الأداء التفاضلي لمفردات اختبار رافن للمصفوفات المتتابعة في ضوء نظرية الاستجابة للمفردة وفق متغيري النوع والتخصص

سعود بن شايش بشير العنزي
أستاذ مشارك في القياس والتقويم التربوي
جامعة الحدود الشمالية

(قُدِّم للنشر في 1445/02/05هـ، وقُبِل للنشر في 1445/02/28هـ)

مستخلص البحث : هدف البحث الحالي الكشفَ عن الأداء التفاضلي لمفردات اختبار رافن للمصفوفات المتتابعة في ضوء نظرية الاستجابة للمفردة وفق متغيري النوع والتخصص، تمثلت أداة البحث في اختبار رافن للمصفوفات المتتابعة الملونة، وتكونت عينة البحث من (1179) مشاركاً من طلاب وطالبات جامعة الحدود الشمالية بالمملكة العربية السعودية، بلغ المتوسط الحسابي لأعمارهم الزمنية (21,04) بانحراف معياري (1,57)، وتمثلت أداة البحث في اختبار رافن للمصفوفات المتتابعة الملونة، واعتمد الباحث على البرامج الإحصائية (SPSS؛ BILOG-MG)، واستخدم طريقة نسبة الأرجحية التي تتبع طرق نظرية الاستجابة للمفردة؛ للكشف عن الأداء التفاضلي لمفردات اختبار رافن للمصفوفات المتتابعة، وتوصل البحث إلى: وجود (7) مفردات في اختبار رافن للمصفوفات المتتابعة تظهر أداء تفاضلياً طبقاً للنوع (ذكور، إناث)، حيث أظهرت المفردات أرقام (11-18-21-33-34-35-36) أداءً تفاضلياً للذكور مقارنة بالإناث، حيث كانت هذا المفردات أصعب بالنسبة للإناث، ووجود (9) مفردات في اختبار رافن للمصفوفات المتتابعة تظهر أداءً تفاضلياً طبقاً للتخصص (علمي، أدبي)، حيث أظهرت المفردات أرقام (4-11-12-18-21-31-33-34-36) أداء تفاضلياً للتخصص العلمي مقارنة بالتخصص الأدبي، حيث كانت هذه المفردات أصعب بالنسبة للتخصص الأدبي.

الكلمات المفتاحية: الأداء التفاضلي للمفردة ـ نظرية الاستجابة للمفردة ـ نسبة الأرجحية ـ اختبار رافن للمصفوفات المتتابعة الملونة.

* * *

The Differential Functioning of Raven's Progressive Matrices Test in Light of Item Response Theory According to the Variables of Gender and Specialization

Saud Bin Shaish Basher Alenezi
Northern Border University

(Received 21/8/2023 ; accepted 13/9/2023)

ABSTRACT: The current research aimed at identifying the differential functioning of Raven's progressive matrices test item in the light of item response theory according to the variables of gender and specialization. The instruments of the research were Raven's colored progressive matrices test. The participants were 1179 male and female participants from Northern Border University in the Kingdom of Saudi Arabia. The mean for their ages was 21.04, with a standard deviation of 1.57. Using the statistical programs (SPSS-BILOG-MG) and the odds ratio method that follows the methods of item response theory to detect the differential functioning of Raven's progressive matrices test item, the results revealed that there are seven items (11-18-21-33-34-35-36) in Raven's progressive matrices test that show differential functioning based on gender (male, female). These items were more difficult for female students. In addition, there are nine items (4-11-12-18-21-31-33-34-36) in the Raven progressive matrices test showing differential functioning based on major (science, arts). These items were more difficult for the arts major students.

Keywords: item response theory, differential item functioning, odds ratio, Raven's colored progressive matrices.

(*) للمراسلة:

أستاذ مشارك، قسم علم النفس، جامعة الحدود الشمالية، مدينة عرعر، المملكة العربية السعودية.

(*) Corresponding Author:

Associate Professor, Department of Psychology, Faculty of Education and Arts,Northern Border University. Arar, Kingdom of Saudi Arabia

DOI: 10.12816/0061708

e-mail: dr.saud.shaish@gmail.com

مقدمة:

تقوم عمليات القياس والتقويم في مجال العلوم النفسية والتربوية على تحويل السمات الكامنة إلى مقادير كمية وفق قواعد محددة بحيث يسهل المقارنة بين الأفراد وترتيبهم وتصنيفهم وتوجيههم التوجيه الذي يتناسب مع مستويات قدراتهم بالنسبة للسمة المقيسة، ونتيجة لتعدد سمات الأفراد وتباينها وتشابكها؛ حاول علماء القياس بناء مقاييس واختبارات نفسية متنوعة تضاهي سمات هؤلاء الأفراد بحيث تتسم هذه الاختبارات والمقاييس بالدقة والموضوعية والصدق والثبات وغيرها من الخصائص التي يجب أن تتمتع بها أدوات القياس الجيدة.

وقد سيطرت النظرية الكلاسيكية في القياس وما يرتبط بها من نماذج وأساليب إحصائية على منهجيات القياس النفسي طوال القرن الماضي، وقدمت إطاراً مرجعياً لبناء وتقييم الاختبارات والمقاييس النفسية والتربوية، واستخراج خصائصها، وتفسير درجاتها ، وتستند هذه النظرية إلى عدد قليل من الافتراضات الخاصة بطبيعة البيانات الاختبارية، ولا تتأثر كثيراً بمخالفة هذه الافتراضات في أي من النماذج المنبثقة عنها.(علام، 2005)

ونظراً لاعتماد النظرية الكلاسيكية في القياس على مجموعة قليلة من الافتراضات الضعيفة؛ فقد أدي ذلك إلى ظهور كثير من أخطاء القياس في تقدير قدرات الأفراد وفق هذه النظرية، كحصول الفرد على نسبتي ذكاء مختلفتين مثلاً إذا تم اختباره باختبارين مختلفين يقيسان الذكاء لدى الفرد، وذلك لاختلاف درجة الخطأ المعياري في الحالتين.(Domino& Domino, 2006)

ومن أهم أوجه القصور التي تعاني منها النظرية الكلاسيكية، أن إحصاءات المفردات الاختبارية وخصائص الاختبار المستخدم، تعتمد بدرجة كبيرة على مجتمع البحث الذي طُبق عليه الاختبار، وبالتالي فإن مفردات الاختبار ستكون سهلة جداً إذا طُبقت على مجموعة من الأفراد مرتفعي القدرة، وستكون صعبة إذا طُبقت على مجموعة من الطلاب محدودي القدرة، ونتيجة لتأثر صعوبة المفردة بعينة الأفراد المختبرين؛ فإنه يكون من الصعب تقييم جودة المفردة الاختبارية.(Mayer & Zhu, 2013)

ونتيجة لأوجه القصور الموجودة في النظرية الكلاسيكية؛ فقد تركزت جهود علماء القياس والتقويم إلى تطوير نظرية معاصرة في القياس النفسي والتربوي عرفت باسم نظرية الاستجابة للمفردة Item Response Theory (IRT)، والتي من خلالها تم التغلب على العديد من أوجه قصور النظرية الكلاسيكية، وتُعد الميزة الأكثر أهمية في هذه النظرية هي استقلال إحصائيات المفردة الاختبارية (الصعوبة، التمييز، التخمين) عن قدرة المفحوصين المستخدمين لإيجاد هذه الإحصائيات، وكذلك استقلال تقدير قدرات المفحوصين عن إحصائيات المفردات المستخدمة في عملية التقدير.(الحربي، 2009)

وتوجد نماذج متعددة لنظرية الاستجابة للمفردة، وذلك لاختلاف الافتراضات المتعلقة بالبيانات الاختبارية، إذ يُفترض أن الأداء في الاختبار قد ينطوي على سمة أحادية البعد، أو على سمة متعددة الأبعاد، لذا يمكن تقسيم هذه النماذج إلى قسمين: أحدهما يسمى نماذج نظرية الاستجابة للمفردة أحادية البعد، والآخر يسمي نماذج نظرية الاستجابة للمفردة متعددة الأبعاد ، وينقسم كل منهما إلى نماذج ثنائية الاستجابة ونماذج متعددة الاستجابة،(علام،2005)، لذا توجد نماذج تتعلق بالمفردات ثنائية الاستجابة ونماذج تتعلق بالمفردات متعددة الاستجابة، وفي هذه النماذج فإن حصول المفحوص على درجة معينة على أحد المفردات، هي دالة وظيفية للقدرة الكامنة وراء استجابة المفحوص على هذه المفردة,Wolf) 2013)

ويمكن تمييز نماذج نظرية الاستجابة للمفردة ثنائية الاستجابة إلى ثلاثة نماذج: نموذج Rasch أحادي البرامتر، ونموذج Lord ثنائي

البارامتر، ونموذج Birnbaum ثلاثي البارامتر، وذلك بناء على عدد بارمترات المفردات المتمثلة في الصعوبة والتمييز والتخمين، ويعد النموذج أحادي البارامتر أبسط هذه النماذج، حيث يفرض أن جميع مفردات الاختبار تُميز بين الأفراد بنفس القدر لكنها تختلف في صعوبتها وبالتالي لا تتقاطع المنحيات المميزة لها، بينما يعتمد النموذج ثنائي البارامتر على بارامترين أساسيين، هما الصعوبة والتمييز، أما النموذج ثلاثي البارامتر فيفترض أن مفردات الاختبار تختلف في كل من الصعوبة والتمييز والتخمين.(Boyd, 2003؛ سليمان، 2009)

وتكمن السمة الرئيسة لنظرية الاستجابة للمفردة في النظر إلى كل مفردة على حدة دون الاعتماد على الدرجة الكلية، كما أنها تتميز بمبدأ عدم التغير أي أن بارامترات المفردات تكون مستقلة عن قدرات الأفراد، كما أن قدرات الأفراد تكون مستقلة عن بارامترات المفردات، وتتميز أيضاً بإمكانية مقارنة السمات الكامنة للأفراد من مجموعات مختلفة عندما تطبق عليهم اختبارات ومقاييس تحتوى على مفردات مشتركة، ومقارنة أفراد من نفس المجموعات عندما تقدم لهم اختبارات مختلفة. (Bortolotti et al, 2013)

