# استخدام طريقة أشجار راش للكشف عن الأداء التفاضلي لفقرات الامتحان الوطني في الرياضيات للصف الرابع الأساسي في الأردن

آيات أبو مسرة و يوسف السوالمة*

*Doi: //10.47015/17.1.8*

تاريخ تسلم البحث: 2019/12/31 تاريخ قبوله: 2020/4/29

**ملخص**: هدفت الدراسة الحالية إلى استخدام طريقة أشجار راش للكشف عن الأداء التفاضلي لفقرات الاختبار الوطني لضبط نوعية التعليم في الرياضيات للصف الرابع الأساسي في الأردن، تبعًا للمتغيرات التي توفرها عملية التقييم، وهي: الجنس وموقع المدرسة والسلطة المشرفة على التعليم وطريقة تقديم الاختبار. إضافة إلى معرفة مدى تأثر الطريقة باختلاف حجم العينة. تكون مجتمع الدراسة من استجابات (31136) طالبًا وطالبة في الصف الرابع على (25) فقرة من نوع الاختيار من متعدد، تطابقت مع نموذج راش. وبتطبيق الطريقة على استجابات عينة عشوائية قوامها (1500) من المفحوصين، بينت النتائج أن هناك أداءً تفاضليًا وفقًا لمتغير الجنس فقط؛ إذ تم الكشف عن (8) فقرات ذات أداء تفاضلي، أي ما نسبته (32%) من الفقرات. وكان الأداء التفاضلي لصالح الذكور في (5) فقرات، ولصالح الإناث في (3) فقرات. وأظهرت النتائج حساسية طريقة أشجار راش لحجم العينة؛ إذ زادت فرص وجود أداء تفاضلي للفقرات بازدياد حجم العينة.

(**الكلمات المفتاحية**: أشجار راش، التقسيم المتكرر، الاختبار الوطني، الأداء التفاضلي، تحيز الفقرة)

**مقدمة:** تولي وزارة التربية والتعليم الأردنية الرياضيات اهتمامًا كبيرًا لاتساع نطاق المجالات التي تؤثر فيها. وتنبثق أهمية الرياضيات من حاجة الفرد إلى معرفتها لتوظيفها في تنظيم أمور حياته ومعاملاته المختلفة. وتعد الرياضيات ذات طبيعة تراكمية يبنى بعضها على بعض، فهي كالسلسلة، إذا ضاع منها جزء صعب الوصول إلى الجزء الآخر. وابتداءً من الصف الرابع، يتم الانتقال بالطالب من مرحلة العمليات الملموسة إلى العمليات المجردة.

وتعَد مراقبة مستوى مخرجات النظام التربوي وتقييمه من المهمات الضرورية لدفع عجلة التنمية في المجتمعات الحديثة، وتوفر نتائج الدراسات التقويمية الشاملة بيانات موضوعية تتعلق بمستوى إتقان الطلبة للكفايات الرئيسة التي يسعى النظام التعليمي إلى تنميتها وصقلها لدى النشء. ويمثل الاختبار الوطني مرجعًا مهمًا يمكن الاستفادة من نتائجه في تحديد جوانب القوة والضعف في أداء الطلبة، تمهيدًا لوضع الحلول المناسبة واتخاذ القرارات التربوية حيثما يلزم ذلك. وفي ظل انتشار حركات المساءلة وفق معايير الجودة، غدا من الضروري تنفيذ اختبارات وطنية تكشف عن مستويات جودة التعليم ونوعيته؛ ليصار إلى توظيف نتائجها في التخطيط لعمليات التصحيح التربوي وبرامج التطوير. وتُعد المشاركة في الاختبارات الوطنية والدولية بمثابة خطوات إيجابية في طريق إصلاح العملية التعليمية وتحقيق عدالة التقويم وتكافؤ الفرص.

**Using Rasch Trees Method to Detect Differential Item Functioning in the Jordanian National Test in Mathematics for Fourth Graders**

**Ayat Abu Masarrah and Yousef Al-Sawalmeh,** *Yarmouk University, Jordan.*

**Abstract**: The present study aimed to use Rasch trees method to detect differential item functioning (DIF) in the Jordanian National Test to control educational quality in mathematics, for fourth graders, according to gender, school location, supervising authority, and test modality, as well as to investigate the effect of sample size on the method. The study population consisted of (31136) fourth graders' responses on (25) multiple choice items fitted with the Rasch model. Rasch trees method was used to analyze (1500) randomly selected responses. The results indicated that there is DIF in terms of gender only. Eight items, representing (32%) of the test items, were detected to have DIF, five in favor of males and three in favor of females. The results showed that Rasch trees model is sensitive to sample size, where the DIF rate increased as the sample size increased.

(**Keywords**: Rasch Trees, Recursive Partitioning, National Test, Differential Functioning, Item Bias)

وانطلاقًا من أن تحصيل الطلبة يُعد مؤشرًا أساسيًا لتقييم نوعية التعليم، بدأت وزارة التربية والتعليم بتطبيق اختبار وطني سنوي لضبط نوعية التعليم في العام الدراسي 1999/ 2000م، بهدف قياس مدى امتلاك الطلبة لمهارات التعلم الأساسية وتحديد جوانب القوة والضعف لديهم في مباحث اللغة العربية واللغة الإنجليزية والعلوم والرياضيات في الصفوف الرابع والثامن والعاشر الأساسية (Jordanian Ministry of Education, 2018).

وزاد الاهتمام بهذه الاختبارات بعد عام 2012 لتحاكي الاختبارات الدولية في قياس المهارات والكفايات الأساسية للمراحل الدراسية أو العمرية المستهدفة من جهة، ولتتناغم مع مؤشرات الأداء الوطنية من جهة أخرى. وتصنف الاختبارات الوطنية ضمن الاختبارات واسعة النطاق (Large-scale Assessment). وهي أداة مهمة تزود متخذي القرارات وصانعي السياسات التربوية بمعلومات تمكنهم من الحكم على عمل الأنظمة التربوية. فهي تقدم ما يفيد في تحديد المستوى التحصيلي للطلبة بشكل عام، وبشكل تفصيلي حسب تصنيفات المحتوى العلمي وأهدافه بمستوياتها المختلفة (Alpursan, 2013).

* جامعة اليرموك، الأردن.

© حقوق الطبع محفوظة لجامعة اليرموك، إربد، الأردن.

وتطبق الاختبارات الوطنية لضبط نوعية التعليم في المدارس الحكومية والخاصة والمدارس التابعة لوكالة الغوث الدولية والمدارس التابعة لمديرية التعليم والثقافة العسكرية، بصورة ورقية أو إلكترونية. وتحلل نتائج هذه الاختبارات، وتكتب تقارير بذلك، وتقدم للمسؤولين على المستويات كافة لاتخاذ القرارات المناسبة. وتتضمن التحليلات عمل مقارنات بين مجموعات مختلفة من المفحوصين (Jordanian Ministry of Education, 2018).

ومن أجل أن تكون هذه القرارات والمقارنات صائبة، يجب أن تعتمد على بيانات صادقة؛ أي أن تتمتع أداة القياس التي جمعت من خلالها البيانات بمجموعة من الخصائص كالصدق والثبات والعدالة (Audeh, 2010). وإحدى الخصائص الرئيسة التي يجب الاهتمام بها في تقييم الاختبارات هي خاصية استقرار القياس ( Stability of Measurement) عبر مجموعات المفحوصين الفرعية، التي تعرف باسم اللاتغير في القياس (Measurement Invariance)، و يُعرف انتهاك هذه الخاصية على مستوى الفقرة باسم الأداء التفاضلي للفقرة (Differential Item Functioning: DIF) (Komboz et al., 2018). ويعد الاختبار الذي يحوي مثل هذه الفقرات غير محقق لمبدأ العدالة بين المفحوصين، مما يهدد صدقه (Rover, 2005).

ويقصد بمصطلح الأداء التفاضلي للفقرة (DIF) في الاختبار التربوي والنفسي أن احتمال الاستجابة الصحيحة بين المتقدمين للاختبار المتساوين في القدرة يختلف باختلاف مجموعاتهم الفرعية التي ينتمون لها حسب العرق أو الجنس أو غير ذلك. ومن المهم عندها تحديد هذه الفقرات، للعمل على تحسينها أو حذفها من الاختبار (Westers & Kelderman, 1992).

ويشتمل الأدب التربوي على العديد من الأساليب الإحصائية للكشف عن الأداء التفاضلي للفقرة، إلا أن أغلبها يعتمد على مقارنة تقديرات معلَمَات الفقرة بين مجموعتين أو أكثر من المفحوصين تم تحديدها مسبقًا، مثل ذكور-إناث، أو مجموعة مستهدفة - مجموعة مرجعية، وذلك اعتمادًا على عدد قليل من المتغيرات، مثل الجنس أو العرق. وفي حال المتغيرات الكمية مثل العمر، يتم اعتماد الوسيط – اعتباطيًا- كدرجة قطع لتقسيم العينة ( Komboz et al., 2018).

واقترح ستروبل وآخرون (Strobl et al., 2010) طريقة جديدة للكشف عن الأداء التفاضلي للفقرة ثنائية التدريج في نموذج راش اللوجستي أحادي المعلمة، سميت بأشجار راش ( Rasch Trees). وتصنف أشجار راش باعتبارها طريقة شبه معلمية هدفها الكشف عن الاختلافات في تقديرات معلمات الفقرة بين مجموعات المفحوصين الفرعية، الأمر الذي يدل وجوده على انتهاك خاصية اللاتغير في القياس. وتعتمد الطريقة على الخوارزمية الإحصائية المعروفة بـ "النموذج القائم على التقسيم المتكرر" ( Recursive partitioning based model)، التي توفر إمكانية تقييم مطابقة البيانات لأحد النماذج المعلمية، واختبار عدم استقرار المعلمات عبر مختلف المجموعات الفرعية لمجموعة من المتغيرات. وعندما تظهر البيانات أداءً تفاضليًا، يتم تحديد المتغيرات المسؤولة عن تقسيم عينة الدراسة إلى مجموعات فرعية وتحديد نقاط التقسيم أيضًا، باستخدام اختبار إحصائي يتحكم في معدلات الخطأ من النوع الأول في تحديد تلك المتغيرات. وتضمن التمثيلات البيانية لفروع الشجرة التي يتم الحصول عليها من الخوارزمية المستخدمة صدق التفسيرات الإحصائية (Zeileis et al., 2008).

وينظر إلى أشجار راش على أنها طريقة شاملة للكشف عن الأداء التفاضلي للفقرة، وفيها تعرضُ النتائج على شكل شجرة ذات فروع يسمى كل فرع فيها عقدة (node)، وتمثل كل عقدة مجموعة فرعية من المفحوصين. كما توفر الطريقة تقديرات مَعلَمات الفقرات الخاصة بتلك المجموعة، مما يمكن من ملاحظة أي المجموعات تتأثر بالأداء التفاضلي، وفي أي الفقرات بشكل مباشر. وعندما لا تظهر أي من الفقرات أداءً تفاضليًا، ستتكون الشجرة من فرع واحد، وفي هذه الحالة سيكون نموذج راش أحادي البعد مناسبًا لوصف مجموعة البيانات كاملة. أما إذا أظهرت شجرة راش انقسامًا واحدًا على الأقل، دلّ هذا على وجود أداء تفاضلي للفقرات. وعندها يكون الاختبار غير مناسب لمقارنة المجموعات المختلفة من المفحوصين (Strobl et al., 2010).

