند مئينات معينة، حيث كانت طريقة (EAP) أكثر واقعية من طريقة (MLE)، وقد يعود السبب إلى فقدان بعض المعلومات ضمن طريقة الأرجحية العظمى (MLE)، حيث يتم فيها حذف بيانات المفحوصين الذين حصلوا على الدرجة الكاملة، والذين لم يحصلوا على أي علامة، وكذلك حذف بيانات الفقرات التي أجاب عليها جميع المفحوصين، والفقرات التي لم يجب عنها أي من المفحوصين.

التوصيات

في ضوء النتائج التي توصلت إليها هذه الدراسة يمكن اقتراح مجموعة من التوصيات هي:

1. استخدام النموذج اللوجستي ثنائي المعلمة في تقييم النمو الأكاديمي، في حال توفرت افتراضات نظرية الاستجابة للفقرة، المتمثلة في أحادية البعد والاستقلال الموضعي، واستخدام طريقة هيرونيموس في حال لم تتوفر هذه الافتراضات في البيانات.
2. الاستفادة من نتائج هذه الدراسة للتعرّف على خط النمو الأكاديمي في الرياضيات لدى طلبة الصفوف الرابع والخامس والسادس في الأردن، وكذلك في حال الاختيار بين طرق التدريج العمودي، سواء الطرق الكلاسيكية أو الطرق المبنية على نماذج نظرية الاستجابة للفقرة.
3. إجراء المزيد من الدراسات في موضوع التدريج العمودي، والذي يعدّ من المواضيع متعددة المتغيرات في ميدان القياس والتقويم، وذلك باستخدام متغيرات أخرى غير الواردة في هذه الدراسة، مثل استخدام تصاميم أخرى لجمع البيانات، ومحتويات دراسية أخرى مثل العلوم واللغة العربية واللغة الإنجليزية، ونماذج أخرى من نماذج نظرية الاستجابة للفقرة مثل النماذج متعددة الأبعاد.
4. ضرورة توفير اختبارات وطنية في محتويات مختلفة مثل الرياضيات واللغة والعلوم، ولجميع المراحل الدراسية لا سيما الأساسية، بحيث يتم بناؤها حسب تصاميم جمع المعلومات في التدريج العمودي، لتقييم النمو الأكاديمي لدى الطلبة من قبل وزارة التربية والتعليم والجهات البحثية.

REFERENCES

Asawakun, S. (1987). **The Effects of Data Collection Design and Equating Method on Vertically Equating Achievement Tests**. (Doctoral dissertation, The University of Iowa, 1987). Dissertation Abstracts International, 49(04), 799'A.

Becker, D. F., and Forsyth, R. A. (1992). An Empirical Investigation of Thurstone and IRT Methods of Scaling Achievement Tests. **Journal of Educational Measurement**, 29(4), 341-354.

Chen, W.H., & Thissen, D. (1997). Local Dependence Indices for Item Pairs Using Item Response Theory. **Journal of Educational and Behavioral Statistics.**

Clemans, W. V. (1993). Item Response Theory, Vertical Scaling, and Something's Awry in the State of Test Mark. **Educational Assessment**, 1(4), 329-347.

Ebel, R. L. (1972). **Essentials of Educational Measurement**. New Jersey: Prentice-Hall, INC. Englewood Cliffs.

Harris, D.J. (2007). Practical Issues in Vertical Scaling. In N.J. Dorans, M. Pommerich, & P. M. Holland (Eds.), **Linking and Aligning Scores and Scales**. New York, NY: Springer Science Business Media.

Hendrickson, A., Kolen, M., and Tong, Y. (2004). **Comparison of IRT Vertical Scaling from Scaling-test and Common-item Designs**. Paper presented at the annual conference of National Council on Measurement in Education. San Diego, CA.

Holland, P. W. (2002). Two Measures of Change in the Gaps between the CDFs of Test Score Distributions. **Journal o f Educational and Behavioral Statistics**, 27(1), 3-17.

Holland, P. W., & Dorans, N. J. (2006). Linking and Equating. In R. L. Brennan (Ed.), **Educational Measurement** (4th ed). Westport, CT: American Council on Education: Praeger Publishers.

Hoover, H. D. (1984). The most appropriate scores for measuring educational development in the elementary schools: GEs. **Educational Measurement: Issues and Practice**, 3(4), 8-14.

Hoover, H. D. (1988). Growth Expectations for Low-Achieving Students: A reply to Yen. **Educational Measurement: Issues and Practice**, 7(4), 21-23.

Iowa Testing Programs. (2011). Vertical Scaling and The Assessment of Growth. www.education.uiowa.edu/itp.