وتقوم نظرية الاستجابة للمفردة على مجموعة من الافتراضات التي يجب تحققها في البيانات؛ والتي تتمثل في: افتراض أحادية البعد، والاستقلال الموضعي، ومنحنى خاصية المفردة، والتحرر من السرعة. Reise& Waller, (2003)، حيث تعتمد نماذج نظرية الاستجابة للمفردة على افتراض أحادية البعد Unidimensionality أي أن المفردات الاختبارية تقيس قدرة واحدة على متصل السمة والتي تتراوح نظرياً بين (-∞ : +∞)، يمكن التحقق من ذلك عن طريق التحليل العاملي الاستكشافي على مستوى مفردات الاختبار، حيث يُعتبر الاختبار أحادي البعد إذا كانت النسبة بين الجذر الكامن للعامل الأول والجذر الكامن للعامل الثاني أكبر من (2), أو من خلال كون العامل الأول يفسر نسبة (20%) من نسبة التباين الكلي الذي يفسره جميع العوامل المستخرجة من التحليل العاملي، والتي يكون قيمة الجذر الكامن لها أكبر من الواحد الصحيح. (Krause, 2012؛ Georgiev, 2008)

كما يتطلب تطبيق نظرية الاستجابة للمفردة على البيانات توافر افتراض الاستقلال الموضعي Local Independence والذي يعني أن احتمال الاستجابة الصحيحة للفرد على أحد مفردات الاختبار مستقلاً عن احتمال استجابته على أي مفردة أخرى عند التحكم في قدرة الفرد ومعالم المفردة، والاختبار لا يكون أحادي البعد دون أن تتوفر فيه خاصية الاستقلال الموضعي، ولكن يمكن أن يتوفر في الاختبار خاصية الاستقلال الموضعي حتى في حالة كون الاختبار يقيس سمتين كامنتين أو بعدين مختلفين.(Vincent, 2002)، ويمكن التحقق من هذا الافتراض باستخدام مؤشر (Q3) التي اقترحته Yen (1984)، والذي يتطلب أن تكون جميع معاملات الارتباط بين أزواج مفردات الاختبار أقل من نقطة القطع التي اقترحتها "ين"، وهي القيمة المطلقة للعدد (0،2)، ويكون متوسط قيم مؤشر (Q3) أقل من القيمة المتوقعة للمتوسط الحسابي للتوزيع العيني لهذا المؤشر. (التقي، 2009؛ النفيعي،2012)

ويمثل افتراض المنحنى المميز للمفردة Item Characteristic Curve أحد أهم أسس نظرية الاستجابة للمفردة، وهو عبارة عن دالة رياضية تربط بين احتمال نجاح الفرد في الإجابة على مفردة اختبارية وبين القدرة التي تقيسها مجموعة المفردات التي يشتمل عليها الاختبار، ولذلك يُعبر المنحنى المميز للمفردة عن انحدار الدرجة التي يحصل عليها الفرد في إحدى المفردات على قدرته.(الوليلي، 2002)

أما افتراض التحرر من السرعة فيمكن تقديره عن طريق معرفة عدد الأفراد الذين لم

يتمكنوا من الانتهاء من الإجابة عن جميع مفردات الاختبار الذي طُبق عليهم لمعرفة كم نسبة الذين استجابوا على جميع المفردات، فالاختبار الذي يعتمد على السرعة لا يخُل بافتراض أحادية البعد فقط، بل أنه قد يؤثر في تقدير معالم نماذج نظرية الاستجابة للمفردة، وبشكل خاص المفردات التي تقع في نهاية الاختبار؛ وذلك لأن بعض الأفراد سيتركون بعض المفردات بدون إجابة. (Dawber, 2004؛ علام، 2005)

ويتضح مما سبق، أنه لتطبيق نماذج نظرية الاستجابة للمفردة على بيانات الاختبارات والمقاييس المراد تدريجها، يجب التحقق من افتراضات النظرية المتمثلة في: (أحادية البعد، الاستقلال الموضعي، المنحنى المميز للمفردة، التحرر من السرعة)، حتى يُعطي هذا التدريج نتائج دقيقة وموضوعية وذلك بالاعتماد على نماذج نظرية الاستجابة للمفردة والتي تختلف باختلاف افتراض كل نموذج، حيث يفترض النموذج الأحادي اختلاف المفردات في بارامتر الصعوبة فقط، والنموذج الثنائي يفترض اختلاف المفردات في بارمتري الصعوبة والتمييز، أما النموذج الثلاثي فيفترض اختلاف المفردات في بارمترات الصعوبة والتمييز والتخمين، لذا ينبغي اختيار النموذج الأنسب لتدريج مفردات الاختبار.

وقد ظهرت في العقود الأخيرة تطبيقات عملية متعددة لنظرية الاستجابة للمفردة في مجال القياس والتقويم النفسي والتربوي، منها: تصميم الاختبارات والمقاييس وتدريج مفرداتها، وبناء بنوك الأسئلة، وإعداد اختبارات المواءمة المحوسبة، ومعادلة الاختبارات والمقاييس النفسية، والكشف عن الأداء التفاضلي لمفردات الاختبار.(Embretson& ؛ Shyu, 2001 Reise, 2000)

ويحدث الأداء التفاضلي للمفردة Differential Item Functioning (DIF) عندما يكون لمجموعة من الأفراد الذين طُبق عليهم الاختبار احتمالات مختلفة في الإجابة عن المفردة عن مجموعة أخرى، حيث قد يرجع ذلك إلى كون المفردة تقيس شيئاً آخر بجانب السمة المراد قياسها، وأن هذه المفردة متحيزة لبعض الأفراد المختبرين، وهو ما يهدد الصدق البنائي للاختبار.(Walker, 2011)

فالأداء التفاضلي يمثل دالة إحصائية للتعبير عن التباين في الاستجابة على المفردة لمجموعة من الأفراد عند مستوى قدرة معين نتيجة لوجود فروق في الاستجابة ترجع إلى متغيرات أخرى تجعل هذه المفردة متحيزة إلى مجموعة من المختبرين دون غيرها.(stoneberg, 2004) أي أن الأداء التفاضلي أحد المفاهيم الإحصائية التي تسعى لجعل المقاييس والاختبارات أقل تحيزاً لفئة معينة أو مجموعة ما، وجعل الفروق الكامنة بين المجموعات المدروسة عائداً إلى مستوى القدرة أو السمة المقاسة وليس إلى أي شيء آخر، ويشار إلى هذه المجموعات عادة بالمجموعة المستهدفة (وهي المجموعة التي تتأثر بالأداء التفاضلي للمفردات بمعنى أن المفردات تتحيز ضدها وتمثل الأقلية) والمجموعة المرجعية (وهي المجموعة المقارنة وعادة تكون الأكثر عدداً). (ضعضع، 2017)

ويمكن التمييز بين نوعين من الأداء التفاضلي للمفردة، هما: الأداء التفاضلي المنتظم والأداء التفاضلي غير المنتظم، ويحدث الأداء التفاضلي المنتظم عندما تظهر احتمالات مختلفة في صعوبة المفردة لمجموعتين من الأفراد في كل مستويات القدرة، بمعني أن أداء إحدى المجموعتين على مفردة ما يكون أفضل من المجموعة الأخرى عند جميع مستويات القدرة، أما الأداء التفاضلي غير المنتظم فيحدث عندما تظهر احتمالات مختلفة في صعوبة المفردة لمجموعتين من الأفراد عند مستوى معين من القدرة، بمعني أن أداء إحدى المجموعتين على مفردة ما يكون أفضل من الأخرى عند مستوى قدرة معين ويختلف الفرق بين المجموعتين عند مستوى قدرة آخر. Lai et al., 2005 ؛ Ozdemir, 2015)

وتوجد عدة طرق للكشف عن الأداء التفاضلي للمفردات، منها ما يعتمد على النظرية الكلاسيكية في القياس مثل طريقة مانتل-هانسزل، وطريقة الانحدار اللوجستي، ومنها ما يعتمد على نظرية الاستجابة للمفردة مثل طريقة نسبة الأرجحية وطريقة مربع كاي للورد، وحسب منظور نظرية الاستجابة للمفردة؛ فإن وجود الأداء التفاضلي في مفردة ما يعني أن تقديرات بارامترات هذه المفردة لا توصف باللاتغاير في المجموعات المدروسة (أي لا توجد مطابقة في بيانات المفردة)، فالافتقار إلى اللاتغاير في أحد مفردات الاختبار هو الذي يفسر كدليل على وجود الأداء التفاضلي وفق نظرية الاستجابة للمفردة، (دي أيالا، 2017)

ونظراً لأن نظرية الاستجابة للمفردة تعتمد على افتراضات قوية مقارنة بالنظرية الكلاسيكية أهمها أن هناك استقلالية للمفردات الاختبارية في قياس القدرات المختلفة، وأن لكل مفردة اختبارية من مفردات الاختبار خصائصها السيكومترية الخاصة بها، والمُمثلة في منحنى مُميز لها يصف احتمالية الإجابة بناءً على قدرة الفرد وخصائص المفردات. (Magno, 2009)، فإن هذا يسهم بشكل كبير في تميُز طرق الكشف عن الأداء التفاضلي وفق نظرية الاستجابة للمفردة عن الطرق المستندة إلى النظرية الكلاسيكية في القياس، حيث تستند المقارنة بين المجموعات وفق نظرية الاستجابة للمفردة على مقارنات تتصف بعدم التغير في البارامترات المقدرة، (رشوان، 2015)

ويُعد اختبار نسبة الأرجحية Likelihood Ratio test (وهي الطريقة التي اعتمد عليها البحث الحالي) من أكثر الطرق استخداماً للكشف عن الأداء التفاضلي للمفردة في ضوء نظرية الاستجابة للمفردة، والتي تعتمد على دلالة الفروق بين بارامتري المفردة في المجموعتين المستهدفة والمرجعية، والتي نبعت فكرتها من أفكار لورد في استخدامه لمربع كاي في المقارنة بين بارامترات المفردة للمجموعات المختلفة بالاستناد إلى أن تماثل المنحنى المميز للمفردة لأي مجموعتين يعتمد على تماثل بارامترات هذه المفردة في المجموعتين. (Bechger& Kim& Oshima, 2012؛ Maris, 2015)

وتتميز طريقة نسبة الأرجحية بأنها تختبر الأداء التفاضلي للمفردة بطريقة مباشرة تمكن من اكتشاف الأداء التفاضلي المرتبط ببارامترات صعوبة المفردة وتمييزها ونسبة تخمينها، كما أنها تعبر عن الأداء التفاضلي للمفردة بوحدات كمية واضحة المعني يسهل تفسيرها، بالإضافة إلى أنها لا تعتمد على حساب المجموع الكلي لدرجات الاختبار.(Thissen, 2001)

وتتم هذه الطريقة في ثلاث خطوات: تتمثل الخطوة الأولى في حساب قيمة الترجيح اللوغاريتمي في حالة تدريج مفردات الاختبار في حالة اعتبار المجموعتين مجموعة واحدة حسب نموذج نظرية الاستجابة للمفردة المستخدم، وتتمثل الخطورة الثانية في حساب قيمة الترجيح اللوغاريتمي في حالة اعتبار وجود مجموعتين إحداهما مرجعية وأخرى مستهدفة، أما الخطوة الثالثة فتتمثل في حساب الفرق بين نسبة الأرجحية في الحالتين والحكم على دلالتها حسب توزيع كا^2 ، فإذا كانت دالة إحصائياً فيُعد مؤشراً على وجود أداء تفاضلي في مفردات الاختبار، وللكشف عن المفردات التي تظهر أداء تفاضلي عن طريق حساب الفرق بين بارمترات المفردات في المجموعتين وقسمته على الخطأ المعياري للفرق، ويتبع الإحصائي الناتج التوزيع الطبيعي، فإذا كان الفرق دال إحصائياً؛ فهذا يشير إلى أن المفردة تظهر أداء تفاضلياً لإحدى المجموعتين المرجعية أو المستهدفة.(دي أيالا، 2017؛ ضعضع، 2017)

ويتضح مما سبق أهمية الكشف عن الأداء التفاضلي للمفردات الاختبارية، حتي يتحقق نوع من العدالة وعدم التحيز في تقدير قدرات الأفراد المختبرين، أي أن الفروق بين المجموعات

المختبرة هي فروق حقيقية ترجع إلى الاختلاف في السمة المقاسة وليست بسبب خصائص فرعية تؤثر في الاستجابة على مفردات الاختبار كالنوع والتخصص واللغة والإقامة.