ويمكن أن تساعد طريقة أشجار راش الباحثين في دراسة التفاعلات بين مدى واسع من مصادر الأداء التفاضلي مع الحفاظ على تفسيرات ذات معنى، وعلى قابلية المقارنة مع الطرق التقليدية للكشف عن DIF (Strobl et al., 2015). وهذه خاصية تنفرد بها هذه الطريقة، وتتميز بها على الطرق الأخرى (Kopf, 2013). والأصل في الاختبار أن يقيس قدرة المفحوص بدقة دون أي اعتبار للمجموعة الديموغرافية التي ينتمي لها (Whitmore, 1995). ويتركز الاهتمام في تفسير درجات الاختبار على ثبات القياس عبر وبين المجموعات المختلفة من المفحوصين. ويكون هذا التفسير متحيزًا أو غير صحيح عندما تؤثر متغيرات ليست ذات صلة بالسمة المقاسة في الأداء. ولضمان فائدة الدرجات في عمل مقارنات عادلة بين الأفراد والمجموعات، لابد من تقييم ثبات القياس بين المتقدمين للاختبار عبر المجموعات المختلفة المتعلقة بالجنس والعرق والجنسية والسياق الثقافي وغيرها من الخصائص المتعلقة بخلفية الطالب. وهناك الكثير من المتغيرات المتعلقة بخلفية الطالب، التي قد تكون مصادر مختلفة للأداء التفاضلي في فقرات الاختبار (Albano & Rodriguez, 2013).

وتؤكد شيا (Shea, 2013) على أن التحديد الإحصائي للفقرات التي تظهر DIF بحسب متغير ثنائي واحد لا يوفر الكثير من المعلومات حول أسباب الأداء التفاضلي. ويتطلب تحديد هذه الأسباب مقاربة شاملة لتحليل DIF؛ فخصائص المفحوصين تفاعلية، وغالبًا ما تكون مترابطة. ومن غير الممكن فهمها بوضوح من تلقاء نفسها. فمثلاً، عندما يتم تحديد نوع الجنس كعامل مهم في الأداء التفاضلي، فإنه من المحتمل أن تعمل متغيرات مترابطة

مشتركة أخرى (السكن، والوضع الاقتصادي الاجتماعي، وغيرها) كمصادر لعدم ثبات القياس، وبالتالي تقوّض صدق تفسير درجات الاختبار بالاعتماد على متغيرات فردية فقط (McGraw et al., 2006).

ويرى كامبل (Campbell, 1989) أن تناول نوع الجنس والعرق والحالة الاجتماعية الاقتصادية بشكل منفصل يجعل البحث غير مكتمل في أحسن الأحوال، ومجرد خطأ واضح في أسوأ الأحوال. ويعني ذلك أنه إذا لم يتم بحث هذه المتغيرات معًا، فلن يتم التعرف على الاختلافات داخل المجموعات أو العثور على التفاعلات. فالتحديد المسبق للمجموعة المرجعية وللمجموعة المستهدفة (الذكور مقابل الإناث على سبيل المثال) في تحليل DIF المرتبط بالجنس، لا يسمح بالاستفادة من مجموعة المتغيرات المشتركة الأخرى التي توفرها التقييمات واسعة النطاق في الوقت نفسه، مما يحول دون الكشف الكامل عن الطبيعة المعقدة والمترابطة في DIF (Zhang, 2001).

وقد تم تطوير طريقة أشجار راش للتغلب على هذه المشكلات، ولتقديم أدلة بشأن DIF المحتمل بدلالة مجموعة من المتغيرات المشتركة والتفاعلات فيما بينها. والملاحظ في هذه الطريقة أنها لا تحتاج إلى تعيين مسبق للمجموعات الفرعية، بل يتم تقسيم المفحوصين إلى مجموعات، بالاعتماد على مجموعة من المتغيرات والتفاعل فيما بينها، مما يشكل أنماط مجموعات متنوعة ومعقدة، لكنها واقعية (Aryadoust, 2018). على سبيل المثال، قد يكون الأداء التفاضلي موجودًا لصالح الإناث فوق عمر (40) سنة فقط، دون باقي المفحوصين من الإناث والذكور؛ أي أن الأداء التفاضلي مرتبط بتفاعل متغيري العمر والجنس معًا (Strobl et al., 2015). وتجدر الإشارة إلى أن طريقة التقسيم المتكرر لم تعد تقتصر على الفقرات ثنائية التدريج، بل اتسع مجال استخدامها لتشمل الفقرات متعددة التدريج في نموذج راش (Komboz et al., 2018).

ويتم استنتاج بنية شجرة راش من البيانات باتباع الخطوات الأربع الآتية:

1. **تقدير معلمات الفقرة (Estimating the Item Parameters):** يتم تقدير معلمات الفقرة للعينة كاملة وفق نموذج راش، باستخدام طريقة الأرجحية العظمى الشرطية (Conditional Maximum Likelihood) (Zeileis et al. 2008).

2. **اختبار عدم استقرار المعلمة (Testing for Parameter Instability):** يعرف عدم استقرار المعلمة بانحرافها عن الوسط العام للمعلمة الذي يقدره نموذج راش (Aryadoust, 2018). وهنا يتم البحث فيما إذا كانت معلمات الفقرة تختلف بين مجموعات المفحوصين عبر المتغيرات المحددة، وذلك باستخدام منحى "اختبارات التغير البنائي" (Structural Change Tests Approach) المستمد من الاقتصاد القياسي (Komboz et al., 2018)؛ إذ تُحسب الاختبارات الإحصائية لعدم استقرار المعلمة وقيم $p$ ($p$ values) الخاصة بها، لكل متغير مرشح لتقسيم المفحوصين لمجموعات فرعية، بحيث يعتمد المتغير الذي تكون قيمة $p$ له أقل قيمة من بين قيم $p$ الدالة إحصائيًا، لتقسيم العينة إلى مجموعتين فرعيتين. ثم تتوالى عملية تقسيم المجموعات الفرعية باختيار المتغير الذي له قيمة $p$ الأقل من بين متغيرات الدراسة المتبقية، ليتم بذلك تكوين العقدة الثانية. ويستمر التقسيم الثنائي بشكل متكرر حتى تتجاوز كل قيم $p$ مستوى الدلالة ($\alpha = 0.05$)؛ أي تصبح غير دالة إحصائيًا، أو حتى يقل عدد المشاهدات في العينة الفرعية عن عتبة معينة أقلها (10) (Boehmke & Greenwell, 2020). ويشير حدوث ذلك إلى استقرار دال إحصائيًا للمعلمة.

3. **اختيار نقاط القطع (Selecting the Cutpoints):** بعد أن يتم اختيار المتغير الذي سيتم تقسيم العينة من خلاله، يتم تحديد نقطة القطع مباشرة من البيانات باعتماد النقطة التي تعظّم أرجحية التقسيم (Maximal Partitioned Likelihood) من بين جميع نقاط القطع المرشحة ضمن نطاق ذلك المتغير. ويرتبط تحديد نقطة القطع المثالية للمتغير بعدم الاستقرار الدال إحصائيًا للمعلمة، مما يمنع التحيز في اختيار المتغير (Zeileis et al., 2008).

4. **معايير التوقف (Stopping Criteria):** يُطبق نوعان من معايير التوقف: الأول هو مستوى الدلالة وعادة ما يتم ضبطه عند ($\alpha = 0.05$)؛ إذ يتم إيقاف التقسيم إذا لم يوجد المزيد من عدم الاستقرار الدال إحصائيًا في أي من متغيرات الدراسة. أما معيار التوقف الثاني فهو تحديد الحدّ الأدنى لحجم العينة لكل عقدة، الذي يتم تحديده ليوفر أساسًا مناسبًا لتقدير معلمات الفقرة حسب النموذج المستخدم من نماذج راش للاستجابات الثنائية أو المتعددة. وبالتالي يجب زيادته بزيادة عدد الفقرات المراد تقدير معالمها (Strobl et al., 2010).

والجدير بالذكر أن هذه الطريقة تعالج مشكلة الاختبار المتكرر الذي قد يؤدي إلى زيادة معدل الخطأ عندما يكون عدد المتغيرات كبيرًا، من خلال استخدام تعديل بونفيروني (Bonferroni Adjustment) على معيار التقسيم $p$ ؛ أي أنه يتم فحص أكثر من فرضية بشكل متزامن، بحيث يتم الاحتفاظ بمستوى الدلالة للشجرة بأكملها، مما يضمن أن الأداء التفاضلي الذي يتم اكتشافه بهذه الطريقة لا يتم اكتشافه كخطأ ناتج من تعدد المتغيرات.

وقد تناولت العديد من الدراسات الأداء التفاضلي للفقرات في حال التعيين المسبق للمجموعات. فقد هدفت دراسة ستونبيرغ (Stoneberg, 2004) إلى فحص الأداء التفاضلي في اختبارات ايداهو (Idaho) في التحصيل في القراءة وفنون اللغة والرياضيات في الصفوف الرابع والثامن والعاشر، باستخدام طريقتي

(Simultaneous Item Bias Test :SIBTEST)، ومانتل هانزل (Mantel-Hansel). وتوفر للدراسة عينات عشوائية تمثل حوالي نصف المتقدمين لهذه الامتحانات وتراوحت بين (8659) في رياضيات الصف العاشر و(9676) في اختبارات اللغة للصف الرابع. وكشفت النتائج أن نسبة الفقرات التي تبدي أداءً تفاضليًا للجنس تتراوح بين (7%) في اختبار القراءة للصف الرابع و(37%) في اختبار الرياضيات للصف العاشر.

وهدفت دراسة أبو حماد (Abu Hammad, 2008) إلى تقصي فاعلية طرق منحنى خصائص الفقرة في الكشف عن الأداء التفاضلي لمتغير الجنس في الاختبار الدولي في الرياضيات. تكون الاختبار من (40) فقرة، وجرى تطبيقه على عينة من (800) طالب وطالبة من مدارس بني كنانة في الأردن. وأشارت النتائج إلى وجود (27) فقرة ذات أداء تفاضلي لمتغير الجنس. وسعى العجلوني وسوالمة (Al- Ajlouni and Sawalmeh, 2016) إلى فحص الأداء التفاضلي لفقرات ومموهات اختبار الرياضيات في التقييم الوطني لاقتصاد المعرفة. وقد أظهرت بعض الفقرات والمموهات أداءً تفاضليًا وفقًا لمتغيري الجنس وموقع المدرسة. واستهدفت دراسة أبو شندي وكاظم (Abu Shindi and Kazem, 2018) معرفة الأداء التفاضلي لفقرات اختبار الرياضيات في برنامج التنمية المعرفية بسلطنة عمان. وأظهرت النتائج تأثر عدد الفقرات التي أظهرت أداءً تفاضليًا لمتغير الجنس بحجم عينة المفحوصين.