Jiao, H. Wang, S. (2007, April). **The Effects of The Selection of Vertical Linking Items On Modeling Student Growth**. Paper Presented At The Annual Meeting Of The National Council On Measurement In Education, Chicago.

Kim, J. (2007). **A Comparison of Calibration Methods and Proficiency Estimators For Creating IRT Vertical Scales**, Unpublished Ph.D. Dissertation, The University of Iowa.

Kolen, M. J. (2004). **Common Item Program for Equating, CIPE**. Iowa Testing Programs, The University of Iowa, IA 52242-1529.

Kolen, M. J. (2006). Scaling And Norming. In R. L. Brennan (Ed.), **Educational Measurement** (4th ed., pp. 155-186).

Westport, CT: American Council on Education: Praeger Publishers.

Kolen, M. J. (2011). **Vertical Scaling Methodologies, Applications and Research**, Training Session at NCME, The University of Iowa.

Kolen, M. J., & Brennan, R. L. (2004). **Test Equating, Scaling, and Linking: Methods and Practices**. (2nd ed.). New York: Springer-Verlag.

Linn, R. L. (2001). **The Design and Evaluation of Educational Assessment and Accountability Systems**. (No. 539). Los Angeles: University of California, Los Angeles: National Center for Research on Evaluation, Standards, and Student Testing.

Meng, H., Kolen, M. & Lohman, D. (2006). **An Empirical Investigation of IRT Scaling Methods: How different IRT models, Parameter Estimation Procedures, Proficiency Estimation Methods, and Estimation Programs affect the Results of Vertical Scaling for the Cognitive Abilities Test**. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Patz, R. J. (2007). **Vertical Scaling in Standards-Based Educational Assessment and Accountability Systems**. Prepared for the Technical Issues in Large-Scale Assessment, State Collaborative on Assessment and Student Standards, of the Council of Chief State School Officers.

Thomas, P. P. (2008). **An Investigation of the Effects of Varying the Domain Definition of Science and Method of Scaling on a Vertical Scale**. Unpublished Ph.D. Dissertation, The University of Iowa.

Thorndike, R.W. (1982). **Applied Psychometrics**. Boston: Houghton Mifflin Company.

Tong, Y. (2005). **Comparisons of Methodologies and Results in Vertical Scaling For Educational Achievement Tests**. Unpublished Ph.D. Dissertation, The University of Iowa.

Tong, Y., & Kolen, M.J. (2007). Comparisons of Methodologies and Results in Vertical Scaling for Educational Achievement Tests. **Applied Measurement in Education,** 20, 227- 253.

Tong, Y., Harris, D. J. (2004). **Choice o f Linking Methods and Scales in Vertical Scaling**. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Diego, CA.

Torgerson, S. Warren. (1958). **Methods of Scaling**. John Wiley & sons, Inc, USA.

Williams, V. Pommerich, M. and Thissen, D. (1998). A Comparison of Developmental Scales Based on Thurstone Methods and Item Response Theory. **Journal o f Educational Measurement**, 35(2), 93-107.

Yen, W. M. (1986). The Choice of Scale for Educational Measurement: An IRT perspective. **Journal of Educational Measurement**, 23(4), 299-325.

Yen, W. M. and Burket, G. R. (1997). Comparison of Item Response Theory and Thurstone Methods of Vertical Scaling. **Journal of Educational Measurement**, 34(4), 293-313.

Comparison of Vertical Scaling Methods based on the Procedures of the Classical Theory of Measurement to Methods based on Item Response Theory

Sharif Abdelruhman Alsoudi, Hala Mohammad Alshaw *

ABSTRACT

The main purpose of this study was to compare the Vertical Scaling methods based on the procedures of the Classical Theory in measurement and the methods based on the Item Response theory, To achieve the purposes of the study, two methods, Thurstone and Hieronymus, from the Classical theory in measurement and IRT model methods (two and three Dichotomous Logistic model) have been used. In the two models, two Estimation of Ability methods, the Maximum Likelihood Estimation and the Expected A Posteriori, have been employed using a multi-level Math achievement test for the Fourth, Fifth and Sixth Grades. The test uses the Common Item Design through the Mean, the Standard Deviation, the Effect Size and the Horizontal Distance criteria. The sample of the study consists of (1630) male and female students from the three basic grades mentioned above. Results show that the line of Academic growth was differences as the scaling method has been differences.

Keywords: Vertical Scaling, Thurstone method, Hieronymus method, IRT.

* Department of Curriculum and Instruction, The University of Jordan. Received on 1/12/2014 and Accepted for Publication on 25/5/2015.