ويعد اختبار المصفوفات المتتابعة لرافن من اختبارات الذكاء التي يشيع استخدامها في تقدير القدرة العامة لدى الأفراد. (عودة وعبيدات، 2013)، وقد ظهر اختبار المصفوفات المتتابعة لرافن في انجلترا عام (1983) لقياس العامل العام عند سبيرمان ويتطلب في جوهره إدراك العلاقات بين وحدات مجردة، وهو في رأي معظم علماء النفس في بريطانيا أفضل مقاييس العامل العام، وهو أحد اختبارات الذكاء المتحررة من أثر الثقافة Culture- Free أي الاختبارات التي تصلح لأغراض المقارنة بين الثقافات المختلفة.(أبو حطب وآخرون، 2008)

وقد اهتمت بعض البحوث السابقة بالكشف عن الأداء التفاضلي لمفردات اختبار رافن للمصفوفات المتتابعة في ضوء بعض المتغيرات وباستخدام طرق متعددة، حيث هدف بحث (Arendasy& Sommer, 2012) إلى التحقق من نتائج البحوث التي أظهرت فروقًا في الذكاء لصالح الذكور ترجع إلى فروق حقيقية أم ترجع لوجود أداء تفاضلي لصالح الذكور، وذلك بفحص ثلاثة بحوث استخدمت مصفوفات رافن المتتابعة، وتوصلت نتائج البحث إلى أن تعقد شكل المصفوفة وغموض قاعدة الحل تمثلان سببين رئيسين لزيادة ظهور الأداء التفاضلي للمفردات لصالح الذكور.

وهدف بحث (Chiesi et al., 2012) إلى التحقق من الأداء التفاضلي لاختبار رافن للمصفوفات المتتابعة النسخة المختصرة، والتي تكونت من (12) مفردة، وذلك وفق عدة متغيرات تصنيفية منها النوع (ذكور، إناث)، وقد توصلت نتائج البحث إلى عدم وجود أداء تفاضلي لمفردات النسخة المختصرة من اختبار رافن للمصفوفات المتتابعة وفق متغير النوع (ذكور، إناث).

وهدف بحث (شلبي، 2015) إلى وصف وفهم طبيعة الفروق في الذكاء بين الجنسين في عمر (13-19) باستخدام اختبار رافن للمصفوفات المتتابعة، والتحقق من تكافؤ القياس لاختبار رافن للمصفوفات المتتابعة بين الذكور والإناث، وتوصلت نتائج البحث إلى وجود فروق دالة إحصائية بين الذكور والإناث في الذكاء في عمر السابعة عشر لصالح الإناث، ووجود فروق دالة إحصائية بين الذكور والإناث في الذكاء في عمر الثامنة عشر لصالح الذكور، كما توصلت الدراسة إلى وجود تكافؤ قياس بنوعيه الشكلي والقوي لاختبار رافن للمصفوفات المتتابعة بين الذكور والإناث.

كما هدف بحث (على وآخرون، 2016) إلى الكشف عن الأداء التفاضلي لاختبار رافن للمصفوفات الملونة حسب العمر على دقة المعادلة العمودية لنموذجي الاختبار، وتوصلت نتائج البحث إلى وجود (18) مفردة ذات أداء تفاضلي، منها (8) مفردات في الصورة الأولى، و(10) مفردات في الصورة الثانية، كما توصلت النتائج إلى أن حذف المفردات ذات الأداء التفاضلي يحسن دقة المعادلة العمودية لنموذجي الاختبار.

وهدف بحث (Shibaev et al., 2020) اختبار دلالة الفروق بين العرقيتين الياقوتية والروسية على اختبار رافن للمصفوفات المتتابعة، والكشف عن الأداء التفاضلي لمفردات الاختبار بطريقة الانحدار اللوجستي، وقد أظهرت نتائج البحث وجود فروق دالة إحصائياً بين العرقيتين الياقوتية والروسية لصالح العرقية الياقوتية، كما كشف البحث عن وجود أداء تفاضلي لخمس مفردات من الاختبار، وبحذف المفردات ذات الأداء التفاضلي بقيت الفروق بين العرقيتين الياقوتية والروسية.

كما هدف بحث (حماد، 2021) الكشف عن الأداء التفاضلي لمصفوفات رافن المعيارية تبعاً للنوع (ذكور إناث) باستخدام طريقة مانتل هانزل التي تتبع النظرية الكلاسيكية، وتحديد هل

يؤثر وجود المفردات ذات الأداء التفاضلي على نتائج القياس، وتوصلت نتائج البحث إلى وجود (6) مفردات ذات أداء تفاضلي، أربعة منها متحيزة للذكور واثنتان متحيزتان للإناث، كما توصلت النتائج إلى أن مقياس رافن المعياري لا يتأثر بوجود الأداء التفاضلي للمفردات حيث لم يتغير اتجاه الفروق قبل وبعد حذف هذه المفردات.

وفي ضوء ما سبق يتضح اهتمام البحوث السابقة بالكشف عن الأداء التتفاضلي لمفردات اختبار رافن للمصفوفات المتتابعة تبعا لمتغيرات تصنيفية، معظمها اهتمت بتأثير النوع (ذكور، إناث)، كما يتضح من هذه البحوث اعتمادها على طرق الكشف عن الأداء التفاضلي وفق النظرية الكلاسيكية كما في دراسة (حماد، 2021)، والتحليل العاملي التوكيدي كما في دراسة (شلبي، 2015)، وقد تباينت نتائج هذه الدراسات والبحوث حول وجود مفردات ذات أداء تفاضلي في اختبار رافن للمصفوفات المتتابعة، ويتفق هدف البحث الحالي مع بعض أهداف هذه الدراسات في اهتمامه بالكشف عن الأداء التفاضلي لمفردات اختبار رافن للمصفوفات المتتابعة وفق متغيري النوع والتخصص، بالاعتماد على إحدى طرق نظرية الاستجابة للمفردة المستخدمة في الكشف عن الأداء التفاضلي، وذلك بعد التحقق من افتراضات النظرية في بيانات التطبيق وتدريج مفردات الاختبار وفق نموذج نظرية الاستجابة للمفردة المناسب للبيانات، لذا يسعى البحث الحالي إلى الكشف عن الأداء التفاضلي لمفردات اختبار رافن للمصفوفات المتتابعة في ضوء نظرية الاستجابة للمفردة وفق متغيري النوع والتخصص.

مشكلة البحث:

يشير كل من Bortolotti et al., (2013) و Avcu (2021) إلى أنه نتيجة لوجود العديد من أوجه القصور في تقدير خصائص مفردات الاختبارات بالاعتماد على النظرية الكلاسيكية؛ فإن استخدام نظرية الاستجابة للمفردة في إعداد وتطوير الاختبارات والمقاييس والتحقق من صدقها وثباتها وتحديد بارامتراتها يعد اتجاهاً عالمياً، ويستخدم على نطاق واسع في مجال التربية، فهي تقدر قدرة الأفراد بدقة مع مراعاة بارامترات مفردات الاختبار (الصعوبة، والتمييز، التخمين)، كما أنها تسمح بتدريج المفردات والأفراد على نفس الميزان مما يسمح بإجراء مقارنات موضوعية بين الأفراد، وبين المفردات، وبين المفردات والأفراد معاً.

ويرى Holmefur et al., (2015) أنه ينبغي عدم الاكتفاء باستخدام أساليب النظرية الكلاسيكية في التحقق من الخصائص السيكومترية للاختارات والمقاييس النفسية، بل يجب أن يمتد التحليل إلى الاعتماد على نماذج نظرية الاستجابة للمفردة، وبخاصة الاختبارات التي تستخدم على نطاق واسع في البحوث النظرية والجوانب التطبيقية في تشخيص وتصنيف قدرات الأفراد.

وتعد اختبارات الذكاء من أكثر الاختبارات شيوعاً واستخداماً في مجال تشخيص وتصنيف قدرات الأفراد، ونظراً لتعدد النظريات المفسرة للذكاء؛ فقد تعددت الاختبارات التي تقيسه ، كما اهتم علماء النفس ببناء اختبارات متحررة من أثر الثقافة، والتي تسعى إلى التحكم في الأبعاد التي تتألف منها الثقافة، والمقارنة بين أفراد من ثقافات مختلفة، وتعد اختبارات رافن للمصفوفات المتتابعة أحد الاختبارات المتحررة من أثر الثقافة والتي أعدها العالم الانجليزي رافن J, Raven عام (1938) وقد استمر هو وتلاميذه أكثر من ثلاثين عاماً في تطوير هذه المصفوفة حتى وفاته عام (1970)، وتعد مصفوفات رافن من اختبارات الذكاء غير اللفظي التي تخلو من أثر الثقافة إلى حد كبير وتعتمد أساساً على التطبيق الجمعي، كما يمكن أن تطبق فردياً في ظروف معينة. (أبو حطب وآخرون، 2008؛ علي، 2016)

ويشيع استخدام اختبار المصفوفات المتتابعة لرافن لقياس القدرة العلقية العامة لدى الأفراد لأغراض تصنيفية وتشخيصية، لذا يعد الكشف عن الأداء التفاضلي لمفرداته إجراءً

ضرورياً للتحقق من عدالة المقياس، ففي الآونة الأخيرة أصبح فحص الأداء التفاضلي لمفردات الاختبارات والمقاييس للحصول على مفردات جيدة لقياس السمة المحددة مطلباً مهماً في بناء الاختبار وحساب خصائصه السيكومترية والتحقق من عدالته في القياس والتصنيف. Ismat& Sidiqui, (2015 ؛ حماد، 2021)

وبالرغم من أن العديد من الدراسات أشارت إلى تماثل القدرة العقلية العامة لدى كل من الذكور والإناث(Halpern, 2012 ؛Savage-McGlynn, 2012؛ الحسن، 2017) إلا أن نتائج بعض الدراسات التي استخدمت اختبار رافن للمصفوفات المتتابعة أشارت إلى وجود فروقٍ في الذكاء لصالح الإناث مثل (شلبي، 2015؛ Qiu et al,, 2020)، وبعضها كانت الفروق لصالح الذكور مثل(Van der Elst, 2013)، وربما ترجع هذه الفروق بين الذكور والإناث إلى وجود أداء تفاضلي لمفردات الاختبار، حيث أشارت دراسة حماد (2021) إلى وجود (6) مفردات في اختبار رافن للمصفوفات المتتابعة وذلك بالاعتماد على طريقة مانتل-هانسزل تظهر أداءً تفاضلياً منها أربعة متحيزة للذكور واثنتان متحيزتان للإناث.