ويوجد عدد قليل من الدراسات التي استخدمت طريقة أشجار راش للكشف عن الأداء التفاضلي. فقد استخدم إرتونا وجلبال (Ertuna and Gelbal, 2016) طريقة أشجار راش للكشف عن الأداء التفاضلي للفقرات في امتحانات تحديد إنجاز الطالب التركية (ÖBBS) في اللغة التركية والرياضيات والعلوم والعلوم الاجتماعية، باستخدام متغيرات الجنس والفرص التعليمية والاجتماعية-الاقتصادية للوالدين. وقد أشارت النتائج إلى وجود أداء تفاضلي للفقرة في جميع الاختبارات الفرعية وفقا لمتغيرات الدراسة. وقارن ليو (Liu, 2017) طريقة أشجار راش بطريقتين تقليديتين للكشف عن الأداء التفاضلي. وقد كانت النتائج متقاربة للطرق الثلاث. كما أظهرت النتائج وجود أداء تفاضلي مرتبط بالخلفية العرقية للطالب فقط، ولم يظهر أداء تفاضلي مرتبط بمتغيرات الجنس والحالة الاجتماعية-الاقتصادية واللغة الأم للطالب. كما أظهرت النتائج حساسية طريقة أشجار راش لاختلاف حجم العينة؛ إذ تزداد معدلات الكشف عن الأداء التفاضلي بازدياد حجم العينة.

وقد أجرى أريادوست (Aryadoust, 2018) دراسة استخدم فيها طريقة أشجار راش لتحديد مصادر الأداء التفاضلي في اختبار فهم القراءة، بالنسبة لمتغيرات المعرفة بالمفردات والمعرفة بالقواعد والجنس والتخصص الأكاديمي. وقد أشارت النتائج إلى وجود أداء تفاضلي لجميع متغيرات الدراسة باستثناء التخصص الأكاديمي. وهدفت دراسة ألتنتاس وكوتلو ( Altıntaş and Kutlu, 2019) إلى الكشف عن الأداء التفاضلي في فقرات اختبار مهارات التعلم الأساسية للطلبة الأجانب في جامعة أنقرة، وفق متغيرات الجنس والدولة باستخدام طريقة أشجار راش. وأظهرت النتائج وجود (16) فقرة ذات أداء تفاضلي من أصل (60) فقرة بالنسبة لمتغير الدولة، بينما لم تظهر فقرات الاختبار أداءً تفاضليًا بالنسبة لمتغير الجنس.

وفيما يتعلق باختبارات الرياضيات الوطنية لضبط نوعية التعليم في الأردن، فقد درس الطراونة (Altarawneh, 2011) الأداء التفاضلي لفقرات اختبار الصف الثامن بحسب كل من الجنس والسلطة المشرفة (حكومي، خاص) على انفراد، باستخدام طريقة مانتل-هانزل. وبينت دراسته وجود (4) فقرات ذات أداء تفاضلي لصالح الطالبات و(17) فقرة ذات أداء تفاضلي لصالح التعليم الخاص من بين (30) فقرة.

ودرس البرصان (Alpursan, 2013) الأداء التفاضلي لفقرات اختبار الصف العاشر (30 فقرة) بحسب الجنس في مستويات الأداء المختلفة للطلبة (متقدم، ماهر، ماهر جزئيًا، مبتدىء) باستخدام طريقة مانتل-هانزل. وبينت دراسته وجود (9) فقرات ذات أداء تفاضلي في المستوى المتقدم؛ من بينها (5) فقرات لصالح الطالبات، و(14) فقرة ذات أداء تفاضلي في المستوى الماهر؛ من بينها (10) فقرات لصالح الطالبات، و(19) فقرة ذات أداء تفاضلي في المستوى الماهر جزئيًا؛ من بينها (9) فقرات لصالح الطالبات، و(20) فقرة ذات أداء تفاضلي في المستوى المبتدىء؛ من بينها (8) فقرات لصالح الطالبات. وهذا يعني أن عدد الفقرات ذات الأداء التفاضلي يزداد كلما تدنى مستوى الأداء.

وقد درست فريحات (Frehat, 2014) الأداء التفاضلي لفقرات اختبار الصف العاشر (24 فقرة) بحسب طريقة تقديم الاختبار (ورقي، إلكتروني)، باستخدام مؤشر الأداء التفاضلي التعويضي (CDIF). وأظهرت النتائج أن (8.3%) من الفقرات قد أبدت أداءً تفاضليًا لصالح الاختبار الإلكتروني. كما درس العمري وشطناوي (Alomary and Shatnawi, 2016) الأداء التفاضلي لفقرات اختبار الصف العاشر (24 فقرة) بحسب الجنس باستخدام مؤشر الأداء التفاضلي اللاتعويضي (NCDIF). وأظهرت النتائج وجود (9) فقرات ذات أداء تفاضلي؛ منها (6) فقرات لصالح الطالبات توزعت على موضوعات الجبر والإحصاء والقياس و(3) فقرات لصالح الطلاب توزعت على موضوعات الهندسة والعمليات على الأعداد.

وهكذا تمت دراسة الأداء التفاضلي لاختباري الصفين الثامن والعاشر من خلال عدة دراسات اختلفت فيما بينها بطريقة الكشف عن الأداء التفاضلي. وقد أكدت هذه الدراسات وجود فقرات ذات أداء تفاضلي بحسب كل من الجنس والسلطة المشرفة على المدارس وطريقة تقديم الاختبار للطلبة بشكل منفصل. وتجدر

الإشارة إلى أن فحص الأداء التفاضلي بحسب متغير واحد أو أكثر بشكل منفصل لا يسمح بدراسة التفاعلات بين هذه المتغيرات ولا يوفر معلومات كافية تفسر أسباب الأداء التفاضلي ( Campbell, 1989).

أما امتحان الصف الرابع، فلم يحظَ بأية دراسة حول استقرار القياس عبر مجموعات المفحوصين المختلفة بالرغم من أهمية الفئة العمرية والمرحلة الدراسية اللتين يستهدفهما. كما أنه من غير الممكن توفير أدلة على إمكانية تعميم النتائج التي تم الحصول عليها من اختباري الصفين الثامن والعاشر على اختبار الصف الرابع لأسباب مختلفة، أبرزها اختلاف الفئة العمرية. كما يلاحظ عدم وجود دراسة عربية استخدمت طريقة أشجار راش – حسب علم الباحثين – للكشف عن الأداء التفاضلي لفقرات الاختبار. وعليه جاءت الدراسة الحالية للكشف عن الأداء التفاضلي لفقرات اختبار الصف الرابع في الرياضيات باستخدام طريقة أشجار راش التي تتميز عن كثير من الطرق الأخرى بتناولها لعدة متغيرات بشكل متزامن.

**مشكلة الدراسة وسؤالاها**

تستخدم نتائج الاختبار الوطني لضبط نوعية التعليم في الرياضيات للصف الرابع في مقارنة مستوى الأداء على الاختبار لمجموعات مختلفة من المفحوصين. ولا تتوفر أية معلومات حول خاصية ثبات القياس لهذا الاختبار بين وعبر تلك المجموعات. ويعتمد تفسير درجات الاختبار وعمل مقارنات عادلة بين المجموعات على توفر هذه الخاصية. ولضمان فائدة الدرجات في عمل مقارنات عادلة بين الأفراد والمجموعات، هناك حاجة ماسة لتقييم ثبات القياس بين المتقدمين للاختبار عبر المجموعات المختلفة من خلال دراسة الأداء التفاضلي لفقراته. وتتوافر من عملية تطبيق الاختبار معلومات ديموغرافية متعددة حول خلفيات الطلبة تتعلق بالجنس وموقع المدرسة والسلطة المشرفة على التعليم وطريقة تقديم الاختبار لهم. وقد تعمل هذه المتغيرات كمصادر محتملة للأداء التفاضلي في فقرات الاختبار. وقد تم تناول بعض هذه المتغيرات بصورة منفصلة في دراسات سابقة للأداء التفاضلي لفقرات بعض الاختبارات في الرياضيات. وهذا التناول المنفرد لا يوفر معلومات كافية للتعرف على الاختلافات داخل المجموعات، مما يحول دون الكشف الكامل عن الطبيعة المعقدة والمترابطة في DIF. لذلك سيتم في هذه الدراسة تناول هذه المتغيرات معًا بصورة متزامنة في تحليل الأداء التفاضلي لفقرات الاختبار من خلال استخدام طريقة أشجار راش. لذلك جاءت الدراسة الحالية للكشف عن الأداء التفاضلي في فقرات الاختبار الوطني لضبط نوعيــة التعليم في الرياضيات للصف الرابع باستخدام طريقة أشجار راش. ونظرًا لحداثة هذه الطريقة، تمت دراسة مدى تأثرها بحجم العينة.

وبالتحديد، فقد سعت الدراسة للإجابة عن السؤالين الآتيين:

1. ما الفقرات التي تظهر أداءً تفاضليًا في الاختبار الوطني لضبط نوعيــة التعليم لمبحث الرياضيات للصف الرابع للعام الدراسي 2018/2017م لمتغيرات الجنس وموقع المدرسة والسلطة المشرفــة عليه وطريقة تقديم الاختبار، باستــخدام طريقة أشجار راش؟
2. هل تختلف المتغيرات التي تظهر فقرات الاختبار الوطني لضبط نوعيــة التعليم لمبحث الرياضيات للصف الرابع للعام الدراسي 2018/2017م أداءً تفاضليًا بالنسبة لها باستــخدام طريقة أشجار راش، باختلاف حجم العينة (500، 2500، 4000 طالب)؟

**أهمية الدراسة**

تكتسب الدراسة الحالية أهميتها من أهمية التحقق من خاصية استقرار القياس للاختبار الوطني لضبط نوعيــة التعليم في الرياضيات للصف الرابع، الذي يشكل الصف الأعلى في المرحلة الأساسية الدنيا، نظراً لأهمية هذا الاختبار. وزادت أهمية هذا الاختبار بعد ظهور مؤشرات متعددة على ضعف الطلبة في الصفوف المبكرة (الثلاثة الأولى) في القراءة والرياضيات ( RTI International, 2012). وتعتمد وزارة التربية والتعليم على الاختبار في تشخيص واقع تعليم الرياضيات في الصفوف المبكرة في الأردن، وعمل مقارنات بين المجموعات المختلفة من المفحوصين. وهذا يستلزم المزيد من الاهتمام بخصائص الاختبار السيكومترية، وبالأخص ما يتعلق بتحليل الأداء التفاضلي لفقراته كوسيلة للتحقق من خاصية استقرار القياس عبر مجموعات المفحوصين، بما يضمن عدالة المقارنات وصحة القرارات القائمة على الاختبار. وبالتالي، توفر الدراسة الحالية مساهمة تساعد العاملين في مجال الاختبارات في بناء اختبارات موضوعية وعادلة في العملية التعليمية. كما تكتسب أهميتها من استخدام طريقة راش التي يتوقع منها توفير أدلة عملية بشأن الأداء التفاضلي لفقرات الاختبار بدلالة مزيج من المتغيرات المشتركة والمترابطة، ولكونها طريقة فعالة في الكشف عن الأداء التفاضلي للفقرة في غياب التحديد المسبق لمجموعات المقارنة. كما تفتح الدراسة الحالية الباب أمام الباحثين للتعرف على هذه الطريقة ولاستخدامها في تحليلات الأداء التفاضلي المستقبلية، مقارنة بالطرق الأخرى شائعة الاستخدام.