ويشير French et al., (2012) أن الاختلافات في نوعية الخبرات التعليمية التي يتعرض لها الفرد في المؤسسات التعليمية قد يكون لها تأثير مختلف في نمو بعض المتغيرات المعرفية بصفة عامة، ويعد الذكاء العام أحد هذه المتغيرات التي قد يلعب التخصص دوراً في إظهار فروق في الذكاء، وقد يكون السبب في ظهور الفروق في الذكاء بين التخصصات العلمية والأدبية راجعاً إلى أن بعض مفردات الاختبار تظهر أداء تفاضلياً ضد طلاب أحد التخصصين.

ويتضح مما سبق أهمية الكشف عن الأداء التفاضلي للاختبارات التي تصمم لقياس القدرة العقلية العامة لدى الأفراد والتي تستخدم في أغراض التشخيص والتصنيف، ونتيجة لكون اختبار المصفوفات المتتابعة لرافن أحد الاختبارات التي تستخدم على نطاق واسع في قياس القدرات العقلية كونه أحد اختبارات الذكاء المتحررة من أثر الثقافة، فإن البحث الحالي يسعى إلى الكشف عن الأداء التفاضلي لمفردات اختبار رافن للمصفوفات المتتابعة في ضوء نظرية الاستجابة للمفردة وفق متغيري النوع والتخصص.

وبناءً على ما سبق يمكن صياغة مشكلة البحث الحالي التساؤلات التالية:

1- هل توجد مفردات ذات أداء تفاضلي في اختبار رافن للمصفوفات المتتابعة ترجع للنوع (ذكور، إناث)؟
2- هل توجد مفردات ذات أداء تفاضلي في اختبار رافن للمصفوفات المتتابعة ترجع للتخصص(علمي، أدبي)؟

هدف البحث:

يهدف البحث الحالي إلى الكشف عن الأداء التفاضلي لمفردات اختبار رافن للمصفوفات المتتابعة في ضوء نظرية الاستجابة للمفردة وفق متغيري النوع والتخصص.

أهمية البحث:

يمكن أن تتضح أهمية البحث الحالي في النقاط التالية:

1- إبراز أهمية الكشف عن الأداء التفاضلي لمفردات الاختبارات والمقاييس والذي يعد من أهم المشكلات التي تهدد صدق هذه الاختبارات والقرارات المترتبة عليها.
2- تأتي الدراسة الحالية استجابة لما تنادي به التوجهات الحديثة من ضرورة الكشف عن الأداء التفاضلي لمفردات الاختبارات المستخدمة في تشخيص وتصنيف الأفراد كإجراء أساسي للتحقق من عدالة الاختبار وعدم تحيزه إلى مجموعة دون أخرى.
3- تقديم دلالات إمبريقية يمكن على أساسها تحديد مفردات اختبار رافن للمصفوفات

المتتابعة التي قد تبدي أداءً تفاضلياً وفق متغيري النوع والتخصص.

التعريف الإجرائي لمصطلحات البحث:

1- الأداء التفاضلي للمفردة: Differential Item Functioning (DIF)

الفروق السيكومترية في أداء المفردة لدى مجموعتين مختلفتين، والذي يحدث عندما تظهر مجموعتين متكافئتين في السمة المقاسة احتمالات مختلفة للإجابة الصحيحة على نفس المفردة الاختبارية. (Woods, 2008)، ويعرف إجرائياً بأنه اختلاف احتمالات ظهور الاستجابة الصحيحة على مفردات اختبار رافن للمصفوفات المتتابعة باختلاف النوع (ذكور، إناث)، أو التخصص (علمي، أدبي) في ضوء نظرية الاستجابة للمفردة.

2- نظرية الاستجابة للمفردة: Item Response Theory

مجموعة من النماذج الرياضية التي تحدد العلاقة الاحتمالية بين الاستجابات الملاحظة على مفردات الاختبار، والسمات الكامنة المسببة لها، والتي تُحدد استجابات الأفراد على مفردات هذا الاختبار، وتظهر هذه النماذج بيانياً من خلال دوال رياضية تربط احتمال الإجابة الصحيحة على المفردة الاختبارية بالقدرة التي تقيسها هذه المفردات.

حدود البحث:

تتمثل حدود البحث الحالي في اختبار رافن للمصفوفات المتتابعة الملونة والمكون من (36) مصفوفة المستخدمة في البحث الحالي، والنموذج ثنائي البارامتر المستخدم في تدريج مفردات الاختبار، وطريقة نسبة الأرجحية المستخدمة في الكشف عن الأداء التفاضلي للمفردة في ضوء نظرية الاستجابة للمفردة، والبرامج الإحصائية المستخدمة في البحث الحالي، وطلاب مرحلة البكالوريوس في كليات جامعة الحدود الشمالية في مدينة عرعر للعام الجامعي 2023/2022 التي اشتقت منهم عينة البحث.

منهج البحث:

استخدم الباحث المنهج الوصفي التحليلي لمناسبته لأهداف البحث الحالي؛ وذلك للكشف عن الأداء التفاضلي لمفردات اختبار رافن للمصفوفات المتتابعة في ضوء نظرية الاستجابة للمفردة وفق متغيري النوع (ذكور، إناث) والتخصص (علمي، أدبي).

مجتمع البحث وعينته:

تألف مجتمع البحث من طلاب مرحلة البكالوريوس في جامعة الحدود الشمالية بمدينة عرعر للعام الجامعي2023/2022، وبلغ عدد المشاركين في البحث الحالي (1179) مشاركاً من طلاب وطالبات جامعة الحدود الشمالية بالمملكة العربية السعودية، بلغ المتوسط الحسابي لأعمارهم الزمنية (21,04) بانحراف معياري (1,57)، والجدول التالي يوضح وصف عينة البحث في ضوء متغير النوع والتخصص:

جدول (1)
وصف عينة البحث حسب متغير النوع والتخصص

النوع	التخصص		المجموع
	علمي	أدبي	
ذكور	327	307	634
إناث	291	254	545
المجموع	618	561	1179

أداة البحث:

تمثلت أداة البحث في اختبار رافن للمصفوفات المتتابعة الملونة Coloured Progressive Matrices التي أعدها العالم الانجليزي J, Raven لقياس الذكاء عام 1947 وتم تعديلها عام 1056، (تعريب وتقنين: علي، 2016)، وتتكون من (26) مفردة موزعة على ثلاثة أقسام (أ- أب- ب)، كل قسم مكون من (12)

مفردة، ويتكون كل مفردة في المصفوفات من شكل أو نمط رئيسٍ اقتطع منه جزء معين وأسفله ستة أجزاء يختار من بينها المفحوص الجزء الذي يكمل الفراغ في الشكل الأساسي.

وتعتمد مشكلات القسم (أ) على قدرة الفرد على إكمال الأنماط المستمرة، وقرب نهاية المجموعة يغير نمط الاستمرار على أساس بعدين في نفس الوقت، ويعتمد القسم (أب) على قدرة الفرد على إدراك الأشكال المنفصلة في نمط كلي على أساس الارتباط المكاني، أما القسم (ب) فيعتمد حل مشكلاته على فهم القاعدة التي تحكم التغيرات في الأشكال المرتبطة منطقياً أو مكانياً، ويصحح الاختبار بإعطاء درجة واحدة للإجابة التي تتفق مع مفتاح التصحيح، والإجابة صفر للإجابة الخاطئة.(علي، 2016)

وقد استخدمت عدة أساليب للتحقق من صدق الاختبار في صورته الأصلية، منها: الصدق العاملي، والصدق التنبؤي، والصدق التلازمي، وذلك بحساب معاملات الارتباط مع اختبار ستانفورد بينيه واختبار ويكسلر واختبار رسم الرجل، وقد تراوحت قيم معاملات الارتباط بين (0,32: 0,86) وجميعها كانت قيما دالة إحصائياً عند مستوى دلالة (0,01)، وقد قام (علي، 2016) بالتحقق من صدق الاختبار على عينة عمرية تراوحت بين (5,5: 68,4) باستخدام الصدق التلازمي عن طريق حساب معاملات الارتباط مع المقاييس الفرعية لاختبار ويكسلر ومتاهات بورتيوس ولوحة سيجان، وقد جاءت جميع معاملات الارتباط دالة إحصائياً عند مستوى دلالة (0,01).

كما استخدمت عدة طرق للتحقق من ثبات المقياس في صورته الأصلية، منها: إعادة التطبيق، والتجزئة النصفية، والاتساق الداخلي بين الأقسام الفرعية للاختبار، وقد أشارت جميع قيم معاملات ثبات الاختبار إلى تمتع الاختبار بقيم ثبات مرتفعة، وقد قام (علي، 2016) بالتحقق من ثبات المقياس باستخدام معادلة كيودر وريتشاردسون، وقد بلغ قيمة معامل الثبات (0,85).

التحقق من افتراضات نظرية الاستجابة للمفردة:

تم التحقق من عدم وجود بيانات تامة أو صفرية، حيث قام الباحث بفحص استجابات الأفراد للكشف عن المفردات التي تمت الإجابة عنها إجابة صحيحة من قبل جميع الأفراد، أو الأفراد الذين أجابوا عن جميع المفردات إجابات صحيحة أو خاطئة، ولم يُسفر الفحص الأولي لاستجابات الأفراد عن استبعـــــاد أي فـــرد أو أي مفردة من مفردات الاختبار، وفيما يلي إجراءات التحقق من افتراضات نظرية الاستجابة للمفردة:

أ- التحقق من أحادية البعد:

تم التحقق من افتراض أحادية البعد باستخدام التحليل العاملي الاستكشافي بطريقة المكونات الأساسية لاستجابات الأفراد مفردات اختبار رافن للمصفوفات المتتابعة، حيث قام الباحث بحساب مدى ملاءمة حجم العينة لإجراء التحليل العاملي باستخدام اختبار K,M,O Test وقد بلغت قيمته في تحليل هذا الاختبار (0,700)، وبالتالي فإنه يمكن الحكم بكفاية حجم العينة لإجراء هذا التحليل، ثم قام بإجراء التحليل العاملي لمصفوفة الارتباط بطريقة المكونات الأساسية، وقد أخذ الباحث بمحك جيلفورد لمعرفة حد الدلالة الإحصائية للتشبعات الناتجة وهو اعتبار التشبعات التي تصل إلى (0,30) أو أكثر تشبعات دالة، وفي ضوء نتائج التحليل العاملي أمكن استخلاص مجموعة من العوامل الرئيسة الجذر الكامن لكل منهما أكبر من الواحد الصحيح، والجدول التالي يوضح ذلك الجذور الكامنة والتباين المفسر لبعض العوامل المستخرجة من التحليل العاملي:

جدول (2)

قيم الجذور الكامنة ونسبة التباين المفسر والتباين المفسر التراكمي لاختبار رافن للمصفوفات المتتابعة الملونة

الاختبار	الجذر الكامن للعوامل			نسبة التباين المفسر للعوامل		
	الأول	الثاني	النسبة بينهما	الأول	التراكمي	النسبة بينهما
رافن للمصفوفات المتتابعة	9,253	2,459	3,762	25,702	64,219	%40,022

يتضح من جدول (2) أن نتائج التحليل العاملي الاستكشافي أسفرت عن تشبع مفردات اختبار رافن على عامل واحد فقط، حيث إن النسبة بين الجذر الكامن للعامل الأول والجذر الكامن للعامل الثاني جاءت أعلى من القيمة (2)، كما أن العامل الأول يفسر نسبة أكبر من (20%) مقارنة بنسبة التباين الكلي الذي يفسره جميع العوامل المستخرجة من التحليل العاملي، مما يشير إلى تحقق افتراض أحادية البعد في بيانات اختبار كاتل للذكاء.

ب- التحقق من الاستقلال الموضعي:

قام بالتحقق من افتراض الاستقلال الموضعي بالاعتماد على نتائج التحليل العاملي الاستكشافي لمفردات الاختبار لفحص مصفوفة معاملات الارتباط بين مفردات الاختبار والتي تراوحت ما بين (0,014: 0,306) وهي قيم معاملات ارتباط منخفضة، مما يشير إلى عدم وجود مشكلة الاعتماد الخطي بين مفردات المقياس، والذي يعد مؤشراً على تحقق افتراض الاستقلال الموضعي.

كما قام بحساب قيم مؤشر (Q3) لفحص معاملات الارتباط بين البواقي لجميع أزواج مفردات الاختبار وذلك باستخدام برنامج (jMetrik, V,4,1,1)، وقد تراوحت قيم معاملات الارتباط بين البواقي لجميع أزواج مفردات الاختبار بين(0,0016 : 0,1342) وهي قيم أقل من القيمة المطلقة للعدد (0,2) ، كما بلغ متوسط قيم مؤشر (Q3) لجميع أزواج المفردات (0,0176) وهي أقل من القيمة المتوقعة لهذا المؤشر والبالغة (-1/"ن-1"= -0,0285)، مما يشير إلى تحقق الاستقلال الموضعي.

ج- المنحنى المميز للمفردة:

يظهر تحليل مفردات الاختبار منحنى مميز خاص لكل مفردة حسب النموذج المستخدم، وتقترب فيه احتمال الاستجابة الصحيحة من الصفر عند المستويات المنخفضة من القدرة، ويزيد احتمال الاستجابة الصحيحة عند المستويات المرتفعة حتى يقترب من الواحد الصحيح، وقد تم التحقق من هذا الافتراض من خلال مدى ملاءمة مفردات الاختبار للنموذج المستخدم في التدريج، كما تم استخراج المنحنيات المميزة لمفردات الاختبار باستخدام برنامج بايلوج إم جي وفق نماذج نظرية الاستجابة للمفردة أحادية البعد ثنائية الاستجابة.

د- التحرر من السرعة:

قام الباحث بإعطاء الوقت الكافي للإجابة عن مفردات الاختبار، كما قام الباحث بالتحقق من افتراض التحرر من السرعة عن طريق حساب النسبة المئوية للطلاب الذين تمكنوا من الانتهاء من الإجابة عن جميع مفردات الاختبار، حيث بلغت النسبة المئوية للطلاب الذين تمكنوا من الانتهاء من الإجابة على جميع المفردات (100%)، مما يشير إلى تحقق افتراض التحرر من السرعة في الاختبار.

ملاءمة مفردات الاختبار لنماذج نظرية الاستجابة للمفردة:

لتحديد أي النماذج أحادية البعد ثنائية الاستجابة الأكثر ملاءمة لاستجابات العينة، قام الباحث بتدريج استجابات العينة على النماذج الثلاثة للوقوف على عدد المفردات الملاءمة لكل نموذج (كلما قل عدد المفردات غير الملاءمة كان النموذج أنسب)، واستخراج قيم جذر متوسط مربع الخطأ المعياري RMS (كلما كانت أقل كان النموذج

أنسب)، وتقدير الثبات الامبريقي (كلما كانت قيمته أكبر، كان النموذج أنسب)، وبالمقارنة بين النماذج الثلاثة (أحادي البارامتر، ثنائي البارامتر، ثلاثي البارامتر)، أوضحت نتائج التحليل أن النموذج ثنائي البارامتر هو النموذج المناسب للتحليل نتيجة لارتفاع عدد المفردات الملائمة، وانخفاض قيم RMS، وارتفاع قيم الثبات الامبريقي، لذا قام الباحث بالحكم على ملاءمة مفردات الاختبار للنموذج ثنائي البارامتر باستخدام برنامج بايلوج إم جي (BILOG-MG,3,0)، والذي يعتمد على مؤشر مربع كاي (CHI,SQ) والتي تتطلب أن تكون قيمتها غير دالة إحصائياً حتى تكون المفردة ملائمة للنموذج، والجدول التالي يوضح قيم مؤشرات الملاءمة لمفردات الاختبار:

جدول(3)

مؤشرات كا^2 لملاءمة مفردات اختبار رافن للمصفوفات المتتابعة الملونة للنموذج ثنائي البارامتر

المفردة	كا^2(CHI,SQ)	الدلالة	المفردة	كا^2(CHI,SQ)	الدلالة
1	4,7	0,320	19	5,3	0,621
2	1,2	0,876	20	11,7	0,073
3	6,5	0,165	21	5,8	0,240
4	0,7	0,880	22	10,3	0,076
5	5,2	0,393	23	9,5	0,114
6	6,6	0,254	24	10,8	0,096
7	10,1	0,086	25	1,8	0,780
8	11,2	0,079	26	2,6	0,858
9	9,9	0,129	27	3,3	0,347
10	10,7	0,097	28	3,3	0,648
11	11,2	0,130	29	9,7	0,137
12	12,1	0,096	30	5,7	0,453
13	3,3	0,507	31	7,7	0,459
14	5,1	0,280	32	5,1	0,532
15	0,7	0,947	33	11,3	0,079
16	0,5	0,798	34	9,8	0,122
17	2,8	0,731	35	12,2	0,094
18	7,6	0,267	36	11,3	0,127

يتضح من جدول (3) ملاءمة مفردات اختبار رافن للمصفوفات المتتابعة للنموذج ثنائي البارامتر، حيث جاءت قيم "كا^2" لجميع مفردات الاختبار غير دالة إحصائياً. تدريج مفردات الاختبار وفق النموذج ثنائي البارامتر:

قام الباحث بتحليل مفردات الاختبار وفق النموذج ثنائي البارامتر؛ وذلك لاستخراج تقدير بارامترات المفردات باستخدام برنامج بايلوج إم جي (BILOG-MG,3,0)، وقد جاءت نتائج التحليل كما في الجدول التالي:

جدول (4)

بارامترات الصعوبة والتمييز لمفردات اختبار رافن للمصفوفات المتتابعة الملونة وفق النموذج ثنائي البارامتر

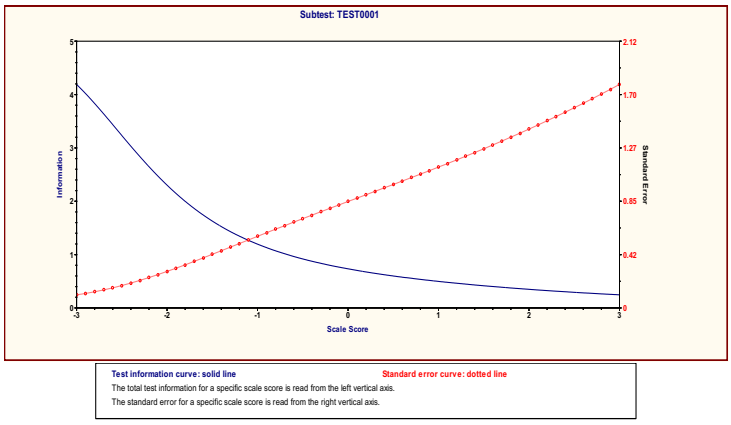
المفردة	الصعوبة (الخطأ المعياري)	التمييز (الخطأ المعياري)	المفردة	الصعوبة (الخطأ المعياري)	التمييز (الخطأ المعياري)
1	3,952- (0,872)	0,712 (0,205)	19	3,744- (0,732)	0,333 (0,070)
2	3,423- (0,620)	0,877 (0,238)	20	1,556- (0,290)	0,290 (0,053)
3	3,138- (0,470)	1,031 (0,261)	21	2,434- (0,573)	0,177 (0,40)
4	5,502- (1,716)	0,555 (0,193)	22	1,707- (0,375)	0,216 (0,045)
5	5,889- (1,662)	0,388 (0,193)	23	1,114- (0,239)	0,259 (0,049)

6	(1,087)4,705-	(0,098)0,389	24	(0,215)0,444-	(0,039)0,178
7	(1,253)5,122-	(0,081)0,312	25	(1,013)4,268-	(0,182)0,625
8	(1,602)5,961-	(0,077)0,275	26	(0,904)4,476-	(0,091)0,408
9	(0,774)3,727-	(0,060)0,276	27	(0,880)3,835-	(0,230)0,742
10	(0,696)3,338-	(0,056)0,287	28	(1,385)5,128-	(0,143)0,475
11	(1,459)5,708-	(0,033)0,132	29	(1,155)5,107-	(0,071)0,300
12	(0,244)0,796-	(0,41)0,190	30	(2,110)7,765-	(0,053)0,195
13	(0,690)3,697-	(0,200)0,782	31	(0,931)4,222-	(0,051)0,226
14	(0,866)4,219-	(0,162)0,643	32	(0,833)3,954-	(0,057)0,260
15	(0,819)3,910-	(0,188)0,691	33	(0,991)4,367-	(0,055)0,237
16	(1,928)5,872-	(0,193)0,540	34	(0,337)1,818-	(0,055)0,288
17	(1,257)5,089-	(0,092)0,348	35	(0,315)1,790-	(0,058)0,315
18	(1,065)4,733-	(0,089)0,366	36	(0,516)1,990-	(0,34)0,146

يتضح من جدول (4) أن قيم بارامترات الصعوبة تراوحت ما بين (-7,765) للمفردة رقم (30)، و(-0,444) للمفردة (24)، وقد بلغ متوسط قيم بارامتر الصعوبة لمفردات الاختبار (-3,847) بانحراف معياري (1,659) كما يتضح أيضاً من جدول (4) أن قيم بارامترات التمييز تراوحت ما بين (0,132) للمفردة رقم (11)، و(1,031) للمفردة (3)، وقد بلغ متوسط قيم بارامتر التمييز لمفردات الاختبار (0,401) بانحراف معياري (0,226).