**مصطلحات الدراسة**

- **أشجار راش:** طريقة إحصائية جديدة للكشف عن الأداء التفاضلي في الفقرة باستخدام الخوارزمية الإحصائية المعتمدة على التقسيم المتكرر للمجموعة الكلية في ضوء مجموعة من المتغيرات.
- **الأداء التفاضلي للفقرة:** دالة مشتقة إحصائيًا للتعبير عن الفروق في الاستجابة على الفقرة بين المجموعات الفرعية للمفحوصين ممن لهم مستوى القدرة نفسه.

- **الاختبار الوطني لضبط نوعية التعليم**: اختبار سنوي تعدّه وزارة التربية والتعليم (مديرية الامتحانات والاختبارات) لقياس مدى امتلاك الطلبة في الصف المستهدف لمهارات التعلم الأساسية ومستويات أداء الطلبة في المباحث التي اختبروا فيها.

**محددات الدراسة**

اقتصرت الدراسة الحالية على طلبة الصف الرابع الأساسي للعام الدراسي 2018/2017م الذين تقدموا لاختبار الرياضيات الوطني لضبط نوعية التعليم والذين تطابقت استجاباتهم مع نموذج راش أحادي المعلمة. كما تم استخدام طريقة أشجار راش المعتمدة على التقسيم المتكرر للكشف عن الأداء التفاضلي للفقرة. وقد اعتمدت الدراسة الحالية على فقرات الاختيار من متعدد بأربعة بدائل التي تطابقت الاستجابات عليها مع نموذج راش أحادي المعلمة.

**الطريقة**

**منهج الدراسة**

تم استخدام المنهج الوصفي التحليلي للكشف عن الفقرات التي تظهر أداءً تفاضليًا في الاختبار الوطني لضبط نوعية التعليم لمبحث الرياضيات للصف الرابع للعام الدراسي 2018/2017م لمتغيرات الجنس وموقع المدرسة والسلطة المشرفة عليه وطريقة تقديم الاختبار، باستخدام طريقة أشجار راش، وذلك لمناسبته لطبيعة الدراسة وأهدافها.

**مصدر البيانات**

المصدر الرئيسي لبيانات الدراسة هو ملف استجابات طلبة الصف الرابع في مدارس المملكة الأردنية الهاشمية على الاختبار الوطني لضبط نوعية التعليم في الرياضيات، الذي جرى تطبيقه في العام الدراسي 2018/2017، وهو أقرب تطبيق إلى التاريخ الذي أجريت فيه الدراسة؛ إذ تم تطبيق (40) فقرة من نوع الاختيار من متعدد. ومن هذا الملف، تم الاحتفاظ بالاستجابات على الفقرات التي تطابقت مع نموذج راش، وبلغ عددها (31136) استجابة على (25) فقرة. وشكّلت هذه الاستجابات ملف بيانات الدراسة.

**عينات التحليل**

تكونت عينة التحليل الأولى لغرض الإجابة عن السؤال الأول من استجابات (1500) طالب وطالبة اختيرت عشوائيًا من ملف بيانات الدراسة. ويبين الجدول (1) توزيع العينة حسب الجنس وموقع المدرسة والسلطة المشرفة وطريقة تقديم الاختبار. وتكونت عينات التحليل لغرض الإجابة عن السؤال الثاني من استجابات (500 و2500 و4000) طالب وطالبة، تم اختيارها بشكل عشوائي من ملف بيانات الدراسة. والاعتبار الذي روعي في هذه الأعداد هو أن تكون ضمن مدى واسع، وأن تختلف عن الأعداد في دراسة ليو (Liu)، وأن يكون من ضمنها ما يزيد على (600) لأغراض الحصول على تقديرات مستقرة لمعلمة الصعوبة.

**الجدول 1**

*توزيع العينة حسب الجنس وموقع المدرسة والسلطة المشرفة وطريقة تقديم الاختبار*

| المتغير | مستويات المتغير | عدد الطلبة | النسبة المئوية |
|---|---|---|---|
| الجنس | ذكور | 701 | %46.7 |
| | إناث | 799 | %53.3 |
| المنطقة | مدينة | 1103 | %73.5 |
| | قرية | 397 | %26.5 |
| السلطة المشرفة | الحكومة | 1188 | %79.2 |
| | وكالة الغوث | 193 | %12.9 |
| | الخاصة | 119 | %7.9 |
| طريقة تقديم الاختبار | ورقي | 1371 | %91.4 |
| | إلكتروني | 129 | %8.6 |
| المجموع | | 1500 | %100 |

**أداة الدراسة**

الأداة الرئيسة المستهدفة هي الاختبار الوطني لضبط نوعية التعليم في الرياضيات للصف الرابع الأساسي. ويجري تطبيق هذا الاختبار سنويًا في الوقت نفسه، بشكل منتظم، في جميع مدارس المملكة. ويتقدم للاختبار حوالي (25%) من طلبة الصف الرابع الأساسي في الأردن. وتستخدم في الاختبار فقرات من نوع الاختيار من متعدد. وتتم كتابة هذه الفقرات من لجنة متخصصة وفق جدول مواصفات وقواعد محددة، وتتم مراجعة الفقرات من لجنة أخرى. ويتم تجريب الفقرات بصورة أولية على عينة ممثلة للصف الرابع في الأردن، ويتم إعداد الصورة النهائية للاختبار من (40) فقرة وتراجع من فريق من المتخصصين. وتضمن هذه المراجعات الصدق المنطقي للاختبار. وتتوزع الفقرات على محاور: الأعداد والعمليات عليها، والجبر، والهندسة والقياس، والإحصاء والاحتمالات. وتقيس هذه الفقرات أداء الطلبة في مهارات: تمثيل الأعداد وقراءتها وكتابتها، ومقارنة الأعداد، وإجراء العمليات الحسابية على الأعداد وتقديرها، وإكمال الأنماط وحل جمل مفتوحة على الأعداد. وقد كان معامل ثبات الاتساق الداخلي للاختبار باستخدام معادلة "كرونباخ ألفا" (0.91) للصورة الورقية و(0.89) للصورة الإلكترونية.

**إجراءات الدراسة**

تم الحصول على ملف إكسل بالبيانات اللازمة من إدارتي الامتحانات والاختبارات ومركز الملكة رانيا العبد الله لتكنولوجيا التعليم والمعلومات في وزارة التربية والتعليم. وقد تمت تنقية

البيانات وحذف المتغيرات والبيانات التي ليست لها علاقة بالدراسة، ثم تحويل الملف إلى ملف (SPSS).

وتم التحقق من افتراضات نظرية الاستجابة للفقرة؛ إذ استخدمت برمجية (SPSS) للتحقق من افتراض أحادية البعد، من خلال المؤشرات التي اعتمدت على التحليل العاملي باستخدام طريقة المكونات الرئيسة. وأظهرت النتائج أن نسبة التباين المفسر للعامل الأول (18.58%)، وهي نسبة قريبة جدًا من (20%)، وأن نسبة التباين المفسر للعامل الثاني (4.53%). كما أظهرت أن ناتج قسمة نسبة التباين المفسر للعامل الأول على نسبة التباين المفسر للعامل الثاني (3.39)، وهي أكبر من (2)، مما يرجح وجود عامل سائد يمكن أن يستدل منه على أحادية البعد (Hattie, 1985 ; Reckase, 1997).

واستخدمت الحزمة الإحصائية (Subscore Package) ضمن برمجية R للحصول على قيم المؤشرQ3 للتحقق من افتراض الاستقلال الموضعي. وقد تبين أن قيم مؤشر Q3 تقل عن (0.02)؛ مما يدل على تحقق الافتراض (Chen & Thissen, 1997).

وهناك مؤشرات على تحقق افتراض التحرر من السرعة لكون مدة الاختبار ساعة ونصف الساعة (مدير إدارة الاختبارات في وزارة التربية والتعليم، اتصال شخصي، 28 نيسان 2019)، وهي مدة كافية للطلبة ليجيبوا عن جميع الفقرات. كما تم التحقق من افتراض اطرادية منحنى خصائص الفقرة من خلال استخدام برمجية (BILOG-MG 3) لجميع الفقرات. ويبن الشكل (1) منحنيات خصائص الفقرة للفقرات (30، 40) كمثال على ذلك. يشير المحور الأفقي في الرسم إلى القدرة (Ability)، ويشير المحور الرأسي إلى احتمال (Probability) إجابة الفقرة إجابة صحيحة.

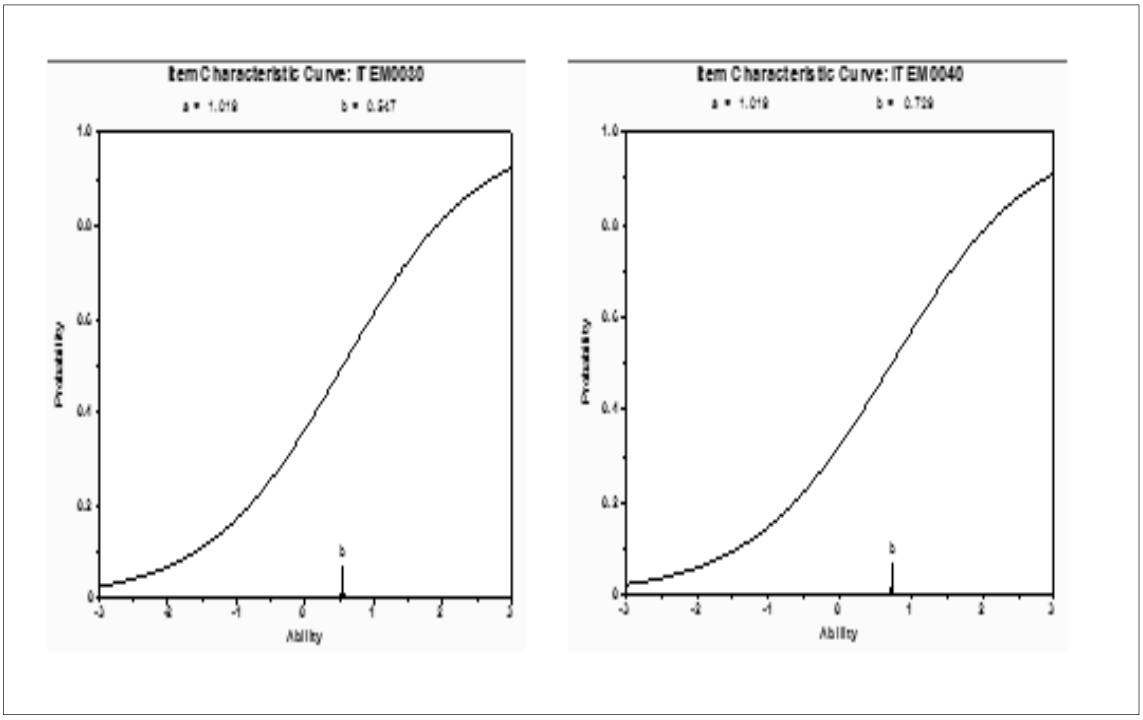
وقد تم تحليل البيانات باستخدام برمجية ( BILOG-MG 3)، للتحقق من مطابقة الأفراد لنموذج راش، باستخدام الإحصائي $\chi^2$ للمطابقة $(\chi^2\ statistic\ fit)$ عند مستوى الدلالة ($\alpha = 0.01$) للتدليل على مطابقة البيانات للنموذج المستخدم. وقد مرت إجراءات المطابقة للأفراد والفقرات بعدة مراحل، انتهت بمطابقة استجابات (31136) طالبًا وطالبة فقط لنموذج راش و25 فقرة من فقرات الاختبار، وهذه الاستجابات شكلت ملف بيانات الدراسة الذي منه اختيرت عينات التحليل.

## نتائج الدراسة ومناقشتها

يبين الجدول (2) الإحصاءات الوصفية لأداء الطلبة في الاختبار حسب مستويات متغيرات الدراسة ودلالة الفروق بين الأوساط الحسابية.

**الشكل 1**

*منحنيات خصائص الفقرة للفقرات (30، 40)*



**الجدول 2**

*الإحصاءات الوصفية لأداء الطلبة في الاختبار حسب مستويات متغيرات الدراسة ودلالة الفروق بين الأوساط الحسابية*

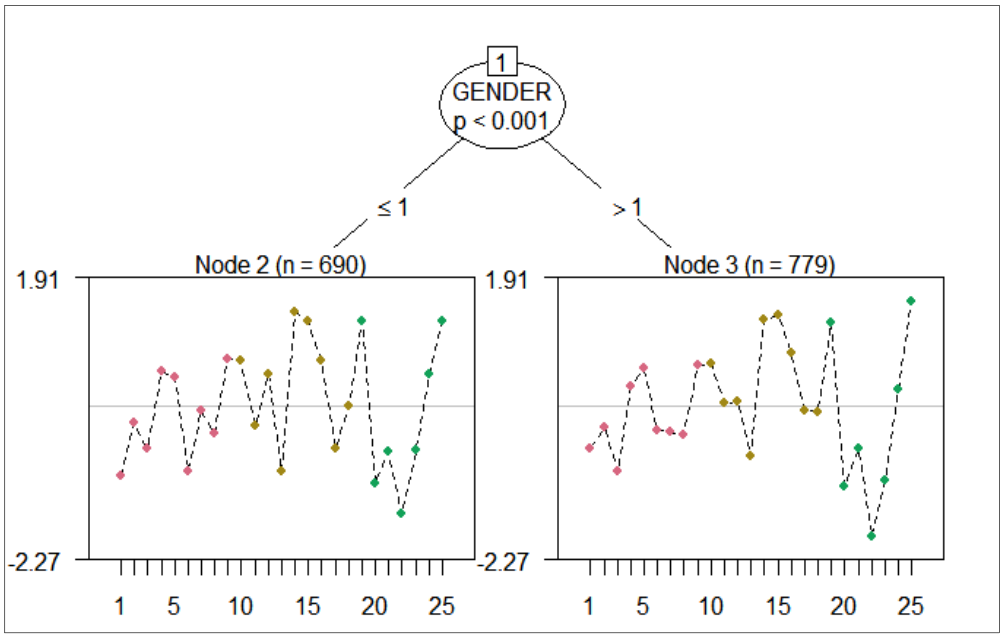
| المتغير | فئات المتغير | العدد | الوسط الحسابي | الانحراف المعياري | قيمة الاختبار | الدلالة الإحصائية |
|---|---|---|---|---|---|---|
| الجنس | ذكور | 14426 | 24.26 | 7.926 | t=-3.47 | 0.001 |
| | إناث | 16710 | 24.57 | 7.667 | | |
| موقع المدرسة | مدينة | 22348 | 24.46 | 7.712 | t=-3.47 | 0.336 |
| | قرية | 8788 | 24.36 | 7.983 | | |
| السلطة المشرفة | الحكومة | 24483 | 23.61 | 7.678 | F=661.57 | 0.000 |
| | الخاصة | 4123 | 27.65 | 7.105 | | |
| | وكالة الغوث | 2530 | 27.09 | 7.978 | | |
| طريقة تقديم الامتحان | ورقي | 28329 | 24.26 | 7.828 | t=-12.45 | 0.000 |
| | إلكتروني | 2807 | 26.17 | 7.164 | | |

تظهر النتائج في الجدول (2) وجود فروق ذات دلالة إحصائية في الأداء بين الجنسين لصالح الإناث؛ وبين طريقتي تقديم الاختبار لصالح الطريقة الإلكترونية؛ وبحسب السلطة المشرفة لصالح المدارس الخاصة ووكالة الغوث مقابل مدارس الحكومة. أما عن وسط الأداء بحسب موقع المدرسة، فقد أظهرت النتائج عدم وجود فرق ذي دلالة إحصائية بين مدارس المدن ومدارس القرى. وتفسير هذه الفروق يتطلب معرفة ما إذا كانت حقيقية أم غير حقيقية، وهذا ما يجيب عنه تحليل DIF.

**للإجابة عن السؤال الأول** المتعلق بالكشف عن الفقرات التي تظهر أداءً تفاضليًا في الاختبار الوطني تبعًا لمتغيرات الجنس وموقع المدرسة والسلطة المشرفة وطريقة تقديم الاختبار، باستخدام طريقة أشجار راش، أدخلت بيانات عينة عشوائية تكونت من (1500) طالب إلى الحزمة Psychotree في برمجية R لاستخراج أشجار راش، واستخراج قيم الاختبار الاحصائي لاستقرار المعلمة وقيم $p$ المرتبطة بها، واستخراج قيم معلمات صعوبة الفقرات للمجموعات الفرعية. ويبين الشكل (2) شجرة راش بالنسبة لمتغيرات الدراسة. وتوضح الأرقام على المحور الأفقي أرقام الفقرات، بينما تشير القيم على المحور الرأسي إلى صعوبة الفقرات مقدرة بحسب نظرية الاستجابة للفقرة.

**الشكل 2**

*شجرة راش للكشف عن الأداء التفاضلي للفقرات بالنسبة لمتغيرات الدراسة*



يُظهر الشكل (2) أن الشجرة تتضمن ثلاث عُقد؛ وفيها يقسم متغير الجنس العقدة (1) إلى عقدتين طرفيتين: العقدة (2) للذكور والعقدة (3) للإناث، مما يشير إلى ظهور أداء تفاضلي وفق متغير الجنس فقط، في حين أن باقي متغيرات الدراسة لم تدخل في عملية التقسيم. ويعني ذلك أن الفقرات لا تظهر أداءً تفاضليًا بحسب تلك المتغيرات. كما يُظهر الشكل مواقع الفقرات في العقدتين الطرفيتين على متصل القدرة (العمود الرأسي). وتكون الفقرات تحت خط الوسط سهلة بحيث تحتاج لقدرات دون الوسط لإجابتها إجابة صحيحة، في حين تكون الفقرات فوق خط الوسط صعبة. وبتفحص الشكل (2) بدقة يتبين وجود اختلاف ظاهري بين أبعاد الفقرات (1، 3، 6، 11، 12، 17، 23، 25) في العقدتين عن خط الوسط؛ بمعنى أن هذه الفقرات مرشحة لان تبدي أداءً تفاضليًا لمتغير الجنس. وتجدر الإشارة إلى أن طريقة راش تقوم بحذف الأفراد الذين يجيبون عن جميع الفقرات إجابة صحيحة أو خطأ، وكذلك حذف الفقرات التي يجيب عنها جميع الطلبة إجابة صحيحة أو خطأ، وبتفحص الشكل (2)، نجد أن مجموع الطلبة في العقدتين الطرفيتين (1469)؛ أي أنه تم حذف (31) طالبًا أجابوا عن جميع الفقرات إجابة صحيحة.

ويبين الجدول (3) قيم الاختبار الإحصائي لاستقرار المعلمة وقيمة $p$ المرتبطة بها وفق متغيرات الدراسة، في العُقد الثلاث لشجرة راش.

**الجدول 3**

*قيم الاختبار الإحصائي لاستقرار المعلمة وقيمة* p *المرتبطة بها لمتغيرات الدراسة*

| | العُقدة | الجنس | المنطقة | السلطة المشرفة | طريقة تقديم الاختبار |
|---|---|---|---|---|---|
| **العقدة 1** | الاختبار الإحصائي | 114.75 | 37.8 | 28.54 | 47.714 |
| | قيمة $p$ | $**< 0.004e^{-5}$ | 0.89 | 0.99 | 0.251 |
| **العقدة 2** | الاختبار الإحصائي | — | 32.90 | 30.25 | 33.68 |
| | قيمة $p$ | — | 0.98 | 0.998 | 0.97 |
| **العقدة 3** | الاختبار الإحصائي | — | 37.18 | 34.71 | 45.23 |
| | قيمة $p$ | — | 0.85 | 0.95 | 0.32 |

** قيمة $p$ المرتبطة بالمتغير الذي تم اختياره للتقسيم.

يلاحظ من الجدول (3) أن متغير الجنس هو المتغير الوحيد الذي تم اختياره لتقسيم الشجرة، وأنه لم يعد متاحًا للتقسيم بدءاً من العقدة الثانية، نظرًا لأنه يوفر نقطة قطع واحدة، ثم توقف التقسيم في العقد الفرعية حيث تتجاوز كل قيم $p$ مستوى الدلالة ($\alpha = 0.05$)، مما يشير إلى وجود استقرار في المعلمة.

ونظرا لأن طريقة أشجار راش لا توفر اختبارًا إحصائيًا للأداء التفاضلي على مستوى الفقرة، فإنه يمكن مقارنة توزيع إجابات الذكور مع توزيع إجابات الإناث باستخدام اختبار مربع كاي للاستقلال بدرجات حرية (1). ويعدّ الاختلاف بين التوزيعين مؤشرًا على الأداء التفاضلي. ويبين الجدول (4) قيم معلمة الصعوبة لفقرات الاختبار في العُقدتين الثانية والثالثة، وقيم مربع كاي للاستقلال وقيم $p$ المرتبطة بها. وتشير قيم معلمة الصعوبة المنخفضة إلى سهولة الفقرة، والقيم المرتفعة إلى صعوبة الفقرة.