ثبات اختبار رافن للمصفوفات المتتابعة الملونة وفق النموذج ثنائي البارامتر:

قام الباحث باستخراج قيم الثبات الامبريقي لاختبار رافن للذكاء وفق النموذج ثنائي البارامتر باستخدام برنامج بايلوج إم جي (BILOG-MG,3,0), وقد بلغت قيمة الثبات الامبريقي للاختبار (0,794)، وهي قيمة ثبات مقبولة إحصائياً، كما قام باستخراج الرسوم البيانية لدالة معلومات الاختبار، ويوضح الشكل التالي دالة معلومات الاختبار والخطأ المعياري للقياس وفق النموذج ثنائي البارامتر:



شكل (1): دالة معلومات اختبار رافن للمصفوفات المتتابعة والخطأ المعياري للقياس عبر مستويات القدرة المختلفة وفق النموذج ثنائي البارامتر

يتضح من شكل (1) أن قيمة دالة معلومات الاختبار تصل إلى أعلى قيمة لها عند مستوى القدرة المنخفضة ثم تبدأ بالانخفاض تدريجياً حتى المستويات المرتفعة من القدرة، وهذا يعني أنه باستخدام مفردات الاختبار يمكن الحصول من على معلومات أكثر دقة عن الأفراد ذوي المستوى المنخفض من القدرة، وهذا ما أشارت إليه قيم صعوبة مفردات الاختبار ومتوسط صعوبة مفردات الاختبار ككل.

نتائج البحث ومناقشتها:

يعرض الباحث في هذا الجزء نتائج فروض البحث والتعليق عليها ومناقشتها وتفسيرها في ضوء الإطار النظري والدراسات والبحوث السابقة:

نتائج الفرض الأول:

ينص الفرض الأول على" لا توجد مفردات ذات أداء تفاضلي في اختبار رافن للمصفوفات المتتابعة ترجع للنوع (ذكور، إناث)".

وللتحقق من هذا الفرض استخدم الباحث برنامج بايلوج إم جي وذلك لحساب نسبة الأرجحية في حالة تدريج مفردات الاختبار في العينة ككل وحالة التدريج في حالة المجموعتين، وحساب الفرق بين نسبة الأرجحية في الحالتين (إذا كانت قيمة الفرق بينهما دالة إحصائياً حسب توزيع كا^2؛ يُشير إلى وجود أداء تفاضلي في مفردات الاختبار) ، وقد بلغت قيمة الفرق بين نسبة الأرجحية في الحالتين (340,075) وهي قيمة دالة إحصائياً عند مستوى دلالة (0,01)، مما يشير إلى وجود مفردات في اختبار رافن للمصفوفات المتتابعة تظهر أداءً تفاضلياً بين مجموعتي (الذكور، الإناث)، وللكشف عن المفردات التي تظهر أداء تفاضلياً؛ تم استخدام مؤشر (G) نسبة الفرق في بارامتر صعوبة المفردات بين المجموعتين إلى الخطأ المعياري للفرق كما يستخرجها برنامج البالوج (إذا كانت أكبر من 1,96؛ يشير إلى أن المفردة يظهر أداءً تفاضلياً)، والجدول التالي يوضح ذلك:

جدول (5)

نتائج الأداء التفاضلي لمفردات اختبار رافن للمصفوفات المتتابعة وفق النوع (ذكور، إناث)

المفردات	الفرق في الصعوبة	الخطأ المعياري	قيمة (G) ودلالتها	المفردات	الفرق في الصعوبة	الخطأ المعياري	قيمة (G) ودلالتها
1	1,76-	1,644	1,071	19	1,139-	0,753	1,513
2	1,522-	1,611	0,945	20	0,602-	0,519	1,160
3	2,685-	1,775	1,513	21	0,872-	0,4	*2,180
4	2,856-	1,483	1,926	22	0,219-	0,392	0,559
5	0,551-	1,239	0,445	23	0,262-	0,39	0,672
6	1,375-	0,903	1,523	24	0,281-	0,385	0,730
7	0,153-	0,734	0,208	25	0,595-	0,952	0,625
8	0,681-	0,755	0,902	26	0,851-	1,192	0,714
9	0,426-	0,509	0,837	27	0,5-	1,132	0,442
10	0,522-	0,468	1,115	28	0,214-	1,388	0,154
11	0,935-	0,444	*2,106	29	0,876-	1,106	0,792
12	0,707	0,376	1,880	30	0,11-	0,719	0,153
13	0,202-	1,502	0,134	31	0,707-	0,588	1,202
14	0,874	1,421	0,615	32	0,009-	0,519	0,017
15	0,918-	1,447	0,634	33	1,269-	0,497	*2,553
16	1,718-	1,974	0,870	34	1,541-	0,455	*3,387
17	1,099-	0,972	1,131	35	0,931-	0,423	*2,201
18	2,197-	0,827	*2,657	36	0,943-	0,421	*2,240

يتضح من جدول (5) أن هناك (7) مفردات في اختبار رافن للمصفوفات المتتابعة تظهر أداء تفاضلياً طبقاً للنوع (ذكور، إناث)، حيث أظهرت المفردات أرقام (11-18-21-33-34-35-36) أداء تفاضلياً للذكور مقارنة بالإناث، حيث كانت هذا المفردات أصعب بالنسبة للإناث،

مما يشير إلى اختلاف احتمالات ظهور الاستجابة الصحيحة على هذه المفردات باختلاف النوع (ذكور، إناث)، فلا يتوقف احتمالات الاستجابة عليها على مستوى قدرة الفرد فقط بل يتأثر أيضاً بانتمائه إلى مجموعة الذكور أو الإناث.

ويفسر الباحث وجود مفردات في اختبار رافن للمصفوفات المتتابعة تُظهر أداء تفاضلياً وفق متغير النوع (ذكور، إناث)، إلى طبيعة الفروق بين الذكور والإناث في القدرات العملية والهندسية والميكانيكية مقارنة الإناث، فنتيجة تشبع اختبار رافن للمصفوفات المتتابعة بالعوامل البصرية واعتماد الاجابة على مفرداته بفهم الأشكال الهندسية وإدراك العلاقات بين مكوناتها ربما يكون من أسباب ظهور مفردات تظهر أداء تفاضلياً في اتجاه الذكور، ويدعم ذلك ما أشارت إليه دراسة (أبو مصطفي، 2010) من تفوق الذكور على الإناث في القدرات المكانية والذي يرجع إلى تأثير السيادة الجانبية والعوامل الاقتصادية والثقافية.

كما أن تفوق الذكور في القدرات البصرية مقارنة بالإناث وتفوقهن في القدرات السمعية واللفظية ربما يجعل أداء الذكور على مفردات الاختبار أفضل من الإناث، كما أن نمط التنشئة الاجتماعية للذكور والدور الاجتماعي الذي يفرضه المجتمع وعاداته وتقاليده بالنسبة للذكور يختلف عن الإناث ، حيث يتعرض أغلب الذكور إلى مثيرات بيئية واجتماعية قد تعمل على تنمية القدرات البصرية والهندسية لديه، وهذا ما يمكن ملاحظته من خلال الألعاب التي قد يتعرض لها كل من الذكور والإناث في مراحل عمرهم الأولى.

ويدعم ذلك ما أشارت إليه نتائج الأداء التفاضلي لمفردات الاختبار، حيث جاءت المفردات التي أظهرت أداءً تفاضلياً في نهاية كل قسم حيث تزداد صعوبة المصفوفات لينتهي القسم الثالث بأصعب المصفوفات وهي ما تركزت فيه أغلب المفردات التي أظهرت أداءً تفاضلياً.

وتتفق هذه النتيجة مع ما أشارت إليه نتائج دراسة كل من (Arendasy& Sommer, 2012؛ على وآخرون، 2016؛ Shibaev et al,, 2020؛ حماد، 2021)، من وجود مفردات في اختبار رافن تظهر أداءً تفاضلياً في اتجاه الذكور مقارنة بالإناث، وتختلف هذه النتيجة مع ما أشارت إليه دراسة كل من Chiesi et al., 2012 ؛ شلبي، 2015) من عدم وجود أداء تفاضلي لمفردات اختبار رافن ووجود تكافؤ قياس بين الذكور ولإناث بالنسبة لاختبار رافن للمصفوفات المتتابعة.