**الجدول 4**

*قيم معلمة الصعوبة لفقرات الاختبار في العُقدتين الثانية والثالثة وقيم مربع كاي للاستقلال وقيم $p$ المرتبطة بها*

| الفقرة | معلمة الصعوبة في العقدة 2 | معلمة الصعوبة في العقدة 3 | اختبار مربع كاي للاستقلال | قيمة $p$ |
|---|---|---|---|---|
| فقرة 1 | -1.019 | -0.626 | 9.05 | 0.003 |
| فقرة 2 | -0.230 | -0.310 | 0.10 | 0.75 |
| فقرة 3 | -0.631 | -0.969 | 4.71 | 0.03 |
| فقرة 4 | 0.522 | 0.309 | 1.66 | 0.20 |
| فقرة 5 | 0.435 | 0.571 | 1.94 | 0.16 |

| الفقرة | معلمة الصعوبة في العقدة 2 | معلمة الصعوبة في العقدة 3 | اختبار مربع كاي للاستقلال | قيمة $p$ |
|---|---|---|---|---|
| فقرة 6 | -0.960 | -0.345 | 21.8 | 0.00 |
| فقرة 7 | -0.069 | -0.366 | 3.83 | 0.05 |
| فقرة 8 | -0.397 | -0.422 | 0.01 | 0.94 |
| فقرة 9 | 0.698 | 0.609 | 0.1 | 0.76 |
| فقرة 10 | 0.676 | 0.641 | 0.01 | 0.92 |
| فقرة 11 | -0.277 | 0.051 | 7.73 | 0.01 |
| فقرة 12 | 0.478 | 0.083 | 7.25 | 0.01 |
| فقرة 13 | -0.970 | -0.741 | 3.4 | 0.07 |
| فقرة 14 | 1.401 | 1.284 | 0.18 | 0.68 |
| فقرة 15 | 1.274 | 1.354 | 1.04 | 0.31 |
| فقرة 16 | 0.683 | 0.789 | 1.39 | 0.24 |
| فقرة 17 | -0.622 | -0.067 | 19.34 | 0.00 |
| فقرة 18 | -0.001 | -0.080 | 0.09 | 0.76 |
| فقرة 19 | 1.274 | 1.236 | 0.02 | 0.88 |
| فقرة 20 | -1.132 | -1.182 | 0.02 | 0.89 |
| فقرة 21 | -0.657 | -0.619 | 0.28 | 0.60 |
| فقرة 22 | -1.580 | -1.925 | 3.33 | 0.07 |
| فقرة 23 | -0.648 | -1.091 | 8.34 | 0.00 |
| فقرة 24 | 0.478 | 0.251 | 1.96 | 0.16 |
| فقرة 25 | 1.274 | 1.565 | 6.46 | 0.01 |

يؤكد الجدول (4) أن الفقرات (1، 3، 6، 11، 12، 17، 23، 25) تبدي أداءً تفاضليًا؛ لصالح الذكور في الفقرات (1، 6، 11، 17 و25)؛ إذ كانت هذه الفقرات أسهل لهم منها للإناث، ولصالح الإناث في الفقرات (3، 12، 23)؛ أي أن هناك (8) فقرات في اختبار الرياضيات الوطني لضبط نوعيــة التعليم تبدي أداءً تفاضليًا لمتغير الجنس حسب طريقة أشجار راش، من أصل (25) فقرة. وكان الأداء التفاضلي لصالح الذكور في (5) فقرات، ولصالح الإناث في (3) فقرات. بينما لم تظهر فقرات الاختبار أداءً تفاضليًا بحسب متغيرات موقع المدرسة والسلطة المشرفة وطريقة تقديم الاختبار.

وقد يكون الاختلاف في الأداء بين الذكور والإناث على الاختبار عائدًا إلى عوامل جينية أو عوامل ثقافية-اجتماعية كتلك التي كونت اعتقادًا لدى الذكور بأن المواد العلمية -خصوصًا الرياضيات - مناسبة لهم أكثر، وأن اللغات والمواد الإنسانية تناسب الإناث أكثر، كما بينت ذلك دراسة هارلين ( as cited in Aladaileh, 1996)، مما يشكل دافعية واتجاهات إيجابية نحو المواد العلمية والرياضيات لدى الذكور أكثر منها لدى الإناث، ومن المعلوم أن إتقان المهارات يرتبط بتقدير الذات والثقة بالنفس.

وقد يعزى الأداء التفاضلي للجنس إلى نوع الفقرة. وهناك دراسات تؤكد أن أداء الذكور أفضل من أداء الإناث في فقرات الاختيار من متعدد (Taylor & Lee, 2012)، والفقرات المستخدمة في الاختبار هي فقرات اختيار من متعدد. وربما يفسر ذلك أن عدد الفقرات المتحيزة للذكور أعلى من عدد الفقرات المتحيزة للإناث. كما قد يعزى الأداء التفاضلي للجنس إلى محتوى الفقرة؛ فالأداء التفاضلي قد يختلف باختلاف مجال المحتوى، فيتفوق الذكور في الهندسة والأشكال والجداول، فيما تتفوق الإناث في الجبر (Garner & Engelhard, 1999).

كما قد يعزى الأداء التفاضلي للجنس إلى أن الفقرات التي أظهرت أداءً تفاضليًا لمتغير الجنس تقيس أبعادًا أخرى، إلى جانب البعد الأساسي الذي صممت لقياسه. وتعمل هذه الأبعاد لصالح طرف على حساب الطرف الآخر. فقد تتصل هذه الأبعاد باستخدام

القوانين أو عملية عقلية أو غموض ما أو خلل في صياغة الفقرة. ومن الملاحظ من النتائج أن أداء الذكور كان أفضل على الفقرات (1 و6 و11) التي تتطلب قدرة عددية: ككتابة الأعداد الصحيحة ومقارنتها وإجراء العمليات الحسابية. وهذا يتفق مع ما أشارت إليه أبو حماد (Abu Hammad, 2008) من تفوق الذكور في القدرة العددية مقارنة بالإناث. ولأن البيئة الشرقية تسمح للذكور بالخروج إلى السوق وحدهم أو برفقة آبائهم أكثر من الإناث، فهذا قد يجعلهم أكثر ألفة مع وحدات قياس الكتلة المستخدمة في السوق كالتي ظهرت في الفقرة (17)، مما قد يفسر تفوق أداء الذكور في هذه الفقرة.

وأظهر تفحص المسائل التي تقيس مهارة تحديد فرص حدوث الحوادث الواردة في مقرر الرياضيات للصف الرابع اقتصارها على عملية السحب من صندوق واحد، في حين أن الفقرة (25) تتطلب السحب من صندوقين على التوالي؛ أي أنها تحتاج إلى تعميم المعرفة السابقة على المسائل غير المألوفة. وهذا يتفق مع ما أشارت إليه أبو حماد (Abu Hammad, 2008) من تفوق الذكور في القدرة على التعميم، بينما تفضل الإناث الاستخدام الروتيني للقوانين، كما تعلمنها في الصف، كما في الفقرة (12) التي كان أداء الإناث فيها أفضل، حيث طلب فيها إيجاد ناتج جمع كسرين لهما المقام نفسه.

وتتطلب الإجابة عن الفقرة (3) قدرة لغوية إلى جانب القدرة العددية المستهدفة في كتابة الكسر بالكلمات، ولا يخفى تفوق الإناث على الذكور في القدرة اللغوية، خصوصًا في المراحل الأولى. وهذا يتفق مع ما أكد عليه تايلر ولي (Taylor & Lee, 2012) من أن الإناث يتفوقن على الذكور في الفقرات المفتوحة. كما تفوقت الإناث في الفقرة (23) حيث تتطلب الاستجابة عليها درجة من الانتباه والحرص في تفحص الأشياء وعَدِّها لتحديد الشيء الأكثر تكرارًا. وهناك شواهد كثيرة على أن الإناث أكثر التزامًا وجديّة من الذكور، خاصة في مثل هذه الاختبارات التي لا تؤخذ نتائجها بالاعتبار في تقييم الطالب. ومن العوامل التي قد تؤثر في الأداء التفاضلي بحسب الجنس اختلاف فرص التخمين بين الجنسين، ولا تتوفر مؤشرات من الدراسة الحالية على مثل هذا الاختلاف.

والجدير بالذكر أن العوامل التي نوقشت ليست عوامل سببية تؤثر في الأداء التفاضلي لمتغير الجنس، بل هي مجرد عوامل مقبولة ظاهريًا، استنادًا إلى ظروف الاختبار ومراجعة نتائج اختبارات سابقة. وهناك حاجة لدراسة إضافية توفر تحليلاً متعمقًا لإجابات الطلبة لتوفير المزيد من الأدلة عن العوامل التي تؤثر في الأداء التفاضلي للفقرة.

وتشير النتائج إلى سلامة بنية فقرات الاختبار وخلوها من الأداء التفاضلي بالنسبة لمتغيرات موقع المدرسة والسلطة المشرفة وطريقة تقديم الاختبار باستخدام طريقة أشجار راش. وهذا يتناقض مع الدراسات التي تمت لامتحاني الصفين الثامن والعاشر. ويعزى هذا الاختلاف بصورة أساسية إلى اختلاف الفئة العمرية واختلاف حجوم العينات. وهناك مؤشرات من نتائج الاختبارات السابقة على تفوق الإناث في الصفين الثامن والعاشر. وقد يعزى ذلك إلى انخفاض جدية واهتمام الذكور بازدياد الفئة العمرية والصف بسبب أن الدرجات في مثل هذا الاختبار لا تدخل في حساب درجات الطلبة.

وتتفق نتائج الدراسة الحالية مع نتائج دراسة ستونبيرج (Stoneberg, 2004) التي أظهرت وجود فقرات ذات أداء تفاضلي بالنسبة لمتغير الجنس بنسبة (37%) على اختبار الرياضيات للصف العاشر. كما تتفق مع نتائج دراسة العجلوني وسوالمة (Al-Ajlouni and Sawalmeh, 2016) التي أظهرت وجود أداء تفاضلي في (27%) من الفقرات لمتغير الجنس في اختبار الرياضيات في التقييم الوطني لاقتصاد المعرفة. وتتفق مع نتائج دراسة إرتونا وجلبال (Ertuna and Gelbal, 2016) التي استخدمت طريقة أشجار راش للكشف عن الأداء التفاضلي للفقرات بالنسبة لعدة متغيرات منها الجنس، التي أظهرت وجود أداء تفاضلي وفقًا لمتغيرات الدراسة. إلا أنها تختلف مع نتائج دراسة ليو (Liu, 2017) التي أظهرت فيها طريقة أشجار راش وجود أداء تفاضلي مرتبط بالخلفية العرقية للطالب فقط، في حين لم تظهر الفقرات أداءً تفاضليًا لمتغيرات الجنس والحالة الاجتماعية-الاقتصادية واللغة الأم للطالب. كما تختلف مع نتائج دراسة التنتاس وكوتلو (Altıntaş and Kutlu, 2019) التي أظهرت فيها طريقة أشجار راش وجود (16) فقرة ذات أداء تفاضلي بالنسبة لمتغير الدولة، بينما لم تظهر فقرات الاختبار أداءً تفاضليًا بالنسبة لمتغير الجنس.

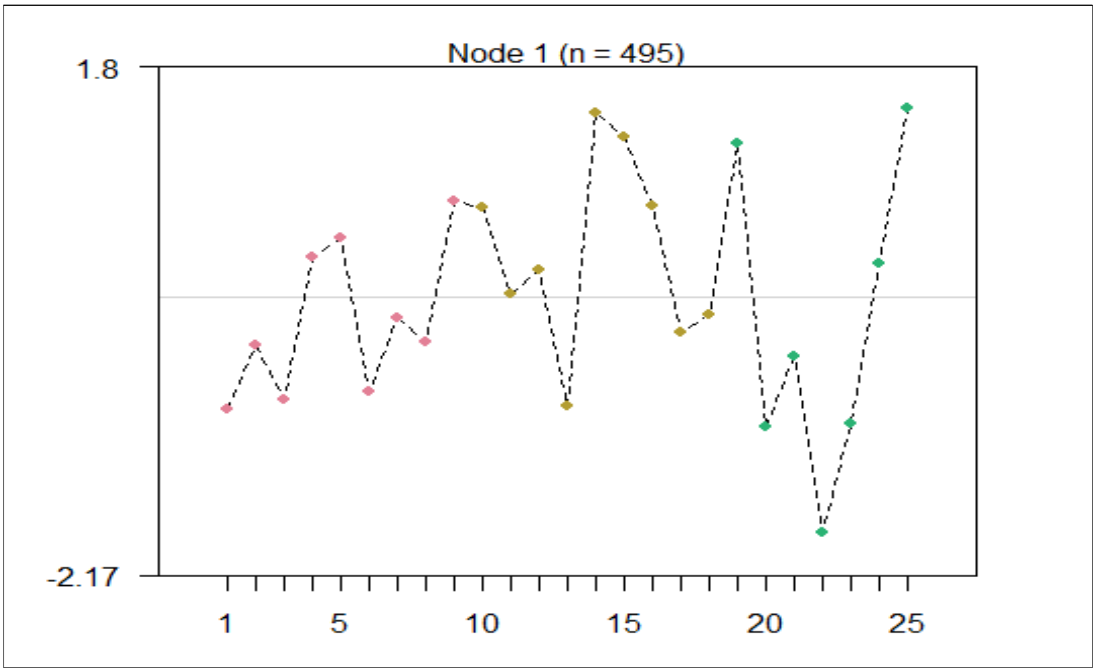
**وللإجابة عن السؤال الثاني** المتعلق بتحديد المتغيرات التي تظهر فقرات الاختبار الوطني لضبط نوعيـة التعليم أداءً تفاضليًا بالنسبة لها باستـخدام طريقة أشجار راش، باختلاف حجم العينة، تم الاكتفاء بعرض أشجار راش لثلاث عينات عشوائية تتكون من (500 و2500 و4000) طالب وطالبة، دون إعطاء تفاصيل عن قيم الاختبار الإحصائي لاستقرار المعلمة وقيم $p$ المرتبطة بها التي أدت إلى تشكيل هذه الأشجار، ودون إعطاء جداول تبين تقديرات لمعالم الصعوبة للفقرات في المجموعات المختلفة. وفيما يلي عرض لنتائج السؤال الثاني.

### العينة الأولى: (500) طالب وطالبة

يبين الشكل (3) شجرة راش للكشف عن الأداء التفاضلي للفقرات، بالنسبة لمتغيرات الجنس وموقع المدرسة والسلطة المشرفة وطريقة تقديم الاختبار، للعينة الأولى (500 طالب وطالبة).

**الشكل 3**

*شجرة راش للكشف عن الأداء التفاضلي للفقرات بالنسبة لمتغيرات الدراسة للعينة الأولى (500 طالب وطالبة)*




يلاحظ من الشكل (3) أن شجرة راش لم تنقسم نهائيًا، وتشكلت من عقدة واحدة لكل أفراد العينة، مما يشير إلى عدم ظهور أداء تفاضلي في فقرات الاختبار وفق متغيرات الدراسة. ويعني ذلك أن الشجرة تتشكل من عقدة طرفية واحدة تعبر عن جميع أفراد هذه العينة. وتعني هذه النتيجة أن الفقرات لم تظهر أداءً تفاضليًا بالنسبة لأي من متغيرات الدراسة في العينة المكونة من (500) طالب وطالبة. وهذا يعزى إلى صغر حجم العينة بالنسبة لما تتطلبه طريقة أشجار راش. ويؤدي صغر حجم العينة إلى زيادة الخطأ المعياري في تقدير المعالم، مما يجعل -من الناحية الإحصائية- الفروق بين قيم المعالم غير دالة إحصائيًا وفقًا لأي من المتغيرات المدروسة. وتختلف هذه النتيجة مع ما تم الوصول إليه في السؤال الأول، حيث كان حجم العينة (1500) طالب وطالبة. فقد أظهرت الفقرات أداءً تفاضليًا لمتغير الجنس، فقط ولم تظهر أداء تفاضليًا لمتغيرات السلطة المشرفة وموقع المدرسة وطريقة تقديم الاختبار. وقد يعزى ذلك إلى عدم التوازن في حجوم العينات الفرعية لتلك المتغيرات؛ إذ كان التفاوت في الحجم أقل ما يمكن في حالة الجنس. لكن مع ذلك لم يكن ممكنًا الكشف عن التفاعلات فيما بين المتغيرات.

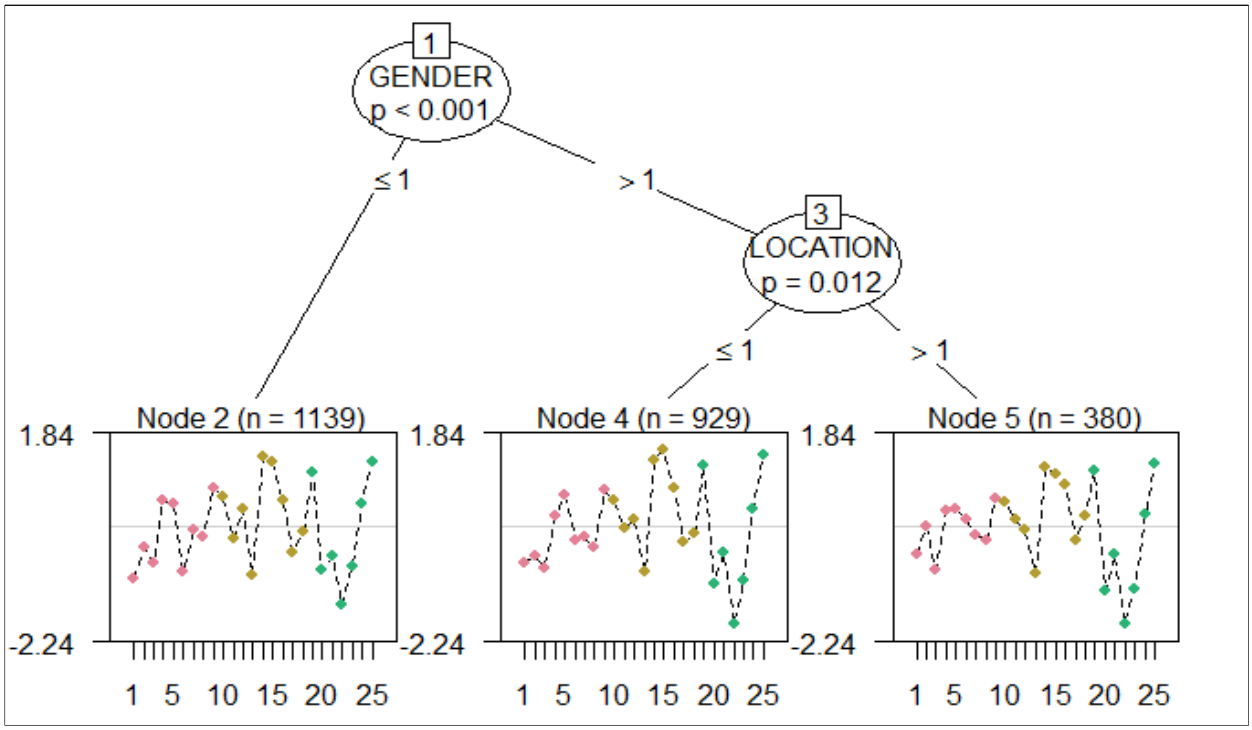
**العينة الثانية: (2500) طالب وطالبة**

يبين الشكل (4) شجرة راش للكشف عن الأداء التفاضلي للفقرات، بالنسبة لمتغيرات الجنس وموقع المدرسة والسلطة المشرفة وطريقة تقديم الاختبار للعينة الثانية (2500 طالب).

يلاحظ من الشكل (4) أن شجرة راش للعينة المكونة من (2500) طالب وطالبة انقسمت إلى خمس عُقد؛ إذ يقسم متغير الجنس العقدة (1) إلى عقدتين: العقدة (2) للذكور والعقدة (3) للإناث، مما يشير إلى ظهور أداء تفاضلي في فقرات الاختبار وفق متغير الجنس. ثم يقسم متغير المنطقة العقدة (3) إلى عقدتين طرفيتين: العقدة (4) للإناث في مدارس المدن والعقدة (5) للإناث في مدارس القرى. وهذا يشير إلى ظهور أداء تفاضلي في فقرات الاختبار وفق متغير المنطقة لدى مجموعة الإناث، في حين لم تدخل باقي المتغيرات في عملية التقسيم؛ أي أن الفقرات لا تظهر أداءً تفاضليًا بالنسبة لها، ويعني ذلك أن الشجرة تضمنت ثلاث عقد طرفية تمثل ثلاث مجموعات من المفحوصين: العقدة (2) وتمثل الذكور، والعقدة (4) وتمثل الإناث في مدارس المدن، والعقدة (5) وتمثل الإناث في مدارس القرى. هذه المجموعات الثلاث هي نتاج التفاعل بين متغيري الجنس وموقع المدرسة؛ أي أن هناك متغيرين (الجنس والموقع) أظهرت الفقرات أداءً تفاضليًا بالنسبة إليهما. وكشفت النتائج هنا عن تفاعل ثنائي بين متغير الجنس ومتغير موقع المدرسة. ويظهر من الشجرة أن موقع المدرسة ليس له تأثير لدى الذكور، بينما كان له تأثير لدى الإناث.

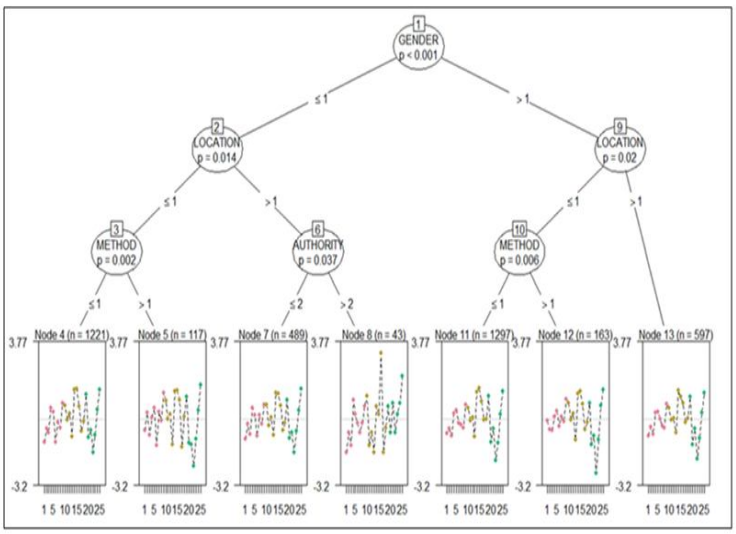
**الشكل 4**

*شجرة راش للكشف عن الأداء التفاضلي للفقرات بالنسبة لمتغيرات الدراسة للعينة الثانية (2500 طالب وطالبة)*




**العينة الثالثة: (4000 طالب وطالبة)**

يبين الشكل (5) شجرة راش للكشف عن الأداء التفاضلي للفقرات، بالنسبة لمتغيرات الدراسة للعينة المكونة من (4000) طالب وطالبة.