نتائج الفرض الثاني:

ينص الفرض الثاني على" لا توجد مفردات ذات أداء تفاضلي في اختبار رافن للمصفوفات المتتابعة ترجع للتخصص (علمي، أدبي)",

وللتحقق من هذا الفرض استخدم الباحث برنامج بايلوج إم جي وذلك لحساب نسبة الأرجحية في حالة تدريج مفردات الاختبار في العينة ككل وحالة التدريج في حالة المجموعتين، وحساب الفرق بين نسبة الأرجحية في الحالتين (إذا كانت قيمة الفرق بينهما دالة إحصائياً حسب توزيع كا^2؛ يُشير إلى وجود أداء تفاضلي في مفردات الاختبار) ، وقد بلغت قيمة الفرق بين نسبة الأرجحية في الحالتين (385,461) وهي قيمة دالة إحصائياً عند مستوى دلالة (0,01)، مما يشير إلى وجود مفردات في اختبار رافن للمصفوفات المتتابعة تظهر أداءً تفاضلياً بين مجموعتي (الذكور، الإناث)، وللكشف عن المفردات التي تظهر أداء تفاضلياً؛ تم استخدام مؤشر (G) نسبة الفرق في بارامتر صعوبة المفردات بين المجموعتين إلى الخطأ المعياري للفرق كما يستخرجها برنامج البالوج (إذا كانت أكبر من 1,96؛ يشير إلى أن المفردة يظهر أداءً تفاضلياً)، والجدول التالي يوضح ذلك:

جدول (6)

نتائج الأداء التفاضلي لمفردات اختبار رافن للمصفوفات المتتابعة وفق التخصص (علمي، أدبي)

المفردات	الفرق في الصعوبة	الخطأ المعياري	قيمة (G) ودلالتها	المفردات	الفرق في الصعوبة	الخطأ المعياري	قيمة (G) ودلالتها
1	1,683-	1,851	0,909	19	1,062-	0,674	1,576
2	1,445-	1,487	0,972	20	0,525-	0,481	1,091
3	2,608-	1,509	1,728	21	0,795-	0,289	*2,751
4	2,779-	1,372	*2,026	22	0,142-	0,375	0,379
5	0,474-	1,238	0,383	23	0,185-	0,279	0,663
6	1,298-	0,792	1,639	24	0,204-	0,274	0,745
7	0,076-	0,706	0,108	25	0,518-	0,918	0,564
8	0,604-	0,644	0,938	26	0,774-	1,081	0,716
9	0,349-	0,398	0,877	27	0,423-	1,313	0,322
10	0,445-	0,485	0,918	28	0,137-	1,277	0,107
11	0,858-	0,333	*2,577	29	0,799-	0,995	0,803
12	0,784-	0,265	*2,958	30	0,033-	0,472	0,070
13	0,125-	1,391	0,090	31	1,11-	0,477	*2,327
14	0,951	1,408	0,675	32	0,068	0,518	0,131
15	0,841-	0,943	0,892	33	1,192-	0,386	*3,088
16	1,641-	1,542	1,064	34	1,464-	0,431	*3,397
17	1,022-	0,861	1,187	35	0,854-	0,531	1,608
18	2,12-	0,716	*2,961	36	0,866-	0,354	*2,446

يتضح من جدول (6) أن هناك (9) مفردات في اختبار رافن للمصفوفات المتتابعة تظهر أداء تفاضلياً طبقاً للتخصص (علمي، أدبي)، حيث أظهرت المفردات أرقام (4-11-12-18-21-31-33-34-36) أداء تفاضلياً للتخصص العلمي مقارنة بالتخصص الأدبي، حيث كانت هذه المفردات أصعب بالنسبة للتخصص الأدبي ، مما يشير إلى اختلاف احتمالات ظهور الاستجابة الصحيحة على هذه المفردات باختلاف التخصص (علمي، أدبي)، فلا يتوقف احتمالات الاستجابة عليها على مستوى قدرة الفرد فقط بل يتأثر أيضاً بانتمائه إلى مجموعة العلمي أو الأدبي.

ويفسر الباحث وجود مفردات في اختبار رافن للمصفوفات المتتابعة تُظهر أداء تفاضلياً وفق متغير التخصص (علمي، أدبي)، إلى طبيعة المناهج الدراسية التي يتعرض لها طلاب التخصصات العلمة مقارنة بالتخصصات الأدبية والتي تلعب دوراً مهماً في تنمية قدراتهم الميكانيكية والهندسية، وإلى تزايد خبراتكم في معرفة أوضاع الأشكال وأمكان تدويرها، والذي قد يسهم بصورة كبيرة في الاستجابة على مصفوفات الاختبار حتى وإن بدت في بدايتها سهلة بالنسبة للتخصصين إلا أن المصفوفات الأكثر تعقيداً قد أظهرت أداء تفاضلياً لطلاب التخصص العلمي.

كما أن طبيعة هذا المقررات وما تحتويه من خبرات تطبيقية وعملية مقارنة بالمقررات النظرية التي يتلقاها طلاب التخصص الأدبي والتي تركز بدرجة كبيرة على التذكر والبحث عن الأسباب والنتائج والتسلسل في عرض المعلومات قد تكون أحد الأسباب التي أظهرت أداءً تفاضليا بين التخصص العلمي والأدبي، وما يدعم ذلك أيضاً أن المفردات التي أظهرت أداء تفاضلياً جاءت أغلبها في المصفوفات الأكثر صعوبة كونها

تحتاج إلى مزيد من التفكير البصري والقدرة المكانية والتعرف على الأشكال وكيفية تدويرها، ومعظمها قدرات يتعرض لها طلاب التخصص العلمي في مقرراتهم الدراسية.

وتتفق هذه النتيجة مع ما أشار إليه French et al., (2012) من أن الاختلافات في نوعية الخبرات التعليمية التي يتعرض لها الفرد في المؤسسات التعليمية قد يكون لها تأثير مختلف في نمو بعض المتغيرات المعرفية.

توصيات البحث:

في ضوء ما أسفرت عنه نتائج البحث الحالي؛ يقدم الباحث التوصيات التالية:

1- ضرورة الاهتمام باختيار اختبارات الذكاء المستخدمة على نطاق واسع في الانتقاء والتصنيف التي لا تظهر مفرداتها أداءً تفاضلياً بالنسبة للمجموعات التي تطبق عليه هذه الاختبارات من أجل تحقيق العدالة بين الأفراد المختبرين.
2- يفضل عند المقارنة بين الأفراد الذين ينتمون لخلفيات تعليمية متنوعة اختيار اختبارات لا تتأثر في الاستجابة على مفرداتها بنوعية الخبرات التعليمية التي يتعرض لها الطلاب أو العمل على تحييد هذه المفردات لدى المجموعتين، فقد أظهر الاختبار المستخدم في البحث الحالي وجود مفردات ذات أداء تفاضلي لبعض التخصصات.
3- عند استخدام اختبار رافن للمصفوفات المتتابعة الملونة للمقارنة بين أفراد مختلفين حسب النوع أوالتخصص يفضل حذف المفردات التي تظهر أداءً تفاضلياً بالنسبة لهذه المتغيرات.

دراسات وبحوث مقترحة:

يقترح البحث الحالي المشكلات والبحوث التالية:

1- أثر وجود أداء تفاضلي في مفردات اختبار ستانفورد بينيه على توزيعات نسب الذكاء.
2- الأداء التفاضلي لاختبار الذكاء المصور وفق متغير الإقامة (ريف ، حضر)
3- المقارنة بين الأداء التفاضلي للمفردة وفق أساليب النظرية الكلاسيكية ونظرية الاستجابة للمفردة.
4- الكشف عن الأداء التفاضلي لاختبارات التفكير الابتكاري لتورانس وفق التخصص (علمي، أدبي).

مراجع البحث

أولاً: المراجع العربية

أبو حطب، فؤاد؛ عثمان، سيد وصادق، آمال (2008). *التقويم النفسي*. (ط4)، القاهرة: الأنجلو المصرية.

أبو مصطفي، سهيلة (2010). *العلاقة بين القدرة المكانية والتحصيل في الرياضيات لدى طلبة الصف السادس الأساسي بمدارس وكالة الغوث*. (رسالة ماجستير غير منشورة)، كلية التربية، الجامعة الإسلامية.

التقي، أحمد (2009). *النظرية الحديثة في القياس*. عمان: دار المسيرة للنشر والتوزيع.

الحربي، عيسى (2009). *أثر تمثيل الفقرات للمحتوى ونسبة الفقرات المشتركة وطرق توزين الفقرات على دقة معادلة درجات الصور للاختبار عندما يكون التصميم المستخدم هو الجذع المشترك (الفقرات المشتركة)*. (رسالة دكتوراه غير منشورة)، كلية الدراسات العليا، الجامعة الأردنية.

الحسن، أسماء (2017). الخصائص السيكومترية لاختبار المصفوفات المتتابعة المعيارية لرافن على تلاميذ الفئية العمرية (9-12) سنة ودراسة الفروق بينهم في ضوء عدة متغيرات. *مجلة جامعة البعث*، *29*(72)، 11-49.

حماد، ديانا (2021). فحص الأداء التفاضلي المرتبط بالنوع لمصفوفات رافن المعيارية وأثر على الاستجابات المعرفية لعينة من المملكة العربية السعودية. *دراسات تربوية ونفسية، كلية التربية، جامعة الزقازيق*، 111، 1-35.

دي إيالا، آر (2017). *النظرية والتطبيق في نظرية الاستجابة للفقرة*. (ترجمة: عبدالله الكيلاني وإسماعيل البرصان)، الرياض: دار جامعة الملك سعود للنشر.(العمل الأصلي 2009).

رشـــوان، ربيع (2015). الأداء التفاضـــلي لبنود اختبار واطسون وجليسر للتفكير الناقد باستخدام النماذج البارامترية واللابارامترية في نظرية الاســتجابة للمفردة وعلاقته بـالأداء التفـاضـــلي للاختبـار بـاســتخدام التحليل العـاملي التوكيـدي متعـدد المجموعات.*مجلة كلية التربية جامعة الأزهر، 165*(2)، 455 -525.

ســـليمـان، أمين (2009). *القيـاس والتقويم في العلوم الانسانية، أسسه وأدواته وتطبيقاته*. القاهرة: دار الكتاب الحديث.

شـــلبي، ســـوســـن (2015). البنية العاملية وتكافؤ القياس لاختبار المصـــفوفات المتتابعة المعياري لرافن لدى طلاب المرحلة الإعدادية والثانوية في ضوء نموذج المعادلة البنائية. *مجلة العلوم التربوية، معهد الدراسات التربوية، جامعة القاهرة، 4*(1)، 47-117.

ضـــعضـــع، هبة (2017). *الأداء التفاضـــلي للمفردة في النظرية الكلاســـيكية ونموذج راش "دراســـة إحصـــائية مقارنة"*. (رســـالة ماجســـتير غير منشورة)، كلية التربية، جامعة حلب.

علام، صـــلاح الدين (2005). *نماذج الاســـتجابة للمفردة الاختبـاريـة أحـاديـة البعـد ومتعـددة الأبعـاد وتطبيقاتها في القياس النفسـي والتربوي*. القاهرة: دار الفكر العربي.

علي، عماد (2016). *اختبار المصـفوفات المتتابعة الملونة لـــــ "Raven" للأطفال والكبار (5,5 – 68,4 سنة)*. القاهرة: مكتبة الأنجلو المصرية.

علي، عماد؛ حســن، ســميرة وحســن، علي (2016). أثر وجود أداء تفاضـــلي في الفقرات المشـــتركة على دقة المعادلة العمودية لاختبار رافن للذكاء في ضوء نموذج فيشر للاستجابة للمفردة. *مجلة كلية التربية، جامعة أسيوط، 32*(4)، 2- 32.