الشكل (5): شجرة راش تكشف عن الأداء التفاضلي للفقرات بالنسبة للجنس وموقع المدرسة والسلطة المشرفة وطريقة تقديم الاختبار (4000) طالب وطالبة

ويُظهر الشكل (5) أن شجرة راش للعينة الثالثة انقسمت إلى (13) عُقدة. وقد قسم متغير الجنس العينة إلى العقدة (2) للذكور والعقدة (9) للإناث، وقسم تفاعل باقي المتغيرات العقدتين (2 و9) إلى (7) عقد طرفية تمثل سبع مجموعات من المفحوصين: العقدة (4) وتمثل الذكور في مدارس المدن الذين تقدموا لامتحان ورقي، والعقدة (5) وتمثل الذكور في مدارس المدن الذين تقدموا لامتحان إلكتروني، والعقدة (7) وتمثل الذكور في مدارس القرى الحكومية، والعقدة (8) وتمثل الذكور في مدارس القرى الخاصة، والعقدة (11) وتمثل الإناث في مدارس المدن اللواتي تقدمن لامتحان ورقي، والعقدة (12) وتمثل الإناث في مدارس المدن

اللواتي تقدمن لامتحان إلكتروني، والعقدة (13) وتمثل الإناث في مدارس القرى. هذه المجموعات السبع هي نتاج التفاعل بين المتغيرات الأربعة؛ أي أن هناك أربعة متغيرات (الجنس وموقع المدرسة والسلطة المشرفة وطريقة تقديم الامتحان) أظهرت الفقرات أداءً تفاضليًا بالنسبة إليها. وكشفت النتائج هنا عن تفاعلات ثنائية وتفاعلات ثلاثية بين المتغيرات.

ومن أمثلة التفاعلات الثنائية التي لها ما يدعمها من شواهد واقعية أن هناك تفاعلاً بين الجنس والسلطة المشرفة (وزارة التربية، وكالة الغوث)؛ إذ تشكل السلطة المشرفة مصدرًا للأداء التفاضلي لدى الذكور، ولا تشكل مصدرًا للأداء التفاضلي لدى الإناث. كما أن هناك تفاعلاً بين أسلوب تقديم الاختبار (ورقي، إلكتروني) وموقع المدرسة حيث يشكل أسلوب تقديم الاختبار مصدرًا للأداء التفاضلي في مدارس المدن ولا يشكل مصدرًا للأداء التفاضلي في مدارس القرى. ومن الأمثلة على التفاعلات الثلاثية أن هناك تفاعلاً بين السلطة المشرفة والموقع والجنس؛ إذ تشكل السلطة المشرفة مصدرًا للأداء التفاضلي في مدارس القرى للذكور ولا تشكل مصدرًا للأداء التفاضلي في مدارس المدن للإناث. كما أن هناك تفاعلاً بين أسلوب تقديم الاختبار والجنس والموقع؛ إذ يشكل أسلوب التقديم مصدرًا للأداء التفاضلي لدى الذكور في مدارس المدن ولا يشكل مصدرًا للأداء التفاضلي لدى الإناث في مدارس القرى.

وبتفحص نتائج السؤال الثاني، نجد أن معدل الكشف عن الفقرات ذات الأداء التفاضلي يزداد بزيادة حجم العينة. وتتفق هذه النتائج مع نتائج دراسة ليو (Liu, 2017) التي لم تكشف فيها طريقة أشجار راش عن أي أداء تفاضلي في العينات التي يقل حجمها عن (600). ويعزى ذلك إلى زيادة الخطأ المعياري في تقدير معالم صعوبة الفقرات في العينات الصغيرة. كما تتفق مع نتائج دراسة أبو شندي وكاظم (Abu Shindi & Kazem, 2018) التي أظهرت حساسية طريقتي مانتل-هانزل ومنحنى خصائص الفقرة لحجم عينة المفحوصين. وتعزى هذه النتيجة إلى أن طريقة راش كغيرها من الطرق تقوم على الإحصاء التحليلي، لذلك يؤثر حجم العينة في القوة الإحصائية لعملية الكشف (Wood, 2011). ويعني ذلك أن هناك اتفاقًا على أن معدلات كشف DIF تزداد بازدياد حجم العينة. وتتميز طريقة راش عن الطرق الأخرى في أنها تكشف عن مصادر DIF التي تزداد كذلك بازدياد حجم العينة. وقد لوحظ ذلك بوضوح في الدراسة الحالية؛ إذ لم يتوفر أي مصدر للأداء التفاضلي عند حجم عينة مقداره (500) طالب وطالبة، وأصبح هناك أربعة مصادر عند حجم عينة مقداره (4000) طالب وطالبة. كما تتميز هذه الطريقة بكشفها عن التفاعلات بين المصادر المتعددة للأداء التفاضلي. وتجدر الإشارة إلى أن هذه النتائج لا توفر إجابة عن السؤال: "ما الحجم المناسب للعينة في حال استخدام أشجار راش للكشف عن الأداء التفاضلي؟"؛ إذ إن الإجابة عن مثل هذا السؤال تحتاج إلى قيود وضوابط من الصعب توافرها في البيانات الإمبريقية.

## التوصيات

في ظل النتائج التي توصلت إليها الدراسة الحالية، تمّ الخروج بالتوصيات الآتية:

- إجراء دراسات محاكاةٍ لتحديد حجوم العينات الأنسب مع اختلاف عدد المتغيرات المستخدمة في التحليل.
- إجراء دراسات عن الأداء التفاضلي لفقرات اختبار من النوع المقالي -متعدد الاستجابة - باستخدام طريقة أشجار راش.
- زيادة الحرص في تدقيق ومراجعة فقرات الاختبار الوطني، خاصة أنها تطبق على عينة كبيرة من الطلبة.

**References**


Abu Hammad, K. (2008). *Investigating the effectiveness of item characteristic curve in detecting differential functioning according to gender in a selected sample from a mathematics test*. Master Thesis, Yarmouk University. Irbid, Jordan.

Abu Shindi, Y. & Kazem, A. (2018). Sex differential item functioning for mathematics test in a cognitive development program in the Sultanate of Oman by Mantel-Haenszel and item characteristic curve method. *International Journal of Learning Management Systems, 6*(2), 61-73.

Aladaileh, S. (1996). *The influence of sex and educational level in confidence in learning mathematics and its relationship to academic achievement in mathematics for sixth and tenth grades in the government schools of Karak Governerate*. Master Thesis, Mutah University, Al-Karak, Jordan.

Al Ajlouni, J. & Sawalmeh, Y. (2016). Examining the differential distractors functioning of math test in the Jordanian national assessment of knowledge economy for sex and school location. *An-Najah University Journal for Research, 30*(11), 2265-2286.

Albano, A. D. & Rodriguez, M. C. (2013). Examining differential math performance by gender and opportunity to learn. *Educational and Psychological Measurement, 73*(5), 836-856.

Alomary, H. & Shatnawi, A. (2016). Differential item functioning of the national education quality control test in mathematics for $10^{th}$-grade students according to gender. *An-Najah University Journal for Research, 30*(8),1529-1554.

Alpursan, I. (2013). Gender-related differential item fumctioning for Jordanian national test to control the quality of education in mathematics for $10^{th}$-grade students. *Zagazig Journal, 28*(79), 1-35.

Altarawneh, M. (2011). *Investigating item bias in mathematics and science national tests for $8^{th}$-grade students.* Master Thesis, Mutah University, Jordan.

Altıntaş, Ö. & Kutlu, Ö. (2019). Investigating differential item functioning of Ankara University examination for foreign students by recursive partitioning analysis in the Rasch model. *International Journal of Assessment Tools in Education, 6*(4), 602-616.

Aryadoust, V. (2018). Using recursive partitioning Rasch trees to investigate differential item functioning in second-language reading tests. *Studies in Educational Evaluation, 56,* 197-204

Audeh, A. (2010). *Measurement and evaluation in the teaching process* ($14^{th}$ ed.). Dar Al-Amal.

Boehmke, B. & Greenwell, B. (2020). *Hands-on machine learning with R*. CRC Press.

Campbell, P. B. (1989). So, what do we do with the poor, non-white female? Issues of gender, race and social class in mathematics and equity. *Peabody Journal of Education, 66*(2), 95-112.

Chen, W. & Thissen, D. (1997). Local dependence indices for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*(3), 265-289.

Ertuna, L. & Gelbal. (2016). *An application of differential item functioning analysis with Rasch trees*. Retrieved from: https://www.researchgate.net/publication.

Frehat, N. (2014*). differential item functioning of the national educational quality control test in mathematics and science for $10^{th}$ grade according to mode of test presentation*. Ph.D. Dissertation, Yarmouk University, Irbid, Jordan.

Garner, M. & Engelhard Jr, G. (1999). Gender differences in performance on multiple- choice and constructed response mathematics items. *Applied Measurement in Education, 12*(1), 29-51.

Hattie, J. (1985). Methodology review: Assessing unidimensionality of tests and ltems. *Applied Psychological Measurement, 9*(2), 139-164.

Jordanian Ministry of Education. (2018). *Results of the national educational quality control test in mathematics for the year 2017/2018*. Amman, Jordan.

Komboz, B., Strobl, C. & Zeileis, A. (2018). Tree-based global model tests for polytomous Rasch models. *Educational and Psychological Measurement*, *78*(1), 128-166.

Kopf, J. (2013). *Model-based recursive partitioning meets item response theory: New statistical methods for the detection of differential item functioning and appropriate anchor selection.* PhD. Dissertation, University of München.

Liu, M. (2017). *Differential item functioning in large-scale mathematics assessments: Comparing the capabilities of the Rasch trees model to traditional approaches.* Ph.D. Dissertation, University of Toledo.

McGraw, R., Lubienski, S. T. & Strutchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race/ethnicity and socioeconomic status. *Journal for Research in Mathematics Education*, *37*(2), 129-150.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, *21*(1), 25-36.

Rover, C. (2005). *That's not fair! Fairnees, bias & differential item functioning in language testing*. Retrieved from the University of Hawai'i Systm. Website: www.hawai'i.edu.

RTI International. (2012). *Executive summary:* S*tudent performance in reading and mathematics, pedagogic practice, and school management in Jordan.* USAID/Jordan.

Shea, C. A., (2013). *Using a mixture IRT model to understand English learner performance on large-scale assessments*. PhD. Ddissertation, University of Massachusetts, Amherst, MA.

Stoneberg Jr, B. D. (2004). *A study of gender-based and ethnic-based differential item functioning (DIF) in the Spring 2003 Idaho standards achievement tests applying the simultaneous bias test (SIBTEST) and the Mantel-Haenszel chi square test*. Retrieved from: https://eric.ed.gov/?id=ED489949.

Strobl, C., Kopf, J. & Zeileis, A. (2010). *A new method for detecting differential item functioning in the Rasch model*. (Tech. Rep. No. 92). Department of Statistics at the University of Munich.

Strobl, C., Kopf, J. & Zeileis, A. (2015). Rasch trees: A new method for detecting differential item functioning in the Rasch model. *Psychometrika*, *80*(2), 289-316.

Taylor, C. S. & Lee, Y. (2012). Gender DIF in reading and mathematics tests with mixed-item formats. *Applied Measurement in Education*, *25*(3), 246-280.

Westers, P. & Kelderman, H. (1992). Examining differential item functioning due to item difficulty and alternative attractiveness. *Psychometrika*, *57*(1), 107-118.

Whitmore, M. L. (1995). *A comparison of two differential item functioning detection methods: Logistic regression and analysis of variance approach using Rasch estimation.* Ph.D. Dissertation, University of North Texas, Denton, TX.

Wood, S. (2011). *Differential item functioning procedures for polytomous items when examinee sample sizes are small.* Ph.D. Dissertation, University of Iowa, Iowa.

Zeileis, A., Hothorn, T. & Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, *17*(2), 492-514.

Zhang, Y. (2001). *Differential item functioning in a large-scale standardized mathematics assessment: The interaction of gender and ethnicity*. PhD. Dissertation, Ohio Univer-sity, Athens, OH.