عودة، أحمد؛ وعبيدات، عمر (2013). فاعلية الاختبار التكيفي المحوســـب في تقدير القدرة العقلية باستخدام مصفوفات رافن. *مجلة دراسات العلوم التربوية، الجامعة الأردنية، 40*(2)، 1602-1621.

النفيعي، عبدالرحمن (2012). الخصـــائص الســـيكومترية لاختبار المصـــفوفات المتتابعة المتقدم في ضـــوء نظرية الاســـتجابة للمفردة الاختبارية. *مجلة كلية التربية، جامعة الأزهر، 2*(147)، 175-214.

الوليلي، اســـماعيل (2002). *دراســـة ســـيكومترية مقارنة لبعض نمـاذج الاســـتجـابـة للمفردة في انتقـاء مفردات الاختبارات مرجعية المحك*. (رســـالة دكتوراه غير منشـــورة)، كلية التربية، جامعة الأزهر.

ثانياً: المراجع الأجنبية

Abu Hatab, F., Othman, S. and Sadiq, A. (2008). *Psychological evaluation*. (in Arabic), (4th ed.), Cairo: Anglo-Egyptian.

Abu Mustafa, S. (2010). *The relationship between spatial ability and achievement in mathematics among sixth grade students in UNRWA schools*. (in Arabic), (Unpublished master's thesis), College of Education, Islamic University.

Al-Harbi, I. (2009). *The effect of the items' representation of the content, the percentage of common items, and the methods of weighting items on the accuracy of the test's form score equation when the design used is anchor test (common items)*. (in Arabic), (Unpublished doctoral dissertation), College of Graduate Studies, University of Jordan.

Al-Hassan, A. (2017). Psychometric properties of the Raven's Standard Progressive Matrices test on students in the age group (9-12) years and studying the differences between them in light of several variables. (in Arabic), *Al-Baath University Journal, 29*(72), 11-49.

Ali, I. (2016). Raven Colored Progressive Matrices Test for children and adults (5.5 – 68.4 years). (in Arabic), Cairo: Anglo-Egyptian Library.

Ali, I., Hassan, S. and Hassan, A. (2016). The effect of differential performance in common items on the accuracy of the vertical equation of the Raven's intelligence test in light of Fisher's item response model. (in Arabic), *Journal of the Faculty of Education, Assiut University, 32*(4), 2- 32.

Allam, S. (2005). Item response models for unidimensional and multidimensional and their applications in psychological and educational measurement. (in Arabic), Cairo: Dar Al-Fikr Al-Arabi.

Al-Nafiey, A. (2012). Psychometric properties of the advanced progressive matrices test in light of the test item response theory. (in Arabic), *Journal of the Faculty of Education, Al-Azhar University, 2*(147), 175-214.

Al-Taqi, A. (2009). Modern theory of measurement. (in Arabic), Amman: Dar Al Masirah for Publishing and Distribution.

Arendasy, M., & Sommer, M. (2012). Gender differences in figural Matrices: the moderating role of item design. *Intelligence, 40*(6), 584- 597.

Auda, A. and Obaidat, O. (2013). The effectiveness of computerized adaptive testing in estimating mental ability using Raven's matrices. (in Arabic), *Journal of Educational Science Studies, University of Jordan, 40*(2), 1602-1621.

Avcu, A. (2021). Item response theory-based psychometric investigation of SWLS for university students. *International Journal of Psychology and Educational Studies, 8*(2), 27-37.

Hammad, D. (2021). Examining the gender-related differential performance of Raven's standard matrices and its effect on cognitive responses in a sample from the Kingdom of Saudi Arabia. (in Arabic), *Educational and psychological studies, Faculty of Education, Zagazig University*, 111, 1-35.

Holmefur, M., Sundberg, K., Wettergren, L. & Langius-Eklof, A. (2015). Measurement properties of the 13-item sense of coherence scale using Rasch analysis. *Quality of Life Research, 24(6),* 1455-1463.

Ismat, S., & Sidiqui, S. (2015). A study of intelligence measure using Raven standard progressive matrices test item by principal components analysis. *FUUAST Journal of Biology, 5*(1), 169-173.

Kim, J. & Oshima, T. (2012). Effect of Multiple Testing Adjustment in Differential Item Functioning Detection. *Educational and Psychological Measurement, 73*(3), 458-470.

Krause, J. (2012). *Assessment of item parameter drift of known items in a University Placement Exam*. (Unpublished Master's dissertation), Arizona State University.

Lai, J.; Teresi, J. & Gershon, R. (2005). Procedures for the Analysis of Differential Item Functioning (DIF) For Small Sample Sizes. *Evaluation & the Health Professions, 28*(3), 283-294.

Magno, C. (2009). Demonstrating the Difference between Classical Test Theory and Item Response Theory Using Derived Test Data, *The International Journal of Educational and Psychological Assessment, 1*(1), 1-11.

Mayer, M. & Zhu, S. (2013). Fair and equitable measurement of student learning in MOOCs: An introduction to item response theory, scale linking, and score equating. *Research& Practice in Assessment. 8*, 26-39.

Ozdemir, B. (2015). A Comparison of IRT-Based Methods for Examining Differential Item Functioning in TIMSS 2011 Mathematics Subtest. *Procedia- Social and Behavioral Sciences*, *174*, 2075 - 2083.

Qiu, C., Hatton, R., & Hou, M., (2020). Variations in Raven Progressive matrices scores among Chinese children and adolescents. *Personality and Individual Differences, 164*, 1-16.

Rashwan, R. (2015). Differential performance on items in the Watson and Glaser test of critical thinking using parametric and nonparametric models in item response theory and its relationship to differential performance of the test using multi-group confirmatory factor analysis. (in Arabic),

Bechger, T. & Maris, G. (2015). A Statistical Test for Differential Item Pair Functioning. *Psychometrika, 80* (2), 317-340.

Bortolotti, S., Tezza, R., de Andrade, D., Bornia, A. & de Sousa Junior, A.(2013). Relevance and advantages of using the item response theory. *Qual Quant, 47*, 2341-2360.

Boyd, A. (2003). *Strategies for controlling testlet exposure rates in computerized adaptive testing systems* [Doctoral dissertation, University of Texas at Austin, United States of America]. ProQuest Dissertations and Theses Global.

Chiesi, F., Ciancaeoni, M., Galli, S., Morsanyi, K., & Primi, C. (2012). Item response theory analysis and differential item functioning across age, gender and country of a short form of advanced progressive matrices. *Learning and Individual Differences, 22*(3), 390- 396.

Dawber, T. (2004). *Robustness of lord formulas for item difficulty and discrimination conversions between classical and item response theory models*. (Unpublished doctoral dissertation), University of Alberta.

De Ayala, R. (2017). Theory and practice of item response theory. (in Arabic), (Translated by: Abdullah Al-Kilani and Ismail Al-Borsan), Riyadh: King Saud University Publishing Dar. (Original work 2009).

Dodoh, H. (2017). *Differential performance of the item in classical theory and the Rasch model, "a comparative statistical study*.". (in Arabic),(Unpublished master's thesis), Faculty of Education, University of Aleppo.

Domino, G. & Domino, M. (2006). *Psychological testing: An introduction*. (2nd ed.), New York: Library of Congress.

El Welily, I. (2002). *A comparative psychometric study of some item response models in selecting item for criterion-referenced tests*. (in Arabic), (Unpublished doctoral dissertation), Faculty of Education, Al-Azhar University.

Embretson, S., & Reise, S. (2000). *Item Response theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates.

French, B.; Hand, B.; Therrien, W. & Vazquez, J. (2012). Detection of Sex Differential Item Functioning in the Cornell Critical Thinking Test. *European Journal of Psychological Assessment, 28*(3), 201-207.

Georgiev, N. (2008): Item analysis of C, D and E Series from Raven′s standard progressive matrices with item response theory two parameter logistic model. *Europe's Journal of Psychology, 8*, 1-17.

Halpern, G. (2012). *A Naval history of World War 1*. New York: Naval Institute Press.

applications. (in Arabic), Cairo: Dar Al-Kitab Al-Hadith.

Thissen, D. (2001). *IRTLRDIF v.2.0b: Software for the Computation of the Statistics Involved in Item Response Theory Likelihood-Ratio Tests for Differential Item Functioning*. http://www.unc.edu/~dthissen/dl.html.

Van der Elst, W., Ouwehand, c., Van, P., Lee, N., Van, M., & Jolles, J. (2013). The shorting raven standard progressive matrices: Item response theory- based Psychometric analysis and normative data. *Assessment, 20*(1), 48- 59.

Vincent, G. (2002). *Investigating the legitimacy of adolescent psychopathy assessments: Contributions of Item Response Theory*. (Unpublished doctoral dissertation), Simon Fraser University.

Walker, C. (2011). What's the DIF? Why Differential Item Functioning Analyses Are an Important Part of Instrument Development and Validation. *Journal of Psycho educational Assessment, 29*(4), 364-376.

Wolf, R. (2013). *Assessing the impact of characteristics of the test, common-items, and examinees on the preservation of equity properties in mixed-format test equating*. (Unpublished doctoral dissertation), University of Pittsburg.

Woods, C. (2008). Likelihood- Ratio DIF Testing: Effects of Nonnormality. *Applied Psychological Measurement, 32*(7), 511-526.

Journal of the Faculty of Education, Al-Azhar University, 165(2), 455- 525.

Reise, S.& Waller, N. (2003). How many IRT parameters does it take to model psychopathology items? *Psychological Methods, 8*(2), 164- 184.

Savage-McMlynn, E. (2012). Sex differences in intelligence in younger and older participants of the Raven standard Progressive Matrices Plus. *Personality and Individual Differences, 53*(2). 137- 141.

Shalabi, S. (2015). The factorial structure and measurement equivalence of the standard Raven's Progressive Matrices test among middle and high school students in light of the structural equation model. (in Arabic), *Journal of Educational Sciences, Institute of Educational Studies, Cairo University, 4*(1), 47-117.

Shibaev, V., Grigoriev, A., Valueva, E., Karlin, A. (2020). Differential item functioning on raven SPM+ amongst two convenience samples of Yakuts and Russians. *Psych, 2*, 44-51.

Shyu, C. (2001). *Estimating error indexes in estimating proficiencies and constructing confidence interval in item response Theory*. (Unpublished doctoral dissertation), University of Lowa.

Stoneberg, D. (2004). *A study of gender-based and ethnic-based differential item functioning (DIF) in the spring 2003 Idaho standards achievement tests applying test*. (Unpublished doctoral dissertation), University of Maryland.

Suleiman, A. (2009). Measurement and evaluation in the human sciences, its foundations, tools, and